Typology of phase transitions in Bayesian inference problems

Federico Ricci-Tersenghi,[1] Guilhem Semerjian,[2] and Lenka Zdeborová[3]

[1]*Dipartimento di Fisica, Sapienza Università di Roma, Nanotec CNR, UOS di Roma, and INFN Sezione di Roma 1,
Piazzale Aldo Moro 5, 00185 Roma, Italy*
[2]*Laboratoire de Physique, Ecole Normale Supérieure, Université PSL, CNRS, Sorbonne Université,
Université Paris–Diderot, Université Sorbonne Paris Cité, Paris, France*
[3]*Institut de Physique Théorique, CNRS, CEA, Université Paris–Saclay, Gif-sur-Yvette, France*

Many inference problems undergo phase transitions as a function of the signal-to-noise ratio, a prominent example of this phenomenon being found in the stochastic block model (SBM) that generates a random graph with a hidden community structure. Some of these phase transitions affect the information-theoretic optimal (but possibly computationally expensive) estimation procedure, others concern the behavior of efficient iterative algorithms. A computational gap opens when the phase transitions for these two aspects do not coincide, leading to a hard phase in which optimal inference is computationally challenging. In this paper we refine this description in two ways. From a qualitative perspective, we emphasize the existence of more generic phase diagrams with a hybrid-hard phase, in which it is computationally easy to reach a nontrivial inference accuracy but computationally hard to match the information-theoretic optimal one. We support this discussion by quantitative expansions of the functional cavity equations that describe inference problems on sparse graphs, around their trivial solution. These expansions shed light on the existence of hybrid-hard phases, for a large class of planted constraint satisfaction problems, and on the question of the tightness of the Kesten-Stigum (KS) bound for the associated tree reconstruction problem. Our results show that the instability of the trivial fixed point is not generic evidence for the Bayes optimality of the message-passing algorithms. We clarify in particular the status of the symmetric SBM with four communities and of the tree reconstruction of the associated Potts model: In the assortative (ferromagnetic) case the KS bound is always tight, whereas in the disassortative (antiferromagnetic) case we exhibit an explicit criterion involving the degree distribution that separates a large-degree regime where the KS bound is tight and a low-degree regime where it is not. We also investigate the SBM with two communities of different sizes, also known as the asymmetric Ising model, and describe quantitatively its computational gap as a function of its asymmetry, as well as a version of the SBM with two groups of q_1 and q_2 communities. We complement this study with numerical simulations of the belief propagation iterative algorithm, confirming that its behavior on large samples is well described by the cavity method.

I. INTRODUCTION

Problems of statistical inference where a signal is observed via noisy measurements appear ubiquitously in situations involving data analysis. It is intuitive that as the signal-to-noise ratio decreases, the recovery of the signal becomes harder. Quantifying the optimal performance is one of the main theoretical tools to provide performance guarantees in practical situations. Numerous models of statistical inference bear formal analogies to models of disordered systems and due to this connection methods from statistical physics were applied to inference problems (for reviews see, e.g., [1–4]). Sharp changes in behavior, known as phase transitions, appear in many situations in physical systems. Due to the above connection, it does not come as a surprise that various types of phase transitions were also identified in inference problems. The statistical and computational implications of the presence of phase transitions have recently attracted attention in several fields.

A number of works, including the present one, focus on inference models defined on sparse graphs or hypergraphs where the signal is related to a (hidden) labeling of the nodes and the graph is generated via a rule that depends on these labels. A paradigmatic example of this class is the stochastic block model (SBM), where nodes belong to communities and the probability of observing an edge between a pair of nodes depends on the communities of the two nodes. There is a long history of studies of this model (see, e.g., [3,5,6] and references therein), particularly because of its relevance in the context of complex networks [7]. In terms of phase transitions in the SBM, the work of [8,9] was particularly influential as it identified three phases, undetectable, hard, and easy, and located quantitatively transitions between them. In the undetectable phase, inference better than random guessing is not possible; in the hard phase it is information-theoretically possible, but known algorithms are not able to perform better than random guesses (depending on

parameters of the model, this phase is sometimes missing); in the easy phase known algorithms match the information-theoretic optimal performance that is in this case strictly better than random guesses. Another example of inference problems on sparse hypergraphs is the planted constraint satisfaction problems with various types of constraints [10–13]. Planted constraint satisfaction problems are instrumental in studies of average algorithmic complexity [14], hardness of finding a solution [15], and refutation of satisfiability formulas. The same three phases explained above in the SBM example have also been found in planted constraint satisfaction problems, as well as in many other "dense" inference problems in which the signal is observed through all pairs of variables, not only the edges of a sparse graph (see [3] for a review).

This paper provides a more detailed picture of these phase transitions. We emphasize indeed the existence of hybrid-hard phases, in which efficient inference better than random guessing is possible, however, matching the information-theoretic optimal performance is computationally hard. This type of phase was identified in dense inference problems (see, e.g., [16]), but was missed in some previous studies of sparse problems [11]. We revisit the latter by means of the cavity method, which relates the analytic description (in particular the optimal errors, as defined below) of inference problems on sparse graphs and hypergraphs to reconstruction problems on trees [17–22] (the local limit of the graphs in the large-size limit). These tree problems are solved by fixed-point equations whose unknowns (called order parameters) are functions (more precisely probability distributions) at variance with their dense counterparts where the order parameters are finite dimensional. We will focus on problems where the undetectable phase exists at a small signal-to-noise ratio, which translates into the existence of a so-called uninformative (or trivial) fixed point of the corresponding cavity equations (in such cases there exists a close connection between the inference problems and the associated uniformly random models [23–25] through the notion of quiet planting). There are several reasons that motivate this restriction of the family of inference problems we consider here: In a certain sense, these are the hardest ones, as no local information can be exploited for the inference of the hidden signal they contain. Moreover, they exhibit a richer phenomenology in terms of phase transitions than the problems where a local property (the degree of a vertex in the SBM, for instance) is correlated to the signal. Finally, from a technical point of view, the existence of a trivial fixed point allows one to set up a systematic perturbative expansion of the cavity equations, in the neighborhood of the so-called Kesten-Stigum transition where the trivial fixed point loses its stability; this perturbative computation will be our main technical contribution in this paper. Similar expansions can be found in Refs. [20,22,25], but our results are either valid for more general models or pushed to a higher order in the perturbation series. This technical tool allows us to clarify several features of the phase diagrams of inference and tree reconstruction problems.

For the symmetric stochastic block model with q groups (or the tree reconstruction of the q-state Potts model), it was known that the hard phase does not exist for two and

three groups (in other words, the Kesten-Stigum bound on reconstruction is tight), whereas it always exists for five and more groups [22]. In the intermediate case of $q = 4$, whose status has remained unclear up to now, we provide an explicit condition on the degree distribution and the ferromagnetic (or assortative) character of the model to distinguish these two qualitatively different behaviors.

We also study the asymmetric SBM with two groups of different sizes (but equal average degree) and derive conditions on the asymmetry that induces the appearance of the hard phase. We recover the critical fraction $1/2 - 1/\sqrt{12}$ for the size of the smaller group below which the hard phase exists, which appeared in related studies in the dense regime [16,26] or in the limit of large degrees [27,28]; quite strikingly, this critical asymmetry does not depend on the degree distribution. This phenomenon was also studied from the tree reconstruction perspective under the name of the asymmetric Ising model [17,20].

Based on this expansion, we also conclude that the hybrid-hard phase always exists in a very generic class of sparse inference models where Boolean variables are observed through k-uplets, for any k, as long as they preserve a global symmetry between the two possible values of the variables. This class encompasses a large part of the Boolean occupation problems of [11]. More precisely, in that class of models we rule out the existence of a hard phase that would not be accompanied by the hybrid-hard phase.

Our analytical findings are confirmed by running the belief propagation (BP) iterative algorithm on large instances of the corresponding problems, thus providing strong evidence that the analytical solution obtained in the thermodynamical limit is also of practical relevance for problems of large but finite size.

Two recent papers [29,30] have some overlap with ours. Liu *et al.* study indeed the reconstruction problem on regular d-ary trees and provide expansions around the Kesten-Stigum transition for the asymmetric Ising model and a q-state Potts model whose symmetry is partially broken. While the results of [29,30] are rigorous, our expansions are more generic, pushed to a higher order, and put to use in connection with the inference problems on graphs.

The rest of the paper is organized as follows. In Sec. II we present in generic terms the structure describing the performance of optimal and efficient estimators in inference problems, review the well-known phase diagrams that have been deduced from it, and explain the more complicated possible scenarios that can be observed. In Sec. III we define a family of inference problems on sparse (hyper)graphs, introduce the cavity equations that describe them, and explain their link to the tree reconstruction problem; for the convenience of the reader, we then summarize our main results in Sec. III D. The results are obtained by systematic moment expansions of the functional cavity equations, which are presented in Sec. IV for arbitrary discrete variables with pairwise interactions and in Sec. V for Boolean variables with ($k \geqslant 2$)-wise interactions; the more technical parts of these computations are deferred to the Appendixes. Section VI is devoted to numerical experiments with belief propagation on finite-size samples. In Sec. VII we summarize and offer perspectives for future work.

II. TYPOLOGY OF PHASE TRANSITIONS

This section is dedicated to the presentation of our qualitative results on the variety of possible phase diagrams in inference problems. We will first briefly review the Bayesian perspective on inference and the way statistical mechanics handles it, staying at an abstract level of description (concrete examples will be introduced in Sec. III). The typology of phase transitions will then naturally follow from this discussion, as well as the quantitative computations that can be performed on concrete examples as first steps towards this classification (which will form the core of the rest of the paper).

A. Bayesian inference problems

In a statistical inference problem a ground-truth vector $\underline{s}^* \in \mathbb{R}^N$ is to be inferred from some observations (or data) denoted by G that correlate with $\underline{s}^*$. In the Bayesian setting $\underline{s}^*$ is a random variable whose distribution is called prior and G is a random variable with a conditional law $P(G|\underline{s}^*)$ that correlates it with the ground truth. An observer provided with a sample of G and with the knowledge of the model that generated it, i.e., of the prior distribution and of $P(G|\underline{s}^*)$, must base one's inference of $\underline{s}^*$ on the posterior distribution that follows from Bayes' theorem

$$P(\underline{s}|G) = \frac{1}{Z(G)} P(G|\underline{s}) \prod_{i=1}^{N} P_0(s_i), \qquad (1)$$

where we assume for simplicity a prior distribution P_0 factorized on the components of the vector and $Z(G)$ is a normalizing constant. The observer uses an estimator $\hat{\underline{s}}(G)$ that should ideally be close to $\underline{s}^*$. Which estimator is optimal depends on the definition of closeness between $\hat{\underline{s}}(G)$ and the true value $\underline{s}^*$. To clarify this point, let us define the marginal probability distributions of the posterior as

$$\mu_i(s_i) = \sum_{\{s_j\}_{j \neq i}} P(\underline{s}|G). \qquad (2)$$

If the distance between $\hat{\underline{s}}(G)$ and $\underline{s}^*$ is measured in terms of the mean-square error (MSE) $\sum_i (\hat{s}_i - s_i^*)^2 / N$, then the optimal estimator (that minimizes the MSE) is simply given by the means of the marginals

$$\hat{s}_i = \sum_{s_i} s_i \mu_i(s_i), \qquad (3)$$

and the error achieved by this estimator is called the minimum mean-square error (MMSE). When the variables s_i belong to a discrete set χ (as will be the case in this paper) another meaningful definition of similarity between $\hat{\underline{s}}(G)$ and $\underline{s}^*$ is the mean overlap $\sum_i \delta_{\hat{s}_i, s_i^*} / N$. The estimator that maximizes the mean overlap is given by the value of s_i for which the marginal is the largest, namely,

$$\hat{s}_i = \arg\max \mu_i(s_i). \qquad (4)$$

We are interested in this paper in phase transitions, i.e., qualitative changes of the behavior of such inference problems and nonanalyticities in their optimal error. These can only occur in the so-called thermodynamic limit $N \to \infty$; the dimensionality of G must be scaled appropriately for

this limit to be nontrivial, in such a way that the signal-to-noise ratio (SNR) (whose precise definition depends on the specific problem considered) that measures the quantity of information on $\underline{s}^*$ conveyed by G remains finite in the limit. We will focus on a particularly interesting and challenging class of problems for which there exists a nontrivial low-SNR phase in which the inference problem cannot be solved more precisely than by using only the prior information; we will call this phase undetectable, as the ground truth $\underline{s}^*$ has asymptotically no effect on the posterior distribution. We will call the accuracy $a \geqslant 0$ of an estimation procedure a measure of the additional information it exploits in the posterior distribution with respect to the prior in such a way that $a = 0$ in the undetectable phase. The precise definition of the accuracy is a model-dependent problem, to which we will return in Sec. IV K, with rather subtle pitfalls (see, for example, [6] for a discussion of this point); for instance, when the problem admits an exact symmetry, the marginals of the posterior can be strictly equal to the prior distribution for an arbitrary large SNR. This is, for instance, the case of the symmetric SBM, which is invariant under the permutation of the labels; in such a case the definition of the distance between $\underline{s}^*$ and $\hat{\underline{s}}(G)$ must be adapted to break explicitly the symmetry by considering, for instance, the overlap modulo of an arbitrary global permutation of the labels of the vertices.

B. Statistical mechanics description

Several Bayesian inference problems have recently been studied via statistical mechanics methods, and many of the predictions thus obtained have been confirmed rigorously; our goal here is not to review these results (see, for instance, [3–6]) but to emphasize their common formal structure that is at the origin of the few possible universality classes of phase diagrams. Consider, from this abstract perspective, an inference problem parametrized by a signal-to-noise ratio c. In the statistical mechanics approach the accuracy of the Bayes-optimal estimator is expressed as a variational problem: One derives (with the cavity or replica method) a so-called free energy f, which is a function of c and of an order parameter a (an object whose nature depends on the problem under study). This free energy has to be minimized with respect to the order parameter, and the optimal accuracy is then a function of the location of this global minimum (the minimal free energy yields instead the mutual information between the signal and the observations). In dense models, i.e., those in which there are much more than N measurements, each giving weak information on the signal (see, e.g., [16]), the order parameter is a scalar (or a finite-dimensional vector); in the sparse models studied later in this paper the order parameter is functional, yet the qualitative behavior is the same. To simplify this formal discussion let us stick to a scalar order parameter and furthermore use the accuracy $a \geqslant 0$ as the order parameter itself, with the minimal value $a = 0$ corresponding to the trivial uninformative estimator (exploiting only the prior and not the observations). We write the free energy as

$$f(a, c) = \Phi(a, c) - \frac{a^2}{2}; \qquad (5)$$

the motivation for the subtraction of the second term will be clarified below. The location of the global minimum of f with respect to a, denoted by $a^*(c)$, is thus a prediction for the accuracy of the optimal estimator, irrespective of its computational complexity.

The cavity method also enlightens the computational complexity of the inference problems. As a matter of fact, the derivation of the free energy is tightly linked to the analysis of an efficient iterative algorithm (called approximate message passing [31,32] in the dense case or belief propagation in the sparse regime [33]) and provides a dynamical map that describes the discrete time evolution of the order parameter during the execution of the algorithm. In our abstract setting this yields an accuracy $a^{(n)}$ after n steps of the algorithm, which evolves according to

$$a^{(n)} = \varphi(a^{(n-1)}, c), \qquad (6)$$

with the initial condition $a^{(0)} = \varepsilon$, where $\varepsilon > 0$ is an infinitesimal positive constant. Initially, the algorithm only knows the prior on the signal and at each time step uses the observation to iteratively update and improve the accuracy of its belief on the true value of the signal. We define $a_{\mathrm{alg}}(c) = \lim_{n \to \infty} a^{(n)}$ as the accuracy ultimately reached by this algorithm.

The fundamental connection between the information-theoretic optimal accuracy and the description of the iterative algorithm is expressed mathematically in our setting by the relation

$$\varphi(a, c) = \Phi'(a, c). \qquad (7)$$

Here and in the following the prime denotes a derivative with respect to the variable a. This equation implies indeed a tight connection between the static description in terms of the free energy $f(a, c)$ and the dynamical one in terms of the iteration function $\varphi(a, c)$: The stationary points of the former correspond to the fixed points of the latter. As mentioned previously, we focus in this paper on problems which admit an undetectable phase; this translates in the abstract formalism employed here into the assumption that $a = 0$ is, for all values of the SNR c, a fixed point of the iteration (6), which we call the trivial or uninformative fixed point. Equivalently, $a = 0$ is a stationary point of the free energy $f(a, c)$.

C. Possible phase diagrams

It should be clear at this point that the description of the phase transitions and the classification of the possible phase diagrams is nothing but a bifurcation analysis. As the SNR c is varied, the functions $f(a, c)$ [$\varphi(a, c)$] evolve in a smooth way; their stationary (fixed) points also do so, except at bifurcations where their number and nature can change in a singular way. The qualitative aspects of such a bifurcation diagram do not depend on the details of the function $f(a, c)$, but only on its behavior close to the bifurcations (essentially the order and sign of the first nonvanishing derivatives), which explains the high level of universality among inference problems which can have very different origins.

Let us now describe from this perspective the phase transitions in inference problems, interpreted as bifurcations for the stationary points of $f(a, c)$. Among the possibly many stationary points, two (which can coincide) will play a

particular role: $a^*(c)$, the global minimum of f, which gives the information-theoretic best possible accuracy, and $a_{\mathrm{alg}}(c)$, the fixed point reached by iterations starting infinitesimally close to the trivial one, as it corresponds to the accuracy reachable by an efficient algorithm.

As $a = 0$ is assumed here to be a fixed point for all values of c and as the algorithm starts infinitesimally close to it, the local stability of $a = 0$ under the iterations (6) yields crucial information on the accuracy $a_{\mathrm{alg}}(c)$ reached by the algorithm: If $a = 0$ is stable the iterations will drive the algorithm to the trivial accuracy and then $a_{\mathrm{alg}}(c) = 0$; if not, the dynamical system flows away from this fixed point and one reaches $a_{\mathrm{alg}}(c) > 0$. This stability can be determined very easily by expanding (6) at first order in a,

$$a^{(n)} \approx \varphi'(0, c) a^{(n-1)}, \qquad (8)$$

which shows that $a = 0$ is stable if and only if $\varphi'(0, c) < 1$ [as $a^{(n)} \geqslant 0$ we can assume $\varphi'(0, c) \geqslant 0$]. As c represents here a signal-to-noise ratio, increasing values of c tend to destabilize the trivial fixed point. We thus define the so-called Kesten-Stigum (KS) threshold (from [34] and elucidated below) c_{KS} as the largest signal-to-noise ratio c for which $a = 0$ is a stable fixed point. As mentioned above, for $c < c_{\mathrm{KS}}$ one has $a_{\mathrm{alg}}(c) = 0$ and the algorithm will not be able to infer the underlying signal better than a random guess from the prior. On the other hand, for $c > c_{\mathrm{KS}}$ the iterative algorithm reaches a nontrivial accuracy $a_{\mathrm{alg}}(c) > 0$ and provides a strictly positive correlation between the ground truth signal and its estimator [but not necessarily as good as the optimal estimator (discussed below)].

A moment of thought reveals that this change of stability of the trivial fixed point at c_{KS} must be accompanied by a modification in the total number of fixed points, as the free energy $f(a, c)$ is smooth. Keeping in mind the restriction $a \geqslant 0$, the two simplest global bifurcation diagrams that can arise are depicted in Figs. 1(c) and 1(d), where all the fixed points are represented as a function of the SNR c, with solid (dashed) lines when they are stable (unstable).

The bifurcation diagrams presented in Figs. 1(a) and 1(c) correspond to a so-called continuous (second-order) phase transition. For all values of c there is a single stable fixed point, with a trivial accuracy when $c < c_{\mathrm{KS}}$ and a nontrivial one for $c > c_{\mathrm{KS}}$. The optimal accuracy $a^*(c)$ and the algorithmic one $a_{\mathrm{alg}}(c)$, depicted, respectively, by red and blue lines in Figs. 1(a) and 1(b), coincide for all values of the SNR c. The transition at c_{KS} thus separates an undetectable phase [for $c < c_{\mathrm{KS}}$ no estimator can detect the signal as $a^*(c) = 0$] from an easy phase [for $c > c_{\mathrm{KS}}$ the iterative efficient algorithm matches the optimal performance as $a_{\mathrm{alg}}(c) = a^*(c) > 0$].

The next possible bifurcation diagrams, in order of increasing complexity, are presented in Figs. 1(b) and 1(d) and present a discontinuous (first-order) phase transition. What happens by increasing c around the KS transition is now the disappearance of an unstable nontrivial fixed point (instead of the appearance of a stable nontrivial fixed point for a continuous phase transition). There thus exists an interval of the SNR $c \in [c_{\mathrm{sp}}, c_{\mathrm{KS}}]$ where two stable fixed points coexist (the trivial one $a = 0$ and a nontrivial one). A bifurcation occurs at the spinodal transition c_{sp}, where the high-accuracy branch appears discontinuously. The consequences of this
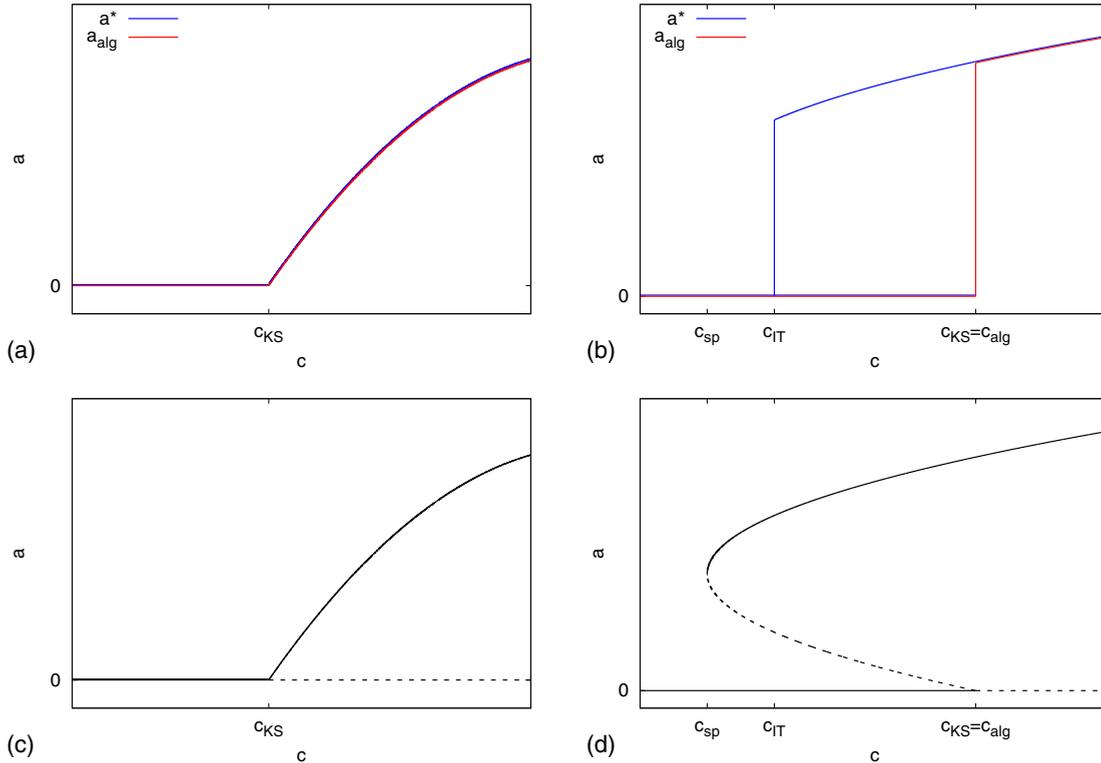
FIG. 1. Two "standard" scenarios with either (a) and (c) a second-order (continuous) transition or (b) and (d) a first-order (discontinuous) transition for (a) and (b) the Bayes-optimal (a^*, in blue) and algorithmic (a_{alg}, in red) performance and (c) and (d) the fixed points of Eq. (6), with solid (dashed) lines for stable (unstable) fixed points. In all panels the vertical axis is the accuracy a and the horizontal axis is the signal-to-noise ratio c. The thresholds c_{sp}, c_{IT}, c_{KS}, and c_{alg}, as defined in the text, are marked on the horizontal axes.

bifurcation diagram on the optimal and algorithmic accuracies $a^*(c)$ and $a_{\mathrm{alg}}(c)$ are presented in Fig. 1(b); the latter is only affected by the change of stability of the trivial fixed point at c_{KS}, above which it jumps to the only stable fixed point which must thus coincide with the optimal accuracy $a^*(c)$. In the interval $c \in [c_{\mathrm{sp}}, c_{\mathrm{KS}}]$ the two stable fixed points correspond to two local minima of the free energy $f(a, c)$; their values cross each other at the information-theoretic phase transition $c_{\mathrm{IT}} \in [c_{\mathrm{sp}}, c_{\mathrm{KS}}]$, the trivial (nontrivial) fixed point being the global minimum for $c \in [c_{\mathrm{sp}}, c_{\mathrm{IT}}]$ ($c \in [c_{\mathrm{IT}}, c_{\mathrm{KS}}]$). One concludes in this case that the undetectable phase (for $c < c_{\mathrm{IT}}$) and the easy phase (for $c > c_{\mathrm{KS}}$) are separated by a hard phase (for $c \in [c_{\mathrm{IT}}, c_{\mathrm{KS}}]$) in which a nontrivial accuracy is information-theoretically possible [$a^*(c) > 0$] yet the iterative algorithm does not provide any correlation with the signal [$a_{\mathrm{alg}}(c) = 0$].

One can easily imagine more and more complicated bifurcation diagrams that will be exhibited by free energies with more and more stationary points; we will content ourselves with the next possibility, depicted in Fig. 2. As shown in Figs. 2(c) and 2(d), the branch of the nontrivial stable fixed point that appears continuously for $c > c_{\mathrm{KS}}$ undergoes a bifurcation at an algorithmic spinodal $c_{\mathrm{alg}} > c_{\mathrm{KS}}$, while the high-accuracy branch disappears discontinuously at the spinodal $c_{\mathrm{sp}} < c_{\mathrm{alg}}$ [the spinodal c_{sp} can occur after or before c_{KS}, as shown in Figs. 2(a) and 2(c) and Figs. 2(b) and 2(d), respectively]. The algorithmic accuracy $a_{\mathrm{alg}}(c)$ is in this case vanishing for $c < c_{\mathrm{KS}}$, growing continuously on the interval $[c_{\mathrm{KS}}, c_{\mathrm{alg}}]$, and jumping discontinuously at c_{alg}. As we assumed here the existence of a single stable fixed

point for $c > c_{\mathrm{alg}}$, the algorithmic accuracy coincides then with $a^*(c)$ (i.e., the jump reaches the optimal value). To determine the information-theoretic optimal accuracy in the presence of two stable fixed points of the recursion, one has to compare the values of the corresponding local minima of the free energy $f(a, c)$. Let us call c_{IT} the information-theoretic transition at which these two free energies cross each other, inducing a discontinuity in the accuracy $a^*(c)$ of the global minimum of the free energy. Figures 2(a) and 2(c) and Figs. 2(b) and 2(d) distinguish further two different scenarios. In Figs. 2(a) and 2(c) one has $c_{\mathrm{IT}} > c_{\mathrm{KS}}$ and the four regimes separated by the transitions $c_{\mathrm{KS}} < c_{\mathrm{IT}} < c_{\mathrm{alg}}$ will be called, for increasing values of c, undetectable, easy, hybrid-hard, and easy; in Figs. 2(b) and 2(d) $c_{\mathrm{IT}} < c_{\mathrm{KS}}$ and the four phases separated by $c_{\mathrm{IT}} < c_{\mathrm{KS}} < c_{\mathrm{alg}}$ are undetectable, hard, hybrid-hard, and easy. In this terminology we define an undetectable phase by the condition $a^*(c) = a_{\mathrm{alg}}(c) = 0$, an easy phase by $a^*(c) = a_{\mathrm{alg}}(c) > 0$, a hard phase by $a^*(c) > a_{\mathrm{alg}}(c) = 0$, and an hybrid-hard phase by $a^*(c) > a_{\mathrm{alg}}(c) > 0$. The hybrid character of this phase arises from the simultaneous easiness to beat the trivial accuracy of the uninformative estimator [$a_{\mathrm{alg}}(c) > 0$] and hardness of achieving the optimal accuracy [$a^*(c) > a_{\mathrm{alg}}(c)$].

Let us make a few remarks on the classification of possible phase diagrams for inference problems we just presented.

(i) The scenarios described in this section and more generally in this paper are the simplest ones and they do not by far cover all possibilities. One can always devise more complicated free energies $f(a, c)$ with more distinct stationary points
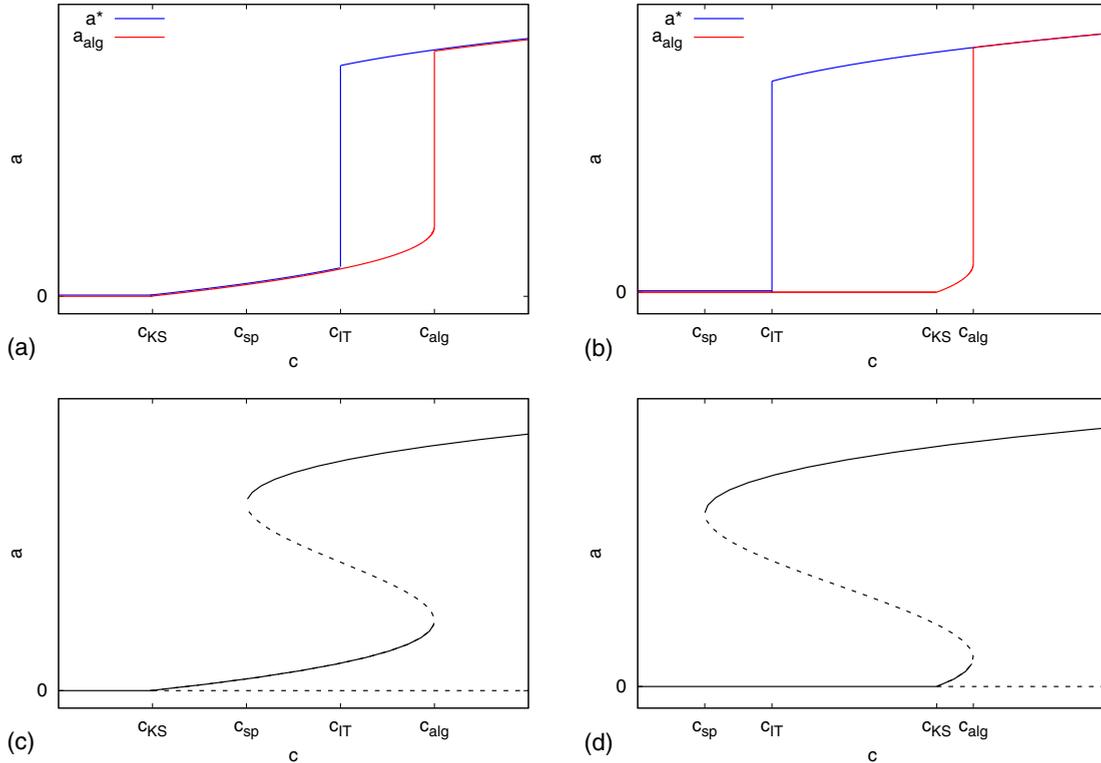
FIG. 2. Two "new" scenarios of phase transitions for inference problems, one in (a) and (c) and one in (b) and (d) for (as in Fig. 1) (a) and (b) the Bayes-optimal (a^*, in blue) and algorithmic (a_{alg}, in red) performance and (c) and (d) the fixed points of Eq. (6), with solid (dashed) lines for stable (unstable) fixed points. In all panels the accuracy a is plotted as a function of the signal-to-noise ratio c. The thresholds c_{sp}, c_{IT}, c_{KS}, and c_{alg}, as defined in the text, are marked on the x axes. The KS threshold c_{KS} here is strictly smaller than the algorithmic one c_{alg}. We note that the thresholds can also be ordered as $c_{\text{sp}} < c_{\text{KS}} < c_{\text{IT}} < c_{\text{alg}}$; in terms of a^* and a_{alg} this still corresponds to the behavior shown in (a).

and/or more bifurcations, yielding, for instance, several hard phases separated by easy regimes; we will not discuss these more complicated situations as they did not arise in any of the cases we analyzed (see [35] for such examples in the related context of error-correcting codes).

(ii) We defined the computational easiness or hardness of inference with respect to one specific efficient algorithm (approximate message passing or belief propagation), as to this day there are no algorithms more efficient (i.e., running in polynomial time) known for the problems that are NP-hard from the worst-case computational complexity point of view. It remains of course a very challenging open question to prove or disprove (under some computational complexity hypothesis) the existence of more accurate, efficient algorithms than these message-passing ones.

(iii) The two simplest scenarios presented in Fig. 1 are well known in the context of inference on sparse random graphs; for instance, the study of the symmetric stochastic block model in Ref. [9] demonstrated a continuous [Figs. 1(a) and 1(c)] phase transition for $q \leqslant 3$ communities and a discontinuous one [Figs. 1(b) and 1(d)] for $q \geqslant 5$. The more complicated phase diagrams of Fig. 2 have only been discussed in the context of dense inference problems (in particular for constrained low-rank matrix estimation; see Figs. 3 and 6 in Ref. [16]). However, we will see in the rest of the paper that they also occur naturally in sparse problems.

(iv) It is known that for inference problems for which $a = 0$ is a fixed point of (6) there is a close relation between

the inference (planted) model and the model with random uncorrelated disorder (see, for instance, [24]). In random optimization and constraint satisfaction problems some of the equilibrium phase transitions described in the literature correspond to phase transitions defined above. Notably, the dynamical phase transition [23], sometimes referred to as the mode-coupling theory transition [36] in the literature on the mean-field theory of structural glasses and referred to as the reconstruction threshold [19] in the theory of reconstruction on tree, corresponds to $c_{\text{d}} = \min(c_{\text{KS}}, c_{\text{sp}})$. This is indeed the smallest SNR for which a nontrivial fixed point does exist. The condensation phase transition from random constraint satisfaction problems [23], sometimes referred to as the Kauzmann ideal glass transition in the mean-field theory of structural glasses [36], corresponds to $c_{\text{cond}} = \min(c_{\text{KS}}, c_{\text{IT}})$. The KS phase transition is referred to as the de Almeida–Thouless instability in the theory of spin glasses [37] and corresponds to the point starting from which iterative message-passing algorithms such as belief propagation cease to converge on random instances of the corresponding problems. This correspondence between planted and random instances does not hold for all phase transitions. For instance, the satisfiability [38] of random constraint satisfaction problems has no counterpart in the planted or inference setting, nor does the algorithmic spinodal $c_{\text{alg}} > c_{\text{KS}}$ of inference problems in the random ensembles. This correspondence also breaks down for models where $a = 0$ is not a fixed point of (6), most notably the satisfiability of random Boolean formulas.

D. Expansions around the trivial fixed point

The classification of the phase diagrams presented above relies on a global bifurcation analysis, which requires the identification of all fixed points of Eq. (6) and the study of their domain of existence as a function of the SNR parameter c. This is relatively easy to do when the order parameter is a scalar or a finite-dimensional object and when the recursion function φ can be studied explicitly. This global analysis becomes much more difficult for sparse inference models, because the order parameter is infinite dimensional in this case. The main methodological contribution of the present paper is a generalization to the sparse case of a local bifurcation analysis of the trivial fixed point in the neighborhood of the Kesten-Stigum transition. For pedagogical reasons, let us first explain here how this analysis is performed in the case of a scalar order parameter (as done, for instance, in Sec. IV C 3 of [16]), which will help one understand the strategy followed in the functional case.

The location of the Kesten-Stigum transition c_{KS}, defined by the condition $\varphi'(0, c_{\mathrm{KS}}) = 1$, was obtained through the linearization of (6). In order to study the nontrivial fixed point in the neighborhood of this transition we will expand the fixed-point equation $a = \varphi(a, c)$ at the next order in a, to obtain

$$a = \varphi'(0, c)a + \tfrac{1}{2}\varphi''(0, c)a^2 + O(a^3). \tag{9}$$

In the neighborhood of c_{KS}, the derivative $\varphi'(0, c)$ is close to 1; we will hence trade the SNR c for a parameter ϵ defined by $\varphi'(0, c) = 1 + \epsilon$ such that $\epsilon = 0$ corresponds to the KS transition and $\epsilon > 0$ ($\epsilon < 0$) to the high-SNR (low-SNR) regime. If the second-order derivative of φ does not vanish exactly at c_{KS} we can rewrite this equation at lowest order as

$$a = a + \epsilon a + va^2 + O(a^3), \tag{10}$$

where we have defined $v = \varphi''(0, c_{\mathrm{KS}})/2 \neq 0$. The nontrivial solution of this equation is obviously $a = -\epsilon/v$; at this point it is crucial to remember the positivity condition $a \geqslant 0$ (without it the bifurcation diagrams would be qualitatively different). Indeed, this requirement implies that if $v < 0$ the nontrivial perturbative fixed point exists for $\epsilon > 0$, in the high-SNR regime; this case corresponds both to the continuous transition [Figs. 1(a) and 1(c)] and to the more complicated phase diagrams of Fig. 2. In contrast, if $v > 0$ it is in the low-SNR regime ($\epsilon < 0$), as in the first-order transition case depicted in Figs. 1(b) and 1(d).

Let us emphasize the limitations of such a local study. By definition, it can only provide perturbative information on the phase diagram, in the neighborhood of the KS transition and for the branch of fixed point that coalesces with the trivial one. If this is enough to distinguish the first-order transition scenario [Figs. 1(b) and 1(d)] and exclude the three other ones when $v > 0$, in the opposite case one cannot decide between the second-order transition [Figs. 1(a) and 1(c)] and the scenarios of Fig. 2 solely from the condition $v < 0$, as nonperturbative (finite a) features further differentiate these cases.

We will develop in the sparse case an expansion that goes one order further, corresponding to an equation of the form

$$a = a + \epsilon a + va^2 + wa^3 + O(a^4). \tag{11}$$

The motivation for continuing the expansion to this order is twofold. On the one hand, there are important inference problems for which $v = 0$ exactly (most notably the symmetric stochastic block model for $q = 4$ groups, both in the sparse [22] and in the dense regimes [16]; we will return to the details of this point in Sec. IV H). In that case the nontrivial solution is $a = \sqrt{-\epsilon/w}$ and it is the sign of w (if $w \neq 0$) that allows us to determine whether the perturbative nontrivial fixed point exists in the low- or high-SNR phase (the argument of the square root must be positive). On the other hand, some problems have, in addition to the SNR, another continuously varying parameter; this is in particular the case of the asymmetric two-group stochastic block model (or asymmetric Ising model in the tree reconstruction language), where the asymmetry between the sizes of the two groups can be tuned independently of the SNR. Depending on this asymmetry parameter, which we denote by $\overline{m}$ in the following, the type of transition changes from second to first order, which comes from a change of sign of v when $\overline{m}$ crosses a critical value $\overline{m}_{\mathrm{c}}$. Having pushed the expansion to the third order will thus allow us to study the neighborhood of the tricritical point $(\epsilon, \overline{m}) = (0, \overline{m}_{\mathrm{c}})$ in parameter space, where the nature of the Kesten-Stigum transition changes. Taking simultaneously the limits $\epsilon \to 0$ and $\overline{m} \to \overline{m}_{\mathrm{c}}$ with well-chosen scalings, the last three terms of (11) will indeed be of the same order (this will be discussed further in Sec. IV I).

III. CAVITY EQUATIONS AND MAIN RESULTS

In the preceding section we discussed the classification of phase transitions and phase diagrams of Bayesian inference problems, relying on the bifurcation analysis for a scalar order parameter; dense inference problems can indeed be reduced to such a scalar (or more generically finite-dimensional) representation in the large-size limit. In the rest of the paper we will turn instead to inference problems defined on sparse graphs, for which the corresponding order parameter becomes functional. In this section we will introduce two exemplary cases of these sparse inference problems, which cover the main applications of this paper, state in a rather generic way the cavity formalism that can be used to handle them, briefly explain its interpretation, and finally present our main results, as obtained in the later sections via an expansion of the cavity equations around their trivial solution.

