

Interacting Copies of Random Constraint Satisfaction Problems

Maria Chiara Angelini,^{1,2} Louise Budzynski,^{3,1} and Federico Ricci-Tersenghi^{1,2}

¹*Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, Rome 00185, Italy*

²*CNR-Nanotec, Rome unit and INFN, sezione di Roma1, Piazzale Aldo Moro 5, Rome 00185, Italy*

³*DI ENS, École Normale Supérieure, PSL, CNRS, INRIA*

We study a system of $y = 2$ coupled copies of a well-known constraint satisfaction problem (random hypergraph bicoloring) to examine how the ferromagnetic coupling between the copies affects the properties of the solution space. We solve the replicated model by applying the cavity method to the supervariables taking 2^y values. Our results show that a coupling of strength γ between the copies decreases the clustering threshold $\alpha_d(\gamma)$, at which typical solutions shatters into disconnected components, therefore preventing numerical methods such as Monte Carlo Markov Chains from reaching equilibrium in polynomial time. This result needs to be reconciled with the observation that, in models with coupled copies, denser regions of the solution space should be more accessible. Additionally, we observe a change in the nature of the clustering phase transition, from discontinuous to continuous, in a wide γ range. We investigate how the coupling affects the behavior of the Belief Propagation (BP) algorithm on finite-size instances and find that BP convergence is significantly impacted by the continuous transition. These results highlight the importance of better understanding algorithmic performance at the clustering transition, and call for a further exploration into the optimal use of re-weighting strategies designed to enhance algorithmic performances.

I. INTRODUCTION

Combinatorial optimization problems are widespread in real life as well as in many scientific disciplines: from physics, in the computation of ground-state configurations, to statistical inference with likelihood maximization, and in many areas of computer science. Among the many types of combinatorial optimization problems, random Constraint Satisfaction Problems (CSPs) stand out as ideal prototypical problems for studying the average-case hardness of algorithms. Prominent examples of CSPs include the q -coloring problem on graphs, and the Boolean satisfiability problem. In an instance of CSP, a set of N variables is subjected to M constraints, and the decision version of this problem consists in finding an assignment to the variables satisfying all constraints simultaneously.

The average-case hardness of a CSP can be analyzed by introducing random ensembles of instances. Thanks to a formal analogy between CSPs and spin-glasses, the application of methods coming from statistical physics of disordered systems, such as the replica and the cavity method, has led to a detailed description of the solution space of random instances [1–6]. Many of these predictions were later proven rigorously [7–10]. In this context, a particularly interesting regime is the large size (or thermodynamic) limit, where both the number of constraints M and the number of variables N are sent to ∞ , at a fixed ratio $\alpha = M/N$. Random CSPs exhibit threshold phenomena (or phase transitions) in this limit, as α increases. The most prominent of these phase transitions is the SAT/UNSAT one, at α_{sat} , above which no solutions exists with high probability. In the satisfiable phase $\alpha < \alpha_{\text{sat}}$, many other phase

transitions occur, affecting the geometrical structure of the solution set.

One could hope that this detailed description can shed light on the average-case algorithmic hardness of CSPs, helping understanding the behavior of algorithms in the satisfiable phase. However, many of these phase transitions affect the equilibrium properties of the solution-set, making difficult the connection with algorithms – working mostly in the out-of-equilibrium regime (either because they do not satisfy detailed balance, or because they are run on time scales shorter than their relaxation time).

Recently, the line of work on the Overlap Gap Property [11–15] could rule out a large class of *stable* algorithms, for problems exhibiting a strong form of topological discontinuity in the set of distances between near optimal solutions. This approach has the advantage of describing the properties of all solutions, and not only the typical ones (dominating the uniform measure over the solution-set). Hence, it does not suffer from the aforementioned discrepancy between the geometrical structure of *equilibrium* and *out-of-equilibrium* solutions. On the other hand, widely used algorithms such as Simulated Annealing (SA) [16, 17], and Belief-Propagation guided decimation [18, 19] might fall out of the class of *stable* algorithms when used in the regime such that the number of iterations scale quadratically in the system size [20].

In this paper, we pay particular attention to the clustering (or dynamic) threshold α_d , above which the set of solutions splits into a large number of distinct groups of solutions, called clusters, which are internally well connected but well separated from each other. This transition is also manifested by the appearance of a specific form of long-range correlations between variables,

called point-to-set correlations, in the probability distribution defined as the uniform measure over the set of solutions. These correlations prohibit the rapid equilibration of stochastic processes that satisfy the detailed balance condition [21], which justifies the alternative name “dynamic” for the clustering transition. The clustering threshold α_d hence gives a lower bound to the algorithmic threshold α_{alg} above which no algorithm can find solutions to a typical instance, given that below α_d Monte Carlo based algorithms can sample uniformly the solution-set in polynomial time down to arbitrarily small temperatures. Of course, the lower bound is generally not tight, as e.g., SA is able to find solutions non-uniformly even before reaching equilibrium [22].

Many structural phase transitions occurring in the satisfiable regime, and in particular the clustering threshold α_d , depend on correlations between variables defined over a specific probability distribution over the configuration space, namely the uniform measure across solutions. In a series of works [22–28], it was demonstrated that introducing a re-weighting of the solution set could significantly move the location of these structural phase transitions, and that this strategy could be used to improve the performance of algorithms searching for a solution to a random CSP instance.

Following this line of thought, in this paper, we study a model of coupled copies (or real replicas) of a CSP instance, where the coupling strength between copies is an external parameter that can be varied to re-weight the solution space. The model with coupled replicas has a deep connection with the local-entropy approach [24] counting the number of solutions in a given neighborhood: once marginalized over the other copies of the model, the resulting re-weighted probability distribution over a single replica favors solutions living in dense regions (i.e., having a large local entropy).

Our investigation builds upon recent works exploring the effects of coupled copies of CSPs in the planted q -coloring on graphs [17], and in the binary perceptron [29] in the teacher-student scenario, where it was shown that algorithms sampling from the system of interacting copies (such as replicated SA) outperforms the classical approach using a single copy of the CSP instance when one wants to infer the planted known solution. Another motivation for studying a system of interacting copies is that it induces long-range interactions between variables of a single copy (once marginalized over the other ones), that could be tuned to decrease the long-range point-to-set correlation between distant variables that appears in the clustered phase. It was indeed observed that extending the range of interactions between variables in the re-weighted distribution over the solution-set [22, 27] can delay the clustering threshold α_d to higher values, therefore calling for a generic strategy to extend the interaction range.

In this work, we concentrate our efforts on the bi-

coloring problem on random k -hypergraphs [30], in which the variables can take two values, and each constraint acts on a k -uplet of variables, forbidding monochromatic configurations. We introduce a model of y copies of the same instance, interacting site-by-site through a ferromagnetic coupling of strength γ , and examine how the dynamical phase transition behaves as a function of γ . Remarkably, we find that turning on the interaction between copies has the effect of *decreasing* the dynamical threshold $\alpha_d(\gamma)$, effectively shrinking the region where algorithms such as SA and message-passing algorithms can sample instances in polynomial time. In addition to the characterization of the phase diagram, we also provide a numerical study of the effect of the coupling strategy on the Belief Propagation algorithm run on finite size graph instances, confirming that in the coupled model the region of convergence of BP get reduced.

These results are surprising, and challenge prior conjectures about the benefits of introducing interacting copies of a system in optimization problems. They open the door to further investigations into the optimal use of re-weighting strategies. However, we will also discuss the possible beneficial effects of the coupling that, modifying a discontinuous transition into a continuous one, can make solution in the clustered phase easier to approximate.

While in the present work the analysis is limited to the case of two copies, $y = 2$, our study can be straightforwardly extended to larger values of y .

II. SET-UP OF THE PROBLEM

A. Definition of the model

1. The k -uniform hyper-graph bi-coloring problem

We consider in this paper the k -uniform hyper-graph bi-coloring problem [30]. An instance of this Constraint Satisfaction Problem (CSP) is defined by an hyper-graph $G = (V, E)$, with a set V of N vertices, and a set E of M hyper-edges, each hyper-edge involving a subset of k vertices (see Figure 1, top panel (a)). A set of N spin variables $\sigma_1, \dots, \sigma_N$ live on the vertices of the graph, with $\sigma_i \in \{-1, 1\}$. We denote by $\underline{\sigma} = (\sigma_1, \dots, \sigma_N)$ the global configuration of the variables, and $\underline{\sigma}_S$ the configuration of a subset $S \subseteq V$ of the vertices. A constraint (or clause) is associated to each hyper-edge. The a -th constraint is satisfied if and only if there is at least one $+1$ and one -1 among the k variables of $\underline{\sigma}_{\partial a}$, where ∂a is the set of vertices contained in an hyper-edge $a \in E$ (and similarly ∂i the set of hyper-edges adjacent to vertex $i \in V$). A configuration $\underline{\sigma}$ is a solution of the CSP if and only if it satisfies

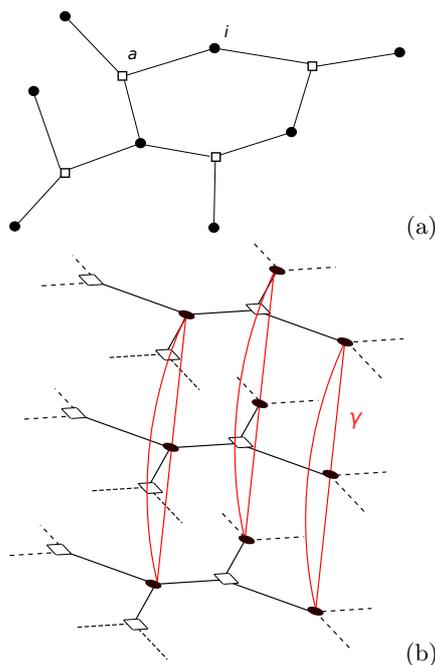


FIG. 1. Top (a): An instance of the k -uniform hyper-graph bi-coloring problem with $k = 3$, $N = 8$, $M = 4$. Vertices are represented by black circles, hyper-edges by white squares. An edge is drawn between the a -th hyper-edge and the vertex i if and only if i is linked to a ($i \in \partial a$). Bottom (b): A system of $y = 3$ interacting copies. Each pair of copies interact site-by-site with a coupling strength γ (red edges).

all the M constraints simultaneously.

A convenient way of studying the solution-set $\mathcal{S}(G)$ of a given instance of this problem is to introduce the uniform probability measure:

$$\mu(\underline{\sigma}) = \frac{1}{Z(G)} \prod_{a=1}^M \omega(\underline{\sigma}_{\partial a}) \quad (1)$$

where the normalization factor $Z(G) = |\mathcal{S}(G)|$ counts the number of solutions, and the function $\omega(\sigma_1, \dots, \sigma_k)$ is the indicator function of the event “the k variables $\sigma_1, \dots, \sigma_k$ are not all equal”.

2. Interacting copies

In this paper, we introduce y copies of a given instance $G = (V, E)$ of the bicoloring problem. The system is represented in Figure 1, bottom panel (b). Each pair of copies is interacting site-by site, with a coupling strength γ . The probability measure representing this

system is:

$$\mu_y(\underline{\sigma}^1, \dots, \underline{\sigma}^y) = \frac{1}{Z_y} \prod_{a=1}^M \left(\prod_{s=1}^y \omega(\underline{\sigma}_{\partial a}^s) \right) \prod_{i=1}^N e^{\frac{\gamma}{2y} \sum_{s \neq t} \sigma_i^s \sigma_i^t} \quad (2)$$

where the spin variables associated with the copy $s \in \{1, \dots, y\}$ are denoted $\underline{\sigma}^s = (\sigma_1^s, \dots, \sigma_N^s)$. A ferromagnetic coupling $\gamma > 0$ favors copies $\underline{\sigma}^1, \dots, \underline{\sigma}^y$ in similar configurations, while an anti-ferromagnetic coupling favors distant configurations.

The case of a single copy ($y = 1$) is retrieved from (2) in the case of independent copies (setting $\gamma = 0$), or when the coupling forces the copies to be identical (sending $\gamma \rightarrow \infty$).

Our aim is to study the properties of μ_y for typical hyper-graphs, by studying random instances. In this paper, we concentrate our efforts on Erdős Rényi (ER) random hyper-graphs. An instance of this problem is generated by drawing, independently for each hyper-edges $a \in \{1, \dots, M\}$, the set of adjacent vertices ∂a uniformly at random among the $\binom{N}{k}$ possible k -uplets. We will be interested in the large size (thermodynamic) limit, where both N and M go to infinity at a fixed ratio $\alpha = M/N$ (called the density of constraints). In this limit, ER random hyper-graphs converge locally to hyper-trees, and the degree distribution follows a Poisson law of parameter αk .

