

# Phase transitions in the $q$-coloring of random hypergraphs

Marylou Gabrié, Varsha Dani, Guilhem Semerjian, Lenka Zdeborová

▶ **To cite this version:**

Marylou Gabrié, Varsha Dani, Guilhem Semerjian, Lenka Zdeborová. Phase transitions in the $q$-coloring of random hypergraphs. Journal of Physics A: Mathematical and Theoretical, 2017, 50, pp.505002. 10.1088/1751-8121/aa9529 . cea-01614263

**HAL Id: cea-01614263**

**https://cea.hal.science/cea-01614263**

Submitted on 10 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Phase transitions in the $q$-coloring of random hypergraphs

Marylou Gabrié
*Laboratoire de physique statistique, Département de physique de l'ENS, Ecole normale supérieure, Université Paris Diderot, UPMC Univ. Paris 06, CNRS, PSL Research University, 75005 Paris, France*
*Sorbonne Universités, UPMC Univ. Paris 06, Université Paris Diderot, Ecole normale supérieure, CNRS, Laboratoire de Physique Statistique (LPS ENS), 75005 Paris, France and*
*Université Paris Diderot, Sorbonne Paris Cité, Laboratoire de Physique Statistique de l'ENS (LPS ENS), Ecole normale supérieure, UPMC Univ. Paris 06, CNRS, 75005 Paris, France*

Varsha Dani
*University of New Mexico.*

Guilhem Semerjian
*Laboratoire de physique théorique, Département de physique de l'ENS, Ecole normale supérieure, UPMC Univ. Paris 06, CNRS, PSL Research University, 75005 Paris, France and*
*Sorbonne Universités, UPMC Univ. Paris 06, Ecole normale supérieure, CNRS, Laboratoire de Physique Théorique (LPT ENS), 75005 Paris, France*

Lenka Zdeborová
*Institut de Physique Théorique, CNRS, CEA, Université Paris-Saclay, F-91191, Gif-sur-Yvette, France.*

(Dated: July 10, 2017)

We study in this paper the structure of solutions in the random hypergraph coloring problem and the phase transitions they undergo when the density of constraints is varied. Hypergraph coloring is a constraint satisfaction problem where each constraint includes $K$ variables that must be assigned one out of $q$ colors in such a way that there are no monochromatic constraints, i.e. there are at least two distinct colors in the set of variables belonging to every constraint. This problem generalizes naturally coloring of random graphs ($K = 2$) and bicoloring of random hypergraphs ($q = 2$), both of which were extensively studied in past works. The study of random hypergraph coloring gives us access to a case where both the size $q$ of the domain of the variables and the arity $K$ of the constraints can be varied at will. Our work provides explicit values and predictions for a number of phase transitions that were discovered in other constraint satisfaction problems but never evaluated before in hypergraph coloring. Among other cases we revisit the hypergraph bicoloring problem ($q = 2$) where we find that for $K = 3$ and $K = 4$ the colorability threshold is not given by the one-step-replica-symmetry-breaking analysis as the latter is unstable towards more levels of replica symmetry breaking. We also unveil and discuss the coexistence of two different 1RSB solutions in the case of $q = 2$, $K \geq 4$. Finally we present asymptotic expansions for the density of constraints at which various phase transitions occur, in the limit where $q$ and/or $K$ diverge.

PACS numbers: Classification, Keywords

arXiv:1707.01983v1 [cond-mat.dis-nn] 6 Jul 2017

## I. INTRODUCTION

Constraint satisfaction problems (CSPs) where values are assigned to a set of variables in such a way to satisfy a given set of constraints arise in many areas of science. In particular they stand at the core of the theory of computational complexity [1] and analysis of algorithms. The study of random constraint satisfaction problems became popular in the endeavor to analyze the computational complexity of typical instances, instead of the worst possible case [2, 3]. In random CSPs the set of constraints is chosen from some simply defined probability distribution, typical instances being those sampled with high probability from this distribution. Random CSPs are formally closely related to models considered in statistical physics of disordered systems and spin glasses [4, 5], and the application of statistical mechanics methods to random CSPs turned out to be particularly fruitful.

The most widely studied random CSP is satisfiability of random Boolean formulas ($K$-SAT). A series of works coming from statistical physics brought an explicit prediction for the so-called satisfiability threshold – a density of constraints beyond which the formula is with high probability unsatisfiable [6, 7]. These studies also unveiled a peculiar structure of the satisfiable assignments in the satisfiable phase and additional phase transitions, besides the satisfiability one, at which the qualitative structure of the set of solutions changes drastically [6, 8–10]. This understanding inspired the development of a new class of message-passing-based satisfiability solvers that outperform existing ones for randomly generated formulas [6]. While many of these results were origi-

nally obtained using non-rigorous methodology (namely the cavity and the replica method), in recent years a large part of these results and the associated picture was established rigorously, see e.g. [11–15], leading to a new methodology in probability theory able to deal with probability distributions associated to randomly generated constraints on a large set of Boolean variables.

Studies of random CSPs were of course not restricted to random $K$-SAT, the statistical mechanics methodology being very versatile and easily adapted to other type of constraints and of variables. For instance random graph coloring is an eminent example in which the variables can take more than two values [16–18]. On the other hand in graph coloring every constraint corresponds to an edge and hence contains only two variables. Some of these results have been rigorously established, see e.g. in [19–23], although the larger size of variables domain does complicate the situation. Some of the techniques are simpler for a version of the random graph coloring problem where every edge forbids a random set of neighboring colors [24]. This "permutated" version of the coloring problem is conjectured to be equivalent to the usual one in terms of the structure of solutions and the phase transitions. However, proving this rigorously remains an open problem.

Another popular random CSP with Boolean variables and $K$-ary constraints is Not-All-Equal satisfiability (NAE-SAT) where each constraint contains $K$ random literals (i.e. variables randomly negated or not). This problem possesses a symmetry that is absent in random $K$-SAT and renders the problem simpler to analyze. A conjecture that follows from the statistical physics analysis is that the structure of solutions and associated phase transitions in random NAE-SAT are the same for bicoloring of random hypergraphs, where each constraint contains $K$ variables and states that they should not all take the same value. Hypergraphs generalize the concept of a graph in the sense that hyperedges connect to an arbitrary number of variables ($K$ in our case) whereas in graph an edge connects two variables. Statistical physics studies of the hypergraph bicoloring can be found in [25–27]. Again, large part of the resulting picture was established rigorously [12, 14, 28–32].

Hypergraph coloring is another very natural CSP that generalized both graph coloring and hypergraph bicoloring. Coloring of hypergraphs as defined in [33] consists of requiring that in the set of $K$ nodes belonging to a given constraint each of the $q$ colors appears at most $\gamma$ times, with $1 \leq \gamma \leq K-1$. The authors of [34, 35] then analyzed coloring of random hypergraphs for $\gamma = K-1$, i.e. the constraints only forbid configurations where all the nodes that belong to one constraint have the same color. This is also the version of the problem we consider in the present paper, in which the statistical mechanics methodology is applied for the first time to the random hypergraph coloring with arbitrary values of $K$ and $q$, yielding results on the structure of solutions and phase transitions comparable to the state-of-the-art for random graph coloring and random $K$-SAT.

Besides the quantitative computation of the thresholds for various phase transitions in this problem, we shall present a number of qualitatively new results. While the bicoloring ($q = 2$) problem was studied previously in a number of papers, we found indeed some behaviors and results that were missed. Our first main result concerns the fact that on Erdős-Rényi random $K$-uniform hypergraphs with $K = 3$ and $K = 4$ the colorability threshold is not given by the one-step-replica-symmetry breaking (1RSB) analysis. An explicit prediction for the colorability threshold in those cases thus remains an open problem. This contrasts with the conjectures existing for random graph coloring, and random satisfiability [17, 36]. The second main result is the coexistence of two distinct 1RSB solutions for $q = 2$ and $K = 4$ in the satisfiable region. Finally, another important outcome of this paper consists in asymptotic expansions for the phase transition thresholds in a model where both the parameters $K$ and $q$ can diverge, as far as we know no such case was discussed previously in the statistical mechanics literature.

The rest of the paper is organized as follows. In Sec. II we define precisely the model under study, phrase it as a statistical mechanics problem and discuss qualitatively the various phase transitions it undergoes. In Sec. III we derive the equations that describe the problem in the formalism of the cavity method (including the phenomenon of replica symmetry breaking). Our main results are presented in Sec. IV, where we discuss the numerical resolution of these equations, as well as their asymptotic expansions for large $K$ and/or $q$. We present our conclusions and perspectives for future work in Section V, while a series of Appendices contain some more technical details.

## II. STATISTICAL PHYSICS FORMULATION OF HYPERGRAPH $q$-COL

### A. Model

#### 1. *Constraint satisfaction problems as interacting spin models*

In a constraint satisfaction problem (CSP) $N$ variables $\{x_i\}_{i=1}^N$, each taking values in a finite set $\mathcal{Q}$, are required to satisfy a set of $M$ constraints. The $\mu$-th constraint (also called clause, and indexed with $\mu \in \{1, \dots, M\}$) involves a subset $\partial\mu \subset \{1, \dots, N\}$ of the variables, and is defined by a function $\Delta_\mu(\mathbf{x}_{\partial\mu})$ equal to 1 if the constraint is satisfied, and 0 otherwise (here and in the following we denote $\mathbf{x}_S = \{x_i\}_{i\in S}$ the configuration of a subset $S$ of the variables). An assignment $\mathbf{x} = \{x_1, .., x_N\} \in \mathcal{Q}^N$ of the variables is called a solution if it simultaneously satisfies all the constraints. A CSP is said to be satisfiable if it admits at least one solution.

Such a CSP can be conveniently represented by a factor

graph of $N$ vertices and $M$ factors, drawn respectively as circles and squares in Fig. 1. Each vertex $i$ carries the variable $x_i$, each factor $\mu$ is associated to the clause $\Delta_\mu$, an edge being drawn between a factor $\mu$ and a variable $i$ if and only if $\Delta_\mu$ depends on $x_i$. This gives a graphical interpretation of $\partial\mu$ as the variables adjacent to $\mu$ in the factor graph, and we shall denote similarly $\partial i$ the set of contraints the variable $x_i$ is involved in.

A bridge towards statistical mechanics is built by interpreting the variables $x_i$ as spins taking a finite number of distinct values ($\mathcal{Q} = \{-1, 1\}$ for Ising spins, $\mathcal{Q} = \{1, \dots, q\}$ for Potts ones) and defining an Hamiltonian (or energy, or cost function) on the set of spin configurations as

$$\mathcal{H}(\mathbf{x}) = \sum_{\mu=1}^{M} \left(1 - \Delta_\mu(\mathbf{x}_{\partial\mu})\right) , \tag{1}$$

which counts the number of unsatisfied constraints. Each spin interacts with its nearest neighbors along the CSP factor graph: a clause $\mu$ involving $K$ spins models a $K$-wise interaction, so that the energy shift for breaking the constraint is equal to 1. Solutions of the CSP corresponds to zero energy configurations, hence an instance of a CSP is satisfiable if and only if its groundstate (minimal) energy is equal to zero. The Gibbs Boltzmann probability density over the configuration space associated to this Hamiltonian is

$$p(\mathbf{x}) = \frac{e^{-\beta\mathcal{H}(\mathbf{x})}}{\mathcal{Z}(\beta)} , \tag{2}$$

where $\beta$ is a parameter known as the inverse temperature and the partition function $\mathcal{Z}(\beta)$ normalizes this probability law. In the zero temperature limit ($\beta \to \infty$) $p$ concentrates on the optimal configurations minimizing the number of unsatisfied constraints. In particular if the instance is satisfiable $p$ becomes in this limit the uniform measure over the solutions, and $\mathcal{Z}$ counts then the number of these satisfying assignments.

### 2. Hypergraph q-Col

We shall consider in this paper the hypergraph $q$-coloring problem: it consists in assigning one color, among $q$ available, to each vertex $i$ so that none of the hyperedges $\mu$ is monochromatic (an hyperedge is said to be monochromatic if all the vertices it involves carry the same color).

This definition fits naturally in the framework defined above, by choosing the alphabet $\mathcal{Q} = \{1, 2, .., q\}$ to encode the possible colors of the vertices, denoting $N$ and $M$ the number of vertices and hyperedges of the hypergraph, and associating to each hyperedge $\mu$ a constraint function $\Delta_\mu(\mathbf{x}_{\partial\mu}) = 1 - \delta\left(\mathbf{x}_{\partial\mu}\right)$, with $\delta(x_1, \dots, x_K) = 1$ if all elements in $x_1, \dots, x_K$ are equal and $\delta(x_1, \dots, x_K) = 0$ otherwise, in such a way that $\Delta_\mu$ equals 1 if and only if the $\mu$-th hyperedge is non-monochromatic. The Hamiltonian thus becomes

$$\mathcal{H}(\mathbf{x}) = \sum_{\mu=1}^{M} \delta(\mathbf{x}_\mu) , \tag{3}$$

a generalization of the usual Potts antiferromagnet from graphs ($K = 2$) to hypergraph ($K \geq 3$) interactions.

### 3. Random ensembles for typical case analysis

As explained in the introduction random ensembles of instances provide useful benchmarks to assess the typical difficulty of a CSP; in the case of hypergraph $q$-coloring an instance is fully specified by an hypergraph, we shall thus consider ensembles of random hypergraphs, and in particular the following two distributions:

(i) the $K$-uniform $\ell$-regular random hypergraph, for which each hyperedge links $K$ vertices, and each vertex belongs to $\ell$ hyperedges, all the hypergraphs fulfilling these conditions being equiprobable in this ensemble. The numbers $N$ of vertices and $M$ of hyperedges must obviously obey the relationship $N\ell = MK$.

(ii) the $K$-uniform Erdős - Rényi (ER) hypergraph, where $M$ hyperedges are chosen uniformly at random among all $\binom{N}{K}$ possible $K$-uplets of distinct variables.

Explicit analytic results can be obtained for the typical properties of such random structures in the large size limit (called thermodynamic limit in physics jargon), in which both $N$ and $M$ go to infinity at a fixed ratio denoted $\alpha = M/N$. A crucial property shared by these two ensembles is the local convergence in this limit to an hypertree: the neighborhood within a fixed distance around an uniformly chosen vertex is, with a probability going to 1 in the thermodynamic limit, acyclic. This limit tree can be described in terms of a Galton-Watson branching process, in which the root has a number $d$ of neighboring hyperedges with probability $p_d$, each of the $d(K-1)$ vertices at distance 1 from the root giving themselves birth to a number of offsprings with a probability distribution denoted $r_d$, and so on and so forth. The excess degree distribution $r_d$ is the probability that one vertex in an uniformly chosen hyperedge has degree $d+1$, and is related to $p_d$ through

$$r_d = \frac{(d+1)p_{d+1}}{\sum_{d'=1}^{\infty} d'\, p_{d'}} . \tag{4}$$

In the regular ensemble, case (i) above, all vertices have degree $\ell$, hence $p_d = \delta_{d,\ell}$ and $r_d = \delta_{d,\ell-1}$ with $\delta$ denoting here the Kronecker symbol. The ratio $\alpha$ of constraints by variables is given in terms of the parameters of this ensemble as $\alpha = \ell/K$.

In the Erdős - Rényi case both $p_d$ and $r_d$ are found to be Poisson laws of average $\alpha K$. With a slight abuse of notation we shall denote $\ell = \alpha K$ the average degree in

this case, there should be no possibility of confusion with the degree of the regular ensemble.

Some of our results will be obtained in the large degree limit (taken after the thermodynamic limit), in which a Poisson distributed random variable concentrates around its mean and hence the two ensembles become essentially equivalent, which will allow us to use the one in which the computations are the simplest.

### B. Phase transitions and thresholds

The random ensembles of hypergraph $q$-coloring are parametrized by $q$, the number of colors, $K$ the arity of the constraints, and $\alpha$ the ratio of the number of constraints per variable (or equivalently $\ell = \alpha K$ the average degree of the variables). For a given choice of $q$ and $K$ the problem becomes more and more constrained as $\ell$ increases, and one can naturally expect the existence of a critical value $\ell_{\rm col}$ below which typical random hypergraphs are $q$-colorable, while with high probability they are un-colorable for $\ell > \ell_{\rm col}$. This satisfiability phase transition (or threshold phenomenon) is actually accompanied by a series of phase transitions affecting the structure of the space of solutions in the satisfiable phase $\ell < \ell_{\rm col}$. These structural phase transitions have been discovered by the applications of statistical mechanics methods, they are essentially the same as those discussed in the mean field theory of structural glasses [37], and are largely independent of the details in the definition of the random CSP [5, 10]. In the next section we first explain these phase transitions in a qualitative way, before moving gradually to a more quantitative and technical description.

#### 1. *Qualitative picture of the phase diagram*

At low density of constraints all solutions are found in one "connected" portion of configuration space, meaning that it is possible to go from any solution to any other solution by "flipping" only one (or a number sub-linear in $N$) variable at each step without breaking any constraint along the path. Equivalently, the set of solutions is an ergodic subspace and Monte Carlo Markov chains can achieve efficient sampling of solutions.

As the density of constraints is increased a first remarkable transition is encountered: at the *clustering* transition the set of solutions decomposes in an exponential (in $N$) number of "disconnected" smaller sets, called *clusters*. Following the physics formulation of CSPs presented in the previous section, there are either energetic or entropic barriers between different clusters while there exist zero energy paths between solutions in the same cluster. The precise definition of a cluster may vary among authors of previous works, we shall clarify our operational definition of this threshold $\ell_{\rm clust}$ in Sec. III while detailing our computations.

Further increasing $\ell$ beyond $\ell_{\rm clust}$, a *condensation* transition occurs at $\ell_{\rm cond}$: it separates a regime in which an exponential number of clusters contain most of the solutions, to a situation for $\ell > \ell_{\rm cond}$ in which most of the solutions are found in an only sub-exponential number of clusters.

Eventually, the disappearance of the last surviving clusters of solutions marks the *colorability* (or *satisfiability* for a generic CSP) threshold, at $\ell_{\rm col}$.

The partition of the set of solutions into clusters allows to define the notion of the *frozen* variables of a solution: these are the variables which take the same color in all the solutions of the corresponding cluster. One can then further refine the description of the clustered phase and introduce two new phase transitions: the *rigidity* transition, denoted $\ell_{\rm r}$, above which the typical solutions contain an extensive number of frozen variables, and the *freezing* transition $\ell_{\rm f}$, above which all solutions have this property (this latter threshold is much more difficult to compute as it requires a control of all solutions, including atypical ones, and has been estimated analytically only for the bicoloring of hypergraphs [27]).

As various notations have been used across different works, we recap the definition of the commonly encountered thresholds in the study of random CSPs along with our notations in Appendix A .

By definition of the different thresholds the inequalities $\ell_{\rm clust} \le \ell_{\rm cond} \le \ell_{\rm col}$ and $\ell_{\rm clust} \le \ell_{\rm r} \le \ell_{\rm col}$ must hold. A common phenomenon observed in many CSPs, such as $q$-coloring [18], bicoloring of $K$-regular graphs [26, 27], or $K$-SAT [38], is the saturation of some of these bounds when the parameters $q$ or $K$ are small (for instance $\ell_{\rm clust} = \ell_{\rm cond}$ in the 3-coloring of random graphs [18]). On the contrary when $q$ or $K$ are sufficiently large the generic order for the thresholds becomes $\ell_{\rm clust} < \ell_{\rm r} < \ell_{\rm cond} < \ell_{\rm col}$. Additionally, asymptotic expansions performed in the limit of diverging $q$ or $K$ [7, 18] revealed the existence of two main scales for these various thresholds, namely $\ell_{\rm clust} \sim \ell_{\rm r} \ll \ell_{\rm cond} \sim \ell_{\rm col}$.

Note that the asymptotic expansions of the thresholds predicted by the statistical mechanics methods have been performed up to now in models in which only one parameter could be taken to infinity ($K$ was fixed to 2 for the $q$-coloring of random graphs, and $q$ was fixed to 2 for the bicoloring of random hypergraphs). In the presently studied model one can perform asymptotic expansions in which $q$ and/or $K$ can be taken to infinity. We will come back in Sec. IV C on these expansions and on the interplay between these two possible limits.

In the following section we review how the cavity approach reveals the location of the different transitions.