A. Definitions of two exemplary inference problems on sparse (hyper)graphs

1. Stochastic block model

The stochastic block model is a random graph ensemble which generates networks with a community structure [39–41]. It is specified by the following parameters: N, the number of vertices (or nodes) of the graph $G = (V, E)$; q, the number of communities; $\overline{\eta} = (\overline{\eta}_1, \ldots, \overline{\eta}_q)$, a prior probability distribution on the communities; and $c = \{c_{\sigma\tau}\}$, a $q \times q$ symmetric matrix (not to be confused with the SNR which was also denoted by c in Sec. II). A graph of this ensemble is generated by drawing for each vertex $i \in V$ a label $\sigma_i \in \{1, \ldots, q\}$, independently at random with probability $\overline{\eta}_{\sigma_i}$; this

label represents the community of the vertex. Once the labels $\underline{\sigma} = (\sigma_1, \ldots, \sigma_N)$ are chosen, each of the $N(N-1)/2$ possible edges $\langle i, j \rangle$ between pairs of distinct vertices is included in the set of edges E of the graph with probability $c_{\sigma_i, \sigma_j}/N$. The inference problem we will consider is the determination of the labels $\underline{\sigma}$ (up to a potential symmetry between communities) from the mere observation of the graph G and from the knowledge of the parameters $\overline{\eta}$ and c.

We are in particular interested in the large-size limit $N \to \infty$, taken for a fixed value of the affinity matrix c. In this limit the degree of a vertex labeled σ becomes a Poisson random variable with average $d_\sigma = \sum_{\sigma'} c_{\sigma\sigma'} \overline{\eta}_{\sigma'}$; the finiteness of these degrees justifies the name "sparse" given to this type of inference problem. The phase transitions in the Bayes-optimal inference in the sparse stochastic block model were studied in detail in Refs. [8,9]. Many of these results were confirmed rigorously in subsequent works (see, e.g., [6,42–45]).

We will assume in the following that $d_\sigma = d$ independently of σ in such a way that the degree of a vertex is uninformative of its label. This is the condition for the existence of an undetectable phase at small d, in which case the SBM is asymptotically contiguous to a purely random Erdős-Rényi graph of the same average degree and the optimal estimation of the labels can rely only on the prior distribution $\overline{\eta}$.

2. Planted occupation models

We consider now the family of constraint satisfaction problems (CSPs) in which N Boolean variables, defined by $\underline{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \{0, 1\}^N$, are required to satisfy simultaneously M constraints of the following form: The ath constraint bears on a subset of k variables, defined by $\partial a \subset \{1, \ldots, N\}$, and is satisfied by $\underline{\sigma}$ if and only if $\sum_{i \in \partial a} \sigma_i \in S$, where $S \subset \{0, 1, 2, \ldots, k\}$ defines the type of problem. The interpretation of $\sigma_i = 1$ ($\sigma_i = 0$) as a site occupied by a particle (that is empty) justifies the name of "occupation" model, each constraint restricting the number of particles adjacent to it (in the bipartite graph on $N + M$ vertices with an edge put between i and a if and only if $i \in \partial a$) to belong to S. This family of CSPs contains as special cases the bicoloring of hypergraphs (when $S = \{1, \ldots, k-1\}$) and XORSAT problems (also known as parity checks, when $S = \{0, 2, 4, \ldots, \}$ or $S = \{1, 3, 5, \ldots, \}$). Random ensembles of occupation problems, in which the neighborhoods ∂a are drawn uniformly at random among the $\binom{N}{k}$ possible ones, have been studied, e.g., in Ref. [46].

An inference problem can be associated with these occupation models if instead of this random ensemble of CSPs one considers its planted version, in which one first chooses a configuration $\underline{\sigma}$, drawing independently each σ_i with a prior probability distribution $\overline{\eta}_{\sigma_i}$ (with $\overline{\eta}_0 + \overline{\eta}_1 = 1$), and then draws the M neighborhoods ∂a uniformly among those that are satisfied by $\underline{\sigma}$ (i.e., such that $\sum_{i \in \partial a} \sigma_i \in S$). The inference

problem is then to reconstruct the planted configuration $\underline{\sigma}$ solely from the observation of the bipartite graph linking variables to constraints and the knowledge of the subset S. Special cases of planted occupation problems and their corresponding phase transition have been studied previously in Refs. [11,47]. In this paper we will focus on symmetric planted occupation models, in the sense that $\overline{\eta}_0 = \overline{\eta}_1 = 1/2$ and the set S of the number of allowed particles is invariant under the exchange of occupied and empty sites (i.e., we assume that $n \in S$ if and only if $k - n \in S$); this condition ensures the existence of an undetectable phase if the density of constraints $\alpha = M/N$ is small enough.

B. Cavity equations and free-entropy functional

We will now present a formalism, known as the cavity method [2,48], that allows us to deal with inference problems on sparse random (hyper)graphs. We will first state the approach in a formal way, with a level of generality that encompasses the two examples above, before discussing its interpretation and justification in the next section.

The cavity equations depend on the following parameters: χ, a discrete alphabet of spins of cardinality q, taken for concreteness to be $\chi = \{1, \ldots, q\}$; a probability distribution $\overline{\eta}$ on χ, i.e., $(\overline{\eta}_1, \ldots, \overline{\eta}_q)$, with $\overline{\eta}_\sigma \geqslant 0$ and $\sum_\sigma \overline{\eta}_\sigma = 1$; a probability law p_ℓ on non-negative integers representing the degree distribution of variables and its size-biased version

$$\tilde{p}_\ell = \frac{(\ell+1) p_{\ell+1}}{\mathbb{E}[\ell]}, \tag{12}$$

where $\mathbb{E}[\ell] = \sum_\ell p_\ell \ell$ is the average degree; an integer $k \geqslant 2$; and a joint probability law $p_j(\sigma_1, \ldots, \sigma_k)$ on χ^k (the subscript j standing for joint) that we assume to be invariant under all the permutations of its arguments and to have $\overline{\eta}$ as marginal laws for a single argument

$$p_j(\sigma_1, \ldots, \sigma_k) = p_j(\sigma_{\pi(1)}, \ldots, \sigma_{\pi(k)}) \, \forall \pi \in \mathcal{S}_k,$$
$$\sum_{\sigma_2, \ldots, \sigma_k} p_j(\sigma_1, \ldots, \sigma_k) = \overline{\eta}_{\sigma_1}. \tag{13}$$

We will also need the conditional version of this probability law, to be defined by $p_c(\sigma_1, \ldots, \sigma_{k-1}|\sigma) = p_j(\sigma_1, \ldots, \sigma_{k-1}, \sigma) \frac{1}{\overline{\eta}_\sigma}$, where the subscript c stands for conditional. As a consequence of the hypothesis made on p_j, this conditional law is invariant under the permutations of its $k-1$ first arguments and fulfills the reversibility property:

$$p_c(\sigma_1, \ldots, \sigma_{k-1}|\sigma) \overline{\eta}_\sigma = p_c(\sigma, \sigma_2, \ldots, \sigma_{k-1}|\sigma_1) \overline{\eta}_{\sigma_1}. \tag{14}$$

We denote by η and ν probability distributions on χ (i.e., q-dimensional vectors of positive reals that sum to one). We call the following recursive relations on the $2q$ distributions $\{P_\tau^{(n)}(\eta), \hat{P}_\tau^{(n)}(\nu)\}_{\tau \in \chi}$ the conditional versions of the cavity equations associated with the parameters $(q, \overline{\eta}, p_\ell, k, p_j)$:

$$P_\tau^{(n+1)}(\eta) = \sum_{\ell=0}^{\infty} \tilde{p}_\ell \int d\hat{P}_\tau^{(n)}(\nu^1) \cdots d\hat{P}_\tau^{(n)}(\nu^\ell) \delta(\eta - f(\nu^1, \ldots, \nu^\ell)), \tag{15}$$

$$\hat{P}_\tau^{(n)}(\nu) = \sum_{\tau_1, \ldots, \tau_{k-1}} p_c(\tau_1, \ldots, \tau_{k-1}|\tau) \int dP_{\tau_1}^{(n)}(\eta^1) \cdots dP_{\tau_{k-1}}^{(n)}(\eta^{k-1}) \delta(\nu - \hat{f}(\eta^1, \ldots, \eta^{k-1})). \tag{16}$$

Here and in the following, unspecified summation over spin indices σ or τ are understood to run over χ, and n should be thought of as a time index along the recursion (the initial condition for $n = 0$ will be discussed below). The functions f and $\hat{f}$ appearing above are called belief propagation recursions and are defined as follows: $\eta = f(\nu^1, \ldots, \nu^\ell)$ means

$$\eta_\sigma = \frac{1}{z(\nu^1, \ldots, \nu^\ell)} \overline{\eta}_\sigma \prod_{i=1}^{\ell} \frac{\nu_\sigma^i}{\overline{\nu}_\sigma}, \tag{17}$$

where z enforces the normalization of η and we have denoted by $\overline{\nu}$ the uniform distribution on χ (i.e., $\overline{\nu}_\sigma = \frac{1}{q}$), and the other recursion function $\nu = \hat{f}(\eta^1, \ldots, \eta^{k-1})$ is spelled out as

$$\nu_\sigma = \frac{1}{\hat{z}(\eta^1, \ldots, \eta^{k-1})} \frac{\overline{\nu}_\sigma}{\overline{\eta}_\sigma} \sum_{\sigma_1, \ldots, \sigma_{k-1}} p_{\mathrm{j}}(\sigma, \sigma_1, \ldots, \sigma_{k-1}) \prod_{i=1}^{k-1} \frac{\eta_{\sigma_i}^i}{\overline{\eta}_{\sigma_i}}, \tag{18}$$

with again $\hat{z}$ a normalizing factor. One can check easily that the hypothesis (13) made on the joint probability law p_{j} implies that $f(\overline{\nu}, \ldots, \overline{\nu}) = \overline{\eta}$ and $\hat{f}(\overline{\eta}, \ldots, \overline{\eta}) = \overline{\nu}$, hence that $P_\tau^{(n)}(\eta) = \delta(\eta - \overline{\eta})$ and $\hat{P}_\tau^{(n)}(\nu) = \delta(\nu - \overline{\nu})$ are a stationary solution of the cavity equations (15) and (16), which will be called a trivial fixed point in the following.

Let us also define what we call the unconditional version of the cavity equations associated with the parameters $(q, \overline{\eta}, p_\ell, k, p_{\mathrm{j}})$ that bear on the two sequences of distributions $P^{(n)}(\eta)$ and $\hat{P}^{(n)}(\nu)$,

$$P^{(n+1)}(\eta) = \sum_{\ell=0}^{\infty} \tilde{p}_\ell \int d\hat{P}^{(n)}(\nu^1) \cdots d\hat{P}^{(n)}(\nu^\ell) \delta(\eta - f(\nu^1, \ldots, \nu^\ell)) z(\nu^1, \ldots, \nu^\ell), \tag{19}$$

$$\hat{P}^{(n)}(\nu) = \int dP^{(n)}(\eta^1) \cdots dP^{(n)}(\eta^{k-1}) \delta(\nu - \hat{f}(\eta^1, \ldots, \eta^{k-1})) \hat{z}(\eta^1, \ldots, \eta^{k-1}). \tag{20}$$

The functions $f, z, \hat{f}$, and $\hat{z}$ have been defined in Eqs. (17) and (18) and the distributions $P^{(n)}(\eta)$ and $\hat{P}^{(n)}(\nu)$ are required to have means $\overline{\eta}$ and $\overline{\nu}$, respectively:

$$\int dP^{(n)}(\eta)\eta = \overline{\eta}, \quad \int d\hat{P}^{(n)}(\nu)\nu = \overline{\nu}. \tag{21}$$

With the choice of normalization made in Eqs. (17) and (18) one has $z(\overline{\nu}, \ldots, \overline{\nu}) = \hat{z}(\overline{\eta}, \ldots, \overline{\eta}) = 1$, hence the normalization of the distributions $P^{(n)}$ and $\hat{P}^{(n)}$, as well as the conditions (21) on their averages, are preserved by the recursions of Eqs. (19) and (20). Note that the conditional and unconditional versions of the cavity equations are actually equivalent: One can go from one form to the other according to the relations

$$P^{(n)}(\eta) = \sum_\tau \overline{\eta}_\tau P_\tau^{(n)}(\eta), \quad P_\tau^{(n)}(\eta) = \frac{\eta_\tau}{\overline{\eta}_\tau} P^{(n)}(\eta) \tag{22}$$

for all n and similarly for $\hat{P}^{(n)}$,

$$\hat{P}^{(n)}(\nu) = \sum_\tau \overline{\nu}_\tau \hat{P}_\tau^{(n)}(\nu), \quad \hat{P}_\tau^{(n)}(\nu) = \frac{\nu_\tau}{\overline{\nu}_\tau} \hat{P}^{(n)}(\nu). \tag{23}$$

In particular, the trivial fixed point, in the unconditional version of the cavity equations, corresponds to $P^{(n)}(\eta) = \delta(\eta - \overline{\eta})$ and $\hat{P}^{(n)}(\nu) = \delta(\nu - \overline{\nu})$.

Let us conclude this formal statement of the cavity formalism by the introduction of a functional, called free entropy, that associates, in its unconditional form, a real value with a pair $(P, \hat{P})$ of distributions that satisfy (21) according to

$$\phi(P, \hat{P}) = -\mathbb{E}[\ell] \int dP(\eta) d\hat{P}(\nu) z_{\mathrm{e}}(\eta, \nu) \ln z_{\mathrm{e}}(\eta, \nu) + \frac{\mathbb{E}[\ell]}{k} \int dP(\eta^1) \cdots dP(\eta^k) z_{\mathrm{c}}(\eta^1, \ldots, \eta^k) \ln z_{\mathrm{c}}(\eta^1, \ldots, \eta^k)$$

$$+ \sum_{\ell=1}^{\infty} p_\ell \int d\hat{P}(\nu^1) \cdots d\hat{P}(\nu^\ell) z_{\mathrm{v}}(\nu^1, \ldots, \nu^\ell) \ln z_{\mathrm{v}}(\nu^1, \ldots, \nu^\ell), \tag{24}$$

where

$$z_{\mathrm{e}}(\eta, \nu) = \sum_\sigma \eta_\sigma \frac{\nu_\sigma}{\overline{\nu}_\sigma}, \tag{25}$$

$$z_{\mathrm{c}}(\eta^1, \ldots, \eta^k) = \sum_{\sigma_1, \ldots, \sigma_k} p_{\mathrm{j}}(\sigma_1, \ldots, \sigma_k) \prod_{i=1}^{k} \frac{\eta_{\sigma_i}^i}{\overline{\eta}_{\sigma_i}}, \tag{26}$$

$$z_{\mathrm{v}}(\nu^1, \ldots, \nu^\ell) = \sum_\sigma \overline{\eta}_\sigma \prod_{i=1}^{\ell} \frac{\nu_\sigma^i}{\overline{\nu}_\sigma}. \tag{27}$$

An equivalent form in the conditional formalism reads

$$\phi(\{P_\sigma, \hat{P}_\sigma\}_{\sigma \in \chi}) = -\mathbb{E}[\ell] \sum_\sigma \overline{\eta}_\sigma \int dP_\sigma(\eta) d\hat{P}_\sigma(\nu) \ln z_{\mathrm{e}}(\eta, \nu)$$

$$+ \frac{\mathbb{E}[\ell]}{k} \sum_{\sigma_1, \ldots, \sigma_k} p_{\mathrm{j}}(\sigma_1, \ldots, \sigma_k) \int dP_{\sigma_1}(\eta^1) \cdots dP_{\sigma_k}(\eta^k) \ln z_{\mathrm{c}}(\eta^1, \ldots, \eta^k)$$

$$+ \sum_{\ell=1}^{\infty} p_\ell \sum_\sigma \overline{\eta}_\sigma \int d\hat{P}_\sigma(\nu^1) \cdots d\hat{P}_\sigma(\nu^\ell) \ln z_{\mathrm{v}}(\nu^1, \ldots, \nu^\ell). \tag{28}$$

Note that these expressions of ϕ are variational in the sense that their derivatives with respect to their arguments vanish when the latter are fixed-point solutions of the cavity recursions [i.e., (15) and (16) in the conditional version and (19) and (20) in the unconditional one]. They are however well defined even if their arguments are not solutions of the cavity equations [as long as the condition (21) is satisfied]. This property was recently exploited in Refs. [24,25]; we will return to this point in Sec. IV E. Let us also mention that our choice of normalization leads to $\phi = 0$ on the trivial fixed point of the cavity equations.

C. Interpretation of the cavity equations

The cavity equations written above arise in the context of several slightly different and delicately intertwined problems, in particular in the study of graphical models on random structures, tree reconstruction problems, and inference on sparse graphs. We will not attempt here an exhaustive discussion of these various interpretations and refer the reader to the literature [3,12,19,23,24] for more details on these various perspectives and the connections between them, and instead give a brief description of two interpretations in order to clarify the meaning of our main results to be discussed below.

1. Reconstruction on trees

The first interpretation of the cavity equations we will discuss concerns the tree reconstruction problem [17–22]. Consider a rooted bipartite tree, with two types of vertices, called variable nodes and interaction nodes, the latter all of degree k and the root and the leaves being variable nodes. This tree is used as an information channel in the following way: Each variable node i bears a spin $\sigma_i \in \chi$ and information is sent from the root to the leaves by first drawing the spin of the root vertex with a prior probability $\overline{\eta}$ on χ and then recursively assigning the values of the spins at distance $n = 1, 2, \ldots$ from the root. This last step is done independently for each interaction node a: Denoting by $\sigma_1, \ldots, \sigma_{k-1}$ the $k-1$ variables adjacent to a that are at a distance $n+1$ from the root and by σ the variable at a distance n (which has been assigned in the previous step of the induction), one draws the former conditional on the latter, with a law denoted by $p_{\mathrm{c}}(\sigma_1, \ldots, \sigma_{k-1}|\sigma)$. The tree reconstruction question is now the following: Given the observation of the spins at distance n, is it possible to infer the value of the spin at the root that started this broadcast process? Due to the tree structure of the channel, it is possible to compute easily the posterior probability η of the root given this observation. When the

distance n becomes very large two possibilities can arise, depending on the level of noise in the channel: Either η becomes very close to $\overline{\eta}$ and then all information on the root value is lost or η keeps a trace of the correlation with the root and an inference of its value with a better chance of success than the one allowed by the prior probability $\overline{\eta}$ is possible.

It is not too difficult to convince oneself that the distribution $P_\tau^{(n)}(\eta)$ that solves the cavity equations (15) and (16) is precisely the one described above in this tree reconstruction problem (see, e.g., [19] for a detailed derivation with similar notation), conditional on the root variable being τ in the broadcast process and when the tree used as a channel is a random Galton-Watson tree with offspring degree probability $\tilde{p}_\ell$. In this light the unconditional version of the cavity equations (19) and (20) is seen to describe the distribution of the marginal probability law of the root η, conditional on the observation of the leaves, when the law of the latter is not conditioned on the value of the root in the broadcast, and the relations (22) are mere traductions of the Bayes theorem.

Following this interpretation, the initial condition of the induction on n should be

$$P_\tau^{(n=0)}(\eta) = \delta(\eta - \delta_{\tau, \cdot}), \tag{29}$$

with $\delta_{\tau, \cdot}$ the probability law on χ supported solely on τ, which expresses the perfect knowledge of the variables observed at the leaves of the tree. As mentioned above, the distribution $\delta(\eta - \overline{\eta})$ is a fixed point of the cavity recursions (15) and (16). The tree reconstruction problem can thus be rephrased, in this setting, as determining whether the recursion leads, for $n \to \infty$, the distributions $P_\tau^{(n)}$ towards this trivial fixed point, in which case all information on the root is lost, or towards a nontrivial fixed point that contains more information on the root than the prior probability $\overline{\eta}$.

It is actually useful to consider a more generic initial condition, namely,

$$P_\tau^{(n=0)}(\eta) = \varepsilon \delta(\eta - \delta_{\tau, \cdot}) + (1 - \varepsilon)\delta(\eta - \overline{\eta}), \tag{30}$$

with $\varepsilon \in [0, 1]$. The iterations of (15) nd (16), starting from this initial condition, describe now the inference problem of the tree where the spin value of each leaf is observed with probability ε and kept hidden otherwise. The usual tree reconstruction problem is recovered for $\varepsilon = 1$; the variant known as the robust reconstruction problem [18] corresponds to the limit $\varepsilon \to 0$, taken after the $n \to \infty$ limit, and again asks whether the iterations of the cavity equation return the distributions $P_\tau^{(n)}$ to the trivial fixed point (corresponding to $\varepsilon = 0$) or lead it to a nontrivial one. Robust reconstruction is thus possible if an infinitesimal amount of information on a

far away boundary is amplified in the reconstruction process and yields nonvanishing information on the root. It was shown in Ref. [18] that the threshold for robust reconstruction corresponds to the Kesten-Stigum condition [34]; in other words, the robust reconstruction problem is solvable if and only if the trivial fixed point is locally unstable. This provides a bound on the original reconstruction problem, which is certainly solvable whenever the robust variant is. The converse is not true in general; the tightness of the Kesten-Stigum bound depending on the channel will be a point that our main results will explain.

2. Link with inference problems on graphs

Let us now turn to the second interpretation of the cavity formalism, which corresponds to the inference problems on random graphs described in Sec. III A. For concreteness we concentrate on the stochastic block model example, parametrized by the probability law $\overline{\eta}$ and the affinity matrix c, which verify the condition $d_\sigma = \sum_{\sigma'} c_{\sigma\sigma'} \overline{\eta}_{\sigma'} = d$ independently of the label σ. Consider a graph G drawn from the SBM ensemble, along with its labels $\underline{\sigma}$, and choose uniformly at random one of its vertices. One can then show (see, for instance, Proposition 2 in Ref. [49] for a formal statement) that its local neighborhood of fixed radius n converges in the distribution, in the large-size limit, to a Galton-Watson tree with an offspring distribution $\tilde{p}_\ell$ that is a Poisson law of average d, decorated by labels σ_i that have the law described above from the tree reconstruction perspective: The label of the root is drawn with probability $\overline{\eta}$ and broadcasted along the edges of the tree with the conditional law

$$p_c(\sigma'|\sigma) = \frac{1}{d}c_{\sigma\sigma'}\overline{\eta}_{\sigma'}, \tag{31}$$

which verifies the reversibility property (14). This simple observation unveils a link between the inference problem on a graph and its tree counterpart; of course from a rigorous point of view, the interplay between these two problems is subtler, in particular because there is no explicit observation of any label in the graph problem, contrary to the tree case. Another difference is the (weak) interaction between vertices that are not linked by an edge in the posterior probability law of the graph SBM problem, which is not explicitly present in the cavity formalism but can be traced to the condition (21). The cavity method approach leads to the following conjectures.

(i) When the robust reconstruction tree problem is solvable, the labels of the graph inference problem can be inferred with better accuracy than using only the prior information $\overline{\eta}$, via a polynomial time message-passing algorithm known as belief propagation, whose accuracy is described in the large-size limit by the fixed point of the cavity equations reached through the robust reconstruction initial condition, i.e., with $\varepsilon \to 0$ after $n \to \infty$ in Eq. (30).

(ii) The accuracy of the optimal estimator of the labels (which may require an exponential time to be computed) is described by the fixed point of the cavity equations (15) and (16) that yields the largest possible value of the free entropy (28), the latter being related to the mutual information between $\underline{\sigma}$ and G. In particular, if (15) and (16) only admit the

trivial fixed point, then the hidden labels cannot be inferred better than with their prior distributions.

(iii) No polynomial time algorithm is able to beat the estimation accuracy of the belief propagation algorithm, hence the existence of the hard phase, where according to (ii) it is information-theoretically possible to infer the labels with a nontrivial accuracy, but this is not possible with belief propagation. We stress here that in problems where the hybrid-hard phase (with BP giving a better estimate than random, but still being largely suboptimal) exists, the nonoptimality of BP occurs for a SNR above the Kesten-Stigum threshold.

Parts of points (i) and (ii) have been established rigorously, in particular for the stochastic block model as reviewed in Ref. [6]. The most detailed results have been obtained for the symmetric two-group SBM [42,43]. Point (ii) was established in Refs. [24,25,50] for a large set of models; roughly speaking, these works show that the existence of probability distributions $(P,\hat{P})$ with $\phi(P,\hat{P}) > 0$ implies the possibility of nontrivial inference of the labels in the graph problem (a similar variational principle for the tree reconstruction problem can be found in Ref. [19]). The optimality of belief propagation for a large signal-to-noise ratio (above an unspecified constant multiplying the KS threshold) was established in Ref. [44], but remains an open problem in the low signal-to-noise ratio regime. Point (i) is very much related to Theorem 2.12 in Ref. [44], where an auxiliary algorithm provides the initial condition of BP on a finite graph that emulates the observation of a few labels in the tree robust reconstruction problem.

D. Main results

The main technical results of this paper, to be presented in the subsequent sections, are expansions of the functional cavity equations (15) and (16) and free entropy (28) around their trivial fixed point (that exists because we consider models having an undetectable phase). These computations allow us to generalize the local bifurcation analysis sketched in Sec. II D for a finite-dimensional order parameter, to the functional infinite-dimensional case. When the parameters $(\overline{\eta}, p_j, \tilde{p}_\ell)$ of the models are varied in the neighborhood of the Kesten-Stigum transition, at which the trivial fixed point changes stability, a nontrivial fixed point bifurcates from it continuously. The local bifurcation analysis allows us to determine whether it bifurcates as an unstable fixed point in the direction of lower SNR, in which case the Kesten-Stigum threshold and the algorithmic threshold coincide [Fig. 1(b)] and the hard phase must exist, or the solution bifurcates as a stable fixed point in the direction of higher SNR [Figs. 1(a) or 2]. To be able to further distinguish between the three scenarios in the latter case, we have also solved the fixed-point cavity equations numerically using the population dynamics technique (a presentation of this numerical algorithm and of the subtleties in its implementation can be found in Appendix A). For the convenience of the reader we summarize here our main findings for the stochastic block model and the planted occupation models and defer the detailed technical explanations to Sec. IV and V, respectively. We also show the numerical confirmation that BP run on large single instances of the inference problem indeed follows the behavior derived from the distributional cavity equations.

1. Symmetric SBM: Marginal case of four groups

The symmetric SBM is a special case of the model introduced in Sec. III A 1, where every community has the same size, i.e., $\overline{\eta}_\sigma = 1/q$, and the probability of an edge between two vertices of labels σ and σ' only depends on whether $\sigma = \sigma'$ or not: $c_{\sigma\sigma} = a$ and $c_{\sigma\sigma'} = b$ for $\sigma \neq \sigma'$. The average degree is then $d = [a + (q-1)b]/q$ and it is convenient to parametrize the difference between the intracommunity and intercommunity connectivity by defining $\theta = (a-b)/[a + (q-1)b]$. This parameter is positive in the assortative case ($a > b$) and negative in the disassortative case ($a < b$). As explained in Sec. III C, the study of this problem by the cavity method [8,9] is tightly linked to the tree reconstruction problem for the symmetric Potts model [19,22,51,52], with an offspring distribution of the Galton-Watson tree $\tilde{p}_\ell$ Poissonian of average d; in this context $\theta > 0$ ($\theta < 0$) is called the ferromagnetic (antiferromagnetic) case. The Kesten-Stigum transition is easily seen from the tree reconstruction perspective to occur when $d\theta^2 = 1$, where d is the average of $\tilde{p}_\ell$. This combination of parameters can thus be regarded as a relevant signal-to-noise ratio (keeping in mind the sign of θ that disappears from this definition).

The results of our local bifurcation analysis, performed for an arbitrary distribution $\tilde{p}_\ell$, can be summarized as follows (the details can be found in Sec. IV H): for $q \leqslant 3$, for both ferromagnetic and antiferromagnetic models, and for any $\tilde{p}_\ell$, the Kesten-Stigum transition is continuous [second order, as in Fig. 1(a)]; for $q \geqslant 5$, for both ferromagnetic and antiferromagnetic models, and for any $\tilde{p}_\ell$, the transition is discontinuous [first order; cf. Fig. 1(b)]; and for $q = 4$, the order of the phase transition depends explicitly on the degree distribution and on the ferromagnetic or antiferromagnetic character of the model, through the sign of the quantity

$$w = -\frac{7}{3}\frac{\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]}{\tilde{\mathbb{E}}[\ell]^3} + \left(\frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{\tilde{\mathbb{E}}[\ell]^2}\right)^2$$
$$\times \left(\frac{5}{\tilde{\mathbb{E}}[\ell]-1} - \frac{12}{\mathrm{sgn}(\theta)\sqrt{\tilde{\mathbb{E}}[\ell]}-1}\right), \qquad (32)$$

where $\tilde{\mathbb{E}}$ denotes the average with respect to the offspring degree distribution $\tilde{p}_\ell$. A first-order phase transition is predicted to arise if $w > 0$, while the transition is continuous (second order) for $w < 0$. In the ferromagnetic or assortative case [$\mathrm{sgn}(\theta) = 1$], one can check easily that $w < 0$ for all degree distributions, which yields a continuous transition (one has necessarily $d = \tilde{\mathbb{E}}[\ell] > 1$ for the problem to make sense). In contrast, in the antiferromagnetic or disassortative case, the sign of w changes according to the degree distribution. For instance, when $\tilde{p}_\ell$ is a Poisson law with average d, as in the application to the graph SBM problem, there is a critical degree $d_c \approx 22.2694$ such that $w > 0$ for small degrees $d < d_c$, leading to a first-order phase transition, while for large degrees $d > d_c$ the coefficient w becomes negative and hence the transition is of second-order type.

This problem was extensively studied in the literature from various perspectives, notably tree reconstruction [19,22,51,52], the stochastic block model [8,9], and the Potts model on random graphs [53], and part of this characterization was already known (in particular, the tightness of the KS

transition for $q = 2$ was established in Refs. [51,52], and [22] proved the nontightness of the KS transition for $q \geqslant 5$ for all degrees and its tightness for $q = 3$ and large enough degrees). However, the situation of the $q = 4$ case has remained rather obscure (for instance, Fig. 2 of [53] missed the crossover towards a continuous transition at large degrees), even if it could have been anticipated that the ferromagnetic and antiferromagnetic models must behave in the same way in the large-degree limit (see, for instance, [22,54] and the discussion in Sec. IV F). The explicit condition (32) that characterizes sharply the degree distributions for which the antiferromagnetic four-group Potts model (or disassortative symmetric SBM) has a gap between the information-theoretic performance and the best known algorithmic one is thus one of our main results.

In order to test this prediction we performed a numerical resolution of the cavity equations (see Appendix A for more details) for the $q = 4$ antiferromagnet on Poissonian Galton-Watson trees. In Fig. 3(a) one can see clearly that for low average degrees the spinodal and information-theoretic transitions are below the Kesten-Stigum bound and that they get closer to it as d grows. A simple scaling argument, to be explained in Sec. IV H, predicts that the signal-to-noise ratio $d\theta^2$ of the spinodal and IT transitions reaches the Kesten-Stigum value 1 with a correction scaling as $(d_c - d)^3$ as $d \to d_c^-$. In Fig. 3(b) we thus plotted the same phase diagram with the change of variable $d\theta^2 \to (1 - d\theta^2)^{1/3}$ in such a way that the deviation from the KS bound should vanish linearly when $d \to d_c^-$ in these units. It is fair to say that our numerical data are compatible with our analytical prediction of d_c, without providing an independent confirmation of it. As a matter of fact, the largest degrees for which we could estimate reliably the location of the spinodal and IT transitions are still rather far from our prediction of d_c; it is indeed very difficult to distinguish numerically a continuous transition from a very weakly discontinuous one, and the relatively large exponent 3 in the behavior of the signal-to-noise ratio is unfavorable for this numerical study. As a consequence, we cannot perform a linear fit from the plot of Fig. 3(b); however, the line we drew as a guide to the eye looks like a possible interpolation towards our analytically computed value for d_c.

2. Asymmetric balanced two-group SBM

We have also dedicated specific attention, detailed in Sec. IV I, to the case of the ($q = 2$)-community SBM with different communities sizes ($\overline{\eta}_1 \neq \overline{\eta}_2$) yet with an affinity matrix such that the average degree of the vertices in the two communities is the same (hence the degree of a vertex is uninformative of its label and the model exhibits an undetectable phase at low degrees). We will parametrize this model by a magnetization parameter $\overline{m} \in [0, 1]$ such that the size of the smaller group is $\overline{\eta}_1 = \frac{1-\overline{m}}{2}$, the symmetric case being recovered for $\overline{m} = 0$. With the condition of both groups having average degree d, we are left with one more free parameter, which we label θ and define as $\theta = 1 - \frac{c_{12}}{d}$, where $c_{12} = c_{21}$ is N times the probability that two nodes in different groups are connected by an edge.

The cavity method relates this model to a tree reconstruction problem known as the asymmetric Ising model
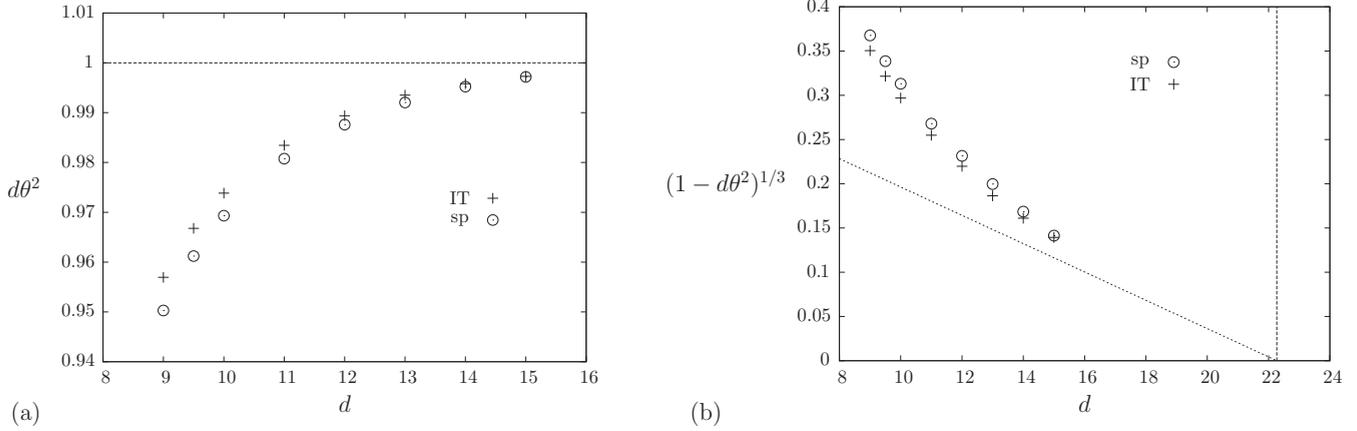
FIG. 3. Phase diagram of the four-group disassortative symmetric SBM (equivalently the reconstruction of the antiferromagnetic $q = 4$ Potts model on Poissonian Galton-Watson trees), as a function of the average degree d. (a) The spinodal and IT transitions (see the text and Fig. 1 for their definitions) are plotted in terms of the signal-to-noise parameter $d\theta^2$, which equals 1 at the Kesten-Stigum transition (horizontal dashed line). (b) Same data as in (a) drawn with a change of variable on the vertical axis to better appreciate the crossover from the first- to the second-order transition as d reaches d_c; as explained in the text and with more details in Sec. IV H, we expect the curves in (b) to vanish linearly when $d \to d_c^-$. The vertical line marks our analytical prediction for $d_c \approx 22.2694$; the other line is only a guide to the eye.

[17,20,29], for which the KS transition occurs at $d\theta^2 = 1$, with $d = \tilde{\mathbb{E}}[\ell]$ the average offspring degree. Our expansions of the cavity equations around this point reveals a striking universality phenomenon: For all degree distributions and for both ferromagnetic and antiferromagnetic models, the qualitative nature of the transition changes when the parameter $\overline{m}$ that quantifies the asymmetry of the model crosses the same value $\overline{m}_c = 1/\sqrt{3}$. The transition is indeed second order [Fig. 1(a)] for $\overline{m} < \overline{m}_c$ and first order [Fig. 1(b)] for $\overline{m} > \overline{m}_c$. Until very recently it was only known from [20] that the Kesten-Stigum bound was tight (i.e., the transition second order) for small enough asymmetries and that it was not tight for large enough asymmetries [17], but these papers did not estimate the critical asymmetry separating these two

regimes. The value $\overline{m}_c = 1/\sqrt{3}$ was deduced independently in Ref. [29], which provides a rigorous proof of the nontightness of the Kesten-Stigum bound for $\overline{m} > \overline{m}_c$ on regular trees, via moment expansions similar to ours. Our (nonrigorous) computations presented in Sec. IV I are more generic, as they encompass arbitrary offspring distributions $\tilde{p}_\ell$, but more importantly have been pushed to a higher order in the expansion. This allows us to predict not only the critical asymmetry $\overline{m}_c$ above which the Kesten-Stigum bound is not tight, but also the leading-order behavior of the spinodal (reconstruction) and IT transition lines in the regime $\overline{m} \to \overline{m}_c^+$ (which is not universal and depends on the degree distribution and on the ferromagnetic or antiferromagnetic character of the model). As an illustration we present in Fig. 4 the phase diagram for
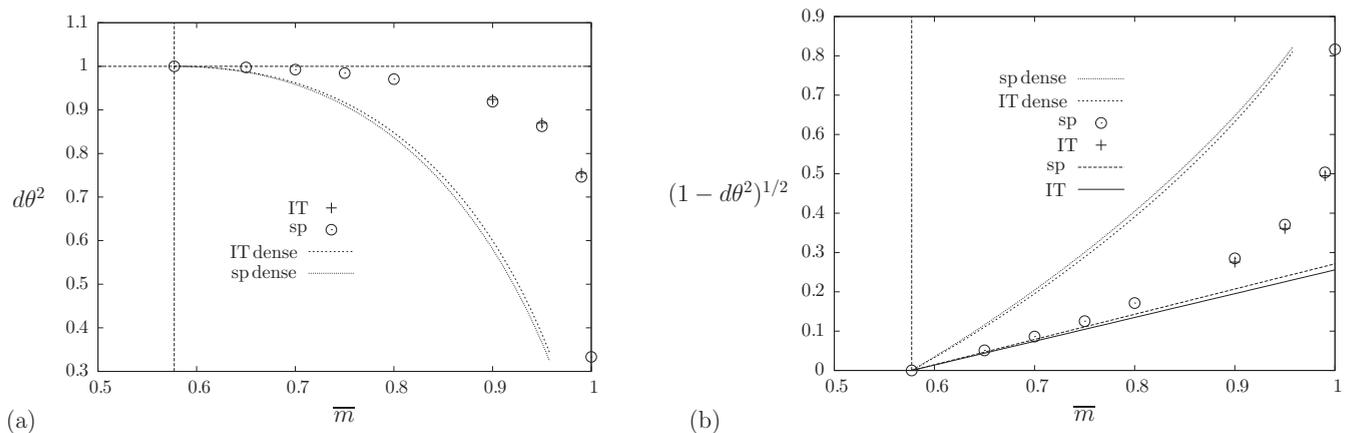
FIG. 4. Phase diagram of the ferromagnetic ($\theta > 0$) asymmetric Ising model on a regular tree with offspring degree distribution $\tilde{p}_\ell = \delta_{\ell,d}$, $d = 3$. The magnetization parameter $\overline{m}$ encodes the size of the smaller community, in the SBM interpretation, as $(1 - \overline{m})/2$. (a) The spinodal and IT transitions are plotted in terms of the signal-to-noise parameter $d\theta^2$, which equals 1 at the Kesten-Stigum transition (horizontal dashed line). Within our numerical accuracy we cannot distinguish the spinodal and IT points for $\overline{m}$ smaller than 0.8; we use a single symbol in this case. The two lines correspond to the large-degree limit, as computed for the dense models in Refs. [16,26,27]. (b) Same data as in (a) drawn in rescaled units to better appreciate the crossover from the first- to the second-order transition as $\overline{m}$ reaches its critical value $\overline{m}_c = 1/\sqrt{3} \approx 0.577$, marked as a vertical line. The two straight lines are the analytical predictions of Eqs. (130) and (133) for the leading-order behavior of the spinodal and IT transitions as $\overline{m} \to \overline{m}_c^+$, for the sparse case $\tilde{p}_\ell = \delta_{\ell,3}$.
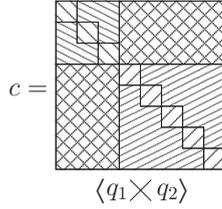
FIG. 5. Shape of the connectivity matrix in the $q_1 + q_2$ Potts model. Matrix elements filled with the same pattern are equal.

TABLE I. Thresholds for the bicoloring on Poissonian random hypergraphs; see the text and Fig. 2 for their definitions.

k	$\alpha_{\rm KS}$	$\alpha_{\rm sp}$	$\alpha_{\rm IT}$	$\alpha_{\rm alg}$
3	1.5			
4	4.083	4.64	4.673	4.71
5	11.25	9.46	10.296	12.44
6	32.03	18.08	21.425	34.5

the ferromagnetic Ising model on a regular tree, obtained by a numerical resolution of the cavity equations. One can indeed see in Fig. 4(a) that the spinodal and information-theoretic transitions occur before the KS threshold for asymmetries larger than $\overline{m}_c$ and that this discontinuity of the transition vanishes when $\overline{m} \to \overline{m}_c^+$. In Fig. 4(b) we present the same data in rescaled units, chosen such that the approach to $\overline{m}_c$ is linear; the two straight lines of the plot are the results of our analytical predictions to be explained in Sec. IV I and are in agreement with the numerical results within the accuracy we could reach.