B. Belief-Propagation

In order to study the typical properties of the measure μ_y (2), we use the cavity method [31, 32], a method efficient on interacting particle models defined on random sparse structures. The first step of the cavity method amounts to study the model (2) on finite trees, where an exact description of μ_y in terms of marginals and of the free energy $\ln Z_y$, can be obtained with Belief Propagation (BP).

1. Super-spin variables

The factor graph (Fig. 1 bottom panel (b)) represents the interactions between variables in the measure μ_y . It contains small loops due to the on-site interaction term $e^{\frac{\gamma}{2y} \sum_{s \neq t} \sigma_i^s \sigma_i^t}$ (represented by the red edges on the figure) on each vertex $i \in V$ of the hyper-graph $G = (V, E)$. These small loops forbid a direct use of the cavity method – which is well-suited for tree-like problems – on this factor graph. A natural strategy to circumvent this difficulty is to define super-spin variables $X_i = (\sigma_i^1, \dots, \sigma_i^y)$ on each vertex $i \in V$. The

associated probability measure now writes:

$$\begin{aligned} \mu_y(\underline{X}) &= \frac{1}{Z_y} \prod_{a=1}^M \Omega(\underline{X}_{\partial a}) \prod_{i=1}^N \phi(X_i) \quad \text{with} \\ \Omega(\underline{X}_{\partial a}) &= \prod_{s=1}^y \omega(\underline{\sigma}_{\partial a}^s) \quad , \quad \phi(X_i) = e^{\frac{\gamma}{2y} \sum_{s \neq t} \sigma_i^s \sigma_i^t} \quad , \end{aligned} \quad (3)$$

and its associated factor graph is the original hyper-graph $G = (V, E)$.

2. Belief Propagation Equations

For each $i \in V$, $a \in \partial i$, we introduce the variable-to-factor and factor-to-variable BP messages $\eta_{i \rightarrow a}$ and $\hat{\eta}_{a \rightarrow i}$, as the marginal probability laws of X_i in the amputated graph where some interactions have been discarded: $\eta_{i \rightarrow a}$ is the marginal of X_i when a has been removed, and $\hat{\eta}_{a \rightarrow i}$ is the marginal of X_i when one removes all hyper-edges in $\partial i \setminus a$. The BP messages obey the following set of equations:

$$\begin{aligned} \eta_{i \rightarrow a}(X_i) &= \frac{\phi(X_i)}{z_{i \rightarrow a}} \prod_{b \in \partial i \setminus a} \hat{\eta}_{b \rightarrow i}(X_i) \\ \hat{\eta}_{a \rightarrow i}(X_i) &= \frac{1}{\hat{z}_{a \rightarrow i}} \sum_{\underline{X}_{\partial a \setminus i}} \Omega(\underline{X}_{\partial a}) \prod_{j \in \partial a \setminus i} \eta_{j \rightarrow a}(X_j) \end{aligned} \quad (4)$$

where $z_{i \rightarrow a}$ and $\hat{z}_{a \rightarrow i}$ are normalization factors. One can compute the marginal probability of the variable X_i from the set of incoming messages $\{\hat{\eta}_{a \rightarrow i}\}$:

$$\mu_i(X_i) = \frac{\phi(X_i)}{z_i} \prod_{a \in \partial i} \hat{\eta}_{a \rightarrow i}(X_i) \quad (5)$$

Note that the size of the super-spin variable $X_i = (\sigma_i^1, \dots, \sigma_i^y)$ grows exponentially with the number of copies, which represent a limitation for a numerical representation of the BP messages with a large number of copies. For this reason in the following we will concentrate our attention to the case $y = 2$.

C. Ensemble average with the cavity method

The BP equations (4) are exact when the hyper-graph G is an hyper-tree, and can be used heuristically on any factor graph, even in the presence of loops. The message-passing iterative algorithm searching for fixed point of these equations is called Belief Propagation algorithm. The cavity method is based on the application of Belief Propagation on random hyper-graphs which are locally tree-like in the thermodynamic limit (such as ER random hyper-graphs).

1. Replica Symmetric cavity method

There are different versions of the cavity method, that rely on self-consistent hypothesis on the effect of the long loops that are present in random graphs. The simplest version, called Replica Symmetric (RS), assumes a fast decay of the correlations between distant variables, in such a way that the probability measure (3) is correctly described by the tree-like approximation, and that the BP equations will converge toward a unique fixed-point on a typical large instance. We give in appendix A the RS equations (A1) for the interacting copies of the bi-coloring problem. These equations can be solved numerically with population dynamics [31], and provide correct predictions for the typical properties of the measure μ_y in the regime of small density of constraints $\alpha = M/N$.

2. Replica Symmetry Breaking

As the density of constraint α increases, the hypothesis underlying the RS cavity method must break down, and a more sophisticated version of the cavity method can be employed to treat the effect of long loops. The first non-trivial level is called 1RSB (for one-step Replica Symmetry Breaking), and postulates the existence of a partition of the configuration space into pure states (or clusters) such that the restriction of the measure to one cluster is accurately described within the RS formalism.

The dynamical threshold α_d separates the regime where the RS approximation is valid (for $\alpha < \alpha_d$) from a region where the 1RSB formalism is needed to correctly describe the typical properties of the measure μ_y . Technically, the dynamical threshold can be computed within the 1RSB formalism, by deriving and solving the 1RSB equations at Parisi parameter $x = 1$ (see appendix B, equation (B11)). These equations always admit a trivial fixed-point: the RS solution given in equation (B13). In the RS phase, this is the unique solution to the 1RSB equation, correctly describing the case of a single cluster. An RSB phase is unveiled by the appearance of a non-trivial solution different from the RS one to the 1RSB equation.

Depending on the situation, this RS/RSB dynamical transition can occur in a continuous or a discontinuous way. In the continuous case, the dynamical threshold α_d can be computed by studying the local instability of the RS solution (see eq. B13) under a small perturbation toward the space of 1RSB solutions. It is called the Kesten-Stigum instability in the context of tree-reconstruction [33, 34], or the Almeida-Thouless transition for mean-field spin glasses [35]. The Kesten-Stigum threshold coincides with the dynamical threshold in the

continuous case, and provides an upper bound in the discontinuous case (in which a non-trivial solution to the 1RSB equations appears while the RS solution is still stable).

In order to probe the dynamical phase transition, one can compute the difference between:

- the *intra-state* overlap q_1 (see appendix B, equations (B23)) measuring the overlap between two typical configurations sampled from *the same* cluster.
- and the *inter-state* overlap q_0 (equation (B22)), measuring the overlap between two typical configurations (in the clustered phase $\alpha > \alpha_d$, two typical configurations are likely to be in two different clusters). In the case of the bi-coloring problem, the inter-state overlap is trivially equal to 0 due to the spin-flip invariance of the measure (2)

While in the RS phase, $q_1 - q_0 = 0$ since the solution space is correctly described by a single cluster, we have $q_1 - q_0 > 0$ in the 1RSB phase (see e.g. Figure 3 the evolution of the intra-state overlap as a function of γ). If the transition is continuous, the value $q_1 - q_0$ grows continuously from 0, otherwise it displays a jump.

III. RESULTS

A. Phase diagram

In this section, we provide a detailed analysis of the effect of the coupling strength γ on the dynamical phase transition occurring at α_d . We restrict our analysis to the case of two copies ($y = 2$), and will treat the case of a larger number of copies in a future work. We also fix the number of variable-per-clause to $k = 5$, which is a representative value for the bi-coloring problem for a single copy [30] (the cases $k = 3$ and $k = 4$ being peculiar with a continuous dynamical transition, while it is discontinuous as soon as $k \geq 5$).

We provide our results in the form of a phase diagram in Fig. 2, in the (α, γ) plane. In the case $y = 2$, the probability distribution (2) satisfy the following symmetry:

$$\mu_y(\underline{\sigma}^1, \underline{\sigma}^2; G, \gamma) = \mu_y(\underline{\sigma}^1, -\underline{\sigma}^2; G, -\gamma). \quad (6)$$

In words, a ferromagnetic system ($\gamma > 0$) favoring configurations in which the two copies $\underline{\sigma}^1, \underline{\sigma}^2$ are aligned, is equivalent to the anti-ferromagnetic system ($-\gamma$) favoring the alignment of $\underline{\sigma}^1$ and $-\underline{\sigma}^2$. This implies that any average quantity computed from this distribution, and in particular the intra-state overlap q_1 (B23), is invariant under the transformation $\gamma \rightarrow -\gamma$. Therefore, the dynamical threshold is such that

$$\alpha_d(\gamma) = \alpha_d(-\gamma) \quad (7)$$

and we don't need to explore the region $\gamma < 0$. This symmetry does not hold however for a larger number of copies $y > 2$, where both $\gamma > 0$ and $\gamma < 0$ regions have to be studied.

The RS phase, painted in gray, is on the left of the dynamical threshold $\alpha_d(\gamma)$. The blue points correspond to the Kesten-Stigum threshold $\alpha_{KS}(\gamma)$, at which a non-trivial solution to the 1RSB equations emerge continuously from the trivial one upon increasing α . The orange squares mark the discontinuous appearance (at $\alpha_{disc}(\gamma)$) of a non-trivial solution to the 1RSB cavity equations, upon increasing α . Details on the numerical computation of these thresholds are given in appendix B 4. The dynamical threshold $\alpha_d(\gamma)$, defined as the appearance of a non-trivial solution to the 1RSB equations (B11), is given as:

$$\alpha_d(\gamma) = \min(\alpha_{KS}(\gamma), \alpha_{disc}(\gamma)). \quad (8)$$

Note that we recover the threshold values for a single

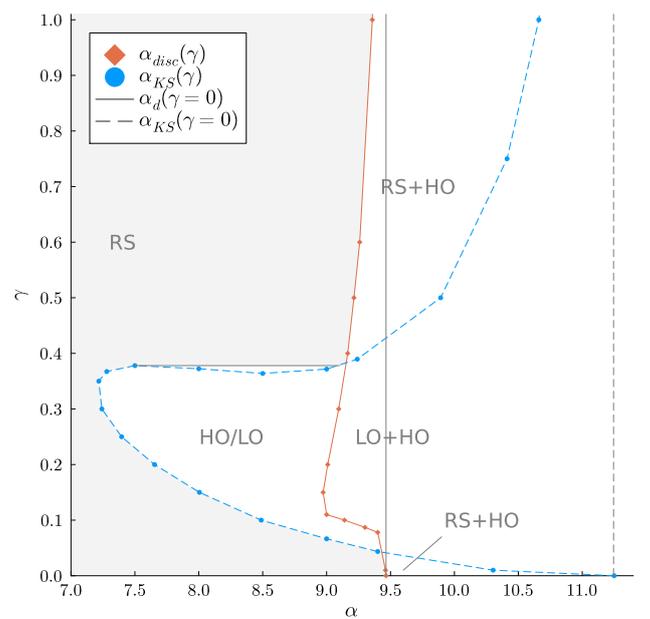


FIG. 2. Phase diagram in the plane (γ, α) for the replicated bi-coloring problem on random k -hypergraphs, with $k = 5$ and $y = 2$ copies. The RS phase is shaded in gray, the RSB phase is in white. The blue circles correspond to the Kesten-Stigum threshold $\alpha_{KS}(\gamma)$. The orange squares correspond to the discontinuous appearance of a non-trivial solution to the 1RSB equations at $\alpha_{disc}(\gamma)$. For each value of γ , the dynamical threshold $\alpha_d(\gamma) = \min\{\alpha_{KS}(\gamma), \alpha_{disc}(\gamma)\}$ separates the RS phase from the RSB phase. The vertical solid and dashed lines mark respectively the dynamical and Kesten-Stigum threshold the for the non-interacting case ($\gamma = 0$) and in the $\gamma \rightarrow \infty$ limit. We identify several RSB phases, denoted ‘LO/HO’, ‘RS+HO’, ‘LO+HO’ (see Sec. III A 1 for a precise definition).

copy ($y = 1$) at $\gamma = 0$ and in the large γ limit: $\alpha_d(y = 1) = 9.465$ and $\alpha_{KS}(y = 1) = 11.25$ (see [22, 30, 36, 37] for numerical values of the hyper-graph bi-coloring problem' thresholds).

