PAPER: Disordered systems, classical and quantum

Optimization of the dynamic transition in the continuous coloring problem

Angelo Giorgio Cavaliere[1,*], Thibault Lesieur[1] and Federico Ricci-Tersenghi[1,2]

[1] Dipartimento di Fisica, Università 'La Sapienza', P.le A. Moro 5, 00185, Rome, Italy
[2] INFN, Sezione di Roma1, and CNR-Nanotec, Rome unit, P.le A. Moro 5, 00185, Rome, Italy
E-mail: angelog.cavaliere@gmail.com

Abstract. Random constraint satisfaction problems (CSPs) can exhibit a phase where the number of constraints per variable α makes the system solvable in theory on the one hand, but also makes the search for a solution hard, meaning that common algorithms such as Monte Carlo (MC) method fail to find a solution. The onset of this hardness is deeply linked to the appearance of a dynamical phase transition where the phase space of the problem breaks into an exponential number of clusters. The exact position of this dynamical phase transition is not universal with respect to the details of the Hamiltonian one chooses to represent a given problem. In this paper, we develop some theoretical tools in order to find a systematic way to build a Hamiltonian that maximizes the dynamic $\alpha_{\rm d}$ threshold. To illustrate our techniques, we will concentrate on the problem of continuous coloring, where one tries to set an angle $x_i \in [0; 2\pi]$ on each node of a network in such a way that no adjacent nodes are closer than some threshold angle θ, that is $\cos(x_i - x_j) \leqslant \cos \theta$. This problem can be both seen as a continuous version of the discrete graph coloring problem or as a one-dimensional version of the Mari–Krzakala–Kurchan model. The relevance of this model stems from the fact that continuous CSPs on sparse random graphs remain largely unexplored in statistical physics. We show that for sufficiently small angle θ this model presents

*Author to whom any correspondence should be addressed.

a random first order transition and compute the dynamical, condensation and Kesten–Stigum transitions; we also compare the analytical predictions with MC simulations for values of $\theta = 2\pi/q$, $q \in \mathbb{N}$. Choosing such values of q allows us to easily compare our results with the renowned problem of discrete coloring.

Contents

1. Introduction

In this work we study in detail a constraint satisfaction problem (CSP) with continuous variables and defined on sparse random graphs, that goes under the name of *continuous coloring*. The problem is essentially the following: given a sparse random graph $G = (V, E)$, which can be, e.g. Erdős–Rényi or random regular, the aim is to assign a 'color' $x_i \in [0, 2\pi)$ to each vertex in V, such that the colors assigned to any pair of connected vertices $(ij) \in E$ are different enough, that is they satisfy $\cos(x_i - x_j) \leqslant \cos\theta$ for some threshold value θ.

This model can be regarded as a continuous version of the well-known problem of discrete coloring of random graphs [1]. Moreover, it also corresponds to the one-dimensional version of the Mari–Kurchan–Krzakala (MKK) model [2], a mean-field approximation to models of hard spheres showing jamming. In the MKK model, d-dimensional hard spheres living in a d-dimensional box with periodic boundary conditions interact with only a finite number of other particles, according to an underlying sparse graph network. From this point of view, continuous coloring can thus be interpreted also as an 'angle-packing' problem, with the obvious identification of θ with the diameter of the particles [3].

The MKK model for $d = 2$ was numerically studied in [2], exhibiting the presence of a random first order transition (RFOT) when increasing the diameter of the spheres (the packing fraction), for sufficiently high connectivities. A generalized version of the MKK model, accounting also for p-body interactions, was studied in [4] for different values of p and d through the cavity (or belief propagation) formalism. The case $d = 1$ and $p = 2$, which we are interested in this work, was shown to undergo an RFOT for large connectivities. However, the dynamic or clustering threshold was not computed.

Let us start explaining why we believe it is very useful to study this model. The continuous coloring problem (CCP) possesses *all* the following features:

- it is a CSP,
- having continuous variables,
- defined on a sparse random graph,
- showing a RFOT.

The most famous and well-studied model showing an RFOT is the spherical p-spin model: in this case one has N real (unbounded) variables x_i, subject only to a global constraint $\sum_i x_i^2 = N$. Unfortunately, this model is well defined only on very dense graphs, because as soon as one makes the interactions slightly sparser the model ground state condensates on a small subset of the variables, becoming meaningless from the physical point of view. In order to have a very sparse model, one needs to avoid the

above condensation phenomenon, and this can be achieved either adding a Lagrange multiplier to each real variable, or more simply by using variables defined on a bounded domain. Among the latter models we have all CSPs with discrete variables (e.g. q-col, k-SAT, k-XORSAT). Willing to use continuous variables with bounded domain, the simplest way is to choose models with vector spins, e.g. XY or Heisenberg spins. The CCP we study can indeed be seen as an XY model where variables are unit norm vectors of 2 components, $\vec{s}_i = (\cos(x_i), \sin(x_i))$, the constraints that we enforce being written as $\vec{s}_i \cdot \vec{s}_j \leqslant \cos\theta$. A recent work addresses the glass transition of this kind of rotational degrees of freedom, but focusing on the opposite limit of a large number of vector components [5].

The results of [4] are particularly encouraging, since it is not obvious at all that taking a discrete CSP, like q-coloring, and transforming it to a version with continuous variables, the physics, in this case the RFOT, is preserved. The change in the symmetry of variables, from $\mathbb{Z}_q$ to $O(2)$, is drastic and may change the nature of the phase transition. In the successful case we do find a model with all the above features (and this is what we are going to confirm in section 2), we have in our hand a model which is very useful in many aspects. Let us list those aspects we find more interesting:

- being defined on a locally tree-like graph, the model can be solved analytically via the cavity method, and the location of phase transitions can be computed with high accuracy;

- having bounded variables interacting in a sparse way, the numerical simulation of the model can be made very efficiently, and thus a meaningful comparison between numerical results and analytical solution can be done;

- being a CSP with continuous variables, and where the density of constraints can be varied continuously, the model would show a jamming transition, for which it can be considered as a very simple sparse mean-field model for jamming (please notice that dense mean-field models for jamming, as the perceptron, have problems when generalized to the sparse case[3]);

- having an RFOT, we expect the model to exhibit a dynamical phase transition and, below the dynamic transition temperature, the energy relaxation to get stuck at some threshold energy connected in some way to the topological properties of the energy landscape (e.g. to the spectrum of the energy Hessian at the stationary points) that can be computed, for sufficiently smooth interaction potentials, thanks to the continuous nature of the variables;

- being defined on a sparse random graph, one can change the mean degree, thus testing very different regimes from the dense one to the very sparse one, thus better understanding how much of the classical (dense) mean-field physical behavior is preserved in the sparse regime (this is particularly relevant to understand why it is so difficult to find good glass models in low dimensions where the number of nearest neighbors becomes small).

---

[3] The delicate point is once more given by the fact that one generally considers continuous models with unbounded variables. In the case of the perceptron, if a variable is subjected to K constraints, the probability for all the random obstacles' components relative to that variable to have the same sign, e.g. to be all positive, is 2^{-K}. In this case that variable can satisfy all of its constraints by taking arbitrarily large values (irrespective of other variables) and the model is ill defined. While in the perceptron $K = O(N)$ and this event does never occur, in the diluted case K is finite and the model is ill defined with a finite probability.

To the best of our knowledge, the above aspects have not been investigated in detail in a *single* model before.

The second part of this work will focus on a very interesting aspect, namely the possibility of optimizing the dynamical threshold by reweighting properly the solutions of the problem. It is well known that in complex CSP, like the one we are studying, solutions may have very different features. For example, in discrete random CSP, solutions organize in clusters of very different sizes and of different nature (e.g. with or without frozen variables) [1, 6, 7]. Solutions that dominate the thermodynamics may be very different from the ones which are found via the best solving algorithms [3], and this makes difficult the connection between thermodynamics and the behavior of dynamical processes searching for solutions. Reweighting the solutions, that is giving them a non-uniform weight, is a simple way to count the atypical solutions that would not weigh enough in the uniform measure (which is usually adopted in order to derive the equilibrium phase diagram of random CSPs).

This program has been pursued in the literature following slightly different approaches [8–10]: favor solutions which are surrounded by a higher number of other solutions in configuration space; the relevance of subdominant clusters of solutions with high internal entropy has been recently pointed out also in [11]; on the contrary [12] uses the number of frozen variables inside each cluster in order to weight differently the solutions. The simplest approach one can resort to is, however, to directly bias the model Hamiltonian. We will perform an optimization of the interaction potential with the aim of postponing as more as possible the dynamical phase transition. This has been done recently in [13] for the discrete CSP of bicoloring random hypergraphs (a generalization to larger interaction range for the bias is considered in [14]) and was done in [15, 16] for models of hard spheres in infinite dimensions. The idea behind this optimization is that, by working with the reweighted interaction potential, the long range correlations leading to the ergodicity breaking at the dynamic phase transition will appear later, and algorithms should find solutions in an easier way in these 'biased landscapes'. The innovative aspect in the present work with respect to [13] is that the optimization of the potential is performed in a semi-automatic way (similarly in spirit to the method of [16]), by modifying the interaction potential which is a function in $[0, 2\pi)$, so formally with an infinite number of parameters (in practice we discretize it with a very large number of points).

1.1. Main results and structure of the paper

In order to help the reader, we start with a summary of the main results, referring to the parts of the paper were they are discussed in detail.

Section 2 is devoted to the definition of the CCP and the corresponding model Hamiltonian, that will allow us to perform a statistical physics study both in the temperature T and in the constraints per variable ratio α. Special attention is devoted to highlight the differences and similarities with the discrete coloring problem (DCP). In particular, given that solutions to the DCP are a subset of the solutions to the CCP, one would expect the latter to have a much larger entropy of solutions and thus being simpler to solve. Our first, unexpected result is that the computation of the phase diagrams we

perform in section 2.2 tells us a different story: the dynamical phase transition in the CCP, when present, happens much before that in the DCP, so there is a large range of α values where finding a solution to DCP is easy, while in CCP ergodicity is already broken and some classes of algorithms may undergo a dynamical arrest in the search for solutions.

We explain the apparent paradox by interpreting the different coloring problems as Bayes optimal inference problems (see section 2.1.3, *planted model*) only differing in the *a priori* distribution, which is flat on $[0, 2\pi)$ in the CCP, while it is the sum of delta functions in the DCP. Although it is true that a flat prior produces a coloring problem with a larger entropy of solutions, it is also evident that it will tend to set variables so that constraints are satisfied in a very loose (i.e. inefficient) way. We conclude section 2 by anticipating the result of our optimization of the model Hamiltonian and showing the new phase diagram in the biased measure.

Section 3 discusses the numerical strategies adopted in order to obtain accurate estimates of the thermodynamic thresholds. In particular, we address the numerical solution via population dynamics of the saddle point equation derived under the one-step replica symmetry breaking (1RSB) ansatz with Parisi parameter $m = 1$. This provides independent estimates of α_d by considering both the divergence of the overlap relaxation time for $\alpha \to \alpha_d^-$ and the birth of instability in the non-trivial solution for $\alpha \to \alpha_d^+$. The equilibrium relaxation time from Monte Carlo (MC) simulations on real instances is also considered, and found in reasonable agreement with the analytical predictions.

In section 4 we present how we solved the problem of maximizing the dynamical phase transition by modifying optimally the interaction potential. With respect to previous works on the subject, where the optimization was performed on very few parameters, we managed to optimize the entire interaction potential, which is a function in $[0, 2\pi)$ (we discretize the interval with a very large number of points for numerical convenience), in a semi-automatic way. We find that the best optimization procedure relies on trying to maximize the complexity slightly above the dynamical threshold. In this way we can increase the dynamical threshold by more than 8%. Finally, we show how the optimization of the dynamical phase transition through the complexity maximization procedure can be performed without the need to solve the belief propagation equations, but only relying on MC measurements.

We relegate to the appendices most of the technical details. In particular, in appendix A we give a detailed interpretation of the planted model of section 2.1.3 as a Bayesian inference problem and define a mixed model which continuously interpolates between the DCP and CCP, while in appendix B we write the belief propagation equations valid in the planted setting.

2. The model

2.1. Definition of the model

In the continuous coloring, N real angular variables $x_i \in [0, 2\pi)$, $i = 1, \ldots, N$ interact via a sparse graph $G = (V, E)$ composed by a set V of variable nodes (vertices), with size $|V| = N$, and a set E of M pairwise interactions between variables, $|E| = M$. In this

work, graphs are random instances drawn from the Erdős–Rényi ensemble, with average connectivity $\alpha = M/N$ (the average degree of variable nodes is hence equal to 2α). Neighbors on the graph have to satisfy the excluded-volume constraint $\cos(x_i - x_j) \leqslant \cos\theta$, for some excluded angle (or 'diameter') $\theta \in [0, \pi)$.

We associate to the CSP a Hamiltonian which exactly counts the number of violated constraints, that in the case of the continuous coloring reads

$$\mathcal{H}_{\text{flat}}(\{x\}) = \sum_{(ij) \in E} \mathbb{I}\left(\cos(x_i - x_j - \omega_{ij}) > \cos\theta\right), \tag{1}$$

where $\mathbb{I}(\ldots)$ is equal to 1 if the condition in the argument is satisfied, and 0 otherwise. The subscript 'flat' refers to the fact that $\mathcal{H}_{\text{flat}} = 0$ on all the solutions to the problem. In the limit $\beta \to \infty$, the associated Boltzmann–Gibbs measure $Z^{-1} e^{-\beta \mathcal{H}_{\text{flat}}}$ is hence uniform over the whole space of solutions, while configurations not simultaneously satisfying all the constraints have zero statistical weight.

The parameters $\omega_{ij} \in [0, 2\pi)$ are quenched random shifts living on each directed edge $i \to j$, with $\omega_{ji} = 2\pi - \omega_{ij}$. The original problem, which is recovered for $\omega_{ij} = 0$, is found to undergo a 'crystallization' phenomenon when increasing connectivity (or packing fraction, by increasing θ), or equivalently when lowering temperature, as already noted in [4]. In the crystal phase, variables condense around a discrete set of values on the interval $[0, 2\pi)$, mimicking a q'-coloring packing, with $2\pi/q' > \theta$. On a real tree, with open boundary conditions, the homogeneous $\omega_{ij} = 0$ and non-homogeneous models are equivalent, since the latter can be mapped onto the former by the following transformation: $\forall (ij) \in E$, $\omega_{ij} \to 0$, $x_j \to x_j + \omega_{ij}$ and $x_k \to x_k + \omega_{ij}$ for all the $\{x_k\}$ belonging to the subtree starting from j (that is excluding the whole branch containing i). However, in the case of random graphs, where loops are always present (even if very large), or on a finite tree with fixed boundary conditions on the leaves, this construction is not guaranteed to hold for every choice of the ω_{ij}'s: loops, together with random shifts, induce frustration of the periodic order.

In order to suppress the crystal phase [2], adopted a small degree of polydispersity. Here we will follow two strategies. In the analytical derivation of the 1RSB belief propagation equations for the homogeneous case $\omega_{ij} = 0$, we will impose translational (rotational) invariance: this allows to simplify the equations when the replica parameter is equal to $m = 1$, obtaining an RS-like scheme at the price of introducing *planted* random shifts in the messages (see appendix B for details on the derivation). By imposing translational invariance, one can disregard the ordered solution while maintaining the tree-like recursive structure of the BP equations. On the other hand, we will also support the BP predictions with direct numerical MC simulations of the non-homogeneous model with planted random shifts in section 3.2.

The parameter $\theta \in [0, \pi)$ represents half the excluded angle around each 'particle', playing the same role of the diameter in hard spheres systems. In order to make direct contact with the very well known q-coloring, it is useful to consider 'discretized' values of θ defined as $\theta = \frac{2\pi}{q}$, with $q \in \mathbb{N}_{>1}$. The meaning of this relation is straightforward: in the q-coloring, colors $u \in \{1, \ldots, q\}$ can be associated to q-states Potts angular variables $x = \frac{2\pi}{q}(u - 1)$, naturally satisfying $\cos(x_i - x_j) \leqslant \cos(2\pi/q)$ if $x_i \neq x_j$. In the following, we will express the diameter θ in terms of integer q's. This mapping allows for the
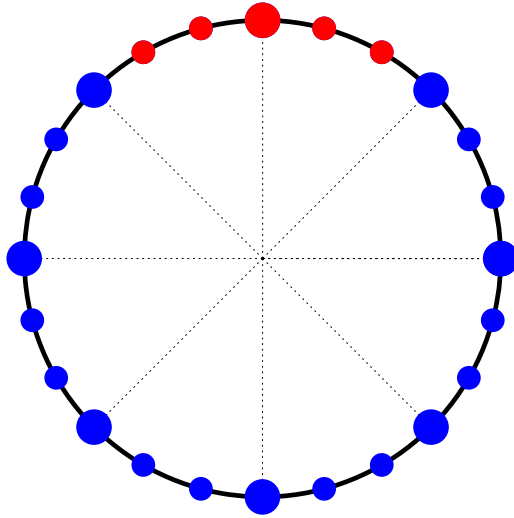
Figure 1. Discretization of precision $d = 3$ for the continuous coloring. The excluded angle (on each side of the particle) is equal to $\theta = 2\pi/q$ in the continuous limit, with $q = 8$. As a consequence of discretization, $2d - 1$ states are forbidden (red points), among the total $p = qd = 24$ possible ones. Big dots, together with the dashed lines, are meant only to identify the q multiples of θ along the circumference. The increase of precision d with q fixed corresponds to the proliferation of small dots inside each sector.

following observation: given a solution to the discrete coloring with q colors, it also represents a solution to the continuous coloring with excluded angle $\theta = \frac{2\pi}{q}$ (actually, the solution is valid for any $\theta' \leqslant \frac{2\pi}{q}$). Conversely, the set of solutions to the continuous coloring for $\theta = \frac{2\pi}{q}$ contains all the solutions to the q'-coloring with $q' \leqslant q$.

2.1.1. Discretization. In order to numerically solve the BP equations, a discretization of the interval $[0, 2\pi)$ is needed. To this end, we introduce a clock approximation with p states by imposing

$$x_i = \frac{2\pi}{p} t_i, \quad t_i \in \{0, \dots, p - 1\}, \quad \forall\, i \in V. \tag{2}$$

Calling d the number of clock states inside the excluded region $0 \leqslant x < \theta$, i.e the discretization precision, then it follows $p = qd$ (as we said, we restrict to integer q for convenience). Each variable precludes $2d - 1$ values to its neighbors, as depicted in figure 1.

2.1.2. Biased measure. A CSP can be recast as the study of the $\beta \to \infty$ limit of a Boltzmann–Gibbs distribution which in some generality takes the form

$$P(\{x_i\}) = \frac{1}{Z(\beta)} \exp\left(-\beta H_0(\{x_i\}) + H_1(\{x_i\})\right). \tag{3}$$

The function H_0 enforces the hard constraints and is defined by the condition to be equal to zero on satisfied configurations *only*. The simplest choice, that we will adopt

throughout the paper, is $H_0 = \mathcal{H}_{\text{flat}}$ given by equation (1). Note that a different definition of H_0 (e.g. a linear or quadratic potential) only affects the behavior of the model as long as temperature is involved or, even at zero temperature, in the UNSAT phase $\alpha > \alpha_{\text{sat}}$, but does not influence the equilibrium $\beta \to \infty$ limit in the SAT phase.

