Gibbs Measures and Phase Transitions on Sparse Random Graphs

Amir Dembo * **, Andrea Montanari** *

Stanford University.
e-mail: amir@math.stanford.edu*;* montanari@stanford.edu

Abstract: Many problems of interest in computer science and information theory can be phrased in terms of a probability distribution over discrete variables associated to the vertices of a large (but finite) sparse graph. In recent years, considerable progress has been achieved by viewing these distributions as Gibbs measures and applying to their study heuristic tools from statistical physics. We review this approach and provide some results towards a rigorous treatment of these problems.

AMS 2000 subject classifications: Primary 60B10, 60G60, 82B20.
Keywords and phrases: Random graphs, Ising model, Gibbs measures, Phase transitions, Spin models, Local weak convergence..

Contents

1 Introduction 2
1.1 The Curie-Weiss model and some general definitions 4
1.2 Graphical models: examples 13
1.3 Detour: The Ising model on the integer lattice 18
2 Ising models on locally tree-like graphs 21
2.1 Locally tree-like graphs and conditionally independent trees . 22
2.2 Ising models on conditionally independent trees 29
2.3 Algorithmic implications: belief propagation 32
2.4 Free entropy density, from trees to graphs 33
2.5 Coexistence at low temperature 36
3 The Bethe-Peierls approximation 40
3.1 Messages, belief propagation and Bethe equations 41
3.2 The Bethe free entropy 44
3.3 Examples: Bethe equations and free entropy 47
3.4 Extremality, Bethe states and Bethe-Peierls approximation . 52
4 Colorings of random graphs 57
4.1 The phase diagram: a broad picture 58

*Research partially funded by NSF grant #DMS-0806211.

4.2 The COL-UNCOL transition 60
4.3 Coexistence and clustering: the physicist's approach 62
5 Reconstruction and extremality 71
5.1 Applications and related work 73
5.2 Reconstruction on graphs: sphericity and tree-solvability . . . 76
5.3 Proof of main results 79
6 XORSAT and finite-size scaling 85
6.1 XORSAT on random regular graphs 87
6.2 Hyper-loops, hypergraphs, cores and a peeling algorithm . . . 91
6.3 The approximation by a smooth Markov kernel 93
6.4 The ODE method and the critical value 98
6.5 Diffusion approximation and scaling window 101
6.6 Finite size scaling correction to the critical value 102
References 106

1. Introduction

Statistical mechanics is a rich source of fascinating phenomena that can be, at least in principle, fully understood in terms of probability theory. Over the last two decades, probabilists have tackled this challenge with much success. Notable examples include percolation theory [49], interacting particle systems [61], and most recently, conformal invariance. Our focus here is on another area of statistical mechanics, the theory of Gibbs measures, which provides a very effective and flexible way to define collections of 'locally dependent' random variables.

The general abstract theory of Gibbs measures is fully rigorous from a mathematical point of view [42]. However, when it comes to understanding the properties of specific Gibbs measures, i.e. of specific models, a large gap persists between physicists heuristic methods and the scope of mathematically rigorous techniques.

This paper is devoted to somewhat non-standard, family of models, namely Gibbs measures on sparse random graphs. Classically, statistical mechanics has been motivated by the desire to understand the physical behavior of materials, for instance the phase changes of water under temperature change, or the permeation or oil in a porous material. This naturally led to three-dimensional models for such phenomena. The discovery of 'universality' (i.e. the observation that many qualitative features do not depend on the microscopic details of the system), led in turn to the study of models on three-dimensional lattices, whereby the elementary degrees of freedom (spins) are

associated with the vertices of of the lattice. Thereafter, d-dimensional lattices (typically $\mathbb{Z}^d$), became the object of interest upon realizing that significant insight can be gained through such a generalization.

The study of statistical mechanics models 'beyond $\mathbb{Z}^d$' is not directly motivated by physics considerations. Nevertheless, physicists have been interested in models on other graph structures for quite a long time (an early example is [36]). Appropriate graph structures can simplify considerably the treatment of a specific model, and sometimes allow for sharp predictions. Hopefully some qualitative features of this prediction survive on $\mathbb{Z}^d$.

Recently this area has witnessed significant progress and renewed interest as a consequence of motivations coming from computer science, probabilistic combinatorics and statistical inference. In these disciplines, one is often interested in understanding the properties of (optimal) solutions of a large set of combinatorial constraints. As a typical example, consider a linear system over GF[2], $A\,\underline{x} = \underline{b} \bmod 2$, with A an $n \times n$ binary matrix and $\underline{b}$ a binary vector of length n. Assume that A and $\underline{b}$ are drawn from random matrix/vector ensemble. Typical questions are: What is the probability that such a linear system admits a solution? Assuming a typical realization does not admit a solution, what is the maximum number of equations that can, typically, be satisfied?