Note that the independence of $\overline{m}_c$ on the degree distribution explains why the same phenomenology, as well as the same critical asymmetry, was already obtained in the corresponding dense inference problems [16,26] or in the large-degree limit [27,28]. The community detection problem in networks with unequal groups was also recently investigated in Refs. [55,56], but these works treat the case where the degree of a vertex is correlated to its label; in contrast, in all the problems we treat in this paper the degree of a vertex is not informative of its label.

3. The $q_1 + q_2$ SBM

As we have seen in the two cases above, the qualitative properties of the phase diagram of the SBM strongly depend on the structure of the prior distribution $\overline{\eta}$ and of the affinity matrix c, in particular on their symmetry properties. It is of course impossible to investigate explicitly all the possible ways to break the full permutation symmetry between the q labels, for arbitrary q, and to determine the typology of the phase transitions of the associated models. We have studied a class of SBMs that generalizes and encompasses both the symmetric SBM model for q arbitrary and the asymmetric $q = 2$ case, by dividing the q possible labels in two supergroups G_1 and G_2 of cardinality q_1 and q_2 (with $q = q_1 + q_2$) and keeping the permutation symmetry inside each of these two sets. This leads to a prior $\overline{\eta}_\sigma$ that is constant inside G_1 and G_2, and an affinity matrix $c_{\sigma\tau}$ which depends only on the super-groups σ and τ belong to and on whether or not $\sigma = \tau$ (see Fig. 5 for an illustration). One can view the graph generated in this way as a superposition of two symmetric SBMs on the vertices in the groups G_1 and G_2 and an asymmetric Ising model that coarse grains all the labels in a group G_i as a single symbol. Once q_1 and q_2 are chosen the model is defined by five parameters: the average degree of a vertex (assumed to be independent of its label), the fraction of vertices in the two supergroups, and the three SNRs quantifying the information provided by the edges of each SBM in the decomposition

explained above. Depending on the choice of parameters, the Kesten-Stigum transition can arise from any of these three subgraphs; in Sec. IV J we present a classification of the typology of the corresponding bifurcation. Moreover, we will argue there that for well-chosen parameters the more complicated bifurcation depicted in Fig. 2 must occur in this model. Note that a special case of this problem (with an additional symmetry between the two subgroups) was recently studied in Rfes. [30,57].

4. Hybrid-hard phase in planted occupation models

In Sec. V we will study in detail the cavity equations for Ising spin ($q = 2$) variables, defining the spin alphabet $\chi = \{-1, +1\}$, with k-wise interactions, for $k \geqslant 2$ arbitrary. We concentrate on models that have a global spin-flip symmetry, i.e., such that $\overline{\eta}_+ = \overline{\eta}_- = 1/2$ and $p_{\rm j}(\sigma_1, \ldots, \sigma_k) = p_{\rm j}(-\sigma_1, \ldots, -\sigma_k)$ in the notation of Sec. III B. Quite strikingly, we will prove that all the problems fulfilling this symmetry property have, around their Kesten-Stigum threshold, a bifurcating nontrivial solution on the large signal-to-noise ratio side of the KS transition. In other words, none of these models can exhibit a first-order transition as in Figs. 1(b) and 1(d); depending on the model, the phase diagram either is of the second-order type [cf. Figs. 1(a) and 1(c)] or exhibits an hybrid-hard phase as in Fig. 2.

As an illustrative example we present in Fig. 6 the bifurcation diagrams for the planted k-uniform hypergraph bicoloring, as defined in Sec. III A 2, in which each hyperedge forbids the k variables around it from being all in the same state. The random version of this constraint satisfaction problem has been studied in some details (see, for instance, [58] and references therein), but we are not aware of previous studies of its planted version for small values of k. The plots of Fig. 6 show that for $k = 3$ the transition is second order, while $k = 4$ and $k = 5$ realize the two scenarios sketched in Figs. 2(a) and 2(c) and Figs. 2(b) and 2(d), respectively, whose interpretation was discussed in general terms in Sec. II C. In these latter two cases the hybrid-hard phase occupies a sizable part of the phase diagram and can be easily identified numerically; we expect all $k \geqslant 5$ to behave qualitatively in the same way and checked this for $k = 6$. We report in Table I the corresponding values of the thresholds $\alpha_{\rm KS}$ for the Kesten-Stigum transition, $\alpha_{\rm sp}$ and $\alpha_{\rm alg}$ for the spinodals of the high- and low-accuracy branches, respectively, and $\alpha_{\rm IT}$, corresponding to the crossings of the free entropies of these two fixed points (the corresponding numerical data will be shown in Sec. V).

The hybrid-hard phase in which belief propagation is able to reach a nontrivial accuracy, yet cannot reach the optimal one, has not been identified in sparse inference problems that
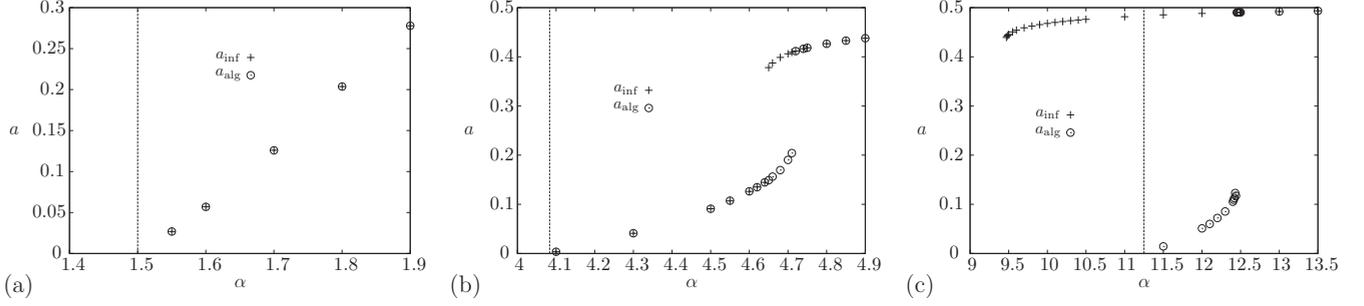
FIG. 6. Stable fixed points of the cavity equations in terms of their accuracies for the planted bicoloring of k-uniform hypergraphs for (a) $k = 3$, (b) $k = 4$, and (c) $k = 5$. The accuracy is defined here as the difference between the probability that a sample from the posterior marginal outputs the correct label and the same probability when only the prior is used (see Sec. IV K for more details). The vertical line marks the KS threshold. These curves have been obtained by a numerical resolution of the cavity equations (15) and (16), a_{inf} corresponds to the fixed point reached from an informative initialization [corresponding to the reconstruction problem, $\varepsilon = 1$ in Eq. (30)], and a_{alg} was obtained from an uninformative initialization [corresponding to the robust reconstruction problem, $\varepsilon \to 0$ in Eq. (30)]. The parameter α controls the offspring degree distribution $\tilde{p}_\ell$, which is a Poisson law of average αk. The optimal accuracy a^* coincides with a_{inf} above the IT transition, whose location can be found in Table I.

have an undetectable phase. This can sound surprising in view of our statement above: Hybrid-hard phases appear as soon as the phase diagram is not a purely second-order transition. The resolution of this apparent paradox is that the hybrid-hard phase is often very narrow: For instance, the so-called 2-in-4 planted SAT problem was reported, wrongly, to have a first-order transition in Refs. [11,47]. As we show in Sec. V, the hybrid-hard phase does exist in this model, but its width is so tiny that it was easily missed in previous numerical studies.

5. Numerical experiments with belief propagation

Our main technical results are based on numerical resolutions and analytic expansion of the distributional cavity equations (19) and (20). These equations describe directly the tree reconstruction problem, but their interpretations in terms of the inference problems on graphs and hypergraphs relies on conjectured connections summarized in Sec. III C 2. We therefore devote Sec. VI to a numerical test of this connection and show that the behavior of the BP algorithm on large but finite-size samples is well described by the (infinite-size) tree reconstruction problems. As a striking example let us anticipate that the curves of Fig. 6 will be reproduced in Fig. 10 from the output of single-sample experiments, including the coexistence of nontrivial branches in the hybrid-hard phases.

IV. MOMENT EXPANSIONS FOR PAIRWISE INTERACTING POTTS VARIABLES

This section is devoted to the first of the two specializations of the formalism introduced in Sec. III: We will only study here pairwise interacting models (i.e., $k = 2$). In other words, we consider graphs instead of hypergraphs, but keep an arbitrary q-state variable alphabet (in Sec. V we will instead turn to k-wise interactions with arbitrary k but restrict ourselves to Ising spin variables).

A. Cavity equations and free-entropy functional

Let us start by rewriting the cavity equations and free-entropy functional introduced in Sec. III in a simplified form, exploiting the pairwise character of the interactions considered here. Our problem is now defined in terms of a probability distribution $\bar{\eta}$ on $\chi = \{1, \dots, q\}$, a $q \times q$ transition matrix $M_{\sigma\sigma'} = p_c(\sigma'|\sigma)$, and the degree distributions p_ℓ and $\tilde{p}_\ell$ related by the size bias (12). The two equations (15) and (16) can be joined into a single recursion on the conditional distribution

$$P_\tau^{(n+1)}(\eta) = \sum_{\ell=0}^{\infty} \tilde{p}_\ell \sum_{\tau_1, \dots, \tau_\ell} M_{\tau\tau_1} \cdots M_{\tau\tau_\ell} \int dP_{\tau_1}^{(n)}(\eta^1) \cdots$$
$$dP_{\tau_\ell}^{(n)}(\eta^\ell) \delta(\eta - f(\eta^1, \dots, \eta^\ell)), \quad (33)$$

where the belief propagation recursion function f is obtained by concatenating (17) and (18). In an explicit form, $\eta = f(\eta^1, \dots, \eta^\ell)$ means that for all labels $\sigma \in \{1, \dots, q\}$,

$$\eta_\sigma = \frac{z_\sigma(\eta^1, \dots, \eta^\ell)}{z(\eta^1, \dots, \eta^\ell)},$$
$$z(\eta^1, \dots, \eta^\ell) = \sum_\gamma z_\gamma(\eta^1, \dots, \eta^\ell), \quad (34)$$
$$z_\sigma(\eta^1, \dots, \eta^\ell) = \bar{\eta}_\sigma \prod_{i=1}^{\ell} \sum_{\sigma'} \hat{M}_{\sigma\sigma'} \eta_{\sigma'}^i,$$

where we have introduced the notation

$$\hat{M}_{\sigma\sigma'} = M_{\sigma\sigma'} \frac{1}{\bar{\eta}_{\sigma'}}. \quad (35)$$

An equivalent and more compact form of (33) can be written as an equality in the distribution of random variables

$$\eta^{(n+1,\tau)} \overset{d}{=} f(\eta^{(n,\tau_1)}, \dots, \eta^{(n,\tau_\ell)}), \quad (36)$$

where all random variables are independent, ℓ has the distribution $\tilde{p}_\ell$, $\eta^{(n,\tau)}$ stands for a random draw from $P_\tau^{(n)}$, and the τ_i on the right-hand side are generated with probability $M_{\tau\tau_i}$. From Eqs. (19) and (20) one obtains the recursion for

the unconditional distribution

$$P^{(n+1)}(\eta) = \sum_{\ell=0}^{\infty} \tilde{p}_\ell \int dP^{(n)}(\eta^1) \cdots$$
$$dP^{(n)}(\eta^\ell)\delta(\eta - f(\eta^1, \ldots, \eta^\ell))z(\eta^1, \ldots, \eta^\ell), \tag{37}$$

where z is the quantity defined in Eq. (34).

We gave in Sec. III an expression of the free entropy for generic random factor graph models. Specializing it to the case of pairwise interactions, one obtains

$$\phi(\{P_\tau\}) = \sum_{\ell=1}^{\infty} p_\ell \sum_{\tau} \overline{\eta}_\tau \sum_{\tau_1, \ldots, \tau_\ell} M_{\tau\tau_1} \cdots$$
$$M_{\tau\tau_\ell} \int dP_{\tau_1}(\eta^1) \cdots dP_{\tau_\ell}(\eta^\ell) \ln z_v(\eta^1, \ldots, \eta^\ell)$$
$$- \frac{1}{2}\mathbb{E}[\ell] \sum_{\tau_1, \tau_2} \overline{\eta}_{\tau_1} M_{\tau_1\tau_2} \int dP_{\tau_1}(\eta^1) dP_{\tau_2}(\eta^2)$$
$$\times \ln z_e(\eta^1, \eta^2), \tag{38}$$

where $\mathbb{E}[\ell] = \sum_\ell p_\ell \ell$ is the average degree. The variable contribution z_v reads exactly as z in Eq. (34), while the edge contribution is

$$z_e(\eta^1, \eta^2) = \sum_{\sigma_1, \sigma_2} \eta^1_{\sigma_1} \hat{M}_{\sigma_1\sigma_2} \eta^2_{\sigma_2}. \tag{39}$$

An equivalent form in terms of the unconditional distribution is

$$\phi(P) = \sum_{\ell=1}^{\infty} p_\ell \int dP(\eta^1) \cdots dP(\eta^\ell) z_v(\eta^1, \ldots, \eta^\ell)$$
$$\times \ln z_v(\eta^1, \ldots, \eta^\ell) - \frac{1}{2}\mathbb{E}[\ell] \int dP(\eta^1) dP(\eta^2)$$
$$\times z_e(\eta^1, \eta^2) \ln z_e(\eta^1, \eta^2). \tag{40}$$

B. Properties of the Markov matrix M

Let us review the properties of the transition matrix $M_{\sigma\sigma'} = p_c(\sigma'|\sigma)$ that follows from the hypothesis on p_c made in Sec. III.

(i) The normalization of p_c, which ensures the conservation of probabilities in the broadcast process, implies that M is a stochastic matrix $\sum_{\sigma'} M_{\sigma\sigma'} = 1$ for all σ.

(ii) The assumption (14) yields the reversibility of M with respect to $\overline{\eta}$ (the detailed balance condition in physics jargon), $\overline{\eta}_\sigma M_{\sigma\sigma'} = \overline{\eta}_{\sigma'} M_{\sigma'\sigma}$.

(iii) This reversibility is a sufficient condition for $\overline{\eta}$ to be stationary for the Markov chain with transition probability matrix M, i.e., $\sum_\sigma \overline{\eta}_\sigma M_{\sigma\sigma'} = \overline{\eta}_{\sigma'}$ for all σ'. Hence M has an eigenvalue $\theta_1 = 1$, with right eigenvector $(1, \ldots, 1)$ and left eigenvector $\overline{\eta}$.

Assuming that M is irreducible and aperiodic, the Perron-Frobenius theorem ensures that all its other eigenvalues are strictly smaller in absolute value; we will order them as $1 > |\theta_2| \geqslant \cdots \geqslant |\theta_q|$. The reversibility assumption simplifies the study of the eigendecomposition of M: Consider indeed the matrix $W_{\sigma\sigma'} = \overline{\eta}_\sigma^{1/2} M_{\sigma\sigma'} \overline{\eta}_{\sigma'}^{-1/2}$. It can be easily seen that W is real and symmetric, hence diagonalizable in an orthonormal

basis, and that M has the same eigenvalues as W. Let us write the eigendecomposition of W as

$$W_{\sigma\sigma'} = \sum_{j=1}^{q} \theta_j v_\sigma^{(j)} v_{\sigma'}^{(j)}, \tag{41}$$

where the vectors v of the orthonormal basis verify

$$\sum_\sigma v_\sigma^{(j)} v_\sigma^{(k)} = \delta_{j,k}, \quad \sum_j v_\sigma^{(j)} v_{\sigma'}^{(j)} = \delta_{\sigma,\sigma'}. \tag{42}$$

Here and in the following $\delta_{j,k}$, the Kronecker symbol, is equal to 1 if the two arguments are equal and 0 otherwise. The Perron eigenvector of W reads in this basis

$$\theta_1 = 1, \quad v_\sigma^{(1)} = \overline{\eta}_\sigma^{1/2}. \tag{43}$$

Transforming back from W to M, we introduce the basis of the left and right eigenvectors of M,

$$r_\sigma^{(j)} = \overline{\eta}_\sigma^{-1/2} v_\sigma^{(j)}, \quad l_\sigma^{(j)} = \overline{\eta}_\sigma^{1/2} v_\sigma^{(j)}, \tag{44}$$

which obey the following relations:

$$\sum_\sigma l_\sigma^{(j)} r_\sigma^{(k)} = \delta_{j,k}, \quad \sum_j l_\sigma^{(j)} r_{\sigma'}^{(j)} = \delta_{\sigma,\sigma'},$$
$$r_\sigma^{(j)} = \frac{1}{\overline{\eta}_\sigma} l_\sigma^{(j)}, \quad l_\sigma^{(1)} = \overline{\eta}_\sigma, \quad r_\sigma^{(1)} = 1. \tag{45}$$

The matrix M can thus be expressed as

$$M_{\sigma\sigma'} = \sum_{j=1}^{q} \theta_j r_\sigma^{(j)} l_{\sigma'}^{(j)} = \overline{\eta}_{\sigma'} + \sum_{j=2}^{q} \theta_j r_\sigma^{(j)} l_{\sigma'}^{(j)}. \tag{46}$$

The matrix $\hat{M}$ introduced in Eq. (35) is symmetric due to the reversibility condition and it can be expressed in terms of the eigenvector decomposition as

$$\hat{M}_{\sigma\sigma'} = \sum_{j=1}^{q} \theta_j r_\sigma^{(j)} r_{\sigma'}^{(j)} = 1 + \sum_{j=2}^{q} \theta_j r_\sigma^{(j)} r_{\sigma'}^{(j)}. \tag{47}$$

C. Symmetry properties

Let us recall that we stated in Eq. (22) some consequences of the Bayes theorem that bridge the conditional and unconditional distributions. These will play an important role in the following, hence we first spell out some consequences of these symmetry relations. We will denote by $\mathbb{E}_\tau^{(n)}$ and $\mathbb{E}^{(n)}$ the averages with respect to $P_\tau^{(n)}$ and $P^{(n)}$, respectively. One has $\mathbb{E}^{(n)}[\eta] = \overline{\eta}$ for all n. For any function g, the relationship between the densities of $P_\tau^{(n)}$ and $P^{(n)}$ given in Eq. (22) yields

$$\overline{\eta}_\tau \mathbb{E}_\tau^{(n)}[g(\eta)] = \mathbb{E}^{(n)}[\eta_\tau g(\eta)]. \tag{48}$$

Applying this identity with $g(\eta) = 1$ gives the already mentioned property $\mathbb{E}^{(n)}[\eta] = \overline{\eta}$. If one uses instead $g(\eta) = \eta_\sigma$, this yields

$$\overline{\eta}_\tau \mathbb{E}_\tau^{(n)}[\eta_\sigma] = \mathbb{E}^{(n)}[\eta_\tau \eta_\sigma]. \tag{49}$$

Let us define a q-dimensional vector δ as the difference $\delta = \eta - \overline{\eta}$ between a message and the stationary value. One has in this way $\mathbb{E}^{(n)}[\delta] = 0$, and from (49) one obtains

$$\overline{\eta}_\tau \mathbb{E}_\tau^{(n)}[\delta_\sigma] = \mathbb{E}^{(n)}[\eta_\tau \eta_\sigma] - \overline{\eta}_\tau \overline{\eta}_\sigma = \mathbb{E}^{(n)}[\delta_\tau \delta_\sigma]. \tag{50}$$

D. Stability analysis of the trivial fixed point via moment expansions

By construction, the probability distribution $\overline{\eta}$ is a fixed point of the BP equation (34), i.e., $f(\overline{\eta}, \ldots, \overline{\eta}) = \overline{\eta}$, reflecting the stationarity of $\overline{\eta}$ under the Markov chain generated by M. This implies that $P_\tau(\eta) = \delta(\eta - \overline{\eta})$ is a fixed point of the cavity recursion (33) [and similarly $P(\eta) = \delta(\eta - \overline{\eta})$ is invariant under the recursion (37) for the unconditional distributions]. We will now study the stability of this trivial fixed point and determine perturbatively the nontrivial fixed points that arise in the neighborhood of the limit of stability. This bifurcation analysis will follow the steps explained in Sec. II D for scalar recursions. The additional complication due to the functional character of the recursions (33) and (37) will be dealt with by concentrating on the low-order moments of the distributions P_τ, thereby reducing the infinite-dimensional bifurcation analysis to a finite-dimensional one.

1. Linear analysis

We define by $\delta = \eta - \overline{\eta}$ the (q-dimensional vector) difference between the BP message and the trivial fixed point. As both η and $\overline{\eta}$ are normalized one has $\sum_\sigma \delta_\sigma = 0$. The BP equation (34) can be expressed as

$$\delta_\sigma = \frac{\overline{\eta}_\sigma \prod_{i=1}^\ell \left(1 + \hat{\delta}_\sigma^i\right)}{\sum_\gamma \overline{\eta}_\gamma \prod_{i=1}^\ell \left(1 + \hat{\delta}_\gamma^i\right)} - \overline{\eta}_\sigma, \tag{51}$$

where we defined

$$\hat{\delta}_\sigma \equiv \sum_{\sigma'} \hat{M}_{\sigma\sigma'} \delta_{\sigma'} \tag{52}$$

as a linearly transformed version of δ. The normalization condition becomes, under this transformation, $\sum_\sigma \overline{\eta}_\sigma \hat{\delta}_\sigma = 0$. For this reason, the denominator of (51) is equal to 1 up to quadratic corrections in δ and hence the linearization of the BP equation gives

$$\delta_\sigma \approx \overline{\eta}_\sigma \sum_{i=1}^\ell \hat{\delta}_\sigma^i. \tag{53}$$

Taking averages according to Eq. (33), we thus obtain the evolution of the first moments of the distributions $P_\tau^{(n)}$, within this linearized approximation, as

$$\mathbb{E}_\tau^{(n+1)}[\delta_\sigma] = \tilde{\mathbb{E}}[\ell] \overline{\eta}_\sigma \sum_{\tau'} M_{\tau\tau'} \mathbb{E}_{\tau'}^{(n)}[\hat{\delta}_\sigma], \tag{54}$$

where $\tilde{\mathbb{E}}[\ell] = \sum_\ell \tilde{p}_\ell \ell$ is the average offspring degree; note that in Eq. (54), and in many equations that will follow, we keep implicit higher-order correction terms on the right-hand side for the sake of readability. To put this evolution equation in a more convenient form let us first exploit the moment identities to write

$$\sum_{\tau'} M_{\tau\tau'} \mathbb{E}_{\tau'}^{(n)}[\hat{\delta}_\sigma] = \sum_{\tau'} \hat{M}_{\tau\tau'} \overline{\eta}_{\tau'} \mathbb{E}_{\tau'}^{(n)}[\hat{\delta}_\sigma]$$

$$= \sum_{\tau'} \hat{M}_{\tau\tau'} \mathbb{E}^{(n)}[\hat{\delta}_\sigma \eta_{\tau'}], \tag{55}$$

where in the first step we used the definition (47) of $\hat{M}$ and in the second one we exploited the relation (48) between condi-

tional and unconditional averages. Noting that $\sum_{\tau'} \hat{M}_{\tau\tau'} \overline{\eta}_{\tau'} = 1$ and that $\mathbb{E}^{(n)}[\hat{\delta}_\sigma] = 0$, we obtain the identity

$$\sum_{\tau'} M_{\tau\tau'} \mathbb{E}_{\tau'}^{(n)}[\hat{\delta}_\sigma] = \mathbb{E}^{(n)}[\hat{\delta}_\sigma \hat{\delta}_\tau]. \tag{56}$$

Multiplying the linearized evolution equation (54) by $\overline{\eta}_\tau$ and using (50) to transform the left-hand side and (56) to treat the right-hand side, we obtain, without making further approximations,

$$\mathbb{E}^{(n+1)}[\delta_\sigma \delta_\tau] = \tilde{\mathbb{E}}[\ell] \overline{\eta}_\sigma \overline{\eta}_\tau \mathbb{E}^{(n)}[\hat{\delta}_\sigma \hat{\delta}_\tau]. \tag{57}$$

We define $A_{\sigma\tau}^{(n)} = \mathbb{E}^{(n)}[\delta_\sigma \delta_\tau]$; it can be viewed as a matrix $A^{(n)}$ of size $q \times q$, which is symmetric positive semidefinite with vanishing row sums. We also define

$$\hat{A}_{\sigma\tau}^{(n)} = \mathbb{E}^{(n)}[\hat{\delta}_\sigma \hat{\delta}_\tau] = \sum_{\sigma',\tau'} \hat{M}_{\sigma\sigma'} \hat{M}_{\tau\tau'} A_{\sigma'\tau'}^{(n)}, \tag{58}$$

which defines a matrix $\hat{A}$, symmetric positive semidefinite with $\sum_\sigma \overline{\eta}_\sigma \hat{A}_{\sigma\tau}^{(n)} = 0$. With this notation, the linearized recursion (57) becomes

$$A_{\sigma\tau}^{(n+1)} = \tilde{\mathbb{E}}[\ell] \overline{\eta}_\sigma \overline{\eta}_\tau \hat{A}_{\sigma\tau}^{(n)}, \tag{59}$$

which, considering now A as a q^2-dimensional vector, can be viewed as a matrix multiplication

$$A^{(n+1)} = NA^{(n)},$$

$$N_{\sigma\tau,\sigma'\tau'} = \tilde{\mathbb{E}}[\ell] \overline{\eta}_\sigma \overline{\eta}_\tau \hat{M}_{\sigma\sigma'} \hat{M}_{\tau\tau'}$$

$$= \tilde{\mathbb{E}}[\ell] \overline{\eta}_\sigma M_{\sigma\sigma'} \frac{1}{\overline{\eta}_{\sigma'}} \overline{\eta}_\tau M_{\tau\tau'} \frac{1}{\overline{\eta}_{\tau'}}. \tag{60}$$

In this equation N is a $q^2 \times q^2$ symmetric matrix; its eigenvalues are easily computed, exploiting its tensor product form, and seen to be $\{\tilde{\mathbb{E}}[\ell] \theta_i \theta_j\}_{i,j=1,\ldots,q}$, where the θ are the eigenvalues of M (see Sec. IV B). As $A^{(n)}$ must obey the normalization condition $\sum_\sigma A_{\sigma\tau}^{(n)} = 0$, the relevant eigenvalues of N are only those with $i, j \geqslant 2$. The stability of $A = 0$ under its multiplication by N is equivalent to the largest absolute value among the (relevant) eigenvalues of N being smaller than 1: We have thus recovered, as expected, the well-known Kesten-Stigum [34] condition $\tilde{\mathbb{E}}[\ell] \theta_2^2 < 1$ for the trivial fixed point $P(\eta) = \delta(\eta - \overline{\eta})$ to be stable under small perturbations.

2. Second-order expansion

Our goal now is to characterize the nontrivial fixed points of the functional recursions (33) and (37) that bifurcate continuously from the trivial one at the Kesten-Stigum transition; we will hence drop the iteration index n in the following and denote by $\mathbb{E}_\tau[\bullet]$ and $\mathbb{E}[\bullet]$ the averages with respect to the P_τ and P fixed-point solutions of (33) and (37), respectively. It is rather intuitive, and can be easily seen in the scalar case treated in Sec. II D, that this task will rely on an expansion of (51) beyond the linear order. However, the functional character of the fixed-point equation complicates the determination of the relevant nonlinear terms in this expansion. In other words, one has to specify in which precise sense the sought fixed point is close to the trivial one. Let us consider the unconditional distribution $P(\eta)$ and discuss the relative scaling of its centered moments. By definition, the first one

$\mathbb{E}[\delta_\sigma]$ must vanish; the second one $A_{\sigma\tau} = \mathbb{E}[\delta_\sigma \delta_\tau]$ must be nonzero for the distribution to be nontrivial, and we assume

that it is small, of the order of some parameter $\kappa \ll 1$. We will make the following ansatz for the higher-order moments:

$$
\begin{aligned}
B_{\sigma\tau\gamma} &= \mathbb{E}[\delta_\sigma \delta_\tau \delta_\gamma] = O(\kappa^2), \\
C_{\sigma\tau\gamma\beta} &= \mathbb{E}[\delta_\sigma \delta_\tau \delta_\gamma \delta_\beta] = O(\kappa^2), \\
&\vdots \\
\mathbb{E}[\delta_{\sigma_1}, \dots, \delta_{\sigma_p}] &= O(\kappa^{\lceil p/2 \rceil}).
\end{aligned}
\tag{61}
$$

This ansatz was inspired by the numerical resolution of the functional equations and we refer the reader to the Appendix B for a proof of its self-consistency.

We will now proceed with these assumptions and compute the leading behavior of $A_{\sigma\tau}$ by considering the following truncated expansion of (51):

$$
\delta_\sigma = \overline{\eta}_\sigma \sum_{i=1}^{\ell} \hat{\delta}_\sigma^i + \overline{\eta}_\sigma \sum_{1 \leqslant i < j \leqslant \ell} \left[\hat{\delta}_\sigma^i \hat{\delta}_\sigma^j - \sum_\gamma \overline{\eta}_\gamma \hat{\delta}_\gamma^i \hat{\delta}_\gamma^j - \sum_\gamma \overline{\eta}_\gamma (\hat{\delta}_\sigma^i + \hat{\delta}_\sigma^j) \hat{\delta}_\gamma^i \hat{\delta}_\gamma^j - \sum_\gamma \overline{\eta}_\gamma \hat{\delta}_\sigma^i \hat{\delta}_\sigma^j \hat{\delta}_\gamma^i \hat{\delta}_\gamma^j + \sum_{\gamma,\beta} \overline{\eta}_\gamma \overline{\eta}_\beta \hat{\delta}_\gamma^i \hat{\delta}_\beta^i \hat{\delta}_\gamma^j \hat{\delta}_\beta^j \right]. \tag{62}
$$

We have included nonlinear terms that are of order 2, 3, and 4 in terms of δ, but that will all be of order κ^2 once averaged, according to the ansatz (61). Taking averages with respect to the fixed-point conditional distributions yields, without further approximations,

$$
\mathbb{E}_\tau[\delta_\sigma] = \tilde{\mathbb{E}}[\ell] \overline{\eta}_\sigma \sum_{\tau'} M_{\tau\tau'} \mathbb{E}_{\tau'}[\hat{\delta}_\sigma] + \frac{1}{2} \tilde{\mathbb{E}}[\ell(\ell-1)] \overline{\eta}_\sigma \sum_{\tau',\tau''} M_{\tau\tau'} M_{\tau\tau''} \left[\mathbb{E}_{\tau'}[\hat{\delta}_\sigma] \mathbb{E}_{\tau''}[\hat{\delta}_\sigma] - \sum_\gamma \overline{\eta}_\gamma \mathbb{E}_{\tau'}[\hat{\delta}_\gamma] \mathbb{E}_{\tau''}[\hat{\delta}_\gamma] \right.
$$

$$
\left. - 2 \sum_\gamma \overline{\eta}_\gamma \mathbb{E}_{\tau'}[\hat{\delta}_\sigma \hat{\delta}_\gamma] \mathbb{E}_{\tau''}[\hat{\delta}_\gamma] - \sum_\gamma \overline{\eta}_\gamma \mathbb{E}_{\tau'}[\hat{\delta}_\sigma \hat{\delta}_\gamma] \mathbb{E}_{\tau''}[\hat{\delta}_\sigma \hat{\delta}_\gamma] + \sum_{\gamma,\beta} \overline{\eta}_\gamma \overline{\eta}_\beta \mathbb{E}_{\tau'}[\hat{\delta}_\gamma \hat{\delta}_\beta] \mathbb{E}_{\tau''}[\hat{\delta}_\gamma \hat{\delta}_\beta] \right]. \tag{63}
$$

In order to simplify this equation, we first state an identity similar to (56),

$$
\sum_{\tau'} M_{\tau\tau'} \mathbb{E}_{\tau'}[\hat{\delta}_\sigma \hat{\delta}_\gamma] = \sum_{\tau'} \hat{M}_{\tau\tau'} \mathbb{E}[\hat{\delta}_\sigma \hat{\delta}_\gamma \eta_{\tau'}] = \mathbb{E}[\hat{\delta}_\sigma \hat{\delta}_\gamma] + \mathbb{E}[\hat{\delta}_\sigma \hat{\delta}_\gamma \hat{\delta}_\tau]. \tag{64}
$$

Note that, according to our ansatz, the second term is negligible with respect to the first one (it is of order $\kappa^2 \ll \kappa$).

We will now multiply (63) by $\overline{\eta}_\tau$ in such a way that its left-hand side becomes $A_{\sigma\tau}$; the right-hand side can be simplified by using the identities (56) and (64). Keeping only the terms of order κ and κ^2 yields

$$
A_{\sigma\tau} = \tilde{\mathbb{E}}[\ell] \overline{\eta}_\sigma \overline{\eta}_\tau \hat{A}_{\sigma\tau} + \frac{1}{2} \tilde{\mathbb{E}}[\ell(\ell-1)] \overline{\eta}_\sigma \overline{\eta}_\tau \left[(\hat{A}_{\sigma\tau})^2 - \sum_\gamma \overline{\eta}_\gamma (\hat{A}_{\sigma\gamma} + \hat{A}_{\gamma\tau})^2 + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\hat{A}_{\gamma\beta})^2 \right], \tag{65}
$$

$$
\hat{A}_{\sigma\tau} = \sum_{\sigma',\tau'} \hat{M}_{\sigma\sigma'} \hat{M}_{\tau\tau'} A_{\sigma'\tau'}, \tag{66}
$$

where we recalled the equation linking A and $\hat{A}$ in Eq. (66).

An equivalent form of these equations, which will turn out to be more convenient in some cases, can be given in the basis that diagonalizes the matrix M defined in Sec. IV B. Using the eigenvectors defined in Eq. (44), we can trade the matrix A for a matrix A', now indexed by these basis vectors, in an invertible way:

$$
A_{\sigma\tau} = \sum_{jk} l_\sigma^{(j)} l_\tau^{(k)} A'_{jk} \leftrightarrow A'_{jk} = \sum_{\sigma\tau} r_\sigma^{(j)} r_\tau^{(k)} A_{\sigma\tau}. \tag{67}
$$

The normalization condition that implied the vanishing of the row and column sums of A shows up now as $A'_{jk} = 0$ whenever $j = 1$ or $k = 1$. The transformation from A to $\hat{A}$, i.e., the matrix multiplication by $\hat{M}$ for each of the two indices, becomes for A' a simple multiplication by the eigenvalues θ_j of the matrix M. Equation (65) thus becomes, in this basis,

$$
A'_{jk} = \tilde{\mathbb{E}}[\ell] \theta_j \theta_k A'_{jk} + \frac{1}{2} \tilde{\mathbb{E}}[\ell(\ell-1)] \left[\sum_{j_1 j_2 k_1 k_2} f_{j j_1 j_2} f_{k k_1 k_2} \theta_{j_1} \theta_{j_2} \theta_{k_1} \theta_{k_2} A'_{j_1 k_1} A'_{j_2 k_2} - 2 \theta_j \theta_k \sum_l A'_{jl} A'_{lk} \theta_l^2 \right], \tag{68}
$$

where we have defined

$$f_{j_1 j_2 j_3} = \sum_\sigma \overline{\eta}_\sigma r_\sigma^{(j_1)} r_\sigma^{(j_2)} r_\sigma^{(j_3)}. \tag{69}$$

We have achieved here our first technical goal: This quadratic equation (65) on the matrix A [or equivalently (68) on the matrix A'] constitutes the equivalent of (10) for the scalar bifurcation analysis, but now for a generic model with pairwise interaction between Potts variables in the sparse regime. In particular, imposing the positive definiteness of A will allow us to discriminate between different bifurcation scenarios, as explained in Sec. II D in the simpler scalar case. We will see in the following various applications of this formula for different choices of the matrix M and stationary distribution $\overline{\eta}$.