We call $H_1 = O(1)$ for $\beta \to \infty$ the 'soft' part of the interaction, as it alters the statistical weight of only the acceptable configurations when $\beta \to \infty$. In this way, it is possible to sample such configurations with a measure that is not necessarily flat (uniform over the solutions), as in the case of $H_1 = 0$. We limit ourselves to the introduction of a first-neighbours bias as in [13]. This implies that our bias factorizes, and one only needs to replace the local interaction term $f^{\text{flat}}(x_i, x_j) \propto \exp[-\beta\mathbb{I}\left(\cos(x_i - x_j - \omega_{ij}) > \cos\theta\right)]$ in all the equations with a suitable biasing function $f(x_i, x_j)$, which for convenience we also take to satisfy translational invariance as the original flat Hamiltonian H_0. From the practical point of view, then, there is no substantial difference in solving the biased problem rather than the original one, and all the relevant equilibrium transition lines can be straightforwardly computed for any choice of the biased interaction f.

In this paper, we are particularly interested in the so called dynamic or clustering transition α_{d}. By tuning f, i.e. H_1, one can move the dynamical threshold to higher α values up to $\alpha_{\text{d}}^{\text{opt}} = \alpha_{\text{d}}(f^{\text{opt}})$. On the contrary, the location of the SAT/UNSAT transition, that is where the volume of the configurations that satisfy all the constraints becomes zero in the thermodynamic limit, does not depend on H_1 nor on the particular choice of H_0 by definition.

2.1.3. Planted model. Throughout this work, we will make extensive use of the planting technique [17]. According to this procedure, one first draws a planted configuration $\{x_i^0\}$ from some prior probability distribution, which is assumed to be factorized, $P_X(\{x_i\}) = \prod_i P_X(x_i)$. Then, a random graph can be constructed in two ways, depending on whether random shifts are explicitly considered or not. In the latter case, one can run through all the pairs of vertices (ij) and add an edge with probability $2\alpha f(x_i^0, x_j^0; \beta)/(N-1)$, where $f(x_i, x_j; \beta) = e^{-\beta\mathcal{H}(x_i,x_j)}/\int dx\, dy P_X(x) P_X(y) e^{-\beta\mathcal{H}(x,y)}$ and $\mathcal{H}(x_i, x_j)$ is the pairwise term in the Hamiltonian of the model (it can refer to $\mathcal{H}_{\text{flat}}$ given by equation (1), or to any biasing function as well). This ensures that the average degree c of variable nodes is always equal to 2α

$$c = (N-1)\int dx\, dy P_X(x) P_X(y) \frac{2\alpha f(x,y)}{(N-1)} = 2\alpha. \tag{4}$$

However, we would also like different nodes to follow the same degree distribution. To this end, one can derive the average degree $c(x)$ conditional on the knowledge that $x_i^0 = x$,

$$c(x) = (N-1)\int dy\, P_X(y) \frac{2\alpha f(x,y)}{(N-1)} = 2\alpha \int dy\, P_X(y) f(x,y). \tag{5}$$

Then, we must also ask for $\int dy\, P_X(y) f(x,y) = 1$, so that $c(x) = 2\alpha$ independently from x. One also recognizes that a bona fide Erdős–Rényi random graph ensemble is recovered for $f = 1$.

In the presence of random shifts, we can plant the system in a slightly different manner: given the configuration $\{x_i^0\}$ and a graph realization extracted from the proper distribution (our choice is the Erdős–Rényi ensemble), we associate a random variable $\omega_{ij} \in [0, 2\pi)$ to each directed edge $i \to j$, and $\omega_{ji} = 2\pi - \omega_{ij}$ on $j \to i$, with probability $p_{ij}(\omega_{ij}) = e^{-\beta \mathcal{H}(x_i - x_j - \omega_{ij})} / \int d\omega\, e^{-\beta \mathcal{H}(x_i - x_j - \omega)}$. In the following, we will use this second method in order to plant the system when running MC numerical simulations (we indeed require random shifts to suppress crystallization). Notice that one can choose, without loss of generality, $\{x_i^0\} = 0$, this being equivalent to a set of local gauge transformations (valid for any i): $x_i \to x_i - x_i^0$, $\omega_{ij} \to \omega_{ij} - x_i^0 \; \forall j \in \partial i$.

Both the planted models are different from the original ones (with or without the random shifts), since planting a solution changes the graph ensemble, and, generally, its properties. This is particularly evident in the $\beta \to \infty$ limit: via planting, one can always construct for arbitrary α, even beyond the SAT/UNSAT threshold, a pair graph-configuration of exactly zero energy. However, before the condensation transition α_c, the planted ensembles are equivalent (provided that the original random model displays a uniform paramagnetic BP fixed point) to the original ones: this is known as *quiet planting* [17]. We can understand this fact by considering that the superimposition of a single planted cluster in the region $\alpha_d < \alpha < \alpha_c$, which is already dominated by exponentially many other clusters, should be thermodynamically undetectable as long as the planted cluster exhibits the properties of typical $m = 1$ clusters from the random ensemble. This provides a very powerful tool to study the average properties of the whole clustered phase $\alpha_d < \alpha < \alpha_c$ by focusing on a single, easy to build planted cluster.

Finally, both the DCP and CCP can be derived from the same planted model of Hamiltonian (1) by allowing for a different definition of the prior distribution P_X over $x \in [0, 2\pi)$, namely $P_X(x) = \frac{1}{q}\sum_{k=0}^{q-1}\delta\left(x - \frac{2\pi k}{q}\right)$ for DCP and $P_X(x) = \frac{1}{2\pi}$ for CCP, where q is the number of colors. This enables us to treat them in a unified Bayesian inference setting by defining a mixed model in appendix A which continuously interpolates between DCP and CCP. We also give an explicit definition of the interaction function $f(x_i, x_j; \beta)$ in all the different cases, also taking into account the discretization of the interval $[0, 2\pi)$.

2.2. Phase diagram

2.2.1. Transitions in random CSP. One of the achievement of the spin glass theory is the understanding of the nature and the structure of the phase space of CSPs as the number of constraints per variable α increases [6]. This description allows us to understand both the random model and the planted model, which for $\alpha < \alpha_c$ have the same physical properties. For simplicity, we will describe this phase diagram in the case of a system that presents both a paramagnetic solution to the cavity equations along with a RFOT.

- For $\alpha < \alpha_d$ the Gibbs measure is a Bethe measure [18]. For both the planted and the random ensembles there is only one fixed point to the BP equations, the paramagnetic fixed point, and sampling from the Gibbs distribution can be achieved by MC algorithms in polynomial time. The threshold α_d is usually referred to as dynamical

or clustering transition, because at α_{d} the phase space breaks into an exponential number of pure states (or clusters), which MC is unable to sample uniformly in a polynomial time.

- For $\alpha_{\mathrm{d}} < \alpha < \alpha_{\mathrm{c}}$ the Gibbs measure is decomposed into an exponential number of pure states, each one corresponding to a different fixed point to the BP equations. BP initialised in the planted solution will converge to a fixed point that is different from the paramagnetic fixed point. The complexity Σ is a measure of the number of pure states that dominate the Gibbs measure. In practice there is no known way to sample one of these pure states in polynomial time. Nevertheless, one can estimate the complexity as

$$N\Sigma = F(\{\nu_{\mathrm{para}}\}) - F(\{\nu_{\mathrm{planted}}\}), \tag{6}$$

where F is the Bethe free entropy as in equation (B25), $\{\nu_{\mathrm{para}}\}$ is the paramagnetic BP fixed point (we denote by $\{\nu\}$ a set of BP messages) and $\{\nu_{\mathrm{planted}}\}$ is the fixed point obtained by initializing BP in the planted solution. As α increases Σ diminishes up to the point α_{c}, the condensation transition, where $\Sigma = 0$.

- For $\alpha > \alpha_{\mathrm{c}}$ it is possible to distinguish between the planted and the random model:

  * in the random model the Gibbs distribution is dominated by the sum of a sub-exponential number of Bethe measures [18]. The Gibbs distribution is not a Bethe measure anymore.

  * In the planted problem the Gibbs distribution is dominated by one Bethe measure (up to some symmetry) corresponding to the planted solution.

In this paper we make use of the planting technique in order to derive the position of the dynamical transition α_{d}. This is equivalently done by initializing MC in a planted state or by solving the 1RSB $m = 1$ cavity equations discussed in appendix B. Other important thresholds are the following:

- α_{KS}: it is the value of α for which the paramagnetic solution ceases to be stable. It is known as the Kesten–Stigum transition in the tree reconstruction literature [19–21], or also as the de Almeida–Thouless local instability of the RS solution in the context of spin glasses [22]. We compute it analytically at the end of this section.

- The *rigidity* transition, that in CSPs with discrete variables, such as the q-coloring, signals the point beyond which all the dominant pure states contain a finite fraction of *frozen* variables. A variable is frozen if it can take only one value in all the configurations within a given cluster. This transition can be located before or after α_c.

- The coloring threshold (COL/UNCOL or more generally SAT/UNSAT transition): it is the value of α beyond which no proper solution exists with probability one in the thermodynamic limit.

The stability analysis of the paramagnetic solution can be carried out and one finds that the Kesten–Stigum transition of the system is linked to the top eigenvalue of the following matrix [23]

$$T_{ab} = n_a \left(f_{ab} - 1\right), \tag{7}$$

where $\{n_a\} \in \mathbb{R}^p$ is our discretized notation for the prior $P_X(x)$ and the matrix $\{f_{ab}\} \in \mathbb{R}^{p \times p}$ is the discretized version of the interaction function $f(x, y; \beta)$, see appendix A.2.3 for details. The instability condition of the paramagnetic fixed point is found to be

$$2\alpha\lambda^2 > 1, \tag{8}$$

where λ is the top eigenvalue of T. For the DCP and CCP the top eigenvalue of T can be derived explicitly. This yields us with the following transition lines

$$\alpha_{\mathrm{KS}}^{\mathrm{DCP}} = (q-1)^2/2, \tag{9}$$

$$\alpha_{\mathrm{KS}}^{\mathrm{CCP}} = \frac{1}{2} \left[\frac{\sin(\pi(1 - 2/q))}{\pi(1 - 2/q)} \right]^{-2} \underset{q \gg 1}{=} O(q^2/8), \tag{10}$$

$$\alpha_{\mathrm{KS}}^{d-\mathrm{CCP}} = \frac{1}{2} \left[\frac{\sin(((q-2)d+1)\pi/(dq))}{\sin(\pi/(dq))((q-2)d+1)} \right]^{-2}. \tag{11}$$

Equation (9) represents the Kesten–Stigum transition line for the q-coloring, as was already derived in [1]. Equation (10) is the Kesten–Stigum transition line for the continuous coloring model where $\theta = 2\pi/q$, if one could deal with continuous messages. Finally, equation (11) represents the Kesten–Stigum transition line for the discretized version (2) of the continuous coloring, for arbitrary values of q and of the discretization precision d. By taking $d = 1$ or $d \to +\infty$, one respectively recovers (9) and (10). It is worth noticing the presence of a factor $1/4$ between the Kesten–Stigum bounds for the continuous and the discrete coloring, $\alpha_{\mathrm{KS}}^{\mathrm{CCP}} \approx \alpha_{\mathrm{KS}}^{\mathrm{DCP}}/4$.

2.2.2. Uniform measure: continuous vs q-coloring. In this section we discuss the phase diagram of the continuous coloring, as obtained by numerically solving the 1RSB $m = 1$ population dynamics equations up to α_{c}. This means that we will focus on the detection of the dynamic and condensation thresholds $\alpha_{\mathrm{d}} = \alpha_{\mathrm{d}}(T = 0)$ and $\alpha_{\mathrm{c}} = \alpha_{\mathrm{c}}(T = 0)$ by studying the formation of *typical* clusters (with respect to the uniform measure under consideration). The same analysis will be performed in temperature in the next section, allowing the calculation of the transition lines $\alpha_{\mathrm{d}}(T)$ and $\alpha_{\mathrm{c}}(T)$, or equivalently $T_{\mathrm{d}}(\alpha)$, $T_{\mathrm{c}}(\alpha)$. All the transition lines for the continuous coloring here presented are computed using a discretization $d = 10$; we refer the reader to section 3.3 for an analysis on the corrections to the continuous limit, that are shown to scale as $1/d$.

The three $T = 0$ transition lines α_{d}, α_{c} and α_{KS} are shown in figure 2 as a function of q for both the q-coloring on the left and the discretized version of the continuous coloring ($d = 10$) on the right. Looking at the values of α in the figure, it is evident how all the transitions appear in the continuous coloring before (i.e. at lower α) than in the DCP. This is interesting and somehow surprising, as it results that despite *enlarging* the set of solutions for each value of α (all of the solutions to the q-coloring also satisfy the continuous problem with $\theta = \frac{2\pi}{q}$ by definition), the clustering point is anticipated. As a practical consequence, since the ergodicity breaking point α_{d} is generally connected (even if it does not rigorously coincide) with the onset of hardness in the constraint satisfaction optimization problem, this would imply for $q > 14$ the existence of a whole α region

(a) q-coloring

(b) Continuous coloring, $d = 10$
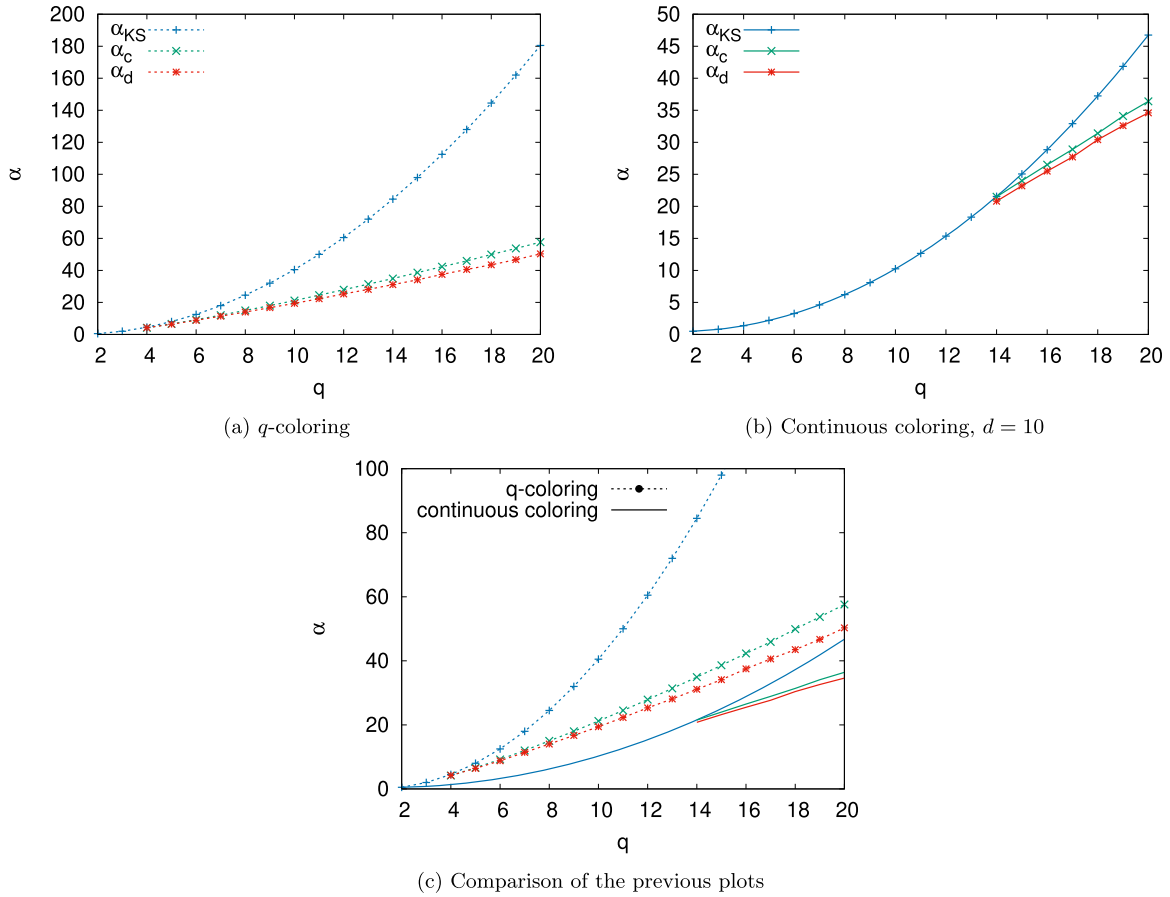

(c) Comparison of the previous plots

Figure 2. $T = 0$ phase diagram in the (q, α) plane for both the discrete (a) and continuous coloring (b). (c) A direct comparison between the two models. In the discrete coloring q can only take integer values, the dotted transition lines in the figure being just meant to guide the eye. The discrete model exhibits for $q \geqslant 4$ a RFOT (the three transition lines merge at $\alpha = 2$ only for $q = 3$) [1]. In the (discretized) CCP, a RFOT is found for $q \geqslant 14$, where q is defined by the relation $\theta = 2\pi/q$. This is however just an upper bound to the actual point where the three transition lines are expected to merge. This point is likely to be located at a value of q that is not an integer. We notice that, for a given q, all the transitions in the continuous model exhibit a lower value of α than in the q-coloring, see also the discussion in the text.

where local algorithms searching for solutions to the continuous coloring get stuck and fail, while solutions to the discrete model are less and yet still easy to find.

We interpret this fact by arguing that in the q-coloring the discrete prior forces the solutions to satisfy all the constraints in a more tight, i.e. efficient, way. In this case, variables can only take q values, and the angular distance between two neighbors is either 0 (violated constraint) or a multiple of $2\pi/q$. The minimum allowed distance between variables is hence $2\pi/q$, which corresponds to a *contact* in the particle-system jargon. The fraction of contacts is in this case very relevant: the equilibrium probability distribution of (discrete) angular differences in the paramagnetic phase, which is valid

for $\alpha < \alpha_c$, is by definition equal to 0 if two angles take the same values and it is uniform otherwise, so that the probability of having a contact is simply equal to $2/(q-1)$, where the factor 2 comes from taking into account both the possibilities to form a contact in one dimension, $x \pm 2\pi/q$. This same analysis can be repeated in the case of continuous coloring. For clarity, let us consider the discretized version with $p = qd$ possible states. The fraction of exact contacts now becomes $2/(p - 2d + 1) \sim 2/[d(q-2)]$, which is as expected heavily suppressed for $d \to \infty$. In this situation, one should rather ask for an angular interval $\Delta \equiv n\frac{2\pi}{p}$ with associated probability $1/(q-1)$ on each side of the 'particle' by integrating the previous probability. In our discretized notation we can write $\frac{n}{d(q-2)+1} = \frac{1}{q-1}$, from which it follows $\Delta \approx \frac{2\pi}{q}\frac{q-2}{q-1} \approx \frac{2\pi}{q}$. In the continuous coloring, typical solutions satisfy the constraints in a very loose way: exact contacts are 'smeared' over an interval comparable to the diameter of the particles. Finally, notice that reversing this perspective the discrete model can be considered, for some qualitative aspect, akin to a continuous sphere system where particles are encouraged to stick together. It is indeed very well known in the literature that, by adding a short range attraction to a hard core repulsion, one can greatly extend the liquid phase of the system [24]. This phenomenon is at the very heart of the results from the biased thermodynamics approach as discussed in the following.

From figure 2 it is also evident that the region where the transition is continuous (small q region) results much more extended in the continuous coloring than in the discrete problem: the discrete q-coloring on Erdős–Rényi random graphs is known to display a RFOT already for $q > 3$ [1], while in the continuous coloring the continuous transition ranges up to $q \approx 14$. The actual point where the transition changes its nature is in this case likely located at a value of q which is not an integer, close to but smaller than $q = 14$ (we only have points for integer q's). Below this point the three transition lines do coincide. In particular, since $\alpha_d = \alpha_c$, the phase dominated by an exponential number of clusters is missing.