The main observation to extract from this plot is that turning on the coupling between copies has the effect of shrinking the RS phase, for all values of $\gamma \neq 0$. Recalling that large instances can be solved in polynomial time in the RS phase [21], this is a negative result: it indicates that the strategy of introducing coupled copies of a CSP might not improve the performance of solving algorithms. This observation is corroborated in the next section III B, where we will see that the BP algorithm fails at converging on finite size instances above the Kesten-Stigum threshold $\alpha_{KS}(\alpha)$.

In addition, we observe a change in the nature of the phase transition, as the coupling strength γ is varied. For values of γ between 0.04 and 0.38, the non-trivial solution to the 1RSB equation appears continuously ($\alpha_{KS}(\gamma) < \alpha_{disc}(\gamma)$), and the dynamical transition is therefore continuous. Outside of this range, we have $\alpha_{disc}(\gamma) < \alpha_{KS}(\gamma)$, and the dynamical transition is discontinuous. In particular, we recover a discontinuous transition in the non-interacting case (at $\gamma = 0$ and in the large γ limit), as already observed in [22, 30] for $k = 5$. On finite-size instances, the effect of a continuous phase transition is more drastic and prevent BP algorithm to converge above α_d (see next section III B).

1. A detailed picture of the phase diagram

In this paragraph, we explain the different regions in the phase diagram given in Fig. 2. For some values of the parameters α, γ , there exist (at least) two different solutions to the 1RSB equations at $x = 1$ (B11). This type of behavior was already observed in [37], [22], and its consequences for inference problems (and planted CSPs) have been discussed in [38]. We use two different initial conditions (see appendix B 4) for the iterative resolution of the 1RSB equations:

- Starting from a high intra-state overlap $q_1 = 1$ and denoted 'HO' initial condition
- Starting with a low intra-state overlap $q_1 = 0.01$ and denoted 'LO' initial condition

Depending on the parameters (α, γ) , the two initial conditions can lead to two different solutions, or not. More precisely, the different phases illustrated in Fig. 2 are defined as follows:

- RS: both HO and LO initial conditions lead to the trivial RS solution ($q_1 = 0$);
- HO/LO: both HO and LO initial conditions lead to the same non-trivial solution ($q_1 > 0$);

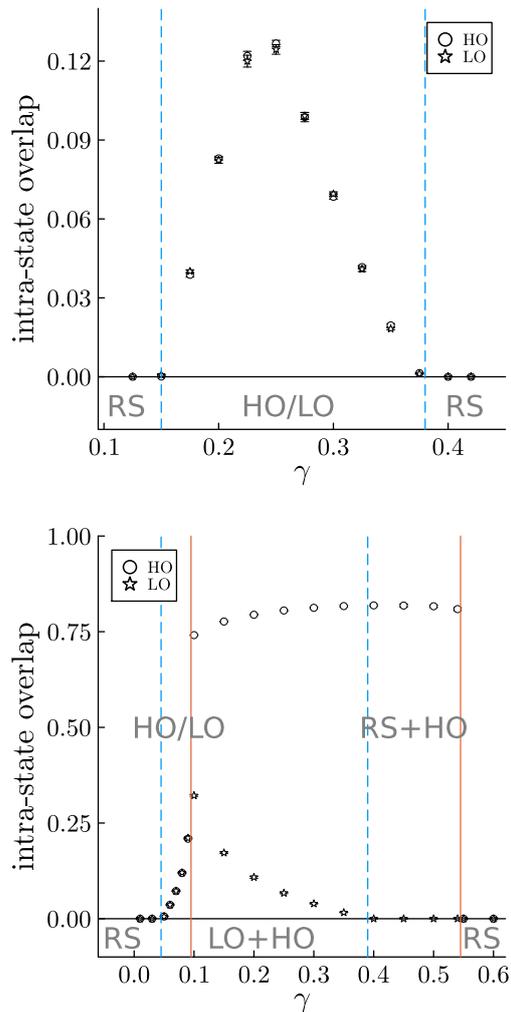


FIG. 3. Evolution of the intra-state overlap as a function of the coupling strength γ , at $y = 2, k = 5$, for $\alpha = 8.0$ (top panel) and $\alpha = 9.24$ (bottom panel). The blue dashed line marks the Kesten-Stigum threshold, and the orange solid line marks the discontinuous appearance of a non-trivial solution (see Fig. 2).

- LO+HO: LO initial condition leads to a non-trivial solution, HO initial condition leads to a different non-trivial solution with a higher overlap;
- RS+HO: LO initial condition leads to the trivial RS solution, while HO initial condition leads to a non-trivial solution.

The frontiers between these different phases are:

- The Kesten-Stigum line: $\alpha_{KS}(\gamma)$, the limit of the stability of the trivial fixed-point under a perturbation toward the space of 1RSB solutions. In the

unstable phase, the LO initial condition leads systematically to a non-trivial solution, while in the stable phase it leads to the trivial RS solution.

- The discontinuous appearance of a high-overlap (HO) non-trivial solution, at $\alpha_{\text{disc}}(\gamma)$, reached from the HO initial condition.

In order to get a better understanding of the various phases, we show in Fig. 3 the evolution of the intra-state overlap as a function of the coupling strength γ , for two values of α .

B. Results on large graph instances

In this section, we compare the results obtained in the large size limit from the cavity method in III A, with numerical results on finite size instances.

We start by comparing the BP marginals $\mu_i(\sigma_i^1, \sigma_i^2)$ found on finite size instances, with the RS solution in the large size limit. In Fig. 4 we report this comparison at $\gamma = 0.2$, for two values of α , one below the dynamical threshold $\alpha_d(\gamma = 0.2) = 7.654$ (top panel) and one above α_d (bottom panel). We compute the BP marginals in Eq. (5) for each value of $X_i = (\sigma_i^1, \sigma_i^2) \in \{(+1, +1), (+1, -1), (-1, +1), (-1, -1)\}$, and average over the vertices of the graph, and over 50 graph instances of size $N = 10^4$. The error bars reported in Fig. 4 represent the standard deviation of $\mu_i(X_i)$ over the population for the RS solution (in black) and over the instances and the vertices of the graph for the BP solutions (in orange and blue).

For any number of copies y , a simple criterion for distinguishing between a ferromagnetic (polarized) and a paramagnetic BP fixed-point is given by the following quantity

$$\begin{aligned} \text{FP} &= \frac{1}{N} \sum_{i=1}^N \text{FP}_i \quad \text{with} \\ \text{FP}_i &= \frac{1}{y} \sum_{s=1}^y (\mu_i^s(1)^2 + \mu_i^s(-1)^2), \end{aligned} \quad (9)$$

where $\mu_i^s(\sigma) = \mathbb{P}[\sigma_i^s = \sigma]$ is the single variable marginal on site i in replica s . In a paramagnetic solution, we have $\mu_i^s = 1/2$ for each vertex $i \in \{1, \dots, N\}$ and for each copy $s \in \{1, \dots, y\}$, and therefore $\text{FP} = 0.5$, while $\text{FP} > 0.5$ for a ferromagnetic solution.

At $\alpha = 7.5$, we observe two types of BP fixed-points, ferromagnetic (in blue) and paramagnetic (in orange). At $\alpha = 7.7$, a ferromagnetic BP fixed-point is always found on all instances. Furthermore, there is a good compatibility between the *paramagnetic* BP marginal found on finite size-instances at $\alpha = 7.5$ (Fig. 4, top panel, in orange), and the RS prediction in the large size

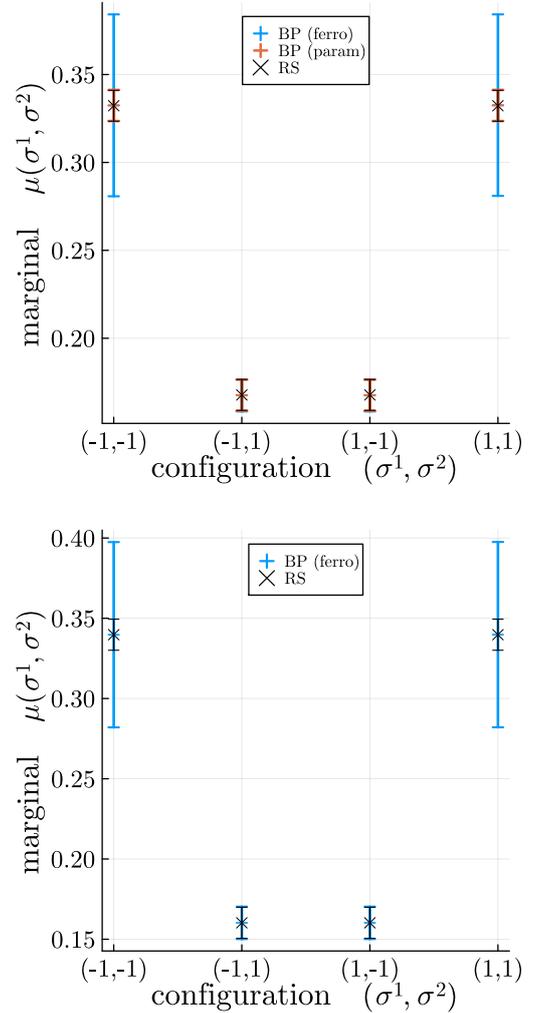


FIG. 4. Comparison of the BP marginal averaged over the vertices and over 50 instances of size $N = 10^4$, with BP marginals predicted by the RS formalism. Here $y = 2$, $k = 5$, $\gamma = 0.2$. Top panel: for $\alpha = 7.5$, a *paramagnetic* BP solution is found on a large fraction (0.84) of instances, and the corresponding paramagnetic BP marginals (in orange) are comparable with the RS prediction (in black). A *ferromagnetic* BP solution is found only on a small fraction (0.16) of instances, and their BP marginals (in blue) are quantitatively different from RS. Bottom panel: for $\alpha = 7.7$, BP marginals are ferromagnetic for all instances (in blue), and quantitatively different from the RS predictions (in black)

limit (in black). The small error on the BP marginals is a finite size effect that should go to zero for larger graphs.

Instead, the *ferromagnetic* BP marginals are quantitatively different from the RS prediction. Although, for each value of $X \in \{(+1, +1), (+1, -1), (-1, +1), (-1, -1)\}$ the aver-

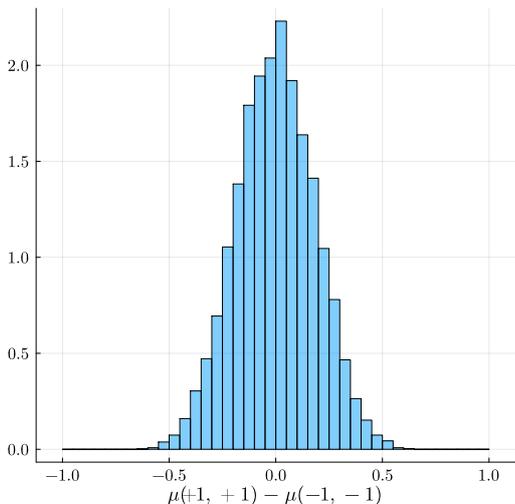


FIG. 5. Histogram of the quantity $\mu_i(+1, +1) - \mu_i(-1, -1)$ for each site of a given instance of size $N = 10^4$. Here $y = 2$, $k = 5$, $\gamma = 0.2$, $\alpha = 7.7$.

age value of $\mu(X)$ is remarkably close to the RS prediction, the fluctuations of $\mu(1, 1)$ and $\mu(-1, -1)$ over the graph sites and over the instances are much larger. In practice the population representing the RS solution is much more homogeneous than the BP marginals on given large graphs.

Given that the largest fluctuations are in values of $\mu(1, 1)$ and $\mu(-1, -1)$, and are highly anti-correlated (i.e., the sum of the two values remains almost constant, close to 0.68), we report in Fig. 5 the histogram of the difference $\mu(1, 1) - \mu(-1, -1)$. We see such a difference span the whole possible range, and gives evidence that on a given large graph the BP solution predicts the presence of strongly polarized vertices, that is vertices whose marginal probability is concentrated on a single value $(1, 1)$ or $(-1, -1)$.

The presence of these polarized vertices is a clear sign of an RSB phase transition, as already found for spin glass models at low temperatures [39]. Indeed, due to the global spin-flip invariance of the probability measure (3), we have $\mu_y(\underline{X}) = \mu_y(-\underline{X})$, and also the marginal probability on any vertex $i \in V$ is also invariant under spin-flip, $\mu_i(X_i) = \mu_i(-X_i)$. As a consequence, on each vertex i , the marginal probability μ_i is not polarized, and one should always observe a paramagnetic solution with $\text{FP} = 0.5$.