#### 2. *Overview of the cavity approach and key observables*

A central objective of the statistical mechanics approach to random CSP is the computation of the

quenched free-entropy density

$$\phi(\beta) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[\ln \mathcal{Z}(\beta)] \ , \tag{5}$$

where $\mathcal{Z}$ is the partition function normalizing the probability law (2), and the average is over the random hypergraph ensemble. Indeed the free-entropy $\ln \mathcal{Z}$ is expected to concentrate around its mean (a property called self-averaging in physics), which is thus representative of the properties of the typical hypergraphs of the ensemble. The behavior of $\phi$ in the zero-temperature limit $\beta \to \infty$ depends on whether the typical graphs are colorable or not. If they are the limit of $\phi$, to be denoted $s$, is the exponential rate of growth of the number of solutions, called entropy. If they are not $\phi$ diverges with $\beta$, with a proportionality constant that gives the minimal number of unsatisfied constraints. For simplicity in the rest of this section we assume that we are in the satisfiable phase and we keep the zero-temperature limit implicit, the reader can adapt this discussion to the positive temperature case without difficulties.

The belief propagation (BP) algorithm [39] provides a tractable procedure (see Sec. III) to compute the Bethe approximation of the entropy density $s = \ln \mathcal{Z}/N$ for a given hypergraph. The replica symmetric (RS) cavity method is a way to average this prediction over the random hypergraph ensemble in the thermodynamic limit. BP yields an exact result for a tree factor graph. As we have seen above the random graph ensembles studied in this paper, with the sparsity assumption we made keeping $\alpha = M/N$ finite, are locally tree-like, their shortest loops are typically of length $O(\ln N)$. As a consequence the predictions of BP are asymptotically exact provided the longer loops do not induce long range correlations between far away variables. This is precisely the case in the unclustered phase $\ell < \ell_{\rm clust}$, but becomes wrong for larger average degrees.

The clustered structure of the configuration space causes indeed the birth of a form of correlation between variables (measured by a so called point-to-set correlation function), the connection between these two properties being reminiscent of the link between spatial and temporal mixing in the studies of Monte Carlo Markov Chain dynamics. It is thus necessary to correct the RS cavity method above the clustering threshold, by promoting it to the Replica Symmetry Breaking (RSB) cavity method. A crucial idea of this method is to exploit the partition of the solution space into disjoint clusters $\{\mathcal{C}\}$, assuming that the measure (2) restricted to the solutions $\mathbf{x}$ of one cluster $\mathcal{C}$ enjoy the decorrelation property described above. This restricted measure can thus be treated within the RS formalism, and what remains to do is to describe the statistical properties of the cluster decomposition. More precisely, this revised formalism introduces a measure over the set of clusters $\{\mathcal{C}\}$,

$$p_1(\mathcal{C}) = \frac{\mathcal{Z}(\mathcal{C})^m}{\mathcal{Z}_1(m)} = \frac{e^{mNs(\mathcal{C})}}{\mathcal{Z}_1(m)} \ , \tag{6}$$

where $\mathcal{Z}(\mathcal{C})$ and $s(\mathcal{C}) = \ln \mathcal{Z}(\mathcal{C})/N$ are respectively the number of solutions in $\mathcal{C}$ and the *internal* entropy of $\mathcal{C}$, and where $m$ is a parameter playing a role similar to the temperature called the *Parisi parameter*. The 1RSB normalizing function $\mathcal{Z}_1$ is thus given by

$$\mathcal{Z}_1(m) = \sum_{\{\mathcal{C}\}} e^{mNs(\mathcal{C})} = \int \mathrm{d}s \ \ e^{N(ms+\Sigma(s))} \tag{7}$$

where we introduced the *complexity* $\Sigma(s)$, defined as the logarithm of the average number of clusters of internal entropy $s$ (divided by $N$). The saddle point evaluation of this integral yields a thermodynamic potential that we shall call *replicated entropy*,

$$\Phi(m) = \lim_{N\to\infty} \frac{1}{N} \ln \mathcal{Z}_1(m) = m\, s(m) + \Sigma(m) \, , \tag{8}$$

where we defined

$$s(m) = \operatorname*{argmax}_s \left[m\, s + \Sigma(s)\right] , \tag{9}$$

and $\Sigma(m) = \Sigma(s(m))$.

Exploiting the decomposition of the solution set into clusters we can express the total entropy as

$$\begin{aligned} s_{\rm tot} &= \lim_{N\to\infty} \frac{1}{N} \ln \int \mathrm{d}s \ \ e^{N(s+\Sigma(s))} \\ &= \sup_{s/\Sigma(s)\geq 0} \left[s + \Sigma(s)\right] \ , \end{aligned} \tag{10}$$

where the second equality results again from a saddle-point evaluation, the condition $\Sigma(s) \geq 0$ ensuring that the corresponding clusters exist in the thermodynamic limit. The total entropy thus results from a competition between most numerous clusters (that have a large complexity) and bigger clusters (that have a large internal entropy). Within the exponential approximation the total entropy is dominated by the contribution of the clusters of internal entropy hereafter called $s^\star$ achieving the supremum of Eq. (10), we thus call a solution typical if it belongs to a cluster of internal entropy $s^\star$.

If $\Sigma(s^\star) > 0$, there is an exponential number of clusters contributing to the total entropy, we are thus in the *clustered phase* as defined in the previous section. On the contrary if $\Sigma(s^\star) = 0$, either it corresponds to connectivities smaller than $\ell_{\rm clust}$ where there is only one giant cluster, or more interestingly, to the *condensed phase*, where a sub-exponential number of clusters contains the vast majority of solutions.

The computation of the complexity at $m = 1$ gives access to $\ell_{\rm clust}$ and $\ell_{\rm cond}$: in (9) the maximization is not constrained by the condition $\Sigma(s) \geq 0$, hence the total entropy $s_{\rm tot}$ is equal to $\Phi(m = 1)$ if and only if $\Sigma(m = 1) > 0$, while $\Sigma(m = 1) < 0$ is the signature of the condensation phenomenon. In the clustered uncondensed regime the prediction $\Phi(m = 1)$ actually coincides with the RS prediction: the superposition of the exponentially large number of clusters create relatively

weak correlations, that can be only detected by a point-to-set correlation function and not correlation between a finite number of variables, see [40, 41] for recent rigorous proofs of the exactness of the RS predictions up to the condensation phase transition.

Another equivalent and compact way of summarizing the thermodynamics prescription is the formula:

$$s_{\rm tot} = \inf_{m\in[0,1]} \frac{\Phi(m)}{m} \ , \tag{11}$$

which is valid irrespectively of the type of phase: in an unclustured situation, with a single cluster of entropy $s_{\rm tot}$, equation (7) shows that $\Phi(m) = m s_{\rm tot}$, hence the parameter $m$ is irrelevant in the right hand side of (11). In presence of clusters the minimizer in (11), that we shall call $m_{\rm s}$ for the *static* value of the Parisi parameter, is either $m_{\rm s} = 1$ if an exponential number of clusters contribute to the total entropy, or some value $m_{\rm s} < 1$ if the measure condensates on a sub-exponential number of clusters. Indeed, one can see that $\frac{\rm d}{{\rm d}m}(\Phi(m)/m) = -\Sigma(m)/m^2$, hence in the condensed phase the minimization selects the largest existing clusters, with $\Sigma(m_{\rm s}) = 0$, as in (10).

Finally the colorability threshold can be obtained from the computation of $\Phi(m = 0) = \Sigma(m = 0) = \max_s \Sigma(s)$: this gives the exponential rate of growth for the total number of clusters of solutions, irrespectively of their sizes. The colorability transition occurs when this number vanishes, marking the disappearance of the last clusters of solutions.

## III. CAVITY EQUATIONS AND THRESHOLDS CHARACTERIZATIONS

In this section we derive the cavity equations of hypergraph $q$-coloring at finite temperature $\beta$. The zero temperature equations follow straightforwardly in the limit $e^{-\beta} \to 0$. These equations cover the special cases of graph $q$-coloring (when $K = 2$) [17, 18], and hypergraph bi-coloring or NAESAT (when $q = 2$) [25–27].

### A. Replica symmetric formalism

#### 1. Cavity equations and Bethe entropy

*a. Messages* Consider first the problem of computing the partition function $\mathcal{Z}(\beta)$ and characterizing the marginals of the probability law $p(\mathbf{x})$ defined in (2) when the factor graph encoding the dependencies between variables is a tree. Then the problem can be solved exactly and very easily by exploiting the recursive nature of trees, with a procedure called dynamic programming or Belief Propagation [39].

We denote by $\chi_{\nu\to i}(x_i)$ the marginal probability of the variable $x_i$ in the law encoded by the amputated factor graph in which all edges around $i$ have been cut except the one connecting it to $\nu$, and similarly $\psi_{i\to\mu}(x_i)$ for the amputated factor graph where only the edge between $i$ and $\mu$ has been cut (see the illustration in Fig. 1). If the factor graph is a tree these "messages" sent between neighboring variable and factor nodes are easily seen to obey the following exact equations:

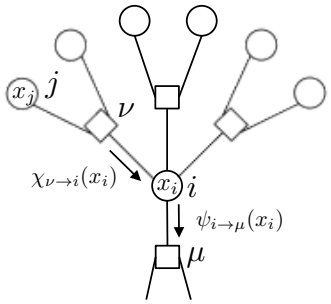



FIG. 1: Sketch of a local architecture of hypergraph illustrating notations for the derivations of the cavity equation. Vertices are represented with circles, and hyperedges with squares.

$$\chi_{\nu\to i}(x_i) = \frac{1}{\widehat{\mathcal{Z}}_0^{\nu\to i}} \sum_{\mathbf{x}_{\partial\nu\setminus i}} e^{-\beta(1-\Delta_\nu(\mathbf{x}_{\partial\nu}))} \prod_{j\in\partial\nu\setminus i} \psi_{j\to\nu}(x_j) \ , \tag{12}$$

and

$$\psi_{i\to\mu}(x_i) = \frac{1}{\widehat{\mathcal{Z}}_0^{i\to\mu}} \prod_{\nu\in\partial i\setminus\mu} \chi_{\nu\to i}(x_i) \, , \tag{13}$$

where we recall that $\partial i$ and $\partial\mu$ denote respectively the set of factor nodes around variable $i$ and the set of variable nodes around factor $\mu$, and $\widehat{\mathcal{Z}}_0^{\nu\to i}$ and $\widehat{\mathcal{Z}}_0^{i\to\mu}$ are normalization factors ensuring that both $\chi_{\nu\to i}$ and $\psi_{i\to\mu}$ sum up to one, i.e.

$$\sum_{x=1}^{q} \chi_{\nu\to i}(x) = \sum_{x=1}^{q} \psi_{i\to\mu}(x) = 1. \tag{14}$$

Equation (12) has been written for a generic CSP; specializing it to the hypergraph $q$-coloring case yields

$$\chi_{\nu\to i}(x_i) = \frac{1}{\widehat{\mathcal{Z}}_0^{\nu\to i}} \left( 1 + (e^{-\beta} - 1) \prod_{j\in\partial\nu\setminus i} \psi_{j\to\nu}(x_i) \right) \ , \tag{15}$$

as given $x_i$ the only penalized configuration is the one in which all neighbors $j$ of clause $\nu$ take this same color $x_j = x_i$.

From (13) and (15) we deduce a closed recursion on

messages from nodes to hyperedges

$$\psi_{i\to\mu}(x_i) = \frac{1}{\mathcal{Z}_0^{i\to\mu}} \prod_{\nu\in\partial i\backslash\mu} \left(1 + (e^{-\beta}-1) \prod_{j\in\partial\nu\backslash i} \psi_{j\to\nu}(x_i)\right) ,$$
$$\mathcal{Z}_0^{i\to\mu} = \sum_{x=1}^{q} \prod_{\nu\in\partial i\backslash\mu} \left(1 + (e^{-\beta}-1) \prod_{j\in\partial\nu\backslash i} \psi_{j\to\nu}(x)\right) , \tag{16}$$

and define the shorthand notation

$$\psi_{i\to\mu} \equiv \mathcal{F}^{\rm RS}(\{\psi_{j\to\nu}\}_{\nu\in\partial i\backslash\mu, j\in\partial\nu\backslash i}) \equiv \mathcal{F}^{\rm RS}(\{\psi_{j\to\nu}\}) , \tag{17}$$

for this relation.

The marginal probability of variable $x_i$ in the full factor graph can then be obtained from the messages as

$$p_i(x_i) = \frac{1}{\mathcal{Z}_0^i} \prod_{\nu\in\partial i} \chi_{\nu\to i}(x_i) , \tag{18}$$

where $\mathcal{Z}_0^i$ is the marginal partition function

$$\mathcal{Z}_0^i = \sum_{x=1}^{q} \prod_{\nu\in\partial i} \chi_{\nu\to i}(x) . \tag{19}$$

If the factor graph is a tree there exists a single solution to the consistency equations between messages, and the prediction thus obtained for the marginals of $p(\mathbf{x})$ is exact. One can however look for a fixed point of Eq. (17) on an arbitrary hypergraph even if it has loops: this is precisely the BP algorithm, that we expect to converge for typical random hypergraphs of small enough average degree.

The replica symmetric cavity method provides a way to deal with such random (hyper)graphs ensembles. Consider the thought experiment in which one generates at random a large hypergraph with a degree distribution $p_d$, finds the fixed point of BP on this particular hypergraph (assuming it exists and is unique), and choose uniformly at random a directed edge $i\to\mu$ in this hypergraph. The message $\psi_{i\to\mu}$ thus obtained is random, and we denote $\mathcal{P}^{\rm RS}(\psi)$ its probability distribution. For consistency reasons this distribution has to obey the following equation

$$\mathcal{P}^{\rm RS}(\psi) = \sum_{d=0}^{\infty} r_d \int \prod_{\nu=1}^{d} \prod_{j=1}^{K-1} {\rm d}\psi_{j\to\nu} \mathcal{P}^{\rm RS}(\psi_{j\to\nu}) \, \delta\left[\psi - \mathcal{F}^{\rm RS}(\{\psi_{j\to\nu}\})\right] , \tag{20}$$

where $\delta[]$ is the Dirac distribution, and $r_d$ the excess degree distribution associated to $p_d$ according to (4).

*b. Bethe free energy* On a tree factor graph the messages solution of (17) can be used to compute exactly the associated partition function. One finds indeed that

$$\phi^{\rm Bethe} = \frac{1}{N}\ln\mathcal{Z} = \frac{1}{N}\sum_{i=1}^{N} \ln\mathcal{Z}_0^{i+\partial i} - \frac{(K-1)}{N}\sum_{\mu=1}^{M} \ln\mathcal{Z}_0^{\mu} , \tag{21}$$

where in the general CSP formalism

$$\mathcal{Z}_0^{\mu} = \sum_{\mathbf{x}_{\partial\mu}} e^{-\beta(1-\Delta_\mu(\mathbf{x}_{\partial\mu}))} \prod_{i\in\partial\mu} \psi_{i\to\mu}(x_j) , \tag{22}$$

$$\mathcal{Z}_0^{i+\partial i} = \sum_{x_i=1}^{q} \prod_{\mu\in\partial i} \sum_{\mathbf{x}_{\partial\mu\backslash i}} e^{-\beta(1-\Delta_\mu(\mathbf{x}_{\partial\mu}))} \prod_{j\in\partial\mu\backslash i} \psi_{j\to\mu}(x_j) , \tag{23}$$

which yields for the hypergraph $q$-coloring interactions

$$\mathcal{Z}_0^{\mu} = 1 + (e^{-\beta}-1) \sum_{x=1}^{q} \prod_{i\in\partial\mu} \psi_{i\to\mu}(x) , \tag{24}$$

$$\mathcal{Z}_0^{i+\partial i} = \sum_{x=1}^{q} \prod_{\mu\in\partial i} \left(1 + (e^{-\beta}-1) \prod_{j\in\partial\mu\backslash i} \psi_{j\to\mu}(x)\right) . \tag{25}$$

If the factor graph is not a tree, but if BP converges, one can use the same expressions with the messages solution of the BP equations: this yields the Bethe approximation for the free entropy density.

The prediction of the cavity method at the RS level for the quenched free entropy density $\phi(\beta)$ defined in (5) is then obtained by averaging this Bethe approximation with respect to the message distribution $\mathcal{P}^{\rm RS}(\psi)$ and the degree distribution $p_d$, yielding

$$\phi^{\rm RS}(\beta) = \sum_{d=0}^{\infty} p_d \int \prod_{\mu=1}^{d} \prod_{j=1}^{K-1} \mathrm{d}\psi_{j\to\mu} \mathcal{P}^{\rm RS}(\psi_{j\to\mu}) \ln \mathcal{Z}_0^{i+\partial i}(\{\psi_{j\to\mu}\}) - \frac{\ell(K-1)}{K} \int \prod_{i=1}^{K} \mathrm{d}\psi_i \mathcal{P}^{\rm RS}(\psi_i) \ln \mathcal{Z}_0^{\mu}(\psi_1, \dots, \psi_K) \,, \tag{26}$$

with $\ell$ the average of $p_d$.

#### 2. Replica symmetric solution at zero temperature

The replica symmetric cavity equation (17) admits as a trivial fixed point, independently of the precise hypergraph, the uniform distribution,

$$\forall i, \mu, x_i \quad \psi_{i\to\mu}(x_i) = \chi_{\mu\to i}(x_i) = \overline{\psi}(x_i) = \frac{1}{q} \,. \tag{27}$$

It is actually the relevant one: as the interactions are antiferromagnetic the symmetry between colors cannot be broken in a pattern compatible with the existence of loops of all lengths in the hypergraph.

The RS prediction for the quenched entropy at zero temperature is thus found by inserting the solution $\mathcal{P}^{\rm RS}(\psi) = \delta[\psi - \overline{\psi}]$ and setting $e^{-\beta} = 0$ in (26), yielding

$$s^{\rm RS} = \ln q + \frac{\ell}{K} \ln \left( 1 - \frac{1}{q^{K-1}} \right) \,. \tag{28}$$

This expression is actually equal to the annealed entropy $\lim_{N\to\infty} \frac{1}{N} \ln \mathbb{E}[\mathcal{Z}(\beta = \infty)]$ counting the average number of solutions, and is hence an upper bound on the true quenched entropy. As a consequence the average degree above which the RS entropy is negative is an upper bound on the coloring threshold

$$\ell_{\rm RS} = -K \ln q / \ln \left( 1 - 1/q^{K-1} \right) > \ell_{\rm col} \,. \tag{29}$$

This corresponds to the first moment argument extensively used in computer science to obtain such upperbounds on the satisfiability transition, see for instance [33–35] for its use in the context of hypergraph $q$-coloring.

### B. One step replica symmetry broken formalism

As already discussed in Sec. II B the hypothesis underlying the RS formalism are violated for sufficiently large average degree, as the clustering phenomenon induce long-range correlations between variables that are neglected at the RS level. We shall now describe a more elaborate version of the cavity method, called 1RSB for "one step of replica symmetry breaking", that is able to tackle this clustering phenomenon, and then explain how to quantitatively determine the clustering, condensation and colorability thresholds.