2.2.3. Biased measure for the CCP. We have performed a computation analogous in spirit to the one of [16] in order to postpone as much as possible the location of the $T = 0$ dynamical transition α_d. Unfortunately, we lack in this case an analytical expression for α_d. A possible way round is to consider a gradient descent with respect to some observable which is hopefully related with the location of α_d: we consider the maximization of the complexity Σ, as explained in section 4. We will call the resulting optimized threshold α_d^{opt}, and the optimized interaction f^{opt}. We stress, however, that α_d^{opt} should be more safely interpreted as a lower bound to the actual optimal threshold, since there is still no rigorous proof linking the maximization of the dynamical transition with the one of Σ. On the other hand, the gradient of the complexity will be shown to be relatively simple to compute, and also accessible through MC sampling. If the complexity-maximization approach actually turns out to be valid in more general models with an RFOT, then we would have in our hands a very powerful and versatile tool to compute the optimized interaction also in more computer-memory consuming settings, such as in higher spatial dimensions, where BP really struggle.

The optimized interaction f^{opt} is given in figure 3 for the usual choice of $q = 20$ and a discretization of precision $d = 10$ ($p = 200$ clock-states). The interaction f is in our
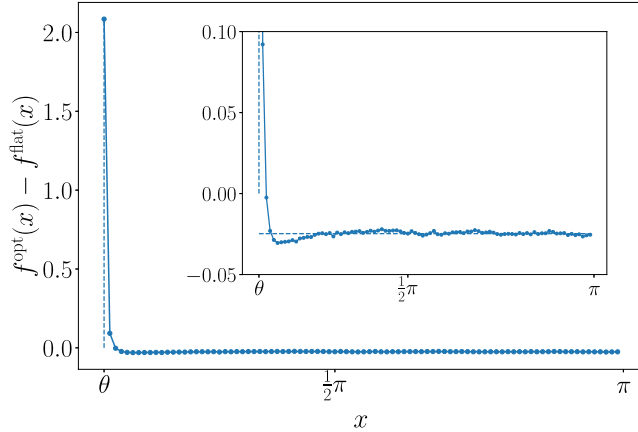
Figure 3. Soft part $|x| \geqslant \theta$ of the optimized interaction at $T = 0$ for the continuous coloring ($\theta = 2\pi/q$, with $q = 20$) as a function of the interparticle angular distance. The function is symmetric for $x \to 2\pi - x$. The difference $f^{\mathrm{opt}} - f^{\mathrm{flat}}$ (f^{flat} being a simple step function around $x = \theta$) does not go precisely to zero at large angles because we are subtracting two normalized probabilities, both satisfying $\sum_a f_a = p$ in the discretized case. The inset shows a zoom.

setting proportional to the Boltzmann–Gibbs weight, rather than to an energy, for this reason a peak in the probability for $x = \theta$ has to be interpreted as a very short range attraction favoring contacts between angles/particles. The fact that by perturbing the hard spheres Hamiltonian with a short-range attractive potential one can dramatically extend the liquid phase of the system is very well known in the literature [15, 16, 24–26]. This is also very relevant for soft matter colloidal systems, where such interaction potentials can be experimentally engineered [27, 28]. With our choice for the potential, the zero temperature dynamic threshold estimated from BP is moved from $\alpha_{\mathrm{d}}^{\mathrm{flat}} = 34.63(2)$ to $\alpha_{\mathrm{d}}^{\mathrm{opt}} = 37.71(1)$, thus exhibiting an increase of about 8%.

The numerical optimization has been performed for $T = 0$, when the hard part of the interaction f (for $|x| < \theta$) is exactly zero. By taking the logarithm of the soft part of f (for $\theta \leqslant x \leqslant 2\pi - \theta$), one can obtain the pairwise contribution to the biasing Hamiltonian H_1 up to an additive constant, due to the arbitrariness in the normalization. When switching on temperature, the hard core excluded volume interaction is softened, the cost for violating a constraint being now proportional to $\exp(-\beta)$. The soft part of the interaction should instead be independent from β in our approach, apart for a global β-dependent normalization. In the following, we will always adopt implicitly the definition $f^{\mathrm{opt}}(x; \beta) \propto A\,\mathrm{e}^{-\beta}\mathbb{I}[|x| < \theta] + f_{T=0}^{\mathrm{opt}}\mathbb{I}[\theta \leqslant x \leqslant 2\pi - \theta]$. We fix the relative amplitude A of the hard part with respect to the soft part of the interaction f by requiring (using the notation for the discretized model)

$$\frac{\sum_{(-\theta; \theta)} f_a^{\mathrm{opt}}}{\sum_{[\theta; 2\pi-\theta]} f_a^{\mathrm{opt}}} = \frac{\sum_{(-\theta; \theta)} f_a^{\mathrm{flat}}}{\sum_{[\theta; 2\pi-\theta]} f_a^{\mathrm{flat}}} = \frac{(2d-1)\mathrm{e}^{-\beta}}{p - (2d-1)}. \tag{12}$$

With this choice, the annealed energy $e_{\mathrm{ann}}(\beta)$, counting the number of violated constraints per spin as a function of temperature in the paramagnetic phase, is equal for
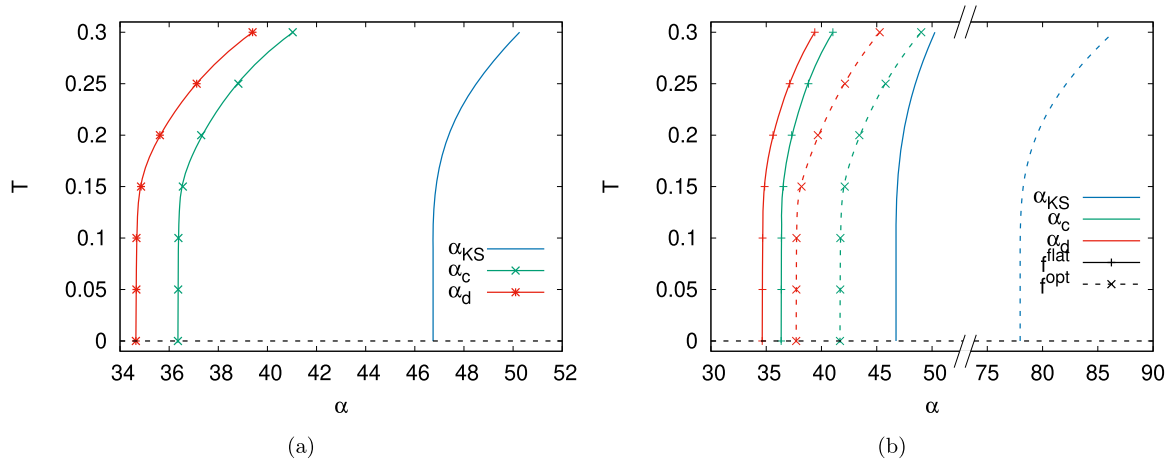
Figure 4. Phase diagram in the (α, T) plane for the continuous coloring with fixed $q = 20$ and discretization precision $d = 10$. The dynamic and condensation thresholds are obtained by numerically solving the BP equations for different values of T (points in the figure). (Left) Uniform measure. (Right) Comparison between the two choices for the interaction f^{flat} (solid) and f^{opt} (dashed). The $T = 0$ dynamical transition is moved from $\alpha_{\text{d}}^{\text{flat}} = 34.63(2)$ to $\alpha_{\text{d}}^{\text{opt}} = 37.71(1)$.

both f^{flat} and f^{opt} to

$$e_{\text{ann}}(\beta) \equiv \frac{M}{N} \frac{\sum_a f_a \mathbb{I}\left[\cos\left(\frac{2\pi a}{p}\right) > 0\right]}{\sum_a f_a} = \frac{\alpha(2d-1)\text{e}^{-\beta}}{(2d-1)\text{e}^{-\beta} + p - (2d-1)}. \tag{13}$$

The phase diagram in the (α, T) plane is shown in figure 4, solid lines corresponding to f^{flat}, while dashed ones referring to f^{opt}. We stress that in our case temperature tunes the strength of the violated constraints (we recover the infinite hard sphere potential only in the limit $\beta \to \infty$), and thus plays a fundamentally different role than, for instance, in [15, 16]. On the right, we directly compare the phase diagram according to the uniform measure (solid lines) with the one obtained by using the optimized interaction f^{opt} (dashed lines). It is interesting to notice that also the Kesten–Stigum bound α_{KS} is sensitive to the modification of the interaction, and it actually increases the most with respect to the other thresholds, going from the $T = 0$ value $\alpha_{\text{KS}} = 46.743$ to $\alpha_{\text{KS}} = 77.972$, with an increase of more than 60%. The location of α_{KS} is particularly relevant for the associated Bayesian inference problem, since it signals the α-point below which inference is practically unfeasible. In order to improve the performance of inference protocols, one should probably try to optimize the interaction in the opposite direction (lowering α_{KS}).

3. Numerical determination of the dynamic transition

In this section, the question of how to obtain the most accurate numerical estimation of α_d is addressed. The typical way of detecting α_d is by verifying the existence of a non-trivial solution to the planted population dynamics BP equations: usually, one starts from $\alpha > \alpha_d$ and decreases α until the non-trivial solution is lost. This method, however, is prone to inaccuracies. On the one hand, because of the finite number of BP iterations one can perform, one may simply not give the algorithm enough time to forget the planted solution and conclude that the non-trivial fixed point survives for values of α slightly smaller than α_d, thus leading to an underestimation of α_d. On the other hand, fluctuations due to the finite size of the population may become relevant close to α_d, where the non-trivial fixed point looses its stability, this leading to an early loss of the fixed point and to a consequent overestimation of α_d.

For these reasons, a more robust approach taking into account how things scale *around* α_d is very desirable. Following [13], we give independent, compatible estimations of α_d coming from both below and above the dynamic threshold, by studying respectively the (power-law) divergence of BP relaxation time and the birth of instability in the non-trivial fixed point.

The situation is more complicated when one considers real instances of the problem, where the finite size of the system determines the presence of loops in the graph and in general smooths the transitions, which are sharp only in the thermodynamic limit. For large enough sizes, however, we certainly expect the population dynamics BP predictions to be of some use. This can be directly tested through MC simulations, where we observe an MCT-like behavior of the dynamic correlation (the overlap) compatible with $\alpha_d \equiv \alpha_d^{BP} = 34.63(2)$, at least up to timescales where finite size effects are negligible. In this section we present the results of a detailed analysis at $T = 0$, the same procedure being straightforwardly applicable to any temperature.

3.1. BP population dynamics

In figure 5 we show the data from the numerical solution of population dynamics equations for the Erdős–Rényi ensemble. The size of the population here considered is $U = 10^6$, each message being a probability array of $p = 200$ values ($q = 20$, $d = 10$). We choose the value $q = 20$ because, as it is suggested from figure 2, the transition should be in this case appreciably first order. On the left side, the overlap decay $q(t)$ is displayed for different values of $\alpha < \alpha_d$. Given a set of U marginals $\{\nu_i\}^t$, the overlap is defined as (using for convenience continuous notation)

$$q(t) = \frac{1}{U} \sum_{i=1}^{U} \int dx_1\, dx_2\, \nu_i(x_1)\nu_i(x_2) \cos(x_1 - x_2)$$

$$= \frac{1}{U} \sum_{i=1}^{U} \left[\langle \cos(x) \rangle_i^2 + \langle \sin(x) \rangle_i^2 \right], \tag{14}$$
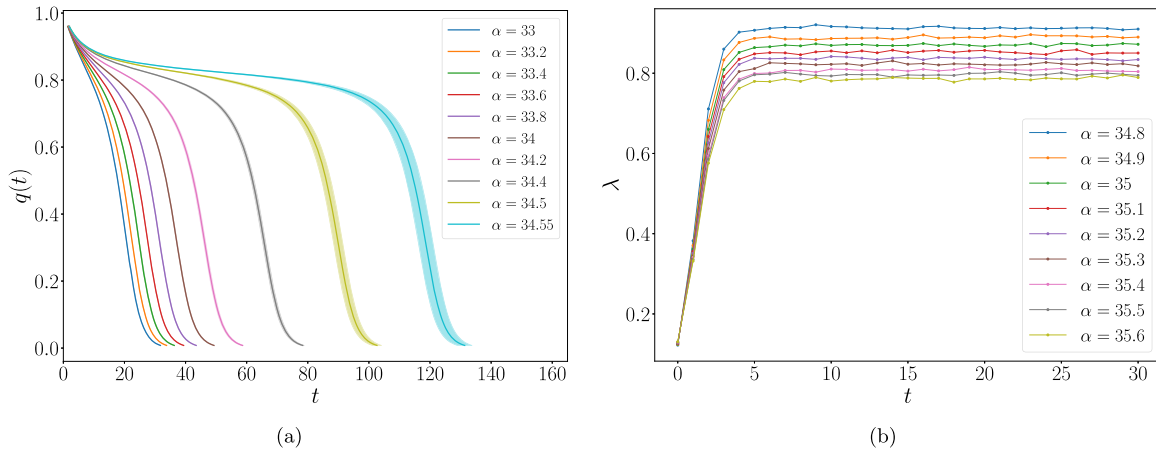
Figure 5. (Left) Overlap as a function of the number t of BP iterations. As long as $\alpha < \alpha_{\rm d}$, the overlap decays to zero, thus implying that any memory of the planted initialization is lost and one recovers the paramagnetic solution with uniform messages. (Right) Ratio $\lambda(t)$ between the magnitude of perturbations to the non-trivial solution at times $t + 1$ and t for $\alpha > \alpha_{\rm d}$. After a transient, stationarity is reached, thus identifying the stability parameter $\lambda(\alpha) < 1$.

where $\langle \cdot \rangle_i$ represents the average over the marginal distribution ν_i. We considered 10 different initializations of the random generator, which we refer to as samples. Each sample has a very smooth behavior, almost identical to the others except for a slightly difference in the time it leaves the forming plateau. For this reason, the most natural way to present the data is by averaging over t at fixed q, once a spline interpolation of the single sample is performed.

On the right side, we consider the case $\alpha > \alpha_{\rm d}$, where the BP equations admit a non-trivial solution with $\lim_{t \to \infty} q(t) \neq 0$. Once convergence in the overlap is reached, we compute the ratio of the magnitude between two subsequent perturbations around the fixed point (by using the linearized version of the BP equations derived in appendix B.3), as a function of time

$$\lambda(t) = \frac{\|\epsilon(t + 1)\|}{\|\epsilon(t)\|}, \tag{15}$$

where we adopted the L_2 norm of the perturbation parameters averaged over the population, $\|\epsilon\| = \frac{1}{pU} \sum_{i=1}^{U} \sum_{a=0}^{p-1} \epsilon_{i,a}^2$. After a sufficient but finite number of steps, $\lambda(t)$ is actually selecting the contribution from the biggest in modulus eigenvalue of the linearized matrix which enforces the BP iterative procedure (see [13] for a detailed discussion), being the slowest mode to decay. We define the stability parameter $\lambda(\alpha)$ as essentially the modulus of this eigenvalue, with $\lambda(\alpha) < 1$ for the solution to be stable. In practice, we average over t the $\lambda(t)$ data of figure 5(b) once the plateau is reached. In this case we consider just one sample for each value of $\alpha > \alpha_{\rm d}$.

Let us first consider the region $\alpha < \alpha_{\rm d}$. From figure 5(a), it is clear that approaching $\alpha_{\rm d}$ the overlap relaxation becomes increasingly slower. Following [13], this behavior

can be directly connected to the discontinuous bifurcation occurring to the fixed-point solution of the BP equations, which is of the form $\mathbb{P} = \mathcal{F}(\mathbb{P}, \alpha)$. To fix the ideas, it is convenient to consider an analogous relation in the scalar case, $q = f(q, \alpha)$, where q is real (the similarity with the overlap, which is indeed our scalar representative for $\mathbb{P}$, will be readily evident). Assuming that $q \geqslant 0$ and that $q = 0$ is always a solution, the dynamic threshold α_{d} is defined as the subitaneous birth of a second solution $q_{\mathrm{d}} = f(q_{\mathrm{d}}, \alpha_{\mathrm{d}})$, with $q_{\mathrm{d}} > 0$. Our problem reduces to the study of the dynamical process $q^{(t+1)} = f(q^{(t)}, \alpha)$, with initial condition $q^{(t=0)} > q_{\mathrm{d}}$ (in analogy with the planted initialization). In this context, it can be proved that the number of steps t around the plateau q_{d} diverges as $\tau \sim K(\alpha_{\mathrm{d}} - \alpha)^{-\frac{1}{2}}$ to leading order for $\alpha \to \alpha_{\mathrm{d}}^{-}$, where K is fixed by some derivative of $f(q, \alpha)$ at the special point $(q_{\mathrm{d}}, \alpha_{\mathrm{d}})$.