Therefore, a *ferromagnetic* BP fixed-point with $\text{FP} > 0.5$ is not describing correctly the full probability distribution (3). This ferromagnetic solution rather describes a restricted part of the probability distribution, that is a *cluster*, in which some vertices are polarized toward one specific value (while all vertices should be not polarized once averaged over the clusters). Finding a ferromag-

netic BP fixed-point on some instances therefore could be a signal of the RSB phase transition occurring at the dynamical threshold $\alpha_d = 7.654$, and the appearance of glassy states even below this value could be understood as a finite-size effect (on some graph instances, a polarized solution might naturally appear).

We want to stress that standard BP with a single replica never finds ferromagnetic fixed points, neither in the clustered phase, nor in the paramagnetic phase. These ferromagnetic FPs are only found by the replicated BP: it could be possible that the replicated measure enlarges the basin of attraction of the glassy states with larger entropy that become thus accessible for the replicated BP.

Fig. 6 illustrates the appearance of the ferromagnetic solution as α approaches the dynamical threshold $\alpha_d(\gamma)$, for two values of γ :

- (a) $\gamma = 0.2$: at which the dynamical phase transition is continuous (see phase diagram Fig. 2 and explanations in Sec. III A).
- (b) $\gamma = 0.01$: at which the dynamical phase transition is discontinuous.

We considered two graph sizes ($N = 10^4, 10^5$), and run BP on 50 random graphs for each size. We show in Fig. 6 the fraction of graphs for which BP iterations converged towards a fixed-point (squares), and among the converged runs, the fraction of ferromagnetic BP fixed-points (circles). In general, at the BP algorithmic transition, the fraction of converged BP runs drastically decreases and the few converged runs achieve a ferromagnetic fixed point.

In the case of a continuous dynamical transition at $\gamma = 0.2$ (top panel), we observe the BP algorithmic transition to take place slightly below the dynamical threshold $\alpha_d(\gamma = 0.2) = \alpha_{\text{KS}}(\gamma = 0.2) = 7.65$, the small difference being possibly a finite size effect.

In the case of a discontinuous transition at $\gamma = 0.01$, the dynamical threshold coincides with the discontinuous appearance of a non-trivial RSB solution at $\alpha_d(\gamma = 0.01) = \alpha_{\text{disc}}(\gamma = 0.01) = 9.46$, but it does not affect the behavior of BP. The BP algorithmic behavior is again ruled by the Kesten-Stigum threshold: only when the paramagnetic fixed point becomes locally unstable, BP fails to converge or converges to a ferromagnetic fixed point.

We recall that the Kesten-Stigum threshold decreases as soon as the coupling between copies is turned on with $\gamma \neq 0$ (see the phase diagram in Fig. 2, blue dashed line). Introducing a coupling between copies has therefore the effect of shrinking the range of α for which BP can converge on finite size instances.

The above results may be interpreted as an indication that the coupling strategy can not increase the performance of message-passing algorithms in solving an in-

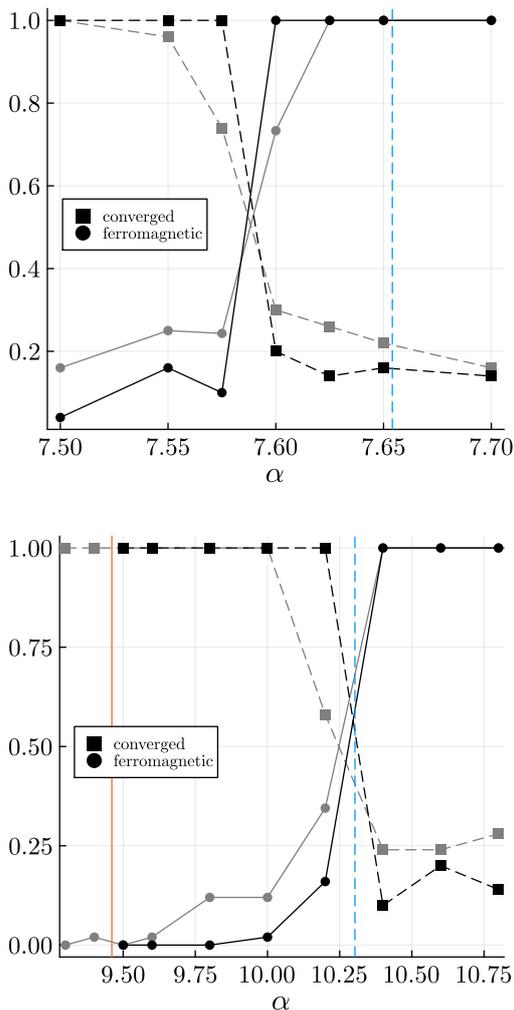


FIG. 6. Solutions found by BP on finite-size instances, for $y = 2$, $k = 5$, at $\gamma = 0.2$ (top panel) and $\gamma = 0.01$ (bottom panel), for two graph sizes: $N = 10^4$ (gray points), and $N = 10^5$ (black points). We display the fraction of instances, over 50 runs, for which BP converged (squares), and, among converged instances, the fraction of ferromagnetic BP fixed-points (circles). The vertical lines correspond to the thresholds computed in Sec. III A: $\alpha_{KS}(\gamma)$ (blue dashed line), $\alpha_{disc}(\gamma)$ (orange solid line). The last threshold is not displayed on the top panel, as it occurs at larger values $\alpha_{disc}(\gamma = 0.2) = 9.01$.

stance of CSP. However, we have to remind that the BP algorithm by itself does not output a solution, but only a set of marginal probability distributions (5). In order to obtain a solution to a given instance of CSP, the BP algorithm should be followed by a decimation procedure. An analysis of the convergence of BP during the decimation procedure can be made following the steps of [18, 19], and we plan to do it in a future work.

IV. DISCUSSION AND PERSPECTIVES

We have introduced a model of y interacting copies of a constraint satisfaction problem, the bi-coloring problem on k -hyper-graphs, and studied the effect of the coupling strength between copies on the dynamical phase transition. We focused on the case of two copies ($y = 2$), and leave for future work the analysis of this model for a larger (but finite) number of copies, as well as an analytical study in the large y limit.

We showed that turning on the coupling strength has the effect of shrinking the Replica Symmetric phase, where algorithms exist that can sample solutions in polynomial time. This result is quite surprising, because it was conjectured that a re-weighting strategy favoring solutions living in dense clusters (i.e., with a large local entropy [24]) should enhance algorithmic performances [25]. This conjecture was further confirmed by numerical investigations mainly in inference problems, with a planted solution [17, 29]. However, the conjecture was not tested at large in optimization problems, as in the present case. Only in Ref. [40] it was shown that for the largest independent set optimization problem, introducing coupled replicas does not improve standard algorithms such as Simulated Annealing.

An interesting feature that we have observed is the change of the nature of the clustering phase transition, from discontinuous to continuous (in the range $\gamma \in [0.04, 0.38]$). This may have important algorithmic consequences. Indeed, in a continuous phase transition, we expect that approximating the RSB solution slightly beyond the phase transition is easier than finding the RSB solution that appears in a discontinuous phase transition. It would be very interesting to check whether a Monte Carlo based algorithm, like SA, could work around the continuous phase transition, while we expect it to fail dramatically at a discontinuous phase transition.

We also studied the effect of the re-weighting strategy induced by the coupling on the behavior of the BP algorithm on finite-size instances, and find that the convergence of BP is strongly affected by the continuous transition occurring at the Kesten-Stigum threshold. These findings call for a better understanding of the behavior of algorithms at the clustering transition. A promising direction could be to perform an analysis of the BP plus the decimation procedure, following the formalism developed in [18, 19], adapting it to the case of y interacting copies.

Another interesting direction will be to consider the associated planted problem (or inference problem), as it was shown [17] that an algorithmic strategy based on replicated interacting copies of the problem (replicated SA) could outperform classical SA and reach the Bayes optimal threshold.

In addition, we also plan to extend our results when finite temperature is added to the problem. This should be particularly useful to explain the behaviour of algorithms, such as Simulated Annealing, that do not work directly at zero temperature.

ACKNOWLEDGMENTS

We thank Guilhem Semerjian for insightful discussions. The research has received financial support from the “National Centre for HPC, Big Data and Quantum Computing - HPC”, Project CN_00000013, CUP B83C22002940006, NRP Mission 4 Component 2 Investment 1.5, Funded by the European Union - NextGenerationEU, and from the 2021 FIS (Fondo Italiano per la Scienza) funding scheme (FIS783 - SMaC - Statistical Mechanics and Complexity) and the PRIN funding scheme (2022LMHTET - Complexity, disorder and fluctuations: spin glass physics and beyond), both from the Italian MUR (Ministry of University and Research).

-
- [1] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky. 2+p-SAT: Relation of typical-case complexity to the nature of the phase transition. *Random Structures and Algorithms*, **15**, 414 (1999).
- [2] G. Biroli, R. Monasson, and M. Weigt. A variational description of the ground state structure in random satisfiability problems. *Eur. Phys. J. B*, **14**, 551 (2000).
- [3] M. Mézard, G. Parisi, and R. Zecchina. Analytic and Algorithmic Solution of Random Satisfiability Problems. *Science*, **297**, 812–815 (2002).
- [4] S. Mertens, M. Mézard, and R. Zecchina. Threshold values of random K-SAT from the cavity method. *Random Struct. Algorithms*, **28**(3), 340–373 (2006).
- [5] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborova. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, **104**(25), 10318–10323 (2007).
- [6] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, 2009.
- [7] D. Achlioptas and F. Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *Proc. of 38th STOC*, pages 130–139, New York, NY, USA, 2006. ACM.
- [8] D. Achlioptas and A. Coja-Oghlan. Algorithmic Barriers from Phase Transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802, 2008.
- [9] M. Molloy. The freezing threshold for k-colourings of a random graph. In *Proceedings of the 44th symposium on Theory of Computing*, page 921. ACM, 2012.
- [10] J. Ding, A. Sly, and N. Sun. Proof of the Satisfiability Conjecture for Large K. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC ’15, pages 59–68, 2015.
- [11] D. Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, **118**(41), e2108492118 (2021).
- [12] D. Gamarnik and M. Sudan. Performance of Sequential Local Algorithms for the Random NAE-K-SAT Problem. *SIAM Journal on Computing*, **46**(2), 590–619 (2017).
- [13] A. Coja-Oghlan, A. Haqshenas, and S. Hetterich. Walksat Stalls Well Below Satisfiability. *SIAM Journal on Discrete Mathematics*, **31**(2), 1160–1173 (2017).
- [14] G. Bresler and B. Huang. The Algorithmic Phase Transition of Random k-SAT for Low Degree Polynomials. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 298–309, 2022.
- [15] D. Gamarnik. Turing in the shadows of Nobel and Abel: an algorithmic story behind two recent prizes. *arXiv preprint arXiv:2501.15312*, (2025).
- [16] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, **220**, 671–680 (1983).
- [17] M. C. Angelini and F. Ricci-Tersenghi. Limits and Performances of Algorithms Based on Simulated Annealing in Solving Sparse Hard Inference Problems. *Phys. Rev. X*, **13**, 021011 (2023).
- [18] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Solving Constraint Satisfaction Problems through Belief Propagation-guided decimation. *45th Annual Allerton Conference on Communication, Control, and Computing 2007*, **1** (2007).
- [19] F. Ricci-Tersenghi and G. Semerjian. On the cavity method for decimated random constraint satisfaction problems and the analysis of belief propagation guided decimation algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, **2009**(09), P09001 (2009).
- [20] M. Angelini, M. Avila-González, F. D’Amico, D. Machado, R. Mulet, and F. Ricci-Tersenghi. Algorithmic thresholds in combinatorial optimization depend on the time scaling. *arXiv preprint arXiv:2504.11174*, (2025).
- [21] A. Montanari and G. Semerjian. Rigorous Inequalities Between Length and Time Scales in Glassy Systems. *Journal of Statistical Physics*, **125**(1), 23–54 (2006).