While probabilistic combinatorics developed a number of ingenious techniques to deal with these questions, significant progress has been achieved recently by employing novel insights from statistical physics (see [65]). Specifically, one first defines a Gibbs measure associated to each instance of the problem at hand, then analyzes its properties using statistical physics techniques, such as the cavity method. While non-rigorous, this approach appears to be very systematic and to provide many sharp predictions.

It is clear at the outset that, for 'natural' distributions of the binary matrix A, the above problem does not have any d-dimensional structure. Similarly, in many interesting examples, one can associate to the Gibbs measure a graph that is sparse and random, but of no finite-dimensional structure. Non-rigorous statistical mechanics techniques appear to provide detailed predictions about general Gibbs measures of this type. It would be highly desirable –and in principle possible– to develop a fully mathematical theory of such Gibbs measures. The present paper provides a unified presentation of a few results in this direction.

In the rest of this section, we proceed with a more detailed overview of the topic, proposing certain fundamental questions the answer to which plays an important role within the non-rigorous statistical mechanics analysis. We illustrate these questions on the relatively well-understood Curie-Weiss (toy)

model and explore a few additional motivating examples.

Section 2 focuses on a specific example, namely the ferromagnetic Ising model on sequences of locally tree-like graphs. Thanks to its monotonicity properties, detailed information can be gained on this model.

A recurring prediction of statistical mechanics studies is that Bethe-Peierls approximation is asymptotically tight in the large graph limit, for sequences of locally tree-like graphs. Section 3 provides a mathematical formalization of Bethe-Peierls approximation. We also prove there that, under an appropriate correlation decay condition, Bethe-Peierls approximation is indeed essentially correct on graphs with large girth.

In Section 4 we consider a more challenging, and as of now, poorly understood, example: proper colorings of a sparse random graph. A fascinating 'clustering' phase transition is predicted to occur as the average degree of the graph crosses a certain threshold. Whereas the detailed description and verification of this phase transition remains an open problem, its relation with the appropriate notion of correlation decay ('extremality'), is the subject of Section 5.

Finally, it is common wisdom in statistical mechanics that phase transitions should be accompanied by a specific 'finite-size scaling' behavior. More precisely, a phase transition corresponds to a sharp change in some property of the model when a control parameter crosses a threshold. In a finite system, the dependence on any control parameter is smooth, and the change and takes place in a window whose width decreases with the system size. Finite-size scaling broadly refers to a description of the system behavior within this window. Section 6 presents a model in which finite-size scaling can be determined in detail.

1.1. The Curie-Weiss model and some general definitions

The Curie-Weiss model is deceivingly simple, but is a good framework to start illustrating some important ideas. For a detailed study of this model we refer to [37].

1.1.1. A story about opinion formation

At time zero, each of n individuals takes one of two opinions $X_i(0) \in \{+1, -1\}$ independently and uniformly at random for $i \in [n] = \{1, \ldots, n\}$. At each subsequent time t, one individual i, chosen uniformly at random,

computes the opinion imbalance

$$M \equiv \sum_{j=1}^{n} X_j \,, \tag{1.1}$$

and $M^{(i)} \equiv M - X_i$. Then, he/she changes his/her opinion with probability

$$p_{\text{flip}}(\underline{X}) = \begin{cases} \exp(-2\beta|M^{(i)}|/n) & \text{if } M^{(i)}X_i > 0 \ , \\ 1 & \text{otherwise.} \end{cases} \tag{1.2}$$

Despite its simplicity, this model raises several interesting questions.

(a). How long does is take for the process $\underline{X}(t)$ to become approximately stationary?
(b). How often do individuals change opinion in the stationary state?
(c). Is the typical opinion pattern strongly polarized (*herding*)?
(d). If this is the case, how often does the popular opinion change?

We do not address question (a) here, but we will address some version of questions (b)–(d). More precisely, this dynamics (first studied in statistical physics under the name of *Glauber* or *Metropolis* dynamics) is an aperiodic irreducible Markov chain whose unique stationary measure is

$$\mu_{n,\beta}(\underline{x}) = \frac{1}{Z_n(\beta)} \exp\Big\{\frac{\beta}{n}\sum_{(i,j)} x_i x_j\Big\}\,. \tag{1.3}$$

To verify this, simply check that the dynamics given by (1.2) is *reversible* with respect to the measure $\mu_{n,\beta}$ of (1.3). Namely, that $\mu_{n,\beta}(\underline{x})\mathbb{P}(\underline{x} \to \underline{x}') = \mu_{n,\beta}(\underline{x}')\mathbb{P}(\underline{x}' \to \underline{x})$ for any two configurations $\underline{x}$, $\underline{x}'$ (where $\mathbb{P}(\underline{x} \to \underline{x}')$ denotes the one-step transition probability from $\underline{x}$ to $\underline{x}'$).