In order to test this prediction on our real data, we operatively define a family of relaxation times $\tau_{q^*}(\alpha)$ representing, for any given threshold $q^* < q_{\mathrm{d}}$, the time needed for the overlap $q(t)$ to decrease down to q^*. By looking at the data in figure 5(a), one notices that different sets of $\tau_{q^*}(\alpha)$ for different q^* should essentially differ by only a finite amount of steps, which is reasonably constant in α for any given q^* since the curves have approximately the same behavior (same slope) on a linear x-scale. However, if we also consider the relatively small values of t needed to complete the decay and hence of $\tau_{q^*}(\alpha)$, it follows that relevant subleading corrections to the asymptotic scaling may be due just to the choice of q^*. This is evidenced by the behavior of the first function that we use to fit the data, neglecting any scaling correction,

$$\tau^{(1)}(\alpha) = \frac{A}{(\alpha_{\mathrm{d}} - \alpha)^{1/b}}. \tag{16}$$

This function depends on three free parameters: α_{d}, b and A. Even though at first glance it may appear to reasonably fit the data, see figure 6(a) (colors), all the values of the parameters, presented in table 1 (left), exhibit some dependence on q^*, which is very noticeable in the case of the exponent b (that was argued to be $b = 2$). To account for subleading corrections, we also consider the two following power laws, still depending on three parameters,

$$\tau^{(2)}(\alpha) = \frac{B}{(\alpha_{\mathrm{d}} - \alpha)^{1/2}} + C, \tag{17}$$

$$\tau^{(3)}(\alpha) = \frac{D}{(\alpha_{\mathrm{d}} - \alpha)^{1/2}} + \frac{E}{(\alpha_{\mathrm{d}} - \alpha)^{1/3}}. \tag{18}$$

The rescaled data is presented respectively in figures 6(b) and (c). For what concerns the results of fit $\tau^{(2)}$, see again table 1, we obtain good χ^2 and a much less q^*-dependent slope (i.e. B) especially for intermediate values of $0.1 \leqslant q^* \leqslant 0.5$. We conclude that in this range the scalar bifurcation scaling form with exponent $b = 2$ adequately describes our $\tau_{q^*}(\alpha)$ data. On the other hand, the function $\tau^{(2)}$ is shown to fail for values of q^* close to the plateau, $q^* = 0.8, 0.7$ (compare the values of χ^2 in table 1). In this case fit $\tau^{(3)}$ does a better job, thus suggesting that the scaling $b = 2$ is just apparently altered by some non-trivial subleading corrections. In general, we observe that function $\tau^{(3)}$, even though outperformed by $\tau^{(2)}$ when q^* takes the central values, always performs better than fit $\tau^{(1)}$, apart for the case $q^* = 0.01$ for which already $\tau^{(1)}$ produces a value of b

Table 1. Results from fit (16) (left), (17) (center) and (18) (right) for the divergence of the overlap relaxation time $\tau_{q^*}(\alpha)$ as obtained from population dynamics.

|	$\tau^{(1)}$					$\tau^{(2)}$					$\tau^{(3)}$			
q^*	α_{d}	b	A	χ^2/dof	q^*	α_{d}	B	C	χ^2/dof	q^*	α_{d}	D	E	χ^2/dof
0.8	34.639(2)	1.384(5)	12.94(2)	4.39	0.8	34.592(1)	17.4(1)	−4.77(5)	22.81	0.8	34.607(1)	27.9(2)	−15.2(2)	2.89
0.7	34.672(2)	1.280(4)	19.57(4)	10.83	0.7	34.604(1)	27.3(1)	−8.58(6)	30.19	0.7	34.623(1)	46.6(2)	−27.7(2)	3.20
0.6	34.680(2)	1.345(4)	23.31(4)	21.21	0.6	34.615(1)	31.1(1)	−8.76(7)	6.62	0.6	34.636(1)	51.3(3)	−28.6(2)	1.69
0.5	34.677(2)	1.427(4)	25.65(4)	23.81	0.5	34.621(1)	32.8(1)	−7.99(7)	1.46	0.5	34.641(1)	51.4(3)	−26.3(3)	4.37
0.4	34.671(2)	1.507(4)	27.43(4)	23.41	0.4	34.624(1)	33.7(1)	−7.05(7)	0.42	0.4	34.643(1)	50.3(3)	−23.3(3)	6.64
0.3	34.664(2)	1.588(5)	28.99(4)	20.64	0.3	34.627(1)	34.4(1)	−6.01(7)	0.80	0.3	34.643(1)	48.6(3)	−19.9(3)	7.79
0.2	34.657(2)	1.679(5)	30.63(4)	17.18	0.2	34.629(1)	35.0(1)	−4.75(7)	1.68	0.2	34.642(1)	46.2(3)	−15.8(3)	8.26
0.1	34.647(2)	1.807(5)	32.85(3)	12.21	0.1	34.632(1)	35.6(1)	−2.94(8)	3.42	0.1	34.639(1)	42.4(3)	−9.7(3)	7.99
0.01	34.627(1)	2.164(5)	39.48(3)	3.83	0.01	34.638(1)	37.2(1)	2.51(8)	9.80	0.01	34.631(1)	31.0(3)	8.5(3)	5.34

(a)

(b)

(c)
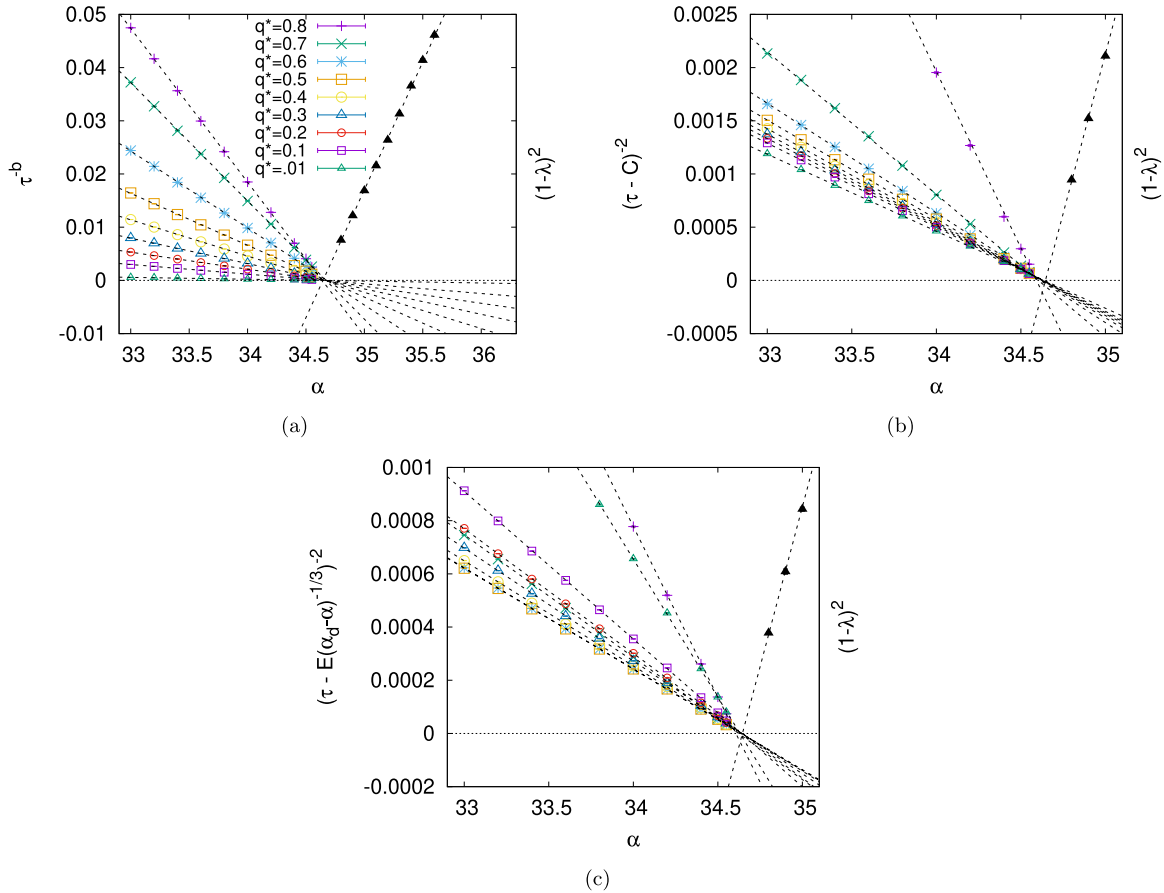
Figure 6. Colors: divergence of the overlap relaxation time $\tau_{q^*}(\alpha)$, defined as the time needed for $q(t)$ to decrease down to q^*. (a)–(c) The data rescaled according to functions (16)–(18) respectively. Black: behavior of $(1 - \lambda(\alpha))^2$, where $\lambda(\alpha)$ is the stability parameter of the non-trivial solution for $\alpha > \alpha_d$. The correct scale for λ is given in the first figure (in both (b) and (c), black points follow a different y-scale than colored ones and are just meant to guide the eye toward the transition point). The dynamic transition α_d is located at the intersection of the color and black lines with the x-axis.

close to $b = 2$. This corroborates the argument supporting $b = 2$; a cautious estimate for α_d from our data would be at this point $\alpha_d = 34.625(20)$.

A second independent estimation of α_d can be obtained by studying the loss of stability for the non-trivial solution when approaching α_d from above. A quantitative prediction from scalar bifurcation theory [13] states that $(1 - \lambda(\alpha))^2 \sim (\alpha - \alpha_d)$ vanishes linearly for $\alpha \to \alpha_d^+$. This is shown in figure 6(a) in black points (beware that points reproduced in figures 6(b) and (c) are out of scale), while the estimated α_d from the fit is given in table 2. Comparing it with the result from the analysis of the overlap decay for $\alpha < \alpha_d$, we conclude $\alpha_d = 34.64(2)$. The quoted uncertainty comes from the reasonable assumption that the difference in the two independent estimations is mostly due to systematic rather than statistical errors.

Table 2. Results from the fit on the stability parameter $\lambda(\alpha)$.

α_{d}	χ^2/dof
34.645(2)	1.39
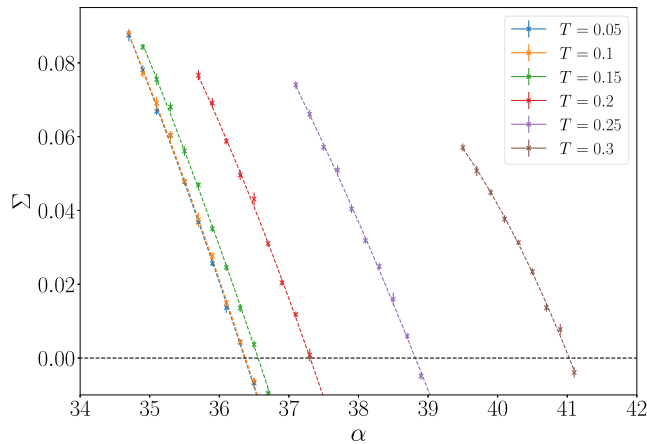
Figure 7. Complexity curves $\Sigma(\alpha, T)$. Dashed lines are a second order polynomial fit. The intersection of each curve with the x-axis returns the value of the condensation transition $\alpha_{\mathrm{c}}(T)$.

3.1.1. Numerical determination of α_c. The condensation transition line $\alpha_{\mathrm{c}}(T)$ is defined by the condition $\Sigma(\alpha, T) = 0$, where Σ is the complexity of the states dominating the partition function at temperature T. Recalling equation (6), Σ can be written as the difference of Bethe free entropies

$$\Sigma = \frac{1}{N} \left[F_{\mathrm{Rec\,II}}(\mathrm{para}) - F_{\mathrm{Rec\,II}}(\mathrm{planted}) \right], \tag{19}$$

where $F_{\mathrm{Rec\,II}}$ is obtained for any BP fixed point from equation (B25), the temperature dependence being encoded into the affinity function f_a. In particular, $F_{\mathrm{Rec\,II}}(\mathrm{para}) = -\alpha$. A number of curves $\Sigma(\alpha, T)$ for different values of T are plotted in figure 7 as a function of α. The $T = 0$ limit is under control thanks to the fact that the curves for the two lowest values of T are practically indistinguishable. Notice that the step $\Delta\alpha = 0.2$ in the data is quite large, and the first point in each curve does not exactly correspond to $\alpha_{\mathrm{d}}(T)$. The estimation of Σ is on the contrary much more precise and allows for a reliable determination of $\alpha_{\mathrm{c}}(T)$. We also observe that the maximum of the complexity appears to decrease when increasing temperature. We cannot exclude the possibility that it goes to zero for some finite temperature, i.e. that the transition becomes continuous for high enough α.
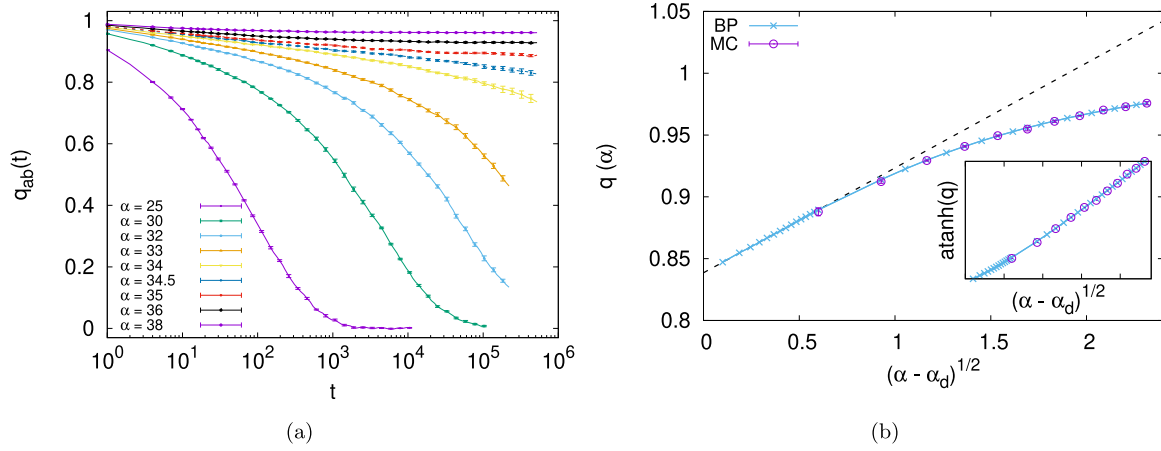
(a)

(b)

Figure 8. (Left) Overlap as a function of time between two independently evolving replicas starting from the same equilibrated (planted) initial condition. Different curves correspond to different values of α. Dashed lines for $\alpha = 34.5$ and $\alpha = 35$ indicate the region where the overlap supposedly starts to develop a plateau with $q = q_{\mathrm{d}}$. (Right) Long time value of the overlap for $\alpha > \alpha_{\mathrm{d}}$ from MC and population dynamics BP simulations. The birth of the non-trivial solution follows a square root singularity, allowing us to extrapolate from the BP data $q_{\mathrm{d}} = 0.838(3)$. The data is scaled according to $\alpha_{\mathrm{d}} = 34.63$ from BP population dynamics. (Inset) The trivial bending of data points for high α can be straightened by plotting $\tanh^{-1}(q(\alpha))$

3.2. Equilibrium autocorrelation time from Monte Carlo

MC simulations have the advantage of being much less eager of computer memory as compared to solving the BP equations, where one needs to work with entire distributions (for each of the messages) even in our RS-like case. On the other hand, relaxation times can be very long (we are approaching an MCT-like glass transition, indeed). This can be noticed by looking at figure 8(a), where we plot the overlap decay for several orders of magnitude in the number of MC sweeps. Before analyzing the results, we give a precise definition of the overlap and discuss the details of the MC simulations at zero temperature.

The overlap between two configurations of angles $\underline{x}_a$ and $\underline{x}_b$ needs to be maximized over rotations due to the global rotational symmetry of the model, which becomes continuous in the limit $d \to \infty$ of infinite discretization precision,

$$q(t) = \max_{A \in [0, 2\pi)} \langle \cos(x_a - x_b - A) \rangle = \sqrt{\langle \cos(x_a - x_b) \rangle^2 + \langle \sin(x_a - x_b) \rangle^2}, \quad (20)$$

where $\langle \cdot \rangle = \frac{1}{N} \sum_{i=1}^{N}$ stands for the average over the system. We consider both the overlap $q_{ab}(t)$ between two independent replicas of the same system, in which case $x_a = x_i^{(a)}(t)$ and $x_b = x_i^{(b)}(t)$, and the overlap $q(t)$ of each replica with the starting configuration, in which case $x_a = x_i(t)$ and $x_b = x_i(t = 0)$. The latter situation can be further simplified by initially planting the system in a configuration with all the variables identically set equal to zero, which is possible thanks to the introduction of random shifts associated

with each edge of the graph (that are required in the first place in order to suppress crystallization). Of course, we expect the equilibrium long time limit of both the overlaps $q(t)$ and $q_{ab}(t)$ to be identical. However, it is also clear that $q_{ab}(t)$ will in general decorrelate faster.

For what concerns the algorithm, we adopt in this case a simple zero temperature heat bath rule: at each step, a randomly selected spin is updated with uniform probability among the Potts states which satisfy all the constraints with the neighbors. Since we start from a planted configurations of exactly zero energy, at least one such a state is guaranteed to exist for every spin, and the procedure is well posed. The advantage with respect to standard Metropolis is that moves are always accepted, this coming at the price of computing the local energy for each of the p states a variable can assume. Notice, however, that (for f^{flat}) this can be performed in a smart way by running just one time over the neighbors of the variable to update and increasing by 1 the energy of the $2d-1$ states inside the excluded region.

Obtaining a precise estimate of α_{d} from MC data is more difficult than from population dynamics. Figure 8(a) shows as a function of time (MC sweeps) the average of the overlap over 10 different systems of size $N = 10^4$. We checked finite size corrections to be smaller than the statistical errors by simulating a few samples of size $N = 10^5$. For $\alpha > \alpha_{\mathrm{d}}$, planting allows us to initialize the system *inside* the planted cluster, which is indistinguishable from a typical cluster for the random model, up to α_{c}. Unfortunately, the $\alpha > \alpha_{\mathrm{d}}$ data is of little use in the estimation of α_{d}, since we do not have in this case an analogous of the BP stability parameter. A possible way round would be to consider the long time limit $q(\alpha)$ of the overlap at the plateau, which is expected [13] to approach a non zero value $q_{\mathrm{d}} > 0$ with a square root singularity at α_{d}, as shown in figure 8(b). Obtaining α_{d} from the $q(\alpha)$ data, on the other hand, would imply a fit where also the parameter q_{d} is unknown. For this reason we prefer first to estimate α_{d} from the overlap decay for $\alpha < \alpha_{\mathrm{d}}$, and then use it to verify the square root singularity and eventually obtain q_{d}. Notice that the behavior $(q(\alpha) - q_{\mathrm{d}}) \sim (\alpha - \alpha_{\mathrm{d}})^{1/2}$ can be valid only in the vicinity of α_{d}, since we know by definition that $q(\alpha) < 1$. The strong deviation from a straight line behavior in figure 8(b) is in this sense totally expected. What is inconvenient, though, is the fact that MC points are located in the region where these corrections are already relevant. This is due to the fact that obtaining a reliable, size independent estimation of $q(\alpha)$ by MC simulations becomes very difficult close to α_{d}, where one has to increase both the time of the simulation and N. For this reason, we can only rely on the BP data, which fit reasonably well the square root singularity, to obtain $q_{\mathrm{d}} = 0.838(3)$.

The last approach we can resort to is the study of the overlap correlation time for $\alpha < \alpha_{\mathrm{d}}$. To this end, we once again define $\tau_{q^*}(\alpha)$ as the time needed for $q(t)$ or $q_{ab}(t)$ to decrease down to q^*, with $q^* < q_{\mathrm{d}} \approx 0.83$. The $T = 0$ approach to α_{d} is essentially governed by MCT-like laws [29], predicting a power-law divergence $\tau(\alpha) \sim (\alpha_{\mathrm{d}} - \alpha)^{-\gamma_0}$ for $\alpha \to \alpha_{\mathrm{d}}^-$, where γ_0 is a specific $T = 0$ exponent. Here we prefer to use instead the parameter $b = 1/\gamma_0$, and fit the data according to equation (16). The outcome is given in figure 9. On the left side, we exhibit the rescaled data for $q^* = 0.8$, showing a value of α_{d} compatible with the BP prediction. On the other hand, from figure 8(a) it is already evident that in order to decrease q^* one may be forced to exclude some of the biggest
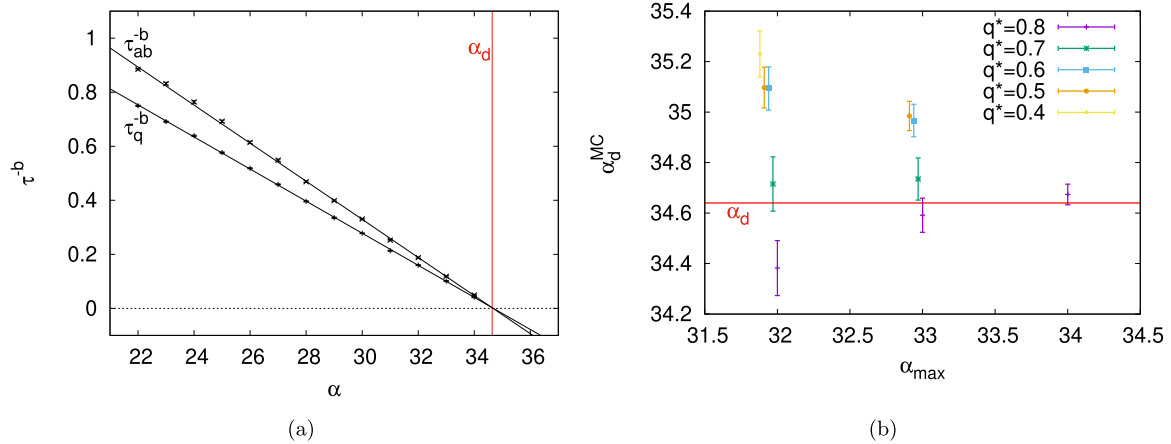
(a)

(b)

Figure 9. (Left) Power-law divergence of $\tau_{q^*}(\alpha)$, for $q^* = 0.8$. Both the overlap with the initial configuration and the one between two independent replicas are considered (the correlation times are named respectively τ_q and τ_{ab}). In this case we obtain $b \approx 0.26$, corresponding to an MCT-like exponent $\gamma_0 \approx 3.85$. Notice however that the value of b strongly depends on q^*, see table 3. (Right) The value of α_d resulting from the fit depends both on q^* and on $\alpha_{\max}$, corresponding to the biggest value of α included in the fit. The red line indicates in both figures the estimate of α_d as obtained from BP in the previous section.

values of α. For this reason, in figure 9(b) we study how the choice of the maximum value $\alpha_{\max}$ included in the fit affects the estimation of α_d, for different values of q^*. From the picture, one can notice that estimates on α_d are the lesser precise the lower $\alpha_{\max}$; in particular, the dispersion of data points over q^* at fixed $\alpha_{\max}$ and the size of error bars visibly decrease from $\alpha_{\max} = 32$ to $\alpha_{\max} = 33$ (for $\alpha_{\max} = 34$ we have only one point, so we cannot make any strong assertion). It is interesting, by the way, to highlight how the high dependence of α_d from q^*, for $\alpha_{\max}$ fixed, is accompanied by a relevant variation of also the exponent b with q^*, see table 3, similarly to what happened when analyzing BP data in the previous section. In this case, however, we lack at the present moment an analytical prediction for b, and an accurate treatment of subleading corrections to the simple power-law behavior is beyond our reach. For this reason, we can only conclude from figure 9(b) that the behavior of real MC data is reasonably well described by the population dynamics BP prediction, considering the level of the systematic biases affecting our determination of α_d from MC simulations.