- [22] L. Budzynski, F. Ricci-Tersenghi, and G. Semerjian. Biased landscapes for random constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, **2019**(2), 023302 (2019).
- [23] A. Braunstein, L. Dall’Asta, G. Semerjian, and L. Zdeborová. The large deviations of the whitening process in random constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, **2016**(5), 053401 (2016).
- [24] C. Baldassi, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina. Local entropy as a measure for sampling solutions in constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, **2016**(2), 023301 (2016).
- [25] C. Baldassi, C. Borgs, J. T. Chayes, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina. Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes. *Proceedings of the National Academy of Sciences*, **113**(48), E7655–E7662 (2016).
- [26] T. Maimbourg, M. Sellitto, G. Semerjian, and F. Zamponi. Generating dense packings of hard spheres by soft interaction design. *SciPost Phys.*, **4**, 39 (2018).
- [27] L. Budzynski and G. Semerjian. Biased measures for random constraint satisfaction problems: larger interaction range and asymptotic expansion. *Journal of Statistical Mechanics: Theory and Experiment*, **2020**(10), 103406 (2020).
- [28] H. Zhao and H.-J. Zhou. Maximally flexible solutions of a random K -satisfiability formula. *Phys. Rev. E*, **102**, 012301 (2020).
- [29] G. Catania, A. Decelle, and B. Seoane. Copycat perceptron: Smashing barriers through collective learning. *Phys. Rev. E*, **109**, 065313 (2024).
- [30] T. Castellani, V. Napolano, F. Ricci-Tersenghi, and R. Zecchina. Bicolouring random hypergraphs. *Journal of Physics A: Mathematical and General*, **36**(43), 11037 (2003).
- [31] M. Mézard and G. Parisi. The Bethe lattice spin glass revisited. *Eur. Phys. J. B*, **20**, 217 (2001).
- [32] M. Mézard and G. Parisi. The cavity method at zero temperature. *J. Stat. Phys.*, **111**, 1–34 (2003).
- [33] H. Kesten and B. P. Stigum. Additional Limit Theorems for Indecomposable Multidimensional Galton-Watson Processes. *The Annals of Mathematical Statistics*, **37**(6), 1463 – 1481 (1966).
- [34] E. Mossel and Y. Peres. Information flow on trees. *The Annals of Applied Probability*, **13**(3), 817 – 844 (2003).
- [35] J. R. L. de Almeida and D. J. Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, **11**(5), 983 (1978).
- [36] L. Dall’Asta, A. Ramezanpour, and R. Zecchina. Entropy landscape and non-Gibbs solutions in constraint satisfaction problems. *Phys. Rev. E*, **77**, 031118 (2008).
- [37] M. Gabrié, V. Dani, G. Semerjian, and L. Zdeborová. Phase transitions in the q -coloring of random hypergraphs. *Journal of Physics A: Mathematical and Theoretical*, **50**(50), 505002 (2017).
- [38] F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Typology of phase transitions in Bayesian inference problems. *Phys. Rev. E*, **99**, 042109 (2019).
- [39] G. Perrupato, M. C. Angelini, G. Parisi, F. Ricci-Tersenghi, and T. Rizzo. Ising spin glass on random graphs at zero temperature: Not all spins are glassy in the glassy phase. *Physical Review B*, **106**(17), 174202 (2022).
- [40] M. C. Angelini and F. Ricci-Tersenghi. Monte Carlo algorithms are very effective in finding the largest independent set in sparse random graphs. *Physical Review E*, **100**(1), 013302 (2019).
- [41] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Clusters of solutions and replica symmetry breaking in random k -satisfiability. *Journal of Statistical Mechanics: Theory and Experiment*, **2008**(04), P04004 (2008).

Appendix A: Cavity method under the Replica Symmetric ansatz

In the RS formalism, one assumes that the effect of long loops in a sample of the hyper-graph ensemble is negligible, and that the BP equations (4) admit a unique fixed-point that correctly describes the probability distribution (3). We consider a uniformly chosen directed edge $i \rightarrow a$ on a random hyper-graph G , and let $\mathcal{P}^{\text{RS}}(\eta)$ be the probability distribution of the fixed-point variable-to-factor BP message $\eta_{i \rightarrow a}$ thus obtained. Similarly, we define the distribution $\widehat{\mathcal{P}}^{\text{RS}}(\widehat{\eta})$ of the factor-to-variable BP message $\widehat{\eta}_{a \rightarrow i}$. Then, under the RS hypothesis, the incoming messages on a given variable node $i \in V$ (resp. a factor node $a \in E$) are i.i.d. with the law $\widehat{\mathcal{P}}^{\text{RS}}$ (resp. \mathcal{P}^{RS}). This implies that the laws \mathcal{P}^{RS} and $\widehat{\mathcal{P}}^{\text{RS}}$ must obey the following equations:

$$\begin{aligned} \mathcal{P}^{\text{RS}}(\eta) &= \sum_{d=1}^{\infty} r_d \int \prod_{a=1}^d d\widehat{\mathcal{P}}^{\text{RS}}(\widehat{\eta}_a) \delta[\eta - f^{\text{BP}}(\widehat{\eta}_1, \dots, \widehat{\eta}_d)] \\ \widehat{\mathcal{P}}^{\text{RS}}(\widehat{\eta}) &= \int \prod_{i=1}^{k-1} d\mathcal{P}^{\text{RS}}(\eta_i) \delta[\widehat{\eta} - \widehat{f}^{\text{BP}}(\eta_1, \dots, \eta_{k-1})] \end{aligned} \quad (\text{A1})$$

where $\eta = f^{\text{BP}}(\widehat{\eta}_1, \dots, \widehat{\eta}_d)$ and $\widehat{\eta} = \widehat{f}^{\text{BP}}(\eta_1, \dots, \eta_{k-1})$ are shorthand notations for the BP equations (4), and where r_d is the residual degree distribution:

$$r_d = \frac{(d+1)p_d}{\sum_d (d+1)p_d} \quad (\text{A2})$$

with p_d the degree distribution. For the Erdős Rényi ensemble, the degree distribution follows a Poisson law of mean αk , and the residual distribution therefore follows the same law.

1. RS Overlap

We define the following component-wise overlap between two super-spin configurations $\underline{X}, \underline{X}'$ for each $s \in \{1, \dots, y\}$:

$$O_s(\underline{X}, \underline{X}') = \frac{1}{N} \sum_{i=1}^N \sigma_i^s \sigma_i'^s \quad (\text{A3})$$

where $\underline{X} = (X_1, \dots, X_N)$, and $X_i = (\sigma_i^1, \dots, \sigma_i^y) \in \{-1, +1\}^y$. The averaged overlap over the super-spin probability distribution (3) writes:

$$\begin{aligned} \langle O_s(\underline{X}, \underline{X}') \rangle_{\mu_y} &= \sum_{\underline{X}} \sum_{\underline{X}'} \mu_y(\underline{X}) \mu_y(\underline{X}') O_s(\underline{X}, \underline{X}') \\ &= \frac{1}{N} \sum_{i=1}^N \left(\sum_{X_i} \mu_i(X_i) \sigma_i^s \right)^2 \end{aligned} \quad (\text{A4})$$

with μ_i the marginal probability distribution of X_i .

In order to compute the overlap averaged over the random ensemble of hyper-graph instances, we need to average over the distribution of BP marginals μ_i . In the RS formalism, we obtain (with $X = (\sigma^1, \dots, \sigma^y) \in \{-1, 1\}^y$):

$$q_s^{\text{RS}} = \int d\mathcal{Q}^{\text{RS}}(\mu) \left(\sum_X \mu(X) \sigma^s \right)^2 \quad (\text{A5})$$

with \mathcal{Q}^{RS} probability distribution for the BP marginal μ_i , for a uniformly chosen node i on a random hyper-graph G , satisfying the equation:

$$\mathcal{Q}^{\text{RS}}(\mu) = \sum_d p_d \int \prod_{a=1}^d d\widehat{\mathcal{P}}^{\text{RS}}(\widehat{\eta}_a) \delta[\mu - g^{\text{BP}}(\widehat{\eta}_1, \dots, \widehat{\eta}_d)] \quad (\text{A6})$$

with $\mu = g^{\text{BP}}(\widehat{\eta}_1, \dots, \widehat{\eta}_d)$ a shorthand notation for equation (5).

2. Invariances of the RS solution

Note that the probability distribution (3) is invariant under a global spin-flip symmetry. The RS prediction

for the overlap q_s^{RS} is therefore trivially equal to 0.5. It is however a useful sanity check to compute this quantity in order to verify the correctness of the RS solution found numerically with population dynamics.

Similarly, the probability distribution (3) is invariant under a permutation of the y copies $\underline{X}^1, \dots, \underline{X}^y$:

$$\mu_y(\underline{X}^1, \dots, \underline{X}^y) = \mu_y(\underline{X}^{\pi(1)}, \dots, \underline{X}^{\pi(y)}) \quad \forall \pi \in \mathcal{S}(y) \quad (\text{A7})$$

As a result, the marginal probabilities $\mu_i(X_i)$ are also invariant under a permutation of its components. This is verified for instance in Fig. 4, and it is also a good sanity-check for the RS solution found numerically.

Appendix B: One-step Replica Symmetry Breaking cavity method

1. 1RSB cavity equations

Under the 1RSB hypothesis, one assumes that the probability distribution (3) is partitioned into clusters (or pure states):

$$\mu_y(\underline{X}) = \sum_{\gamma} p(\gamma) \mu_{y,\gamma}(\underline{X}) \quad (\text{B1})$$

with $p(\gamma)$ the distribution over the clusters. The restriction of the distribution μ_y to one cluster can be described by the RS formalism, i.e. can be described by a fixed-point of the BP equations.

We define $P_{i \rightarrow a}$ (resp. $\widehat{P}_{a \rightarrow i}$) as the probability law of the message $\eta_{i \rightarrow a}^\gamma$ (resp. $\widehat{\eta}_{a \rightarrow i}^\gamma$), for a cluster γ being chosen randomly with probability $p(\gamma)$. Then the 1RSB messages $P_{i \rightarrow a}, \widehat{P}_{a \rightarrow i}$ satisfy the following self-consistent equations:

$$\begin{aligned}
P_{i \rightarrow a}(\eta_{i \rightarrow a}) &= \frac{1}{Z_{i \rightarrow a}^{\text{1RSB}}} \int \prod_{b \in \partial i \setminus a} d\hat{P}_{b \rightarrow i}(\hat{\eta}_{b \rightarrow i}) \delta(\eta_{i \rightarrow a} - f^{\text{BP}}(\{\hat{\eta}_{b \rightarrow i}\}_{b \in \partial i \setminus a})) z_{i \rightarrow a}(\{\hat{\eta}_{b \rightarrow i}\}_{b \in \partial i \setminus a})^x \\
\hat{P}_{a \rightarrow i}(\hat{\eta}_{a \rightarrow i}) &= \frac{1}{\hat{Z}_{a \rightarrow i}^{\text{1RSB}}} \int \prod_{j \in \partial a \setminus i} dP_{j \rightarrow a}(\eta_{j \rightarrow a}) \delta(\hat{\eta}_{a \rightarrow i} - \hat{f}^{\text{BP}}(\{\eta_{j \rightarrow a}\}_{j \in \partial a \setminus i})) \hat{z}_{a \rightarrow i}(\{\{\eta_{j \rightarrow a}\}_{j \in \partial a \setminus i}\})^x
\end{aligned} \tag{B2}$$

where x is the Parisi parameter, allowing to weight differently the various clusters according to their size. In the above equation, $Z_{i \rightarrow a}^{\text{1RSB}}$ and $\hat{Z}_{a \rightarrow i}^{\text{1RSB}}$ are normalization factors for $P_{i \rightarrow a}$ and $\hat{P}_{a \rightarrow i}$. The terms $z_{i \rightarrow a}(\{\hat{\eta}_{b \rightarrow i}\}_{b \in \partial i \setminus a})$ and $\hat{z}_{a \rightarrow i}(\{\{\eta_{j \rightarrow a}\}_{j \in \partial a \setminus i}\})$ are the normalization factors in equations (4).

In order to average over the disorder, one introduces the probability distributions over the 1RSB messages: $\mathcal{P}^{\text{1RSB}}(P)$ and $\hat{\mathcal{P}}^{\text{1RSB}}(\hat{P})$. They obey the 1RSB equations (similar to the RS equations (A1)):

$$\begin{aligned}
\mathcal{P}^{\text{RSB}}(P) &= \sum_d r_d \int \prod_{a=1}^d d\hat{\mathcal{P}}^{\text{1RSB}}(\hat{P}_a) \delta(P - F(\hat{P}_1 \dots \hat{P}_d)) \\
\hat{\mathcal{P}}^{\text{RSB}}(\hat{P}) &= \int \prod_{i=1}^{k-1} d\mathcal{P}^{\text{1RSB}}(P_i) \delta(Q - \hat{F}(P_1, \dots, P_{k-1}))
\end{aligned} \tag{B3}$$

where $P = F(\hat{P}_1, \dots, \hat{P}_d)$ and $\hat{P} = \hat{F}(P_1, \dots, P_{k-1})$ are shorthand notations for the equations (B2).