We are mostly interested in the large-n (population size), behavior of $\mu_{n,\beta}(\cdot)$ and its dependence on β (the interaction strength). In this context, we have the following 'static' versions of the preceding questions:

(b'). What is the distribution of $p_{\text{flip}}(\underline{x})$ when $\underline{x}$ has distribution $\mu_{n,\beta}(\,\cdot\,)$?
(c'). What is the distribution of the opinion imbalance M? Is it concentrated near 0 (evenly spread opinions), or far from 0 (herding)?
(d'). In the herding case: how unlikely are balanced ($M \approx 0$) configurations?

1.1.2. Graphical models

A graph $G = (V, E)$ consists of a set V of vertices and a set E of edges (where an edge is an unordered pair of vertices). We always assume G to

be finite with $|V| = n$ and often make the identification $V = [n]$. With $\mathcal{X}$ a finite set, called the *variable domain*, we associate to each vertex $i \in V$ a variable $x_i \in \mathcal{X}$, denoting by $\underline{x} \in \mathcal{X}^V$ the complete assignment of these variables and by $\underline{x}_U = \{x_i : i \in U\}$ its restriction to $U \subseteq V$.

Definition 1.1. *A* bounded specification $\underline{\psi} \equiv \{\psi_{ij} : (i,j) \in E\}$ *for a graph* G *and variable domain* $\mathcal{X}$ *is a family of functionals* $\psi_{ij} : \mathcal{X} \times \mathcal{X} \to [0, \psi_{\max}]$ *indexed by the edges of* G *with* $\psi_{\max}$ *a given finite, positive constant (where for consistency* $\psi_{ij}(x,x') = \psi_{ji}(x',x)$ *for all* $x, x' \in \mathcal{X}$ *and* $(i,j) \in E$*). The specification may include in addition functions* $\psi_i : \mathcal{X} \to [0, \psi_{\max}]$ *indexed by vertices of* G*.*

A bounded specification $\underline{\psi}$ *for* G *is* permissive *if there exists a positive constant* κ *and a 'permitted state'* $x_i^{\mathrm{p}} \in \mathcal{X}$ *for each* $i \in V$*, such that* $\min_{i,x'} \psi_i(x') \geq \kappa\psi_{\max}$ *and*

$$\min_{(i,j)\in E, x'\in\mathcal{X}} \psi_{ij}(x_i^{\mathrm{p}}, x') = \min_{(i,j)\in E, x'\in\mathcal{X}} \psi_{ij}(x', x_j^{\mathrm{p}}) \geq \kappa\psi_{\max} \equiv \psi_{\min}\,.$$

The *graphical model* associated with a graph-specification pair $(G, \underline{\psi})$ is the *canonical probability measure*

$$\mu_{G,\underline{\psi}}(\underline{x}) = \frac{1}{Z(G,\underline{\psi})} \prod_{(i,j)\in E} \psi_{ij}(x_i, x_j) \prod_{i\in V} \psi_i(x_i) \tag{1.4}$$

and the corresponding *canonical stochastic process* is the collection $\underline{X} = \{X_i : i \in V\}$ of $\mathcal{X}$-valued random variables having joint distribution $\mu_{G,\underline{\psi}}(\cdot)$.

One such example is the distribution (1.3), where $\mathcal{X} = \{+1, -1\}$, $\overline{G}$ is the complete graph over n vertices and $\psi_{ij}(x_i, x_j) = \exp(\beta x_i x_j/n)$. Here $\psi_i(x) \equiv 1$. It is sometimes convenient to introduce a 'magnetic field' (see for instance Eq. (1.9) below). This corresponds to taking $\psi_i(x_i) = \exp(Bx_i)$.

Rather than studying graphical models at this level of generality, we focus on a few concepts/tools that have been the subject of recent research efforts.

Coexistence. Roughly speaking, we say that a model $(G, \underline{\psi})$ exhibits coexistence if the corresponding measure $\mu_{G,\underline{\psi}}(\cdot)$ decomposes into a convex combination of well-separated lumps. To formalize this notion, we consider sequences of measures μ_n on graphs $G_n = ([n], E_n)$, and say that coexistence occurs if, for each n, there exists a partition $\Omega_{1,n}, \ldots, \Omega_{r,n}$ of the configuration space $\mathcal{X}^n$ with $r = r(n) \geq 2$, such that

(a). The measure of elements of the partition is uniformly bounded away

from one:

$$\max_{1\le s\le r} \mu_n(\Omega_{s,n}) \le 1-\delta\,. \tag{1.5}$$

(b). The elements of the partition are separated by 'bottlenecks'. That is, for some $\epsilon > 0$,

$$\max_{1\le s\le r} \frac{\mu_n(\partial_\epsilon \Omega_{s,n})}{\mu_n(\Omega_{s,n})} \to 0\,, \tag{1.6}$$

as $n\to\infty$, where $\partial_\epsilon\Omega$ denotes the ϵ-boundary of $\Omega\subseteq\mathcal{X}^n$,

$$\partial_\epsilon\Omega \equiv \{\underline{x}\in\mathcal{X}^n:\, 1\le d(\underline{x},\Omega)\le n\epsilon\}\,, \tag{1.7}$$

with respect to the Hamming [1] distance. The normalization by $\mu_n(\Omega_{s,n})$ removes 'false bottlenecks' and is in particular needed since $r(n)$ often grows (exponentially) with n.
Depending on the circumstances, one may further specify a required rate of decay in (1.6).