3. Third-order expansion

Before turning to these applications let us first state the results of this moment expansion at the next order (which in some cases will be crucial to discriminate between different bifurcation scenarios). According to our ansatz (61), we now have to determine the centered moments of $P(\eta)$ up to the fourth moment. We thus define

$$A_{\sigma\tau} = \mathbb{E}[\delta_\sigma \delta_\tau], \quad B_{\sigma\tau\gamma} = \mathbb{E}[\delta_\sigma \delta_\tau \delta_\gamma], \quad C_{\sigma\tau\gamma\beta} = \mathbb{E}[\delta_\sigma \delta_\tau \delta_\gamma \delta_\beta]. \tag{70}$$

Generalizing the derivation of (65) above by including higher-order terms, we have obtained the following set of equations for A, B, and C (see Appendix B for a justification through a formal expansion at all orders):

$$
A_{\sigma\tau} = \tilde{\mathbb{E}}[\ell]\overline{\eta}_\sigma \overline{\eta}_\tau \hat{A}_{\sigma\tau} + \frac{1}{2}\tilde{\mathbb{E}}[\ell(\ell-1)]\overline{\eta}_\sigma \overline{\eta}_\tau \left[(\hat{A}_{\sigma\tau})^2 - \sum_\gamma \overline{\eta}_\gamma (\hat{A}_{\sigma\gamma} + \hat{A}_{\gamma\tau})^2 + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\hat{A}_{\gamma\beta})^2 \right]
$$

$$
+ \tilde{\mathbb{E}}[\ell(\ell-1)]\overline{\eta}_\sigma \overline{\eta}_\tau \left[-\sum_\gamma \overline{\eta}_\gamma \hat{B}_{\sigma\tau\gamma}(\hat{A}_{\sigma\gamma} + \hat{A}_{\tau\gamma}) + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\hat{B}_{\sigma\gamma\beta} + \hat{B}_{\tau\gamma\beta})\hat{A}_{\gamma\beta} + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta \hat{C}_{\sigma\tau\gamma\beta}\hat{A}_{\gamma\beta} \right]
$$

$$
+ \tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\overline{\eta}_\sigma \overline{\eta}_\tau \left[\frac{1}{6}(\hat{A}_{\sigma\tau})^3 - \frac{1}{6}\sum_\gamma \overline{\eta}_\gamma (\hat{A}_{\sigma\gamma} + \hat{A}_{\tau\gamma})^3 - \frac{1}{2}\hat{A}_{\sigma\tau}\sum_\gamma \overline{\eta}_\gamma (\hat{A}_{\sigma\gamma} + \hat{A}_{\tau\gamma})^2 \right.
$$

$$
+ \frac{1}{6}\sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\hat{A}_{\gamma\beta})^3 + \frac{1}{2}\sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\hat{A}_{\gamma\beta})^2 (\hat{A}_{\sigma\tau} + \hat{A}_{\sigma\gamma} + \hat{A}_{\tau\gamma} + \hat{A}_{\sigma\beta} + \hat{A}_{\tau\beta})
$$

$$
\left. + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta \hat{A}_{\gamma\beta}(\hat{A}_{\sigma\gamma} + \hat{A}_{\tau\gamma})(\hat{A}_{\sigma\beta} + \hat{A}_{\tau\beta}) - \sum_{\gamma\beta\alpha} \overline{\eta}_\gamma \overline{\eta}_\beta \overline{\eta}_\alpha \hat{A}_{\gamma\beta}\hat{A}_{\beta\alpha}\hat{A}_{\alpha\gamma} \right], \tag{71}
$$

$$
B_{\sigma\tau\gamma} = \tilde{\mathbb{E}}[\ell]\overline{\eta}_\sigma \overline{\eta}_\tau \overline{\eta}_\gamma \hat{B}_{\sigma\tau\gamma} + \tilde{\mathbb{E}}[\ell(\ell-1)]\overline{\eta}_\sigma \overline{\eta}_\tau \overline{\eta}_\gamma \left[\hat{A}_{\sigma\tau}\hat{A}_{\tau\gamma} + \hat{A}_{\sigma\gamma}\hat{A}_{\gamma\tau} + \hat{A}_{\tau\sigma}\hat{A}_{\sigma\gamma} - \sum_\beta \overline{\eta}_\beta (\hat{A}_{\sigma\beta}\hat{A}_{\beta\gamma} + \hat{A}_{\sigma\beta}\hat{A}_{\beta\tau} + \hat{A}_{\tau\beta}\hat{A}_{\beta\gamma}) \right], \tag{72}
$$

$$
C_{\sigma\tau\gamma\beta} = \tilde{\mathbb{E}}[\ell]\overline{\eta}_\sigma \overline{\eta}_\tau \overline{\eta}_\gamma \overline{\eta}_\beta \hat{C}_{\sigma\tau\gamma\beta} + \tilde{\mathbb{E}}[\ell(\ell-1)]\overline{\eta}_\sigma \overline{\eta}_\tau \overline{\eta}_\gamma \overline{\eta}_\beta (\hat{A}_{\sigma\tau}\hat{A}_{\gamma\beta} + \hat{A}_{\sigma\gamma}\hat{A}_{\tau\beta} + \hat{A}_{\sigma\beta}\hat{A}_{\tau\gamma}), \tag{73}
$$

$$\hat{A}_{\sigma\tau} = \sum_{\sigma'\tau'} \hat{M}_{\sigma\sigma'}\hat{M}_{\tau\tau'} A_{\sigma'\tau'}, \tag{74}$$

$$\hat{B}_{\sigma\tau\gamma} = \sum_{\sigma'\tau'\gamma'} \hat{M}_{\sigma\sigma'}\hat{M}_{\tau\tau'}\hat{M}_{\gamma\gamma'} B_{\sigma'\tau'\gamma'}, \tag{75}$$

$$\hat{C}_{\sigma\tau\gamma\beta} = \sum_{\sigma'\tau'\gamma'\beta'} \hat{M}_{\sigma\sigma'}\hat{M}_{\tau\tau'}\hat{M}_{\gamma\gamma'}\hat{M}_{\beta\beta'} C_{\sigma'\tau'\gamma'\beta'}. \tag{76}$$

E. Expansion of the free entropy

Let us now focus on the second fundamental object introduced besides the cavity equation recursions, namely, the free-entropy functional, in particular, the expressions (38) and (40) for the case of pairwise interacting Potts variables. With the conventions we used, the value of ϕ for the trivial fixed point $P_{\text{triv}}(\eta) = \delta(\eta - \overline{\eta})$ is $\phi(P_{\text{triv}}) = 0$. We have computed an expansion of $\phi(P)$ when P is close to P_{triv}, i.e., when the centered moments of P are small [imposing the condition $\int dP(\eta)\eta = \overline{\eta}$]. Again, denoting by A, B, and C the tensors encoding these centered moments of P of orders 2, 3, and 4 [as defined in Eq. (70)] and by $\hat{A}$, $\hat{B}$, and

$\hat{C}$ their transformed versions according to (74)–(76), we have found (see Appendix C for the details of the derivation)

$$
\begin{aligned}
\frac{1}{\mathbb{E}[\ell]}\phi(P) = {}& -\frac{1}{4}\sum_{\sigma\tau}A_{\sigma\tau}\hat{A}_{\sigma\tau} + \frac{1}{12}\sum_{\sigma\tau\gamma}B_{\sigma\tau\gamma}\hat{B}_{\sigma\tau\gamma} - \frac{1}{24}\sum_{\sigma\tau\gamma\beta}C_{\sigma\tau\gamma\beta}\hat{C}_{\sigma\tau\gamma\beta} + \frac{1}{4}\tilde{\mathbb{E}}[\ell]\sum_{\sigma\tau}\overline{\eta}_\sigma\overline{\eta}_\tau(\hat{A}_{\sigma\tau})^2 - \frac{1}{12}\tilde{\mathbb{E}}[\ell]\sum_{\sigma\tau\gamma}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma(\hat{B}_{\sigma\tau\gamma})^2 \\
& + \frac{1}{24}\tilde{\mathbb{E}}[\ell]\sum_{\sigma\tau\gamma\beta}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\overline{\eta}_\beta(\hat{C}_{\sigma\tau\gamma\beta})^2 + \frac{1}{12}\tilde{\mathbb{E}}[\ell(\ell-1)]\left[\sum_{\sigma\tau}\overline{\eta}_\sigma\overline{\eta}_\tau(\hat{A}_{\sigma\tau})^3 - 2\sum_{\sigma\tau\gamma}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\hat{A}_{\sigma\tau}\hat{A}_{\tau\gamma}\hat{A}_{\gamma\sigma}\right] \\
& - \frac{1}{2}\tilde{\mathbb{E}}[\ell(\ell-1)]\sum_{\sigma\tau\gamma}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\hat{B}_{\sigma\tau\gamma}\hat{A}_{\sigma\tau}\hat{A}_{\sigma\gamma} + \frac{1}{4}\tilde{\mathbb{E}}[\ell(\ell-1)]\sum_{\sigma\tau\gamma\beta}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\overline{\eta}_\beta\hat{C}_{\sigma\tau\gamma\beta}\hat{A}_{\sigma\tau}\hat{A}_{\gamma\beta} \\
& + \frac{1}{48}\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\left[\sum_{\sigma\tau}\overline{\eta}_\sigma\overline{\eta}_\tau(\hat{A}_{\sigma\tau})^4 - 6\sum_{\sigma\tau\gamma}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma(\hat{A}_{\sigma\tau})^2(\hat{A}_{\sigma\gamma})^2 - 12\sum_{\sigma\tau\gamma}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma(\hat{A}_{\sigma\tau})^2\hat{A}_{\sigma\gamma}\hat{A}_{\tau\gamma}\right. \\
& \left. + 3\left(\sum_{\sigma\tau}\overline{\eta}_\sigma\overline{\eta}_\tau(\hat{A}_{\sigma\tau})^2\right)^2 + 12\sum_{\sigma\tau\gamma\beta}\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\overline{\eta}_\beta\hat{A}_{\sigma\tau}\hat{A}_{\tau\gamma}\hat{A}_{\gamma\beta}\hat{A}_{\beta\sigma}\right]. \quad (77)
\end{aligned}
$$

Of course this expression is a truncation of an infinite series involving moments of all orders. Let us make a few comments about this result.

(i) One can rewrite this expansion in the eigenbasis of M, i.e., in terms of the quantities A'_{jk} instead of $A_{\sigma\tau}$, as defined in Eq. (67). The corresponding expression is given in Appendix C.

(ii) The expression (40) of the free entropy $\phi(P)$ is variational, in the sense that its variation with respect to P vanishes when P is a fixed-point solution of the recursion equation (37). Accordingly, the derivatives of the truncated expansion (77) with respect to A, B, and C vanish when the latter are solutions of the moment equations (71)–(76).

(iii) If the centered moments of P verify the scaling ansatz (61) that we argued was the correct one for fixed points of the cavity recursion, then the expansion (77) contains all the terms of order κ^2, κ^3, and κ^4.

(iv) However, the free-entropy functional $\phi(P)$ of Eq. (40) is well defined even if P is not a fixed point of the recursion equation (37) [as long as it satisfies $\int dP(\eta)\eta = \overline{\eta}$] and accordingly the expansion (77) can be used even if A, B, and C do not satisfy (71)–(76). As a matter of fact, such an expansion was performed in Ref. [25], and we briefly mention the similarities and differences with respect to ours. In the context of the models studied in this section, one of the results of [25] (i.e., Theorem 2.3 therein) can be rephrased as follows: Above the Kesten-Stigum transition, i.e., whenever $\tilde{\mathbb{E}}[\ell]\theta_2^2 > 1$ in such a way that the trivial fixed point is unstable under the cavity recursions, there exists a distribution P with $\phi(P) > 0$ and hence the Kesten-Stigum transition is an upper bound for the information-theoretic transition (or condensation transition in the associated nonplanted random ensemble). We recover this statement from our computations, defining the probability distributions P_λ by

$$
P_\lambda(\eta) = \tfrac{1}{2}\delta(\eta - (\overline{\eta} + \lambda l^{(2)})) + \tfrac{1}{2}\delta(\eta - (\overline{\eta} - \lambda l^{(2)})), \quad (78)
$$

where $l^{(2)}$ is the left eigenvector of M associated with the eigenvalue θ_2 and $\lambda > 0$ should be sufficiently small for $\overline{\eta} \pm \lambda l^{(2)}$ to remain inside the polytope of probability distributions. The average of P_λ is $\overline{\eta}$ for all λ, and the first centered moments

of P_λ are easily seen to be

$$
A_{\sigma\tau} = \lambda^2 l_\sigma^{(2)} l_\tau^{(2)}, \quad B_{\sigma\tau\gamma} = 0, \quad C_{\sigma\tau\gamma\beta} = \lambda^4 l_\sigma^{(2)} l_\tau^{(2)} l_\gamma^{(2)} l_\beta^{(2)}. \quad (79)
$$

More generically, the odd centered moments vanish and the $2p$th one is λ^{2p} times the $2p$th tensor power of $l^{(2)}$. Computing the corresponding value of $\hat{A}_{\sigma\tau}$ and inserting in Eq. (77), one obtains

$$
\frac{1}{\mathbb{E}[\ell]}\phi(P_\lambda) = \frac{1}{4}\theta_2^2\left(\tilde{\mathbb{E}}[\ell]\theta_2^2 - 1\right)\lambda^4 + O(\lambda^6), \quad (80)
$$

which can indeed be made strictly positive as soon as $\tilde{\mathbb{E}}[\ell]\theta_2^2 > 1$ by taking a sufficiently small λ. Note that the centered moments of P_λ do not satisfy our ansatz (61), which is not contradictory as P_λ has no reason to be a fixed point of the cavity recursion [incidentally, there might be no perturbative nontrivial fixed point in such a situation; see Fig. 1(b)]. Only two terms of (77) contribute to $\phi(P_\lambda)$ at order λ^4 (the first ones in the first and second lines), hence our expansion is more detailed than the one in Ref. [25], which in contrast deals with a much more general and complicated setting (k-wise interaction of Potts variables with quenched disorder) and is performed in a mathematically rigorous way.

F. Large-degree limit

We have underlined in the Introduction the existence of two families of inference problems, the dense ones which can be described by a finite-dimensional order parameter and the sparse ones that require the use of a distributional infinite-dimensional description. The focus of this paper is on the latter family, but it is important to realize that the two types of problems are not completely disjoint. As a matter of fact, the sparse models reduce effectively to the simpler dense ones if one takes the limit of large degrees, after the thermodynamic limit. This is a well-known fact that has been derived rigorously in several papers (see, for instance, [22,27,28]). For the sake of self-containedness and to be able to contrast the behavior of the sparse and dense models, we reproduce here this derivation.

We will consider the recursion for the unconditional distribution (37) and simplify it in the large-degree limit $\tilde{\mathbb{E}}[\ell] \to \infty$. We will show that in this limit the evolution of the distribution $P^{(n)}$ can be tracked exactly by following a finite-dimensional object, namely, the covariance matrix $A_{\sigma\tau}^{(n)} = \mathbb{E}^{(n)}[\delta_\sigma \delta_\tau]$, that obeys a recursion equation of the form $A^{(n+1)} = F(A^{(n)})$.

The main idea of the derivation is to exploit a central-limit theorem. We will thus rewrite (37) in a way that makes apparent a sum of a large number of random variables. To reach this goal we will define two functions $\psi(L)$ and $\tilde{z}(L)$ whose arguments are q-dimensional real vectors $L = (L_1, \dots, L_q)$ that are mapped by ψ and $\tilde{z}$ to normalized probability distributions and positive reals, respectively, according to

$$\psi(L)_\sigma = \frac{e^{L_\sigma}}{\tilde{z}(L)}, \quad \tilde{z}(L) = \sum_{\gamma=1}^{q} e^{L_\gamma}. \tag{81}$$

One can translate the unconditional recursion relation (37) as

$$\mathbb{E}^{(n+1)}[g(\eta)] = \mathbb{E}^{(n)}[g(\psi(L))\tilde{z}(L)] \tag{82}$$

for an arbitrary function g, where on the right-hand side the vector L has the distribution

$$(L)_{\sigma=1,\dots,q} \stackrel{d}{=} [\ln z_\sigma(\eta^1, \dots, \eta^\ell)]_{\sigma=1,\dots,q}, \tag{83}$$

with ℓ drawn from $\tilde{p}_\ell$ and the η^i drawn from $P^{(n)}$. From the expression of z_σ given in Eq. (34) one realizes that

$$L_\sigma = \ln \overline{\eta}_\sigma + \sum_{i=1}^{\ell} \ln \left(\sum_{\sigma'} \hat{M}_{\sigma\sigma'} \eta_{\sigma'}^i \right). \tag{84}$$

We now recall that if $Y \stackrel{d}{=} \sum_{i=1}^{n} X^i$ is a sum of independent and identically distributed random vectors, the number of summands n being also random, the two first moments of the sum are given by

$$(\mathbb{E}[Y])_\sigma = \mathbb{E}[n]\mathbb{E}[X_\sigma], \tag{85}$$

$$(\mathrm{Cov}[Y])_{\sigma\tau} = \mathbb{E}[n]\mathbb{E}[X_\sigma X_\tau] + (\mathrm{Var}[n] - \mathbb{E}[n])\mathbb{E}[X_\sigma]\mathbb{E}[X_\tau]. \tag{86}$$

Moreover, the large-degree limit $\tilde{\mathbb{E}}[\ell] \to \infty$ yields a nontrivial result only if the nontrivial eigenvalues of M, $\theta_2, \dots, \theta_q$, vanish in the limit. More precisely, they have to be of the form $\theta_i = \tilde{\theta}_i / \sqrt{\tilde{\mathbb{E}}[\ell]}$, with $\tilde{\theta}_i$ finite. We thus have

$$X_\sigma = \ln \left(\sum_{\sigma'} \hat{M}_{\sigma\sigma'} \eta_{\sigma'} \right) = \ln \left(1 + \frac{1}{\sqrt{\tilde{\mathbb{E}}[\ell]}} \sum_{j=2}^{q} \tilde{\theta}_j r_\sigma^{(j)} \sum_{\sigma'} r_{\sigma'}^{(j)} \delta_{\sigma'} \right) \tag{87}$$

$$= \frac{1}{\sqrt{\tilde{\mathbb{E}}[\ell]}} \sum_{j=2}^{q} \tilde{\theta}_j r_\sigma^{(j)} \sum_{\sigma'} r_{\sigma'}^{(j)} \delta_{\sigma'} - \frac{1}{2} \frac{1}{\tilde{\mathbb{E}}[\ell]} \left(\sum_{j=2}^{q} \tilde{\theta}_j r_\sigma^{(j)} \sum_{\sigma'} r_{\sigma'}^{(j)} \delta_{\sigma'} \right)^2 + o\left(\frac{1}{\tilde{\mathbb{E}}[\ell]} \right). \tag{88}$$

Combining these various observations and assuming that the offspring degree distribution is sufficiently well behaved for the second term in Eq. (86) to be negligible (it vanishes exactly for a Poisson distribution), we find that in the large-degree limit the promised recursion $A^{(n+1)} = F(A^{(n)})$ can be decomposed in two steps: (i) From $A^{(n)}$ compute $\tilde{A}^{(n)}$ with

$$\tilde{A}_{\sigma\tau}^{(n)} = \lim \left(\tilde{\mathbb{E}}[\ell] \sum_{\sigma'\tau'} \hat{M}_{\sigma\sigma'} \hat{M}_{\tau\tau'} A_{\sigma'\tau'}^{(n)} \right) = \sum_{j,k=2}^{q} \tilde{\theta}_j \tilde{\theta}_k r_\sigma^{(j)} r_\tau^{(k)} \sum_{\sigma'\tau'} r_{\sigma'}^{(j)} r_{\tau'}^{(k)} A_{\sigma'\tau'}^{(n)} \tag{89}$$

and then (ii) compute the new covariance matrix $A^{(n+1)}$ from

$$A_{\sigma\tau}^{(n+1)} = \mathbb{E}[(\psi(L)_\sigma - \overline{\eta}_\sigma)(\psi(L)_\tau - \overline{\eta}_\tau)\tilde{z}(L)] = \mathbb{E}\left[\frac{[e^{L_\sigma} - \overline{\eta}_\sigma \tilde{z}(L)][e^{L_\tau} - \overline{\eta}_\tau \tilde{z}(L)]}{\tilde{z}(L)} \right], \tag{90}$$

with L a Gaussian vector characterized by its first two moments

$$(\mathbb{E}[L])_\sigma = \ln \overline{\eta}_\sigma - \tfrac{1}{2}\tilde{A}_{\sigma\sigma}^{(n)}, \quad (\mathrm{Cov}[L])_{\sigma\tau} = \tilde{A}_{\sigma\tau}^{(n)}. \tag{91}$$

As emphasized before, this finite-dimensional recursion $A^{(n+1)} = F(A^{(n)})$ is an exact description of the functional recursion on $P^{(n)}$ in the large-degree limit. The trivial covariance $A = 0$ is a fixed point of F that reproduces the stationarity of $P(\eta) = \delta(\eta - \overline{\eta})$ under the functional recursion. It is instructive to expand the equation $A = F(A)$ around $A = 0$ in order to compare this expansion with the one we performed previously on the full functional equation. Exploiting the properties of Gaussian random variables, it is relatively easy to obtain, from (90),

$$A_{\sigma\tau} = \overline{\eta}_\sigma \overline{\eta}_\tau \tilde{A}_{\sigma\tau} + \frac{1}{2}\overline{\eta}_\sigma \overline{\eta}_\tau \left[(\tilde{A}_{\sigma\tau})^2 - \sum_\gamma \overline{\eta}_\gamma (\tilde{A}_{\sigma\gamma} + \tilde{A}_{\gamma\tau})^2 + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\tilde{A}_{\gamma\beta})^2 \right] + \overline{\eta}_\sigma \overline{\eta}_\tau \left[\frac{1}{6}(\tilde{A}_{\sigma\tau})^3 - \frac{1}{6}\sum_\gamma \overline{\eta}_\gamma (\tilde{A}_{\sigma\gamma} + \tilde{A}_{\tau\gamma})^3 \right.$$

$$- \frac{1}{2}\tilde{A}_{\sigma\tau} \sum_\gamma \overline{\eta}_\gamma (\tilde{A}_{\sigma\gamma} + \tilde{A}_{\tau\gamma})^2 + \frac{1}{6}\sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\tilde{A}_{\gamma\beta})^3 + \frac{1}{2}\sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta (\tilde{A}_{\gamma\beta})^2 (\tilde{A}_{\sigma\tau} + \tilde{A}_{\sigma\gamma} + \tilde{A}_{\tau\gamma} + \tilde{A}_{\sigma\beta} + \tilde{A}_{\tau\beta})$$

$$\left. + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta \tilde{A}_{\gamma\beta}(\tilde{A}_{\sigma\gamma} + \tilde{A}_{\tau\gamma})(\tilde{A}_{\sigma\beta} + \tilde{A}_{\tau\beta}) - \sum_{\gamma\beta\alpha} \overline{\eta}_\gamma \overline{\eta}_\beta \overline{\eta}_\alpha \tilde{A}_{\gamma\beta} \tilde{A}_{\beta\alpha} \tilde{A}_{\alpha\gamma} \right] + O(A^4). \tag{92}$$

This is, as expected, consistent with the expansion performed in the sparse case and presented in Eq. (71). Of course the equations in the sparse regime are richer, with the effects of the third- and fourth-order moments B and C being washed out in the large-degree limit by the Gaussian character of the distribution.

For completeness, let us also give the large-degree limit for the free entropy, which becomes a function of a covariance matrix A,

$$\phi(A) = \mathbb{E}[\tilde{z}(L) \ln \tilde{z}(L)] - \frac{1}{4} \sum_{\sigma\tau} A_{\sigma\tau} \tilde{A}_{\sigma\tau}, \qquad (93)$$

where $\tilde{A}$ is obtained from A according to (89) and in the first term L is a Gaussian vector with the two first moments indicated in Eq. (91).

In the remainder of this section we will investigate the consequences of our expansions, which were performed for an arbitrary choice of q, $\overline{\eta}$, and M, in several specific cases. In particular, we have to analyze the bifurcation of the quadratic equation (65) on the covariance matrix A in the neighborhood of the Kesten-Stigum transition and use it to discriminate between the possible scenarios depicted in Figs. 1 and 2 by imposing the positive definiteness of A.

G. Application 1: The nondegenerate case

Let us first consider the case of a simple second eigenvalue of M at the Kesten-Stigum transition. More explicitly, we consider that the various parameters of the model (degree distribution, $\overline{\eta}$, and M) are functions of a single control parameter ϵ defined in such a way that $\tilde{\mathbb{E}}[\ell]\theta_2^2 = 1 + \epsilon$, with $\epsilon = 0$ at the Kesten-Stigum transition, and we assume that $\tilde{\mathbb{E}}[\ell]\theta_j^2$ are bounded away from 1 for $j = 3, \ldots, q$ in the neighborhood of $\epsilon = 0$.

Consider now the set of quadratic equations (68) on A', the covariance matrix expressed in the basis of eigenvectors of M. With the assumption of simplicity of θ_2, the only coefficient of the linear terms on the right-hand side that crosses 1 at the Kesten-Stigum transition corresponds to $j = k = 2$; the bifurcating solution will thus have $|A'_{22}| \gg |A'_{jk}|$ for $(j, k) \neq (2, 2)$ in the neighborhood of $\epsilon = 0$. We can thus give the equation for this leading term as

$$A'_{22} = \tilde{\mathbb{E}}[\ell]\theta_2^2 A'_{22} + \frac{1}{2}\tilde{\mathbb{E}}[\ell(\ell-1)]\theta_2^4 (A'_{22})^2 [(f_{222})^2 - 2], \qquad (94)$$

where the expression of f_{222} is given in Eq. (69). At the leading order in ϵ we have

$$0 = \epsilon + \frac{1}{2}\tilde{\mathbb{E}}[\ell(\ell-1)]\theta_2^4 [(f_{222})^2 - 2]A'_{22}, \qquad (95)$$

hence A'_{22} varies linearly with ϵ around $\epsilon = 0$, the other matrix elements of A' being at least of order ϵ^2. The crucial point is now to remember that A', as a covariance matrix, must be positive definite and as a consequence A'_{22} must be a positive real. Depending on the sign of $(f_{222})^2 - 2$, this happens for $\epsilon > 0$ ($\epsilon < 0$), corresponding, for instance, to the scenario sketched in Fig. 1(a) [Fig. 1(b)]. Spelling out the definition

of f_{222} in terms of the eigenvectors of M, we have obtained that the first case occurs when

$$\left(\sum_{\sigma} \overline{\eta}_{\sigma} \left(r_{\sigma}^{(2)} \right)^3 \right)^2 < 2. \qquad (96)$$

Let us underline the striking similarity between this criterion and the one obtained in Ref. [16] in the context of matrix factorization. This is a dense inference problem with continuous variables, for which the condition of existence of a bifurcating solution above the Kesten-Stigum transition was found to be $\langle x_0^3 \rangle^2 < 2\langle x_0^2 \rangle^3$, the average being here on the prior distribution of a real-valued variable x_0 [see Eq. (205) in Ref. [16]]. Using the properties of the eigenvalue decomposition of M explained in Sec. IV B, one can rewrite (96) as

$$\left(\sum_{\sigma} \overline{\eta}_{\sigma} \left(r_{\sigma}^{(2)} \right)^3 \right)^2 < 2 \left(\sum_{\sigma} \overline{\eta}_{\sigma} \left(r_{\sigma}^{(2)} \right)^2 \right)^3, \qquad (97)$$

a formally equivalent expression with $\overline{\eta}_{\sigma}$ replacing the prior distribution and the eigenvector $r_{\sigma}^{(2)}$ the continuous variable x_0.

H. Application 2: The symmetric q-state case

We turn now to a case which can be viewed as the opposite of the previous one, with a maximal degeneracy of the second eigenvalue of M. Consider indeed the symmetric q-state Potts model, with a stationary distribution $\overline{\eta}_{\sigma} = \frac{1}{q}$ for all σ and a matrix M which is invariant under all permutations of its rows and columns. As M is stochastic it must be of the form

$$M_{\sigma\tau} = \frac{1}{q} + \theta K_{\sigma\tau} \quad \text{with} \quad K_{\sigma\tau} = \delta_{\sigma,\tau} - \frac{1}{q}; \qquad (98)$$

θ is the nontrivial eigenvalue of M, with a degeneracy $q - 1$. The matrix elements of M must be non-negative and the authorized range of θ is thus $]-\frac{1}{q-1}, 1[$. The model is said to be assortative (ferromagnetic) if $\theta > 0$ and disassortative (antiferromagnetic) if $\theta < 0$. The matrix elements of $\hat{M}$ are $\hat{M}_{\sigma\tau} = 1 + q\theta K_{\sigma\tau}$.

The recursion equation (37), for this choice of $\overline{\eta}$ and M, will respect the permutation symmetry between the q values of σ. If the initial condition is symmetric this will also be the case for $P^{(n)}$ for all n. In particular, the covariance matrix $A^{(n)}$ will be invariant under all row and column permutations; as its row sums must also vanish, it is necessarily of the form $A^{(n)} = a^{(n)}K$, with $a^{(n)}$ a real number and K the matrix defined in Eq. (98). The eigenvalues of K being 0 and 1, the positivity constraint on $A^{(n)}$ thus translates into $a^{(n)} \geqslant 0$. One finds easily with this form of A and M that $\hat{A}^{(n)} = (q\theta)^2 a^{(n)}K$. Plugging these forms of A and $\hat{A}$ in the linearized evolution equation (59) yields $a^{(n+1)} = \tilde{\mathbb{E}}[\ell]\theta^2 a^{(n)}$, hence the Kesten-Stigum condition $\tilde{\mathbb{E}}[\ell]\theta^2 = 1$ for the limit of stability of the trivial solution $a = 0$, as expected.

Looking for a fixed-point solution of the form $A_{\sigma\tau} = aK_{\sigma\tau}$ in the neighborhood of the Kesten-Stigum transition, we obtain from (65), after a brief computation,

$$a = \tilde{\mathbb{E}}[\ell]\theta^2 a + \tilde{\mathbb{E}}[\ell(\ell-1)]\theta^4 \frac{1}{2}q(q-4)a^2. \qquad (99)$$

Defining $\tilde{\mathbb{E}}[\ell]\theta^2 = 1 + \epsilon$ and keeping in mind the crucial positivity condition $a \geqslant 0$, one realizes that for $q < 4$ the nontrivial perturbative fixed point exists for $\epsilon > 0$, whereas if $q > 4$ it exists for $\epsilon < 0$, which implies that the Kesten-Stigum bound is not tight for the tree reconstruction problem in the latter case. This fact was proven rigorously in Ref. [22], with Eq. (99) corresponding to (1.1) in Ref. [22] (with a linear change of variable between our $a^{(n)}$ and x_n of [22]). Let us emphasize the necessity of imposing the permutation symmetry of A in order to reach this conclusion; the bifurcation equations (65) admit indeed, besides the symmetric solution studied above, nonsymmetric spurious solutions A even when $\bar{\eta}$ and M are invariant under permutations.

We have thus justified the statements made in Sec. III D 1 in the cases $q \leqslant 3$ and $q \geqslant 5$. When $q = 4$ the coefficient of a^2 in the bifurcation equation vanishes exactly and one cannot conclude from this lowest-order expansion about the type of transition encountered at the Kesten-Stigum threshold. Due to the next-order expansion presented in Sec. IV D 3, we will now be able to elucidate this case and justify the effort put into this generalization of the expansion. In order to solve Eqs. (71)–(76) on A, B, and C, we first note that B and C vanish when summed over any of their indices; from the expression of $\hat{M}$ in the case under study, we can thus conclude that $\hat{B}_{\sigma\tau\gamma} = (q\theta)^3 B_{\sigma\tau\gamma}$ and $\hat{C}_{\sigma\tau\gamma\beta} = (q\theta)^4 C_{\sigma\tau\gamma\beta}$. Inserting these expressions in Eq. (72) and (73) yields an explicit solution for B and C in terms of $\hat{A}$,

$$B_{\sigma\tau\gamma} = \frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{1 - \tilde{\mathbb{E}}[\ell]\theta^3} \frac{1}{q^3} \left[\hat{A}_{\sigma\tau}\hat{A}_{\tau\gamma} + \hat{A}_{\sigma\gamma}\hat{A}_{\gamma\tau} + \hat{A}_{\tau\sigma}\hat{A}_{\sigma\gamma} - \frac{1}{q}\sum_\beta (\hat{A}_{\sigma\beta}\hat{A}_{\beta\gamma} + \hat{A}_{\sigma\beta}\hat{A}_{\beta\tau} + \hat{A}_{\tau\beta}\hat{A}_{\beta\gamma}) \right], \tag{100}$$

$$C_{\sigma\tau\gamma\beta} = \frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{1 - \tilde{\mathbb{E}}[\ell]\theta^4} \frac{1}{q^4} (\hat{A}_{\sigma\tau}\hat{A}_{\gamma\beta} + \hat{A}_{\sigma\gamma}\hat{A}_{\tau\beta} + \hat{A}_{\sigma\beta}\hat{A}_{\tau\gamma}). \tag{101}$$

Furthermore, we argued above by symmetry arguments that $A = aK$ and $\hat{A} = a(q\theta)^2 K$; we can thus determine completely B and C modulo the still unknown real a. Plugging these expressions in Eq. (71) gives, after some combinatorial evaluations, the equation for a that generalizes (99),

$$a = \tilde{\mathbb{E}}[\ell]\theta^2 a + \tilde{\mathbb{E}}[\ell(\ell-1)]\theta^4 \tfrac{1}{2}q(q-4)a^2 + wa^3, \tag{102}$$

where the coefficient w reads

$$w = \frac{1}{6}q^2(q^2 - 18q + 42)\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\theta^6 - \frac{6q^2(q-2)\tilde{\mathbb{E}}[\ell(\ell-1)]^2\theta^9}{1 - \tilde{\mathbb{E}}[\ell]\theta^3} + \frac{q^2(q+1)\tilde{\mathbb{E}}[\ell(\ell-1)]^2\theta^{10}}{1 - \tilde{\mathbb{E}}[\ell]\theta^4}. \tag{103}$$

In the most interesting case $q = 4$ where the coefficient of a^2 vanishes, one finds, for w [substituting $\theta = \mathrm{sgn}(\theta)/\sqrt{\tilde{\mathbb{E}}[\ell]}$ at lowest order around the Kesten-Stigum transition],

$$w = 16 \left[-\frac{7}{3}\frac{\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]}{\tilde{\mathbb{E}}[\ell]^3} + \left(\frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{\tilde{\mathbb{E}}[\ell]^2} \right)^2 \left(\frac{5}{\tilde{\mathbb{E}}[\ell] - 1} - \frac{12}{\mathrm{sgn}(\theta)\sqrt{\tilde{\mathbb{E}}[\ell]} - 1} \right) \right], \tag{104}$$

which justifies the expression (32) given in Sec. III D 1 [the factor 16 between (32) and (104) is irrelevant for the study of the sign of w]. Depending on the degree distribution $\tilde{p}_\ell$ and on the sign of θ, this coefficient can be positive or negative; as the equation for a reduces here to $0 = \epsilon + wa^2$, with $\epsilon = \tilde{\mathbb{E}}[\ell]\theta^2 - 1$, the sign of w controls the sign of ϵ for which a nontrivial solution of the cavity equations exists in a neighborhood of the trivial one, hence the type of bifurcation scenario as sketched in Fig. 1. As mentioned in Sec. III D 1, one has $w < 0$ in the ferromagnetic case for all degree distributions, which yields a second-order transition; it is indeed easy to check numerically that

$$\frac{5}{d-1} - \frac{12}{\sqrt{d} - 1} < 0 \tag{105}$$

for all average offspring degrees $d = \tilde{\mathbb{E}}[\ell] > 1$.

In the antiferromagnetic case the sign of w depends on the degree distribution. For instance, when $\tilde{p}_\ell$ is a Poisson law of parameter d, the expression of w above [Eq. (104)] can be

simplified to

$$w = 16 \left(-\frac{7}{3} + \frac{5}{d-1} + \frac{12}{\sqrt{d} + 1} \right). \tag{106}$$

A numerical study of this function reveals the existence of a critical degree $d_c \approx 22.2694$ such that $w > 0$ for $d \in (1, d_c)$, hence a discontinuous bifurcation scenario similar to the case $q > 4$, while for $d > d_c$ one has $w < 0$, yielding the scenario of $q < 4$. We have also evaluated the sign of w in the regular case, i.e., for an offspring degree distribution $\tilde{p}_\ell = \delta_{\ell,d}$. Similarly to the Poisson case, one finds (for the antiferromagnetic model) that $w < 0$ for $d \leqslant 24$ and $w > 0$ for $d \geqslant 25$.

Note that for large enough degrees ferromagnetic and antiferromagnetic models behave in the same way; this could be anticipated from the study of the large-degree limit recalled in Sec. IV F, as in this limit the degree distribution and θ only appear in the combination $\tilde{\mathbb{E}}[\ell]\theta^2$, which is clearly independent of the sign of θ. Actually an expansion to the same order as in Eq. (102) can be found in Eq. (4.9) of [22], but in the large-degree limit and not in the sparse regime,

which did not allow one to deduce the modification of the type of phase transition according to the degree distribution.

Let us finally justify the scaling exponent we used in the plot presented in Fig. 3(b). If we had included in the expansion of the cavity equations around the trivial fixed point one more term, one would have found, for $q = 4$, an equation for a of the form

$$0 = \epsilon + w(\tilde{p})a^2 + xa^3, \tag{107}$$

where *a priori* $x \neq 0$ and we emphasized the dependence of w on the degree distribution. Let us consider for concreteness $\tilde{p}$ to be Poissonian with average d in such a way that $w(d)$ is given explicitly by (106). In the neighborhood of the critical degree d_c where w changes sign we have $w(d) \approx w'(d - d_c)$, with w' some constant. Inserting this form in Eq. (106) gives

$$0 = \epsilon + w'(d - d_c)a^2 + xa^3; \tag{108}$$

the scaling of θ_{sp}, which corresponds to the limit of existence of a solution of this equation, must be such that the three terms in Eq. (108) are of the same order. This is easily seen to imply that ϵ is of the order of $(d - d_c)^3$, which is the reason for our choice of exponent in Fig. 3(b): The spinodal and IT curves should, in this rescaled units, behave linearly in the neighborhood of d_c.

I. Application 3: The asymmetric Ising ($q = 2$) case

We consider now the case of binary variables ($q = 2$), which for convenience we define as Ising spins or sign variables $\sigma \in \{-1, 1\} = \{-, +\}$. We parametrize the stationary probability distribution $\overline{\eta}$ as $\overline{\eta}_\sigma = \frac{1 + \sigma \overline{m}}{2}$, with $\overline{m} \in [-1, 1]$. This bias parameter can be interpreted as a magnetization in the perspective of a physical Ising model [20] or in the context of the SBM as controlling the relative size of the two communities of vertices which are $\overline{\eta}_+ = \frac{1 + \overline{m}}{2}$ and $\overline{\eta}_- = \frac{1 - \overline{m}}{2}$. The symmetric case, as studied in Sec. IV H, is recovered for $\overline{m} = 0$. We write the matrix M as

$$
\begin{aligned}
M &= \begin{pmatrix} M_{++} & M_{+-} \\ M_{-+} & M_{--} \end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} 1 + \overline{m} & 1 - \overline{m} \\ 1 + \overline{m} & 1 - \overline{m} \end{pmatrix} + \theta\frac{1}{2}\begin{pmatrix} 1 - \overline{m} & -1 + \overline{m} \\ -1 - \overline{m} & 1 + \overline{m} \end{pmatrix},
\end{aligned}
\tag{109}
$$

i.e., $M_{\sigma\tau} = \frac{1 + \tau\overline{m}}{2} + \sigma\tau\theta\frac{1 - \sigma\overline{m}}{2}$. This is indeed the only Markov matrix reversible with respect to $\overline{\eta}$ with the second eigenvalue θ. To connect with the definition of the SBM in Sec. III A 1 the parameter θ is defined as $\theta = 1 - c_{-+}/d$, where $c_{-+} = c_{+-}$ is N times the probability that nodes in different groups are connected and d is the average degree. Given the group sizes $\overline{\eta}_\pm$, the condition of equal average degree in the two groups then implies uniquely the parameters c_{--} and c_{++}. In the context of the SBM the degree distributions p_ℓ and $\tilde{p}_\ell$ are Poissonian with average d; we will, however, consider generic degree distributions, as they are relevant from the tree reconstruction perspective.

The range of allowed values of the parameters (m, θ) is a subset of $[-1, 1]^2$: Imposing the non-negativity of the matrix

elements of M implies indeed that

$$\theta \geqslant -\frac{1 - |\overline{m}|}{1 + |\overline{m}|}, \tag{110}$$

which is always fulfilled for $\theta \geqslant 0$ but restricts the range of allowed biases $\overline{m}$ when $\theta < 0$. The transformed version $\hat{M}$ of M defined in Eq. (47) reads thus

$$\hat{M}_{\sigma\tau} = M_{\sigma\tau}\frac{1}{\overline{\eta}_\tau} = 1 + \theta\sigma\tau\frac{1 - \sigma\overline{m}}{1 + \tau\overline{m}}. \tag{111}$$

Due to the binary nature of the variables, the probability laws η can be parametrized by a single real and the recursion equation (34) can be rewritten in a simpler form with this parametrization. This expression, along with some more technical details on the numerical resolution of the cavity equations, can be found in Appendix A.

1. Critical asymmetry

Let us now apply the generic results of the moment expansions presented in Sec. IV D to this specific case. Because of the normalization condition $\delta_+ + \delta_- = 0$, the deviation $\delta_\sigma = \eta_\sigma - \overline{\eta}_\sigma$ is equal to σ multiplied by a scalar random variable. We can thus write the matrix $A^{(n)}_{\sigma\tau} = \mathbb{E}^{(n)}[\delta_\sigma \delta_\tau]$ as $A^{(n)}_{\sigma\tau} = a^{(n)}\frac{1}{2}\sigma\tau$, where $a^{(n)}$ is a scalar; the factor $1/2$ is chosen in such a way that $\frac{1}{2}\sigma\tau$ coincides with the matrix $K_{\sigma\tau}$ of (98), in the case $q = 2$. Once again we have a positivity requirement on this covariance, $a^{(n)} \geqslant 0$. From the expression of $\hat{M}$ given in Eq. (111) one sees that

$$\sum_{\sigma\sigma'}\hat{M}_{\sigma\sigma'}\sigma' = \theta\frac{1}{\overline{\eta}_\sigma}\sigma, \tag{112}$$

hence with the relation (58) one obtains

$$\hat{A}^{(n)}_{\sigma\tau} = \theta^2 a^{(n)}\frac{1}{\overline{\eta}_\sigma\overline{\eta}_\tau}\frac{1}{2}\sigma\tau. \tag{113}$$

The linear evolution of (59) thus becomes

$$a^{(n+1)}\frac{1}{2}\sigma\tau = \tilde{\mathbb{E}}[\ell]\theta^2 a^{(n)}\frac{1}{2}\sigma\tau, \tag{114}$$

i.e., $a^{(n+1)} = \tilde{\mathbb{E}}[\ell]\theta^2 a^{(n)}$, which reproduces the expected Kesten-Stigum threshold at $\tilde{\mathbb{E}}[\ell]\theta^2 = 1$.

We now look for a fixed-point solution in the neighborhood of the Kesten-Stigum transition, by injecting the above forms of A and $\hat{A}$ in Eq. (65). This yields, after a short computation,

$$a = \tilde{\mathbb{E}}[\ell]\theta^2 a + \tilde{\mathbb{E}}[\ell(\ell - 1)]\theta^4\frac{2(3\overline{m}^2 - 1)}{(1 - \overline{m}^2)^2}a^2. \tag{115}$$

A simple consistency check is provided by the coincidence of (99) and (115) when $q = 2$ and $\overline{m} = 0$. Moreover, when $q = 2$ the nontrivial eigenvalue of M is unique, hence certainly nondegenerate; one can check that (115) is also a consequence of Eq. (94) derived in Sec. IV G under this nondegeneracy assumption for all $\overline{m}$.