#### 1. Cavity equations

The crucial hypothesis of the 1RSB cavity method is that the decomposition of the Gibbs measure (2) in the clusters partitioning the configuration space is a "pure state" decomposition in the mathematical physics sense, namely that the restriction of the Gibbs measure to one cluster $\mathcal{C}$ enjoy the decorrelation properties used at the RS level. Hence each cluster $\mathcal{C}$ is described by a set of BP messages $\{\psi^{\mathcal{C}}_{i\to\mu}\}$, solution of the BP fixed-point equations:

$$\psi^{\mathcal{C}}_{i\to\mu} = \mathcal{F}^{\rm RS}(\{\psi^{\mathcal{C}}_{j\to\nu}\}) \,. \tag{30}$$

Our goal is now to compute the 1RSB partition function (7) that sums the contribution of the different clusters to describe the thermodynamics of the space of solutions. As we assume that each cluster $\mathcal{C}$ is well described by a BP fixed point, we can use the Bethe approximation to compute the partition function restricted to $\mathcal{C}$ and write

$$\mathcal{Z}_1(m) = \sum_{\{\mathcal{C}\}} e^{mN\phi^{\rm Bethe}(\mathcal{C})} \,, \tag{31}$$

where the internal free entropy density $\phi^{\rm Bethe}(\mathcal{C})$ is $\phi^{\rm Bethe}(\{\psi^{\mathcal{C}}_{i\to\mu}\})$ defined in Eq. (21). Using this correspondence between clusters and BP fixed points, $\mathcal{Z}_1$ can be rewritten as an integral over BP messages:

$$\mathcal{Z}_1(m) = \int \prod_{(i\mu)} \mathrm{d}\psi_{i\to\mu} \left( \prod_{(i\mu)} \delta \left[ \psi_{i\to\mu} - \mathcal{F}^{\rm RS}\left(\{\psi_{j\to\nu}\}\right) \right] \right) e^{mN\phi^{\rm Bethe}(\{\psi_{i\to\mu}\})} \,, \tag{32}$$

where $(i\mu)$ spans all the edges of the factor graph, and we used the shorthand notation $\mathcal{F}^{\rm RS}$ defined in (17). The Dirac distributions in Eq. (32) make the integral equivalent to the former sum over clusters of Eq. (31), by selecting only the messages that are BP fixed points. Using the expression of $\phi^{\rm Bethe}$ given in (21) we can interpret Eq. (32) as the partition function of an auxiliary statistical physics problem [5] where now the variables are the messages $\{\psi_{i\to\mu}\}$ instead of the spin variables $\{x_i\}$, and the interactions arise from the partition functions $\mathcal{Z}_0^{i+\partial i}$ and $\mathcal{Z}_0^{\mu}$ in $\phi^{\rm Bethe}$. The factor graph encoding this auxiliary problem on the variables $\{\psi_{i\to\mu}\}$ is as locally tree-like as the original graph, one can thus treat it with the Belief Propagation approach explained above, with the messages in this auxiliary problem becoming probability distributions $P_{i\to\mu}\left(\psi_{i\to\mu}\right)$ over the variables $\{\psi_{i\to\mu}\}$. The 1RSB equations are nothing but the BP equations of this auxiliary problem, that read

$$P_{i\to\mu}\left(\psi_{i\to\mu}\right)=\frac{1}{\mathcal{Z}_1^{i\to\mu}}\int\left(\prod_{\nu\in\partial i\setminus\mu}\prod_{j\in\partial\nu\setminus i}\mathrm{d}\psi_{j\to\nu}P_{j\to\nu}\left(\psi_{j\to\nu}\right)\right)\delta\left[\psi_{i\to\mu}-\mathcal{F}^{\rm RS}\left(\{\psi_{j\to\nu}\}\right)\right]\left(\mathcal{Z}_0^{i\to\mu}(\{\psi_{j\to\nu}\})\right)^m\,, \tag{33}$$

where $\mathcal{Z}_1^{i\to\mu}$ is a normalization that depends on the $\{P_{j\to\nu}\}$ and on $m$; we shall write in short (33) as

$$P_{i\to\mu}\equiv\mathcal{F}^{\rm 1RSB}\left(\{P_{j\to\nu}\}\right). \tag{34}$$

The probability distributions $P_{i\to\mu}$ have to be interpreted as the probability to observe a BP message $\psi^{\mathcal{C}}_{i\to\mu}=\psi_{i\to\mu}$ when a cluster $\mathcal{C}$ is chosen randomly. As explained above we do not expect the symmetry between colors to be spontaneously broken, hence we impose that on average over the various clusters no color is privileged at any vertex, i.e.

$$\int\mathrm{d}\psi_{i\to\mu}P_{i\to\mu}\left(\psi_{i\to\mu}\right)\psi_{i\to\mu}=\overline{\psi} \tag{35}$$

where we had already defined $\overline{\psi}$ as the uniform distribution over $\mathcal{Q}$, with $\overline{\psi}(x)=1/q$.

The Bethe free-entropy of this auxiliary model yields the 1RSB estimate of the replicated free-entropy defined in (8), in terms of the auxiliary messages $\{P_{i\to\mu}\}$:

$$\Phi^{\rm Bethe}(m)=\frac{1}{N}\sum_{i=1}^{N}\ln\mathcal{Z}_1^{i+\partial i}(m)-\frac{K-1}{N}\sum_{\mu=1}^{M}\ln\mathcal{Z}_1^{\mu}(m)\,, \tag{36}$$

where

$$\mathcal{Z}_1^{i+\partial i}(\{P_{j\to\nu}\};m)=\int\prod_{\mu\in\partial i}\prod_{j\in\partial\mu\setminus i}\mathrm{d}\psi_{j\to\mu}P_{j\to\mu}\left(\psi_{j\to\mu}\right)\left(\mathcal{Z}_0^{i+\partial i}\right)^m \tag{37}$$

and

$$\mathcal{Z}_1^{\mu}(\{P_{i\to\mu}\};m)=\int\prod_{i\in\partial\mu}\mathrm{d}\psi_{i\to\mu}P_{i\to\mu}\left(\psi_{i\to\mu}\right)\left(\mathcal{Z}_0^{\mu}\right)^m\,. \tag{38}$$

In order to deal with random graph ensembles we introduce a probability distributions of 1RSB messages $\mathcal{P}^{\rm 1RSB}(P)$ that obey the analog of (20), namely

$$\mathcal{P}^{\rm 1RSB}(P)=\sum_{d=0}^{\infty}r_d\int\prod_{\nu=1}^{d}\prod_{j=1}^{K-1}\mathrm{d}P_{j\to\nu}\mathcal{P}^{\rm 1RSB}(P_{j\to\nu})\delta\left[P-\mathcal{F}^{\rm 1RSB}(\{P_{j\to\nu}\})\right]\,. \tag{39}$$

From the solution of this equation one can compute the 1RSB estimate of the typical replicated free-entropy for a random hypergraph with degree distribution $p_d$. The formula is the analog of the one given in (26) at the RS level, i.e.

$$\begin{aligned}\Phi^{\rm 1RSB}(m)\;&=\;\sum_{d=0}^{\infty}p_d\int\prod_{\mu=1}^{d}\prod_{j=1}^{K-1}\mathrm{d}P_{j\to\mu}\mathcal{P}^{\rm 1RSB}(P_{j\to\mu})\ln\mathcal{Z}_1^{i+\partial i}(\{P_{j\to\mu}\};m)\\&-\;\frac{\ell(K-1)}{K}\int\prod_{i=1}^{K}\mathrm{d}P_i\mathcal{P}^{\rm 1RSB}(P_i)\ln\mathcal{Z}_1^{\mu}(P_1,\ldots,P_K;m)\,.\end{aligned} \tag{40}$$

The estimate of the complexity $\Sigma(m)$ is then deduced as

$$\Sigma(m) = \Phi^{\rm 1RSB}(m) - m\frac{\mathrm{d}}{\mathrm{d}m}\Phi^{\rm 1RSB}(m)\,, \qquad (41)$$

following the discussion of Sec. II B 2. To compute this derivative it is enough to take into account only the explicit dependency on $m$ of the functions $\mathcal{Z}_1$, as the potential (40) is stationary with respect to the variations of $\mathcal{P}^{\rm 1RSB}$ whenever the 1RSB cavity equation (39) is fulfilled.

Note that the 1RSB formalism contains as a special case the RS one when the $P$'s in the support of $\mathcal{P}^{\rm 1RSB}(P)$ are Dirac distributions; in the following we shall call this a trivial solution of the 1RSB equations. Such a solution expresses indeed the absence of clustering in the configuration space, as the randomness over the choice of the clusters becomes deterministic in this case.

### 2. *Hard fields and frozen variables*

We have introduced in Sec. II B 1 the notion of frozen variables as the ones taking the same value in all the solutions of a cluster. In the cavity formalism a cluster is described in terms of BP messages, hence frozen variables are to be associated with "hard fields", i.e. messages that constrain one variable to a single color. More precisely, let us define $\psi^x$ the distribution over $\mathcal{Q}$ concentrated on the color $x$, i.e. $\psi^x(x') = \delta_{x,x'}$, and decompose the 1RSB distributions as

$$P_{i\to\mu}\left(\psi_{i\to\mu}\right) = \sum_{x=1}^{q} \eta^x_{i\to\mu}\delta[\psi_{i\to\mu} - \psi^x] + \eta^0_{i\to\mu}\widetilde{P}_{i\to\mu}(\psi_{i\to\mu})\,, \qquad (42)$$

where $\eta^x_{i\to\mu}$ is to be interpreted as the probability under $P_{i\to\mu}$ of an hard field imposing the color $x$, $\eta^0_{i\to\mu} = 1 - \sum_x \eta^x_{i\to\mu}$ is the complementary probability of "soft fields", and $\widetilde{P}_{i\to\mu}$ the distribution of the latter. When the symmetry between colors is respected we shall denote more simply $\eta^x_{i\to\mu} = \eta_{i\to\mu}$, then $\eta^0_{i\to\mu} = 1 - q\eta_{i\to\mu}$. Obviously this phenomenon of hard-fields can only occur at zero temperature; in the opposite case all configurations have positive probability, hence no marginals can be supported on a single color.

### 3. *Computation of the thresholds and of the free-entropy*

Now that we have set up the cavity formalism at the 1RSB level we can come back on the definitions of the thresholds given in Sec. II B and explain their practical computation:

- $\ell_{\rm clust}$ is the smallest average degree for which there exists a non-trivial solution of the 1RSB equations at $m = 1$. An analysis of the local stability of the trivial RS solution yields an upperbound on this threshold (sometimes called the Kesten-Stigum bound),

$$\ell_{\rm clust} \le \ell^{\rm RS}_{\rm stab} = \frac{\left(q^{K-1}-1\right)^2}{K-1} \qquad (43)$$

  for Poisson degree distributions, as we shall prove in Appendix B. This bound is not tight in general because a bifurcation can cause the non-trivial solution to appear in a discontinuous way, far from the trivial solution, a phenomenon which cannot be detected by a local analysis.

- $\ell_{\rm cond}$ is the smallest average degree for which there exists a non-trivial solution of the 1RSB equations at $m = 1$ with $\Sigma(m = 1) < 0$.

- $\ell_{\rm col}$ is the smallest average degree for which the complexity at $m = 0$ is negative.

- the prediction of the free-entropy that encompasses the RS and 1RSB formalism is

$$\inf_{m\in[0,1]} \frac{\Phi_{\rm 1RSB}(m)}{m}\,, \qquad (44)$$

  as follows from the 1RSB estimate of the potential $\Phi$ in the expression (11) (for some models this prediction was proven rigorously to be an upperbound on $\phi$ [42, 43]). At low degrees the 1RSB cavity equations only admits the trivial RS solution for which $\Phi_{\rm 1RSB}(m) = m\phi_{\rm RS}$ and one recovers the RS estimate; in the presence of clusters the minimizer is reached either in $m_{\rm s}(\ell) = 1$ if $l_{\rm clust} < \ell < \ell_{\rm cond}$, or in some non-trivial value $m_{\rm s}(\ell) < 1$ when $\ell > \ell_{\rm cond}$.

- $\ell_{\rm r}$ is the smallest average degree for which the solution of the 1RSB equation at the static value $m_{\rm s}$ of the Parisi parameter has a positive fraction of hard fields. We will also denote $\ell_{\rm r}(m)$ the threshold defined similarly but at a fixed value of $m$. We will compute in particular $\ell_{\rm r}(m = 1)$, and note that if $\ell_{\rm r}(m = 1) < \ell_{\rm cond}$ then $\ell_{\rm r} = \ell_{\rm r}(m = 1)$.

The numerical resolution of the cavity equations can be done via "population dynamics" algorithms (also known as particle representations in the context of filtering), that rely on the approximation of probability laws by the empirical measure of a large sample (population) of representative elements. For instance to solve the RS equation (20) one considers a sample of $\psi$'s, approximate $\mathcal{P}^{\rm RS}$ in the right hand side by the empirical distribution of the sample, and represent the left hand side as a new population of $\psi$'s. This procedure is repeated until convergence towards a fixed point solution of the equation, the numerical accuracy improving when the sample size increases. The same idea can be followed at the 1RSB level, with an additional difficulty: each $P$ in the sample representing

$\mathcal{P}^{\rm 1RSB}$ is itself a probability distribution, that has to be represented in an approximate way. This leads to a representation of $\mathcal{P}^{\rm 1RSB}$ by a "population of populations of $\psi$'s", whose numerical accuracy is strongly limited by the memory available on current computers. Fortunately we have seen above that the determination of the thresholds $\ell_{\rm clust}$, $\ell_{\rm cond}$ and $\ell_{\rm col}$ involves the resolution of the 1RSB equations only at $m = 1$ and $m = 0$. It turns out that these two specific values of $m$ allow great analytical simplifications, which we explain in the rest of this section before presenting our numerical results.

#### 4. *Simplifications at $m = 1$*

The special role played by the value $m = 1$, as well as the connection between the 1RSB equations in this case and the tree reconstruction problem [44], were first discussed in [10, 18, 38, 45], to which we refer the reader for more extensive discussions.

The first step in the simplification of (39) consists in considering the normalization $\mathcal{Z}_1^{i\to\mu}$ in (33). At $m = 1$, and thanks to the condition (35), one sees that $\mathcal{Z}_1^{i\to\mu} = \overline{\mathcal{Z}}_0^{i\to\mu}$ where $\overline{\mathcal{Z}}_0^{i\to\mu}$ is the value of (16) computed with $\psi_{j\to\nu} = \overline{\psi}$ for all incoming messages. As the denominator $\mathcal{Z}_1^{i\to\mu}$ is thus independent on the $\{P_{j\to\nu}\}$ one realizes that the functional $\mathcal{F}^{\rm 1RSB}$ is linear in each of its argument. This allows us to introduce the average of $\mathcal{P}^{\rm 1RSB}$,

$$\overline{P} = \int \mathrm{d}P\, \mathcal{P}^{\rm 1RSB}(P)\, P \ , \tag{45}$$

and write a closed equation on it,

$$\overline{P}(\psi) = \sum_{d=0}^{\infty} r_d \int \prod_{\nu=1}^{d}\prod_{j=1}^{K-1} \mathrm{d}\psi_{j\to\nu}\overline{P}(\psi_{j\to\nu}) \, \delta\left[\psi - \mathcal{F}^{\rm RS}(\{\psi_{j\to\nu}\})\right] \frac{\mathcal{Z}_0(\{\psi_{j\to\nu}\})}{\overline{\mathcal{Z}}_0} \ . \tag{46}$$

that easily follows from (39).

This is already a great simplification with respect to the general 1RSB formalism, as the unknown to solve for, $\overline{P}$, is simply a distribution over $\psi$'s instead of the distribution of distributions $\mathcal{P}^{\rm 1RSB}$. There is still an annoying feature in this equation, namely the reweighting factor $\mathcal{Z}_0$, that we shall now get rid of.

To do this let us define, for all $x = 1, \dots, q$, a biased version of $\overline{P}$ according to

$$\overline{P}_x(\psi) = q\, \psi(x)\, \overline{P}(\psi) \ , \tag{47}$$

that we shall call a conditional version of $\overline{P}$. Each of these $q$ measures are normalized to 1 thanks to (35), and can thus be interpreted as probability distributions. Reciprocally one can express $\overline{P}$ in terms of its conditional versions according to

$$\overline{P}(\psi) = \frac{1}{q}\sum_{x=1}^{q} \overline{P}_x(\psi) \ . \tag{48}$$

Plugging these definitions into (46) leads to the following equations on the $\overline{P}_x$'s,

$$\overline{P}_x^{(n+1)}(\psi) = \sum_{d=0}^{\infty} r_d \sum_{\mathbf{x}} p(\mathbf{x}|x) \int \prod_{\nu=1}^{d}\prod_{j=1}^{K-1} \mathrm{d}\psi_{j\to\nu}\overline{P}^{(n)}_{x_{\nu,j}}(\psi_{j\to\nu}) \, \delta\left[\psi - \mathcal{F}^{\rm RS}(\{\psi_{j\to\nu}\})\right] \ , \tag{49}$$

where we have added for future convenience some indices $n+1$ and $n$ in the left and right hand side respectively, and where $\mathbf{x} = \{x_{\nu,j}\}_{\nu=1,\dots,d}^{j=1,\dots,K-1}$, with

$$p(\mathbf{x}|x) = \prod_{\nu=1}^{d} \frac{1 + (e^{-\beta}-1)\mathbb{I}(x = x_{\nu,1} = \cdots = x_{\nu,K-1})}{q^{K-1} + e^{-\beta} - 1} \ ; \tag{50}$$

here and in the following $\mathbb{I}(E)$ the indicator function of the event $E$. This is now a very convenient form for numerical resolution, as it only involves a finite number $q$ of probability distributions, without any reweighting factor. It is also sufficient to compute the thermodynamic properties of the model. Besides $\Phi^{\rm 1RSB}(m = 1)$ which is simply equal to $\phi^{\rm RS}$ one needs the derivative of the replicated potential in order to deduce the complexity. This can be obtained by taking the derivative with respect to the explicit dependence in $m$ of (40); expressing it in terms of the $\overline{P}_x$ yields

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}m}\Phi^{\rm 1RSB}(m)\right|_{m=1} &= \sum_{d=0}^{\infty} p_d \frac{1}{q}\sum_{x,\mathbf{x}} p(\mathbf{x}|x) \int \prod_{\mu=1}^{d}\prod_{j=1}^{K-1} \mathrm{d}\psi_{j\to\mu}\overline{P}_{x_{\mu,j}}(\psi_{j\to\mu}) \mathcal{Z}_0^{i+\partial i} \ln \mathcal{Z}_0^{i+\partial i} \\ &- \frac{\ell(K-1)}{K} \sum_{x_1,\dots,x_K} p(x_1,\dots,x_K) \int \prod_{i=1}^{K} \mathrm{d}\psi_i \overline{P}_{x_i}(\psi_i) \mathcal{Z}_0^{\mu} \ln \mathcal{Z}_0^{\mu} \ , \end{aligned} \tag{51}$$

where in the second term

$$p(x_1,\dots,x_K) = \frac{1 + (e^{-\beta}-1)\mathbb{I}(x_1 = \cdots = x_K)}{q^K + q(e^{-\beta}-1)} \ . \tag{52}$$

In addition to its numerical convenience this formu-

lation has an enlightening interpretation in terms of the tree reconstruction problem [44, 45] that we shall now describe. Consider indeed a Galton Watson hypertree with offspring distribution $r_d$, and draw at random a coloring of its vertices from the free boundary Gibbs measure at inverse temperature $\beta$. By this we mean the broadcasting procedure in which one draws the color $x$ of the root uniformly at random, then independently for each of the hyperedges around the root the colors of the $K-1$ other variables are chosen with probability proportional to $e^{-\beta}$ if the hyperedge is monochromatic, 1 otherwise, and so on and so forth for the next levels of the tree. Suppose now that the color of the vertices at distance $n$ from the root are revealed to an observer, whom is asked to guess the value of $x$ from this sole information. The best strategy the observer can follow is a Bayesian one, computing the marginal probability of the root conditional on the observations. As the problem is defined on a tree this computation can be done exactly with the Belief Propagation algorithm. A moment of though should convince the reader that the distribution of this marginal probability, conditional on the root being of color $x$ during the broadcast process, is nothing but the distribution $\overline{P}_x^{(n)}$ obtained after $n$ iterations of (49) from the initial condition

$$\overline{P}_x^{(0)}(\psi) = \delta[\psi - \psi^x]\,. \tag{53}$$

Let us denote $\overline{P}_x = \lim_{n\to\infty} \overline{P}_x^{(n)}$ the fixed point solution of (49) reached for infinitely deep trees. Two situations can arise: either $\overline{P}_x(\psi) = \delta(\psi - \overline{\psi})$, or the limit is non-trivial. In the first case one says that the tree problem is not reconstructible, as no information on the color of the root can be deduced from the observation of far away vertices, in the second it is reconstructible, as an observer can estimate the color of the root with a probability larger than from a random guess. As we have defined the clustering transition in terms of the existence of a non-trivial solution of the 1RSB equations at $m = 1$, which we showed to be equivalent to (49), this non-reconstructible to reconstructible transition for the tree problem coincides with the clustering transition.

To quantify the distance between $\overline{P}_x$ and the trivial fixed-point we define the following overlap,

$$C^{(n)} = \frac{1}{1-1/q} \int \mathrm{d}\psi \overline{P}_x^{(n)}(\psi) \sum_{x'=1}^{q} (\psi(x') - 1/q)^2\ , \tag{54}$$

which is independent of $x$ thanks to the symmetry between colors, and in which the prefactor ensures the normalization $C^{(0)} = 0$. The reconstructibility of the tree problem, i.e. the condition $\ell > \ell_{\mathrm{clust}}$, is then equivalent to $\lim_{n\to\infty} C^{(n)} > 0$.

We shall now complete our study of the special case $m = 1$ of the 1RSB formalism by studying the rigidity threshold $\ell_{\mathrm{r}}(m = 1)$. We recall that we defined this transition in terms of the appearance of hard-fields in the 1RSB solution at zero temperature, see in particular (42). We have thus to compute the probability of an hard-field $\psi^x$ under the distribution $\overline{P}$, that we shall denote $\eta$, or equivalently the probability $q\eta$ of $\psi^x$ under $\overline{P}_x$ (which does not contain hard-fields of colors $x' \neq x$, see (47)). In terms of the tree reconstruction problem explained above, the question is now whether an observer can deduce the color of the root without any probability of error (rather than more accurately than a random guess) from the colors of far away vertices; this is sometimes called a naive reconstruction algorithm, as the observer only uses part of the information available by concentrating on the hard-fields.