We often consider families of models indexed by one (or more) continuous parameters, such as the inverse temperature β in the Curie-Weiss model. A phase transition will generically be a sharp threshold in some property of the measure $\mu(\,\cdot\,)$ as one of these parameters changes. In particular, a phase transition can separate values of the parameter for which coexistence occurs from those values for which it does not.

Mean field models. Intuitively, these are models that lack any (finite-dimensional) geometrical structure. For instance, models of the form (1.4) with ψ_{ij} independent of (i,j) and G the complete graph or a regular random graph are mean field models, whereas models in which G is a finite subset of a finite dimensional lattice are not. To be a bit more precise, the Curie-Weiss model belongs to a particular class of mean field models in which the measure $\mu(\underline{x})$ is exchangeable (that is, invariant under coordinate permutations). A wider class of mean field models may be obtained by considering *random* distributions[2] $\mu(\cdot)$ (for example, when either G or $\underline{\psi}$ are chosen at random in (1.4)). In this context, given a realization of μ, consider k i.i.d.

[1] The Hamming distance $d(\underline{x},\underline{x}')$ between configurations $\underline{x}$ and $\underline{x}'$ is the number of positions in which the two configurations differ. Given $\Omega\subseteq\mathcal{X}^n$, $d(\underline{x},\Omega)\equiv\min\{d(\underline{x},\underline{x}'):\ \underline{x}'\in\Omega\}$.

[2] A random distribution over $\mathcal{X}^n$ is just a random variable taking values on the $(|\mathcal{X}|^n-1)$-dimensional probability simplex.

configurations $\underline{X}^{(1)}, \ldots, \underline{X}^{(k)}$, each having distribution μ. These 'replicas' have the unconditional, joint distribution

$$\mu^{(k)}(\underline{x}^{(1)}, \ldots, \underline{x}^{(k)}) = \mathbb{E}\left\{\mu(\underline{x}^{(1)}) \cdots \mu(\underline{x}^{(k)})\right\} . \tag{1.8}$$

The random distribution μ is a candidate to be a mean field model when for each fixed k the measure $\mu^{(k)}$, viewed as a distribution over $(\mathcal{X}^k)^n$, is exchangeable (with respect to permutations of the coordinate indices in $[n]$). Unfortunately, while this property suffices in many 'natural' special cases, there are models that intuitively are not mean-field and yet have it. For instance, given a non-random measure ν and a uniformly random permutation π, the random distribution $\mu(x_1, \ldots, x_n) \equiv \nu(x_{\pi(1)}, \ldots, x_{\pi(n)})$ meets the preceding requirement yet should not be considered a mean field model. While a satisfactory mathematical definition of the notion of mean field models is lacking, by focusing on selective examples we examine in the sequel the rich array of interesting phenomena that such models exhibit.

Mean field equations. Distinct variables may be correlated in the model (1.4) in very subtle ways. Nevertheless, mean field models are often tractable because an effective 'reduction' to local marginals[3] takes place asymptotically for large sizes (i.e. as $n \to \infty$).

Thanks to this reduction it is often possible to write a closed system of equations for the local marginals that hold in the large size limit and determine the local marginals, up to possibly having finitely many solutions. Finding the 'correct' mathematical definition of this notion is an open problem, so we shall instead provide specific examples of such equations in a few special cases of interest (starting with the Curie-Weiss model).

1.1.3. Coexistence in the Curie-Weiss model

The model (1.3) appeared for the first time in the physics literature as a model for ferromagnets[4]. In this context, the variables x_i are called *spins* and their value represents the direction in which a localized magnetic moment (think of a tiny compass needle) is pointing. In certain materials the different magnetic moments favor pointing in the same direction, and physicists want to know whether such interaction may lead to a macroscopic magnetization (imbalance), or not.

[3] In particular, single variable marginals, or joint distributions of two variables connected by an edge.

[4] A ferromagnet is a material that acquires a macroscopic spontaneous magnetization at low temperature.

In studying this and related problems it often helps to slightly generalize the model by introducing a linear term in the exponent (also called a 'magnetic field'). More precisely, one considers the probability measures

$$\mu_{n,\beta,B}(\underline{x}) = \frac{1}{Z_n(\beta,B)} \exp\Big\{\frac{\beta}{n}\sum_{(i,j)} x_i x_j + B\sum_{i=1}^n x_i\Big\}\,. \tag{1.9}$$

In this context $1/\beta$ is referred to as the 'temperature' and we shall always assume that $\beta \geq 0$ and, without loss of generality, also that $B \geq 0$.

The following estimates on the distribution of the magnetization per site are the key to our understanding of the large size behavior of the Curie-Weiss model (1.9).