The crucial property of (115) we would like to emphasize is the change of sign of the coefficient of a^2 depending on whether $|\overline{m}|$ is larger or smaller than $\overline{m}_c = 1/\sqrt{3}$. Because of the condition $a \geqslant 0$ this implies that the nontrivial perturbative solution exists when $\tilde{\mathbb{E}}[\ell]\theta^2 > 1$ at small asymmetry ($|\overline{m}| < \overline{m}_c$) and when $\tilde{\mathbb{E}}[\ell]\theta^2 < 1$ at large asymmetry ($\overline{m} > \overline{m}_c$). In the latter case one has, in the tree reconstruction
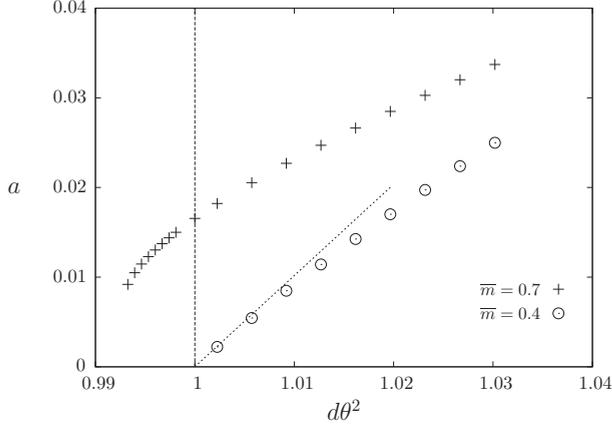
FIG. 7. Value of a in the fixed-point solution of the cavity equations reached from an informative initial condition, as a function of the signal-to-noise ratio, for the ferromagnetic ($\theta > 0$) asymmetric Ising model. We used a regular degree distribution $\tilde{p}_\ell = \delta_{\ell,d}$ with $d = 3$, each point corresponding to a different value of $\theta > 0$. The two sets of symbols correspond to $\overline{m} = 0.4 < \overline{m}_c$ and $\overline{m} = 0.7 > \overline{m}_c$. The vertical dashed line indicates the location of the Kesten-Stigum transition, while the inclined line is our analytical prediction from (115) for $\overline{m} = 0.4$.

language, a reconstructible phase below the Kesten-Stigum threshold, which is not tight in this case. Quite strikingly, the critical asymmetry $\overline{m}_c$ does not depend on the degree distribution; this explains why this value of $1/\sqrt{3}$ was also obtained previously in the large-degree limit [27,28] and in related dense inference problems [16,26]. Very recently this critical asymmetry was also discovered for the reconstruction of the Ising model on regular trees [29]; this paper proved rigorously the nontightness of the Kesten-Stigum bound for $\overline{m} > \overline{m}_c$ and its tightness for $\overline{m} < \overline{m}_c$ at large enough degrees.

As an illustration of this phenomenon we present in Fig. 7 the results of a numerical resolution of the cavity equations for two asymmetries below and above $\overline{m}_c$. One can clearly see that in the latter case reconstruction is possible below the Kesten-Stigum threshold and that in the former case our analytical prediction (115) is in agreement with the numerical results.

2. Further expansions around the critical asymmetry

Let us now further describe the phase diagram of the problem in the $(\theta, \overline{m})$ plane (for simplicity of the discussion we

assume the degree distribution to be held fixed) (see Fig. 4 for an example). In the low-asymmetry ($|\overline{m}| < \overline{m}_c$) situation there are only two phases, the tree reconstruction being possible if and only $\tilde{\mathbb{E}}[\ell]\theta^2 > 1$. This tightness of the Kesten-Stigum bound at small nonzero asymmetry was proven in Ref. [20], but that paper did not estimate the value of $\overline{m}_c$. In terms of the SBM inference problem, the Kesten-Stigum transition separates an easy phase from an information-theoretic impossible phase; this was proven in the symmetric ($\overline{m} = 0$) case in Refs. [42,43].

The large-asymmetry ($|\overline{m}| > \overline{m}_c$) part of the phase diagram is richer. In terms of the tree reconstruction problem there will be two lines of transition $\theta_{\mathrm{sp},\pm}(\overline{m})$ such that reconstruction is possible for $\theta < \theta_{\mathrm{sp},-}(\overline{m}) < 0$ and $\theta > \theta_{\mathrm{sp},+}(\overline{m}) > 0$, with $\tilde{\mathbb{E}}[\ell]\theta_{\mathrm{sp},\pm}(\overline{m})^2 < 1$; the subscript $\pm$ indicates the ferromagnetic ($\theta > 0$) or antiferromagnetic ($\theta < 0$) part of the phase diagram. Note that the average offspring degree $\tilde{\mathbb{E}}[\ell]$ has to be larger than $(2+\sqrt{3})/(2-\sqrt{3}) \approx 13.928$ for the large asymmetry antiferromagnetic part of the phase diagram to be nonempty, because of the condition (110). This discontinuous transition of the tree reconstruction problem, depicted in Fig. 1(d), is not directly relevant for the SBM graph inference problem. The latter has an easy phase for $\tilde{\mathbb{E}}[\ell]\theta^2 > 1$ and a hard phase for $\theta < \theta_{\mathrm{IT},-}(\overline{m}) < 0$ and $\theta > \theta_{\mathrm{IT},+}(\overline{m}) > 0$, with $\tilde{\mathbb{E}}[\ell]\theta_{\mathrm{IT},\pm}(\overline{m})^2 \in [\tilde{\mathbb{E}}[\ell]\theta_{\mathrm{sp},\pm}(\overline{m})^2, 1]$, while the inference is information-theoretically impossible for $\theta \in [\theta_{\mathrm{IT},-}(\overline{m}), \theta_{\mathrm{IT},+}(\overline{m})]$. This information-theoretic (IT) line is defined by the vanishing of the free entropy computed in the nontrivial solution of the cavity equations.

In the remainder of this section we will describe more precisely the neighborhood of the (ferromagnetic and antiferromagnetic when they exist) points $(\theta, \overline{m}) = (\theta_{\mathrm{KS}}, \overline{m}_c)$ of the phase diagram and in particular present an analytic description of the lines $\theta_{\mathrm{sp},\pm}(\overline{m})$ and $\theta_{\mathrm{IT},\pm}(\overline{m})$ when $|\overline{m}| \to \overline{m}_c^+$. To reach this goal we will exploit the next order in our generic moment expansions, as summarized in Eqs. (71)–(76). We first note that for symmetry reasons the tensors B and C depend on their spin indices as $B_{\sigma\tau\gamma} = b\sigma\tau\gamma$ and $C_{\sigma\tau\gamma\beta} = c\sigma\tau\gamma\beta$, with b and c two reals. The form of the matrix $\hat{M}$ given in Eq. (111) leads to

$$\hat{B}_{\sigma\tau\gamma} = \theta^3 b \frac{1}{\overline{\eta}_\sigma \overline{\eta}_\tau \overline{\eta}_\gamma} \sigma\tau\gamma, \quad \hat{C}_{\sigma\tau\gamma\beta} = \theta^4 c \frac{1}{\overline{\eta}_\sigma \overline{\eta}_\tau \overline{\eta}_\gamma \overline{\eta}_\beta} \sigma\tau\gamma\beta.$$

(116)

Plugging these expressions into (71)–(73) leads, after a short computation, to a set of equations for a, b, and c, namely,

$$a = \tilde{\mathbb{E}}[\ell]\theta^2 a + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{2(3\overline{m}^2-1)}{(1-\overline{m}^2)^2}\theta^4 a^2 + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{32\overline{m}}{(1-\overline{m}^2)^2}\theta^5 ab + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{16}{(1-\overline{m}^2)^2}\theta^6 ac$$

$$+ \tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\frac{4}{3}\frac{5-42\overline{m}^2+45\overline{m}^4}{(1-\overline{m}^2)^4}\theta^6 a^3,$$

(117)

$$b = \tilde{\mathbb{E}}[\ell]\theta^3 b - \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{3\overline{m}}{1-\overline{m}^2}\theta^4 a^2,$$

(118)

$$c = \tilde{\mathbb{E}}[\ell]\theta^4 c + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{3}{4}\theta^4 a^2.$$

(119)

Alternatively, these equations can be obtained via the computation of the free entropy (77), which is found to be

$$\frac{1}{\mathbb{E}[\ell]}\phi(a,b,c) = \frac{\tilde{\mathbb{E}}[\ell]\theta^2 - 1}{(1-\overline{m}^2)^2}\theta^2 a^2 - \frac{16}{3}\frac{\tilde{\mathbb{E}}[\ell]\theta^3 - 1}{(1-\overline{m}^2)^3}\theta^3 b^2 + \frac{32}{3}\frac{\tilde{\mathbb{E}}[\ell]\theta^4 - 1}{(1-\overline{m}^2)^4}\theta^4 c^2 + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{4}{3}\frac{3\overline{m}^2 - 1}{(1-\overline{m}^2)^4}\theta^6 a^3$$

$$+ \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{32\overline{m}}{(1-\overline{m}^2)^4}\theta^7 a^2 b + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{16}{(1-\overline{m}^2)^4}\theta^8 a^2 c$$

$$+ \tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\frac{2}{3}\frac{5 - 42\overline{m}^2 + 45\overline{m}^4}{(1-\overline{m}^2)^6}\theta^8 a^4; \tag{120}$$

its derivatives with respect to a, b, and c do indeed vanish when Eqs. (117)–(119) are fulfilled.

Equations (118) and (119) can be immediately solved to obtain b and c as a function of a; reinjecting these results in Eq. (117), one obtains a quadratic equation for a,

$$0 = u + va + wa^2, \tag{121}$$

with

$$u(\theta) = \tilde{\mathbb{E}}[\ell]\theta^2 - 1, \tag{122}$$

$$v(\theta, \overline{m}) = \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{2(3\overline{m}^2 - 1)}{(1-\overline{m}^2)^2}\theta^4, \tag{123}$$

$$w(\theta, \overline{m}) = \tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\frac{4}{3}\frac{5 - 42\overline{m}^2 + 45\overline{m}^4}{(1-\overline{m}^2)^4}\theta^6 \tag{124}$$

$$+ \tilde{\mathbb{E}}[\ell(\ell-1)]^2\frac{12\theta^9}{(1-\overline{m}^2)^2}\left(-\frac{8\overline{m}^2}{1-\overline{m}^2}\frac{1}{1-\mathbb{E}[\ell]\theta^3} + \frac{\theta}{1-\mathbb{E}[\ell]\theta^4}\right). \tag{125}$$

We want to study this equation in the neighborhood of the point $(\theta_{\text{KS}}, \overline{m}_{\text{c}})$, taking simultaneously the limits $\theta \to \theta_{\text{KS}}$ and $\overline{m} \to \overline{m}_{\text{c}}$; we define as before $\epsilon = \tilde{\mathbb{E}}[\ell]\theta^2 - 1$. The coefficients $u = \epsilon$ and v will thus both be small in this regime, while $\tilde{w} = w(\theta_{\text{KS}}, \overline{m}_{\text{c}}) \neq 0$. In order to have the three terms in Eq. (121) of the same order, one realizes that the two simultaneous limits must be taken with $\epsilon = t(\overline{m} - \overline{m}_{\text{c}})^2$, where t is finite and constitutes the relevant control parameter in this scaling regime. Then, defining $v(\theta_{\text{KS}}, \overline{m}) = \tilde{v}(\overline{m} - \overline{m}_{\text{c}}) + O((\overline{m} - \overline{m}_{\text{c}})^2)$, we reduce (121) at lowest order to

$$0 = t(\overline{m} - \overline{m}_{\text{c}})^2 + \tilde{v}(\overline{m} - \overline{m}_{\text{c}})a + \tilde{w}a^2. \tag{126}$$

Defining a reduced unknown $\tilde{a} = \frac{a}{\overline{m} - \overline{m}_{\text{c}}}$, which will be finite in the limit, we obtain finally

$$0 = t + \tilde{v}\tilde{a} + \tilde{w}\tilde{a}^2, \tag{127}$$

hence $\tilde{a} = -\frac{\tilde{v}}{2\tilde{w}} \pm \frac{1}{2\tilde{w}}\sqrt{\tilde{v}^2 - 4\tilde{w}t}$. The coefficients $\tilde{v}$ and $\tilde{w}$ only depend on the degree distribution; to determine them at dominating order one replaces θ by its value at the Kesten-Stigum transition, i.e., $\text{sgn}(\theta)/\sqrt{\tilde{\mathbb{E}}[\ell]}$, and obtains

$$\tilde{v} = 9\sqrt{3}\frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{\tilde{\mathbb{E}}[\ell]^2}, \tag{128}$$

$$\tilde{w} = -27\frac{\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]}{\tilde{\mathbb{E}}[\ell]^3} - 27\left(\frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{\tilde{\mathbb{E}}[\ell]^2}\right)^2$$

$$\times \left(\frac{4}{\text{sgn}(\theta)\sqrt{\tilde{\mathbb{E}}[\ell]} - 1} - \frac{1}{\tilde{\mathbb{E}}[\ell] - 1}\right). \tag{129}$$

It is rather easy to check that $\tilde{w} < 0$ in the ferromagnetic case [$\text{sgn}(\theta) > 0$] for all degree distributions with $\tilde{\mathbb{E}}(\ell) > 1$ (a

necessary condition for the Galton-Watson tree to be infinite with positive probability and for the Kesten-Stigum transition to exist). We believe that $\tilde{w} < 0$ also in the antiferromagnetic case whenever the point $(-1/\sqrt{\tilde{\mathbb{E}}[\ell]}, \overline{m}_{\text{c}})$ belongs to the authorized domain of parameters according to the condition (110) for all offspring degree distributions, but could only check it explicitly in the regular and Poisson cases. Note that in the large-degree limit both coefficients $\tilde{v}$ and $\tilde{w}$ remains finite, namely, $\tilde{v} \to 9\sqrt{3}$ and $\tilde{w} \to -27$.

As a consequence of $\tilde{w} < 0$, the real solutions of (127) exist for $t \geqslant t_{\text{sp}} = \tilde{v}^2/4\tilde{w}$. This is precisely the condition that defines the line θ_{sp}; translating back from rescaled units yields the lowest-order expansion of $\theta_{\text{sp}}(\overline{m})$ in the limit $\overline{m} \to \overline{m}_{\text{c}}^+$,

$$\theta_{\text{sp}}(\overline{m}) = \theta_{\text{KS}}\left(1 - \frac{\tilde{v}^2}{8|\tilde{w}|}(\overline{m} - \overline{m}_{\text{c}})^2 + o((\overline{m} - \overline{m}_{\text{c}})^2)\right). \tag{130}$$

Note that $\tilde{w}$ depends on $\text{sgn}(\theta)$, giving two distinct expansions for the ferromagnetic and antiferromagnetic lines $\theta_{\text{sp},+}$ and $\theta_{\text{sp},-}$.

Using the expression (120) of ϕ, one can also expand the IT line θ_{IT} around $(\theta, \overline{m}) = (\theta_{\text{KS}}, \overline{m}_{\text{c}})$. Indeed, in the scaling regime described above one finds that

$$\phi = \frac{9}{2}\frac{\mathbb{E}[\ell]}{\tilde{\mathbb{E}}[\ell]}(\overline{m} - \overline{m}_{\text{c}})^4\left(\frac{1}{2}t\tilde{a}^2 + \frac{1}{3}\tilde{v}\tilde{a}^3 + \frac{1}{4}\tilde{w}\tilde{a}^4\right). \tag{131}$$

The IT line corresponds to $\phi = 0$ (this is the value on the trivial fixed point), hence the IT threshold corresponds to the
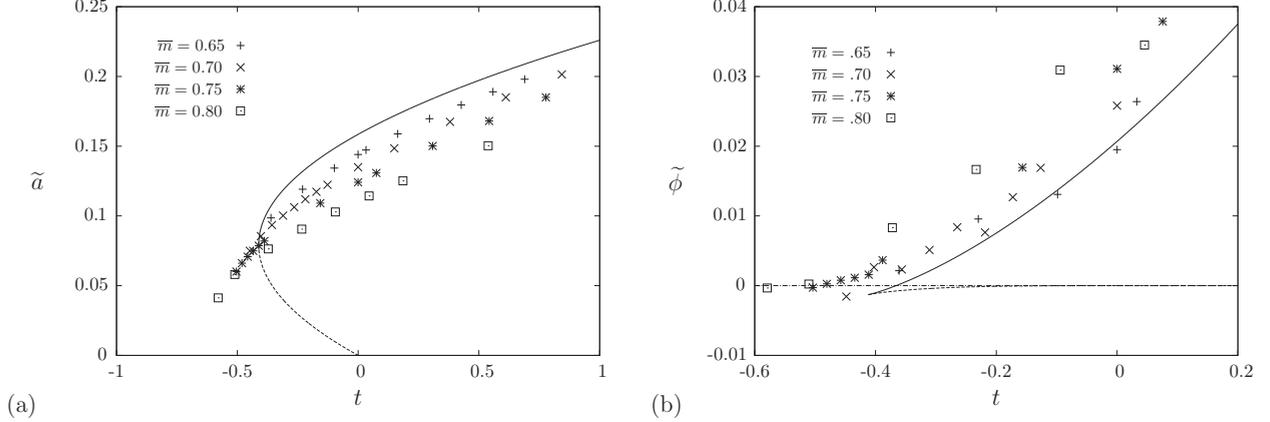
FIG. 8. Study of the scaling regime in the neighborhood of the point $(\theta_{KS}, \overline{m}_c)$ for the ferromagnetic asymmetric Ising model with degree distribution $\tilde{p}_\ell = \delta_{\ell,d}$, $d = 3$. The rescaled signal-to-noise ratio parameter is $t = (d\theta^2 - 1)/(\overline{m} - \overline{m}_c)^2$; the Kesten-Stigum transition thus corresponds to $t = 0$. (a) Rescaled accuracy $\tilde{a} = a/(\overline{m} - \overline{m}_c)$ as a function of t. The lines are the analytic predictions for the stable (solid) and unstable (dashed) branches of fixed points obtained in Eq. (127) and the symbols are numerical results that approach the analytic line when $\overline{m} \to \overline{m}_c^+$. (b) Rescaled free entropy $\tilde{\phi} = \phi/(\overline{m} - \overline{m}_c)^4$ as a function of t. The analytic lines have been obtained by inserting in Eq. (131) the two branches of $\tilde{a}$ from (127); the threshold t_{IT} corresponds to the crossing with the horizontal line $\tilde{\phi} = 0$.

solution of

$$0 = t + \tilde{v}\tilde{a} + \tilde{w}\tilde{a}^2,$$

$$0 = \frac{1}{2}t + \frac{1}{3}\tilde{v}\tilde{a} + \frac{1}{4}\tilde{w}\tilde{a}^2 \Rightarrow \tilde{a} = -\frac{2\tilde{v}}{3\tilde{w}}, \qquad (132)$$

$$t = \frac{2\tilde{v}^2}{9\tilde{w}}.$$

The IT threshold $t_{IT} = \frac{2\tilde{v}^2}{9\tilde{w}}$ is thus distinct at this order from the spinodal one $t_{sp} = \frac{\tilde{v}^2}{4\tilde{w}}$ (even though the difference in the coefficient is small). Translating back in terms of the parameters $(\theta, \overline{m})$ gives

$$\theta_{IT}(\overline{m}) = \theta_{KS}\left(1 - \frac{\tilde{v}^2}{9|\tilde{w}|}(\overline{m} - \overline{m}_c)^2 + o((\overline{m} - \overline{m}_c)^2)\right), \qquad (133)$$

with again two different expressions for the ferromagnetic and antiferromagnetic transitions.

These two expansions (130) and (133) on the behavior of the spinodal and IT lines in the neighborhood of the critical asymmetry where the transition crosses over from second to first order constitute our main results for the asymmetric Ising model. We have shown in Fig. 4(b) that they are in agreement with our numerical data, within the accuracy we could reach. As a further illustration we show in Fig. 8 the numerical determination of a and ϕ in the scaling regime and compare it to our analytical formulas.

J. Application 4: The $q_1 + q_2$ case

We now turn our attention to the model introduced in Sec. III D 3, which breaks the symmetry between q labels in a minimal way, generalizing the two previous cases (i.e., q symbols with maximal symmetry studied in Sec. IV H and $q = 2$ without any symmetry in Sec. IV I).

Let us consider indeed an alphabet of q labels $\chi = \{1, \dots, q\}$, divided into two supergroups (or supercommuni-

ties) G_1 and G_2 containing, respectively, q_1 and q_2 symbols (in such a way that $q = q_1 + q_2$), namely, $G_1 = \{1, \dots, q_1\}$ and $G_2 = \{q_1 + 1, \dots, q\}$. We break the symmetry among the q labels in a minimal way, according to this subdivision in two groups, by taking $\overline{\eta}_\sigma$ constant in G_1 and G_2: $\overline{\eta}_\sigma = \overline{\eta}_1$ if $\sigma \in G_1$ and $\overline{\eta}_\sigma = \overline{\eta}_q$ if $\sigma \in G_2$, with the normalization condition $q_1\overline{\eta}_1 + q_2\overline{\eta}_q = 1$. We parametrize the fraction of vertices in G_1 and G_2 by $\overline{m} \in [-1, 1]$, according to

$$\overline{\eta}_\sigma = \begin{cases} \frac{1}{q_1}\frac{1+\overline{m}}{2} & \text{if } \sigma \in G_1 \\ \frac{1}{q_2}\frac{1-\overline{m}}{2} & \text{if } \sigma \in G_2; \end{cases} \qquad (134)$$

this notation is reminiscent of the asymmetric Ising case, which would be obtained by coarse graining the labels $\sigma \in G_1$ ($\sigma \in G_2$) as $\sigma = +$ ($\sigma = -$). We also perform this minimal symmetry breaking on the matrix M, assuming that its matrix elements $M_{\sigma\tau}$ only depend on the group to which σ and τ belong and on whether $\sigma = \tau$ or not (the shape of the associated connectivity matrix in the SBM interpretation is sketched in Fig. 5).

Imposing in addition the reversibility of the stochastic matrix M with respect to $\overline{\eta}$ (this condition corresponds to the equality of the average degrees in G_1 and G_2 in the SBM interpretation), one realizes that such a matrix M is necessarily of the form

$$M_{\sigma\tau} = \frac{1+\overline{m}}{2q_1}\mathbb{I}(\tau \in G_1)$$

$$+ \frac{1-\overline{m}}{2q_2}\mathbb{I}(\tau \in G_2) + \mu_1\mathbb{I}(\sigma, \tau \in G_1)\left(\delta_{\sigma,\tau} - \frac{1}{q_1}\right)$$

$$+ \mu_2\mathbb{I}(\sigma, \tau \in G_2)\left(\delta_{\sigma,\tau} - \frac{1}{q_2}\right)$$

$$+ \mu_0\left(\frac{1-\overline{m}}{2}\mathbb{I}(\sigma \in G_1) - \frac{1+\overline{m}}{2}\mathbb{I}(\sigma \in G_2)\right)$$

$$\times \left(\frac{1}{q_1}\mathbb{I}(\tau \in G_1) - \frac{1}{q_2}\mathbb{I}(\tau \in G_2)\right). \qquad (135)$$

One can check that M admits 1 as an eigenvalue, with eigenvectors $\bar{\eta}$ on the left and constant vector on the right; the parameters μ_0, μ_1, and μ_2 introduced in Eq. (135) correspond to the nontrivial eigenvalues of M. There are $q_1 - 1$ ($q_2 - 1$) degenerate eigenvalues μ_1 (μ_2) with eigenvectors supported on G_1 (G_2) and a simple eigenvalue μ_0 whose eigenvector is constant inside each of the two groups. For a given choice of $\bar{m} \in [-1, 1]$ the requirement of positivity of the matrix elements of M restricts the allowed domain of the parameters (μ_0, μ_1, μ_2) to the subset of $[-\frac{1-|\bar{m}|}{1+|\bar{m}|}, 1] \times [-\frac{1}{q_1-1}, 1] \times [-\frac{1}{q_2-1}, 1]$ that fulfills the condition

$$\mu_0 \geqslant \max\left(\frac{2\mu_1 - 1 - \bar{m}}{1 - \bar{m}}, \frac{-2(q_1 - 1)\mu_1 - 1 - \bar{m}}{1 - \bar{m}}, \frac{2\mu_2 - 1 + \bar{m}}{1 + \bar{m}}, \frac{-2(q_2 - 1)\mu_2 - 1 + \bar{m}}{1 + \bar{m}}\right). \tag{136}$$

One sees that the relevant eigenvalue for the Kesten-Stigum transition, i.e., the maximal one in absolute value, can be either μ_1, μ_2, or μ_0 depending on the choice of the parameters. In other words, the signal on the labels transmitted through the edges of the SBM can be stronger inside G_1, inside G_2, or in the difference of behavior between G_1 and G_2.

We will now specialize the equations of Sec. IV D to this particular case. We first have to describe the matrices $A_{\sigma\tau} = \mathbb{E}[\delta_\sigma \delta_\tau]$. Because of the permutation invariance inside each of the two groups of labels, the matrix element $A_{\sigma\tau}$ should only depend on whether $\sigma = \tau$ or not and on the group of the two communities σ and τ. In addition, the matrix A has to be symmetric and the sum of every row or column has to vanish. A moment of thought reveals that the vector space of such matrices is three dimensional and is spanned by the three matrices K_1, K_2, and K_0 defined as

$$(K_1)_{\sigma\tau} = \mathbb{I}(\sigma, \tau \in G_1)\left(\delta_{\sigma,\tau} - \frac{1}{q_1}\right),$$

$$(K_2)_{\sigma\tau} = \mathbb{I}(\sigma, \tau \in G_2)\left(\delta_{\sigma,\tau} - \frac{1}{q_2}\right), \tag{137}$$

$$(K_0)_{\sigma\tau} = \frac{1}{qq_1q_2}[q_2\mathbb{I}(\sigma \in G_1) - q_1\mathbb{I}(\sigma \in G_2)][q_2\mathbb{I}(\tau \in G_1) - q_1\mathbb{I}(\tau \in G_2)]. \tag{138}$$

These matrices, which are defined similarly to the matrix K of the symmetric Potts model introduced in Eq. (98), are linearly independent and obey in addition the simple algebra

$$K_1^2 = K_1, \quad K_2^2 = K_2, \quad K_0^2 = K_0, \quad K_i K_j = 0 \quad \text{if } i \neq j. \tag{139}$$

We can thus parametrize $A^{(n)}$ as $a_1^{(n)} K_1 + a_2^{(n)} K_2 + a_0^{(n)} K_0$; the nonzero eigenvalues of $A^{(n)}$ are then found to be $a_1^{(n)}$, $a_2^{(n)}$, and $a_0^{(n)}$. These three reals must thus be non-negative for $A^{(n)}$ to be semipositive definite.

The matrix $\hat{M}$ defined in Eq. (47) can then be expressed as

$$\hat{M}_{\sigma\tau} = 1 + \mu_1 \frac{1}{\bar{\eta}_1}(K_1)_{\sigma\tau} + \mu_2 \frac{1}{\bar{\eta}_q}(K_2)_{\sigma\tau} + \mu_0 q\bar{\eta}_1\bar{\eta}_q \frac{1}{\bar{\eta}_\sigma}(K_0)_{\sigma\tau}\frac{1}{\bar{\eta}_\tau} \tag{140}$$

and a simple computation based on the algebraic properties stated above reveals that

$$\bar{\eta}_\sigma \bar{\eta}_\tau \hat{A}_{\sigma\tau}^{(n)} = \left(\mu_1^2 a_1^{(n)} K_1 + \mu_2^2 a_2^{(n)} K_2 + \mu_0^2 a_0^{(n)} K_0\right)_{\sigma\tau}. \tag{141}$$

Hence the linearized evolution equation (59) yields

$$a_i^{(n+1)} = \tilde{\mathbb{E}}[\ell]\mu_i^2 a_i^{(n)} \quad \text{for} \quad i = 0, 1, 2. \tag{142}$$

The trivial fixed point $a_1 = a_2 = a_0 = 0$ becomes unstable as soon as one of the three coefficients a_i grows under these iterations, hence we recover the Kesten-Stigum criterion $\tilde{\mathbb{E}}[\ell]\max(\mu_1^2, \mu_2^2, \mu_0^2) = 1$ at the limit of stability of the trivial fixed point, the parameters μ_i being the nontrivial eigenvalues of M.

Following our usual program, we incorporate the next-order correction and look for a perturbative nontrivial fixed point around the Kesten-Stigum transition. Inserting the form given above for the matrices A and $\hat{A}$ into the generic equation (65) yields, after some computations, the following system of quadratic equations for a_0, a_1, and a_2:

$$a_1 = \tilde{\mathbb{E}}[\ell]\mu_1^2 a_1 + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{2q_1}{(1+\bar{m})^2}\left((q_1 - 3 - \bar{m})\mu_1^4 a_1^2 + \frac{2q_2}{q}\mu_0^2\mu_1^2 a_0 a_1\right), \tag{143}$$

$$a_2 = \tilde{\mathbb{E}}[\ell]\mu_2^2 a_2 + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{2q_2}{(1-\bar{m})^2}\left((q_2 - 3 + \bar{m})\mu_2^4 a_2^2 + \frac{2q_1}{q}\mu_0^2\mu_2^2 a_0 a_2\right), \tag{144}$$

$$a_0 = \tilde{\mathbb{E}}[\ell]\mu_0^2 a_0 + \tilde{\mathbb{E}}[\ell(\ell-1)]\left(\frac{4q_1q_2}{q}\frac{3\bar{m}^2 - 1}{(1-\bar{m}^2)^2}\mu_0^4 a_0^2 + \frac{qq_1(q_1-1)}{2q_2}\frac{(1-\bar{m})^2}{(1+\bar{m})^2}\mu_1^4 a_1^2 + \frac{qq_2(q_2-1)}{2q_1}\frac{(1+\bar{m})^2}{(1-\bar{m})^2}\mu_2^4 a_2^2\right). \tag{145}$$

From this system of equations one can recover, as a consistency check, the equations (99) of the symmetric q-state model and (115) of the asymmetric Ising case. The latter is indeed obtained with $q_1 = q_2 = 1$, in which case $K_1 = K_2 = 0$ and the only parameter is a_0, which obeys indeed (115) with the identification $\mu_0 = \theta$. To recover the former case one takes $\mu_1 = \mu_2 = \mu_0 = \theta$, $a_1 = a_2 = a_0$, and $\overline{m} = (q_1 - q_2)/2$; then the three equations (143)–(145) reduce to (99). Note also that these equations coincide, in the special case $q_1 = q_2$, $\mu_1 = \mu_2$, and $\overline{m} = 0$, for which $a_1 = a_2$, with the moment recursions derived in Ref. [30].

Let us now return to an arbitrary choice of parameters in Eq. (143)–(145) and discuss the bifurcation of the nontrivial solution of this system of equations at the Kesten-Stigum transition. This discussion must be divided according to which eigenvalue μ_i causes the transition by fulfilling the condition $\tilde{\mathbb{E}}[\ell]\mu_i^2 = 1$ (we will assume for simplicity that only one among μ_1, μ_2, and μ_0 becomes critical).

1. Bifurcation driven by μ_0

Let us first consider the case where the bifurcation is driven by μ_0, i.e., where $\tilde{\mathbb{E}}[\ell]\mu_0^2 = 1 + \epsilon$ with ϵ small, while $\tilde{\mathbb{E}}[\ell]\mu_1^2 < 1$ and $\mathbb{E}[\ell]\mu_2^2 < 1$. Since the dominant direction of the bifurcation is a_0 one can simplify the system (143)–(145) at lowest order into a single equation for a_0:

$$0 = \epsilon + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{4q_1q_2}{q}\frac{3\overline{m}^2 - 1}{(1 - \overline{m}^2)^2}\mu_0^4 a_0. \qquad (146)$$

Recalling the positivity condition $a_0 \geqslant 0$, one realizes that the bifurcating solution exists above the Kesten-Stigum transition (i.e., for $\epsilon > 0$) if and only if the asymmetry between the two groups of labels is small enough, namely, if $\overline{m} < \overline{m}_c = 1/\sqrt{3}$, as in the asymmetric Ising case. As on the right-hand sides of (143) and (144) there are no terms proportional to a power of a_0 (without further multiplication by a_1 or a_2), the nonbifurcating unknowns a_1 and a_2 remain strictly equal to 0 in this solution.

2. Bifurcation driven by μ_1

Suppose now that μ_1 is the critical eigenvalue (the case where μ_2 becomes critical can be deduced from this one by exchanging the two groups), i.e., that $\tilde{\mathbb{E}}[\ell]\mu_1^2 = 1 + \epsilon$ with ϵ small, while $\tilde{\mathbb{E}}[\ell]\mu_2^2 < 1$ and $\tilde{\mathbb{E}}[\ell]\mu_0^2 < 1$. The lowest-order equation for the bifurcating direction a_1 thus becomes

$$0 = \epsilon + \tilde{\mathbb{E}}[\ell(\ell-1)]\frac{2q_1}{(1 + \overline{m})^2}(q_1 - 3 - \overline{m})\mu_1^4 a_1, \quad (147)$$

with the subdominant coefficient a_0 being $O(a_1^2) = O(\epsilon^2)$, while a_2 remains strictly zero [within the system (143)–(145)]. The sign of ϵ for which the nontrivial solution satisfies the constraint $a_1 \geqslant 0$ is thus the one of $-(q_1 - 3 - \overline{m})$. To analyze the sign of this quantity we can exclude the cases

where $\overline{m} = \pm 1$, as these reduce to purely symmetric models with either q_1 or q_2 labels. The type of bifurcation thus depends on the (integer) value of q_1 as follows.

(i) If $q_1 \geqslant 4$, for any value of $\overline{m}$, the bifurcating solution exists for $\epsilon < 0$, leading to the nontightness of the Kesten-Stigum bound for the reconstruction. This is the conclusion reached in Ref. [30], in the special case $q_1 = q_2$ and $\overline{m} = 0$.

(ii) If $q_1 \in \{1, 2\}$, for any value of $\overline{m}$, the continuous solution is present above the Kesten-Stigum transition (for $\epsilon > 0$).

(iii) If $q_1 = 3$ the scenario depends on the asymmetry parameter $\overline{m}$: For $\overline{m} < 0$ the bifurcating solution exists for $\epsilon < 0$, yielding a first-order transition with the nontightness of the Kesten-Stigum bound. On the other hand, if $\overline{m} > 0$ the nontrivial solution appears continuously in the large-SNR regime, above the Kesten-Stigum transition.

Note that this classification is independent of the number q_2 of labels in the group, which does not become critical at the Kesten-Stigum transition.

3. Higher-order terms and the existence of algorithmic spinodals

We have computed the next order in the expansion of (143)–(145), computing the coefficients of the terms that are cubic in the a_i, by specializing the generic equations (71)–(73) to the symmetry pattern of the $q_1 + q_2$ model. This requires in particular the determination of the most generic symmetric tensors with three and four indices, $B_{\sigma\tau\gamma}$ and $C_{\sigma\tau\gamma\beta}$, that are invariant under permutations of the labels inside G_1 and G_2. The resulting equations are rather long and hence we will not write them completely but concentrate on the additional predictions they led us to.

As long as the coefficient of the quadratic term in the equation for the bifurcating a_i is nonzero on the right-hand side of (143)–(145), these higher-order terms can only affect the solution quantitatively, but not qualitatively. Suppose first that the bifurcation is driven by μ_0, and that $\overline{m} = \overline{m}_c = 1/\sqrt{3}$, in such a way that the first nonlinear term in a_0 vanishes. In this case a_0 is the solution of a quadratic equation of the form $0 = \epsilon + wa_0^2$; our computation yields explicitly the value of this coefficient w, which turns out to be proportional (with a positive constant depending only on q_1 and q_2) to the corresponding expression found in the asymmetric Ising case and written in Eq. (129), with μ_1 playing the role of θ. We argued this coefficient to be always negative for all degree distributions and sign of μ_1; this case thus offers no novelty with respect to the situations investigated previously.

Suppose now that the bifurcation is driven by μ_1, and that $q_1 = 3$ and $\overline{m} = 0$, in such a way that the coefficient of a_1^2 in Eq. (143) vanishes. In this much more interesting case the leading-order bifurcation equation becomes $0 = \epsilon + wa_1^2$, with the following expression for the coefficient w (with $\tilde{\mathbb{E}}[\ell]\mu_1^2 = 1$ at lowest order):

$$w = 36\left[-3\frac{\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]}{\tilde{\mathbb{E}}[\ell]^3} + \left(\frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{\tilde{\mathbb{E}}[\ell]^2}\right)^2\left(\frac{4}{\tilde{\mathbb{E}}[\ell]-1} - \frac{12}{\mathrm{sgn}(\mu_1)\sqrt{\tilde{\mathbb{E}}[\ell]}-1} - \frac{2\mu_0}{1 - \mu_0} + \frac{\tilde{\mathbb{E}}[\ell]\mu_0^2}{1 - \tilde{\mathbb{E}}[\ell]\mu_0^2}\right)\right]. \quad (148)$$

The crucial point we want to emphasize is that w can be made positive for well-chosen values of the parameters of the model, in particular when $|\mu_0|$ is close to $|\mu_1|$ (but still strictly smaller for the bifurcation to be driven by μ_1). In such a case the nontrivial solution a_1 of the equation exists for $\epsilon < 0$, hence the Kesten-Stigum bound is not tight. Let us now argue that the situation depicted in Fig. 2 must occur in some part of the phase diagram of the $q_1 + q_2$ model; for simplicity, let us consider that the degree distribution is Poissonian with average c, which we take as the SNR. Suppose that $q_1 = 3$ and fix the values of the μ_i in such a way that $w > 0$ in the expression of (148). For $\overline{m} = 0$ the analysis above shows the existence of reconstruction down to $c_{sp}(\overline{m} = 0) < c_{KS}$, while $c_{alg}(\overline{m} = 0) = c_{KS}$ as the bifurcating solution exists only for $\epsilon < 0$. Suppose now that $\overline{m}$ is slightly increased to a small positive value; from (147) we conclude that a nontrivial perturbative solution exists for some range of $c > c_{KS}$, but by continuity with the situation for $\overline{m} = 0$ must undergo a bifurcation at some $c_{alg}(\overline{m}) > c_{KS}$. Also, by continuity the spinodal of the high-accuracy branch $c_{sp}(\overline{m})$ must persist for small enough $\overline{m} > 0$, hence the existence of a bifurcation diagram as in Fig. 2. When $\overline{m}$ is further increased the two spinodals $c_{sp}(\overline{m})$ and $c_{alg}(\overline{m})$ collide and for even larger values of $\overline{m}$ the bifurcation diagram becomes the one in Fig. 1(a). Indeed, when $\overline{m} \to 1$ the model reduces to a symmetric SBM with $q_1 = 3$ communities. Preliminary numerical results obtained by a resolution of the cavity equations via the population dynamics algorithm suggest a rather narrow domain of parameters for which this phenomenon of coexistence of two nontrivial stable solutions is observable; for this reason we do not present numerical data to illustrate this model.

K. Definition of accuracy and the maximal overlap estimator

The covariance matrix $A_{\sigma\tau} = \mathbb{E}[\delta_\sigma \delta_\tau]$ appeared naturally in our expansions as a measure of the deviation between the fixed-point distribution $P(\eta)$ solution of the cavity equations and the trivial fixed point $\delta(\eta - \overline{\eta})$. Let us now briefly comment on its interpretation as the accuracy of an estimation procedure and its connection with the maximal overlap estimator.

Consider the tree reconstruction problem explained in Sec. III C 1 and assume that the root had value τ in the broadcast process; an observer, who has no direct knowledge of τ, computes its posterior probability distribution η given the values of spins on far away vertices of the tree. If the observer proposes as an estimator of τ a random spin value chosen with probability η, the probability of success of the reconstruction is, on average with respect to the broadcast, the estimation and the tree, $\mathbb{E}_\tau[\eta_\tau]$. The probability of success if one had discarded all the observations, i.e., if one draws the estimator with the prior probability $\overline{\eta}$, is $\overline{\eta}_\tau$; the accuracy, defined as the difference of these two probabilities, is thus $\mathbb{E}_\tau[\delta_\tau]$. Averaging finally over the value τ of the unknown spin, we obtain the accuracy

$$a = \sum_\tau \overline{\eta}_\tau \mathbb{E}_\tau[\delta_\tau] = \sum_\tau A_{\tau\tau} = \mathrm{Tr} A, \quad (149)$$

where we used the identity (50) between conditional and unconditional distributions. For binary spins $\sigma = \pm 1$ there is an affine mapping between the MMSE and the accuracy.

Consider now the maximal overlap estimator defined in Sec. II A, for which the observer estimates the value of the root as $\mathrm{argmax}_\sigma \eta_\sigma$. The probability of correct estimation can be expressed in different forms involving the conditional or unconditional distributions of η, namely,

$$P_{corr} = \sum_\tau \overline{\eta}_\tau \mathbb{E}_\tau\left[\mathbb{I}\left(\mathrm{argmax}_\sigma \eta_\sigma = \tau\right)\right]$$

$$= \mathbb{E}\left[\sum_\tau \eta_\tau \mathbb{I}\left(\mathrm{argmax}_\sigma \eta_\sigma = \tau\right)\right] = \mathbb{E}\left[\max_\sigma \eta_\sigma\right], \quad (150)$$

the first step resulting from the general change of density between condition and unconditional distributions expressed in Eq. (48). This quantity can easily be evaluated numerically from the resolution of the cavity equations by the population dynamics algorithm; it is however much more difficult to characterize it analytically than the accuracy (149). If we could indeed determine systematically $A_{\sigma\tau}$ as a perturbative expansion in the small parameter κ in the neighborhood of the Kesten-Stigum transition, a similar determination is not possible for P_{corr}. The scaling ansatz (61) means that around the KS transition, the dominant behavior of δ under P is described by $\delta/\sqrt{\kappa} \xrightarrow{d} X$ (where X has a symmetric distribution). We can thus conclude that P_{corr} should behave as $\max_\sigma \overline{\eta}_\sigma + C\sqrt{\kappa}$, but with a prefactor C that involves the whole distribution of the rescaled random vector X, and thus cannot be computed from a finite number of moments. An explicit determination of P_{corr} can be achieved in the large-degree limit, due to the Gaussian simplifications explained in Sec. IV F. One finds indeed in this limit

$$P_{corr} = \mathbb{E}\left[\exp\left(\max_\sigma L_\sigma\right)\right], \quad (151)$$

where the expectation is over a Gaussian vector L defined by its moments in Eq. (91). For instance, in the symmetric Ising case ($q = 2$ and $\overline{m} = 0$), defining $\lambda = \lim(\tilde{\mathbb{E}}[\ell]\theta^2)$ as the SNR, a short computation yields

$$P_{corr} = \frac{1}{2} + \int_0^{\sqrt{2\lambda a}} \frac{dt}{\sqrt{2\pi}} e^{-t^2/2}, \quad (152)$$

where $a = a(\lambda)$ is the solution of

$$a = e^{-\lambda a} \mathbb{E}\left[\frac{e^{2\sqrt{2\lambda a}Z}}{e^{\sqrt{2\lambda a}Z} + e^{-\sqrt{2\lambda a}Z}}\right] - \frac{1}{2}, \quad (153)$$

with Z a standard Gaussian random variable (of zero mean and unit variance). Expanding now these expressions around the Kesten-Stigum transition, i.e., setting $\lambda = 1 + \epsilon$ with $\epsilon \to 0^+$, one obtains the leading behavior $a \sim \epsilon/2$, while $P_{corr} \sim \frac{1}{2} + \sqrt{\epsilon/2\pi}$.