2. RS trivial fixed-point

On a given graph, in the Replica Symmetric phase, there is only one fixed point to the BP equation (4), that we denote $\{\bar{\eta}_{i \rightarrow a}, \bar{\hat{\eta}}_{a \rightarrow i}\}_{i \in V, a \in \partial i}$. The solution to the equations (B2) is therefore a trivial Dirac delta:

$$\begin{aligned}
P_{i \rightarrow a}(\eta_{i \rightarrow a}) &= \delta(\eta_{i \rightarrow a}, \bar{\eta}_{i \rightarrow a}) \\
\hat{P}_{a \rightarrow i}(\hat{\eta}_{a \rightarrow i}) &= \delta(\hat{\eta}_{a \rightarrow i}, \bar{\hat{\eta}}_{a \rightarrow i})
\end{aligned} \tag{B4}$$

Once averaged over the disorder, we can see that the 1RSB equations (B3) always admit the RS trivial fixed-point:

$$\begin{aligned}
\mathcal{P}^*(P) &= \int d\mathcal{P}^{\text{RS}}(\bar{\eta}) \delta[P(\eta) - \delta(\eta, \bar{\eta})] \\
\hat{\mathcal{P}}^*(\hat{P}) &= \int d\hat{\mathcal{P}}^{\text{RS}}(\bar{\hat{\eta}}) \delta[\hat{P}(\hat{\eta}) - \delta(\hat{\eta}, \bar{\hat{\eta}})]
\end{aligned} \tag{B5}$$

For small values of the density of constraints α , this trivial solution is the only one, and the predictions given

by the RS and 1RSB cavity method coincide: we are in the RS phase. Increasing α , non-trivial solutions of (B3) can appear. The clustering threshold $\alpha_d(y, \gamma)$ is defined as the smallest value of α for which the 1RSB equations at $x = 1$ admit a non-trivial solution.

3. Simplifications of the 1RSB equations at Parisi parameter $x = 1$

The complete 1RSB equations (B3) can be simplified for the special value of $x = 1$. As explained in [41], the first step is to note that the normalization factor $Z_{i \rightarrow a}^{\text{1RSB}}$ (resp. $\hat{Z}_{a \rightarrow i}^{\text{1RSB}}$) depend only on the distributions $\{\hat{P}_{b \rightarrow i}\}_{b \in \partial i \setminus a}$ (resp. on $\{P_{j \rightarrow a}\}_{j \in \partial a \setminus i}$) through their mean value. One defines $\bar{\eta}[P]$, $\bar{\hat{\eta}}[\hat{P}]$ as the averages:

$$\begin{aligned}
\bar{\eta}[P](X) &= \int dP(\eta) \eta(X) \\
\bar{\hat{\eta}}[\hat{P}](X) &= \int d\hat{P}(\hat{\eta}) \hat{\eta}(X)
\end{aligned} \tag{B6}$$

Then, one can check that $Z_{i \rightarrow a}^{\text{1RSB}}$ depends on the distributions $\hat{P}_1, \dots, \hat{P}_d$ only through the averages $\bar{\eta}[\hat{P}_1], \dots, \bar{\hat{\eta}}[\hat{P}_d]$ (and similarly for $\hat{Z}_{a \rightarrow i}^{\text{1RSB}}$):

$$\begin{aligned}
Z_{i \rightarrow a}^{\text{1RSB}}(\hat{P}_1, \dots, \hat{P}_d) &= z_{i \rightarrow a}(\bar{\eta}[\hat{P}_1], \dots, \bar{\hat{\eta}}[\hat{P}_d]) \\
\hat{Z}_{a \rightarrow i}^{\text{1RSB}}(P_1, \dots, P_{k-1}) &= \hat{z}_{a \rightarrow i}(\bar{\eta}[P_1], \dots, \bar{\eta}[P_{k-1}])
\end{aligned} \tag{B7}$$

Then, one can check that the random variables $\bar{\eta}[P]$, $\bar{\hat{\eta}}[\hat{P}]$ obtained by drawing P (resp \hat{P}) from $\mathcal{P}^{\text{1RSB}}(P)$ (resp $\hat{\mathcal{P}}^{\text{1RSB}}(\hat{P})$) satisfy the RS equations (A1), and therefore are distributed according to the RS distributions \mathcal{P}^{RS} , $\hat{\mathcal{P}}^{\text{RS}}$. One defines the conditional averages:

$$\begin{aligned}
\bar{P}(\eta|\bar{\eta}) &= \frac{1}{\mathcal{P}^{\text{RS}}(\bar{\eta})} \int d\mathcal{P}^{\text{1RSB}}(P) P(\eta) \delta(\bar{\eta}, \bar{\eta}[P]) \\
\bar{\hat{P}}(\hat{\eta}|\bar{\hat{\eta}}) &= \frac{1}{\hat{\mathcal{P}}^{\text{RS}}(\bar{\hat{\eta}})} \int d\hat{\mathcal{P}}^{\text{1RSB}}(\hat{P}) \hat{P}(\hat{\eta}) \delta(\bar{\hat{\eta}}, \bar{\hat{\eta}}[\hat{P}])
\end{aligned} \tag{B8}$$

We can get closed equations for these distributions:

$$\begin{aligned}
\bar{P}(\eta|\bar{\eta})\mathcal{P}^{\text{RS}}(\bar{\eta}) &= \sum_d r_d \int \prod_{a=1}^d d\hat{\mathcal{P}}^{\text{RS}}(\hat{\eta}_a)\delta[\bar{\eta} - f^{\text{BP}}(\hat{\eta}_1, \dots, \hat{\eta}_d)] \int \prod_{a=1}^d d\tilde{P}(\hat{\eta}_a|\hat{\eta}_a)\delta[\eta - f^{\text{BP}}(\hat{\eta}_1, \dots, \hat{\eta}_d)] \frac{z_{i \rightarrow a}(\hat{\eta}_1, \dots, \hat{\eta}_d)}{z_{i \rightarrow a}(\hat{\eta}_1, \dots, \hat{\eta}_d)} \\
\tilde{P}(\hat{\eta}|\hat{\eta})\hat{\mathcal{P}}^{\text{RS}}(\hat{\eta}) &= \int \prod_{i=1}^{k-1} d\mathcal{P}^{\text{RS}}(\bar{\eta}_i)\delta[\hat{\eta} - \hat{f}^{\text{BP}}(\bar{\eta}_1, \dots, \bar{\eta}_{k-1})] \int \prod_{i=1}^{k-1} d\bar{P}(\eta_i|\bar{\eta}_i)\delta[\hat{\eta} - \hat{f}^{\text{BP}}(\eta_1, \dots, \eta_{k-1})] \frac{\hat{z}_{a \rightarrow i}(\eta_1, \dots, \eta_{k-1})}{\hat{z}_{a \rightarrow i}(\bar{\eta}_1, \dots, \bar{\eta}_{k-1})}
\end{aligned} \tag{B9}$$

These equations are simpler than equations (B3): the distribution $\bar{P}(\eta|\bar{\eta})\mathcal{P}^{\text{RS}}(\bar{\eta})$ can be seen as a joint distribution over $\eta, \bar{\eta}$, and can be represented by a population of couples $(\{\eta_i, \bar{\eta}_i\}_{i=1}^N)$. However, the factor $\frac{z_{i \rightarrow a}(\hat{\eta}_1, \dots, \hat{\eta}_d)}{z_{i \rightarrow a}(\hat{\eta}_1, \dots, \hat{\eta}_d)}$ is still hard to represent with a population. To get rid of it, we can define the following distributions (following the steps of [41]):

$$\begin{aligned}
P_X(\eta|\bar{\eta}) &= \frac{\eta(X)}{\bar{\eta}(X)} \bar{P}(\eta|\bar{\eta}) \\
\hat{P}_X(\hat{\eta}|\hat{\eta}) &= \frac{\hat{\eta}(X)}{\hat{\eta}(X)} \tilde{P}(\hat{\eta}|\hat{\eta})
\end{aligned} \tag{B10}$$

One can check that these distributions now satisfy the following self-consistent equations:

$$\begin{aligned}
P_X(\eta|\bar{\eta})\mathcal{P}^{\text{RS}}(\bar{\eta}) &= \sum_d r_d \int \prod_{a=1}^d d\hat{\mathcal{P}}^{\text{RS}}(\hat{\eta}_a)\delta[\bar{\eta} - f^{\text{BP}}(\hat{\eta}_1, \dots, \hat{\eta}_d)] \int \prod_{a=1}^d d\hat{P}_X(\hat{\eta}_a|\hat{\eta}_a)\delta[\eta - f^{\text{BP}}(\hat{\eta}_1, \dots, \hat{\eta}_d)] \\
\hat{P}_X(\hat{\eta}|\hat{\eta})\hat{\mathcal{P}}^{\text{RS}}(\hat{\eta}) &= \int \prod_{i=1}^{k-1} d\mathcal{P}^{\text{RS}}(\bar{\eta}_i)\delta[\hat{\eta} - \hat{f}^{\text{BP}}(\bar{\eta}_1, \dots, \bar{\eta}_{k-1})] \sum_{\{X_i\}_{i=1}^k} \nu(\{X_i\}_{i=1}^{k-1}|X, \{\bar{\eta}_i\}_{i=1}^{k-1}) \int \prod_{i=1}^{k-1} dP_{X_i}(\eta_i|\bar{\eta}_i) \\
&\quad \times \delta[\hat{\eta} - \hat{f}^{\text{BP}}(\eta_1, \dots, \eta_{k-1})]
\end{aligned} \tag{B11}$$

with the probability distribution:

$$\nu(\{X_i\}_{i=1}^{k-1}|X, \{\bar{\eta}_i\}_{i=1}^{k-1}) = \frac{\Omega(X, \{X_i\}_{i=1}^{k-1}) \prod_{i=1}^{k-1} \bar{\eta}_i(X_i)}{\sum_{\{X_i\}_{i=1}^{k-1}} \Omega(X, \{X_i\}_{i=1}^{k-1}) \prod_{i=1}^{k-1} \bar{\eta}_i(X_i)}. \tag{B12}$$

Equations (B11) can be solved with population dynamics. One employs a population of tuples $\{\bar{\eta}_i, \{\eta_i^X\}_{X \in \{+1, -1\}^y}\}_{i=1}^N s$ and $\{\hat{\eta}_i, \{\hat{\eta}_i^X\}_{X \in \{+1, -1\}^y}\}_{i=1}^N s$. As each tuple contains $2^y + 1$ elements, this representation can become too heavy to be implemented numerically with a large value of y . We will see in the following (see section B3b) how to decrease the number of elements to be stored by using the symmetries of the model.

a. RS trivial fixed-point

The RS trivial fixed-point (B5) can be written in terms of the distributions (B10):

$$\begin{aligned}
P_X(\eta|\bar{\eta})\mathcal{P}^{\text{RS}}(\bar{\eta}) &= \delta(\eta, \bar{\eta})\mathcal{P}^{\text{RS}}(\bar{\eta}) \\
\hat{P}_X(\hat{\eta}|\hat{\eta})\hat{\mathcal{P}}^{\text{RS}}(\hat{\eta}) &= \delta(\hat{\eta}, \hat{\eta})\hat{\mathcal{P}}^{\text{RS}}(\hat{\eta}).
\end{aligned} \tag{B13}$$

b. *Further simplifications using symmetries*

Let $\pi : \{1, \dots, y\} \rightarrow \{1, \dots, y\}$ be a permutation of the super-spin components. Let $X = (\sigma^1, \dots, \sigma^y) \in \{+1, -1\}^y$ be a super-spin vector. We will use the notation $\pi(X) = (\sigma^{\pi(1)}, \dots, \sigma^{\pi(y)})$. Furthermore, let η^π be the BP message η after permutation of the components:

$$\eta^\pi(X) = \eta(\pi(X)). \quad (\text{B14})$$

Starting from equations (B8) one can check that the probability distribution $\bar{P}(\eta|\bar{\eta})$ satisfies the following invariance:

$$\bar{P}(\eta|\bar{\eta}) = \bar{P}(\eta^\pi|\bar{\eta}^\pi) \quad (\text{B15})$$

(and similarly for $\tilde{P}(\hat{\eta}|\tilde{\eta})$).

Using the definitions (B10) for the distributions P_X and \hat{P}_X , this invariance translates into the relations:

$$\begin{aligned} P_X(\eta|\bar{\eta}) &= P_{\pi^{-1}(X)}(\eta^\pi|\bar{\eta}^\pi) \\ \hat{P}_X(\hat{\eta}|\tilde{\eta}) &= \hat{P}_{\pi^{-1}(X)}(\hat{\eta}^\pi|\tilde{\eta}^\pi) \end{aligned} \quad (\text{B16})$$

One can use this relation to decrease the number of distributions involved in the self-consistent equations (B11).