To be more precise, let us call $q\eta^{(n)} = \overline{P}_x^{(n)}(\psi^x)$ and derive from (49) a recursion relation of the form $\eta^{(n+1)} = f(\eta^{(n)})$. From (16) we see that $\mathcal{F}^{\mathrm{RS}}(\{\psi_{i\to\mu}\}) = \psi^x$ if and only if each of the $q-1$ colors $x'$ distinct from $x$ is forbidden by at least one adjacent constraint $\mu$; the latter event means that the $K-1$ other variables in $\mu$ are forced to the color $x'$, i.e. that $\psi_{i\to\mu} = \psi^{x'}$. For concreteness let us assume that $x = 1$ and denote $d_{x'}$ the number of adjacent constraints forbidding the color $x' = 2, \ldots, q$, and $d_0 = d - (d_2 + \cdots + d_q)$ the number of adjacent constraints that do not forbid any color. As the various constraints are independent in (49) the vector $(d_0, d_2, \ldots, d_q)$ has a multinomial distribution, the probability of one constraint $\mu$ forbidding one color $x'$ being $1/(q^{K-1}-1)$, the probability of $x_{\mu,1} = \cdots = x_{\mu,K-1}$ under $p(\mathbf{x}|x)$, multiplied by $(q\eta^{(n)})^{K-1}$, the probability of picking only hard fields from the $\overline{P}_{x'}^{(n)}$ . Combining these observations leads easily to the following formula:

$$q\eta^{(n+1)} = \sum_{d=0}^{\infty} r_d \sum_{\substack{d_0,d_2,\ldots,d_q \\ d_0+d_2+\cdots+d_q=d}} \frac{d!}{d_0!d_2!\ldots d_q!} \left(1 - \frac{(q-1)(q\eta^{(n)})^{K-1}}{q^{K-1}-1}\right)^{d_0} \left(\frac{(q\eta^{(n)})^{K-1}}{q^{K-1}-1}\right)^{d_2+\cdots+d_q} \mathbb{I}(d_2 > 0)\ldots\mathbb{I}(d_q > 0) \tag{55}$$

To put this recursion under a more tractable form we use the following identity:

$$\mathbb{I}(d_2>0)\dots\mathbb{I}(d_q>0) = (1-\delta_{d_2,0})\dots(1-\delta_{d_q,0}) = \sum_{p=0}^{q-1}(-1)^p\binom{q-1}{p}\delta_{d_2,0}\dots\delta_{d_{2+p-1},0}\ , \quad (56)$$

where we made a slight abuse of notation in the second line, symmetrizing all possible choices of $p$ colors. This identity is an analytical expression for the combinatorial inclusion-exclusion principle, expressing the conjunction of $q-1$ events in terms of conjuctions of the complementary events. Inserting (56) in (55) one can perform the multinomial sums and obtain

$$\eta^{(n+1)} = \frac{1}{q}\sum_{d=0}^{\infty} r_d \sum_{p=0}^{q-1}(-1)^p\binom{q-1}{p}\left(1-\frac{p\,(q\eta^{(n)})^{K-1}}{q^{K-1}-1}\right)^d . \quad (57)$$

In the case of a Poissonian degree distribution of average $\ell$ this expression can be further simplified into

$$\eta^{(n+1)} = \frac{1}{q}\left(1-\exp\left[-\ell\frac{(q\eta^{(n)})^{K-1}}{q^{K-1}-1}\right]\right)^{q-1} . \quad (58)$$

This recursion always admits the fixed-point solution $\eta = 0$; the rigidity thrshold $\ell_{\rm r}(m=1)$ can be computed as the smallest $\ell$ such that a non-trivial fixed-point $\eta>0$ exists.

### 5. *Simplifications at $m=0$ and $\beta=\infty$*

We shall now explain the simplifications of the 1RSB cavity formalism that appears in the limit $m\to 0$, taken after the zero-temperature limit $\beta\to\infty$, and re-obtain the Survey Propagation (SP) equations first obtained for the $K$-SAT problem in [6], then for the $q$-coloring for graphs in [17] (the order of the limits in $\beta$ and $m$ is slightly different in these papers but the final result is the same).

*a. SP equations and complexity* The first point to notice is that the Parisi parameter $m$ always appear in quantities of the form $\mathcal{Z}_0^m$, that can be thus replaced in the $m\to 0$ limit by $\mathbb{I}(\mathcal{Z}_0>0)$. The probability that $\mathcal{Z}_0$ vanishes turns out to depend only on the intensity of hard-fields in the 1RSB distributions, which is the key to the simplifications in the SP formalism.

More precisely, let us first focus on the expression of $\mathcal{Z}_0^{i\to\mu}$ given in (16). This quantity vanishes if and only if each of the $q$ colors $x$ is forbidden by at least one constraint around it, which in turns means that the $K-1$ other variables are forced by an hard-field to be of color $x$. With the same kind of reasoning as the one we did to arrive at (55-57), we see that the normalization $\mathcal{Z}_1^{i\to\mu}$ of equation (33) becomes in the $m=0$ limit:

$$\mathcal{Z}_1^{i\to\mu} = \sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\nu\in\partial i\setminus\mu}\left(1-p\prod_{j\in\partial\nu\setminus i}\eta_{j\to\nu}\right) \quad (59)$$

Similarly the integral in (33) will give rise to $\delta[\psi_{i\to\mu}-\psi^x]$ if the $q-1$ colors distinct from $x$ are forbidden while $x$ is not (the ill-defined situation with the $q$ colors forbidden is cancelled out by the factor $\mathcal{Z}_0$ in the integral). Putting together these two observations allows us to project (33) onto the intensity of hard-fields alone, that obey the recursion:

$$\eta_{i\to\mu} = \frac{\displaystyle\sum_{p=0}^{q-1}(-1)^p\binom{q-1}{p}\prod_{\nu\in\partial i\setminus\mu}\left(1-(p+1)\prod_{j\in\partial\nu\setminus i}\eta_{j\to\nu}\right)}{\displaystyle\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\nu\in\partial i\setminus\mu}\left(1-p\prod_{j\in\partial\nu\setminus i}\eta_{j\to\nu}\right)} \equiv \mathcal{F}^{\rm SP}(\{\eta_{j\to\nu}\})\,, \quad (60)$$

The probability distribution $\mathcal{P}^{\rm 1RSB}(P)$ of the generic 1RSB treatment of random hypergraph ensembles can be simplified into a distribution $\mathcal{P}^{\rm SP}(\eta)$ for the intensity of hard-fields. From (39) one obtains easily the fixed-point equation it obeys, namely

$$\mathcal{P}^{\rm SP}(\eta) = \sum_{d=0}^{\infty} r_d \int\prod_{\nu=1}^{d}\prod_{j=1}^{K-1}{\rm d}\eta_{j\to\nu}\mathcal{P}^{\rm SP}(\eta_{j\to\nu}) \,\delta\left[\eta-\mathcal{F}^{\rm SP}(\{\eta_{j\to\nu}\})\right]\ . \quad (61)$$

The $m=0$ complexity, whose vanishing will yield the colorability threshold, can be computed from the solution of this equation. From (41) we see that $\Sigma(m=0)=\Phi^{\rm 1RSB}(m=0)$, because when $m=0$ all the clusters are counted in the same way, irrespectively of their sizes. Taking the limit $m\to 0$ in the expression (40) of $\Phi^{\rm 1RSB}$ gives

$$\Sigma(m=0) = \sum_{d=0}^{\infty} p_d\int\prod_{\mu=1}^{d}\prod_{j=1}^{K-1}{\rm d}\eta_{j\to\mu}\mathcal{P}^{\rm SP}(\eta_{j\to\mu})\ln\mathcal{Z}_{\rm SP}^{i+\partial i} - \frac{\ell(K-1)}{K}\int\prod_{i=1}^{K}{\rm d}\eta_i\mathcal{P}^{\rm SP}(\eta_i)\ln\mathcal{Z}_{\rm SP}^{\mu}\ , \quad (62)$$

where the expressions of $\mathcal{Z}_{\rm SP}$ are obtained from (37,38) in the limit $m\to 0$. Analyzing once more the probabilities that $\mathcal{Z}_0^{i+\partial i}$ and $\mathcal{Z}_0^{\mu}$ does not vanish because of contradicting hard-fields yields

$$\mathcal{Z}_{\rm SP}^{i+\partial i}(\{\eta_{j\to\mu}\}) = \sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\mu=1}^{d}\left(1-p\prod_{j=1}^{K-1}\eta_{j\to\mu}\right)\ , \quad (63)$$

which is one minus the probability that the $q$ colors are forbidden by neighboring constraints, and

$$\mathcal{Z}_{\rm SP}^{\mu}(\eta_1,\dots,\eta_K) = 1-q\prod_{i=1}^{K}\eta_i\ , \quad (64)$$

where one recognizes in $\prod\eta_i$ the probability that an hyperedge is forced to be monochromatic of one given color.

*b. Stability analysis* As its name suggests the 1RSB version of the cavity method is only the first level of a hierarchy of refinements. In the $p$-RSB version there are $p$ steps of replica symmetry breaking, which means that the pure states (clusters) introduced at the 1RSB level are themselves organized in a hierarchical structure with $p$ levels. The equations at the $p$-th level admit as special cases the solutions of the equations at all the $p' < p$ levels (we saw above that the RS formalism was recovered as a trivial solution of the 1RSB equations). In principle the correct (in the sense of the cavity method) description of a mean-field model should be sought for in all the levels of RSB; in simple situations this hierarchy is expected to collapse on the RS and 1RSB level only, i.e. in such cases the $p$-RSB equations do not admit solutions that do not reduce to the RS or 1RSB ones. Solving the cavity equations with an arbitray level of RSB in a sparse mean-field model is an extremely challenging task, hence this program cannot be implemented in practice. One can however test the local stability of the 1RSB solution into the larger set of 2RSB solutions, which allows to assess the existence, or not, of nearby proper 2RSB solutions which presumably strictly improve the 1RSB bound (44). This local stability analysis of the 1RSB solutions for sparse models was introduced in [36, 46], see also [7, 17, 47] for other presentations of the method. The 2RSB description of the set of solutions of a CSP introduces clusters of configurations, which are themselves organized into groups of clusters. As a consequence there are two ways in which this can be reduced to the 1RSB description: either there is a single group, or each cluster contains a single configuration. These two reductions yield two different type of instabilities of the 1RSB solution, termed respectively type I or noise propagation instability, and type II or bug proliferation instability. To quantify these phenomena one introduces two positive numbers $\lambda_{\rm I}$ and $\lambda_{\rm II}$, in such a way that one type of perturbation of the 1RSB solution is stable (resp. unstable) if the corresponding $\lambda$ is strictly smaller (resp. larger) than 1. One finds generically that $\lambda_{\rm I}$ grows with the density of constraints $\ell$, crossing the critical value 1 at a threshold denoted $\ell_{\rm I}^{\rm SP}$, the 1RSB solution being thus stable for $\ell < \ell_{\rm I}^{\rm SP}$. Conversely $\lambda_{\rm II}$ is usually found to be decreasing with $\ell$, the stability regime being thus of the form $\ell > \ell_{\rm II}^{\rm SP}$. We defer the details of the computations of the $\lambda$ parameters for the hypergraph coloring problem to Appendix C, the numerical results being presented in the next Section.

## IV. RESULTS

We present in this Section the results obtained on the coloring of random hypergraphs through the application of the cavity method presented above. We shall first give the numerical values of the various thresholds of the Erdős-Rényi ensemble in Sec. IV A, then study briefly a regular case (cf. Sec. IV B), and finally present analytical asymptotic expansions of the thresholds for large values of $K$ and/or $q$ in Sec. IV C.

### A. The thresholds of the Erdős-Rényi ensembles

Our main results are summarized in Table I in which we give, for a few small values of $K \geq 3$ (the graph case $K = 2$ has already been studied in [17, 18]) and $q$, the average degree $\ell$ at which the various phase transitions occur in the Erdős-Rényi ensemble, i.e. when the degree and excess degree distributions $p_d$ and $r_d$ are Poissonian laws of average $\ell$.

Some of these thresholds have been obtained analytically or by solving numerically a simple scalar equation: the upperbound $\ell_{\rm RS}$ on the colorability threshold and the Kesten-Stigum bound $\ell_{\rm stab}^{\rm RS}$ can be read directly from (29) and (43), and the rigidity threshold $\ell_{\rm r}(m = 1)$ is the smallest $\ell$ such that (58) admits a positive fixed-point solution.

The determination of the other thresholds have required an heavier numerical work. As explained in Sec. III B 3 they are defined in terms of the solution of the 1RSB equations for the values $m = 1$ and $m = 0$ of the Parisi parameter. For these two special cases we have derived above some fixed-point equations where the unknowns are probability distributions of scalars (or $q$-dimensional vectors), see in particular (49) and (61). These equations can thus be handled numerically by population dynamics algorithms where the unknown probability distributions are approximated by the empirical distribution of large samples of random representants, that are iteratively updated until numerical convergence (see for instances the appendices of [48] for more details on algorithmic implementation issues).

To be more specific we shall first consider the last line of Table I, i.e. the case $q = 4$, $K = 3$, and explain the steps that led us to the numbers displayed in the Table. On the left panel of Figure 2 we plot the overlap $C^{(n)}$ (also known as the point-to-set function) defined in Eq. (54), as a function of the number of iterations $n$, for several values of $\ell$. By definition the clustering threshold is the smallest value of $\ell$ such that this function does not decay to zero at large $n$, as it marks the appearance of a non-trivial solution of the 1RSB equations at $m = 1$. One can clearly see on this plot that the bifurcation is here discontinuous: the large $n$ limit of $C^{(n)}$ jumps from zero to a strictly positive value when $\ell$ crosses $\ell_{\rm clust}$. Accordingly we have here $\ell_{\rm clust} < \ell_{\rm stab}^{\rm RS}$: this transition is unrelated to the Kesten-Stigum local instability of the trivial solution. A precursor of the transition when $\ell$ approaches $\ell_{\rm clust}$ from below is the divergence of the length of the "plateau" in $C^{(n)}$, that we quantify by $n_*(C)$, the number of iterations necessary to make the correlation drop below a specified level $C$. As illustrated on the right panel of Fig. 2 we have thus estimated $\ell_{\rm clust}$ by fitting the divergence of $n_*(C)$ as $(\ell - \ell_{\rm clust})^{-1/2}$ close to the transition (see [49] for a justification of this critical

| | $\ell_{\rm clust}$ | $\ell_r(m=1)$ | $\ell_{\rm cond}$ | $\ell_{\rm col}$ | $\ell_{\rm RS}$ | $\ell_{\rm stab}^{\rm RS}$ | $\ell_{\rm I}^{\rm SP}$ | $\ell_{\rm II}^{\rm SP}$ |
|---|---|---|---|---|---|---|---|---|
| $q=2$ $K=3$ | 4.50 | (7.37) | 4.50 | (6.32) | 7.23 | **4.50** | **6.07** | 6.16 |
| $q=2$ $K=4$ | 16.33 | (21.62) | 16.33 | (19.62) | 20.76 | **16.33** | **19.35** | 17.84 |
| $q=2$ $K=5$ | 47.4 | (52.63) | 51.5 | 52.32 | 53.70 | 56.25 | 59.42 | 43.9 |
| $q=3$ $K=3$ | 25.06 | (28.07) | 26.2 | 26.92 | 27.98 | 32.00 | 33.62 | 23.9 |
| $q=3$ $K=4$ | 97.7 | 105.88 | 114.3 | 115.04 | 116.44 | 225.33 | 225.51 | 90.7 |
| $q=4$ $K=3$ | 56.20 | 61.09 | 62.7 | 63.3 | 64.44 | 112.50 | 112.78 | 52.7 |

TABLE I: The thresholds of $q$-coloring of $K$-uniform Erdős-Rényi random hypergraphs, given in terms of their average degree $\ell$. The definitions of the thresholds are given in the text and summarized in Table II. Bold font numbers indicate significant instabilities. Numbers given by unstable or invalid ansatz are put between parenthesis: (i) The prediction of the colorability threshold for the two first rows is hindered by the SP type I instability. (ii) The $m=1$ rigidity threshold does not coincide with the true rigidity threshold for the four first rows because $\ell_{\rm cond} < \ell_{\rm r}(m=1)$: the rigidity should be evaluated at the static Parisi parameter $m_{\rm s} < 1$ in those cases.

exponent). We turn now our attention to Figure 3. On its left part we continue the presentation of the results at $m=1$ initiated in Figure 2; the upper left panel displays indeed the large $n$ limit of the overlap $C^{(n)}$ for $\ell > \ell_{\rm clust}$. The bottom left panel shows the value of the associated complexity $\Sigma(m=1)$: it is strictly positive at the birth of the non-trivial solution, and decreases with increasing $\ell$. The condensation threshold $\ell_{\rm cond}$ is deduced from this data as the point where the complexity vanishes. The right part of Figure 3 is devoted to the results of the Survey Propagation equations (i.e. the 1RSB equations at $m=0$). On the upper right panel we plotted the complexity $\Sigma(m=0)$, whose vanishing marks the colorability transition $\ell_{\rm col}$, while the lower right panel demonstrates that this prediction falls into a regime where the SP solution is locally stable, according to the analysis presented in Appendix C.

Let us underline the main qualitative features that arise from the analysis of the data we just presented in the case $q=4$, $K=3$: (i) the clustering transition occurs discontinuously, hence $\ell_{\rm clust} < \ell_{\rm stab}^{\rm RS}$ (ii) at this point the $m=1$ complexity is strictly positive, yielding $\ell_{\rm clust} < \ell_{\rm cond}$ (iii) the SP formalism is stable at $\ell = \ell_{\rm col}$, we have thus no reason to discard this prediction as a conjecture for the exact colorability threshold. We found the same qualitative features in the analysis of the cases $(q,K) = (2,5), (3,3), (3,4)$, see the Table for the numerical values which are of course quantitatively different, and we expect this to be the generic scenario for all larger values of $q$ and $K$ (i.e. for $q=2$ and $K \geq 5$, and for $q \geq 3$ and $K \geq 3$).

There are however two cases which do not correspond to this generic scenario, namely the bicoloring ($q=2$) of $K=3$ and $K=4$-uniform hypergraphs, see the first two lines of Table I. The first aspect in which they differ is that the SP formalism is unstable (with respect to the noise propagation instability) for $\ell = \ell_{\rm col}$: the 1RSB computation does not give the correct prediction for the colorability threshold in such a situation, however one expects on the basis of the bounds of [42, 43, 50] the 1RSB result to be an upperbound on the location of the colorability transition (in fact one expects this bound to be strict, assuming that the instability implies the existence of a proper 2RSB solution that strictly improves the 1RSB bound of (44)).

They also have a different pattern of transitions in the solution of their 1RSB equations at $m=1$. To demonstrate this more clearly we present in Figure 4 the overlap and complexity at $m=1$ for the bicoloring ($q=2$) of $K=3,4$ and 5-uniform ER hypergraphs. The case $K=3$ on the left exhibits a continuous transition, with $\ell_{\rm clust} = \ell_{\rm cond} = \ell_{\rm stab}^{\rm RS}$. The non-trivial solution of the 1RSB equation bifurcates continuously from the trivial one at the Kesten-Stigum threshold, its complexity is always negative hence there is no clustered un-condensed phase; this scenario was also found in the 3-coloring of ER graphs ($K=2$, $q=3$) in [18].