Lemma 1.2. *Let $H(x) = -x\log x - (1-x)\log(1-x)$ denote the binary entropy function and for $\beta \geq 0$, $B \in \mathbb{R}$ and $m \in [-1,+1]$ set*

$$\varphi(m) \equiv \varphi_{\beta,B}(m) = Bm + \frac{1}{2}\beta m^2 + H\left(\frac{1+m}{2}\right)\,. \tag{1.10}$$

Then, for $\overline{X} \equiv n^{-1}\sum_{i=1}^n X_i$, a random configuration $(X_1,\ldots,X_n)$ from the Curie-Weiss model and each $m \in S_n \equiv \{-1,-1+2/n,\ldots,1-2/n,1\}$,

$$\frac{e^{-\beta/2}}{n+1}\,\frac{1}{Z_n(\beta,B)}e^{n\varphi(m)} \leq \mathbb{P}\{\overline{X}=m\} \leq \frac{1}{Z_n(\beta,B)}\,e^{n\varphi(m)}\,. \tag{1.11}$$

Proof. Noting that for $M = nm$,

$$\mathbb{P}\{\overline{X}=m\} = \frac{1}{Z_n(\beta,B)}\binom{n}{(n+M)/2}\exp\Big\{BM + \frac{\beta M^2}{2n} - \frac{1}{2}\beta\Big\}\,,$$

our thesis follows by Stirling's approximation of the binomial coefficient (for example, see [24, Theorem 12.1.3]). □

A major role in determining the asymptotic properties of the measures $\mu_{n,\beta,B}$ is played by the *free entropy density* (the term 'density' refers here to the fact that we are dividing by the number of variables),

$$\phi_n(\beta,B) = \frac{1}{n}\log Z_n(\beta,B)\,. \tag{1.12}$$

Lemma 1.3. *For all n large enough we have the following bounds on the free entropy density $\phi_n(\beta,B)$ of the (generalized) Curie-Weiss model*

$$\phi_*(\beta,B) - \frac{\beta}{2n} - \frac{1}{n}\log\{n(n+1)\} \leq \phi_n(\beta,B) \leq \phi_*(\beta,B) + \frac{1}{n}\log(n+1)\,,$$

where

$$\phi_*(\beta, B) \equiv \sup\{\varphi_{\beta,B}(m) : m \in [-1, 1]\} . \tag{1.13}$$

Proof. The upper bound follows upon summing over $m \in S_n$ the upper bound in (1.11). Further, from the lower bound in (1.11) we get that

$$\phi_n(\beta, B) \geq \max\left\{\varphi_{\beta,B}(m) : m \in S_n\right\} - \frac{\beta}{2n} - \frac{1}{n}\log(n+1) .$$

A little calculus shows that maximum of $\varphi_{\beta,B}(\cdot)$ over the finite set S_n is not smaller that its maximum over the interval $[-1, +1]$ minus $n^{-1}(\log n)$, for all n large enough. □

Consider the optimization problem in Eq. (1.13). Since $\varphi_{\beta,B}(\cdot)$ is continuous on $[-1, 1]$ and differentiable in its interior, with $\varphi'_{\beta,B}(m) \to \pm\infty$ as $m \to \mp 1$, this maximum is achieved at one of the points $m \in (-1, 1)$ where $\varphi'_{\beta,B}(m) = 0$. A direct calculation shows that the latter condition is equivalent to

$$m = \tanh(\beta m + B) . \tag{1.14}$$

Analyzing the possible solutions of this equation, one finds out that:

(a). For $\beta \leq 1$, the equation (1.14) admits a unique solution $m_*(\beta, B)$ increasing in B with $m_*(\beta, B) \downarrow 0$ as $B \downarrow 0$. Obviously, $m_*(\beta, B)$ maximizes $\varphi_{\beta,B}(m)$.

(b). For $\beta > 1$ there exists $B_*(\beta) > 0$ continuously increasing in β with $\lim_{\beta\downarrow 1} B_*(\beta) = 0$ such that: (*i*) for $0 \leq B < B_*(\beta)$, Eq. (1.14) admits three distinct solutions $m_-(\beta, B), m_0(\beta, B), m_+(\beta, B) \equiv m_*(\beta, B)$ with $m_- < m_0 \leq 0 \leq m_+ \equiv m_*$; (*ii*) for $B = B_*(\beta)$ the solutions $m_-(\beta, B) = m_0(\beta, B)$ coincide; (*iii*) and for $B > B_*(\beta)$ only the positive solution $m_*(\beta, B)$ survives.
Further, for $B \geq 0$ the global maximum of $\varphi_{\beta,B}(m)$ over $m \in [-1, 1]$ is attained at $m = m_*(\beta, B)$, while $m_0(\beta, B)$ and $m_-(\beta, B)$ are (respectively) a local minimum and a local maximum (and a saddle point when they coincide at $B = B_*(\beta)$). Since $\varphi_{\beta,0}(\cdot)$ is an even function, in particular $m_0(\beta, 0) = 0$ and $m_\pm(\beta, 0) = \pm m_*(\beta, 0)$.