For each $n \in \{0, 1, \dots, y\}$, let

$$U_n = (\underbrace{1, \dots, 1}_n, \underbrace{-1, \dots, -1}_{y-n}) \quad (\text{B17})$$

be the configuration with the n first components being 1, the remaining components being -1 . For each $X \in \{+1, -1\}^y$, define π_X such that:

$$\pi_X(X) = U_{n(X)} \quad (\text{B18})$$

where $n(X) = |\{s \in \{1, \dots, y\} : \sigma(X)^s = 1\}|$ is the number of components equal to 1 in X . In other words, π_X is the permutation of the components such that, when applied to the vector X , it places all the 1-components first.

Define the $y + 1$ distributions, for each $n \in \{0, 1, \dots, y\}$:

$$P_n = P_{U_n} \quad (\text{and similarly: } \hat{P}_n = \hat{P}_{U_n}) \quad (\text{B19})$$

Then, the invariance relation (B16) applied to $\pi = \pi_X$ gives:

$$P_X(\eta|\bar{\eta}) = P_{n(X)} \left(\eta^{(\pi_X)^{-1}} | \bar{\eta}^{(\pi_X)^{-1}} \right) \quad (\text{B20})$$

Inserting this relation into the second equation of (B11), this allows us to write a set of self-consistent equations only between the $2(y+2)$ distributions $\{\hat{P}_n\}_{n \in \{0, \dots, y\}}$, $\{\tilde{P}_n\}_{n \in \{0, \dots, y\}}$, \mathcal{P}^{RS} , and $\hat{\mathcal{P}}^{\text{RS}}$:

$$\begin{aligned} P_n(\eta|\bar{\eta})\mathcal{P}^{\text{RS}}(\bar{\eta}) &= \sum_d r_d \int \prod_{a=1}^d d\hat{\mathcal{P}}^{\text{RS}}(\tilde{\eta}_a) \delta[\bar{\eta} - f^{\text{BP}}(\tilde{\eta}_1, \dots, \tilde{\eta}_d)] \int \prod_{a=1}^d d\hat{P}_n(\hat{\eta}_a|\tilde{\eta}_a) \delta[\eta - f^{\text{BP}}(\hat{\eta}_1, \dots, \hat{\eta}_d)] \\ \hat{P}_n(\hat{\eta}|\tilde{\eta})\hat{\mathcal{P}}^{\text{RS}}(\tilde{\eta}) &= \int \prod_{i=1}^{k-1} d\mathcal{P}^{\text{RS}}(\tilde{\eta}_i) \delta[\tilde{\eta} - \hat{f}^{\text{BP}}(\tilde{\eta}_1, \dots, \tilde{\eta}_{k-1})] \sum_{\{X_i\}_{i=1}^k} \nu(\{X_i\}_{i=1}^{k-1} | X, \{\tilde{\eta}_i\}_{i=1}^{k-1}) \int \prod_{i=1}^{k-1} dP_{n(X_i)}(\eta_i|\tilde{\eta}_i) \\ &\quad \times \delta[\hat{\eta} - \hat{f}^{\text{BP}}(\eta_1^{\pi_{X_1}}, \dots, \eta_{k-1}^{\pi_{X_{k-1}}})] \end{aligned} \quad (\text{B21})$$

We solved these equations iteratively with population dynamics, see appendix B 4 for more details on the numerical resolution.

c. *Inter-state and intra-state overlap*

The inter-state overlap q_0 and intra-state overlap q_1 , defined in section II C 2 can be written in terms of the

1RSB distributions. Starting from the definition of the component-wise overlap (A3) and its average (A4) over the super-spin probability distribution (3), we obtain:

- the (component-wise) *inter-state overlap* between two configurations sampled independently from

(3):

$$q_0^s = \int d\mathcal{P}^{\text{1RSB}}(P) \left[\left(\int dP(\eta) \sum_{X \in \{\pm 1\}^y} \sigma^s \eta(X) \right)^2 \right] \quad (\text{B22})$$

with $X = (\sigma^1, \dots, \sigma^y) \in \{+1, -1\}^y$.

- the (component-wise) *intra-state overlap* between two configurations sampled from the same cluster:

$$q_1^s = \int d\mathcal{P}^{\text{1RSB}}(P) \left[\int dP(\eta) \left(\sum_{X \in \{\pm 1\}^y} \sigma^s \eta(X) \right)^2 \right] \quad (\text{B23})$$

The intra-state and inter-state overlap q_0 and q_1 are simply $q_0 = \sum_{s=1}^y q_0^s$, $q_1 = \sum_{s=1}^y q_1^s$. In particular, q_1 is the quantity plotted in Fig. 3.

One can check that the inter-state overlap is trivially equal to the RS overlap q_s^{RS} . Using the definition of $\bar{\eta}[P]$ (B6), and recalling that $\bar{\eta}[P]$ satisfies the RS equations (A1), therefore is distributed according to \mathcal{P}^{RS} , one obtains:

$$\begin{aligned} q_0^s &= \int d\mathcal{P}^{\text{1RSB}}(P) \left(\sum_{X \in \{\pm 1\}^y} \sigma^s \bar{\eta}[P](X) \right)^2 \\ &= \int d\mathcal{P}^{\text{RS}}(\eta) \left(\sum_{X \in \{\pm 1\}^y} \sigma^s \eta(X) \right)^2 = q_s^{\text{RS}}. \end{aligned} \quad (\text{B24})$$

Furthermore, one can express the intra-state overlap in terms of the distribution P_X :

$$\begin{aligned} q_1^s &= \int d\mathcal{P}^{\text{RS}}(\bar{\eta}) \left[\sum_X \sigma^s \bar{\eta}(X) \int dP_X(\eta|\bar{\eta}) \left(\sum_{X'} \eta(X') \sigma'^s \right) \right] \\ &= \int d\mathcal{P}^{\text{RS}}(\bar{\eta}) \left[\sum_X \sigma^s \bar{\eta}(X) \int dP_{n(X)}(\eta|\bar{\eta}) \left(\sum_{X'} \eta^{\pi_X}(X') \sigma'^s \right) \right]. \end{aligned} \quad (\text{B25})$$

With $X = (\sigma^1, \dots, \sigma^y)$ and $X' = (\sigma'^1, \dots, \sigma'^y)$. In the second equation, we used the invariance relations (B16) to write the intra-state overlap only in terms of the distributions $\{P_n\}_{n \in \{0, \dots, y\}}$.

4. Numerical resolution of the 1RSB equations

The numerical resolution of the 1RSB equations (B21) can be done with population dynamics. A first possibility is to represent both the distributions

$\{\mathcal{P}, P_0, \dots, P_y\}$ and $\{\hat{\mathcal{P}}, \hat{P}_0, \dots, \hat{P}_y\}$, with two population of tuples: $\{(\bar{\eta}^{(i)}, \eta_0^{(i)}, \dots, \eta_y^{(i)}) : i \in \{1, \dots, \mathcal{N}\}\}$ and $\{(\hat{\eta}^{(i)}, \hat{\eta}_0^{(i)}, \dots, \hat{\eta}_y^{(i)}) : i \in \{1, \dots, \mathcal{N}\}\}$. However, an iterative solution using this implementation suffers from oscillations that prevent convergence to a fixed point. We therefore preferred to use a single population of elements $\{(\bar{\eta}^{(i)}, \eta_0^{(i)}, \dots, \eta_y^{(i)}) : i \in \{1, \dots, \mathcal{N}\}\}$. This implementation amounts to solve the set of self-consistent equations on the distributions $\{\mathcal{P}, P_0, \dots, P_y\}$ only (that can be obtained by plugging the second equation of (B21) into the first one).

At each iteration, one constructs a new population of elements by repeating for $i \in \{1, \dots, \mathcal{N}\}$, independently, the following steps:

- draw a random number d from the residual distribution r_d (in the case of ER graphs, it is a Poisson law of parameter αk).

- draw $d(k-1)$ indices $\{j_{i,a}\}_{i \in \{1, \dots, k-1\}, a \in \{1, \dots, d\}}$ i.i.d. from $\{1, \dots, \mathcal{N}\}$

- Compute the new message:

$$\bar{\eta}^{(i)} = \tilde{f}^{\text{BP}}(\{\bar{\eta}^{(j_{i,a})}\}_{i \in \{1, \dots, k-1\}, a \in \{1, \dots, d\}}) \quad (\text{B26})$$

with \tilde{f}^{BP} a shorthand notation for the BP equation on the η messages only (obtained by plugging the second equation of (4) in the first one). This can be done efficiently with convolutions, see next section B 4 a for more details.

- For each $n \in \{0, \dots, y\}$:

- Sample, for each $a \in \{1, \dots, d\}$ independently, the super-spin configurations: $\{X_{i,a}\}_{i \in \{1, \dots, k-1\}}$ from the distribution (cf. equation (B12))

$$\nu(\cdot | U_n, \{\bar{\eta}_{j_{i,a}}\}_{i \in \{1, \dots, k-1\}})$$

with $U_n = (\underbrace{1, \dots, 1}_n, \underbrace{-1, \dots, -1}_{y-n})$ be the con-

figuration with the n first components being 1, the remaining components being -1 (see section B 4 b. for an efficient implementation).

- Apply the permutation to each incoming message:

$$\eta'_{n(X_{i,a})} = \left(\eta_n^{(j_{i,a})} \right)^{\pi_{X_{i,a}}}$$

- Compute the new message:

$$\eta_n^{(i)} = \tilde{f}^{\text{BP}}(\{\eta'_{n(X_{i,a})}\}_{i \in \{1, \dots, k-1\}, a \in \{1, \dots, d\}}) \quad (\text{B27})$$

- Fill the population with the new element $\{(\bar{\eta}^{(i)}, \eta_0^{(i)}, \dots, \eta_y^{(i)}) : i \in \{1, \dots, \mathcal{N}\}\}$.

Iterating these steps many times, one converges to a fixed point solution of (B21).

a. *Convolutions*

In this sub-section, one explains how to compute a new message $\hat{\eta} = \hat{f}^{\text{BP}}(\{\eta_i\}_{i \in \{1, \dots, k-1\}})$ from equation (4) efficiently using convolutions.

For this purpose, we adopt the binary representation of a super-spin variable $X = (x^1, \dots, x^y) \in \{0, 1\}^y$ (where one uses the spin-bit correspondence $\sigma = 2x - 1$).

We re-write the second BP equation (4) here for convenience:

$$\hat{\eta}(X) = \frac{1}{\hat{z}} \sum_{X_1, \dots, X_{k-1}} \Omega(X, X_1, \dots, X_{k-1}) \prod_{i=1}^{k-1} \eta_i(X_i) \quad (\text{B28})$$

with

$$\Omega(X, X_1, \dots, X_{k-1}) = \prod_{s=1}^y \mathbb{I}[x^s, x_1^s, \dots, x_{k-1}^s \text{ not all equal}] \quad (\text{B29})$$

with $\mathbb{I}[A]$ the indicator function of the event A . The sum $\sum_{X_1, \dots, X_{k-1}}$ contains $2^{y(k-1)}$ terms. Using convolution allows to decrease the number of operations needed in order to compute this sum.

Let $X, X' \in \{0, 1\}^y$ be two super-bit configurations, and let $X \oplus X'$ be the result of the XOR operation applied component-wise. Similarly, let $X \vee X'$ be the result of OR operation applied component-wise.

Let $h_1 : \{0, 1\}^y \rightarrow \mathbb{R}$, $h_2 : \{0, 1\}^y \rightarrow \mathbb{R}$ be a pair of discrete functions, and define the convolution as follows:

$$\begin{aligned} h_1 \otimes h_2 &: \{0, 1\}^y \rightarrow \mathbb{R} \\ h_1 \otimes h_2(X) &= \sum_{X_1, X_2} h_1(X_1) h_2(X_2) \mathbb{I}[X = X_1 \vee X_2] \end{aligned} \quad (\text{B30})$$

At fixed X , one can define the super-spin configurations $W_i = X \oplus X_i$, that stores the super-spin components that are satisfied by X_i . One can re-write the

function Ω in terms of W_1, \dots, W_{k-1} :

$$\Omega(X, X_1, \dots, X_{k-1}) = \mathbb{I}[W_1 \vee W_2 \vee \dots \vee W_{k-1} = U_y] \quad (\text{B31})$$

with $U_y = (1, \dots, 1)$ the super-spin variable with all components equal to 1.