The middle panel presents the rather peculiar results of the $K=4$ case. One sees the overlap growing continuously for $\ell > \ell_{\rm stab}^{\rm RS} = \ell_{\rm clust}$, as in the $K=3$ case, but it undergoes an upward jump for a larger value of the average degree. More precisely there is a domain of coexistence of two distinct stable solutions of the $m=1$ 1RSB equations, for $\ell$ roughly between 18.5 and 19. We call these two solutions the high and low overlap (HO/LO) branches, with two spinodals at which one of these branches disappear, both spinodals occuring for some $\ell > \ell_{\rm clust}$. In this domain of coexistence one obtains the HO branch by initializing the resolution of (49) with the initial condition (53), corresponing to the usual tree reconstruction problem, while the LO solution is reached from an initialization very close to the unstable trivial fixed-point $\delta[\psi - \overline{\psi}]$ (that we perturb by a small fraction of hard-fields), that implements the so-called robust version of the reconstruction problem [51]. One can see on the middle bottom panel of Fig. 4 that the $m=1$ complexity of the LO branch is always negative, while a small part of the HO branch, for $\ell$ slightly above its spinodal, is positive. This phenomenon is puzzling at first sight, and its interpretation requires a moment of thought. One

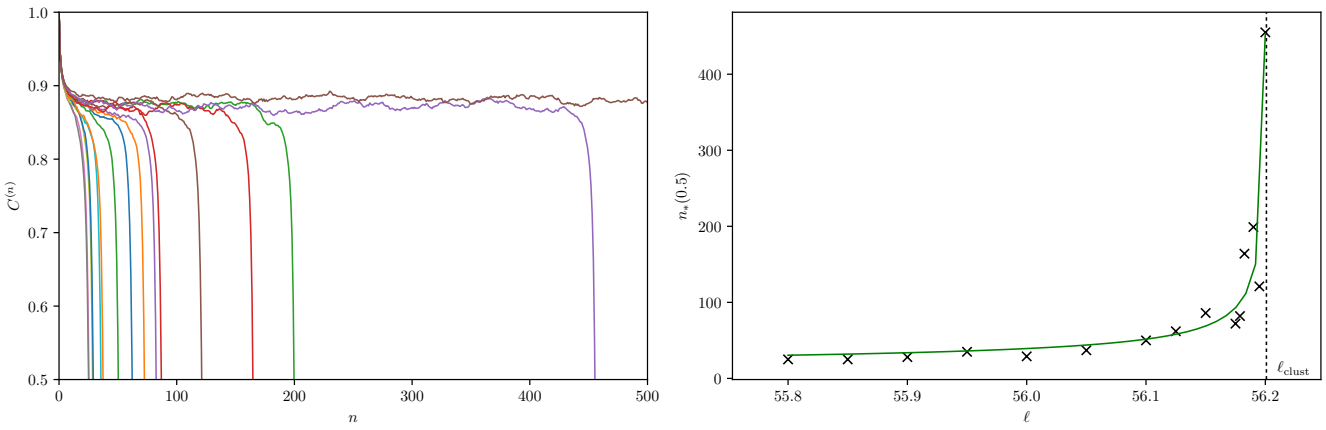



FIG. 2: Left: the overlap (or point-to-set correlation function) $C^{(n)}$ of equation (54) as a function of the number $n$ of iterations, for the $q = 4$-coloring of $K = 3$-uniform ER hypergraphs. The different curves corresponds to the following values of the average degree $\ell$, from left to right : 55.8, 55.85, 55.9, 55.95, 56.0, 56.05, 56.1, 56.125, 56.15, 56.175, 56.1788, 56.1825, 56.19, 56.195, 56.2, 56.3. This last value corresponds to $\ell > \ell_{\rm clust}$, hence the overlap does not decay to 0. Right: the divergence of the number $n_*(0.5)$ of iterations after which the overlap drops below 0.5 (black crosses), along with a fit of the form $A \times (\ell - \ell_{\rm clust})^{-1/2}$ (green line) over the adjustable parameters $A$ and $\ell_{\rm clust}$, which yields the estimate $\ell_{\rm clust} \approx 56.2$ reported in Table I.

could think that in this regime the HO branch is the relevant one, hence that a small clustered un-condensed regime lies between two condensed ones. It turns out that this interpretation is wrong: in the variational prescription for the computation of the free-entropy recalled in (44) one should also minimize over the different solutions of the cavity equations in case several do coexist (the rigorous bounds of [42, 43, 50] actually hold for arbitrary order parameters, not necessarily solutions of the cavity equations). It is useful at this point to take a look at the sketch of the function $\Phi(m)/m$ presented in Figure 5, recalling that the slope of this function is given in terms of the complexity by $\frac{\rm d}{{\rm d}m}(\Phi(m)/m) = -\Sigma(m)/m^2$. This should convince the reader that the existence of one solution of the 1RSB equations with $\Sigma(m = 1) < 0$ implies that $\phi < \phi_{\rm RS}$ and that the model is in a condensed phase, even if other solutions have a positive complexity. The minimum of $\Phi(m)/m$ occurs indeed at a non-trivial static Parisi parameter $m_{\rm s} < 1$, and most probably on the branch continuously connected to the solution with the most negative complexity at $m = 1$ (if one excludes additional crossings of the branches as functions of $m$). The conclusion of this analysis is that $\ell_{\rm clust} = \ell_{\rm cond} = \ell^{\rm RS}_{\rm stab}$ in the case $q = 2$, $K = 4$, and that the coexistence of solutions of the 1RSB equations at $m = 1$ is not thermodynamically relevant here: the minimum in (44) is reached at a non-trivial value $m_{\rm s} < 1$, hence neither of the solutions at $m = 1$ yields the correct 1RSB prediction for the entropy. Note that this peculiar bifurcation scenario remained, as far as we know, unobserved previously, and that it caused a slight mistake in [27].

Finally the right panel of Figure 4 shows that there are also coexisting 1RSB solutions for $K = 5$, but at variance with the $K = 4$ case the spinodal of the HO branch occurs before the Kesten-Stigum transition, as well as the vanishing of the associated complexity. Hence in this case $\ell_{\rm clust} < \ell_{\rm cond} < \ell^{\rm RS}_{\rm stab}$: from a thermodynamics point of view we are in the generic case explained above for $q = 4$, $K = 3$, and the existence of the LO branch bifurcating continuously from the trivial fixed point at the Kesten-Stigum transition is completely irrelevant.

Let us briefly mention that the coexistence of solutions has much more important consequences in so-called planted ensembles of inference problems, which will be discussed in [52]. In this perspective determining the conditions for the existence of a continuously bifurcating solution above the Kesten-Stigum transition is an important question; it is shown in [52] that a large class of $K$-wise interacting Ising spin models (including the bicoloring of hypergraphs) always exhibit this continuous transition, as confirmed by the plots of Fig. 4. Moreover the analysis of dense inference problems that mimick the large degree limit of the $q$-coloring of $K$-uniform hypergraphs [53] show that the criterion for the existence (resp. absence) of the continuous solution is $qK - 2K - q < 0$ (resp. $qK - 2K - q > 0$). This is agreement with the previous statement for $q = 2$, and implies that when $K \geq 3$ and $q \geq 3$ there should not be a stable solution bifurcating continuously above Kesten-Stigum (except possibly in the marginal case $q = K = 3$ which cannot be decided from the above criterion). Our numerical simulations seem to confirm this statement (and to indicate the absence of continuous solution when $q = K = 3$ for Poisson hypergraphs), however we cannot formally exclude from them the existence of a very short branch with a spinodal closely above the Kesten-Stigum transition.

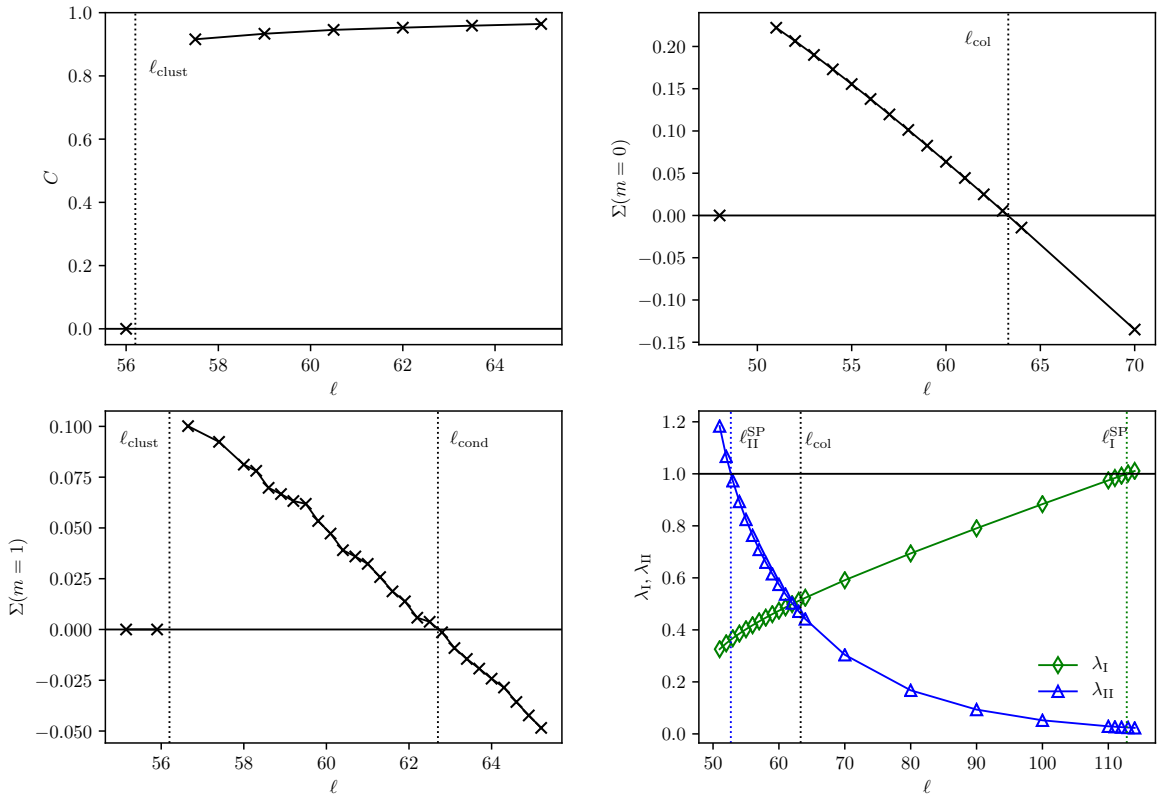



FIG. 3: Numerical results for the $q = 4$-coloring of $K = 3$-uniform ER random hypergraphs. Left: Overlap (top) and complexity (bottom) for the solution of the 1RSB equation at $m = 1$ as a function of the average degree $\ell$. Right: Complexity (top) and stability criteria (bottom) for the solution at $m = 0$ (Survey Propagation) as a function of the average degree $\ell$. The stability of the SP solution is assessed by the quantities $\lambda_{\rm I}$ and $\lambda_{\rm II}$ defined in Appendix C, unstable regimes corresponding to $\lambda > 1$.

### B. A finite $m$ study of the bicoloring of regular hypergraphs ($q = 2$, $K = 4$)

In order to confirm our interpretation of the coexisting solutions we found in the Erdős-Rényi ensemble with $q = 2$, $K = 4$ and average degree between 18.5 and 19 (roughly) we should solve the 1RSB equations for arbitrary values of $m$ in $[0, 1]$ to find explicitly the behavior sketched in Fig. 5. Unfortunately this is a rather difficult numerical task: apart from the two special cases $m = 0$ and $m = 1$ the solution of the 1RSB equations for arbitrary $m$ and non-trivial degree distributions have to be represented by populations of populations, a precise computation of $\Phi(m)$ is thus extremely challenging. To circumvent this difficulty we turned instead to regular ensembles, with deterministic degree distributions $p_d = \delta_{d,\ell}$ and $r_d = \delta_{d,\ell-1}$. In such a case the cavity equations presented in Sec. III B admit a "translationally invariant" (sometimes called factorized) solution with

$$\mathcal{P}^{\rm 1RSB}(P) = \delta[P - P_{\rm reg}] \ , \tag{65}$$

traducing the equivalence of all local neighborhoods in a regular random graph. The 1RSB equation with arbitrary $m$ can thus be handled with a single population of messages $\psi$ representing the distribution $P_{\rm reg}$ fixed-point solution of (34). Solving this equation is slightly more difficult than, for instance, (61), because of the reweighting factor $\mathcal{Z}_0^m$. This point can be tackled by representing $P_{\rm reg}$ as a weighted sample of representants, or by using resampling techniques, we refer the reader to appendix E3 of [48] for more details on the implementation issues.

We performed this study for the bicoloring ($q = 2$) of $K = 4$-uniform $\ell = 19$-regular random hypergraphs. It is indeed for this connectivity that we found the coexistence of two solutions of the 1RSB equations at $m = 1$ (and zero temperature), one with a high value of the overlap and a positive complexity, one with a low value of the overlap and a negative complexity, exactly as in the

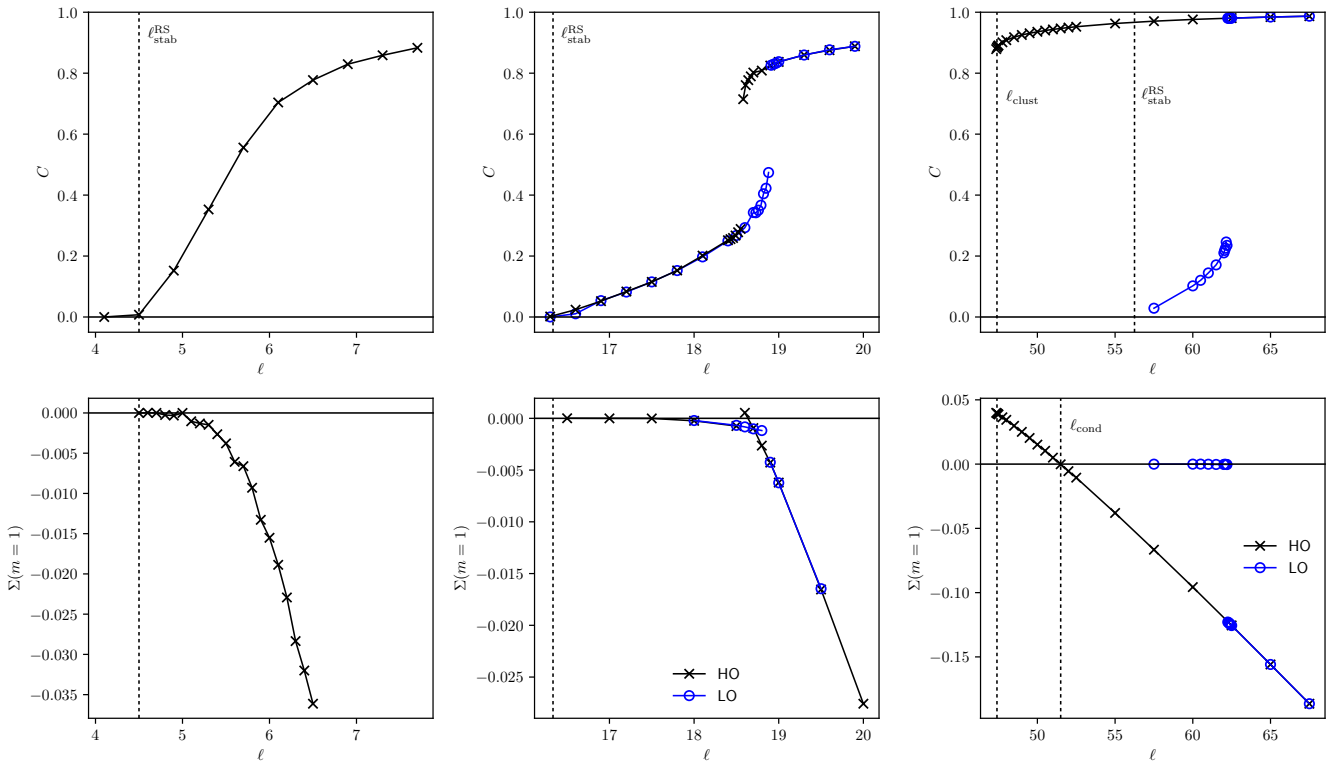



FIG. 4: Overlaps (top) and complexities (bottom) for the solutions of the 1RSB equations at $m = 1$ as a function of the average connectivity $\ell$ of the ER random hypergraph ensemble, for $q = 2$, $K = 3$ (left), $K = 4$ (middle) and $K = 5$ (right). In the middle and right panels the two branches are selected by different initial conditions, see the text for details.

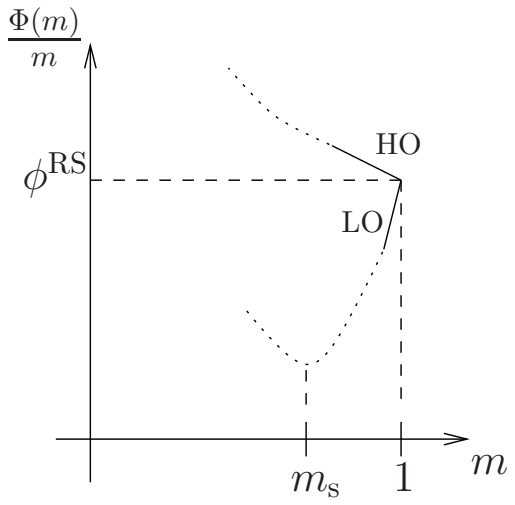



FIG. 5: A sketch of the behavior of $\Phi(m)/m$ when two solutions of the 1RSB equations coexist at $m = 1$, the one denoted LO (resp. HO) has a negative (resp. positive) complexity, which fixes the sign of the slope in $m = 1$. The dotted part of the curves are educated guesses, to be confirmed in Sec. IV B

.

coexistence region of the ER ensemble. The results presented in Fig. 6 show the evolution of these two solutions as $m$ is varied (we used here a small positive temperature $T = 0.1$, but the $T = 0$ case is qualitatively equivalent). We see that the coexistence phenomenon persists in an interval of $m$, roughly for $m \in [0.35, 1]$, and that the HO branch disappears for smaller values of $m$. More importantly one sees that the complexity of the HO branch remains positive for $m \leq 1$ (and even increases when $m$ decreases), while the complexity of the LO branch changes sign at the static value of the Parisi parameter $m_{\rm s} \approx 0.26$. This confirms the sketch of Fig. 5, and in particular the fact that the solution that minimizes $\Phi(m)/m$ is the continuation of the LO branch (or more generically of the one with the most negative complexity at $m = 1$).

We also studied the same regular ensemble as a function of temperature, see the results in Fig. 7. Upon lowering the temperature a non-trivial solution of the 1RSB equation at $m = 1$ appears continuously at the Kesten-Stigum threshold (B5) with a negative complexity; this solution persists down to zero temperature and corresponds to the LO branch. The other branch appears discontinuously, with a positive complexity, around $T = 0.1$, and is always thermodynamically irrelevant according to our analysis. Lowering the temperature in the regular

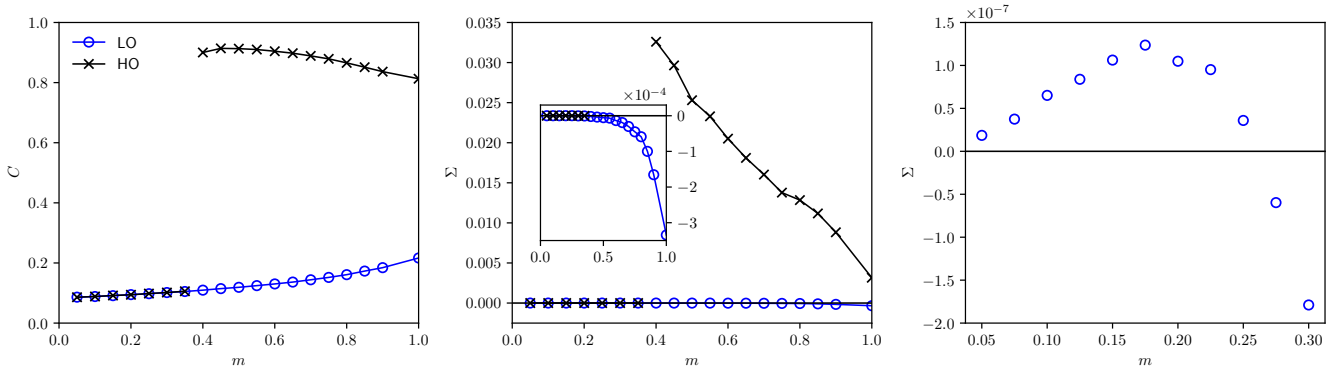



FIG. 6: Overlap (left) and complexity (middle, inset and right) of the solutions of the 1RSB cavity equations for the bicoloring of 19-regular 4-uniform random hypergraphs at temperature $T = 0.1$, as a function of the Parisi parameter $m$. Inset of middle panel and right panel are zoomed views of middle pannel. Black lines with crosses corresponds to the high overlap branch (HO) and blues lines with blue circles to the low overlap (LO) branch, reached from the two different initial conditions explained in Sec. IV A.