V. MOMENT EXPANSIONS FOR k-WISE INTERACTING ISING VARIABLES

We turn now to the second specialization of the formalism of Sec. III: We will consider generic k-wise interactions, but we restrict the discussion now to binary variables ($q = 2$), which for convenience we will represent as Ising spins $\chi = \{-1, 1\}$. As most of the reasonings are similar to the

ones presented in Sec. IV, we will give fewer details of the computations and underline the main differences with the case of pairwise interacting Potts variables.

A. Fourier transforms of Boolean functions

The main ingredient defining the models under study here is a joint probability $p_{\rm j}(\sigma_1, \ldots, \sigma_k)$ over $\{-1, +1\}^k$, invariant under all permutations of its arguments. This symmetry implies that $p_{\rm j}$ is a function of $(\sigma_1 + \cdots + \sigma_k)$ only and can thus be specified by k real numbers (taking into account the normalization condition). The occupation models of [46], whose definition was recalled in Sec. III, correspond to the special case of a $p_{\rm j}$ vanishing for some values of $(\sigma_1 + \cdots + \sigma_k)$ and constant otherwise. A convenient way to specify a generic permutation invariant $p_{\rm j}$ is via the representation

$$
\begin{aligned}
&p_{\rm j}(\sigma_1, \ldots, \sigma_k) \\
&= \frac{1}{2^k}\left[1 + \gamma_1 \sum_{i=1}^{k} \sigma_i + \gamma_2 \sum_{i<j} \sigma_i \sigma_j + \cdots + \gamma_k \sigma_1 \cdots \sigma_k \right],
\end{aligned}
\tag{154}
$$

where the γ_n are the Fourier coefficients of $p_{\rm j}$. They can be expressed as

$$
\gamma_n = \mathbb{E}[\tau_1, \ldots, \tau_n],
\tag{155}
$$

where the average is over a configuration $(\tau_1, \ldots, \tau_k)$ drawn with probability $p_{\rm j}(\tau_1, \ldots, \tau_k)$. Let us describe in these terms the conditional distribution $p_{\rm c}(\tau_1, \ldots, \tau_{k-1}|\tau)$ obtained from $p_{\rm j}$. Because of its invariance under the permutations of its $k-1$ first arguments, it is fully described by the averages of products of n spins. These values are easily expressed in terms

of the Fourier coefficients of $p_{\rm j}$ as

$$
\mathbb{E}[\tau_1, \ldots, \tau_n|\tau] = \frac{\gamma_n + \tau \gamma_{n+1}}{1 + \tau \gamma_1} \quad \text{for} \quad n = 1, \ldots, k-1.
\tag{156}
$$

In what follows we will assume that the stationary probability distribution $\overline{\eta}$ is unbiased, i.e., $\overline{\eta}_+ = \overline{\eta}_- = \frac{1}{2}$. For the reversibility assumption of Sec. III to hold in this case one needs the marginal probability of a single variable drawn from $p_{\rm j}$ to also be unbiased. In terms of the Fourier coefficient representation this is equivalent to $\gamma_1 = 0$ [see Eq. (155)]. We will actually make a stronger hypothesis on $p_{\rm j}$, namely, that it is invariant under a global spin reversal: $p_{\rm j}(\sigma_1, \ldots, \sigma_k) = p_{\rm j}(-\sigma_1, \ldots, -\sigma_k)$. This implies that not only γ_1 but all the Fourier coefficients γ_p with p odd vanish. With this assumption, (156) can be simplified into

$$
\mathbb{E}[\tau_1, \ldots, \tau_{2p-1}|\tau] = \tau \gamma_{2p}, \quad \mathbb{E}[\tau_1, \ldots, \tau_{2p}|\tau] = \gamma_{2p}.
\tag{157}
$$

B. Cavity equations and free-entropy functional

We have given in Sec. III recursive equations obeyed by the distributions $P_\tau^{(n)}(\eta)$ and $\hat{P}_\tau^{(n)}(\nu)$ and their unconditional versions $P^{(n)}(\eta)$ and $\hat{P}^{(n)}(\nu)$ [see (15)–(20)]. As we are dealing now with Ising variables, we can use more succinct notation, a probability distribution over a Boolean variable being parametrized by a single real. We will use the notation

$$
\eta_\sigma = \frac{1 + m\sigma}{2}, \quad \nu_\sigma = \frac{1 + u\sigma}{2},
\tag{158}
$$

with m and u in $[-1, 1]$, and hence denote by $P_\tau^{(n)}(m)$, $P^{(n)}(m)$, $\hat{P}_\tau^{(n)}(u)$, and $\hat{P}^{(n)}(u)$ the above distributions expressed with this parametrization. The belief propagation equation (17) can then be reformulated as

$$
m = f(u^1, \ldots, u^\ell) = \frac{\prod_{i=1}^{\ell}(1+u^i) - \prod_{i=1}^{\ell}(1-u^i)}{\prod_{i=1}^{\ell}(1+u^i) + \prod_{i=1}^{\ell}(1-u^i)} = \tanh\left(\sum_{i=1}^{\ell} \operatorname{arctanh}(u^i)\right),
\tag{159}
$$

with the normalization factor

$$
z(u^1, \ldots, u^\ell) = \frac{1}{2}\prod_{i=1}^{\ell}(1+u^i) + \frac{1}{2}\prod_{i=1}^{\ell}(1-u^i).
\tag{160}
$$

The other BP equation (18) reads, in this notation,

$$
u = \hat{f}(m^1, \ldots, m^{k-1}) = \frac{\gamma_1 + \gamma_2 \sum_{i=1}^{k-1} m^i + \gamma_3 \sum_{i<j} m^i m^j + \cdots + \gamma_k m^1 \cdots m^{k-1}}{1 + \gamma_1 \sum_{i=1}^{k-1} m^i + \gamma_2 \sum_{i<j} m^i m^j + \cdots + \gamma_{k-1} m^1 \cdots m^{k-1}}.
\tag{161}
$$

With the assumption of invariance of $p_{\rm j}$ under global spin reversal, the odd Fourier coefficients vanish and one can simplify this equation into

$$
\begin{aligned}
u = \hat{f}(m^1, \ldots, m^{k-1}) &= \frac{1}{\hat{z}(m^1, \ldots, m^{k-1})}\left(\gamma_2 \sum_{i=1}^{k-1} m^i + \gamma_4 \sum_{i_1<i_2<i_3} m^{i_1} m^{i_2} m^{i_3} + \cdots \right), \\
\hat{z}(m^1, \ldots, m^{k-1}) &= 1 + \gamma_2 \sum_{i<j} m^i m^j + \gamma_4 \sum_{i_1<i_2<i_3<i_4} m^{i_1} m^{i_2} m^{i_3} m^{i_4} + \cdots.
\end{aligned}
\tag{162}
$$

Let us now discuss the symmetry properties of the distributions $P_\tau^{(n)}(m)$ and $P^{(n)}(m)$ and the relationships between them. The consequences of the Bayes theorem stated in Eq. (22) become, for an unbiased stationary distribution $\overline{\eta}$,

$$P^{(n)}(m) = \frac{1}{2}P_+^{(n)}(m) + \frac{1}{2}P_-^{(n)}(m), \quad P_\tau^{(n)}(m) = (1 + \tau m)P^{(n)}(m), \quad \int dP^{(n)}(m)m = 0. \tag{163}$$

Actually the assumption of invariance under global spin reversal has further consequences: Not only does $P^{(n)}$ have zero average, but it is also symmetric. Hence one has

$$P^{(n)}(m) = P^{(n)}(-m), \quad P_+^{(n)}(m) = P_-^{(n)}(-m). \tag{164}$$

Combining these two set of properties yields particularly simple identities between moments of these distributions. The change of densities between $P_\tau^{(n)}$ and $P^{(n)}$ means indeed that $\mathbb{E}_+^{(n)}[f(m)] = \mathbb{E}^{(n)}[(1 + m)f(m)]$ for any function $f(m)$. Applying this identity with $f(m) = m^{2p}$ and $f(m) = m^{2p-1}$ yields

$$\mathbb{E}_+^{(n)}[m^{2p}] = \mathbb{E}^{(n)}[m^{2p}] + \mathbb{E}^{(n)}[m^{2p+1}] = \mathbb{E}^{(n)}[m^{2p}], \quad \mathbb{E}_+^{(n)}[m^{2p-1}] = \mathbb{E}^{(n)}[m^{2p-1}] + \mathbb{E}^{(n)}[m^{2p}] = \mathbb{E}^{(n)}[m^{2p}], \tag{165}$$

where we exploited the symmetry of $P^{(n)}$ that makes its odd moments vanish. These identities are well known in the context of low-density parity check codes (see, for instance, Lemma 3 in Ref. [59]). Spelling out these identities for the lowest-order moments, we obtain

$$\mathbb{E}_+^{(n)}[m] = \mathbb{E}_+^{(n)}[m^2] = \mathbb{E}^{(n)}[m^2], \quad \mathbb{E}_+^{(n)}[m^3] = \mathbb{E}_+^{(n)}[m^4] = \mathbb{E}^{(n)}[m^4]. \tag{166}$$

The distributions $\hat{P}_\tau^{(n)}$ and $\hat{P}^{(n)}$ have exactly the same symmetry properties as $P_\tau^{(n)}$ and $P^{(n)}$, hence the random variables u enjoy the same identities as the m.

The identity $P_+^{(n)}(m) = P_-^{(n)}(-m)$ that follows from the invariance under global spin flip allows one to close the equations (15) and (16) on the two distributions $P_+^{(n)}(m)$ and $\hat{P}_+^{(n)}(u)$, which are found to evolve with n according to

$$P_+^{(n+1)}(m) = \sum_{\ell=0}^{\infty} \tilde{p}_\ell \int d\hat{P}_+^{(n)}(u^1) \cdots d\hat{P}_+^{(n)}(u^\ell)\delta(m - f(u^1, \ldots, u^\ell)), \tag{167}$$

$$\hat{P}_+^{(n)}(u) = \sum_{\tau_1, \ldots, \tau_{k-1}} p_c(\tau_1, \ldots, \tau_{k-1}|+) \int dP_+^{(n)}(m^1) \cdots dP_+^{(n)}(m^{k-1})\delta(u - \hat{f}(\tau_1 m^1, \ldots, \tau_{k-1}m^{k-1})), \tag{168}$$

where the function f and $\hat{f}$ are given in Eq. (159) and (162), respectively. This form is very convenient for a numerical resolution by the population dynamics algorithm (see Appendix A for more details on this point). Equivalently, the unconditional versions of the cavity equations read

$$P^{(n+1)}(m) = \sum_{\ell=0}^{\infty} \tilde{p}_\ell \int d\hat{P}^{(n)}(u^1) \cdots d\hat{P}^{(n)}(u^\ell)\delta(m - f(u^1, \ldots, u^\ell))z(u^1, \ldots, u^\ell), \tag{169}$$

$$\hat{P}^{(n)}(u) = \int dP^{(n)}(m^1) \cdots dP^{(n)}(m^{k-1})\delta(u - \hat{f}(m^1, \ldots, m^{k-1}))\hat{z}(m^1, \ldots, m^{k-1}). \tag{170}$$

The free-entropy functional can be expressed in these two versions of the cavity formalism from (24) and (28) as

$$\phi(P, \hat{P}) = -\mathbb{E}[\ell] \int dP(m)d\hat{P}(u)z_e(m, u) \ln z_e(m, u)$$

$$+ \frac{\mathbb{E}[\ell]}{k} \int dP(m^1) \cdots dP(m^k)z_c(m^1, \ldots, m^k) \ln z_c(m^1, \ldots, m^k)$$

$$+ \sum_{\ell=1}^{\infty} p_\ell \int d\hat{P}(u^1) \cdots d\hat{P}(u^\ell)z_v(u^1, \ldots, u^\ell) \ln z_v(u^1, \ldots, u^\ell), \tag{171}$$

$$\phi(P_+, \hat{P}_+) = -\mathbb{E}[\ell] \int dP_+(m)d\hat{P}_+(u) \ln z_e(m, u)$$

$$+ \frac{\mathbb{E}[\ell]}{k} \sum_{\sigma_1, \ldots, \sigma_k} p_j(\sigma_1, \ldots, \sigma_k) \int dP_+(m^1) \cdots dP_+(m^k) \ln z_c(\sigma_1 m^1, \ldots, \sigma_k m^k)$$

$$+ \sum_{\ell=1}^{\infty} p_\ell \int d\hat{P}_+(u^1) \cdots d\hat{P}_+(u^\ell) \ln z_v(u^1, \ldots, u^\ell), \tag{172}$$

where

$$z_e(m, u) = 1 + mu, \tag{173}$$

$$z_c(m^1, \ldots, m^k) = 1 + \gamma_2 \sum_{i<j} m^i m^j + \gamma_4 \sum_{i_1<i_2<i_3<i_4} m^{i_1} m^{i_2} m^{i_3} m^{i_4} + \cdots, \tag{174}$$

$$z_v(u^1, \ldots, u^\ell) = \frac{1}{2}\prod_{i=1}^{\ell}(1 + u^i) + \frac{1}{2}\prod_{i=1}^{\ell}(1 - u^i). \tag{175}$$

C. Stability analysis of the trivial fixed point via moment expansions

The functions $f(\{u^i\})$ and $\hat{f}(\{m^i\})$ defined in Eqs. (159) and (162) vanish when all their arguments are equal to 0, which traduces the stationarity of the unbiased distribution $\overline{\eta}$. As a consequence, the distributions $P^{(n)}(m) = \delta(m)$ and $\hat{P}^{(n)}(u) = \delta(u)$ form a fixed point of the cavity equations. Following the same strategy as in Sec. IV D, we will now investigate its stability, by first locating the Kesten-Stigum transition where the trivial fixed point goes from stable to unstable and then looking for a bifurcating nontrivial fixed point in the neighborhood of the transition.

1. Linear analysis

Linearizing for small arguments the functions $f(\{u^i\})$ and $\hat{f}(\{m^i\})$ defined in Eq. (159) and (162) and inserting this expansion in the conditional distribution recursions (167) and (168) yields very easily

$$\mathbb{E}_+^{(n+1)}[m] = \tilde{\mathbb{E}}[\ell]\mathbb{E}_+^{(n)}[u],$$
$$\mathbb{E}_+^{(n)}[u] = (k-1)\gamma_2\mathbb{E}[\tau_1|+]\mathbb{E}_+^{(n)}[m], \tag{176}$$

with $\mathbb{E}[\tau_1|+] = \gamma_2$ as proven before in Eq. (157). Putting these two equations together and expressing the conditional first moments in terms of the unconditional second moments according to (166) gives

$$\mathbb{E}^{(n+1)}[m^2] = \tilde{\mathbb{E}}[\ell](k-1)\gamma_2^2\mathbb{E}^{(n)}[m^2]. \tag{177}$$

This implies that the Kesten-Stigum transition occurs when

$$\tilde{\mathbb{E}}[\ell](k-1)\gamma_2^2 = 1; \tag{178}$$

in the following we will assume that the second Fourier coefficient does not vanish, $\gamma_2 \neq 0$, for this bifurcation to occur at a finite value of $\tilde{\mathbb{E}}[\ell]$ (this excludes notably the case of a XORSAT constraint for p_j whenever $k \geqslant 3$).

2. Second-order expansion: Continuity of the Kesten-Stigum transition

We look now for a fixed-point distribution $P(m)$ in the neighborhood of the Kesten-Stigum transition. We need to make an ansatz for the behavior of its moments in order to reduce the functional bifurcation problem to a finite-dimensional one, as already explained for the case of Potts variables in Sec. IV D. Due to the spin-reversal symmetry, we known that for all odd moments $\mathbb{E}[m^{2p+1}] = \mathbb{E}[u^{2p+1}] = 0$; we assume the existence of a small parameter κ such that the even moments scale as $\mathbb{E}[m^{2p}] = O(\kappa^p)$ and $\mathbb{E}[u^{2p}] = O(\kappa^p)$. Pushing the expansion of $f(\{u^i\})$ and $\hat{f}(\{m^i\})$ to the order needed to obtain the first correction to the variances of u and m yields, after a short computation,

$$\mathbb{E}[m^2] = \tilde{\mathbb{E}}[\ell]\mathbb{E}[u^2] - \tilde{\mathbb{E}}[\ell(\ell-1)](\mathbb{E}[u^2])^2 + O(\kappa^3), \tag{179}$$

$$\mathbb{E}[u^2] = (k-1)\gamma_2^2\mathbb{E}[m^2]$$
$$- (k-1)(k-2)\gamma_2^3(\mathbb{E}[m^2])^2 + O(\kappa^3). \tag{180}$$

Eliminating the variance of u gives us a single equation for $\mathbb{E}[m^2]$; we will define $a = \mathbb{E}[m^2]/2$ in such a way that this quantity corresponds to the one of Sec. IV D, which obeys

$$a = \tilde{\mathbb{E}}[\ell](k-1)\gamma_2^2 a - 2\tilde{\mathbb{E}}[\ell](k-1)(k-2)\gamma_2^3 a^2 - 2\tilde{\mathbb{E}}[\ell(\ell-1)]\big[(k-1)\gamma_2^2 a\big]^2 + O(\kappa^3). \tag{181}$$

Defining $\tilde{\mathbb{E}}[\ell](k-1)\gamma_2^2 = 1 + \epsilon$, with ϵ parametrizing the distance to the Kesten-Stigum transition, dividing the equation by a, and taking $\epsilon = 0$ in the correction term gives

$$0 = \epsilon - 2\left[(k-2)\gamma_2 + \frac{\tilde{\mathbb{E}}[\ell(\ell-1)]}{\tilde{\mathbb{E}}[\ell]^2}\right]a. \tag{182}$$

We will show now that the expression in the large square brackets is non-negative for all the models encompassed by the study of this section and hence that the bifurcating solution (which must certainly obey the positivity condition $a = \mathbb{E}[m^2] > 0$) always exists for $\epsilon > 0$, i.e., in the regime of parameters where the trivial fixed point is unstable. To prove our claim we first notice that the Fourier coefficient γ_2 lies necessarily in the interval $[-\frac{1}{k-1}, 1]$. Indeed, from the interpretation of these coefficients as spin averages given in Eq. (155) we obtain

$$\mathbb{E}[(\tau_1 + \cdots + \tau_k)^2] = \mathbb{E}[k\tau_1^2 + k(k-1)\tau_1\tau_2] = k[1 + (k-1)\gamma_2] \geqslant 0 \Rightarrow \gamma_2 \geqslant -\frac{1}{k-1}, \tag{183}$$

where we have merely exploited the permutation invariance and the fact that Ising spins square to 1. Moreover, the upper bound $\gamma_2 = \mathbb{E}[\tau_1\tau_2] \leqslant 1$ is obvious. Then we rewrite the term in large square brackets in Eq. (182) as

$$\frac{\tilde{\mathbb{E}}[\ell^2] - \tilde{\mathbb{E}}[\ell]^2}{\tilde{\mathbb{E}}[\ell]^2} + 1 - \frac{1}{\tilde{\mathbb{E}}[\ell]} + (k-2)\gamma_2 = \frac{\tilde{\mathbb{E}}[\ell^2] - \tilde{\mathbb{E}}[\ell]^2}{\tilde{\mathbb{E}}[\ell]^2} + \big[1 - (k-1)\gamma_2^2 + (k-2)\gamma_2\big], \tag{184}$$

where we used the Kesten-Stigum condition $\tilde{\mathbb{E}}[\ell](k-1)\gamma_2^2 = 1$. The first term in Eq. (184) is proportional to the variance of the offspring distribution and hence is non-negative; as $1 - (k-1)x^2 + (k-2)x \geqslant 0$ for $x \in [-1/(k-1), 1]$, the authorized interval of variation of γ_2, the second term in Eq. (184), is also non-negative.

3. Third-order expansion

We have also pushed the expansion in moments to the next order; compared to the Potts case, the computations are much easier, due, on the one hand, to the binary nature of the variable (hence the moments to be determined are scalar quantities that do not bear any spin index) and, on the other hand, to the spin-flip symmetry (which cancels the moments of odd order). Let us introduce for convenience some more compact notation that is reminiscent of that used in Sec. IV:

$$a = \tfrac{1}{2}\mathbb{E}[m^2], \quad b = \mathbb{E}[m^4], \quad \hat{a} = \tfrac{1}{2}\mathbb{E}[u^2], \quad \hat{b} = \mathbb{E}[u^4]. \tag{185}$$

These quantities obey the set of equations

$$a = \tilde{\mathbb{E}}[\ell]\hat{a} - 2\tilde{\mathbb{E}}[\ell(\ell-1)]\hat{a}^2 + \tilde{\mathbb{E}}[\ell(\ell-1)]\hat{a}\hat{b} + \frac{20}{3}\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\hat{a}^3,$$

$$b = \tilde{\mathbb{E}}[\ell]\hat{b} + 12\tilde{\mathbb{E}}[\ell(\ell-1)]\hat{a}^2,$$

$$\hat{a} = (k-1)\gamma_2^2 a - 2(k-1)(k-2)\gamma_2^3 a^2 + (k-1)(k-2)\gamma_2^4 ab + 4(k-1)(k-2)(k-3)\left(\frac{\gamma_4^2}{6} - \gamma_2^2\gamma_4 + \frac{5\gamma_2^4}{2}\right)a^3,$$

$$\hat{b} = (k-1)\gamma_2^4 b + 12(k-1)(k-2)\gamma_2^4 a^2, \tag{186}$$

where in a and $\hat{a}$ (b and $\hat{b}$) we have neglected terms of order κ^4 (κ^3).

The expansion of the free entropy to the corresponding order yields

$$\phi(a, b, \hat{a}, \hat{b}) = (k-1)\gamma_2^2 a^2 - \frac{4}{3}(k-1)(k-2)\gamma_2^3 a^3 + \frac{1}{24}(k-1)\gamma_2^4 b^2 + (k-1)(k-2)\gamma_2^4 a^2 b$$

$$+ 2(k-1)(k-2)(k-3)\left(\frac{\gamma_4^2}{6} - \gamma_2^2\gamma_4 + \frac{5\gamma_2^4}{2}\right)a^4 + \tilde{\mathbb{E}}[\ell]\hat{a}^2 - \frac{4}{3}\tilde{\mathbb{E}}[\ell(\ell-1)]\hat{a}^3$$

$$+ \frac{1}{24}\tilde{\mathbb{E}}[\ell]\hat{b}^2 + \tilde{\mathbb{E}}[\ell(\ell-1)]\hat{a}^2\hat{b} + \frac{10}{3}\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\hat{a}^4 - 2a\hat{a} - \frac{1}{12}b\hat{b}; \tag{187}$$

one can check that the derivatives of this expression vanish when the cavity equations (186) are fulfilled, a property which arises from the variational character of the free entropy. Another short computation also reveals that the free entropy, computed on the nontrivial perturbative fixed point which here always exists for $\epsilon > 0$, behaves as ϵ^3 (with a positive prefactor) when $\epsilon \to 0^+$.

D. Numerical results for two examples

In order to confirm our analytical expansions and to complement them with a global bifurcation analysis, we have solved numerically the cavity equations for two cases of occupation models, as we will now detail.

1. Bicoloring

We consider first the bicoloring problem, for which the joint probability $p_j(\sigma_1, \ldots, \sigma_k)$ is 0 if $\sigma_1 = \cdots = \sigma_k = \pm 1$ and $1/(2^k - 2)$ otherwise. This constraint forbids monochromatic hyperedges and gives an equal weight to all configurations with at least one positive and at least one negative variable around it. One easily finds the corresponding value of the Fourier coefficients,

$$\gamma_n = -\frac{1}{2^{k-1} - 1} \tag{188}$$

for all even n between 2 and k (the odd coefficients vanish due to the up-down symmetry). We use for the degree distributions

$p_\ell = \tilde{p}_\ell$ Poisson laws of average αk, which corresponds to a random hypergraph with $M = \alpha N$ constraints. The criterion (178) shows then that the Kesten-Stigum transition happens at

$$\alpha_{\text{KS}} = \frac{(2^{k-1} - 1)^2}{k(k-1)}. \tag{189}$$

Figure 6, discussed in Sec. III D 4, presents the variance $a = \mathbb{E}[m^2]/2$ of the fixed-point solutions of the cavity equations reached from the reconstruction (informative) and robust reconstruction (uninformative) initial conditions. As discussed there, our analytical prediction on the existence of a bifurcating solution above the Kesten-Stigum transition is confirmed by these numerical results, and for $k = 3, 4, 5$ we find explicit realizations of the bifurcation diagrams presented in Figs. 1(a) and 2. In Fig. 9 we plot the complexity (i.e., minus the free entropy) of these fixed points; the location of the discontinuity in the derivative of the solution that minimizes the complexity defines the threshold α_{IT} reported in Table I. We chose here to plot the opposite of the free entropy as it is a more familiar quantity in the context of random, instead of planted, constraint satisfaction problems. The interpretation of this coexistence of solutions in this context has been discussed, also in the example of the bicoloring, in Ref. [58].

2. The 2-in-4 satisfiability model

We have also considered the so-called 2-in-4 satisfiability model in the nomenclature of [46], defined by
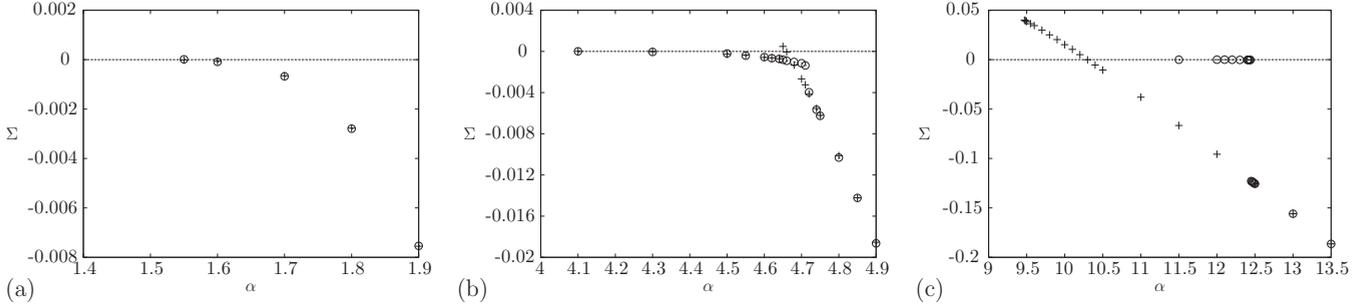
FIG. 9. Complexity $\Sigma = -\phi$ for the fixed-point solutions of the cavity equations for the hypergraph bicoloring problem for (a) $k = 3$, (b) $k = 4$, and (c) $k = 5$. The symbols are the same as in Fig. 6; pluses correspond to an informative initial condition and circles to an uninformative one.

$k = 4$ and $p_{\mathrm{j}}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = \mathbb{I}(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4 = 0)/6$: The constraint imposes that among the four variables exactly two are positive and two negative. The corresponding Fourier coefficients are

$$\gamma_1 = \gamma_3 = 0, \quad \gamma_2 = -\tfrac{1}{3}, \quad \gamma_4 = 1. \qquad (190)$$

Note that this case saturates the lower bound on γ_2.

In our numerical investigation we have used truncated Poisson distributions for p_ℓ and $\tilde{p}_\ell$, namely,

$$p_\ell = \frac{1}{e^c - 1 - c} \frac{c^\ell}{\ell!} \mathbb{I}(\ell \geqslant 2), \quad \tilde{p}_\ell = \frac{1}{e^c - 1} \frac{c^\ell}{\ell!} \mathbb{I}(\ell \geqslant 1). \qquad (191)$$

The averages of these distribution are

$$\mathbb{E}[\ell] = c \frac{1 - e^{-c}}{1 - (c+1)e^{-c}}, \quad \tilde{\mathbb{E}}[\ell] = \frac{c}{1 - e^{-c}}. \qquad (192)$$

In terms of this parameter c, the Kesten-Stigum transition is thus found to happen at $c_{\mathrm{KS}} \approx 2.821\,44$.

Note that this is a locked model in the sense of [46], as the minimal degree of a variable is 2 and no pair of configurations allowed by p_{j} differs from one single spin flip: The typical solutions of such a constraint satisfaction problem on a locally treelike factor graph are separated by a Hamming distance that diverges with the graph size. Nevertheless, this property does not have any effect on the existence of a continuously bifurcating solution above the Kesten-Stigum transition. In some sense the locked property is a high-frequency condition on p_{j}, whereas we have shown that the criterion (184) only depends on the low-frequency Fourier coefficient γ_2.

Indeed, our numerical resolution of the cavity equations showed that a continuously growing solution exists right above the Kesten-Stigum transition and disappears at $c_{\mathrm{alg}} \approx 2.84$. Note that the interval $[c_{\mathrm{KS}}, c_{\mathrm{alg}}]$ is very small, which explains why it remained unnoticed in Ref. [11]. This model thus falls in the scenario sketched in Figs. 2(b) and 2(d), with the other thresholds $c_{\mathrm{sp}} \approx 1.256$ and $c_{\mathrm{IT}} \approx 1.853$ correctly determined in Ref. [11]. Actually, for locked models the high-accuracy fixed point is perfectly informative, as it corresponds to $P_\tau(m) = \delta(m - \tau)$; the transition at c_{sp} corresponds to a change of stability of this perfectly informative fixed point, which can be tested by taking $\varepsilon \to 1$ after $n \to \infty$ in Eq. (30), a procedure termed small noise reconstruction in Ref. [46].

VI. NUMERICAL EXPERIMENTS ON SINGLE SAMPLES

This section is devoted to a study of the behavior of the belief propagation algorithm run on given instances of two inference problems defined and studied analytically in the previous sections, namely, the planted hypergraph bicoloring and the asymmetric SBM with two groups. As a matter of fact, most of the analytical results presented in the rest of the paper directly apply to (infinite) tree reconstruction problems and their interpretation in terms of inference on (large but) finite-size graphs relying on delicate conjectures of the cavity method, as briefly discussed in Sec. III C 2. The experiments reported in this section will allow us to test this connection and to confront quantitatively the predictions of the tree reconstruction problem, studied numerically via the population dynamics algorithm, and the results of BP on single large instances. Because of the local convergence of the graph problems towards trees, it is quite simple to see that a finite number of iterations of BP can be described analytically, in the thermodynamic limit, by a finite number of the tree distributional iterations (15) and (16). There are however two aspects that make the graph-tree connection much less trivial and justify these numerical tests: (i) On finite graphs BP can be run until convergence to a fixed point, i.e., for a number of iterations much larger than the girth of the graph, this regime cannot *a priori* be described in terms of treelike local properties and could be sensitive to the long cycles of the graph, and (ii) in the graph problems there is initially no observation of the labels on the vertices and the infinitesimal information of the tree robust reconstruction problem must thus arise from the amplification of noise in the initial condition of BP.

A. Generation of planted problems, BP equations, and observables

The models that we study numerically, namely, random hypergraph bicoloring and asymmetric SBM with two groups, have already been defined in detail in Sec. III A. Here we just give additional information about the generation of the instances with a planted solution and about the BP equations and the way we solve them, for the convenience of the reader who would like to repeat our numerical tests. Both models have Ising (i.e., binary) variables $\sigma_i \in \{-1, 1\}$ and thus the marginals can be written in terms of a single scalar variable.

1. Random hypergraph bicoloring

The random ensemble of hypergraph bicoloring problems has three parameters: the number of variables N, the number of constraints $M = \alpha N$, and the degree of the factor nodes k (i.e., the number of variables per constraint). The N variables are divided into two groups of equal size and the planted configuration is defined as $\sigma_i^* = 1$ in the first group and $\sigma_i^* = -1$ in the second one. For each of the M constraints we extract k variables uniformly at random, conditional on the k variables not all belonging to the same group (we implemented this condition by a rejection method). The random graph thus obtained can be described by the set E of edges (ia) connecting variable nodes and factor nodes ($1 \leqslant i \leqslant N$, $1 \leqslant a \leqslant M$, and $|E| = KM$).

One can write the posterior distribution of $\underline{\sigma}$ given the observation of the graph and treat this probability measure with the BP algorithm, which can be put, after some simplification, in the form of messages $\eta_{i \to a}$ and $\hat{\eta}_{a \to i}$ passed between variables and interactions obeying the equations

$$\eta_{i \to a}^{(t)} = \frac{\prod_{b \in \partial i \backslash a} \hat{\eta}_{b \to i}^{(t-1)}}{\prod_{b \in \partial i \backslash a} \hat{\eta}_{b \to i}^{(t-1)} + \prod_{b \in \partial i \backslash a} \left(1 - \hat{\eta}_{b \to i}^{(t-1)}\right)}, \quad (193)$$

$$\hat{\eta}_{a \to i}^{(t)} = \gamma \hat{\eta}_{a \to i}^{(t-1)} + (1 - \gamma)$$
$$\times \frac{1 - \prod_{j \in \partial a \backslash i} \eta_{j \to a}^{(t)}}{2 - \prod_{j \in \partial a \backslash i} \eta_{j \to a}^{(t)} - \prod_{j \in \partial a \backslash i} \left(1 - \eta_{j \to a}^{(t)}\right)}, \quad (194)$$

where $\partial i = \{a : (ia) \in E\}$ and $\partial a = \{i : (ia) \in E\}$ are the local neighborhoods of variable and factor nodes, respectively, and γ is a damping factor set to 0.5 in most of our numerical simulations. The BP messages $\eta_{i \to a}$ and $\hat{\eta}_{a \to i}$ represent the probability that variable i belongs to a given group (say, the first group) in a modified graph where some of the edges have been removed: in particular $\eta_{i \to a}$ considers the graph where constraint a has been removed, while $\hat{\eta}_{a \to i}$ considers the graph where constraints in $\partial i \setminus a$ have been removed.

Belief propagation messages are initialized in the following way: With probability q_0 we set $\hat{\eta}_{a \to i}^{(0)} = \eta_{i \to a}^{(0)} = \mathbb{I}[\sigma_i^* = 1]$ and with probability $1 - q_0$ we set $\hat{\eta}_{a \to i}^{(0)} = 0.5$ and $\eta_{i \to a}^{(0)} \in [0.45, 0.55]$ uniformly at random. The parameter q_0 (the q_0 and q of this section should not be confused with the number of states of the Potts model in the rest of the paper) thus controls the amount of direct information on the planted configuration we use in this initial condition. Of course only $q_0 = 0$ should be considered if we want to study BP as an inference algorithm that does not use any information on the hidden labels. It is however useful to allow arbitrary values of q_0 as a tool to investigate the connections between the tree and graph problems. The choice of distributing the messages in the interval [0.45,0.55] for the uninformed vertices is made to avoid the trivial fixed point with all messages $\eta = \hat{\eta} = 0.5$.

Belief propagation messages are not updated all in parallel, as the time indices in Eqs. (193) and (194) may suggest. In each step of BP we visit all the variables once in a random order: For each variable i we compute all the outgoing messages $\eta_{i \to a}$ according to Eq. (193) and immediately update all the BP messages leaving neighboring factor nodes according to Eq. (194). By these tricks (and the use of damping) we avoid any undesirable oscillation and improve convergence to fixed

points. Convergence is declared achieved if, in a given step of BP, all messages change by less than a pre-fixed threshold, i.e., $|\eta_{i \to a}^{(t+1)} - \eta_{i \to a}^{(t)}| < 10^{-8}$.

At a fixed point $\{\eta_{i \to a}^{\star}, \hat{\eta}_{a \to i}^{\star}\}$ of the BP equations, the local magnetizations are given by

$$m_i^{\star} = \frac{\prod_{a \in \partial i} \hat{\eta}_{a \to i}^{\star} - \prod_{a \in \partial i}(1 - \hat{\eta}_{a \to i}^{\star})}{\prod_{a \in \partial i} \hat{\eta}_{a \to i}^{\star} + \prod_{a \in \partial i}(1 - \hat{\eta}_{a \to i}^{\star})}. \quad (195)$$

Because of the global spin-flip symmetry of the model, the mean magnetization $m = \sum_i m_i^{\star}/N$ is zero, while detection of the planted solution is signaled by the following order parameters: the staggered magnetization m_s (i.e., the absolute value of the mean overlap with the planted configuration); the magnetization variance m_2 (recall that $\overline{m} = 0$ here), which equals the mean overlap between two real replicas; and the maximum overlap q with the planted configuration, defined as

$$m_s = \frac{1}{N} \left| \sum_i m_i^{\star} s_i^* \right|,$$

$$m_2 = \frac{1}{N} \sum_i (m_i^{\star})^2, \quad (196)$$

$$q = \frac{1}{N} \left| \sum_i \text{sgn}(m_i^{\star}) s_i^* \right|.$$

The Bethe (replica symmetric) entropy is given by the expression

$$S = \frac{1}{N} \sum_i \ln \left[\prod_{a \in \partial i} \hat{\eta}_{a \to i}^{\star} + \prod_{a \in \partial i}(1 - \hat{\eta}_{a \to i}^{\star}) \right]$$
$$+ \frac{1}{N} \sum_a \ln \left[1 - \prod_{i \in \partial a} \eta_{i \to a}^{\star} - \prod_{i \in \partial a}(1 - \eta_{i \to a}^{\star}) \right]$$
$$- \frac{1}{N} \sum_{(ia)} \ln[\hat{\eta}_{a \to i}^{\star} \eta_{i \to a}^{\star} + (1 - \hat{\eta}_{a \to i}^{\star})(1 - \eta_{i \to a}^{\star})], \quad (197)$$

which reduces to $S_{\text{para}} = [\ln(2) + \alpha \ln(1 - 2^{1-k})]$ on the paramagnetic (trivial) BP fixed point, where all messages are uninformative $\hat{\eta}_{a \to i}^{\star} = \eta_{i \to a}^{\star} = 1/2$.

2. Asymmetric two-group SBM

An instance of the asymmetric SBM with two groups is generated using the following four parameters: the number of variables N, the mean degree d, the asymmetry $\overline{m}$, and the parameter θ related to the SNR. The N variables are divided into two groups of sizes $N_1 = \frac{1+\overline{m}}{2}N$ and $N_2 = \frac{1-\overline{m}}{2}N$ such that the planted configuration is $\sigma_i^* = 1$ in the first group and $\sigma_i^* = -1$ in the second group. Then the edge set E of the interaction graph between the variable nodes is generated by selecting uniformly at random $M_{11} = \frac{N_1^2}{2}\frac{c_{++}}{N}$ edges among the variables in the first group, $M_{22} = \frac{N_2^2}{2}\frac{c_{--}}{N}$ edges among the variables in the second group, and $M_{12} = N_1 N_2 \frac{c_{+-}}{N}$ edges joining variables in different groups, where $c_{\sigma\tau} = d(1 + \theta\frac{(\sigma - \overline{m})(\tau - \overline{m})}{1 - \overline{m}^2})$. It is easy to check that $\frac{M_{11}}{N_1} + \frac{M_{12}}{2N_1} = \frac{M_{22}}{N_2} + \frac{M_{12}}{2N_2} = \frac{d}{2}$, i.e., the mean degree in each group is the same. It is worth noting that these rules define a microcanonical ensemble of random graphs

having a *fixed* number of edges. In the canonical ensemble of random graphs, each possible edge is chosen independently with a given probability and thus the number of edges is a random variable. The difference between the two ensemble vanishes in the large-N limit and produces fluctuations of $O(1/\sqrt{N})$ in intensive observables measured in finite graphs. The BP equations for the posterior measure of the SBM were written in the general case in Refs. [8,9]; we reproduce here this derivation for the asymmetric two-group model.