Finally, it is interesting to consider in figure 10 the long time tail of the overlap decay as approaching α_d from below. In the top panel, the BP data is shown to possess a simple exponential behavior. The MC dynamics of real systems, on the contrary, exhibits a stretched exponential form $A \, e^{-(t/\hat{\tau})^\beta}$, typical of MCT systems close to the MCT transition. The value of the exponent β from the different fits (black lines in figure 10(b)) is reasonably constant on the whole considered α-interval, varying between 0.48 and 0.53 without any recognizable trend. The parameter $\hat{\tau}(\alpha)$ from the fit is in principle proportional to the structural relaxation time of the system $\tau(\alpha)$. However, we believe the definition of $\tau(\alpha)$ in terms of $\tau_{q^*}(\alpha)$ to be somewhat more reliable, since

Table 3. Value of the MCT-like exponent $b = 1/\gamma_0$ from MC simulations.

q^*	$\alpha_{\max}$	$b\ (\tau_q)$	$b\ (\tau_{ab})$
0.8	34	0.260(2)	0.263(2)
	33	0.262(3)	0.266(3)
	32	0.266(4)	0.274(5)
0.7	33	0.223(2)	0.225(3)
	32	0.221(3)	0.226(4)
0.6	33	0.206(2)	0.204(2)
	32	0.205(2)	0.200(3)
0.5	33	0.197(2)	0.197(2)
	32	0.197(2)	0.194(2)
0.4	32	0.188(2)	0.186(2)

it does not depend on the quality of a phenomenological fit and on the considered fitting interval. On the right side, we show the correlation for different values of α as a function of the rescaled time $t/\tau_{q^*=0.5}(\alpha)$, resulting in a decent collapse of all the curves. The collapse is not perfect as the tail becomes slightly more pronounced when increasing α.

3.3. Effects of discretization

All the analysis performed so far was only limited to a discretized version of the model, with precision $d = 10$ to be exact. This coincides, having chosen for definiteness to focus on the fixed $q = 20$ case, for which the transition is appreciably first order, with approximating the continuous interval $[0, 2\pi)$ with $p = qd = 200$ discrete clock-states. It is thus essential to assess how much of the quantitative predictions given so far, in particular for what concerns the computation of the transition lines, is still valid in the limit $p \to \infty$. The answer is provided in figure 11, where we plot the $T = 0$ thresholds $\alpha_{\mathrm{d}}(p)$ and $\alpha_{\mathrm{c}}(p)$ for $p = 200, 240, 280, 320$, as obtained from the population dynamics BP numerical procedure. The corrections, for the values of p considered, appear to be very relevant. Even worse, they show a very slow $1/p$ scaling. For these reasons, we believe a precise extrapolation of the continuous limit of the thresholds, starting only from the solution of the discretized BP equations, to be rather delicate. In this spirit we report the results from the fits of figure 11, to be taken with caution: $\alpha_{\mathrm{d}}^\infty \approx 32.37$ and $\alpha_{\mathrm{c}}^\infty \approx 33.86$. On a more bright side, however, some way to better control the extrapolation error could be hopefully provided by the study of the p dependence of α_{KS}, which is known analytically (11), and that also shows an approximate $1/p$ behavior.

The $1/p$ scaling of discretization corrections for the model at hand was already reported in [4]. Unfortunately, the situation is very different from other planar spin-models on random graphs, exhibiting a very fast exponential convergence [30]. This is presumably due to the peculiarity of the excluded volume interaction, that sharply depends on the particle diameter θ. Among the effects of discretization, we have indeed that the forbidden states are contained by all means into an effective diameter which is smaller than θ by a quantity of order $1/p$ (see for example the red points in figure 1).

(a)

(b)

(c)
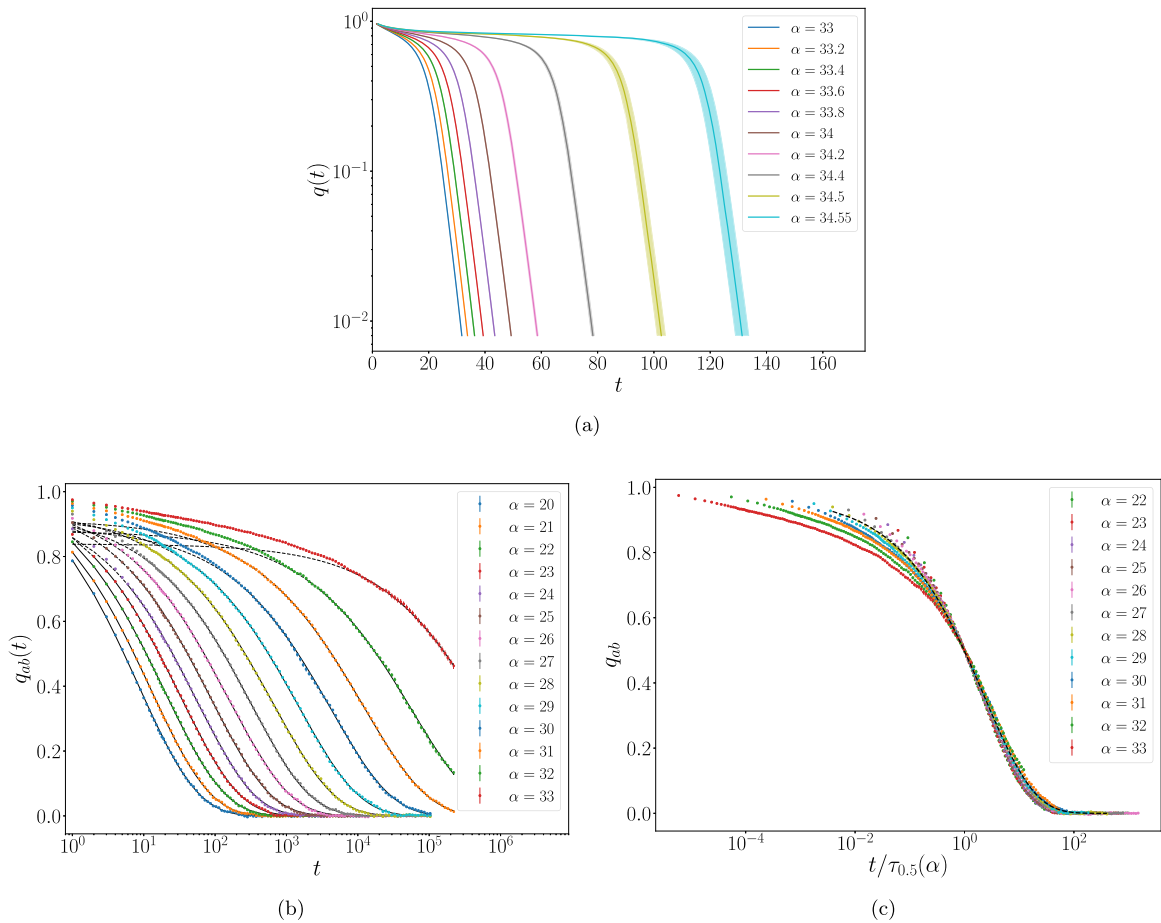
Figure 10. (Top) Overlap decay from population dynamics BP equations for $\alpha < \alpha_{\mathrm{d}}$, using the same data of figure 5(a). The log-scale on the y-axis highlights a purely exponential tail for the decay at small overlap. (Bottom left) Overlap decay from MC data. Black lines represent fits according to a stretched exponential of the form $A\,e^{-(t/\hat{\tau})^{\beta}}$. The values of β vary randomly between 0.48 and 0.53. (Bottom right) The same data is plotted against $t/\tau_{q^*=0.5}(\alpha)$. Black dashed line is a stretched exponential of exponent $\beta = 0.48$.

(a)

(b)

Figure 11. Corrections to the $T = 0$ transition thresholds scale as $1/p$.

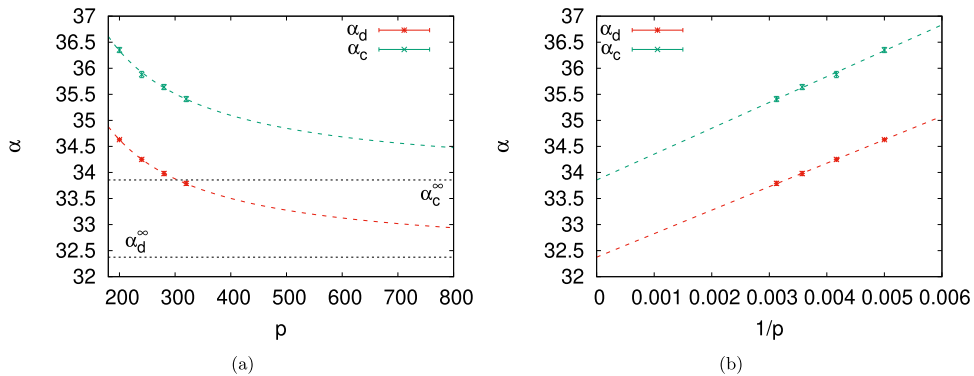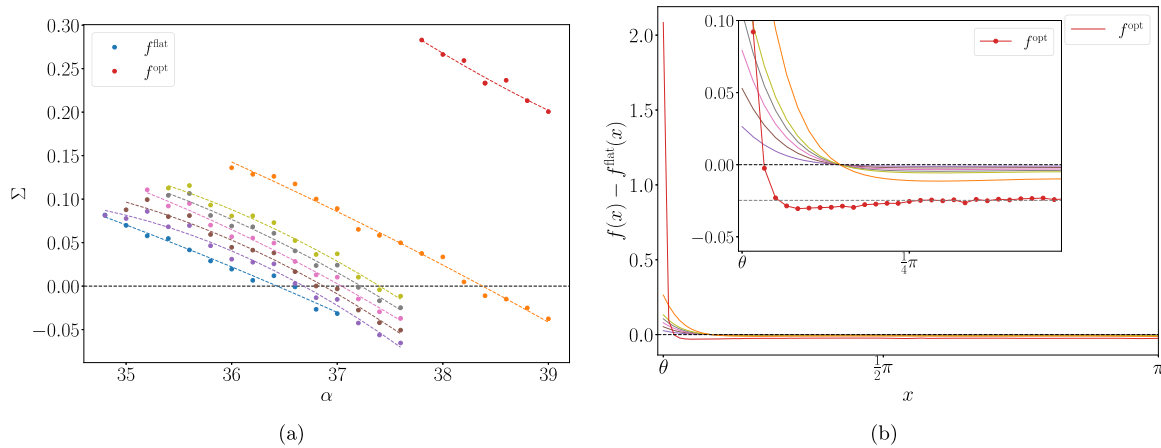**Figure 12.** (Left) Complexity curves as a function of α for different interactions (that are depicted on the right with the corresponding color) along the optimization procedure. Dotted lines are meant only to guide the eye. Blue curve refers to the starting point f^{flat}, while red curve to the result of the optimization f^{opt}. Curves in between all refer to the first step in the procedure (we move along the direction of the gradient of the complexity starting from f^{flat}), for increasing sizes of the 'step' in the direction of the gradient, i.e. the speed rate of the gradient descent.

When we increase discretization, we are hence actually increasing also the effective diameter (i.e. decreasing the effective q), which may play a direct role in the $1/p$ decrease of the transition thresholds, since the system is more constrained for bigger diameters, see for instance figure 2.

4. Optimizing the dynamical transition

4.1. Complexity maximization

The 'typical' complexity Σ (i.e. the one associated with dominant states) is an observable that can be estimated via BP as a difference of free entropies, see equation (19). The general behavior of Σ as a function of α in the case of a discontinuous transition, as shown in figure 7, is the following: $\Sigma = 0$ below α_{d}, then Σ has a jump at α_{d} to a finite (maximum) value, and hence starts to monotonically decrease with α up to α_{c}, where $\Sigma = 0$. One may then wonder if there is a way to postpone the dynamical threshold α_{d} by directly acting on the shape of the complexity. A strategy that we found to be successful is to iteratively maximize the value of the complexity close to its maximum at α_{d}: the idea is that by shifting the complexity curve to higher values of α, also the value of Σ at fixed $\alpha > \alpha_{\text{d}}$ will increase, see figure 12(a). However, since the maximum value of the complexity can take any strictly positive value, there is no reason to exclude the existence of particular functions f for which the complexity curve is still postponed, but it is also lowered. In this sense, we lack a physical intuition (not to mention a rigorous

proof) supporting the fact that the interaction f (not to be confused with a free energy) for which $\alpha_{\mathrm{d}}(f)$ has its maximum should coincide with the function f for which also $\Sigma(\alpha_{\mathrm{d}}(f)\,;f)$ has its maximum.

Having said that, this method is the best chance to our knowledge to automatically optimize the potential when an analytical expression for α_{d} is not known, and it yields promising results. We compared it with other strategies, such as the maximization of the stability parameter λ of the BP non-trivial solution, which is $\lambda = 1$ at the dynamical transition and then decreases for $\alpha > \alpha_{\mathrm{d}}$ (the solution being more stable). Since it takes a fixed value at α_{d} for any f, then the maximization procedure is in this case justified. Another strategy is the minimization of the so called Kullback–Leibler divergence, measuring the distance of the non-trivial distribution of messages $\mathbb{P}^{\mathrm{eq}}(\nu)$ from the paramagnetic one, which in the continuous notation reads

$$\mathbb{E}_{\mathbb{P}^{\mathrm{eq}}(\nu)}\left[D_{\mathrm{KL}}(\nu^{\mathrm{para}}\|\nu)\right] = \mathbb{E}_{\mathbb{P}^{\mathrm{eq}}(\nu)}\left[-\frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}x\,\log\left(\frac{\nu(x)}{(2\pi)^{-1}}\right)\right]. \tag{21}$$

For $\alpha < \alpha_{\mathrm{d}}$ this quantity is zero, since the only solution is $\mathbb{P}^{\mathrm{eq}} = \mathbb{P}^{\mathrm{para}}$, it takes a non-trivial value at α_{d} and then increases for $\alpha > \alpha_{\mathrm{d}}$. Following the same line of thought as in the previous cases, one should minimize it. Furthermore, the minimization of this quantity has a nice physical implication: since (21) becomes bigger the more the fixed point messages depart from the uniform solution, this implies that solutions containing frozen fields (which are highly irregular distributions) are strongly penalized by the procedure. However, the numerical estimation of the gradient of λ and of the Kullback–Leibler divergence is much more involving than computing $\partial\Sigma/\partial f$, which reduces to computing the derivative of a free energy. Moreover, the λ estimator becomes very noisy as the optimization proceeds, thus requiring the use of an increasing number of samples, while the Kullback–Leibler divergence appears also to decrease the condensation threshold α_{c}, making convergence difficult near the tri-critical point. We did not further investigate this phenomenon.

In the end, the best result is obtained through the complexity maximization. Despite the caveats already expressed, it has some practical advantages. In particular, this method does not rely exclusively on BP to estimate messages. Being a derivative of a difference of free energies, the gradient $\partial\Sigma/\partial f$ can be computed also as some two-point correlation function, e.g. via MC sampling, as outlined in the following. Moreover, the method is completely general, and it can in principle be applied also to mean-field hard-spheres systems as studied in [16], where an exact method can be independently developed, hopefully contributing to shed some light on the physics behind the complexity maximization.

The maximization is implemented through a gradient descent: we move in the space of f (subjected to the normalization constraint $\sum_a f_a = p$) along the gradient $\partial\Sigma/\partial f$ at fixed $\alpha = \alpha_{\mathrm{d}} + \delta\alpha$, compute the new α_{d} and repeat the procedure until stationarity is reached. Thanks to discretization, the gradient descent in the space of functions f (which depend only on one argument, the angular distance) can be numerically treated as an optimization in the finite set of parameters $\{f_a\}$. The procedure is the following

$$f_a^{(t+1)} = f_a^{(t)} + \mu \left(\frac{\partial \Sigma \left(\alpha_{\mathrm{d}}(f^{(t)}) + \delta\alpha; f^{(t)} \right)}{\partial f_a} + \Lambda \right) \mathbb{I}\left(\cos(2\pi a/p) \leqslant \cos(2\pi/q) \right), \qquad (22)$$

where $\Lambda = -[p - (2d-1)]^{-1}\sum_{a=d}^{p-(d-1)} \partial\Sigma^{(t)}/\partial f_a$ is a Lagrange multiplier ensuring that $\{f_a^{(t+1)}\}$ is correctly normalized at each step, and μ is the speed rate of the gradient descent. We found empirically that computing the gradient close to the dynamical transition results in less noise. In all the computations we have used $\delta\alpha = 0.5$. As the optimization proceeds, the support of the gradient concentrates on the forbidden values that encode the constraints of the problem. The optimization procedure ends, when the only way to further increase the complexity would be to relax the constraints, something that is forbidden thanks to the factor $\mathbb{I}\left(\cos(2\pi(a-1)/p) \leqslant \cos(2\pi/q) \right)$ in (22).