Our next theorem answers question (c') of Section 1.1.1 for the Curie-Weiss model.

Theorem 1.4. *Consider $\overline{X}$ of Lemma 1.2 and the relevant solution $m_*(\beta, B)$ of equation (1.14). If either $\beta \le 1$ or $B > 0$, then for any $\varepsilon > 0$ there exists $C(\varepsilon) > 0$ such that, for all n large enough*

$$\mathbb{P}\left\{\left|\overline{X} - m_*(\beta, B)\right| \le \varepsilon\right\} \ge 1 - e^{-nC(\varepsilon)}\,. \tag{1.15}$$

In contrast, if $B = 0$ and $\beta > 1$, then for any $\varepsilon > 0$ there exists $C(\varepsilon) > 0$ such that, for all n large enough

$$\mathbb{P}\left\{\left|\overline{X} - m_*(\beta, 0)\right| \le \varepsilon\right\} = \mathbb{P}\left\{\left|\overline{X} + m_*(\beta, 0)\right| \le \varepsilon\right\} \ge \frac{1}{2} - e^{-nC(\varepsilon)}\,. \tag{1.16}$$

Proof. Suppose first that either $\beta \le 1$ or $B > 0$, in which case $\varphi_{\beta,B}(m)$ has the unique non-degenerate global maximizer $m_* = m_*(\beta, B)$. Fixing $\varepsilon > 0$ and setting $I_\varepsilon = [-1, m_* - \varepsilon] \cup [m_* + \varepsilon, 1]$, by Lemma 1.2

$$\mathbb{P}\{\overline{X} \in I_\varepsilon\} \le \frac{1}{Z_n(\beta, B)}\,(n+1)\,\exp\left\{n \max[\varphi_{\beta,B}(m) : m \in I_\varepsilon]\right\}.$$

Using Lemma 1.3 we then find that

$$\mathbb{P}\{\overline{X} \in I_\varepsilon\} \le (n+1)^3 e^{\beta/2}\,\exp\left\{n \max[\varphi_{\beta,B}(m) - \phi_*(\beta, B) : m \in I_\varepsilon]\right\},$$

whence the bound of (1.15) follows.

The bound of (1.16) is proved analogously, using the fact that $\mu_{n,\beta,0}(\underline{x}) = \mu_{n,\beta,0}(-\underline{x})$. □

We just encountered our first example of coexistence (and of phase transition).

Theorem 1.5. *The Curie-Weiss model shows coexistence if and only if $B = 0$ and $\beta > 1$.*

Proof. We will limit ourselves to the 'if' part of this statement: for $B = 0$, $\beta > 1$, the Curie-Weiss model shows coexistence. To this end, we simply check that the partition of the configuration space $\{+1, -1\}^n$ to $\Omega_+ \equiv \{\underline{x} : \sum_i x_i \ge 0\}$ and $\Omega_- \equiv \{\underline{x} : \sum_i x_i < 0\}$ satisfies the conditions in Section 1.1.2. Indeed, it follows immediately from (1.16) that choosing a positive $\epsilon < m_*(\beta, 0)/2$, we have

$$\mu_{n,\beta,B}(\Omega_\pm) \ge \frac{1}{2} - e^{-Cn}, \qquad \mu_{n,\beta,B}(\partial_\epsilon \Omega_\pm) \le e^{-Cn}\,,$$

for some $C > 0$ and all n large enough, which is the thesis. □

1.1.4. The Curie-Weiss model: Mean field equations

We have just encountered our first example of coexistence and our first example of phase transition. We further claim that the identity (1.14) can be 'interpreted' as our first example of a *mean field equation* (in line with the discussion of Section 1.1.2). Indeed, assuming throughout this section not to be on the coexistence line $B = 0$, $\beta > 1$, it follows from Theorem 1.4 that $\mathbb{E}\, X_i = \mathbb{E}\, \overline{X} \approx m_*(\beta, B)$.[5] Therefore, the identity (1.14) can be rephrased as

$$\mathbb{E}\, X_i \approx \tanh\Big\{B + \frac{\beta}{n} \sum_{j \in V} \mathbb{E}\, X_j\Big\}\,, \tag{1.17}$$

which, in agreement with our general description of mean field equations, is a closed form relation between the local marginals under the measure $\mu_{n,\beta,B}(\cdot)$.