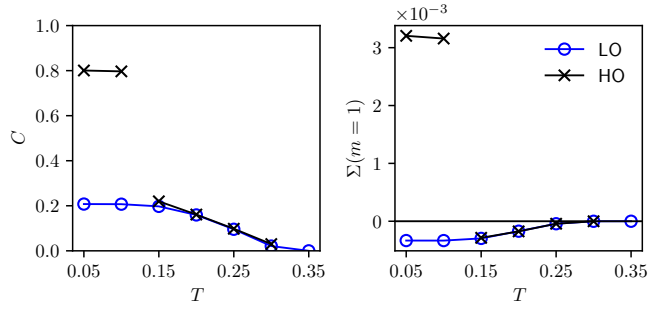



FIG. 7: Overlap (left) and complexity (right) of the solutions of the 1RSB cavity equations at $m = 1$, for the bicoloring of 19-regular 4-uniform random hypergraphs as a function of the temperature. Same symbols as in Fig. 6.

ensemble is thus qualitatively similar to increasing the average degree at zero temperature in the ER ensemble.

### C. Asymptotic expansions

We shall discuss now the asymptotic scalings of the various thresholds described in this paper when $K$ and/or $q$ gets large, the details of the computations being deferred to Appendix D.

The easiest one to begin with is the upperbound on the colorability threshold provided by the RS cavity method, or equivalently with the first moment method. As it is given explicitly (see (29)) it can be expanded as

$$\ell_{\rm RS} = Kq^{K-1}\ln q - \frac{1}{2}K\ln q + O\left(\frac{K\ln q}{q^{K-1}}\right) . \qquad (66)$$

Here the large parameter in which one expands is $q^{K-1}$, hence the expansion is valid when $K$ and/or $q$ gets large, as in all of these cases $q^{K-1} \to \infty$.

We also considered the rigidity tresholds $\ell_{\rm r}(m)$, defined as the smallest degree above which the clusters described by the 1RSB equations with Parisi parameter $m$ contain an extensive number of frozen variables, and found them to behave as

$$\ell_{\rm r}(m) = q^{K-1}[\ln(K-1) + \ln(q-1) + \ln\ln(Kq) + C(m) + o(1)] , \qquad (67)$$

an expansion valid when $q$ and/or $K$ diverges (see Appendix D for details). The constant term in the parenthesis depends on the Parisi parameter $m$, we computed its value in the two special cases $m = 1$ and $m = 0$, finding $C(m = 1) = 1$ and $C(m = 0) = 1 - \ln 2$ (in agreement with previous results for $K = 2$ [17, 18] and $q = 2$ [27]). We also expect the clustering transition $\ell_{\rm clust}$ to have the same asymptotic expansion as in (67), with a different constant term $C$; this has been rigorously proven for coloring of graphs ($K = 2$) in [19, 23].

The colorability and condensation thresholds have been found in previous studies [7, 17, 18, 27, 38] to occur asymptotically very close to the upperbound of the first moment method. This is also the case here, we have indeed obtained:

$$\ell_{\rm col} = Kq^{K-1}\ln q - \frac{1}{2}K\left(\ln q + 1 - \frac{1}{q}\right) + \widetilde{O}\left(\frac{1}{q^{K-1}}\right) , \qquad (68)$$

$$\ell_{\rm cond} = Kq^{K-1}\ln q - \frac{1}{2}K\left(\ln q + \left(1 - \frac{1}{q}\right)2\ln 2\right) + \widetilde{O}\left(\frac{1}{q^{K-1}}\right) . \qquad (69)$$

To be more precise, we have performed these expansions by taking the limit $K \to \infty$ for a fixed value of $q$, the

notation $\widetilde{O}$ hiding terms which are polynomial in $K$, the expansion being naturally organized in inverse powers of the exponential scale $q^{K-1}$. However, setting $K = 2$ in (68) and (69) we reproduce the results presented for the large $q$ limit of the $K = 2$ graph case in [17, 18], provided one removes the $1/q$ terms which are indeed comparable to the neglected terms of order $1/q^{K-1} = 1/q$. We can thus conjecture these expansions to be valid when $K \geq 3$ and/or $q$ gets large, with $\widetilde{O}$ hiding terms at most polynomial in $K$ and $\ln q$. These results are compatible with the asymptotic lowerbound on $\ell_{\rm col}$ presented in [35] where the large $q$ limit is studied for arbitrary $K \geq 3$.

## V. DISCUSSION

We summarize some of the main original features of the hypergraph coloring problem and discuss their potential consequences:

- Survey Propagation, and therefore the 1RSB solution at $m = 0$, is unstable around the colorability threshold for $q = 2$, $K = 3$ and $K = 4$. While the instability of the 1RSB solution was observed in many models (the Sherrington-Kirkpatrick (SK) model and the satisfiability problem in the unsatisfiable phase at low temperature [54, 55] being two notable examples), it is important to note that among previously studied random constraint satisfaction problems we are aware of only one case, the circular 5-coloring of 3-regular graphs [56], where 1RSB has been found unstable at equilibrium in the satisfiable phase. The present case of $K = 3$ or 4 hypergraph bicoloring is another example. In the other random CSPs where the stability of SP has been tested, for instance random 3- and 4-SAT [36], instabilities were found in a part of the satisfiable phase that did not include the SP prediction of the satisfiability threshold.

  Establishing this example is important for the following reason: while at least on the physics level the 1RSB picture is well understood in sparse constraint satisfaction problems, there is no known solvable form of full-step-replica-symmetry-breaking (FRSB) equations for sparse systems (see [57] for a recent attempt in this direction). The nature of the structure of solutions in a CSP where 1RSB is not stable is a widely open problem (even on the physics non-rigorous level). One can conjecture this structure to be similar to the FRSB phase that has been well described and understood in dense models, such as the SK model [4, 58, 59], yet this remains to be put on a more solid ground; having concrete examples on which to study this phenomenon is thus an important starting point. An important difference between FRSB in the satisfiable regime of CSPs and in the SK model is the fact that in the former case the groundstate energy is known to be exactly zero, while in the latter it is a non-trivial quantity. This fact might simplify the analysis, and is reminiscent of the situation in the mean-field treatment of the hard-sphere model [60] where a so-called Gardner transition towards FRSB occurs in the un-jammed (satisfiable) phase.

- We found a coexistence of two 1RSB solutions in the bicoloring ($q = 2$) of $K \geq 4$-uniform hypergraphs. We concluded that this coexistence has no significant bearings for the thermodynamics of random ensembles as studied in this paper. However, we note that this phenomenon is much more interesting in the perspective of planted inference problems, as discussed in detail in [52].

- We have obtained the asymptotic expansions of the rigidity, condensation and coloring thresholds when the parameters $q$ and/or $K$ diverge. These expansions connect smoothly (at least at the order of the expansion we have reached) the one previously obtained for $q = 2$ and $K = 2$.

Concerning other open directions for future work we want to note that the version of hypergraph coloring where each constraint requires no more than $\gamma$ variables to have the same color for $\gamma < K - 1$ [33] is also interesting and might also bring examples of interesting new phenomena.

Establishing rigorously the results presented here is another natural direction for future work. In particular any result for the case of bicoloring with $K = 3$ and $K = 4$ beyond the clustering threshold (equal to the Kesten-Stigum bound in those cases) would be remarkable because it might shed some light about the mysterious FRSB in sparse constraint satisfaction problems.

We also note that the freezing transition where all solutions acquire frozen variables has been so far determined analytically only for the bicoloring of random hypergraph problem [27]. Generalizing the results of [27] to $q > 2$ should in principle be possible, as a first step towards the computation of the freezing transition in arbitrary random CSPs.

### Acknowledgments

We warmly thank the Simons Institute for the Theory of Computing at UC Berkeley, where this work has been initiated as a working group within the program Counting Complexity and Phase Transitions, and in particular Florent Krzakala, Catherine Greenhill and Laura Florescu. M.G. acknowledges funding from "Chaire de recherche sur les modèles et sciences des donnés", Fondation CFM pour la Recherche-ENS and from ERC grant SPARCS 307087.

[1] Stephen A Cook. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, pages 151–158. ACM, 1971.

[2] Peter Cheeseman, Bob Kanefsky, and William M Taylor. Where the really hard problems are. In *IJCAI*, volume 91, pages 331–337, 1991.

[3] David Mitchell, Bart Selman, and Hector Levesque. Hard and easy distributions of sat problems. In *AAAI*, volume 92, pages 459–465, 1992.

[4] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin-Glass Theory and Beyond*, volume 9 of *Lecture Notes in Physics*. World Scientific, Singapore, 1987.

[5] M. Mézard and A. Montanari. *Physics, Information, Computation*. Oxford Press, Oxford, 2009.

[6] M. Mézard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812–815, 2002.

[7] Stephan Mertens, Marc Mézard, and Riccardo Zecchina. Threshold values of random k-sat from the cavity method. *Random Struct. Algorithms*, 28(3):340–373, 2006.

[8] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky. 2+p-sat: Relation of typical-case complexity to the nature of the phase transition. *Random Structures and Algorithms*, 15:414, 1999.

[9] G. Biroli, R. Monasson, and M. Weigt. A variational description of the ground state structure in random satisfiability problems. *Eur. Phys. J. B*, 14:551, 2000.

[10] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborova. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, jun 2007.

[11] Marc Mézard, Thierry Mora, and Riccardo Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94(19):197205, 2005.

[12] Dimitris Achlioptas and Cristopher Moore. Two moments suffice to cross a sharp threshold. *SIAM Journal on Computing*, 36:740–762, 2006.

[13] Dimitris Achlioptas and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *Proc. of 38th STOC*, pages 130–139, New York, NY, USA, 2006. ACM.

[14] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proceedings of FOCS 2008*, page 793, 2008.

[15] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. *arXiv preprint arXiv:1411.0650*, 2014.

[16] Roberto Mulet, Andrea Pagnani, Martin Weigt, and Riccardo Zecchina. Coloring random graphs. *Physical review letters*, 89(26):268701, 2002.

[17] Florent Krzakala, Andrea Pagnani, and Martin Weigt. Threshold values, stability analysis, and high-q asymptotics for the coloring problem on random graphs. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2004.

[18] Lenka Zdeborová and Florent Krzakala. Phase transitions in the coloring of random graphs. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2007.

[19] Allan Sly. Reconstruction of random colourings. *Communications in Mathematical Physics*, 288(3):943–961, 2009.

[20] M. Molloy. The freezing threshold for k-colourings of a random graph. In *Proceedings of the 44th symposium on Theory of Computing*, page 921. ACM, 2012.

[21] Amin Coja-Oghlan and Dan Vilenchik. Chasing the k-colorability threshold. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 380–389. IEEE, 2013.

[22] Victor Bapst, Amin Coja-Oghlan, Samuel Hetterich, Felicia Raßmann, and Dan Vilenchik. The condensation phase transition in random graph coloring. *Communications in Mathematical Physics*, 341(2):543–606, 2016.

[23] A. Sly and Y. Zhang. Reconstruction of colourings without freezing. *ArXiv e-prints arXiv:1610.02770*, 2016.

[24] Varsha Dani, Cristopher Moore, and Anna Olson. Tight bounds on the threshold for permuted k-colorability. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 505–516. Springer, 2012.

[25] Tommaso Castellani, Vincenzo Napolano, Federico Ricci-Tersenghi, and Riccardo Zecchina. Bicoloring random hypergraphs. *J. Phys. A*, 36:11037, 2003.

[26] L. Dall'asta, A. Ramezanpour, and R. Zecchina. Entropy landscape and non-Gibbs solutions in constraint satisfaction problems. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2008.

[27] Alfredo Braunstein, Luca Dall'Asta, Guilhem Semerjian, and Lenka Zdeborová. The large deviations of the whitening process in random constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(5):053401, 2016.

[28] Amin Coja-Oghlan and Lenka Zdeborova. The condensation transition in random hypergraph 2-coloring. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 241–250, 2012.

[29] Amin Coja-Oglan and Konstantinos Panagiotou. Catching the k-naesat threshold. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 899–908. ACM, 2012.

[30] Jian Ding, Allan Sly, and Nike Sun. Satisfiability threshold for random regular nae-sat. *Communications in Mathematical Physics*, 341(2):435–489, 2016.

[31] Michael Molloy and Ricardo Restrepo. Frozen variables in random boolean constraint satisfaction problems. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1306–1318, 2013.

[32] Victor Bapst, Amin Coja-Oghlan, and Felicia Rassmann. A positive temperature phase transition in random hypergraph 2-coloring. *arXiv preprint arXiv:1410.2190*, 2014.

[33] Michael Krivelevich and Benny Sudakov. The chromatic numbers of random hypergraphs. *Random Structures and Algorithms*, 12(4):381–403, 1998.

[34] Martin Dyer, Alan Frieze, and Catherine Greenhill. On the chromatic number of a random hypergraph. *Journal of Combinatorial Theory, Series B*, 113:68 – 122, 2015.

[35] Peter Ayre, Amin Coja-Oghlan, and Catherine Greenhill. Hypergraph coloring up to condensation. *arXiv preprint arXiv:1508.01841*, 2015.

[36] A. Montanari, G. Parisi, and F. Ricci-Tersenghi. Instability of one-step replica-symmetry-broken phase in satisfiability problems. *J. Phys. A*, 37:2073, 2004.

[37] Ludovic Berthier and Giulio Biroli. Theoretical perspective on the glass transition and amorphous materials. *Rev. Mod. Phys.*, 83:587–645, 2011.

[38] Andrea Montanari, Federico Ricci-Tersenghi, and Guilhem Semerjian. Clusters of solutions and replica symmetry breaking in random $k$-satisfiability. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(04):P04004, 2008.

[39] F. R. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory*, 47(2):498–519, 2001.

[40] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborova. Information-theoretic thresholds from the cavity method. *arXiv:1611.00814*, 2016.

[41] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang, and T. Kapetanopoulos. Charting the replica symmetric phase. *arXiv:1704.01043*, 2017.

[42] S. Franz and M. Leone. Replica bounds for optimization problems and diluted spin systems. *J. Stat. Phys.*, 3-4:535–564, 2003.

[43] Dmitry Panchenko and Michel Talagrand. Bounds for diluted mean-fields spin glass models. *Probability Theory and Related Fields*, 130(3):319–336, 2004.

[44] E. Mossel. Survey: Information flow on trees. In J. Nestril and P. Winkler, editors, *Graphs, Morphisms and Statistical Physics*, DIMACS series in discrete mathematics and theoretical computer science, pages 155–170, 2004.

[45] Marc Mézard and Andrea Montanari. Reconstruction on trees and spin glass transition. *J. Stat. Phys.*, 124:1317–1350, september 2006.

[46] A. Montanari and F. Ricci-Tersenghi. On the nature of the low-temperature phase in discontinuous mean-field spin glasses. *European Physical Journal B*, 33(3):339–346, 2003.

[47] O. Rivoire, G. Biroli, O. C. Martin, and M. Mézard. Glass models on bethe lattices. *Eur. Phys. J. B*, 37:55–78, 2004.

[48] Lenka Zdeborova. Constraint satisfaction problems, Random graphs coloring problem, Average computational complexity, Cavity method, Spin glasses,. *Acta Physica Slovaca*, 59(3):169–303, 2009.

[49] A. Montanari and G. Semerjian. On the dynamics of the glass transition on bethe lattices. *J. Stat. Phys.*, 124:103–189, 2006.

[50] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics*, 233(1):1–12, 2003.

[51] Svante Janson and Elchanan Mossel. Robust reconstruction on trees is determined by the second eigenvalue. *Ann. Probab.*, 32:2630–2649, 2004.

[52] Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborova. in preparation. 2017.

[53] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. *ArXiv e-prints arXiv:1701.08010*, 2017.

[54] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[55] Andrea Crisanti, Luca Leuzzi, and Giorgio Parisi. The 3-sat problem with large number of clauses in the $\infty$-replica symmetry breaking scheme. *Journal of Physics A: Mathematical and General*, 35(3):481, 2002.

[56] Christian Schmidt, Nils-Eric Guenther, and Lenka Zdeborová. Circular coloring of random graphs: statistical physics investigation. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(8):083303, 2016.

[57] Giorgio Parisi. The marginally stable bethe lattice spin glass revisited. *Journal of Statistical Physics*, 167(3):515–542, May 2017.

[58] Michel Talagrand. *Spin glasses: a challenge for mathematicians: cavity and mean field models*, volume 46. Springer Science & Business Media, 2003.

[59] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer-Verlag, 2013.

[60] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. Glass and jamming transitions: From exact results to finite-dimensional descriptions. *Annual Review of Condensed Matter Physics*, 8:265–288, 2017.

## Appendix A: Threshold definitions

We summarize the definitions and notations for the various transitions undergone by random ensembles of CSPs in Table II.

## Appendix B: Stability of the RS solution

We shall prove here the Kesten-Stigum bound (43) on the clustering transition. To do so we shall study the stability of the trivial RS solution of the 1RSB equations, and show that for $\ell > \ell^{\rm RS}_{\rm stab}$ it is locally unstable. The iterations of the 1RSB equations must then flow to another fixed-point, which completes the reasoning as $\ell_{\rm clust}$ is defined in terms of the existence of a non-trivial solution of the 1RSB equations.

To put this reasoning on a quantitative basis let us consider a distribution $P(\psi)$ in the support of $\mathcal{P}^{\rm 1RSB}(P)$, that is close to the RS distribution $\delta[\psi - \overline{\psi}]$. To quantify the distance between $P$ and this Dirac delta it is natural to consider its first moments,

$$M(x) = \int \mathrm{d}\psi P(\psi) \left(\psi(x) - \frac{1}{q}\right) , \tag{B1}$$

$$V(x,x') = \int \mathrm{d}\psi P(\psi) \left(\psi(x) - \frac{1}{q}\right)\left(\psi(x') - \frac{1}{q}\right) . \tag{B2}$$

We note however that the condition (35) that enforces the symmetry between colors in the 1RSB formalism implies that $M = 0$. Hence at the lowest order the closeness of $P$ from the Dirac measure in $\overline{\psi}$ can be tracked by the covariance matrix $V(x,x')$. The latter has itself to respect the invariance under color permutations (hence can only depend on whether $x = x'$ or not), the symmetry $V(x,x') = V(x',x)$, and the normalization condition $\sum_x V(x,x') = 0$, as the $\psi$ in the support of $P$ are probability distributions over $\mathcal{Q}$. Consequently $V$ must be of the form $V(x,x') = v(q\delta_{x,x'} - 1)$, with $v$ a scalar quantity quantifying the width of $P$. What remains now to assess is the growth or shrinking of these $v$'s under the iterations of the 1RSB equation (39).

To do so suppose that the distributions $\{P_{i\to\mu}\}$ are characterized by variances $\{v_{i\to\mu}\}$, and let us compute the variance $v$ of $P = \mathcal{F}^{\rm 1RSB}(\{P_{j\to\nu}\})$. As a first step we consider the RS iteration $\psi = \mathcal{F}^{\rm RS}(\{\psi_{j\to\nu}\})$, with arguments $\psi_{j\to\nu}(x) = \frac{1}{q} + \epsilon_{j\to\nu}(x)$ close to the uniform distribution $\overline{\psi}$. Expanding (16) linearly in the $\epsilon$'s yields

$$\psi(x) = \frac{1}{q} + \frac{e^{-\beta}-1}{q^{K-1} + (e^{-\beta}-1)} \sum_{\nu=1}^{d} \sum_{j=1}^{K-1} \epsilon_{j\to\nu}(x) , \tag{B3}$$

where we exploited the normalization condition $\sum_x \epsilon_{j\to\nu}(x) = 0$. Injecting this expansion in (33), and noticing that the normalization $\mathcal{Z}_1$ and the reweighting factor $\mathcal{Z}_0^m$ compensates at lowest order in the variance expansion, we obtain a projection of $P = \mathcal{F}^{\rm 1RSB}(\{P_{j\to\nu}\})$ in terms of the variances as

$$v = \left(\frac{e^{-\beta}-1}{q^{K-1} + (e^{-\beta}-1)}\right)^2 \sum_{\nu=1}^{d} \sum_{j=1}^{K-1} v_{j\to\nu} . \tag{B4}$$

This relation being linear in the $v$'s one can take its average with respect to $\mathcal{P}^{\rm 1RSB}$ and deduce that the instability of the trivial RS solution under the iterations of (39) will occur if and only if

$$\left(\sum_{d=1}^{\infty} r_d d\right)(K-1)\left(\frac{e^{-\beta}-1}{q^{K-1} + (e^{-\beta}-1)}\right)^2 > 1 . \tag{B5}$$

For a Poisson distribution of mean $\ell$, and at zero temperature, this gives the criterion stated in (43).

## Appendix C: Survey propagation instabilities

In this appendix we give some details on the computation of the instability parameters $\lambda_{\rm I}$ and $\lambda_{\rm II}$ introduced in Section III B 5. As explained there these parameters allow to test the existence of proper 2RSB solutions of the cavity equations that are close to the SP solution. As the 2RSB formalism describes the configuration space with two hierarchical levels in the organization of the pure states there are two ways to embed the SP description as a degenerate 2RSB solution: either there is only one group of pure states, or each pure state contains only one configuration. We shall hence study the perturbations of the SP equations in the neighborhood of these two reductions and assess their stability.