We rewrite the posterior probability as the Gibbs measure of an Ising model

$$\mathbb{P}[\underline{\sigma}] \propto \exp\left[\sum_i H_i\sigma_i + \sum_{i<j} J_{ij}\sigma_i\sigma_j\right]. \quad (198)$$

Here and in the following $\propto$ denotes proportionality up to a constant independent of the spin variables. Indeed, the prior is given by

$$\prod_i \frac{1+\overline{m}\sigma_i}{2} \propto \exp\left(H\sum_i \sigma_i\right) \quad \text{with } \tanh(H) = \overline{m}, \quad (199)$$

while the likelihood is proportional to

$$\prod_{(ij)\in E} \frac{1}{N}c_{\sigma_i\sigma_j} \prod_{(ij)\notin E}\left(1-\frac{1}{N}c_{\sigma_i\sigma_j}\right)$$

$$\simeq \prod_{(ij)\in E} \frac{1}{N}c_{\sigma_i\sigma_j} \exp\left(-\frac{1}{N}\sum_{i<j}c_{\sigma_i\sigma_j}\right), \quad (200)$$

where c is the affinity matrix with which the edges have been generated. We have approximated the product over the nonedges (i.e., pairs of vertices not connected by an edge) with a sum over all pairs of variables, the difference being a correction of $O(1/N)$ in the large-N limit. The algebraic relation that holds for $\sigma, \tau \in \{-1, 1\}$,

$$c_{\sigma\tau} = d\left(1 + \theta\frac{(\sigma-\overline{m})(\tau-\overline{m})}{1-\overline{m}^2}\right) \propto e^{J\sigma\tau - K(\sigma+\tau)}, \quad (201)$$

with

$$J = \frac{1}{4}\ln\left[\frac{(1+\theta)^2 - \overline{m}^2(1-\theta)^2}{(1-\theta)^2(1-\overline{m}^2)}\right], \quad (202)$$

$$K = \frac{1}{4}\ln\left[\frac{1-\overline{m}^2 + \theta(1+\overline{m})^2}{1-\overline{m}^2 + \theta(1-\overline{m})^2}\right], \quad (203)$$

allows us to rewrite the first term of the likelihood in exponential form

$$\exp\left[J\sum_{(ij)\in E}\sigma_i\sigma_j - K\sum_i d_i\sigma_i\right]. \quad (204)$$

The second term of the likelihood (the one given by all the edges, including those absent from the graph) provides a term proportional to

$$\exp\left[-\frac{d\theta}{2(1-\overline{m}^2)N}\left(\sum_i \sigma_i - N\overline{m}\right)^2\right]. \quad (205)$$

So the posterior distribution that we have to study is given by the expression

$$\mathbb{P}[\underline{\sigma}] \propto \exp\left[\sum_i\left(H - d_iK + \frac{d\theta\overline{m}}{1-\overline{m}^2}\right)\sigma_i\right.$$

$$\left. + J\sum_{(ij)\in E}\sigma_i\sigma_j - \frac{d\theta}{(1-\overline{m}^2)N}\sum_{i<j}\sigma_i\sigma_j\right], \quad (206)$$

which corresponds to an Ising model with local fields and couplings given by

$$H_i = H - d_iK + \frac{d\theta\overline{m}}{1-\overline{m}^2}, \quad (207)$$

$$J_{ij} = \begin{cases} J - \frac{d\theta}{(1-\overline{m}^2)N} \simeq J & \text{if } (ij) \in E \\ -\frac{d\theta}{(1-\overline{m}^2)N} & \text{if } (ij) \notin E. \end{cases} \quad (208)$$

The corresponding BP equations are very well known,

$$u_{i\to j} = \text{arctanh}\left[\tanh(J_{ij})\tanh\left(H_i + \sum_{k\neq j}u_{k\to i}\right)\right]. \quad (209)$$

Unfortunately, these BP equations involve $N(N-1)$ messages $u_{i\to j}$ with $i \neq j$ (we assume $u_{i\to i} = 0$) and we need to simplify them if we want BP to run in a time linear in the system size. The main observation to achieve such a simplification is that BP messages running along the nonedges, i.e., sent between vertices not connected in E, are $O(1/N)$ and thus very small. The following equations hold up to terms of $O(1/N)$ if $(ij) \notin E$:

$$u_{i\to j} = J_{ij}\tanh\left(H_i + \sum_{k\neq j}u_{k\to i}\right) = J_{ij}m_i. \quad (210)$$

Here the first equality comes from J_{ij} being $O(1/N)$, while the second one comes from ignoring $u_{j\to i} \sim O(1/N)$ in the definition of local magnetization

$$m_i \equiv \tanh\left(H_i + \sum_k u_{k\to i}\right). \quad (211)$$

Eliminating BP messages on nonedges via the substitution $u_{i\to j} = -d\theta m_i/(1-\overline{m}^2)N$, we end up working only with the $O(N)$ BP messages running on the edges in E. These are updated according to the iterative equation

$$u_{i\to j}^{(t+1)} = \gamma u_{i\to j}^{(t)} + (1-\gamma)\text{arctanh}$$

$$\times \left[\tanh(J)\tanh\left(\tilde{H}_i^{(t)} + \sum_{k\in\partial i\setminus j}u_{k\to i}^{(t)}\right)\right], \quad (212)$$

where $\partial i = \{j : (ij) \in E\}$ and we have redefined the local fields including also the effect of the nonedges

$$\tilde{H}_i^{(t)} = H - d_iK + \frac{d\theta}{1-\overline{m}^2}\left(\overline{m} - \frac{1}{N}\sum_\ell m_\ell^{(t)}\right). \quad (213)$$

The local magnetizations are defined via

$$m_i^{(t)} = \tanh\left(\tilde{H}_i^{(t)} + \sum_{k\in\partial i}u_{k\to i}^{(t)}\right). \quad (214)$$
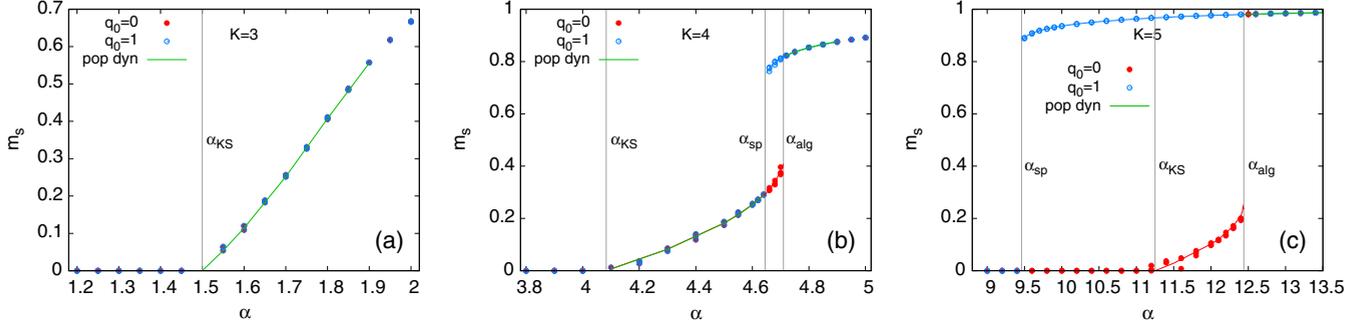
FIG. 10. Order parameter m_s detecting the planted configuration for (a) $k = 3$, (b) $k = 4$, and (c) $k = 5$, obtained by running BP on three samples of size $N = 10^6$ for each value of α. The lines are the predictions obtained by solving the cavity equations for the reconstruction problem on a tree (plotted also in Fig. 6), with the correspondence $m_s = 2a$ explained in the text.

We solve Eq. (212) starting from a set of BP messages strongly biased towards a configuration $\underline{\tau}^0$, that is, $u_{i \to j}^{(0)} = L\tau_i^0$, where L is a large number (we set $L = 100$ in our simulations, but the results are independent of L). Again we call q_0 the similarity between the starting configuration and the planted one, that is, the components of $\underline{\tau}^0$ are independent random variables distributed according to

$$\mathbb{P}\big[\tau_i^0 = s\big] = q_0 \delta_{s,\sigma_i^*} + (1 - q_0)\frac{1 + \overline{m}s}{2}. \quad (215)$$

Belief propagation messages are updated in a random order and convergence is determined by the condition

$$\big|m_i^{(t+1)} - m_i^{(t)}\big| < 10^{-8} \ \forall i. \quad (216)$$

At the BP fixed point, the local magnetizations are given by

$$m_i^\star = \tanh\left(\tilde{H}_i^\star + \sum_{j \in \partial i} u_{j \to i}^\star\right). \quad (217)$$

The parameters m_s, m_2, and q defined in Eq. (196) signal again the detection of the planted configuration (for $\overline{m} \neq 0$ the absolute value in the definition of m_s is not strictly required, but keeping it is not an error). It is worth noting that in this case the uninformative fixed point is characterized by the BP messages $u_{i \to j}^\star = K$ and by the order parameters $m_s = m_2 = \overline{m}^2$ and $m_2 = \overline{m}$, due to the asymmetry in the model. With a little bit of algebra, the Bethe RS free entropy can be simplified to the form

$$\ln(Z) = N(1 - d)\ln(2) - \sum_i \frac{1 - d_i}{2}\ln[1 - (m_i^\star)^2] + \frac{d\theta}{2(1 - \overline{m}^2)N}\left(\sum_i m_i^\star\right)^2$$

$$+ \sum_{(ij) \in E} \ln \sum_{\sigma_i, \sigma_j} \exp\left[\left(\tilde{H}_i^\star + \sum_{k \in \partial i \backslash j} u_{k \to i}^\star\right)\sigma_i + \left(\tilde{H}_j^\star + \sum_{k \in \partial j \backslash i} u_{k \to j}^\star\right)\sigma_j + J\sigma_i\sigma_j\right]. \quad (218)$$

B. Results for random hypergraph bicoloring

We present results for $k = 3, 4, 5$ and system size $N = 10^6$ (some samples of size $N = 10^5$ are shown just for the scaling of the convergence time). Every time we present three plots in a row, the left one refers to $k = 3$, the middle one to $k = 4$, and the right one to $k = 5$.

In Fig. 10 we show the order parameter m_s that detects the planted configuration. Red closed circles correspond to the BP fixed point reached starting with the $q_0 = 0$ initial condition (i.e., with no direct information on the planted configuration), while blue open circles correspond to the BP fixed point reached from the $q_0 = 1$ initial condition (complete information on the planted configuration). For comparison we also draw with lines the results obtained by solving, via population dynamics, the cavity equations: The red (blue) line corresponds to results obtained with an initial condition (30) having $\varepsilon = 10^{-3}$ ($\varepsilon = 1$), while the green line is the result that is obtained with both initial conditions (i.e., it is independent of the initial conditions). These cavity equation results have already been plotted in Fig. 6 in slightly different

units, namely, $a = m_s/2$. Let us explain the reasoning behind this conversion. A technical statement of the tree-graph connection hypothesis of the cavity method is that the empirical distribution of the magnetizations $m_i^\star$ computed at a BP fixed point of a large finite graph, conditional on $\sigma_i^* = \tau$, should be well approximated by the distribution $P_\tau(m)$ (on the infinite tree), in the formula

$$\frac{\sum_i g(m_i^\star)\delta_{\sigma_i^*,\tau}}{\sum_i \delta_{\sigma_i^*,\tau}} \approx \mathbb{E}_\tau[g(m)], \quad (219)$$

for any function g. Applying this identity with $g(m) = m$ and recalling that for models with a global spin-flip symmetry $\mathbb{E}_+[m] = -\mathbb{E}_-[m] = \mathbb{E}[m^2] = 2a$ yields the identification $m_s = 2a$.

For each α value we run BP on three different samples and we report in the plots only the data corresponding to cases where the convergence criterion was met within 10^4 BP steps (Fig. 11 shows that this happens most of the time). Zooming in on the plots, one should be able to see three data points very close; if one fails to see three different points it is because the
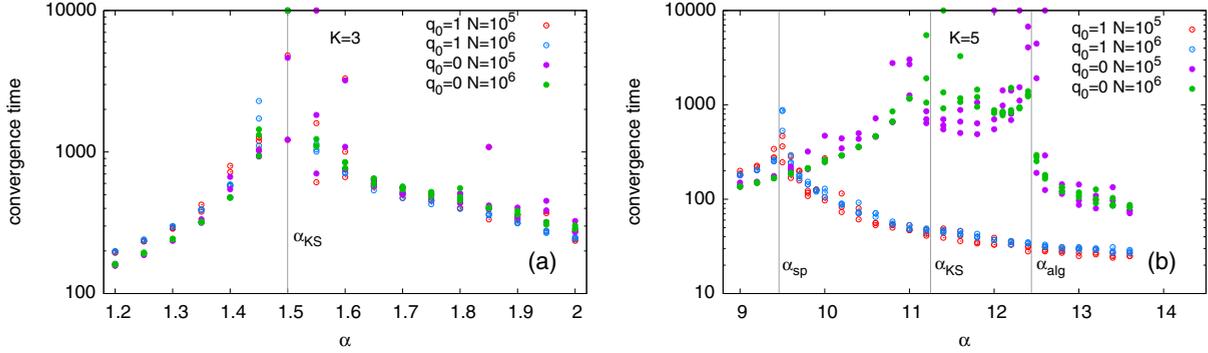
FIG. 11. Times to meet the BP convergence criterion (i.e., any marginal must change by less than 10^{-8}) for (a) $k = 3$ and (b) $k = 5$, two system sizes ($N = 10^5, 10^6$), and two different initial conditions ($q_0 = 0, 1$). A data point with a time 10^4 means BP did not reach convergence.

data from different samples perfectly coincide or because the BP on some samples did not reach convergence (the latter case happens especially at α_{KS} and very close to it).

We note that for all the cases where BP reaches a fixed point, the latter is very well described by the cavity equations derived in the thermodynamic limit for the tree reconstruction problem. When different initial conditions of BP lead to different fixed points one observes, as expected, that the one reached with $q_0 = 0$ is associated with the robust reconstruction version of the tree problem ($\varepsilon \to 0$), while $q_0 = 1$ reproduces the reconstruction one ($\varepsilon = 1$). Sample-to-sample fluctuations can be appreciated only slightly on the right of α_{KS} and close to the spinodal points at α_{sp} and α_{alg}.

We observe that for $\alpha > \alpha_{KS}$ BP with $q_0 = 0$, i.e., the version of the algorithm that does not use any direct information on the planted configuration, can actually detect it in all the samples we have studied, although for $\alpha < \alpha_{alg}$ the detection is suboptimal (hybrid-hard phase). For $\alpha < \alpha_{alg}$ BP can achieve the optimal detection only if initialized with a q_0 value large enough (we will discuss this point later, when searching for the unstable fixed point, separating the two stable fixed points already found).

In Fig. 11 we show the convergence times of BP for $k = 3$ [Fig. 11(a)] and $k = 5$ [Fig. 11(b)]. For each system size ($N = 10^5, 10^6$) and initial condition ($q_0 = 0, 1$) we study three samples.

For $k = 3$ we observe that convergence times strongly increase around the KS threshold, where also strong sample-

to-sample fluctuations arise and the size dependence can be appreciated (consider that at α_{KS} BP did not reach a fixed point for any of the $N = 10^6$ samples). Away from α_{KS} convergence times are size independent and only weakly dependent on the initial condition for $\alpha < \alpha_{KS}$.

For $k = 5$ convergence times increase not only at α_{KS}, but also at the spinodal points α_{sp} and α_{alg}. The behavior of BP strongly depends on the initial condition [recall that in the region $\alpha_{sp} < \alpha < \alpha_{alg}$ BP reaches two different fixed points depending on the value of q_0; see Fig. 10(b)]. The dependence on the system size between $N = 10^5$ and $N = 10^6$ is not evident and sample-to-sample fluctuations for $N = 10^5$ are definitely larger.

In Fig. 12 we show several order parameters that can detect the planted configuration: m_s, m_2, and q. In order to make the plot cleaner we have averaged over the samples where BP reached a fixed point.

We note that the Nishimori equality $m_s = m_2$ is very well satisfied in all the BP fixed points we reached. This is expected from the fact that the planted configuration is a typical configuration of the posterior distribution; the mean overlap with the planted configuration m_s should be equal to the mean overlap between two replicas m_2. One can also derive this property from the correspondence (219) to the tree computation and the consequence of the Bayes theorem stated as a moment identity in Sec. V.

As expected, the maximum overlap q is the largest order parameter (larger than m_s). According to Sec. IV K, q should
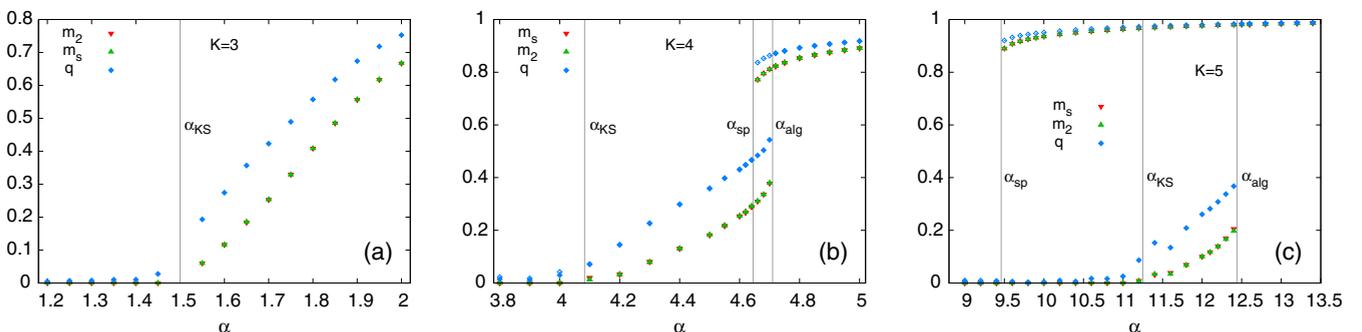
FIG. 12. Comparison between different order parameters detecting the planted configuration for (a) $k = 3$, (b) $k = 4$, and (c) $k = 5$. Data have been averaged over the samples where BP reached a fixed point. As in Fig. 10, closed (open) points correspond to $q_0 = 0$ ($q_0 = 1$).
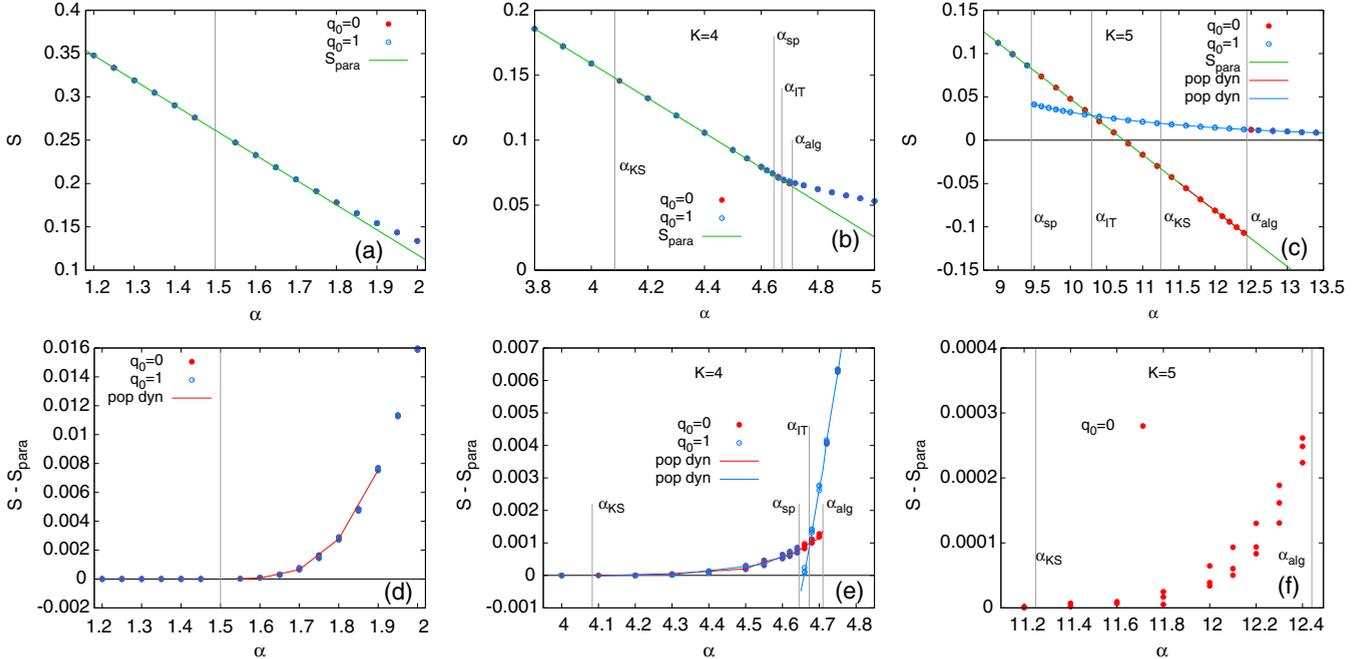
FIG. 13. Bethe RS entropy computed at the BP fixed points for (a) $k = 3$, (b) $k = 4$, and (c) $k = 5$. (d)–(f) The difference with respect to the paramagnetic entropy is shown in order to show better the difference which is tiny in some cases. For $k = 5$ the difference is too small to be estimated reliably via population dynamics and so we report only data obtained via BP.

have a square root singularity at α_{KS}; this is not visible from the data because of strong fluctuations in the neighborhood of α_{KS}.

We investigated further the properties of the BP fixed points by computing their Bethe entropy given in Eq. (197), the results being displayed in Fig. 13. In Figs. 13(a)–13(c) we also show the paramagnetic entropy (the one of the trivial fixed point) with a green straight line. A comparison with the results of the cavity equations is made with the correspondence $S = S_{\text{para}} + \phi$, with ϕ the free entropy of Eq. (172), to compensate for a different choice of normalization. The dominating fixed point, that is, the fixed point providing the right entropy of the posterior probability distribution, is the one with the largest entropy at each value of α. The case $k = 5$ clearly shows that, while on some branches the entropy can be negative, the dominating one is always positive (as it should be by definition of the entropy of a probability measure over a discrete set).

Figures 13(d)–13(f) serve to highlight the way the entropy departs from the paramagnetic curve at α_{KS}. In particular, for $k = 4$ it is possible to appreciate the crossing of the entropies obtained with the two different initial conditions ($q_0 = 0$ and $q_0 = 1$) taking place at α_{IT}, which is hardly visible in the corresponding Figs. 13(a)–13(c). Note that for $k = 5$ the difference $S - S_{\text{para}}$ is extremely small, hence we did not manage to estimate it from the population dynamics algorithm. Nonetheless, the entropy computed on the BP fixed point clearly shows a stable increase with respect to the paramagnetic value, with quite strong sample-to-sample fluctuations (given by the spread of the three points in the figure).

We have shown evidence that for $k \geqslant 4$, in the range $\alpha_{\text{sp}} \leqslant \alpha \leqslant \alpha_{\text{alg}}$ there exist at least two different fixed points

of BP (along with their symmetric partners under the global spin-flip symmetry). At α_{IT} the entropies of these two BP fixed point cross and this corresponds, in the thermodynamic limit, to a first-order transition between two different thermodynamic states. Running BP without information on the planted configuration ($q_0 = 0$), the drastic change in the posterior distribution taking place at α_{IT} is not visible: The algorithm remains confined to the low-informative branch until α_{alg}. By analogy with the bifurcation theory of fixed-point systems of equations (even if α here is not a parameter modifying smoothly a set of equations of constant dimension), we expect a branch of unstable fixed points of BP to exist in the range of parameters $\alpha_{\text{sp}} \leqslant \alpha \leqslant \alpha_{\text{alg}}$ and to connect these two branches as in the scalar toy model sketched in Fig. 2. However, getting evidence of the existence of this unstable fixed point is very difficult, because it is repulsive by its very nature.

We report in Fig. 14 the evidence we have gathered in favor of this hypothesis. In Fig. 14(a) we show the evolution of the order parameter m_s as a function of the number of iterations of the BP updates, for different values of the initialization parameter q_0. During the evolution all variables are updated; in order words, q_0 is *not* the fraction of variables pinned to the value of the planted configuration. So any fixed point we find is a fixed point of the standard BP algorithm. The initial condition is just used to aim the algorithm at different fixed points, but then the BP algorithm is unconstrained. The evolution of BP at $\alpha = 4.68$ with different values of q_0 reported in Fig. 14(a) clearly shows that for q_0 large enough BP converges to the same high-information fixed point reached for $q_0 = 1$, while for q_0 small enough only the low-information fixed point can be reached. Note that it is very hard to precisely define a separatrix value q_0^* that separates the two regimes, because BP is a stochastic algorithm, so its behavior is not deterministically fixed by the
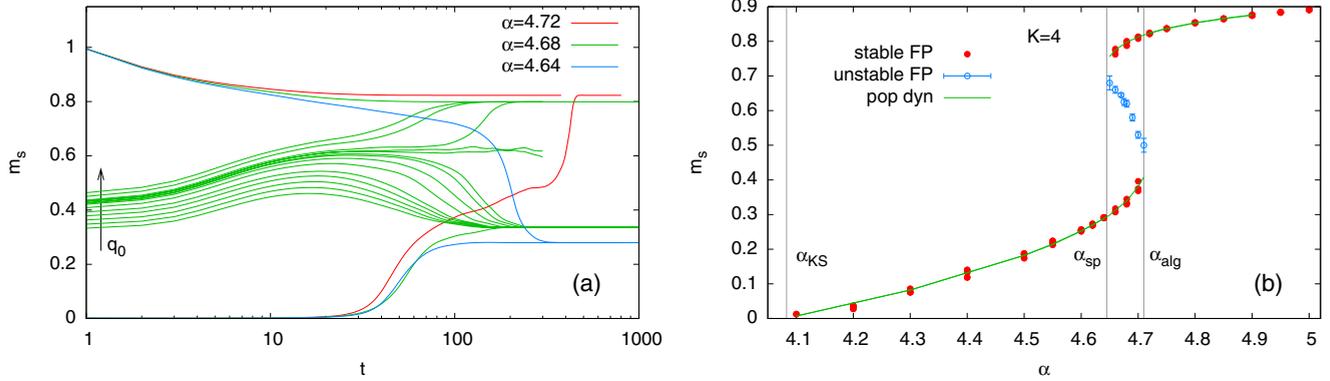
FIG. 14. Search for the unstable fixed point of BP in the range $[\alpha_{sp}, \alpha_{alg}]$. In (a) the lowermost (uppermost) curves have been obtained with $q_0 = 0$ ($q_0 = 1$); the rest of the curves for $\alpha = 4.68$ have been obtained varying q_0 by $\Delta q_0 = 0.01$ in the range $[0.2,0.29]$ and by $\Delta q_0 = 0.002$ in the range $[0.26,0.27]$. (b) The value of m_s at the various fixed points as a function of α.

initial condition, but by the random numbers used during the evolution to choose the order of update of the BP messages. This means that, for the same $q_0 \approx q_0^*$, it may happen that two evolutions converge to different fixed points. Nevertheless, Fig. 14(a) show evidence that some evolutions of BP (those with $q_0 = 0.266$ and $q_0 = 0.268$ in the present figure) remain for a long time on a stationary regime which is neither of the two stable fixed points already found: We define this regime as the unstable fixed point of BP. Although the precise value of q_0 is not significant for the argument we just made, the value of m_s on the unstable fixed point can be measured with a reasonable uncertainty. We repeated the above procedure for several α values in the range $[\alpha_{sp}, \alpha_{alg}]$ and we report in Fig. 14(b) the summary of the results for $k = 4$, with blue open symbols showing the unstable branch in addition to the results already presented in Fig. 10.

C. Results for the asymmetric stochastic block model with two groups

As discussed in Sec. IV I, the asymmetric SBM with two groups undergoes different kinds of phase transitions depending on the level of the asymmetry: For $|\overline{m}| < \overline{m}_c = 1/\sqrt{3}$ the transition is continuous, while for $|\overline{m}| > \overline{m}_c$ the transition is

discontinuous. We focused our analysis on two values of $\overline{m}$, namely, $\overline{m} = 0.4$ belonging to the former case and $\overline{m} = 0.8$ belonging to the latter case. We will present results for a rather small average degree $d = 4$, for which the Kesten-Stigum transition occurs at $\theta_{KS} = 0.5$, but we have checked that the same conclusions apply to the case $d = 8$. We run BP only on problems of a very large size ($N = 10^7$) in order to reduce as much as possible the finite-size effects.

We show in Fig. 15 a comparison of the properties of the fixed point reached by BP on single samples with the results of the population dynamics study of the tree problem, concentrating on values of θ close to the Kesten-Stigum transition, where fluctuations become more significant. In Fig. 15(a) we show data for $\overline{m} = 0.4$ where the transition is continuous: We report the values of both m_s and m_2 at the fixed point in order to show how well the Nishimori condition is satisfied on a given sample. For each sample (we run BP on three different samples for each θ value) both runs with $q_0 = 0$ and $q_0 = 1$ converge to exactly the same fixed point or did not converge within the maximum number of iterations $t_{max} = 10^4$ (the latter happened for all three samples at $\theta = 0.5$ and for one sample at $\theta = 0.502$). We conclude that the main effects of being close to the
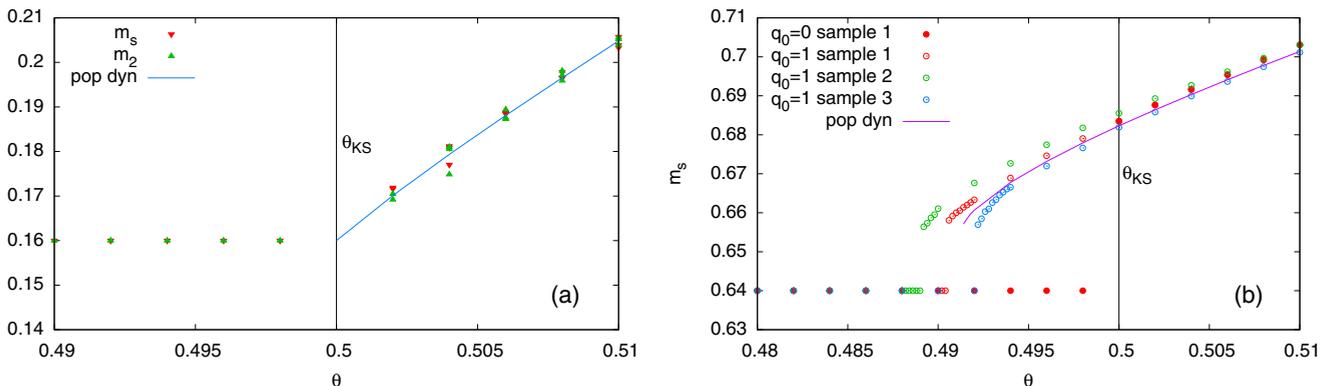
FIG. 15. Comparison between BP and population dynamics in detecting the planted configuration in the asymmetric SBM with $d = 4$. For each θ value we have simulated three samples of size $N = 10^7$. The asymmetry is (a) $\overline{m} = 0.4 < \overline{m}_c$ and (b) $\overline{m} = 0.8 > \overline{m}_c$. In both cases the planted configuration can be detected purely from the graph ($q_0 = 0$) for $\theta > \theta_{KS}$. Despite visible sample-to-sample fluctuations, the prediction by population dynamics is always very faithful.
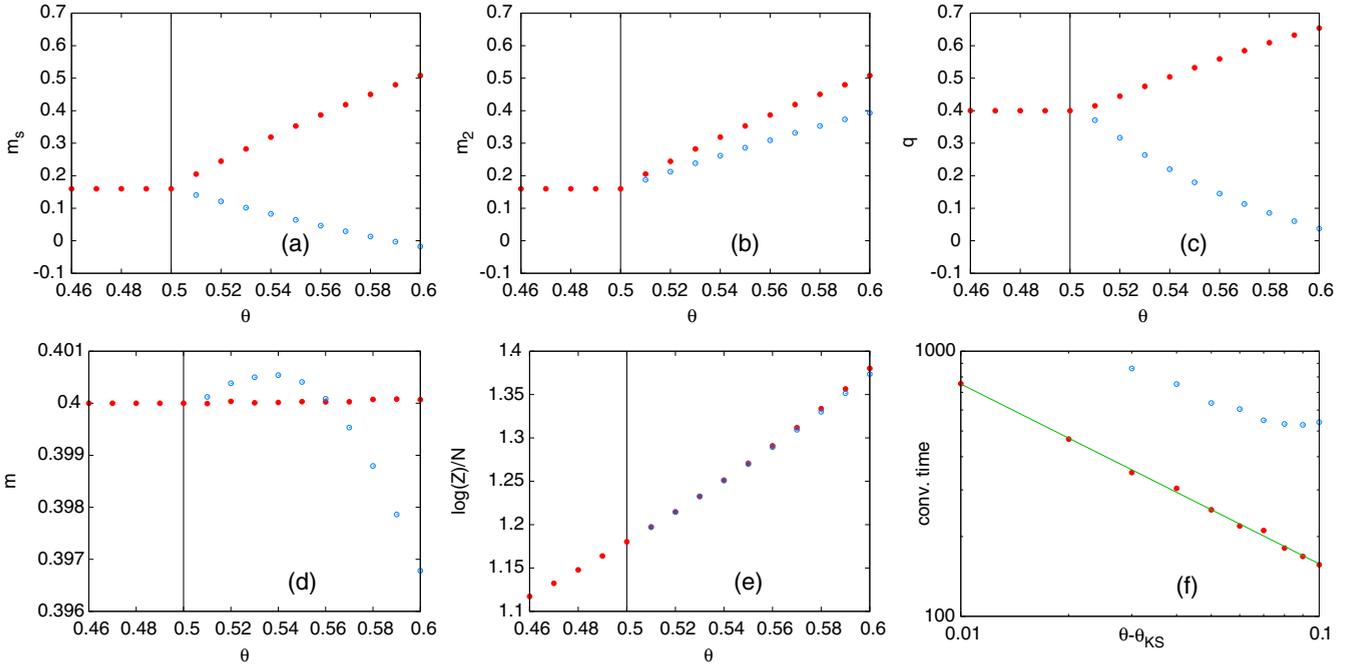
FIG. 16. Different fixed points reached by BP run with an uninformative initial condition on samples of the asymmetric SBM ($d = 4$, $\overline{m} = 0.4$, and $N = 10^7$). Red closed circles represent the dominating fixed point having the largest $\ln(Z)$, while blue open circles represent the subdominating fixed point. (a)–(f) Several observables, indicated on the axis, are displayed for the two fixed points as a function of θ.

critical point are the lack of convergence and some visible sample-to-sample fluctuations (recall that we run BP on samples of size $N = 10^7$). Nevertheless, the prediction from population dynamics is accurate even in the vicinity of θ_{KS}.

In Fig. 15(b) we show data for the overlap with the planted configuration m_s measured in three samples of size $N = 10^7$ and asymmetry $\overline{m} = 0.8 > \overline{m}_c$. Starting from an uninformative initial condition ($q_0 = 0$), the planted configuration can be detected only for $\theta \geqslant \theta_{KS}$, while with an informative initial condition ($q_0 = 1$) the planted configuration can be detected for $\theta \geqslant \theta_{sp}$. The Nishimori condition $m_2 = m_s$ is verified with such good accuracy that we do not plot the data for m_2, the points being almost perfectly superimposed. The prediction of the population dynamics is faithful, although sample-to-sample fluctuations are still clearly visible, even for such large sizes. The quantity that changes the most from sample to sample is the location of the spinodal point θ_{sp}, where the informative fixed point first appears: Given θ_{sp}, the rest of the curve is pretty well conserved and sample independent.

One may wonder how we can follow a sample varying θ [see Fig. 15(b)] given that at different θ values the graph has a different number of links within and between the communities. The construction of the graph for each value of θ is a random process and we identify the sample with the random seed used by the algorithm. The stochastic algorithm that builds the graph is such that when it is run with the same seed (and so with the same sequence of pseudorandom numbers) and two not too different θ values, most of the links are in common between the two graphs. This is the reason why we can follow a sample in θ.

The very good agreement between the BP results and the population dynamics shown in Fig. 15 actually hides a

subtle point that we will now explain in the case $\overline{m} = 0.4$, for which the transition at θ_{KS} is continuous. We have indeed observed that, with the uninformative initial condition ($q_0 = 0$), different runs of BP on the same sample converge to two different fixed points when $\theta > \theta_{KS}$, with comparable probability (as soon as $q_0 > 0$ this issue disappears, but as an inference algorithm BP should not use any information on the planted configuration). The properties of these two fixed points are displayed in Fig. 16 (for $\theta \leqslant \theta_{KS}$ the trivial fixed point is unique and has $m_s = m_2 = \overline{m}^2$ and $m = q = \overline{m}$). We report with red closed circles the data corresponding to the dominating fixed point, i.e., the one having the largest value of $\ln(Z)$, and with blue open circles the data corresponding to the subdominant fixed point. Comparing Figs. 16(a) and 16(b), we see that the dominating fixed point satisfies the Nishimori condition $m_s = m_2$, while the subdominant one does not. Also, the condition $m = \overline{m}$ is satisfied by the dominant fixed point (within finite-size fluctuations), while the subdominant fixed point typically violates such a condition. For all these reasons we consider the dominant fixed point as the correct one, which is why we reported in Fig. 15 only the data corresponding to this one.

The existence of two fixed points beyond the KS threshold can be justified by the following reasoning. Consider first the symmetric case $\overline{m} = 0$, which enjoys an exact global spin-flip symmetry (corresponding to the exchange of the two communities in the SBM interpretation). As a consequence of this symmetry there must be (at least) two nontrivial fixed points that arise at the Kesten-Stigum transition from the bifurcation of the trivial one, these two fixed points having opposite values of the local magnetizations. Consider now a slight increase of $\overline{m}$; by continuity one expects the two
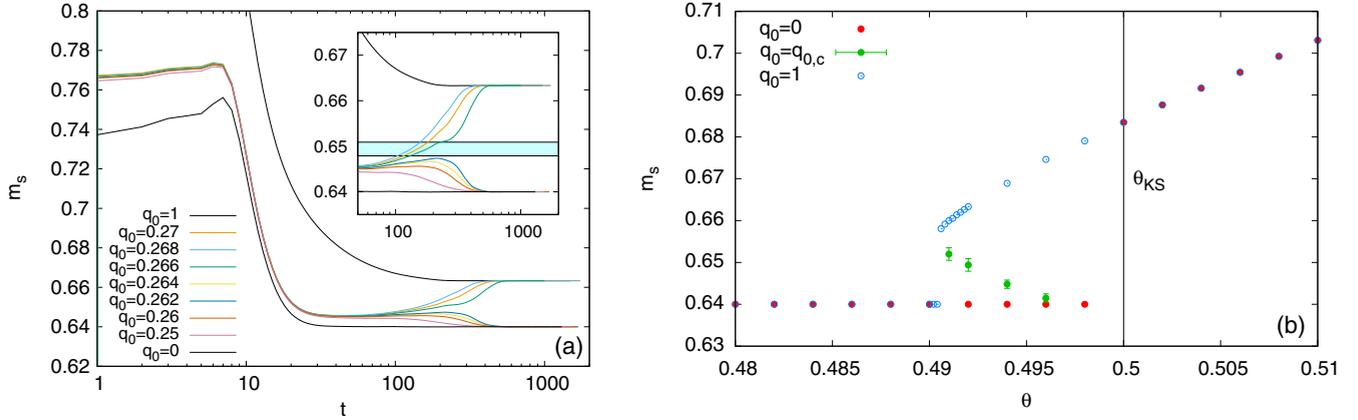
FIG. 17. (a) Illustration of the way we estimate the value of m_s for the unstable fixed point of BP (here $\theta = 0.492$), which plays the role of separatrix for the BP dynamics. (b) This value of m_s obtained in the coexistence region $\theta_{sp} < \theta < \theta_{KS}$ (here $\overline{m} = 0.8$).

fixed points to persist, even if they are no longer symmetric to each other, and in particular they do not have the same value for their free entropy $\ln(Z)$. Somewhat unexpected is the observation that BP reaches both fixed points with similar probabilities. The only difference is in the time BP takes to reach these two fixed points; this is shown in Fig. 16(f), where it is evident that reaching the dominating fixed point requires a smaller number of iterations (the interpolating power-law fit has an exponent approximately equal to $-2/3$). It seems that the small difference in $\ln(Z)$ between the two fixed points does not influence enough the BP evolution at the very beginning, when the choice of the basin of attraction is made; the only effect is to speed up the evolution in case the basin of attraction of the dominating fixed point is eventually chosen.