We next re-derive the equation (1.17) directly out of the concentration in probability of $\overline{X}$. This approach is very useful, for in more complicated models one often has mild bounds on the fluctuations of $\overline{X}$ while lacking fine controls such as in Theorem 1.4. To this end, we start by proving the following 'cavity' estimate.[6]

Lemma 1.6. *Denote by $\mathbb{E}_{n,\beta}$ and $\mathrm{Var}_{n,\beta}$ the expectation and variance with respect to the Curie-Weiss model with n variables at inverse temperature β (and magnetic field B). Then, for $\beta' = \beta(1 + 1/n)$, $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ and any $i \in [n]$,*

$$|\mathbb{E}_{n+1,\beta'} X_i - \mathbb{E}_{n,\beta} X_i| \le \beta \sinh(B + \beta)\sqrt{\mathrm{Var}_{n,\beta}(\overline{X})}\,. \tag{1.18}$$

Proof. By direct computation, for any function $F : \{+1, -1\}^n \to \mathbb{R}$,

$$\mathbb{E}_{n+1,\beta'}\{F(\underline{X})\} = \frac{\mathbb{E}_{n,\beta}\{F(\underline{X})\cosh(B + \beta\overline{X})\}}{\mathbb{E}_{n,\beta}\{\cosh(B + \beta\overline{X})\}}\,.$$

Therefore, with $\cosh(a) \ge 1$ we get by Cauchy-Schwarz that

$$\begin{aligned} &|\mathbb{E}_{n+1,\beta'}\{F(\underline{X})\} - \mathbb{E}_{n,\beta}\{F(\underline{X})\}| \le |\mathrm{Cov}_{n,\beta}\{\mathsf{F}(\underline{X}), \cosh(B + \beta\overline{X})\}| \\ \le\ & ||F||_\infty \sqrt{\mathrm{Var}_{n,\beta}(\cosh(B + \beta\overline{X}))} \le ||F||_\infty \beta \sinh(B + \beta)\sqrt{\mathrm{Var}_{n,\beta}(\overline{X})}\,, \end{aligned}$$

[5] We use $\approx$ to indicate that we do not provide the approximation error, nor plan to rigorously prove that it is small.

[6] Cavity methods of statistical physics aim at understanding thermodynamic limits $n \to \infty$ by first relating certain quantities for systems of size $n \gg 1$ to those in systems of size $n' = n + O(1)$.

where the last inequality is due to the Lipschitz behavior of $x \mapsto \cosh(B+\beta x)$ together with the bound $|\overline{X}| \le 1$. □

The following theorem provides a rigorous version of Eq. (1.17) for $\beta \le 1$ or $B > 0$.

Theorem 1.7. *There exists a constant $C(\beta, B)$ such that for any $i \in [n]$,*

$$\left|\mathbb{E}X_i - \tanh\left\{B + \frac{\beta}{n}\sum_{j\in V}\mathbb{E}\,X_j\right\}\right| \le C(\beta,B)\sqrt{\mathrm{Var}(\overline{X})}\,. \tag{1.19}$$

Proof. In the notations of Lemma 1.6 recall that $\mathbb{E}_{n+1,\beta'}X_i$ is independent of i and so upon fixing $(X_1,\ldots,X_n)$ we get by direct computation that

$$\mathbb{E}_{n+1,\beta'}\{X_i\} = \mathbb{E}_{n+1,\beta'}\{X_{n+1}\} = \frac{\mathbb{E}_{n,\beta}\sinh(B+\beta\overline{X})}{\mathbb{E}_{n,\beta}\cosh(B+\beta\overline{X})}\,.$$

Further notice that (by the Lipschitz property of $\cosh(B+\beta x)$ and $\sinh(B+\beta x)$ together with the bound $|\overline{X}| \le 1$),

$$|\mathbb{E}_{n,\beta}\sinh(B+\beta\overline{X}) - \sinh(B+\beta\mathbb{E}_{n,\beta}\overline{X})| \le \beta\cosh(B+\beta)\sqrt{\mathrm{Var}_{n,\beta}(\overline{X})}\,,$$
$$|\mathbb{E}_{n,\beta}\cosh(B+\beta\overline{X}) - \cosh(B+\beta\mathbb{E}_{n,\beta}\overline{X})| \le \beta\sinh(B+\beta)\sqrt{\mathrm{Var}_{n,\beta}(\overline{X})}\,.$$

Using the inequality $|a_1/b_1 - a_2/b_2| \le |a_1-a_2|/b_1 + a_2|b_1-b_2|/b_1b_2$ we thus have here (with $a_i \ge 0$ and $b_i \ge \max(1,a_i)$), that

$$\left|\mathbb{E}_{n+1,\beta'}\{X_i\} - \tanh\left\{B + \frac{\beta}{n}\sum_{j=1}^{n}\mathbb{E}_{n,\beta}\,X_j\right\}\right| \le C(\beta,B)\sqrt{\mathrm{Var}_{n,\beta}(\overline{X})}\,.$$

At this point you get our thesis by applying Lemma 1.6. □

1.2. Graphical models: examples

We next list a few examples of graphical models, originating at different domains of science and engineering. Several other examples that fit the same framework are discussed in detail in [65].