With these notations, one defines recursively the functions $\{g_i : \{0, 1\}^y \rightarrow \mathbb{R}\}_{i \in \{0, \dots, k-1\}}$ as follows:

$$\begin{aligned} g_i(X) &= (g_{i-1} \otimes \eta_i(X \oplus \cdot))(X) \\ &= \sum_{W_i} \sum_{X'} \eta_i(W_i \oplus X) g_{i-1}(X') \mathbb{I}[X = X' \vee W_i] \end{aligned} \quad (\text{B32})$$

With the initialisation

$$g_0(X) = \mathbb{I}[X = U_0] \quad (\text{B33})$$

with $U_0 = (0, \dots, 0)$.

One can check that:

$$\hat{\eta}(X) = \frac{1}{\hat{z}} g_{k-1}(U_y) \quad (\text{B34})$$

Computing each function g_i requires a sum over two super-spin variables, therefore leading to a total cost of $2^{2y}(k-1)$ operations.

b. *Sampling super-spin configurations*

In this sub-section, one explains how to sample efficiently the super-spin configurations X_1, \dots, X_{k-1} from the probability distribution $\nu(X_1, \dots, X_{k-1} | X, \bar{\eta}_1, \dots, \bar{\eta}_{k-1})$ given in equation (B12). Being a distribution over $k-1$ super-spins, it is a function of $2^{y(k-1)}$ variables. However, one can avoid to compute and store the joint probability over X_1, \dots, X_{k-1} , by sampling them iteratively.

Using the notations introduced in appendix B 4 a, one write the probability distribution of X_{k-1} (marginalized over the remaining super-spin configurations) X_1, \dots, X_{k-2} as:

$$P_\nu(X_{k-1}) = \frac{\bar{\eta}_{k-1}(X_{k-1}) \sum_Z g_{k-2}(Z) \mathbb{I}[(Z \vee W_{k-1}) = U_y]}{g_{k-1}(U_y)} \quad (\text{B35})$$

with the functions g_0, \dots, g_{k-1} defined in the previous section, and with $W_i = X \oplus X_i$. Similarly, the conditional probability distribution of X_{k-i} given $X_{k-i+1}, \dots, X_{k-1}$ is, for each $i \in \{1, \dots, k-2\}$:

$$P_\nu(X_{k-i} | X_{k-i+1}, \dots, X_{k-1}) = \frac{\bar{\eta}_{k-i}(X_{k-i}) \sum_Z g_{k-i-1}(Z) \mathbb{I}[Z \vee W_{k-i} \vee \dots \vee W_{k-1}]}{\sum_Z g_{k-i}(Z) \mathbb{I}[Z \vee W_{k-i+1} \vee \dots \vee W_{k-1}]} \quad (\text{B36})$$

Therefore, in order to sample the super-spin variables

X_1, \dots, X_{k-1} , one can store the functions g_1, \dots, g_{k-1}

computed during the update of $\bar{\eta}$, instead of the full probability distribution ν .

c. Choice of the initial condition

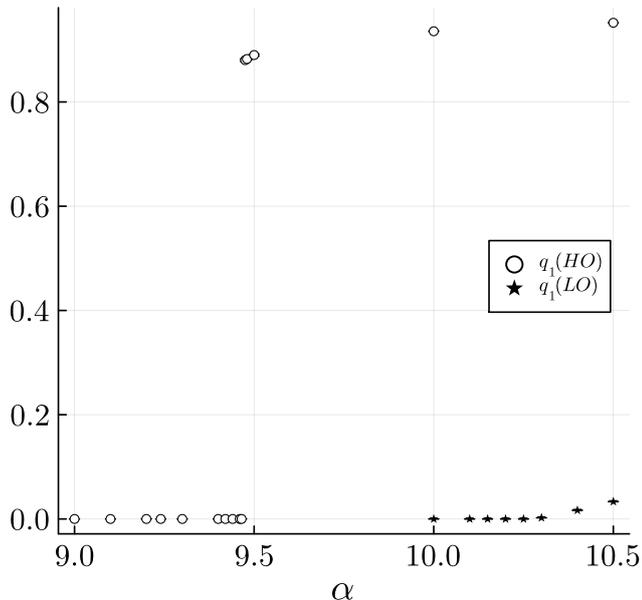


FIG. 7. Intra-state overlap q_1 reached from two initial conditions: HO (white circles) and LO (black stars), see B 4 c, at $\gamma = 0.01$, with population size $\mathcal{N} = 10^5$.

In section III A 1, we unveil the existence of (at least) two different non-trivial solutions to the 1RSB equation (B21). In order to reach numerically these solution, we ran population dynamics with the following initialization:

$$P_n^{(0)}(\eta|\bar{\eta}) = (1 - \epsilon)\mathcal{P}^{\text{RS}}(\bar{\eta}) + \epsilon\delta[\eta(\cdot), \delta(\cdot, U_n)] \quad (\text{B37})$$

with U_n defined in (B17). In other words, with probability ϵ , the message η is polarized on the configuration U_n corresponding to the index of the probability distribution $P_n = P_{U_n}$, and with probability $1 - \epsilon$ it is sampled from the RS distribution \mathcal{P}^{RS} .

For each choice of the parameters we ran twice the population dynamics algorithm, once with $\epsilon = 1$, and once with a small value of $\epsilon > 0$ (in practice we used $\epsilon = 0.01$). We call HO, for high overlap, the initialization with $\epsilon = 1$, and LO (low overlap) the small ϵ one.

d. Numerical determination of the thresholds

In this sub-section, we give details on the numerical computation of the two thresholds α_{KS} , α_{disc} defined in section III A.

We recall that α_{disc} marks the discontinuous appearance of a non-trivial solution to the 1RSB equation (B21). Such transition is depicted in figure 7, at $\gamma = 0.01$, where the intra-state overlap obtained with the HO initial condition abruptly jumps from zero to a positive value at $\alpha_{\text{disc}}(\gamma = 0.01) = 9.461$. The precise numerical determination can be obtained in several ways (see e.g. [22], section III.) by looking at the evolution of q_1 under iterations (see figure 8, top left panel). As α is growing towards α_{disc} , the time $t_*(\alpha)$ needed for $q_1^{(t)}$ to leave the plateau at positive value increases. The time t_* can be computed as the first time $q_1^{(t)}$ crosses an arbitrary value between 0 and the plateau. Plotting t_* as a function of α (figure 8 (top right panel)), one observes a dependence of the form:

$$t_*(\alpha) \simeq K(\alpha_{\text{disc}} - \alpha)^{-1/2} \quad \text{when } \alpha \rightarrow \alpha_{\text{disc}} \quad (\text{B38})$$

with K a finite constant. (as already observed in [22], see [21] for more details on scaling functions describing the overlap evolution). The discontinuous threshold α_{disc} is then obtained from a fit of the data using this scaling function.

The numerical determination of the continuous appearance of a non-trivial solution (or Kesten-Stigum threshold) can be obtained by studying the stability of the trivial RS solution (B5) under the 1RSB equations (B3). Here, we adopted a simpler strategy and observed that the evolution of the intra-state overlap q_1 under iterations is decaying exponentially to 0 for $\alpha < \alpha_{\text{KS}}$ (see figure 8, bottom left panel), i.e.:

$$q_1(t) \simeq A\delta_*^t \quad (\text{B39})$$

The parameter δ_* is extracted from the data $q_1(t)$ with the above scaling function, and is plotted as a function of α in the bottom right panel. Its value is approaching 1 as α increases towards the threshold α_{KS} . A simple linear fit of $\delta_*(\alpha)$ gives the value of the threshold α_{KS} . We checked that this method gives indeed the correct Kesten Stigum threshold at $\gamma = 0$ (i.e. in the non-interacting case): $\alpha_{\text{KS}}(k, \gamma = 0) = \frac{(2^k - 1 - 1)^2}{k(k-1)} = 11.25$ at $k = 5$.

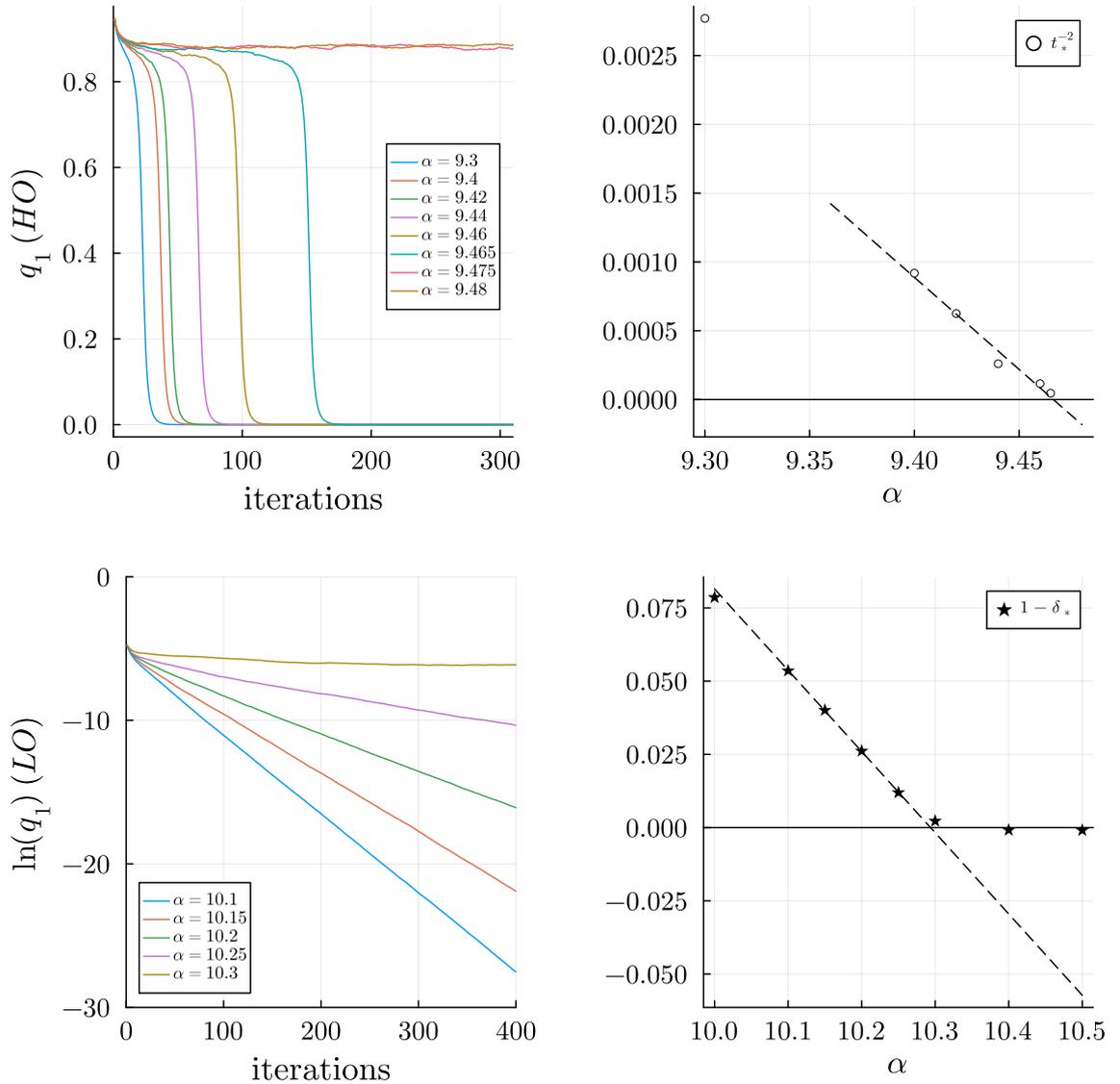


FIG. 8. Numerical determination of the thresholds $\alpha_{\text{disc}}(\gamma)$ (top panels), and $\alpha_{\text{KS}}(\gamma)$ (bottom panels), for $\gamma = 0.01$. Top: the discontinuous transition is computed from the time t_* at which the overlap q_1 leaves the plateau and reaches 0 (left panel). A fit using the scaling function (B38) is used to compute the threshold $\alpha_{\text{disc}}(\gamma = 0.01) = 9.461$ (right panel). Bottom: the continuous transition is computed from the intra-state overlap q_1 exhibiting an exponential decay for $\alpha < \alpha_{\text{KS}}$ (left panel). A fit using the scaling function (B39) is used to compute the threshold $\alpha_{\text{KS}}(\gamma = 0.01) = 10.294$.