4.2. Computation of the gradient of the complexity

The complexity we want to maximize is defined as

$$\Sigma = \frac{1}{N} \left[F_{\mathrm{Rec\,II}}(\mathbb{P}^{\mathrm{para}}, f) - F_{\mathrm{Rec\,II}}(\mathbb{P}^{\mathrm{eq}}, f) \right], \qquad (23)$$

where $\mathbb{P}^{\mathrm{para}}(\nu) = \prod_{a=1}^{p}\delta(\nu_a - n)$, $n = \frac{1}{p}$, is the paramagnetic BP fixed point and $\mathbb{P}^{\mathrm{eq}}$ is the non-trivial one arising for $\alpha > \alpha_{\mathrm{d}}(f)$. In our discretized setting, $f = \{f_a\}$ is a symmetric function under $a \to p - a$ and periodic of period p (we use in this sense the convention $f_{-a} = f_{p-a}$). The population dynamics definition of the free entropy $F_{\mathrm{Rec\,II}}$ is given by equation (B25), which we recall here for convenience as

$$\frac{1}{N}F_{\mathrm{Rec\,II}}(\mathbb{P}, f^{\mathrm{Fixed}}, f^{\mathrm{State}}) = \sum_d P_{\mathrm{Node}}(d) \int \prod_{k=1}^{d} \left[d\nu_k \sum_b f_b^{\mathrm{State}} n \mathbb{P}\left(T(-b)\nu_k \right) \right]$$
$$\times \log \left[Z(\{\nu_k\}_d, f^{\mathrm{Fixed}}) \right]$$
$$- \alpha \int d\nu_1 \, d\nu_2 \mathbb{P}(\nu_1) \sum_b f_b^{\mathrm{State}} n \mathbb{P}(T(-b)\nu_2)$$
$$\times \log \left[\sum_{r,s} \nu_{1,r} f_{s-r}^{\mathrm{Fixed}} \nu_{2,s} \right]. \qquad (24)$$

The distinction between f^{State} and f^{Fixed} is here only introduced in order to discuss separately the different terms coming from the derivative with respect to f_a. Of course in the end one has to evaluate every derivative at $f^{\mathrm{State}} = f^{\mathrm{Fixed}} = f$, with $\sum_a n f_a = 1$, $n = \frac{1}{p}$. Finally, we also recall that

$$Z(\{\nu_k\}_d, f) = \sum_{a=0}^{p-1} n \prod_{k=1}^{d} \sum_{b=0}^{p-1} f_{b-a} \nu_{k,b}. \qquad (25)$$

The derivative $\partial\Sigma/\partial f_a$ contains terms coming from the explicit dependence of Σ from both f^{State} and f^{Fixed}, but also from the change in the non-trivial fixed point distribution

$\mathbb{P}^{\mathrm{eq}}$. The paramagnetic fixed point distribution, on the contrary, is given by $\mathbb{P}^{\mathrm{para}} \equiv \prod_a \delta(\nu_a - n)$ for any choice of f satisfying $\sum_a f_a n = 1$, as can be immediately noticed by a direct inspection of the BP fixed point equation (B23). We can thus write

$$\frac{N\partial\Sigma}{\partial f_a} = \frac{\partial F_{\mathrm{Rec\,II}}(\mathbb{P}^{\mathrm{ppara}}, f^{\mathrm{Fixed}}, f^{\mathrm{State}})}{\partial f_a^{\mathrm{Fixed}}} - \frac{\partial F_{\mathrm{Rec\,II}}(\mathbb{P}^{\mathrm{eq}}, f^{\mathrm{Fixed}}, f^{\mathrm{State}})}{\partial f_a^{\mathrm{Fixed}}}$$
$$+ \frac{\partial F_{\mathrm{Rec\,II}}(\mathbb{P}^{\mathrm{ppara}}, f^{\mathrm{Fixed}}, f^{\mathrm{State}})}{\partial f_a^{\mathrm{State}}} - \frac{\partial F_{\mathrm{Rec\,II}}(\mathbb{P}^{\mathrm{eq}}, f^{\mathrm{Fixed}}, f^{\mathrm{State}})}{\partial f_a^{\mathrm{State}}}$$
$$- \frac{\partial F_{\mathrm{Rec\,II}}}{\partial \mathbb{P}}(\mathbb{P}^{\mathrm{eq}}, f^{\mathrm{Fixed}}, f^{\mathrm{State}}) \left[\frac{\partial \mathbb{P}^{\mathrm{eq}}}{\partial f_a^{\mathrm{Fixed}}} + \frac{\partial \mathbb{P}^{\mathrm{eq}}}{\partial f_a^{\mathrm{State}}} \right]. \quad (26)$$

Most of the terms either simplify or are equal to zero. The first two terms cancel, as we are going to comment in a moment. The third term is trivially equal to zero, since both the logarithms in equation (24) vanish when one considers $\nu_a = n$ and the fact that $n f_a^{\mathrm{Fixed}}$ is correctly normalized. The fifth term is zero since the distribution $\mathbb{P}^{\mathrm{eq}}$ extremizes the Bethe free entropy. In the end, the only remaining term is

$$\frac{\partial\Sigma}{\partial f_a} = - \frac{\partial N^{-1} F_{\mathrm{Rec\,II}}(\mathbb{P}^{\mathrm{eq}}, f^{\mathrm{Fixed}}, f^{\mathrm{State}})}{\partial f_a^{\mathrm{State}}}\bigg|_{f^{\mathrm{State}}=f^{\mathrm{Fixed}}=f}. \quad (27)$$

Before proceeding with the computation, let us show that the first two terms indeed cancel. The derivative $\partial N^{-1} F_{\mathrm{Rec\,II}}/\partial f_a^{\mathrm{Fixed}}$, for generic $\mathbb{P}$, can be written after some manipulations[4] as

$$\frac{\partial N^{-1} F_{\mathrm{Rec\,II}}}{\partial f_a^{\mathrm{Fixed}}}\bigg|_f = \alpha \int \mathrm{d}\nu_1 \, \mathrm{d}\nu_2 \mathbb{P}(\nu_1) \sum_b f_b^{\mathrm{State}} n \mathbb{P}(T(-b)\nu_2) \frac{\sum_r \nu_{1,r} \nu_{2,r+a}}{\sum_{r,s} \nu_{1,r} f_{s-r} \nu_{2,s}}. \quad (28)$$

If we multiply and divide by the same amount f_a, we can make a term $\nu_{1,r} f_a \nu_{2,r+a}$ appear, which is proportional by the definition of messages to the marginal probability $p(x_1 = r, x_2 = r + a)$, as a function of r, of two neighboring variables at a fixed distance a. The whole integral thus represents the total equilibrium probability for two neighbors to be at distance a, which is essentially the pair correlation function of the system. In the entire range $\alpha < \alpha_{\mathrm{c}}$, though, the pair correlation function of the paramagnetic and of the planted fixed point is the same and is simply equal (whenever short loops are absent) to the normalized interaction $n f_a$, that is essentially the Boltzmann–Gibbs exponential of the pairwise potential. Simplifying the extra $1/f_a$ factor, we obtain that the first two terms in equation (26) cancel as they are both equal to a constant α/p. As a final remark, this derivation is clearly valid for any $T \neq 0$, when also $f_a \neq 0 \; \forall a$. The case $f_a = 0$ can nevertheless be recovered from the previous one by taking the limit $T \to 0$.

---

[4] One can use the decomposition (30) in order to recast the derivative of the first term in the rhs of (24) in the form reported in the text (the derivative of the second term is trivially proportional and they can be added).

Coming back to our computation, plugging into equation (27) the definition of $F_{\text{Rec II}}$ given by equation (24), we obtain

$$
\begin{aligned}
\frac{\partial \Sigma}{\partial f_a} = &\; \alpha n \int \mathrm{d}\nu_1 \, \mathrm{d}\nu_2 \mathbb{P}^{\text{eq}}(\nu_1) \mathbb{P}^{\text{eq}}(T(-a)\nu_2) \log \left[\sum_{r,s} \nu_{1,r} f_{s-r} \nu_{2,s} \right] \\
& - 2\alpha n \sum_d P_{\text{Node}}(d) \int \mathrm{d}\nu_{d+1} \mathbb{P}^{\text{eq}}(T(-a)\nu_{d+1}) \\
& \times \prod_{k=1}^d \left[\mathrm{d}\nu_k \sum_b f_b n \mathbb{P}^{\text{eq}}(T(-b)\nu_k) \right] \log \left[Z(\{\nu_k\}_d \cup \nu_{d+1}, f) \right],
\end{aligned}
\tag{29}
$$

where we have used the fact that $d \cdot P_{\text{Node}}(d) = 2\alpha P_{\text{Node}}(d-1)$ and replaced a dummy index d in the second term with $d+1$. The quantity $\log(Z)$ appearing in the previous equation can be iteratively expressed in terms of the update function Φ^{update} given by equation (B8) as

$$
\begin{aligned}
\log \left[Z(\{\nu_k\}_d \cup \nu_{d+1}, f) \right] &= \log \left[\sum_a n \prod_{1 \leqslant k \leqslant d+1} \left(\sum_b f_{b-a} \nu_{k,b} \right) \right] \\
&= \log \left[\sum_a \Phi_a^{\text{update}}(\{\nu_k\}_d, f) Z(\{\nu_k\}_d, f) \sum_b f_{b-a} \nu_{d+1,b} \right] \\
&= \log \left[Z(\{\nu_k\}_d, f) \right] + \log \left[\sum_{ab} \Phi_a^{\text{update}}(\{\nu_k\}_d, f) f_{b-a} \nu_{d+1,b} \right].
\end{aligned}
\tag{30}
$$

Plugging (30) into (29), one recognizes[5] that Φ^{update} can be interpreted as a random message ν distributed according to $\mathbb{P}^{\text{eq}}(\nu)$. The second term from (30) has then the same form of the first term in (29) and they can be directly added. Therefore one obtains

$$
\begin{aligned}
\frac{\partial \Sigma}{\partial f_a} = &- \alpha n \int \mathrm{d}\nu_1 \, \mathrm{d}\nu_2 \mathbb{P}^{\text{eq}}(\nu_1) \mathbb{P}^{\text{eq}}(T(-a)\nu_2) \log \left[\sum_{rs} \nu_{1,r} f_{s-r} \nu_{2,s} \right] \\
& - 2\alpha n \sum_d P_{\text{Node}}(d) \int \prod_{k=1}^d \left[\mathrm{d}\nu_k \sum_b f_b n \mathbb{P}^{\text{eq}}(T(-b)\nu_k) \right] \log \left[Z(\{\nu_k\}_d, f) \right].
\end{aligned}
\tag{31}
$$

We notice that the second term does not depend on the index a. This term is trivially constant and can hence be disregarded, since at each step we already enforce the normalization condition $\sum_a f_a n = 1$. Finally one gets

$$
\frac{\partial \Sigma}{\partial f_a} = -\frac{\alpha}{p} \int \mathrm{d}\nu_1 \, \mathrm{d}\nu_2 \mathbb{P}^{\text{eq}}(\nu_1) \mathbb{P}^{\text{eq}}(T(-a)\nu_2) \log \left[\sum_{rs} \nu_{1,r} f_{s-r} \nu_{2,s} \right]
\tag{32}
$$

---

[5] Formally, one can multiply (29) by a factor $1 = \int \mathrm{d}\nu \, \delta \left(\nu - \Phi_a^{\text{update}}(\{\nu_k\}_d, f) \right)$ in order to make appear the BP expression (B23) for $\mathbb{P}^{\text{eq}}(\nu)$.

4.3. Monte Carlo estimator of the gradient of the complexity

Remarkably, the previous formula (32) can also be estimated using MC sampling. This is very convenient, since it might be the only numerically feasible approach in the case of higher spatial dimensions and/or high discretization. To this purpose, one needs to generate a planted graph A along with a planted configuration, denoted $\{x_i^0\}_{1\leqslant i\leqslant N}$, and then run a MC algorithm initialized in the planted configuration to create samples. Let us assume we have generated T samples $\{x_i^t\}_{1\leqslant i\leqslant N, 1\leqslant t\leqslant T} \in \{0,\ldots,p-1\}^{N\times T}$ of the system. The two cavity messages ν_1 and ν_2 in equation (32) are random variables drawn independently from $\mathbb{P}^{\mathrm{eq}}(\nu)$. We can use at this point the fact that the distribution of cavity messages coincides, for the Erdős–Rényi ensemble, also with the distribution of single-variable local marginal probabilities for the complete graph. This allows one to recast an average over ν_i of the kind of $\sum_a \nu_{i,a} g(a)$ inside the integral in equation (32), where g is a generic function, as the time average over the generated samples $\{x_i^t\}_{1\leqslant t\leqslant T}$, for each fixed site i (eventually averaging over the choice of site i): $\frac{1}{T}\sum_{t=1}^T g(x_i^t)$. Since ν_1 and ν_2 should be extracted independently, we simulate two different graphs (the distribution $\mathbb{P}^{\mathrm{eq}}$ itself encodes the average over the graphs ensemble). Finally, notice that the presence of the shift $\mathbb{P}^{\mathrm{eq}}(T(-a)\nu_2)$ translates into considering for ν_2 a new set of generated samples given by $\{x_i^t + a\}_{1\leqslant i\leqslant N, 1\leqslant t\leqslant T}$. The gradient of the complexity is then estimated using the following formula

$$
\frac{\partial \Sigma}{\partial f_a} = -\frac{\alpha}{p}\frac{1}{U}\sum_{(i,j)} \log\left[\frac{1}{T^2}\sum_{1\leqslant t_1,t_2\leqslant T} f_{(x_j^{t_2}-x_j^0+a)-(x_i^{t_1}-x_i^0)}\right], \tag{33}
$$

where U is the number of pairs (i,j) in the sum. We take $U = N$, by running one time over the index $i = 1,\ldots,N$, and considering $j = i$ to belong to a second different system. Since the MC simulations are subject to coherent rotations of the spin variables due to the global rotational symmetry of the model, which becomes continuous in the $p \to \infty$ limit, one should also maximize at each step the overlap with the planted configuration over a global rotation of the system.

In figure 13, we compare the gradient estimated from MC simulations following equation (33) with the one obtained from equation (32). The gradient is computed by fixing $\alpha = \alpha_{\mathrm{d}} + 0.5$ and in correspondence of $f = f^{\mathrm{flat}}$. For $x < \theta$ not all the data points are defined if the simulation time T is not long enough. This can be understood by directly inspecting equation (33). In general, above α_{d} variables will display a distribution biased toward their initial planted state. However, once we subtract the initial position, the distributions are all shifted on top of each other. This can create problems as for a small the sum inside square brackets in equation (33) gets contributions different from zero only from the tail of the distributions, i.e. when x_i or x_j are sufficiently distant from x_i^0, x_j^0. For this reason, one should give the MC enough time to correctly sample the tails of the distributions. We stress however that the determination of the gradient of the complexity in the region $x < \theta$ is not required in order to perform the complexity maximization procedure, since one is interested in this case in following the gradient only relatively to the soft range $x \geqslant \theta$.
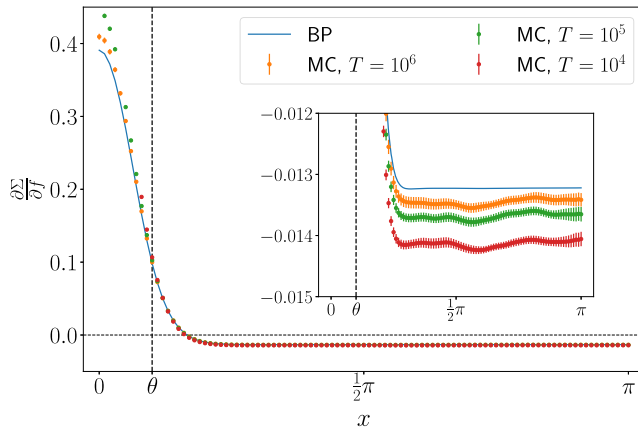
Figure 13. Comparison of the gradient of the complexity around $f = f^{\text{flat}}$ and for $\alpha = \alpha_{\text{d}} + 0.5$, estimated from BP population dynamics using equation (32) and MC simulations using (33). In the latter case, we considered systems of size $N = 10^4$, while error bars accounts for 10 repetitions of the experiment on different graphs. MC data for $x \geqslant \theta$ is less sensible to the total simulation time T (this is convenient as long as one is interested in optimizing the soft part of the interaction only, as in our case). In particular, not all the data points in the region $x < \theta$ are well defined for small T. (Inset) Fine detail of the soft region $x > \theta$.

5. Conclusions and outlook

The CCP we have studied in detail in the present work has several features that make it a unique model: it is a CSP with continuous variables defined on a sparse random graph and presenting an RFOT. To the best of our knowledge this is the only model possessing all these features. The model can be seen as a random CSP, as a mean-field model for the jamming transition and as a solvable glassy model with an RFOT. So it may be useful in several fields of research.

Being defined on a random graph, the model is solvable via the cavity method, although the solution turns out to be quite involved. The first part of this work has been dedicated to the identification of the thermodynamical phase transitions. Previous works already attempted at computing critical thresholds, but under approximations valid in specific limits. Here we provide analytically exact or numerically very accurate values for the critical lines as a function of the 'number of colors'. We clearly identify when the thermodynamical phase transition has a continuous nature and when it is preceded by a dynamical phase transition.

The comparison of the phase diagram of the CCP with that of the much better known DCP reveals some surprises. The space of solutions of the CCP is much larger and contains the space of solutions to the DCP: this could suggest that finding a solution to the CCP is much easier. The critical lines we have computed say the contrary! The CCP undergoes a phase transition at a ratio α of constraints per variables much smaller than for the DCP. The explanation of this counter-intuitive observation can be found in the way constraints are satisfied in the two problems. While in the DCP the constraints

are satisfied in a tight way, in a typical solution to the CCP small gaps remains between variables and this makes harder to satisfy all constraints. Obviously the solution with all constraints satisfied tightly exists also in CCP, but it has a much smaller entropy and so it does not dominate the thermodynamic measure over the solutions, and it is not found by actual algorithms that try to solve the CCP.

Having understood that the phase transitions take place essentially because the uniform measure over the space of solution undergoes a breaking of ergodicity due to entropic reasons, in the second part of this work we have tried to re-weight solutions in order to postpone the phase transitions. To this aim we have modified the interaction potential and found the optimal one, that is the one with phase transitions taking place at the largest α values. The optimization of the interaction potential has been carried on by maximizing the complexity in a fully automatized way (other procedures have been tried and discarded as more noisy and less effective).

We believe this procedure can be of general applicability in any model undergoing an RFOT: indeed any model of that kind in the relevant range of parameters has a non-zero complexity that can be used to optimize the interaction potential. It is worth noticing that the derivative of the complexity with respect to the interaction potential can be written as a correlation function and thus estimated also via MC simulations. This is one more important ingredient to allow generalization of this optimization process to other models with an RFOT.

Interestingly enough the optimized potential has a strongly attractive part at short distances that has the main effect of closing the smallest gaps and allows for better packing (i.e. larger α values at the transitions). In other words, the typical configurations obtained using the optimized potential have much more contacts and fewer small gaps with respect to the original flat measure over the solution space. Preliminary simulations suggest that configurations of this kind (with an excess of contacts with respect to the flat measure) are those often found by smart algorithms trying to find solutions at the largest possible α values. The connection between the algorithmic threshold for smart searching algorithms and the dynamical transition for the optimized interaction potential is currently under study.

The sparseness of the model allows to run efficient MC simulations, in order to measure directly the breaking of ergodicity that takes place at the dynamical transition α_{d}. Our results are compatible with a divergence at α_{d} of the timescale controlling the correlation decay. However, the decay of the correlation function during the MC simulation is quite slow: the tail of the correlation is well fitted by a stretched exponential with exponent close to 0.5 and the decay timescale becomes very large approaching α_{d} (the power law divergence has an exponent close to 4). This makes the estimate of α_{d} from MC data very noisy, especially if compared with data obtained by running the belief propagation algorithm.

We have presented an accurate comparison between the actual behavior of the relaxation dynamics simulated via MC algorithms and the analytical predictions obtained via the cavity method and the BP algorithm. This comparison has not been achieved before in other models (to the best of our knowledge). The reason is simple: most mean-field solvable models are defined on fully connected graphs and thus their MC simulations

are very demanding, and could not achieve sizes and timescales useful for a reliable comparison.

We conclude this work commenting on possible future applications of the present model. Having continuous variables, the model can be used to study the behavior of continuous relaxation dynamics like the stochastic gradient descent; particularly interesting would be to study how the dynamical phase transitions affects this kind of dynamics. Moreover, once a global or local minimum is reached, it would be very interesting to compute the spectral properties of the Hessian. We expect marginal states to play an important role as attracting fixed point for relaxation dynamics and the study of the Hessian spectrum can support this hypothesis. Finally, the jamming transition in this model is still to be studied in detail and connecting it to the behavior of relaxation algorithms would be very enlightening.