From the data shown in Fig. 16 it is clear the dominating fixed point is the one we would like BP to reach, since it is the one better correlated with the planted configuration satisfying the right conditions. However, when running BP from an uninformative initial condition we may likely end up at the subdominating fixed point. The simplest solution is to run BP several times, until both fixed points are found and the one with largest Z is then chosen. Actually this procedure may be slow, and so we have found a more efficient way to reach both fixed points in just one run of BP.

At θ_{KS} the bifurcation of the trivial fixed point produces two fixed points which are on opposite sides, that is, the unstable fixed point stays more or less at the midpoint of the line joining the two stable fixed points. This property remains approximately true also for $\theta > \theta_{KS}$ and we can exploit it in order to jump from one fixed point to the other. Once BP reaches a fixed point with messages $\{u^\star_{i\to j}\}$ we can produce an initial condition for a second BP run with the transformation

$$u^{(0)}_{i\to j} = 2K - u^\star_{i\to j} \qquad (220)$$

(recall that the uninformative fixed point has all messages equal to K). We have checked that this initial condition always leads BP to the other fixed point. Thus we can find both fixed points with just two runs (the second being very fast, due to the initial condition which is pretty close to the fixed point).

The BP initial condition for the first run, $u^{(0)}_{i\to j} = L\tau^0_i$ with $L = 100$ and τ^0_i distributed according to Eq. (215), is quite

drastic (i.e., has very strong messages) in order to be far from the uninformative fixed point. For $\theta > \theta_{KS}$ this is actually not necessary, given that the uninformative fixed point is locally unstable. So, in order to check that all what we have described above does not depend on such a drastic first initial condition we have repeated the numerical experiments in the regime $\theta > \theta_{KS}$ by using a different initial condition, which is much closer to the uninformative fixed point

$$u^{(0)}_{i\to j} = K\big(1 + \varepsilon\tau^0_i\big), \qquad (221)$$

with $\varepsilon = 10^{-3}$ and τ^0_i again distributed according to Eq. (215). The results are identical to those with the drastic first initial condition.

For the sake of completeness, let us also mention that for $\theta > \theta_{KS}$ and $\overline{m}$ large enough (e.g., at $\theta = 0.6$ in the range $\overline{m} \gtrsim 0.46$) we observed a divergence of the BP convergence time to the subdominant fixed point. Increasing $\overline{m}$ further, some runs of BP converge to the dominant fixed point and others keep wandering in a region of the message space that is far from the dominant fixed point but no longer contains a strict fixed point. This phenomenon could be interpreted as a spontaneous replica symmetry breaking of the subdominant fixed point. To fix the algorithm in such a way that it always converges to the dominant fixed point, we adopted the following rule: Stop BP after t_{max} iterations and apply the transformation in Eq. (220) to the current BP messages. Even if the BP messages do not correspond to a true fixed point, their transformation is actually close enough to the other fixed point to make the second BP run converge to the dominant fixed point with high probability (this trick worked for all the hundreds of samples we tried).

Finally, we present in Fig. 17 a study of the unstable branch of fixed points of BP for the discontinuous case $\overline{m} > \overline{m}_c$ in the range $\theta_{sp} < \theta < \theta_{KS}$. Similarly to what we did for the hypergraph bicoloring (cf. Fig. 14), well chosen parameters q_0 for the initial condition allow us to identify the value of m_s that plays the role of separatrix for the BP evolution (this value with its uncertainty is marked by a cyan strip in the inset). In Fig. 17(b) we report the data of the first sample already shown in Fig. 15(b) together with the values of m_s measured at the unstable fixed point in the coexistence region $\theta_{sp} < \theta < \theta_{KS}$.

VII. CONCLUSION

We have discussed in this paper the typology of phase transitions in inference problems, emphasizing in particular the possible existence of hybrid-hard phases. Our main technical contribution is a rather generic expansion of the functional cavity equations for sparse models around their trivial fixed point. Let us briefly sketch some possible directions for future work.

Some of our results could probably be proven rigorously. In particular, the techniques of [22] combined with the higher-order expansions we obtained should yield the nontightness of the Kesten-Stigum bound for the reconstruction of the $q = 4$ antiferromagnetic Potts model on a tree with low enough degree (more directly than using the bound on the information-theoretic threshold of [45]).

We performed our expansions assuming either pairwise interactions between arbitrary discrete spins or k-wise interactions between binary-value variables. One could relax this assumption and consider k-wise interactions between q-value variables, as was done in Ref. [25] (but truncating the expansion to a low order).

We believe the $q_1 + q_2$ SBM we introduced warrants further investigation. We argued that its phase diagram must contain some hybrid-hard phases, but probably in a narrow range of parameters. As a first step the large-degree limit of this model could be studied. The corresponding dense model also exhibits this hybrid-hard phenomenon while being much easier to study, with evolution equations bearing on covariance matrices of Gaussian random variables instead of probability distributions.

ACKNOWLEDGMENTS

APPENDIX A: NUMERICAL RESOLUTION OF THE CAVITY EQUATIONS

1. Population dynamics algorithm

We give in this Appendix some details on the numerical procedure we used to solve the cavity equations and produce the curves presented in the main text. It has become customary to solve recursive distributional equations like (33) by a population dynamics method [48,60]. The main idea of this method is to approximate a probability distribution, for instance, $P_\tau^{(n)}$, by the empirical distribution of a large number $\mathcal{N} \gg 1$ of representatives of $P_\tau^{(n)}$, namely,

$$P_\tau^{(n)}(\eta) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(\eta - \eta^{(n,\tau,i)}), \tag{A1}$$

where the $\eta^{(n,\tau,i)}$ are independent and identically distributed samples from $P_\tau^{(n)}$. The iterative equation (33) is then translated into a rule to generate the representative elements (i.e., the population) at iteration $n + 1$ from the one at iteration n, as follows. For each of the τ and independently for $i = 1, \ldots, \mathcal{N}$, (i) draw ℓ from $\tilde{p}_\ell$, (ii) draw $\tau_1, \ldots, \tau_\ell$ with probability $M_{\tau\tau_1}, \ldots, M_{\tau\tau_\ell}$, (iii) draw $i_1, \ldots, i_\ell$ independently, uniformly in $\{1, \ldots, \mathcal{N}\}$, and (iv) set $\eta^{(n+1,\tau,i)} = f(\eta^{(n,\tau_1,i_1)}, \ldots, \eta^{(n,\tau_\ell,i_\ell)})$.

Average values over the conditional distributions $P_\tau^{(n)}$ or over the unconditional distribution $P^{(n)}$ are then evaluated as empirical averages over the population: For an arbitrary function f one computes

$$\mathbb{E}_\tau^{(n)}[f(\eta)] = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} f(\eta^{(n,\tau,i)}),$$

$$\mathbb{E}^{(n)}[f(\eta)] = \sum_\tau \overline{\eta}_\tau \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} f(\eta^{(n,\tau,i)}). \tag{A2}$$

This approach is very natural, simple to implement, and becomes exact in the limit $\mathcal{N} \to \infty$. Unfortunately, it suffers in many cases from an instability problem at finite $\mathcal{N}$ that requires additional care. In order to explain the origin of this difficulty we recall that the unconditional distribution must obey, for all iterations n, the condition $\mathbb{E}^{(n)}[\eta] = \overline{\eta}$. A consequence of this identity and of the Bayes theorem as stated in Eq. (22) is that, for all τ, the conditional distributions should obey

$$\mathbb{E}_\tau^{(n)}\left[\frac{\overline{\eta}_\tau}{\eta_\tau}\right] = 1. \tag{A3}$$

The elements of the population at iteration $n + 1$ should thus be such that

$$\frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{\overline{\eta}_\tau}{\eta_\tau^{(n+1,\tau,i)}} = 1. \tag{A4}$$

If the condition $\mathbb{E}^{(n)}[\eta] = \overline{\eta}$ is verified exactly then it will also be the case at iteration $n + 1$; however, when $\mathcal{N}$ is finite there are necessarily some small fluctuations around this value that turn out to be amplified under the iteration whenever $\tilde{\mathbb{E}}[\ell]|\theta_2| > 1$ (as in the expression of the KS transition, but without the square on the eigenvalue). In that case the implementation presented above is bound to fail because of this instability.

Depending on the model, there are different ways to treat this problem. If the problem has an explicit symmetry, like the permutation invariance for the symmetric q-state Potts model or the up-down symmetry for the Ising models studied in Sec. V, this can be exploited to tame the instability (see below for more details on the practical implementation).

The most difficult cases arise when there is no explicit symmetry to use, for instance, for the asymmetric Ising model studied in Sec. IV I with $\overline{m} \neq 0$. In such a case one needs to compute, at each iteration, the left-hand side of (A4) from the newly generated elements and if this average is not equal to 1 apply some transformation rules on the population samples in order to bring it closer to this target value. There are still

many ways to implement this idea; we will be more explicit below for the asymmetric Ising case.

2. Symmetric problems

As mentioned above, the population dynamics algorithm can be simplified and stabilized when the problem to solve exhibits some additional symmetries. Consider, for instance, the symmetric Potts model with q states, as defined in Sec. IV H. The invariance under any permutation π of the q states implies that $P^{(n)}(\eta) = P^{(n)}(\eta \circ \pi)$, where $(\eta \circ \pi)_\sigma = \eta_{\pi(\sigma)}$ is the probability measure on χ obtained from η by the reshuffling π of the q states. This observation allows us to express all conditional distributions P_τ in terms of a single one, for instance, $\tau = 1$, and close the iteration equation for P_1 as

$$P_1^{(n+1)}(\eta) = \sum_{\ell=0}^{\infty} \tilde{p}_\ell \sum_{\pi_1,\ldots,\pi_\ell} \rho(\pi_1) \cdots$$

$$\rho(\pi_\ell) \int dP_1^{(n)}(\eta^1) \cdots dP_1^{(n)}(\eta^\ell)$$

$$\times \delta(\eta - f(\eta^1 \circ \pi_1, \ldots, \eta^\ell \circ \pi_\ell)), \quad (A5)$$

where the π are permutations on q colors and ρ is a probability distribution on them such that with probability $\frac{1}{q} + \theta(1 - \frac{1}{q})$ the permutation π is uniform under the condition $\pi(1) = 1$; otherwise it is uniform under the condition $\pi(1) \neq 1$. This equation is much simpler than the generic one, as it involves only one population instead of q, and also much more stable numerically.

Even simpler is the treatment of Ising spin models ($q = 2$) which are symmetric under the spin-reversal operation, for instance, those studied in Sec. V. In that case the probability laws on χ are encoded by a single real m, the magnetization, and the conditional and unconditional distributions verify the identities $P^{(n)}(m) = P^{(n)}(-m)$ and $P_-^{(n)}(m) = P_+^{(n)}(-m)$. This allows us to close the cavity equation for a single population $P_+^{(n)}(m)$ [see, in particular, Eqs. (167) and (168)].

3. Asymmetric Ising model

Let us finally turn to the case of the asymmetric Ising model, defined in Sec. IV I. The general BP equation (34) for η_σ can be written as a scalar recursion by parametrizing $\eta_\sigma = \frac{1+\sigma m}{2}$. The BP equation reads, in this parametrization, $m = f(m^1, \ldots, m^\ell)$, with

$$m = \frac{z_+(m^1, \ldots, m^\ell) - z_-(m^1, \ldots, m^\ell)}{z(m^1, \ldots, m^\ell)},$$

$$z(m^1, \ldots, m^\ell) = z_+(m^1, \ldots, m^\ell) + z_-(m^1, \ldots, m^\ell), \quad (A6)$$

$$z_+(m^1, \ldots, m^\ell) = \frac{1+\overline{m}}{2} \prod_{i=1}^{\ell} \left(1 + \theta \frac{m^i - \overline{m}}{1 + \overline{m}}\right),$$

$$z_-(m^1, \ldots, m^\ell) = \frac{1-\overline{m}}{2} \prod_{i=1}^{\ell} \left(1 - \theta \frac{m^i - \overline{m}}{1 - \overline{m}}\right). \quad (A7)$$

One can see immediately that $z_\sigma(\overline{m}, \ldots, \overline{m}) = \frac{1+\sigma\overline{m}}{2}$, hence $z(\overline{m}, \ldots, \overline{m}) = 1$ and $\overline{m} = f(\overline{m}, \ldots, \overline{m})$ is a fixed point, as expected.

The naive implementation of the population dynamics algorithm explained above corresponds to a representation of $P_+^{(n)}(m)$ [$P_+^{(n)}(m)$] by $\mathcal{N}$ reals $m^{(n,+,i)}$ ($m^{(n,-,i)}$). In order to cure the instability and to enforce the condition (A4) at each iteration, we have proceeded as follows: Once the $2\mathcal{N}$ magnetizations $m^{(\sigma,i)}$ have been generated according to the standard procedure (we remove the iteration index to simplify the notation) we compute

$$\beta^+ = \frac{2}{1-\overline{m}} \left(\frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{1+\overline{m}}{1+m^{(+,i)}} - \frac{1+\overline{m}}{2}\right),$$

$$\beta^- = \frac{2}{1+\overline{m}} \left(\frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{1-\overline{m}}{1-m^{(-,i)}} - \frac{1-\overline{m}}{2}\right) \quad (A8)$$

and replace the elements of the population by

$$m'^{(+,i)} = \frac{\beta^+(1+m^{(+,i)}) - (1-m^{(+,i)})}{\beta^+(1+m^{(+,i)}) + (1-m^{(+,i)})},$$

$$m'^{(-,i)} = \frac{(1+m^{(-,i)}) - \beta^-(1-m^{(-,i)})}{(1+m^{(-,i)}) + \beta^-(1-m^{(-,i)})}. \quad (A9)$$

One can check that this procedure strictly enforces the identity (A4), the coefficients β measuring the deviation of the naively generated population elements from the symmetry respecting probability distributions.

APPENDIX B: MOMENT EXPANSION FOR PAIRWISE INTERACTION POTTS VARIABLES

The goal of this Appendix is to justify some of the statements made in Sec. IV D, in particular the ansatz (61) for the scaling of the centered moments of the perturbative nontrivial fixed point and the third-order expansion given in Eqs. (71)–(73).

Let us consider a fixed-point solution $P(\eta)$ of the recursion equation (37) and look for a hierarchy of equations between its centered moments. As in the main text we define $\delta_\sigma = \eta_\sigma - \overline{\eta}_\sigma$ as the centered random variable with distribution P and $\hat{\delta}_\sigma = \sum_{\sigma'} \hat{M}_{\sigma\sigma'} \delta_{\sigma'}$ as a linearly transformed version of δ. Denoting by $\mathbb{E}[\bullet]$ the average with respect to P, we have by definition $\mathbb{E}[\delta_\sigma] = 0$ and we assume $\mathbb{E}[\delta_\sigma \delta_{\sigma'}]$ to be nonzero but small, of an order denoted by κ. From the self-consistent equation (37) on P and the expression (51) of the function f we can write the pth centered moment of P ($p \geqslant 2$ is understood below) as

$$\mathbb{E}[\delta_{\sigma_1}, \ldots, \delta_{\sigma_p}] = \mathbb{E}\left[\left(\frac{z_{\sigma_1}}{z} - \overline{\eta}_{\sigma_1}\right) \cdots \left(\frac{z_{\sigma_p}}{z} - \overline{\eta}_{\sigma_p}\right) z\right], \quad (B1)$$

where on the right-hand side $z_\sigma = \overline{\eta}_\sigma \prod_{i=1}^{\ell}(1 + \hat{\delta}_\sigma^i)$, $z = \sum_\gamma z_\gamma$, ℓ is drawn from $\tilde{p}_\ell$, and for $i = 1, \ldots, \ell$ the $\hat{\delta}^i$ are linearly transformed [according to (52)] versions of η^i independent and identically distributed samples drawn from $P(\eta)$. To simplify this expression we introduce the notation $\varepsilon_\sigma = \prod_{i=1}^{\ell}(1 + \hat{\delta}_\sigma^i) - 1$ in such a way that $z_\sigma = \overline{\eta}_\sigma(1 + \varepsilon_\sigma)$ and $z = 1 + \sum_\gamma \overline{\eta}_\gamma \varepsilon_\gamma$. Equation (B1)

thus becomes

$$\mathbb{E}\big[\delta_{\sigma_1},\ldots,\delta_{\sigma_p}\big] = \overline{\eta}_{\sigma_1}\cdots\overline{\eta}_{\sigma_p}\mathbb{E}\left[\left(\varepsilon_{\sigma_1}-\sum_{\gamma_1}\overline{\eta}_{\gamma_1}\varepsilon_{\gamma_1}\right)\cdots\left(\varepsilon_{\sigma_p}-\sum_{\gamma_p}\overline{\eta}_{\gamma_p}\varepsilon_{\gamma_p}\right)\left(1+\sum_{\gamma}\overline{\eta}_{\gamma}\varepsilon_{\gamma}\right)^{1-p}\right]$$

$$= \overline{\eta}_{\sigma_1}\cdots\overline{\eta}_{\sigma_p}\sum_{m=0}^{\infty}\binom{1-p}{m}\sum_{\gamma_1,\ldots,\gamma_{p+m}}\big(\delta_{\sigma_1,\gamma_1}-\overline{\eta}_{\gamma_1}\big)\cdots\big(\delta_{\sigma_p,\gamma_p}-\overline{\eta}_{\gamma_p}\big)\overline{\eta}_{\gamma_{p+1}}\cdots\overline{\eta}_{\gamma_{p+m}}\mathbb{E}\big[\varepsilon_{\gamma_1},\ldots,\varepsilon_{\gamma_{p+m}}\big], \quad \text{(B2)}$$

where the binomial coefficient with negative argument takes its conventional value $\binom{1-p}{m}=(-1)^m\binom{p-2+m}{m}$. In order to close these equations we should now express the average of products of the ε in terms of centered moments of P. As an intermediate step let us further define $\mu_\sigma = \varepsilon_\sigma + 1 = \prod_{i=1}^{\ell}(1+\hat{\delta}_\sigma^i)$. Exploiting the independence of the random variables η^i for distinct i, one can easily compute the average of products of the μ,

$$\mathbb{E}\big[\mu_{\sigma_1},\ldots,\mu_{\sigma_p}\big] = \sum_{\ell=0}^{\infty}\tilde{p}_\ell\left(1+\sum_{S}\mathbb{E}\left[\prod_{j\in S}\hat{\delta}_{\sigma_j}\right]\right)^{\ell}, \quad \text{(B3)}$$

where S is summed over subsets of $\{1,\ldots,p\}$ of cardinality $|S|\geqslant 2$ (because of the property $\mathbb{E}[\hat{\delta}_\sigma]=0$). Expanding the last power, we rewrite this as

$$\mathbb{E}\big[\mu_{\sigma_1},\ldots,\mu_{\sigma_p}\big] = 1 + \sum_{r=1}^{\infty}\tilde{\mathbb{E}}\left[\binom{\ell}{r}\right]\sum_{S_1,\ldots,S_r}\mathbb{E}\left[\prod_{j\in S_1}\hat{\delta}_{\sigma_j}\right]\cdots\mathbb{E}\left[\prod_{j\in S_r}\hat{\delta}_{\sigma_j}\right], \quad \text{(B4)}$$

where the subsets S satisfy the same properties as above and the average denoted by $\tilde{\mathbb{E}}$ is over ℓ drawn with probability $\tilde{p}_\ell$, the requirement $\ell\geqslant r$ being kept understood. We can now return to the computation of the average of products of the ε, noting that

$$\mathbb{E}\big[\varepsilon_{\sigma_1},\ldots,\varepsilon_{\sigma_p}\big] = \mathbb{E}\big[(\mu_{\sigma_1}-1)\cdots(\mu_{\sigma_p}-1)\big] = \sum_{T}(-1)^{p-|T|}\mathbb{E}\left[\prod_{j\in T}\mu_{\sigma_j}\right], \quad \text{(B5)}$$

where T runs over all subsets of $\{1,\ldots,p\}$. Before proceeding let us recall the following version of the inclusion-exclusion principle: If E is an arbitrary finite set, T runs over all the subsets of E, and F is a subset of E, then

$$\sum_{T\subset E}(-1)^{|E|-|T|}\mathbb{I}(F\subset T) = (1-1)^{|E|-|F|} = \mathbb{I}(F=E), \quad \text{(B6)}$$

which can be easily proven by enumerating the number of subsets T that contain F and that are contained in E. Consider now an arbitrary function $g(S)$ of the subsets of E, and an integer r; as a consequence of the above identity we have

$$\sum_{T\subset E}(-1)^{|E|-|T|}\sum_{S_1\subset T,\ldots,S_r\subset T}g(S_1)\cdots g(S_r) = \sum_{S_1\subset E,\ldots,S_r\subset E}g(S_1)\cdots g(S_r)\sum_{T\subset E}(-1)^{|E|-|T|}\mathbb{I}(S_1\subset T)\cdots\mathbb{I}(S_r\subset T)$$

$$= \sum_{S_1\subset E,\ldots,S_r\subset E}g(S_1)\cdots g(S_r)\mathbb{I}(S_1\cup\cdots\cup S_r=E). \quad \text{(B7)}$$

Inserting (B4) in Eq. (B5) and using this last form of the inclusion-exclusion principle yields finally

$$\mathbb{E}\big[\varepsilon_{\sigma_1},\ldots,\varepsilon_{\sigma_p}\big] = \tilde{\mathbb{E}}[\ell]\mathbb{E}\big[\hat{\delta}_{\sigma_1},\ldots,\hat{\delta}_{\sigma_p}\big] + \sum_{r=2}^{\infty}\tilde{\mathbb{E}}\left[\binom{\ell}{r}\right]\sum_{S_1,\ldots,S_r}\mathbb{E}\left[\prod_{j\in S_1}\hat{\delta}_{\sigma_j}\right]\cdots\mathbb{E}\left[\prod_{j\in S_r}\hat{\delta}_{\sigma_j}\right], \quad \text{(B8)}$$

where we have treated the term $r=1$ separately as it is simpler than the general terms for $r\geqslant 2$. In the latter the summation is over $S_1,\ldots,S_r$ subsets of $\{1,\ldots,p\}$ of cardinality $|S_i|\geqslant 2$ whose union must cover $\{1,\ldots,p\}$ (but they do not need to be disjoint). Each term in this sum can be associated with a diagram, i.e., an hypergraph on p vertices with r hyperedges that connect at least two vertices in such a way that no vertex remains isolated. Furthermore, in Eq. (B2) some of the ε_σ appear multiplied by $\overline{\eta}_\sigma$ and summed over σ; in

that case the only diagrams that contribute are those in which the corresponding vertex has degree of at least 2, because of the property $\sum_\sigma \overline{\eta}_\sigma\hat{\delta}_\sigma = 0$.

The two expressions (B2) and (B8) form an infinite hierarchy of equations that formally determine all the centered moments of the possible distribution $P(\eta)$ fixed-point solutions of (37); as η is bounded, these moments are enough to characterize P itself. The results presented in Sec. IV D have been obtained by truncating this hierarchy under the

hypothesis that

$$\mathbb{E}\big[\delta_{\sigma_1},\dots,\delta_{\sigma_p}\big] = O(\kappa^{\lceil p/2\rceil}), \quad \mathbb{E}\big[\varepsilon_{\sigma_1},\dots,\varepsilon_{\sigma_p}\big] = O(\kappa^{\lceil p/2\rceil}),$$

(B9)

where κ is a small parameter controlling the distance between the studied fixed point $P(\eta)$ and the trivial one $\delta(\eta - \overline{\eta})$. The self-consistency of this ansatz can be checked in Eq. (B2) and (B8). Using the hypothesis on $\mathbb{E}[\varepsilon_{\sigma_1},\dots,\varepsilon_{\sigma_p}]$ on the right-hand side of (B2) leads to a compatible scaling for $\mathbb{E}[\delta_{\sigma_1},\dots,\delta_{\sigma_p}]$, the dominant contribution coming from the term $m = 0$ if p is even and from $m = 0$ and $m = 1$ if p is odd. Similarly, one can insert the hypothesis in $\mathbb{E}[\delta_{\sigma_1},\dots,\delta_{\sigma_p}]$ on the right-hand side of (B8); the first term corresponding to $r = 1$ is obviously compatible with the ansatz. The order in κ of the generic term with $r \geqslant 2$ is

$$\left\lceil \frac{|S_1|}{2} \right\rceil + \cdots + \left\lceil \frac{|S_r|}{2} \right\rceil \geqslant \left\lceil \frac{|S_1| + \cdots + |S_r|}{2} \right\rceil \geqslant \left\lceil \frac{p}{2} \right\rceil,$$

(B10)

as the sets $S_1, \dots, S_r$ have to cover $\{1, \dots, p\}$ and hence these terms are not strictly dominant with respect to the case $r = 1$, which concludes our verification of the consistency of the ansatz. From (B2) and (B8) it is relatively easy to obtain the third-order expansion given in Eqs. (71)–(73), exploiting the ansatz to truncate the hierarchy of equations at the desired order in κ and the diagrammatic representation of the terms in Eq. (B8) to organize their bookkeeping.

APPENDIX C: EXPANSION OF THE FREE ENTROPY FOR PAIRWISE INTERACTING POTTS VARIABLES

We provide in this Appendix some details on the derivation of the expansion (77) for the free entropy of pairwise interacting models. Let us start from the expression (40), which we rewrite in more compact notation

$$\phi(P) = \mathbb{E}[z_v \ln z_v] - \tfrac{1}{2}\mathbb{E}[\ell]\mathbb{E}[z_e \ln z_e],$$

(C1)

where the expressions of z_v and z_e are given in Eq. (34) and (39), respectively.

To deal with the second term of ϕ we write $z_e = 1 + \sum_\sigma \delta_\sigma^1 \hat{\delta}_\sigma^2$, where $\delta_\sigma^1 = \eta_\sigma^1 - \overline{\eta}_\sigma$ and $\hat{\delta}_\sigma^2 = \sum_{\sigma'} \hat{M}_{\sigma\sigma'}(\eta_{\sigma'}^2 - \overline{\eta}_{\sigma'})$, with η^1 and η^2 two independent samples from $P(\eta)$. Using the power-series expansion of $(1+x)\ln(1+x)$ around $x = 0$ gives

$$\mathbb{E}[z_e \ln z_e] = \sum_{p=2}^{\infty} \frac{(-1)^p}{p(p-1)} \sum_{\sigma_1,\dots,\sigma_p} \mathbb{E}\big[\delta_{\sigma_1},\dots,\delta_{\sigma_p}\big]\mathbb{E}$$
$$\times \big[\hat{\delta}_{\sigma_1},\dots,\hat{\delta}_{\sigma_p}\big].$$

(C2)

Neglecting the terms with $p \geqslant 5$ yields the first line in Eq. (77).

In the first term we write $z_v = 1 + \sum_\sigma \overline{\eta}_\sigma \varepsilon_\sigma$, where the notation $\varepsilon_\sigma = \prod_{i=1}^{\ell}(1 + \hat{\delta}_\sigma^i) - 1$ is the same as in Appendix B, the only difference being that ℓ is now drawn from the distribution p_ℓ instead of $\tilde{p}_\ell$. This yields

$$\mathbb{E}[z_v \ln z_v] = \sum_{p=2}^{\infty} \frac{(-1)^p}{p(p-1)} \sum_{\sigma_1,\dots,\sigma_p} \overline{\eta}_{\sigma_1} \cdots \overline{\eta}_{\sigma_p} \mathbb{E}\big[\varepsilon_{\sigma_1},\dots,\varepsilon_{\sigma_p}\big],$$

(C3)

where the expression of $\mathbb{E}[\varepsilon_{\sigma_1},\dots,\varepsilon_{\sigma_p}]$ can be read off from (B8) with the modification $\tilde{p}_\ell \to p_\ell$. Enumerating the various terms that contribute up to order κ^4 using the diagrammatic representation explained in Appendix B yields the remaining terms of (77). We factored out in these terms a common factor $\mathbb{E}[\ell]$, using the identity

$$\mathbb{E}[\ell(\ell - 1) \cdots (\ell - r + 1)]$$
$$= \mathbb{E}[\ell]\tilde{\mathbb{E}}[\ell(\ell - 1) \cdots (\ell - r + 2)],$$

(C4)

which follows from the size bias between p_ℓ and $\tilde{p}_\ell$ expressed in Eq. (12).

Finally, let us write down an expression that is equivalent to the expansion (77) but where the centered moments are expressed in the basis of eigenvectors of M,

$$\frac{1}{\mathbb{E}[\ell]}\phi(P) = \frac{1}{4} \sum_{jk}(A'_{jk})^2 \theta_j \theta_k (\tilde{\mathbb{E}}[\ell]\theta_j\theta_k - 1) - \frac{1}{12}\sum_{jkl}(B'_{jkl})^2 \theta_j\theta_k\theta_l(\tilde{\mathbb{E}}[\ell]\theta_j\theta_k\theta_l - 1)$$

$$+ \frac{1}{24}\sum_{jklm}(C'_{jklm})^2 \theta_j\theta_k\theta_l\theta_m(\tilde{\mathbb{E}}[\ell]\theta_j\theta_k\theta_l\theta_m - 1)$$

$$+ \frac{1}{12}\tilde{\mathbb{E}}[\ell(\ell-1)]\left[\sum_{\substack{j_1j_2j_3 \\ k_1k_2k_3}} A'_{j_1k_1}A'_{j_2k_2}A'_{j_3k_3}\theta_{j_1}\theta_{k_1}\theta_{j_2}\theta_{k_2}\theta_{j_3}\theta_{k_3}f_{j_1j_2j_3}f_{k_1k_2k_3} - 2\sum_{j_1j_2j_3}A'_{j_1j_2}A'_{j_2j_3}A'_{j_3j_1}\theta_{j_1}^2\theta_{j_2}^2\theta_{j_3}^2 \right]$$

$$- \frac{1}{2}\tilde{\mathbb{E}}[\ell(\ell-1)]\sum_{\substack{jkl \\ j_1j_2}} B'_{jkl}A'_{j_1k}A'_{j_2l}\theta_j\theta_{j_1}\theta_{j_2}\theta_k^2\theta_l^2 f_{jj_1j_2} + \frac{1}{4}\tilde{\mathbb{E}}[\ell(\ell-1)]\sum_{jklm}C'_{jklm}A'_{jk}A'_{lm}\theta_j^2\theta_k^2\theta_l^2\theta_m^2$$

$$+ \frac{1}{48}\tilde{\mathbb{E}}[\ell(\ell-1)(\ell-2)]\sum_{\substack{j_1j_2j_3j_4 \\ k_1k_2k_3k_4}} A'_{j_1k_1}A'_{j_2k_2}A'_{j_3k_3}A'_{j_4k_4}\theta_{j_1}\theta_{k_1}\theta_{j_2}\theta_{k_2}\theta_{j_3}\theta_{k_3}\theta_{j_4}\theta_{k_4}h_{k_1k_2k_3k_4}^{j_1j_2j_3j_4},$$

(C5)

with the definitions of the tensors that encode the structure of the eigenvectors of M,

$$f_{j_1 j_2 j_3} = \sum_\sigma \overline{\eta}_\sigma r_\sigma^{(j_1)} r_\sigma^{(j_2)} r_\sigma^{(j_3)}, \tag{C6}$$

$$g_{j_1 j_2 j_3 j_4} = \sum_\sigma \overline{\eta}_\sigma r_\sigma^{(j_1)} r_\sigma^{(j_2)} r_\sigma^{(j_3)} r_\sigma^{(j_4)},$$

$$h_{k_1 k_2 k_3 k_4}^{j_1 j_2 j_3 j_4} = g_{j_1 j_2 j_3 j_4} g_{k_1 k_2 k_3 k_4} - 6 g_{j_1 j_2 j_3 j_4} \delta_{k_1 k_2} \delta_{k_3 k_4} - 12 f_{j_1 j_2 j_3} f_{k_1 k_2 j_4} \delta_{k_3 k_4} + 3 \delta_{j_1 j_2} \delta_{k_1 k_2} \delta_{j_3 j_4} \delta_{k_3 k_4} + 12 \delta_{k_4 j_1} \delta_{k_1 j_2} \delta_{k_2 j_3} \delta_{k_3 j_4}. \tag{C7}$$

[1] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, Oxford, 2001).

[2] M. Mézard and A. Montanari, *Physics, Information, Computation* (Oxford University Press, Oxford, 2009).

[3] L. Zdeborová and F. Krzakala, Statistical physics of inference: Thresholds and algorithms, Adv. Phys. **65**, 453 (2016).

[4] A. S. Bandeira, A. Perry, and A. S. Wein, Notes on computational-to-statistical gaps: Predictions using statistical physics, Port. Math. **75**, 159 (2018).

[5] C. Moore, The computer science and physics of community detection: Landscapes, phase transitions, and hardness, arXiv:1702.00467.

[6] E. Abbe, Community detection and stochastic block models: Recent developments, J. Mach. Learn. Res. **18**, 6446 (2017).

[7] S. Fortunato, Community detection in graphs, Phys. Rep. **486**, 75 (2010).

[8] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phase Transition in the Detection of Modules in Sparse Networks, Phys. Rev. Lett. **107**, 065701 (2011).

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

[10] F. Krzakala and L. Zdeborová, Hiding Quiet Solutions in Random Constraint Satisfaction Problems, Phys. Rev. Lett. **102**, 238701 (2009).

[11] L. Zdeborová and F. Krzakala, Quiet planting in the locked constraint satisfaction problems, SIAM J. Discrete Math. **25**, 750 (2011).

[12] E. Abbe and A. Montanari, Conditional random fields, planted constraint satisfaction and entropy concentration, in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, edited by P. Raghavendra, S. Raskhodnikova, K. Jansen, and J. D. P. Rolim, Lecture Notes in Computer Science Vol. 8096 (Springer, Berlin, Heidelberg, 2013), pp. 332–346.

[13] V. Feldman, W. Perkins, and S. Vempala, *Proceedings of the 47th Annual ACM Symposium on Theory of Computing, STOC '15* (ACM, New York, 2015), pp. 77–86.

[14] U. Feige, *Proceedings of the 34th Annual ACM Symposium on Theory of Computing, STOC '02* (ACM, New York, 2002), pp. 534–543.

[15] W. Barthel, A. K. Hartmann, M. Leone, F. Ricci-Tersenghi, M. Weigt, and R. Zecchina, Hiding Solutions in Random Satisfiability Problems: A Statistical Mechanics Approach, Phys. Rev. Lett. **88**, 188701 (2002).

[16] T. Lesieur, F. Krzakala, and L. Zdeborová, Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications, J. Stat. Mech. (2017) 073403.

[17] E. Mossel, Reconstruction on trees: Beating the second eigenvalue, Ann. Appl. Probab. **11**, 285 (2001).

[18] S. Janson and E. Mossel, Robust reconstruction on trees is determined by the second eigenvalue, Ann. Probab. **32**, 2630 (2004).

[19] M. Mézard and A. Montanari, Reconstruction on trees and spin glass transition, J. Stat. Phys. **124**, 1317 (2006).

[20] C. Borgs, J. Chayes, E. Mossel, and S. Roch, *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)* (IEEE, Piscataway, 2006), pp. 518–530.

[21] A. Sly, Reconstruction of random colourings, Commun. Math. Phys. **288**, 943 (2009).

[22] A. Sly, Reconstruction for the Potts model, Ann. Probab. **39**, 1365 (2011).

[23] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Gibbs states and the set of solutions of random constraint satisfaction problems, Proc. Natl. Acad. Sci. USA **104**, 10318 (2007).

[24] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová, Adv. in Math. **333**, 694 (2018).

[25] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang, and T. Kapetanopoulos, Charting the replica symmetric phase, Commun. Math. Phys. **359**, 603 (2018).

[26] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Neural Information Processing Systems, San Diego, 2016), pp. 424–432.

[27] F. Caltagirone, M. Lelarge, and L. Miolane, Recovering asymmetric communities in the stochastic block model, IEEE Trans. Netw. Sci. Eng. **5**, 237 (2017).

[28] M. Lelarge and L. Miolane, Fundamental limits of symmetric low-rank matrix estimation, Probab. Theory Relat. Fields, 1 (2018), doi:10.1007/s00440-018-0845-x.

[29] W. Liu and N. Ning, Large degree asymptotics and the reconstruction threshold of the asymmetric binary channels, J. Stat. Phys. **174**, 1161 (2019).

[30] W. Liu, S. R. Jammalamadaka, and N. Ning, The tightness of the Kesten-Stigum reconstruction bound of symmetric model with multiple mutations, J. Stat. Phys. **170**, 617 (2018).

[31] D. L. Donoho, A. Maleki, and A. Montanari, Message-passing algorithms for compressed sensing, Proc. Natl. Acad. Sci. USA **106**, 18914 (2009).

[32] M. Bayati and A. Montanari, The dynamics of message passing on dense graphs, with applications to compressed sensing, IEEE Trans. Inf. Theory **57**, 764 (2011).

[33] F. R. Kschischang, B. Frey, and H.-A. Loeliger, Factor graphs and the sum-product algorithm, IEEE Trans. Inf. Theory **47**, 498 (2001).

[34] H. Kesten and B. P. Stigum, Additional limit theorems for indecomposable multidimensional Galton-Watson processes, Ann. Math. Stat. **37**, 1463 (1966).

[35] C. Measson, A. Montanari, and R. Urbanke, Maxwell construction: The hidden bridge between iterative and maximum *a posteriori* decoding, IEEE Trans. Inf. Theory **54**, 5277 (2008).

[36] G. Parisi and F. Zamponi, Mean-field theory of hard sphere glasses and jamming, Rev. Mod. Phys. **82**, 789 (2010).

[37] J. R. L. de Almeida and D. J. Thouless, Stability of the Sherrington-Kirkpatrick solution of a spin-glass model, J. Phys. A: Math. Gen. **11**, 983 (1978).

[38] M. Mézard, G. Parisi, and R. Zecchina, Analytic and algorithmic solution of random satisfiability problems, Science **297**, 812 (2002).

[39] S. E. Fienberg and S. S. Wasserman, Categorical data analysis of single sociometric relations, Sociol. Methodol. **12**, 156 (1981).

[40] P. W. Holland, K. B. Laskey, and S. Leinhardt, Stochastic blockmodels: First steps, Soc. Networks **5**, 109 (1983).

[41] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs, Random Struct. Alg. **31**, 3 (2007).

[42] E. Mossel, J. Neeman, and A. Sly, A proof of the block model threshold conjecture, Combinatorica **38**, 665 (2018).

[43] L. Massoulié, *Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (ACM, New York, 2014), pp. 694–703.

[44] E. Mossel, J. Neeman, and A. Sly, Belief propagation, robust reconstruction and optimal recovery of block models, Ann. Appl. Probab. **26**, 2211 (2016).

[45] E. Abbe and C. Sandon, Detection in the stochastic block model with multiple clusters: Proof of the achievability conjectures, acyclic BP, and the information-computation gap, arXiv:1512.09080.

[46] L. Zdeborová and M. Mézard, Constraint satisfaction problems with isolated solutions are hard, J. Stat. Mech. (2008) P12004.

[47] M. C. Angelini, F. Caltagirone, F. Krzakala, and L. Zdeborová, *Proceedings of the 53rd Annual Allerton Conference on Communication, Control, and Computing* (IEEE, Piscataway, 2015), pp. 66–73.

[48] M. Mézard and G. Parisi, The Bethe lattice spin glass revisited, Eur. Phys. J. B **20**, 217 (2001).

[49] E. Mossel, J. Neeman, and A. Sly, Reconstruction and estimation in the planted partition model, Probab. Theory Relat. Fields **162**, 431 (2015).

[50] A. Coja-Oghlan, T. Kapetanopoulos, and N. Müller, The replica symmetric phase of random constraint satisfaction problems, arXiv:1802.09311.

[51] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice, J. Stat. Phys. **79**, 473 (1995).

[52] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman, Broadcasting on trees and the Ising model, Ann. Appl. Probab. **10**, 410 (2000).

[53] F. Krzakala and L. Zdeborová, Potts glass on random graphs, Europhys. Lett. **81**, 57005 (2008).

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

[55] A. Montanari, Finding one community in a sparse graph, J. Stat. Phys. **161**, 273 (2015).

[56] P. Zhang, C. Moore, and M. Newman, Community detection in networks with unequal groups, Phys. Rev. E **93**, 012303 (2016).

[57] W. Liu and N. Ning, Big data information reconstruction on an infinite tree for a 4×4-state asymmetric model with community effects, arXiv:1812.10475.

[58] M. Gabrié, V. Dani, G. Semerjian, and L. Zdeborová, Phase transitions in the q-coloring of random hypergraphs, J. Phys. A: Math. Theor. **50**, 505002 (2017).

[59] A. Montanari, Tight bounds for LDPC and LDGM codes under MAP decoding, IEEE Trans. Inf. Theory **51**, 3221 (2005).

[60] R. Abou-Chacra, D. J. Thouless, and P. W. Anderson, A self-consistent theory of localization, J. Phys. C **6**, 1734 (1973).