### 1. Type I instability

The stability analysis of SP under noise propagation (type I instability) is somewhat similar to the linear stability of the RS solution detailed in Appendix B. We shall indeed replace, on a given link $i \to \mu$ of the factor graph, the SP message $\eta_{i\to\mu}$ by a narrow distribution $P_{i\to\mu}$ of such messages, and check whether this distribution shrinks or expands upon iteration. More quantitatively we shall follow the average of $P_{i\to\mu}$ and its covariance, to be denoted $V_{i\to\mu}$. We assume that $P_{i\to\mu}$ is symmetric under the permutation of colors, hence its average is a vector with $q$ equal elements, while its covariance is a $q\times q$ matrix with only two distinct elements (on the diagonal and outside).

To obtain the evolution equation of the covariance matrices we shall first write a generalization of the Survey Propagation equation (60), without assuming the symmetry under permutation of the $q$ colors. Projecting (33) on the intensity of the hard fields one finds

$$\eta^x_{i\to\mu} = \frac{\sum_{S\subset\mathcal{Q}\setminus x} (-1)^{|S|} \prod_{\nu\in\partial i\setminus\mu} \left(1 - \xi^x_{\nu\to i} - \sum_{x'\in S} \xi^{x'}_{\nu\to i}\right)}{\sum_{S\subset\mathcal{Q},|S|\geq 1} (-1)^{|S|+1} \prod_{\nu\in\partial i\setminus\mu} \left(1 - \sum_{x'\in S} \xi^{x'}_{\nu\to i}\right)} \tag{C1}$$

where

$$\xi^x_{\nu\to i} = \prod_{j\in\partial\nu\setminus i} \eta^x_{j\to\nu} , \tag{C2}$$

and we recall that $\mathcal{Q} = \{1,\dots,q\}$ denotes the set of available colors.

A small perturbation of one incoming message $\eta^y_{j\to\nu}$ in (C1) will induce a small perturbation of the outcoming message

| Transition / Instability | Notation | Definition / Cavity method evidence strategy / Other therminology |
|---|---|---|
| Clustering | $\ell_{\rm clust}$ | The set of typical solutions splits into exponentially many clusters, at the threshold the MCMC equilibration time diverges. The 1RSB cavity equations at $m = 1$ acquire a non-trivial solution.<br>Other therminology: dynamical transition (in the context of mean field structural glasses), solvability of the tree reconstruction problem. |
| RS - linear stability | $\ell^{\rm RS}_{\rm stab}$ | The trivial (RS) solution of the 1RSB equations becomes linearly unstable under small perturbations. This provides an upperbound on $\ell_{\rm clust}$.<br>Other therminology: Kesten-Stigum transition for tree reconstruction, stability of the uninformative fixed point (for planted problems). |
| Condensation | $\ell_{\rm cond}$ | A finite number of clusters come to dominate the set of solutions. The total entropy is not equal to the RS entropy anymore and becomes strictly smaller. The complexity (computed at $m = 1$) becomes negative.<br>Other therminology: Kauzmann or ideal glass transition (mean field structural glasses), information-theoretic limit (for inference of planted models). |
| Rigidity | $\ell_{\rm r}$ | Typical solutions start to have a fraction (bounded away from zero) of frozen variables. The 1RSB population dynamics (at the static value of the Parisi parameter $m = m_{\rm s}$) features some hard fields. The notation $\ell_{\rm r}(m)$ indicates the transition for a prescribed value of $m$. It is much easier to compute it for $m = 1$ or $m = 0$ than for arbitrary values of $m$. If $\ell_{\rm r}(m = 1) < \ell_{\rm cond}$ then $\ell_{\rm r} = \ell_{\rm r}(m = 1)$; this happens in the generic scenario for $q, K$ sufficiently large. $\ell_{\rm r}(m = 1)$ is an upperbound for $\ell_{\rm clust}$.<br>Other therminology: sometimes unhappily refered to as freezing; $\ell_{\rm r}(m = 1)$ is the threshold for naive reconstruction. |
| SP | $\ell^{\rm SP}$ | The SP equation starts to have a non trivial solution. Corresponds to $\ell_{\rm r}(m = 0)$. |
| Freezing | $\ell_{\rm f}$ | All solutions start to have a fraction bounded away from zero of frozen variables. |
| Colorability | $\ell_{\rm col}$ | There exists no more valid colorings. The frozen variables start to break constraints. The complexity at $m = 0$ becomes negative as all clusters disappear.<br>Other therminology: satisfiability (generic CSPs) |
| RS | $\ell_{\rm RS}$ | The RS entropy reaches zero. This threshold is an upper bound of the true colorability threshold.<br>Other therminology: In problems with uniform BP fixed point this is the first moment bound |
| SP - type I stability | $\ell^{\rm SP}_{\rm I}$ | Noise propagation: the SP non trivial fixed point is linearly unstable under small perturbations breaking the color symmetry. |
| SP - type II stability | $\ell^{\rm SP}_{\rm II}$ | Bug proliferation: changing the color of one frozen variable implies changes at long range |

TABLE II: Definitions, notation and principal characteristics of the main transitions undergone by the set of solutions of random CSPs.

$\eta^x_{i\to\mu}$. To quantify this effect at linear order we define the $q \times q$ matrix $T^{j\to\nu}_{i\to\mu}$ by

$$\left[T^{j\to\nu}_{i\to\mu}\right]_{x,y} = \left.\frac{\partial\eta^x_{i\to\mu}}{\partial\eta^y_{j\to\nu}}\right|_{\rm sym}, \tag{C3}$$

where the subscript "sym" means that we evaluate the derivative at the color-symmetric point with all the incoming and outcoming $\eta$'s independent of their color index. After a short computation one finds that the matrix elements of $T$ take only two distinct values, depending on whether $x = y$ or not. More precisely, the diagonal and off-diagonal elements of $T$ are, respectively:

$$\left[T_{i\to\mu}^{j\to\nu}\right]_{x,x} = \left(\prod_{k\in\partial\nu\setminus i,j}\eta_{k\to\nu}\right)(1-\eta_{i\to\mu})\,\frac{\sum_{p=0}^{q-1}(-1)^{p+1}\binom{q-1}{p}\prod_{\rho\in\partial i\setminus\mu,\nu}\left(1-(p+1)\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}{\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\rho\in\partial i\setminus\mu}\left(1-p\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}\,, \tag{C4}$$

$$\left[T_{i\to\mu}^{j\to\nu}\right]_{x,y} = \left(\prod_{k\in\partial\nu\setminus i,j}\eta_{k\to\nu}\right)\left(\frac{\sum_{p=0}^{q-2}(-1)^{p}\binom{q-2}{p}\prod_{\rho\in\partial i\setminus\mu,\nu}\left(1-(p+2)\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}{\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\rho\in\partial i\setminus\mu}\left(1-p\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}\right. \tag{C5}$$

$$\left. -\,\eta_{i\to\mu}\frac{\sum_{p=0}^{q-1}(-1)^{p+1}\binom{q-1}{p}\prod_{\rho\in\partial i\setminus\mu,\nu}\left(1-(p+1)\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}{\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\rho\in\partial i\setminus\mu}\left(1-p\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}\right)\,, \tag{C6}$$

for all $x \neq y$. Such a matrix has two distinct eigenvalues, $T_{x,x}+(q-1)T_{x,y}$ with constant eigenvector, and $T_{x,x}-T_{x,y}$, $q-1$-times degenerate, with eigenspace perpendicular to the constant vector. In the following we shall only need the latter eigenvalue, that will be denoted $\theta_{i\to\mu}^{j\to\nu}$. One finds after a short computation that the above expressions of the matrix elements imply

$$\theta_{i\to\mu}^{j\to\nu} = -\left(\prod_{k\in\partial\nu\setminus i,j}\eta_{k\to\nu}\right)\frac{\sum_{p=0}^{q-2}(-1)^{p}\binom{q-2}{p}\prod_{\rho\in\partial i\setminus\mu,\nu}\left(1-(p+1)\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}{\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\rho\in\partial i\setminus\mu}\left(1-p\prod_{j\in\partial\nu\setminus i}\eta_{j\to\rho}\right)}\,. \tag{C7}$$

We can indeed concentrate on the fluctuations of $\eta$ that are perpendicular to the constant vector: the longitudinal one does not bring out of the SP solution studied in (61), an instability in this direction would be associated to a bifurcation of (61). Hence $V$ can be parametrized by a single scalar $v$, as in Appendix B. Exploiting the computation made above on the derivative of $\mathcal{F}^{\rm SP}$ one sees that the joint distribution of the average and variance of the $P_{i\to\mu}$ evolve under iteration as

$$\mathcal{P}^{(n+1)}(\eta,v) = \sum_{d=0}^{\infty} r_d \int \prod_{\nu=1}^{d}\prod_{j=1}^{K-1}\mathrm{d}\eta_{j\to\nu}\mathrm{d}v_{j\to\nu}\mathcal{P}^{(n)}(\eta_{j\to\nu},v_{j\to\nu})\;\delta\left[\eta-\mathcal{F}^{\rm SP}(\{\eta_{j\to\nu}\})\right]\delta\left[v-\sum_{\nu=1}^{d}\sum_{j=1}^{K-1}(\theta^{j\to\nu})^2 v_{j\to\nu}\right]\,. \tag{C8}$$

The SP solution is type I stable if under these iterations the $v$'s shrink to zero, unstable if they grow indefinitely. In practice one solves this distributional equation via a population dynamics algorithm, representing $\mathcal{P}^{(n)}(\eta,v)$ by a sample of pairs $\{(\eta_i^{(n)},v_i^{(n)})\}$. After each iteration one divides the $v_i^{(n+1)}$ by a common constant $\lambda_{\rm I}^{(n)}$ in order to keep $\sum_i v_i^{(n)}$ independent of $n$ (note that the evolution of the $v$'s in (C8) is linear, which allows such a rescaling). To increase the accuracy we define $\lambda_{\rm I}$ as the geometric mean of $\lambda_{\rm I}^{(n)}$ for many consecutive iterations, the solution is thus stable if and only if $\lambda_{\rm I}<1$.

### 2. Type II instability

Let us now turn to the analysis of the second type of instability, and first explain its principle for a generic CSP. The zero-temperature BP equations for generic CSPs can be phrased in terms of "warnings" sent by variables to neighboring constraints, and vice versa. We denote $\chi$ the set of possible warnings (that depends on the specific form of the CSP), with the special value 0 encoding the absence of warning, and consider the situation where one variable node has $n$ variable nodes at distance 1 in the cavity graph. The projection of the BP equations on the warnings gives rise to two functions: $\mathcal{V}(x_1,\dots,x_n)$ which is the indicator function of the event "the $n$ warnings are not contradictory", i.e. they allow at least one configuration for the variable considered, and $h(x_1,\dots,x_n)$, which is the value of the warning sent by this variable node (which is well-defined if and only if $\mathcal{V}(x_1,\dots,x_n)=1$). The SP equation relates probability distributions on $\chi$: suppose that the warning $x_i$ is emitted with probability $\eta_i^{x_i}$, independently for each of the neighbors. Then the probability law $\eta^x$ of the output warning, conditional on the absence of contradiction between in-coming warnings, is:

$$\eta^x = \frac{1}{\mathcal{Z}_{\rm SP}} \sum_{x_1,\dots,x_n} \eta_1^{x_1}\dots\eta_n^{x_n}\delta_{x,h(x_1,\dots,x_n)}\mathcal{V}(x_1,\dots,x_n)\ ,$$
$$\mathcal{Z}_{\rm SP} = \sum_{x_1,\dots,x_n} \eta_1^{x_1}\dots\eta_n^{x_n}\mathcal{V}(x_1,\dots,x_n)\ . \tag{C9}$$

As a first step towards the type II instability analysis we consider a random process in which a pair $(x_i,x_i')$ is emitted by each neighbor, with a joint law denoted $\eta_i^{x_i,x_i'}$; the joint law of the pair of outcoming messages, again conditioning on the absence of contradictions in both copies of the process, reads

$$\eta^{x,x'} = \frac{\sum_{\substack{x_1,\dots,x_n\\x_1',\dots,x_n'}} \eta_1^{x_1,x_1'}\dots\eta_n^{x_n,x_n'}\delta_{x,h(x_1,\dots,x_n)}\mathcal{V}(x_1,\dots,x_n)\delta_{x',h(x_1',\dots,x_n')}\mathcal{V}(x_1',\dots,x_n')}{\sum_{\substack{x_1,\dots,x_n\\x_1',\dots,x_n'}} \eta_1^{x_1,x_1'}\dots\eta_n^{x_n,x_n'}\mathcal{V}(x_1,\dots,x_n)\mathcal{V}(x_1',\dots,x_n')}\ . \tag{C10}$$

It is obvious that if the two copies of the incoming messages are strictly identical, i.e. if $\eta_i^{x_i,x_i}=\eta_i^{x_i}\delta_{x_i,x_i'}$ for all $i$, then this is also the case for the outcoming pair of warnings, with a probability $\eta^x$ given by (C9). The type II instability corresponds to a deviation from this strict coupling: we assume that $\eta_i^{x_i,x_i'}$ is very small for $x_i\neq x_i'$, and denote this quantity $\varepsilon_i^{x_i\to x_i'}$, while for $x_i=x_i'$ we keep the notation $\eta_i^{x_i}$. Expanding at first order in the $\varepsilon_i$ the equation (C10) one finds that the output joint law is also close to diagonal, with the small probabilities $\varepsilon^{x\to x'}$ given for $x\neq x'$ by

$$\varepsilon^{x\to x'} = \frac{1}{\mathcal{Z}_{\rm SP}}\sum_{i=1}^{n}\sum_{\substack{x_1,\dots,x_n\\x_i'\neq x_i}}\left(\prod_{j\neq i}\eta_j^{x_j}\right)\varepsilon_i^{x_i\to x_i'}\delta_{x,h(x_1,\dots,x_n)}\mathcal{V}(x_1,\dots,x_n)\delta_{x',h(x_1,\dots,x_i',\dots,x_n)}\mathcal{V}(x_1,\dots,x_i',\dots,x_n)\ . \tag{C11}$$

The name "bug proliferation" of this instability comes from the following algorithmic interpretation: the first copy of the process $x=h(x_1,\dots,x_n)$ is considered to be the "correct" one, while the primed warnings are corrupted by bugs which occurs with a small probability; the SP solution is unstable if the probabilities of the bugs grows during the propagation of the SP equations along the factor graph.

We shall now come back to the hypergraph coloring problem, and recalls that here the warning alphabet is $\{1,\dots,q,0\}$, the first $q$ values corresponding to a variable being forced to a given colour, while 0 is sent if at least two colours are allowed for this variable. The warnings are propagated first from the variables to the hyperedges, a color being forbidden by an hyperedge if its $K-1$ neighbors are forced to the same color, then from hyperedges to variables, a variable being forced to a given value if the $q-1$ other colors are forbidden (a contradiction arises if the $q$ colors are forbidden). A change in a single warning can be propagated in the first step if the $K-2$ other warnings are forcing to the same color, and in the second step if $q-2$ colors are forbidden by the hyperedges not affected by the bug. One thus finds the following rules for the propagation of the bug probabilities,

$$\varepsilon_{i\to\mu}^{x\to 0} = \frac{1}{\mathcal{Z}_{\rm SP}^{i\to\mu}} \sum_{\substack{\nu\in\partial i\setminus\mu\\ j\in\partial\nu\setminus i}} \sum_{y\neq x} \left(\sum_{p=0}^{q-2}(-1)^p \binom{q-2}{p} \prod_{\rho\in\partial i\setminus\mu,\nu} \left(1-(p+2)\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)\right) \left(\varepsilon_{j\to\nu}^{y\to 0} + \sum_{y'\neq y}\varepsilon_{j\to\nu}^{y\to y'}\right) \prod_{k\in\partial\nu\setminus i,j}\eta_{k\to\nu} \ ,$$

$$\varepsilon_{i\to\mu}^{0\to x} = \frac{1}{\mathcal{Z}_{\rm SP}^{i\to\mu}} \sum_{\substack{\nu\in\partial i\setminus\mu\\ j\in\partial\nu\setminus i}} \sum_{y\neq x} \left(\sum_{p=0}^{q-2}(-1)^p \binom{q-2}{p} \prod_{\rho\in\partial i\setminus\mu,\nu} \left(1-(p+2)\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)\right) \left(\varepsilon_{j\to\nu}^{0\to y} + \sum_{y'\neq y}\varepsilon_{j\to\nu}^{y'\to y}\right) \prod_{k\in\partial\nu\setminus i,j}\eta_{k\to\nu} \ ,$$

$$\varepsilon_{i\to\mu}^{x\to x'} = 0 \ ,$$

where $x, x', y$ and $y'$ denote non-trivial warnings (different from 0). Note that in the last line we assumed $K \geq 3$ (but the final result given below turns out to be valid also for $K = 2$), and that in the first two lines the factor containing the summation over $p$ is the probability that the constraints not affected by the bug forbids the $q-2$ colors distinct from $x$ and $y$. After one iteration one has $\varepsilon^{x\to x'} = 0$, and the $\varepsilon^{x\to 0}, \varepsilon^{0\to x}$ become independent of colors. One can thus simplify these equations and keep a single $\varepsilon$ on each edge, that evolves according to

$$\varepsilon_{i\to\mu} = (q-1) \sum_{\substack{\nu\in\partial i\setminus\mu\\ j\in\partial\nu\setminus i}} \mathcal{A}_{i\to\mu}^{j\to\nu}\varepsilon_{j\to\nu} \tag{C12}$$

where the $(q-1)$ comes from the summation over $y \neq x$ in the above equations, and where we defined

$$\mathcal{A}_{i\to\mu}^{j\to\nu} = \left(\prod_{k\in\partial\nu\setminus i,j}\eta_{k\to\nu}\right) \frac{\displaystyle\sum_{p=0}^{q-2}(-1)^p\binom{q-2}{p}\prod_{\rho\in\partial i\setminus\mu,\nu}\left(1-(p+2)\prod_{j\in\partial\rho\setminus i}\eta_{j\to\rho}\right)}{\displaystyle\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}\prod_{\rho\in\partial i\setminus\mu}\left(1-p\prod_{j\in\partial\nu\setminus i}\eta_{j\to\rho}\right)} \ . \tag{C13}$$

This quantity is very similar to the eigenvalue $\theta_{i\to\mu}^{j\to\nu}$ defined in (C7) during the study of the type I inequality, except for a global sign and a shift from $p+1$ to $p+2$ in the bracket of the numerator.

Finally the instability of typical random hypergraphs with excess degree distribution $r_d$ is studied through a population of couples $(\eta, \varepsilon)$, evolving according to

$$\mathcal{P}^{(n+1)}(\eta,\varepsilon) = \sum_{d=0}^{\infty} r_d \int \prod_{\nu=1}^{d}\prod_{j=1}^{K-1} \mathrm{d}\eta_{j\to\nu}\mathrm{d}\varepsilon_{j\to\nu}\mathcal{P}^{(n)}(\eta_{j\to\nu},\varepsilon_{j\to\nu}) \ \delta\left[\eta - \mathcal{F}^{\rm SP}(\{\eta_{j\to\nu}\})\right] \delta\left[\varepsilon - \sum_{\nu=1}^{d}\sum_{j=1}^{K-1}(q-1)\mathcal{A}^{j\to\nu}\varepsilon_{j\to\nu}\right] \ , \tag{C14}$$

an equation similar to (C8) used for the study of the type I instability. As explained in that case we define $\lambda_{\rm II}$ as the quantity by which the $\varepsilon$ should be divided to keep their norm constant under these iterations.

## Appendix D: Details on the asymptotic expansions

In this section we present some more details on the computations that yield the asymptotic expansion of the thresholds presented in IV C. In these limits of large $K$ and/or $q$ the relevant average degree diverges, the Poisson random graph ensemble with average degree $\ell \gg 1$ and the regular ensemble with degree $\ell \gg 1$ can be considered as equivalent at leading order as the Poisson distribution concentrates around its average value. Depending on the threshold to be determined we shall thus use the most convenient ensemble between these two.