Acknowledgments

Appendix A. Planted CCP and DCP as Bayesian inference problems
A.1. Community detection definition (stochastic block model)

In the planting procedure (without random shifts), one adds an edge (ij) to the graph according to some probability[6] $2\alpha f(x_i^0, x_j^0; \beta)/N$, which depends on the values taken by i and j in the planted configuration $\{x_i^0\}$. This generative model for random graphs is also known as the stochastic block model [23]. We may interpret the different values each variable can take as signaling their membership to a specific 'community', and the stochastic block model (planting) as a rule to establish connections between these communities. A natural question is to understand under what circumstances one is able to recover some knowledge on $\{x_i^0\}$ (original community structure) from the observation of the generated graph A. In the simplest case (Bayes optimal), one also knows the parameters of the model, namely the prior P_X from which $\{x_i^0\}$ was extracted, and the function $f(x, y; \beta)$.

Following [23], this problem can be addressed through Bayesian inference. The conditional probability for the adjacency matrix of the graph ($A_{ij} = 1$ if an edge is drawn between nodes i and j, $A_{ij} = 0$ otherwise), reads

$$P(A|\{x_i^0\}) = \prod_{1 \leqslant i < j \leqslant N} \left(2\alpha f(x_i^0, x_j^0; \beta)/N\right)^{A_{ij}} (1 - 2\alpha f(x_i^0, x_j^0; \beta)/N)^{1-A_{ij}}. \quad \text{(A1)}$$

---

[6] In the following, we approximate $N - 1$ in the denominator with N, which is reasonable as $N \to \infty$.

The function $2\alpha f(x,y;\beta)$ is usually called the affinity function, and in our case is a symmetric function. The Bayesian belief for the original group assignment, conditional on the particular realization of the graph, is thus

$$
P(\{x_i\}|A) \propto P(A|\{x_i\})P_X(\{x_i\})
$$
$$
= \prod_{i=1}^{N} P_X(x_i) \prod_{1 \leqslant i < j \leqslant N} \left(\frac{2\alpha f(x_i, x_j; \beta)}{N} \right)^{A_{ij}} \left(1 - \frac{2\alpha f(x_i, x_j; \beta)}{N} \right)^{1-A_{ij}}.
$$
(A2)

We can finally define an extensive Hamiltonian by considering

$$
\beta\hat{\mathcal{H}}(\{x_i\}|A) = -\ln P(\{x_i\}|A) - M \ln N
$$
$$
= -\sum_{i=1}^{N} \ln P_X(x_i) - \sum_{1 \leqslant i < j \leqslant N} \left[A_{ij} \ln\ 2\alpha f(x_i, x_j; \beta) \right.
$$
$$
\left. + (1 - A_{ij}) \ln \left(1 - \frac{2\alpha f(x_i, x_j; \beta)}{N} \right) \right].
$$
(A3)

We notice, in addition to the expected edge-term proportional to A_{ij}, the presence of a weak $1/N$ interaction between non-edges making the graph fully connected, and an external field given by the prior P_X. However, as was shown in [23], the non-edge interaction can be treated in the large N limit in a mean-field way, thus recovering the sparsity of the graph. The BP equations associated to the planted Hamiltonian model (A3) are discussed in appendix B.

A.2. Mixed model

Using the community detection formalism, we can recast both the discrete and continuous coloring as inference problems, also defining a mixed model that allows us to interpolate between the two.

A.2.1. Discrete and continuous coloring as inference problems. Both the discrete and continuous coloring can be treated in a unified way by defining the following affinity function

$$
f(x - y; \beta) = \frac{\exp\left(-\beta\mathbb{I}(\cos(x - y) > \cos\theta)\right)}{\int \mathrm{d}x\,\mathrm{d}y P_X(x)P_X(y)\exp\left(-\beta\mathbb{I}(\cos(x - y) > \cos\theta)\right)},
$$
(A4)

the only difference being in the prior distributions: $P_X(x) = \frac{1}{2\pi}$ for the continuous coloring and $P_X(x) = \frac{1}{q}\sum_{k=0}^{q-1}\delta\left(x - \frac{2\pi k}{q}\right)$ for q-coloring, with $x_i \in [0; 2\pi)$. In the limit $\beta \to \infty$, this reads

$$f_{q-\text{col}}^{T=0}(x-y) = \begin{cases} 0, & \text{if } \cos(x-y) > \cos(2\pi/q) \\ \dfrac{q}{q-1}, & \text{otherwise,} \end{cases} \tag{A5}$$

and

$$f_{\text{CCP}}^{T=0}(x-y) = \begin{cases} 0, & \text{if } \cos(x-y) > \cos(2\pi/q) \\ \dfrac{q}{q-2}, & \text{otherwise.} \end{cases} \tag{A6}$$

A.2.2. Interpolating (mixed) model. We also consider a mixed model that interpolates between the discrete and the continuous coloring for each value of $\theta = \frac{2\pi}{q}$, $q \in \mathbb{N}$. To this end, we modify the prior by introducing a continuous parameter $\rho \in [0,1]$, so that for $\rho = 0$ one recovers the q-coloring, while for $\rho = 1$ one recovers the continuous version (figure 14). We will be interested in the $\beta \to \infty$ limit of the interaction function, which reads

$$P_X(x) = \frac{\rho}{2\pi} + \frac{1-\rho}{q} \sum_{k=0}^{q-1} \delta\left(x - \frac{2\pi k}{q}\right), \quad \rho \in [0,1] \tag{A7}$$

$$f_{\text{Mixed}}^{T=0}(x,y) = \begin{cases} 0, & \text{if } \cos(x-y) > \cos(2\pi/q) \\ \dfrac{q}{q-1}, & \text{if } x,y \in \{2\pi k/q, k \in \mathbb{Z}\} \text{ and } x \neq y \\ \dfrac{q}{q-2}, & \text{otherwise.} \end{cases} \tag{A8}$$

This peculiar shape for $f(x,y)$ is due to the fact that, in order to ensure the normalization condition $\int \mathrm{d}x\, P_X(x) f(x,y) = 1 \ \forall \ y \in [0; 2\pi)$, a simpler translationally invariant form is not enough. This can be understood by considering that variables are effectively divided into two classes, depending on whether they can or cannot be written as $y = 2\pi k/q$, for $k \in \mathbb{Z}$. If variable y belongs to the former class, then it is compatible with $(q-1)$ of the q possible 'discrete' values that neighbors assume with probability $(1-\rho)$. Conversely, if variable y belongs to the continuous background, by geometrical reasons it is then compatible with only $(q-2)$ of the q possible discrete angles. If the linking probability was a function of only the difference $x - y$, this would induce a difference in the average degree of variable nodes depending on the class, and hence the normalization condition would not hold. One can still work with a uniform average degree by increasing the linking probability of variables belonging to the continuous class with respect to the discrete one. Notice that when $\rho = 0$ or $\rho = 1$, there is just one kind of population (the probability for continuous variables with uniform distribution to take a specific set of discrete values being negligible), and the linking probability correctly matches the $\beta \to \infty$ limits $f_{q-\text{col}}^{T=0}(x-y)$ and $f_{\text{CCP}}^{T=0}(x-y)$ given above.

A.2.3. Discretized version of the inference problems. In the case of discretization, variables x's can take only p values. We will denote the prior $P_X(x)$ as $\{n_a\} \in \mathbb{R}^p$, with $0 \leqslant n_a \leqslant 1$ and $\sum_{a=0}^{p-1} n_a = 1$. The function $f(x,y)$ becomes now in general a matrix $\{f_{ab}\} \in \mathbb{R}^{p \times p}$, with $\sum_b n_b f_{ab} = 1 \ \forall a \in \{0, \ldots, p-1\}$.
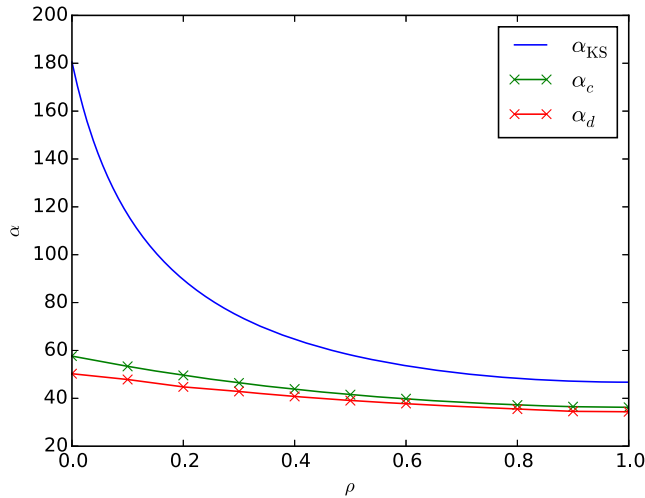
Figure 14. Phase diagram of the mixed model, equations (A8) and (A7) and (A14) and (A13), interpolating between discrete ($\rho = 0$) and continuous ($\rho = 1$) coloring. Parameters are set to $q = 20$ and $d = 10$.

- Discretized version of the continuous model

$$n_a = \frac{1}{p}, \tag{A9}$$

$$f_{ab}^{\text{CCP}}(\beta) = \frac{\exp\left(-\beta\mathbb{I}(\cos(2\pi(a-b)/p) > \cos(2\pi/q))\right)}{[(2d-1)\mathrm{e}^{-\beta} + p - (2d-1)]/p} \tag{A10}$$

- Discretized q-coloring

$$n_a = \frac{1}{q}\sum_{k=0}^{q-1}\delta_a^{kd} \tag{A11}$$

$$f_{ab}^{q-\text{col}}(\beta) = \frac{\exp\left(-\beta\mathbb{I}(\cos(2\pi(a-b)/p) = 1)\right)}{[\exp(-\beta) + (q-1)]/q} \tag{A12}$$

- Discretized mixed model ($\beta \to \infty$)

$$n_a = \frac{\rho}{p} + \frac{(1-\rho)}{q}\sum_{k=0}^{q-1}\delta_a^{kd}, \quad \rho \in [0,1] \tag{A13}$$

$$f_{ab}^{\text{Mixed}} = \begin{cases} 0, & \text{if } \cos(2\pi(a-b)/p) > \cos(2\pi/q) \\ A = \dfrac{q\left[(1-\rho)(q-2) + \frac{\rho}{d}(q-1)\right]}{(q-1)\left(q-2+\frac{\rho}{d}\right)\left(1-\rho+\frac{\rho}{d}\right)}, & \text{if } \sin(\pi a/d) = \sin(\pi b/d) = 0 \text{ and } a \neq b \\ B = \dfrac{q}{q-2+\frac{\rho}{d}} & \text{otherwise.} \end{cases} \tag{A14}$$

In the limit $d \to \infty$ the continuous formulation is correctly recovered, while for $\rho = 0$ and $\rho = 1$ we obtain the $\beta \to \infty$ limit of $f_{ab}^{q-\mathrm{col}}(\beta)$ and $f_{ab}^{\mathrm{CCP}}(\beta)$, respectively. Expression (A14) is obtained in the same way as (A8) by solving for A, B the system $\sum_a n_a f_{ab}^{\mathrm{Mixed}}(A, B) = 1 \ \forall b$.

Appendix B. Belief propagation for CCP

In this section we derive the belief propagation equations for the CCP in the Bayes optimal inference setting, following the discussion of [23]. We denote $\nu^{i \to j}$ the message sent from i to j and ν^i the marginal probability of variable i. Both messages and local marginals are in principle distributions over the continuous interval $[0, 2\pi)$, but thanks to discretization we treat them as p-components arrays $\nu_a^{i \to j}$ and ν_a^i, with $a \in \{0, \ldots, p-1\}$. The belief propagation equations for $\nu^{i \to j}$ associated to the planted Hamiltonian model (A3) reads then

$$\nu_a^{i \to j} = \frac{1}{Z_{i \to j}} n_a \prod_{k \neq j} \left[\sum_{b=1}^{p} (2\alpha f_{ab})^{A_{ik}} (1 - 2\alpha f_{ab}/N)^{1 - A_{ik}} \nu_b^{k \to i} \right], \tag{B1}$$

while local marginals are computed in terms of the messages as

$$\nu_a^i = \frac{1}{Z_i} n_a \prod_{k=1}^{N} \left[\sum_{b=1}^{p} (2\alpha f_{ab})^{A_{ik}} (1 - 2\alpha f_{ab}/N)^{1 - A_{ik}} \nu_b^{k \to i} \right], \tag{B2}$$

where $Z_{i \to j}$ and Z_i ensures that $\nu^{i \to j}$ and ν^i are normalized probability distributions. In the following, we choose to get rid of a term 2α multiplying f_{ab} in the edge-term since it factorizes and just shifts the free energy by a fixed amount.

Since the factor graph of this system is fully connected, one should in principle have to keep track of $N(N-1)$ messages. However, one can notice that the messages $\nu^{i \to j}$ depend weakly on the indices j as long as $A_{ij} = 0$,

$$\nu_a^{i \to j} - \nu_a^i = O(1/N) \quad \forall \ (ij) \notin E. \tag{B3}$$

In the end [23], this allows to simplify the cavity equations to only keep track of the messages between connected nodes by introducing an average external field h_a which accounts for the global message coming on each site from non-edges,

$$\nu_a^{i \to j (t+1)} = \frac{1}{Z_{i \to j}} n_a \, \mathrm{e}^{-h_a^{(t)}} \prod_{k \in \partial i \backslash j} \left[\sum_{b=1}^{p} f_{ab} \nu_b^{k \to i (t)} \right] \tag{B4}$$

$$\nu_a^{i (t+1)} = \frac{1}{Z_i} n_a \, \mathrm{e}^{-h_a^{(t)}} \prod_{k \in \partial i} \left[\sum_{b=1}^{p} f_{ab} \nu_b^{k \to i (t+1)} \right] \tag{B5}$$

$$h_a^{(t+1)} = \frac{2\alpha}{N} \sum_{k=1}^{N} \sum_{b=1}^{p} f_{ab} \nu_b^{k\,(t+1)}, \tag{B6}$$

where we have added time indices in order to define update equations. For future convenience, we can also introduce an update function $\Phi^{\mathrm{update}}(\{\nu_k\}_{1\leqslant k\leqslant d}, f, h) \in \mathbb{R}^p$ and a normalization factor $Z(\{\nu_k\}_{1\leqslant k\leqslant d}, f, h) \in \mathbb{R}$, that both take as inputs d generic *messages* and return

$$Z(\{\nu_k\}_{1\leqslant k\leqslant d}, f, h) = \sum_{a=1}^{p} n_a\, \mathrm{e}^{-h_a} \prod_{1\leqslant k\leqslant d} \left[\sum_{b=1}^{p} f_{ab} \nu_{k,b} \right], \tag{B7}$$

$$\Phi_a^{\mathrm{update}}(\{\nu_k\}_{1\leqslant k\leqslant d}, f, h) = \frac{n_a\, \mathrm{e}^{-h_a}}{Z(\{\nu_k\}_{1\leqslant k\leqslant d}), f, h)} \prod_{1\leqslant k\leqslant d} \left[\sum_{b=1}^{p} f_{ab} \nu_{k,b} \right]. \tag{B8}$$

Once a fixed point to the cavity equations is found, its corresponding Bethe free entropy can be computed as

$$\frac{1}{N} F_{\mathrm{Bethe}}(\{\nu^{i\to j}\}, \{\nu^i\}) = \frac{1}{N} \sum_{i=1}^{N} \log(Z_i) - \frac{1}{N} \sum_{1\leqslant i<j\leqslant N, A_{ij}=1} \log(Z_{ij})$$
$$- \frac{1}{N} \sum_{1\leqslant i<j\leqslant N, A_{ij}=0} \log(\tilde{Z}_{ij}), \tag{B9}$$

where

$$Z_i = \sum_{a=1}^{p} n_a\, \mathrm{e}^{-h_a} \prod_{k\in\partial i} \left[\sum_{b=1}^{p} f_{ab} \nu_b^{k\to i} \right], \tag{B10}$$

$$Z_{ij} = \sum_{a,b} \nu_a^{i\to j} f_{ab} \nu_b^{j\to i}, \tag{B11}$$

$$\tilde{Z}_{ij} = \sum_{a,b} \nu_a^i \left(1 - 2\alpha f_{ab}/N\right) \nu_b^j = 1 - \frac{2\alpha}{N} \sum_{a,b} \nu_a^i f_{ab} \nu_b^j. \tag{B12}$$

The extra term $\tilde{Z}_{ij}$ comes from the contribution of non-edges, and has the same form of Z_{ij} (messages are replaced with local marginals thanks to relation (B3)). The sum $\sum \log(\tilde{Z}_{ij})$ can be further simplified. By expanding the logarithm to order $1/N$ and extending the summation over all the $j \in V$, one gets in the limit $N \to \infty$

$$\frac{1}{N} \sum_{1\leqslant i<j\leqslant N, A_{ij}=0} \log(\tilde{Z}_{ij}) = -\frac{2\alpha}{2} \sum_{a,b} \left[\frac{1}{N} \sum_{1\leqslant i\leqslant N} \nu_a^i \right] f_{ab} \left[\frac{1}{N} \sum_{1\leqslant i\leqslant N} \nu_b^i \right]$$
$$+ o(1) = -\alpha + o(1). \tag{B13}$$

The last equality comes from the fact that $N^{-1}\sum_i \nu_a^i$ converges to the prior n_a (the numerical solutions to the BP equations from the planted initialization being either

$\nu_a^i = n_a \ \forall \ i$, or a non-trivial one characterized by a large overlap with the planted configuration, $\nu_a^i \approx \delta_{a,x_i^0}$, where x_i^0 is extracted according to n_a) and from the normalization condition $\sum_a n_a f_{ab} = 1$. Moreover, from this also follows that the fixed-point auxiliary field h_a defined in (B6) will always be, up to finite size correction, equal to a constant $h_a = 2\alpha, \ \forall a \in \{0, \dots, p-1\}$. The contribution to the Bethe free entropy from $\tilde{Z}_{ij}$ and h_a (inside Z_i) is hence only an additive constant globally amounting to $-\alpha$, and we could disregard it, as already done before. We choose, however, to keep it in the definition of F_{Bethe}, therefore having

$$\frac{1}{N} F_{\text{Bethe}}(\{\nu^{i \to j}\}, \{\nu^i\}) = \frac{1}{N} \sum_{1 \leqslant i \leqslant N} \log(Z_i) - \frac{1}{N} \sum_{1 \leqslant i < j \leqslant N, A_{ij}=1} \log(Z_{ij}) - \alpha. \tag{B14}$$

The presence of the site-independent field h_a can be useful to cure some instability of the BP equations on a given graph. Since h_a becomes equal to a constant in the thermodynamic limit, we will in the following just remove it from every equation and assume $h_a = 0$ everywhere.

B.1. RS population dynamics

In order to analyze the properties of a system in the thermodynamic limit, one is usually interested in computing the quenched average over an ensemble of random graphs. In the replica symmetric cavity equations one assumes that messages incoming to a variable node are independent, and one can therefore describe the properties of a typical fixed point in the large N limit using a unique distribution of messages $\mathbb{P}(\nu)$. The update equation (B4) can then be generalized to

$$\mathbb{P}^{(t+1)}(\nu) = \sum_d P_{\text{Edge}}(d) \int \prod_{k=1}^d \mathrm{d}\nu_k \, \mathbb{P}^{(t)}(\nu_k) \delta \left[\nu - \Phi^{\text{update}}(\{\nu_k\}_d, f) \right], \tag{B15}$$

where Φ^{update} is given by equation (B8), and for Erdős–Rényi random graphs one has

$$P_{\text{Node}}(d) = \exp(-2\alpha) \frac{(2\alpha)^d}{d!} \tag{B16}$$

$$P_{\text{Edge}}(d) = \frac{(d+1) P_{\text{Node}}(d+1)}{\sum_{k=0}^{+\infty} k P_{\text{Node}}(k)} = P_{\text{Node}}(d). \tag{B17}$$

From the equivalence of $P_{\text{Node}}(d)$ and $P_{\text{Edge}}(d)$, it also follows that for Erdős–Rényi random graphs messages $\nu^{i \to j}$ and variable marginals ν^i are subjected to the same statistics $\mathbb{P}(\nu)$ given by equation (B15). In the following we will thus neglect the superscripts and work with a unique family $\{\nu\}$ of messages. The Bethe free energy (B14) can be rewritten as

Algorithm 1. Replica symmetric population dynamics.

1: $\mathbb{P}^{\text{input}} = \{\nu_k\}_{1 \leqslant k \leqslant U}$
2: $\mathbb{P}^{\text{output}} \Leftarrow \{\}$
3: For $i = 1 \ldots U$ do
4:    Sample d according to $P_{\text{Edge}}(d)$
5:    $S \Leftarrow \{\}$
6:    For $j = 1 \ldots d$ do
7:     Sample ν_j uniformly from $\mathbb{P}^{\text{input}}$
8:     $S \Leftarrow S \cup \{\nu_j\}$
9:    End for
10:   $\nu \Leftarrow \Phi^{\text{update}}(S, f)$
11:   $\mathbb{P}^{\text{output}} \Leftarrow \mathbb{P}^{\text{output}} \cup \{\nu\}$
12: End for
13: Return $\mathbb{P}_{\text{output}}$

$$\frac{1}{N} F_{\text{RS}}(\mathbb{P}) = \sum_d P_{\text{Node}}(d) \int \prod_{k=1}^{d} \mathrm{d}\nu_k \, \mathbb{P}(\nu_k) \log \left[Z(\{\nu_k\}_d, f) \right]$$

$$- \alpha \int \mathrm{d}\nu_1 \, \mathrm{d}\nu_2 \mathbb{P}(\nu_1) \mathbb{P}(\nu_2) \log \left[\sum_{ab} \nu_{1,a} f_{ab} \nu_{2,b} \right] - \alpha, \qquad \text{(B18)}$$

where Z is given by (B7).

In practice, one represents $\mathbb{P}(\nu)$ by a finite population of U messages, with U large enough (typically, $U = O(10^6)$ messages), and estimates (B15) by sampling messages from this set. The following algorithm (algorithm 1) updates the starting population $\mathbb{P} = \{\nu_k\}_{1 \leqslant k \leqslant U}$ according to equation (B15).

B.2. Reconstruction equations for $m = 1$

The assumption that incoming messages are independent can end up being false if the density of the constraints α is too large. In a factor graph with locally tree-like structure, this is usually due to the fact that variables start to develop correlations on large distances, together with the presence of long loops, which may then become relevant. When this is the case, one has to resort to the 1RSB cavity equations to study the system. The 1RSB population dynamics equations generally describe the distribution of distributions of messages, and are therefore computationally much heavier to implement. Luckily, limiting oneself to the case of Parisi replica symmetry breaking parameter $m = 1$, and when the BP equations exhibit a para-magnetic fixed point, one can perform a simplification of the 1RSB equations that gives them an RS structure. Since the 1RSB computation at $m = 1$ is correct up to α_c, this allow us to derive the dynamical and Kauzmann transitions.

As a result of this simplification, rather than having only one distribution of messages $\mathbb{P}(\nu)$, one keeps track of p distributions $\mathbb{P}_a(\nu), a \in \{0, \ldots, p-1\}$, where p is the size of the (discrete) alphabet that variables can assume [18]. The update equation (B15) takes

Algorithm 2. General reconstruction equation population dynamics.

1: $\mathbb{P}_a^{\text{input}} = \{\nu_k\}_{1 \leqslant k \leqslant U} \;\; \forall a \in \{0, \ldots, p-1\}$
2: $\mathbb{P}_a^{\text{output}} \Leftarrow \{\} \;\; \forall a \in \{0, \ldots, p-1\}$
3: For $a = 0 \ldots (p-1)$ do
4:      For $i = 1 \ldots U$ do
5:          Sample d according to $P_{\text{Edge}}(d)$
6:          $S \Leftarrow \{\}$
7:          For $j = 1 \ldots d$ do
8:              Sample $b \in \{0, \ldots, p-1\}$ with probability $f_{ab} n_b$
9:              Sample ν_j uniformly from $\mathbb{P}_b^{\text{input}}$
10:             $S \Leftarrow S \cup \{\nu_j\}$
11:          End for
12:          $\nu \Leftarrow \Phi^{\text{update}}(S, f)$
13:          $\mathbb{P}_a^{\text{output}} \Leftarrow \mathbb{P}_a^{\text{output}} \cup \{\nu\}$
14:      End for
15: End for
16: Return $\mathbb{P}_a^{\text{output}}$

---

the following form $\forall a$

$$\mathbb{P}_a^{(t+1)}(\nu) = \sum_d P_{\text{Edge}}(d) \int \prod_{k=1}^d \left[\mathrm{d}\nu_k \sum_b n_b f_{ab} \mathbb{P}_b^{(t)}(\nu_k) \right] \delta\left(\nu - \Phi^{\text{update}}(\{\nu_k\}_d, f)\right). \quad \text{(B19)}$$

The Bethe free entropy becomes now

$$\frac{1}{N} F_{\text{Rec}}(\{\mathbb{P}_a\}) = \sum_d P_{\text{Node}}(d) \sum_a n_a \int \prod_{k=1}^d \left[\mathrm{d}\nu_k \sum_b n_b f_{ab} \mathbb{P}_b(\nu_k) \right] \log\left[Z(\{\nu_k\}_d, f) \right]$$
$$- \alpha \sum_{a,b} n_a n_b f_{ab} \int \mathrm{d}\nu_1 \, \mathrm{d}\nu_2 \, \mathbb{P}_a(\nu_1) \mathbb{P}_b(\nu_2) \log\left[\sum_{r,s} \nu_{1,r} f_{rs} \nu_{2,s} \right] - \alpha.$$
$$\text{(B20)}$$

To iteratively find a non-trivial fixed point to (B19), one starts with p populations $\mathbb{P}_a^{\text{Init}} = \{\nu_k\}_{1 \leqslant k \leqslant U}$ initialized in the planted solution

$$\nu_{k,b} = \delta_{a,b} \quad \forall \, \nu_k \in \mathbb{P}_a^{\text{Init}}, \; a \in \{0, \ldots, p-1\}. \quad \text{(B21)}$$

The algorithm implementing equation (B19) is given in algorithm 2.

As long as one works with a flat prior, as $P_X(x) = \frac{1}{2\pi}$ or $n_a = \frac{1}{p} \equiv n \; \forall a \in \{0, \ldots, p-1\}$, and the matrix f_{ab} depends only on the modulus of the difference between variables $f_{ab} = f_{0,(b-a+p) \bmod p}$, the system has a global rotational symmetry and one can avoid keeping track of p populations P_a, while can just focus on messages centered around zero and then shift them. The presence of a rotational symmetry implies

$$\mathbb{P}_b(\nu) = \mathbb{P}_a(T(a-b)\nu), \quad \text{(B22)}$$

Algorithm 3. Simplified reconstruction equation population dynamics (II).

1: $\mathbb{P}^{\text{input}} = \{\nu_k\}_{1 \leqslant k \leqslant U}$
2: $\mathbb{P}^{\text{output}} \Leftarrow \{\}$
3: For $i = 1 \dots U$ do
4:     Sample d according to $P_{\text{Edge}}(d)$
5:     $S \Leftarrow \{\}$
6:     For $j = 1 \dots d$ do
7:         Sample $b \in \{0, \dots, (p-1)\}$ with probability $n f_b$
8:         Sample ν_j uniformly from $\mathbb{P}^{\text{input}}$
9:         $S \Leftarrow S \cup \{T(-b)\nu_j\}$
10:     End for
11:     $\nu \Leftarrow \Phi^{\text{update}}(S, f)$
12:     $\mathbb{P}^{\text{output}} \Leftarrow \mathbb{P}^{\text{output}} \cup \{\nu\}$
13: End for
14: Return $\mathbb{P}^{\text{output}}$

---

where $T(\Delta x)$ is the linear transformation that shifts a message $\nu(x)$ by an amount Δx. This allows us to simplify the equations, by keeping track of one distribution of messages $\mathbb{P}$ only, as in the RS case. For convenience, we fix the index a in equation (B19) to $a = 0$, corresponding to messages centered around $x = 0$. We also name $f_a \equiv f_{0a}$ and $f_{b-a} \equiv f_{0,(b-a+p) \bmod p}$ (the matrix f_{ab} being a single-variable function f_a for $a \in \{0, \dots p-1\}$, with parity $f_a = f_{p-a}$). The update equation for the distribution of messages $\mathbb{P}$, along with the planted initial condition, then becomes

$$\mathbb{P}^{(t+1)}(\nu) = \sum_d P_{\text{Edge}}(d) \int \prod_{k=1}^d \left[d\nu_k \sum_b n f_b \mathbb{P}^{(t)}\left(T(-b)\nu_k\right) \right]$$
$$\times \delta\left(\nu - \Phi^{\text{update}}(\{\nu_k\}_d, f)\right), \tag{B23}$$

$$\nu_{k,b} = \delta_{0,b} \quad \forall \, \nu_k \in \mathbb{P}^{\text{Init}}. \tag{B24}$$

The implementation of the precedent equation is displayed in algorithm 3. The procedure formally consists in an RS population dynamics, where a planting is enforced through the introduction of random shifts for the messages extracted according to the interaction probability $n f_b$. Finally, the 1RSB $m = 1$ Bethe free entropy takes the form

$$\frac{1}{N} F_{\text{Rec II}}(\mathbb{P}) = \sum_d P_{\text{Node}}(d) \int \prod_{k=1}^d \left[d\nu_k \sum_b n f_b \mathbb{P}^{(t)}\left(T(-b)\nu_k\right) \right] \log\left[Z(\{\nu_k\}, f)\right] \tag{B25}$$
$$- \alpha \int d\nu_1 \, d\nu_2 \, \mathbb{P}^{(t)}(\nu_1) \sum_b n f_b \mathbb{P}^{(t)}(T(-b)\nu_2) \log\left[\sum_{r,s} \nu_{1,r} f_{s-r} \nu_{2,s} \right] - \alpha.$$

Algorithm 4. Simplified reconstruction equation with tracking of first order difference.

1: $\mathbb{P}^{\mathrm{input}} = \{(\nu_k, \epsilon_k)\}_{1 \leqslant k \leqslant U}$
2: $\mathbb{P}^{\mathrm{output}} \Leftarrow \{\}$
3: For $i = 1 \ldots U$ do
4:     Sample d according to $P_{\mathrm{Edge}}(d)$
5:     $S \Leftarrow \{\}$
6:     For $j = 1 \ldots d$ do
7:         Sample $b \in \{0, \ldots, p-1\}$ with probability nf_b
8:         Sample (ν_j, ϵ_j) uniformly from $\mathbb{P}^{\mathrm{input}}$
9:         $S \Leftarrow S \cup \{T(-b)\nu_j, T(-b)\epsilon_j\}$
10:     End for
11:     $\nu \Leftarrow \Phi^{\mathrm{update}}(S, f)$
12:     $\epsilon \Leftarrow \sum_{k=1}^{d} \frac{\partial \Phi^{\mathrm{update}}(S, f)}{\partial \nu_k} \epsilon_k$
13:     $P^{\mathrm{output}} \Leftarrow P^{\mathrm{output}} \cup \{\nu, \epsilon\}$
14: End for
15: Return $\mathbb{P}^{\mathrm{output}}$

B.3. Stability analysis

The stability of the BP fixed points can be analyzed by using the population dynamics equations. One way to do this is by creating two copies of the same population and then by slightly perturbing one of the two, i.e. $P_1 = \{\nu_k\}_{1 \leqslant k \leqslant U}$ and $P_2 = \{\nu_k + \epsilon_k\}_{1 \leqslant k \leqslant U}$, where the ϵ_k's are small perturbations to the fixed point messages. One then keep updating in parallel both of them, using the same choice of messages, and tracks whether and how the difference between the two populations grows or decreases. A cleaner way to perform this task is by only keeping track of the difference between the original and perturbed messages to first order. This leads to some modified version of the reconstruction equations, where one simultaneously evolve both the messages ν_k and their linear perturbations ϵ_k,

$$
\mathbb{P}^{(t+1)}(\nu, \epsilon) = \sum_d P_{\mathrm{Edge}}(d) \int \prod_{k=1}^{d} \left[\mathrm{d}\nu_k \, \mathrm{d}\epsilon_k \sum_b nf_b \mathbb{P}^{(t)} \left(T(-b)\nu_k, T(-b)\epsilon_k \right) \right]
$$
$$
\times \delta \left(\nu - \Phi^{\mathrm{update}}(\{\nu_k\}, f) \right) \delta \left(\epsilon - \sum_{j=1}^{d} \frac{\partial \Phi^{\mathrm{update}}(\{\nu_k\}, f)}{\partial \nu_j} \epsilon_j \right).
$$
(B26)

The implementation of the above equations is given in algorithm 4. In the argument of the second δ function in (B26) we have used a shorthand notation for the following matrix product

$$\left(\frac{\partial \Phi^{\text{update}}(\{\nu_k\}, f)}{\partial \nu_j}\epsilon_j\right)_a = \sum_b \frac{\partial \Phi_a^{\text{update}}(\{\nu_k\}, f)}{\partial m_{j,b}}\epsilon_{j,b}$$

$$= \frac{n}{Z(\{\nu_k\}, f)}\left(\prod_{1 \leqslant k \leqslant d, k \neq j} \hat{\nu}_{k,a}\right)\hat{\epsilon}_{j,a}$$

$$- \frac{\Phi_a^{\text{update}}(\{\nu_k\}, f)}{Z(\{\nu_k\}, f)}\sum_c n\left(\prod_{1 \leqslant k \leqslant d, k \neq j} \hat{\nu}_{k,c}\right)\hat{\epsilon}_{j,c}, \quad \text{(B27)}$$

where we have introduced for convenience the auxiliary (unnormalized) messages

$$\hat{\nu}_{k,a}(\nu_k, f) = \sum_b \nu_{k,b}f_{b-a}, \tag{B28}$$

$$\hat{\epsilon}_{k,a}(\epsilon_k, f) = \sum_b \epsilon_{k,b}f_{b-a}. \tag{B29}$$

References

[1] Zdeborova L and Krzakala F 2007 Phase transitions in the coloring of random graphs *Phys. Rev.* E **76** 031131
[2] Mari R, Krzakala F and Kurchan J 2009 Jamming versus glass transitions *Phys. Rev. Lett.* **103** 025701
[3] Krzakala F and Kurchan J 2007 Landscape analysis of constraint satisfaction problems *Phys. Rev.* E **76** 021122
[4] Mezard M, Parisi G, Tarzia M and Zamponi F 2011 On the solution of a 'solvable' model of an ideal glass of hard spheres displaying a jamming transition (arXiv:1011.5080)
[5] Yoshino H 2018 Disorder-free spin glass transitions and jamming in exactly solvable mean-field models *SciPost Phys.* **4** 040
[6] Krzakala F, Montanari A, Ricci-Tersenghi F, Semerjian G and Zdeborova L 2007 Gibbs states and the set of solutions of random constraint satisfaction problems *Proc. Natl Acad. Sci. USA* **104** 10318–23
[7] Montanari A, Ricci-Tersenghi F and Semerjian G 2008 Clusters of solutions and replica symmetry breaking in random k-satisfiability *J. Stat. Mech.* P04004
[8] Baldassi C, Ingrosso A, Lucibello C, Saglietti L and Zecchina R 2015 Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses *Phys. Rev. Lett.* **115** 128101
[9] Baldassi C, Ingrosso A, Lucibello C, Saglietti L and Zecchina R 2016 Local entropy as a measure for sampling solutions in constraint satisfaction problems *J. Stat. Mech.* 023301
[10] Baldassi C, Borgs C, Chayes J T, Ingrosso A, Lucibello C, Saglietti L and Zecchina R 2016 Unreasonable effectiveness of learning neural networks: from accessible states and robust ensembles to basic algorithmic schemes *Proc. Natl Acad. Sci. USA* **113** E7655–62
[11] Han Z and Zhou H-J 2020 Maximally flexible solutions of a random K-satisfiability formula (arXiv:2006.07023)
[12] Braunstein A, Dall'Asta L, Semerjian G and Zdeborová L 2016 The large deviations of the whitening process in random constraint satisfaction problems *J. Stat. Mech.* 053401
[13] Budzynski L, Ricci-Tersenghi F and Semerjian G 2019 Biased landscapes for random constraint satisfaction problems *J. Stat. Mech.* 023302
[14] Budzynski L and Semerjian G 2020 Biased measures for random constraint satisfaction problems: larger interaction range and asymptotic expansion *J. Stat. Mech.* 103406
[15] Sellitto M and Zamponi F 2013 A thermodynamic description of colloidal glasses *Europhys. Lett.* **103** 46005
[16] Maimbourg T, Sellitto M, Semerjian G and Zamponi F 2018 Generating dense packings of hard spheres by soft interaction design *SciPost Phys.* **4** 039
[17] Krzakala F and Zdeborová L 2009 Hiding quiet solutions in random constraint satisfaction problems *Phys. Rev. Lett.* **102** 238701
[18] Mézard M and Montanari A 2012 *Information, Physics, and Computation* (*Oxford Graduate Texts*) (Oxford: Oxford University Press)

[19] Kesten H and Stigum B P 1966 Additional limit theorems for indecomposable multidimensional Galton–Watson processes *Ann. Math. Stat.* **37** 1463–81

[20] Mossel E and Peres Y 2003 Information flow on trees *Ann. Appl. Probab.* **13** 817–44

[21] Janson S and Mossel E 2004 Robust reconstruction on trees is determined by the second eigenvalue *Ann. Probab.* **32** 2630–49

[22] de Almeida J R L and Thouless D J 1978 Stability of the Sherrington–Kirkpatrick solution of a spin glass model *J. Phys. A: Math. Gen.* **11** 983–90

[23] Decelle A, Krzakala F, Moore C and Zdeborová L 2011 Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications *Phys. Rev.* E **84** 066106

[24] Sciortino F 2002 One liquid, two glasses *Nat. Mater.* **1** 145–6

[25] Dawson K, Foffi G, Fuchs M, Götze W, Sciortino F, Sperl M, Tartaglia P, Voigtmann T and Zaccarelli E 2000 Higher-order glass-transition singularities in colloidal systems with attractive interactions *Phys. Rev.* E **63** 011401

[26] Charbonneau P and Kundu J 2020 Postponing the dynamical transition density using competing interactions *Granular Matter* **22** 55

[27] Poon W C K 2002 The physics of a model colloid polymer mixture *J. Phys.: Condens. Matter* **14** R859–80

[28] Eckert T and Bartsch E 2002 Re-entrant glass transition in a colloid-polymer mixture with depletion attractions *Phys. Rev. Lett.* **89** 125701

[29] Montanari A and Semerjian G 2006 On the dynamics of the glass transition on Bethe lattices *J. Stat. Phys.* **124** 103–89

[30] Lupo C and Ricci-Tersenghi F 2017 Approximating the XY model on a random graph with a q-state clock model *Phys. Rev.* B **95** 054433