### 1. Rigidity threshold at $m = 1$

Consider the equation (58) that gives the probability of a frozen variable in the 1RSB cavity formalism at $m = 1$, for a

Poissonian degree distribution with mean $\ell$. We rewrite it as

$$x = f(x,\ell) = \left(1 - e^{-\ell \frac{x^{K-1}}{q^{K-1}-1}}\right)^{q-1} , \tag{D1}$$

where we defined for simplicity $x = q\eta$. The rigidity threshold $\ell_{\rm r}$ is defined as the smallest value of $\ell$ such that there exists a strictly positive solution $x_{\rm r}$ of this fixed point equation. As $f$ behaves as $x^{(K-1)(q-1)}$ around $x = 0$, the bifurcation at $\ell_{\rm r}$ has to occur discontinuously whenever $(K-1)(q-1) > 1$, a condition that we assume in the following, in which we shall denote $L = (K-1)(q-1)$. Determining $\ell_{\rm r}$ thus amounts to solve for $\ell_{\rm r}$ and $x_{\rm r} > 0$ solution of

$$\begin{cases} x_{\rm r} = f(x_{\rm r},\ell_{\rm r}) \\ 1 = \frac{\partial f}{\partial x}\big|_{(x_{\rm r},\ell_{\rm r})} \end{cases} . \tag{D2}$$

These two equations are easily shown to be equivalent to

$$\begin{cases} \ell_{\rm r} = (q^{K-1}-1)\frac{1}{x_{\rm r}^{K-1}} \ln\left(\frac{1}{1-x_{\rm r}^{1/(q-1)}}\right) \\ 1 = L(1-x_{\rm r}^{-1/(q-1)})\ln(1-x_{\rm r}^{1/(q-1)}) \end{cases} . \tag{D3}$$

It now remains to solve for $x_{\rm r}$ in the second equation and replace in the first one to obtain the asymptotic expansion of $\ell_{\rm r}$. Denoting $y = 1/(1-x_{\rm r}^{1/(q-1)})$, the second equation fixes the value of $y(L)$ according to

$$1 = L\frac{1}{y(L)}\frac{1}{1-\frac{1}{y(L)}}\ln y(L) . \tag{D4}$$

To obtain the asymptotic expansion of $y(L)$ when $L \to \infty$ we rewrite this as

$$\begin{aligned} y(L) &= L\ln y(L)\frac{1}{1-\frac{1}{y(L)}} \qquad (D5)\\ &= L\left(\ln L + \ln\ln y(L) - \ln\left(1-\frac{1}{y(L)}\right)\right)\frac{1}{1-\frac{1}{y(L)}} \end{aligned}$$

From this equation one easily sees that $y(L) \sim L\ln L$, and then more precisely that

$$\begin{aligned} y(L) &= L\left(\ln L + \ln\ln L + O\left(\frac{\ln\ln L}{\ln L}\right)\right) , \qquad (D6)\\ \ln y(L) &= \ln L + \ln\ln L + O\left(\frac{\ln\ln L}{\ln L}\right) . \qquad (D7) \end{aligned}$$

Coming back to the determination of the asymptotic behavior of $\ell_{\rm r}$ we note that

$$\begin{aligned} \frac{1}{x_{\rm r}^{K-1}}\ln\left(\frac{1}{1-x_{\rm r}^{1/(q-1)}}\right) &= \ln(y(L))\left(1-\frac{1}{y(L)}\right)^{-L} \qquad (D8)\\ &= \ln L + \ln\ln L + 1 + O\left(\frac{\ln\ln L}{\ln L}\right) \end{aligned}$$

Inserting this expansion in the expression of $\ell_{\rm r}$ given in (D3) yields the formula we stated in (67), noting finally that $\ln\ln L = \ln\ln(Kq) + o(1)$ as soon as $L$ diverges, whenever $K$ and/or $q$ go to infinity.

### 2. Rigidity threshold at $m = 0$

We turn now to the determination of the rigidity treshold at $m = 0$, in other words the smallest degree for which the SP equation (61) admits a non-trivial solution. This computation turns out to be easier in the regular ensemble, for which $\mathcal{P}^{\rm SP}$ is a Dirac distribution on $\eta$ solution of

$$\eta = \frac{\sum_{p=0}^{q-1}(-1)^p\binom{q-1}{p}(1-(p+1)\eta^{K-1})^{\ell-1}}{\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}(1-p\eta^{K-1})^{\ell-1}} , \tag{D9}$$

as follows from (60). Inspired by the study of the $m = 1$ rigidity we consider the following scale of degrees,

$$\ell = q^{K-1}(\ln L + \ln\ln L + \ell' + o(1)) , \tag{D10}$$

with $\ell'$ a constant, and assume that the solution $\eta$ of (D9) behaves in this case as

$$\eta = \frac{1}{q}\left(1 - \frac{\eta'}{(K-1)\ln L} + o\left(\frac{1}{(K-1)\ln L}\right)\right) , \tag{D11}$$

where $\eta'$ is a constant to be determined as a function of $\ell'$. To do so we first note that as $L \to \infty$ one has

$$\eta^{K-1} = \frac{1}{q^{K-1}}\left(1 - \frac{\eta'}{\ln L} + o\left(\frac{1}{\ln L}\right)\right) . \tag{D12}$$

Then we compute

$$\begin{aligned} (1-p\eta^{K-1})^{\ell-1} &= \exp\Big[-p(\ln L + \ln\ln L + \ell' + o(1)) \\ &\qquad \left(1 - \frac{\eta'}{\ln L} + o\left(\frac{1}{\ln L}\right)\right)\Big] \\ &\sim \left(\frac{1}{L\ln L}e^{\eta'-\ell'}\right)^p . \qquad (D13) \end{aligned}$$

Finally we note that

$$\begin{aligned} \frac{\sum_{p=0}^{q-1}(-1)^p\binom{q-1}{p}u^{p+1}}{\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}u^p} &= \frac{u(1-u)^{q-1}}{1-(1-u)^q} \qquad (D14)\\ &= \frac{1}{q}\left(1 - \frac{q-1}{2}u + O((qu)^2)\right) , \end{aligned}$$

and apply this identity with $u$ the quantity defined in (D13). This gives us, from (D9) and the ansatz made on $\eta$,

$$1 - \frac{\eta'}{(K-1)\ln L} \sim 1 - \frac{q-1}{2}\frac{1}{L\ln L}e^{\eta'-\ell'} . \tag{D15}$$

Recalling that $L = (K-1)(q-1)$ we see that the ansatz was indeed self-consistent, and that $\eta'$ is determined as a function of $\ell'$ through

$$\eta' e^{-\eta'} = \frac{1}{2}e^{-\ell'} . \tag{D16}$$

This equation admits solutions only if $\ell' > 1 - \ln 2$, hence the statement on $C(m = 0)$ made in the main text.

### 3. Colorability threshold

To compute the asymptotics of the colorability transition we shall also work in the regular ensemble, we thus have to find $\eta$ solution of (D9), compute the complexity

$$\Sigma(m=0) = \ln\left(\sum_{p=1}^{q}(-1)^{p+1}\binom{q}{p}(1-p\eta^{K-1})^{\ell}\right) - \frac{\ell(K-1)}{K}\ln\left(1-q\eta^{K}\right) , \tag{D17}$$

and determine the $\ell$ where it vanishes.

We shall fix some $q \geq 2$ and take the limit $K \to \infty$; this allows to organize the asymptotic expansions with exponential dependency on $K$ dominating the polynomial ones. We shall in particular consider degrees on the scale

$$\ell(K) = Kq^{K-1}\ln q - \ell_1(K) - \frac{1}{q^{K-1}}\ell_2(K) + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) \tag{D18}$$

with the $\ell_i$ polynomial in $K$, as we know that the colorability threshold coincides, at its leading order, with the upperbound given by the first moment method. We recall that the notation $\widetilde{O}$ hides polynomial terms in $K$. Here we shall content ourselves with the first correction $\ell_1$, but it is not too difficult to generalize the computation to higher orders. For this scale of degrees we make the following ansatz on the solution $\eta$ of (D9),

$$\eta = \frac{1}{q}\left(1 - \frac{1}{q^{K-1}}\eta_1 - \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right)\right) \tag{D19}$$

with $\eta_1 = \widetilde{O}(1)$. One finds after a short computation that

$$(1-p\eta^{K-1})^{\ell-1} = \left(\frac{1}{q^K}\right)^p \left(1 + \frac{1}{q^{K-1}}\left[p(\ell_1+1+K(K-1)\eta_1\ln q) - \frac{1}{2}p^2K\ln q\right] + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right)\right) \tag{D20}$$

Plugging this result in (D9) we can keep only the terms $p = 0, 1$ in the numerator, and $p = 1, 2$ in the denominator. This shows that the assumption (D19) is indeed self-consistent and fixes the coefficient $\eta_1 = (q-1)/(2q)$. Performing the same kind of expansion on the expression (D17) of the complexity one finds

$$\Sigma(m=0) = \frac{1}{q^{K-1}}\left[\frac{\ell_1}{K} - \frac{1}{2}\left(1 - \frac{1}{q} + \ln q\right)\right] + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) . \tag{D21}$$

It is then trivial to deduce the value of $\ell_1$ for which the complexity vanishes at this order, this yields our asymptotic expansion for $\ell_{\rm col}$ stated in (68).

### 4. Condensation threshold

We finally explain our computation of the asymptotic expansion for the condensation threshold given in (69). We shall work in the Poissonian ensemble, with average degrees on the scale (D18), and determine the value of $\ell_1$ such that the complexity $\Sigma(m=1) = \Phi(m=1) - \Phi'(m=1)$ vanishes.

We can first easily expand $\Phi(m=1) = \phi_{\rm RS}$ on this scale to obtain from (28):

$$\Phi(m=1) = \frac{1}{q^{K-1}}\left[\frac{\ell_1}{K} - \frac{1}{2}\ln q\right] + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) . \tag{D22}$$

The non-trivial part of the computation is the expansion of $\Phi'(m=1)$. We shall use its expression in terms of the distribution $\overline{P}$ defined in (45),

$$\Phi'(m=1) = \sum_{d=0}^{\infty} p_d \int \prod_{\mu=1}^{d}\prod_{i=1}^{K-1} \mathrm{d}\psi_{i\to\mu}\overline{P}(\psi_{i\to\mu})\frac{\mathcal{Z}_0^{i+\partial i}}{\overline{\mathcal{Z}}_0^{i+\partial i}}\ln\mathcal{Z}_0^{i+\partial i} - \frac{\ell(K-1)}{K}\int\prod_{i=1}^{K}\mathrm{d}\psi_i\overline{P}(\psi_i)\frac{\mathcal{Z}_0^{\mu}}{\overline{\mathcal{Z}}_0^{\mu}}\ln\mathcal{Z}_0^{\mu} , \tag{D23}$$

where $p_d$ is the Poissonian distribution of average $\ell$, and

$$\mathcal{Z}_0^{i+\partial i} = \sum_{x=1}^{q}\prod_{\mu=1}^{d}\left(1 - \prod_{i=1}^{K-1}\psi_{i\to\mu}(x)\right) , \tag{D24}$$

$$\overline{\mathcal{Z}}_0^{i+\partial i} = q\left(1 - \frac{1}{q^{K-1}}\right)^d , \tag{D25}$$

$$\mathcal{Z}_0^{\mu} = 1 - \sum_{x=1}^{q}\prod_{i=1}^{K}\psi_i(x) , \tag{D26}$$

$$\overline{\mathcal{Z}}_0^{\mu} = 1 - \frac{1}{q^{K-1}} . \tag{D27}$$

The distribution $\overline{P}$ is the solution of (46) for the excess law $r_d = p_d$. We decompose as usual the contribution of the hard fields with

$$\overline{P}(\psi) = \eta\sum_{x=1}^{q}\delta[\psi - \psi^x] + (1-q\eta)\widetilde{P}(\psi) . \tag{D28}$$

The weight $\eta$ of the hard fields is solution of (58); on this scale of degrees we let the reader check that it has the asymptotic expansion (D19), with $\eta_1 = (q-1)/q$ (notice the factor 2 difference with the expansion at $m=0$). The fraction of soft fields is thus exponentially small in this regime; moreover one can safely assume that the soft fields are perfectly unbiased, i.e. that $\widetilde{P}(\psi) = \delta[\psi - \overline{\psi}]$. This is an approximation that does not change the estimate of the condensation threshold at this order.

Let us first consider the second line in (D23). We denote $K_0$, $K_1$, …, $K_q$ the number of the random $\psi_i$ which are equal, respectively, to $\overline{\psi}$, $\psi^1$, …, $\psi^q$. The distribution of $(K_0, K_1, \ldots, K_q)$ is thus multinomial with parameters $(K; 1-q\eta, \eta, \ldots, \eta)$. One can easily compute the value of $\mathcal{Z}_0^{\mu}$ as a function of $(K_0, K_1, \ldots, K_q)$; one finds that it is equal to 1 in many cases, except when $K_0 = K$ (hence $K_1 = \cdots = K_q = 0$) where one has $\mathcal{Z}_0^{\mu} = 1 - \frac{1}{q^{K-1}}$, or when there is exactly one color $x$ with $K_x > 0$, then $\mathcal{Z}_0^{\mu} = 1 - \frac{1}{q^{K-K_x}}$. We can thus write

$$\int \prod_{i=1}^{K} \mathrm{d}\psi_i \overline{P}(\psi_i) \mathcal{Z}_0^\mu \ln \mathcal{Z}_0^\mu = (1-q\eta)^K \left(1-\frac{1}{q^{K-1}}\right) \ln\left(1-\frac{1}{q^{K-1}}\right) + q \sum_{K_x=1}^{K-1} \binom{K}{K_x} \eta^{K_x} (1-q\eta)^{K-K_x} \left(1-\frac{1}{q^{K-K_x}}\right) \ln\left(1-\frac{1}{q^{K-K_x}}\right) , \tag{D29}$$

the only approximation made up to now being the replacement of $\widetilde{P}$ by $\delta[\psi - \overline{\psi}]$. We can then use the scaling (D19) of $\eta$: it implies that the leading behavior of this expression comes from the term $K_x = K-1$ in the sum, which gives

$$-\frac{\ell(K-1)}{K} \int \prod_{i=1}^{K} \mathrm{d}\psi_i \overline{P}(\psi_i) \frac{\mathcal{Z}_0^\mu}{\overline{\mathcal{Z}}_0^\mu} \ln \mathcal{Z}_0^\mu = \tag{D30}$$
$$-\frac{1}{q^{K-1}} K(K-1) \frac{(q-1)^2}{q} (\ln q) \ln\left(1-\frac{1}{q}\right) + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) .$$

The first line of (D23) can be handled in a similar way. We classify the $d$ constraints in different types according to the type of messages they receive from their $K-1$ variables, writing

$$d = d_{\mathrm{m}} + d_{\mathrm{s}} + \sum_{x=1}^{q} \sum_{K_0=0}^{K-2} d_{x,K_0} . \tag{D31}$$

In this expression $d_{\mathrm{s}}$ is the number of clauses that receive $K-1$ soft messages, $d_{x,K_0}$ counts those that receive $K-1-K_0$ messages of color $x$ and $K_0$ soft messages, and $d_{\mathrm{m}}$ counts all the other situations (i.e. those clauses that receive hard fields of at least two different colors). As $d$ is a random Poisson variable and as the type of each constraint is drawn independently of the others under the integrals in the first line of (D23), we see that these random degrees $\{d_0, d_{\mathrm{s}}, \{d_{x,K_0}\}\}$ are independent Poisson variables, with averages

$$\mathbb{E}[d_{\mathrm{m}}] = \ell p_{\mathrm{m}} \ , \ \mathbb{E}[d_{\mathrm{s}}] = \ell p_{\mathrm{s}} \ , \ \mathbb{E}[d_{x,K_0}] = \ell p_{K_0} . \tag{D32}$$

The $p$'s are here the probabilities of the state of one constraint, from their definition one sees that

$$p_{\mathrm{s}} = (1-q\eta)^{K-1} , \tag{D33}$$

$$p_{K_0} = \binom{K-1}{K_0} \eta^{K-1-K_0} (1-q\eta)^{K_0} , \tag{D34}$$

$$p_{\mathrm{m}} = 1 - p_{\mathrm{s}} - q \sum_{K_0=0}^{K-2} p_{K_0} . \tag{D35}$$

For a given state of this decomposition one can express $\mathcal{Z}_0^{i+\partial i}$ as

$$\mathcal{Z}_0^{i+\partial i} = \left(1-\frac{1}{q^{K-1}}\right)^{d_{\mathrm{s}}} \sum_{x=1}^{q} \mathbb{I}(d_{x,0}=0) \prod_{K_0=1}^{K-2} \left(1-\frac{1}{q^{K_0}}\right)^{d_{x,K_0}} .$$

Indeed the edges counted in $d_{x,0}$ have $K-1$ vertices forced in the color $x$, hence if $d_{x,0} > 0$ the central vertex cannot take this color. On the other hand if at least two vertices are forced in at least two distinct colors the central vertex can take any color indifferently, hence $d_{\mathrm{m}}$ does not appear in this expression. The intermediate situations arise when exactly one color is forced in some of the $K-1$ neighbors in one constraint: the most numerous these are, the more the central vertex is biased to avoid this color.

It remains to compute the average of $(\mathcal{Z}_0 \ln \mathcal{Z}_0)/\overline{\mathcal{Z}}_0$ with the Poisson distribution of the various $d$. We first notice that on this asymptotic scale of degrees $\overline{\mathcal{Z}}_0 = (1/q^{K-1})(1 + \widetilde{O}(1/q^{K-1}))$ and that the fluctuations in the Poisson degree $d$ can be neglected at the leading order, we can thus take $1/\overline{\mathcal{Z}}_0$ out of the average. Then we isolate the number $n$ of allowed colors for the central spin and write

$$\mathbb{E}[\mathcal{Z}_0^{i+\partial i} \ln \mathcal{Z}_0^{i+\partial i}] = \sum_{n=0}^{q} \binom{q}{n} \left(e^{-\ell p_0}\right)^n \left(1-e^{-\ell p_0}\right)^{q-n} \mathbb{E}[\mathcal{Z}(n) \ln \mathcal{Z}(n)] \tag{D36}$$
$$\mathcal{Z}(n) = \left(1-\frac{1}{q^{K-1}}\right)^{d_{\mathrm{s}}} \sum_{x=1}^{n} \prod_{K_0=1}^{K-2} \left(1-\frac{1}{q^{K_0}}\right)^{d_{x,K_0}}$$

At this point we use the asymptotic behavior of $\eta$ and realize that

$$e^{-\ell p_0} = \frac{1}{q} \frac{1}{q^{K-1}} + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) \tag{D37}$$

$$\ell p_1 = K(K-1)(q-1) \ln q \frac{1}{q^{K-1}} + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) \tag{D38}$$

$$\ell p_r = \widetilde{O}\left(\frac{1}{(q^{K-1})^r}\right) \ \text{for } r \geq 2 , \tag{D39}$$

while $\ell p_{\mathrm{s}}$ is neglectible at all perturbative orders. The leading orders of (D36) thus comes from the smallest possible values of $n$. However $\mathcal{Z}(n=0) = 0$ does not contribute to the sum, we have thus to consider $n=1$ and $n=2$. In the first case the most probable configuration with $d_{\mathrm{s}} = d_{1,1} = \cdots = d_{1,K-2} = 0$ will not contribute: this would yield $\mathcal{Z} = 1$, hence cancelling the logarithm. The next-to-most probable configuration of the random variables is $d_{1,1} = 1$, which happens with probability $\ell p_1 e^{-\ell p_1}$. From these observations we thus obtain

$$\mathbb{E}\left[\frac{\mathcal{Z}_0^{i+\partial i}}{\overline{\mathcal{Z}}_0^{i+\partial i}}\ln \mathcal{Z}_0^{i+\partial i}\right] = q^{K-1}\left[q\,e^{-\ell p_0}\ell p_1\left(1-\frac{1}{q}\right)\ln\left(1-\frac{1}{q}\right)+\frac{q(q-1)}{2}\left(e^{-\ell p_0}\right)^2 2\ln 2\right]\widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) \tag{D40}$$

$$= \frac{1}{q^{K-1}}\left[K(K-1)\frac{(q-1)^2}{q}(\ln q)\ln\left(1-\frac{1}{q}\right)+\left(1-\frac{1}{q}\right)\ln 2\right]+\widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right) \tag{D41}$$

Adding up the contribution from (D30) this yields

$$\Phi'(m=1) = \frac{1}{q^{K-1}}\left(1-\frac{1}{q}\right)\ln 2+\widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right)\ , \tag{D42}$$

hence the leading order of the complexity is

$$\Sigma(m=1) = \frac{1}{q^{K-1}}\left[\frac{\ell_1}{K}-\frac{1}{2}\ln q-\left(1-\frac{1}{q}\right)\ln 2\right] + \widetilde{O}\left(\frac{1}{(q^{K-1})^2}\right)\ ,$$

which completes our justification of (69).