1.2.1. Statistical physics

Ferromagnetic Ising model. The ferromagnetic Ising model is arguably the most studied model in statistical physics. It is defined by the Boltzmann distribution

$$\mu_{\beta,B}(\underline{x}) = \frac{1}{Z(\beta,B)} \exp\Big\{\beta \sum_{(i,j)\in E} x_i x_j + B\sum_{i\in V} x_i\Big\}\,, \tag{1.20}$$

over $\underline{x} = \{x_i : i \in V\}$, with $x_i \in \{+1,-1\}$, parametrized by the 'magnetic field' $B \in \mathbb{R}$ and 'inverse temperature' $\beta \geq 0$, where the partition function $Z(\beta,B)$ is fixed by the normalization condition $\sum_{\underline{x}} \mu(\underline{x}) = 1$. The interaction between vertices i,j connected by an edge pushes the variables x_i and x_j towards taking the same value. It is expected that this leads to a global alignment of the variables (spins) at low temperature, for a large family of graphs. This transition should be analogue to the one we found for the Curie-Weiss model, but remarkably little is known about Ising models on *general graphs*. In Section 2 we consider the case of random sparse graphs.

Anti-ferromagnetic Ising model. This model takes the same form (1.20), but with $\beta < 0$.[7] Note that if $B = 0$ and the graph is bipartite (i.e. if there exists a partition $V = V_1 \cup V_2$ such that $E \subseteq V_1 \times V_2$), then this model is equivalent to the ferromagnetic one (upon inverting the signs of $\{x_i, i \in V_1\}$). However, on non-bipartite graphs the anti-ferromagnetic model is way more complicated than the ferromagnetic one, and even determining the most likely (lowest energy) configuration is a difficult matter. Indeed, for $B = 0$ the latter is equivalent to the celebrated max-cut problem from theoretical computer science.

Spin glasses. An instance of the Ising spin glass is defined by a graph G, together with edge weights $J_{ij} \in \mathbb{R}$, for $(i,j) \in E$. Again variables are binary $x_i \in \{+1,-1\}$ and

$$\mu_{\beta,B,\underline{J}}(\underline{x}) = \frac{1}{Z(\beta,B,\underline{J})} \exp\Big\{\beta \sum_{(i,j)\in E} J_{ij} x_i x_j + B\sum_{i\in V} x_i\Big\}\,. \tag{1.21}$$

In a spin glass model the 'coupling constants' J_{ij} are random with even distribution (the canonical examples being $J_{ij} \in \{+1,-1\}$ uniformly and J_{ij} centered Gaussian variables). One is interested in determining the asymptotic properties as $n = |V| \to \infty$ of $\mu_{n,\beta,B,\underline{J}}(\,\cdot\,)$ for a typical realization of the coupling $\underline{J} \equiv \{J_{ij}\}$.

[7] In the literature one usually introduces explicitly a minus sign to keep β positive.

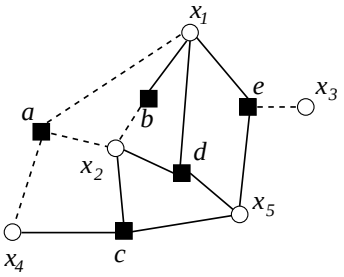

FIG 1. *Factor graph representation of the satisfiability formula* $(\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_4) \wedge (x_1 \vee \bar{x}_2) \wedge (x_2 \vee x_4 \vee x_5) \wedge (x_1 \vee x_2 \vee x_5) \wedge (x_1 \vee \bar{x}_2 \vee x_5)$. *Edges are continuous or dashed depending whether the corresponding variable is directed or negated in the clause.*

1.2.2. Random constraint satisfaction problems

A constraint satisfaction problem (CSP) consists of a finite set $\mathcal{X}$ (called the variable domain), and a class $\mathcal{C}$ of possible constraints (i.e. indicator functions), each of which involves finitely many $\mathcal{X}$-valued variables x_i. An instance of this problem is then specified by a positive integer n (the number of variables), and a set of m constraints involving only the variables $x_1, \ldots, x_n$ (or a subset thereof). A solution of this instance is an assignment in $\mathcal{X}^n$ for the variables $x_1, \ldots, x_n$ which satisfies all m constraints.

In this context, several questions are of interest within computer science:

1. *Decision problem.* Does the given instance have a solution?
2. *Optimization problem.* Maximize the number of satisfied constraints.
3. *Counting problem.* Count the number of solutions.

There are many ways of associating a graphical model to an instance of CSP. If the instance admits a solution, then one option is to consider the uniform measure over all such solutions. Let us see how this works in a few examples.

Coloring. A proper q-coloring of a graph G is an assignment of colors in $[q]$ to the vertices of G such that no edge has both endpoints of the same color. The corresponding CSP has variable domain $\mathcal{X} = [q]$ and the possible constraints in $\mathcal{C}$ are indexed by pairs of indices $(i, j) \in V \times V$, where the constraint (i, j) is satisfied if and only if $x_i \neq x_j$.

Assuming that a graph G admits a proper q-coloring, the uniform measure over the set of possible solutions is

$$\mu_G(\underline{x}) = \frac{1}{Z_G} \prod_{(i,j)\in E} \mathbb{I}(x_i \neq x_j)\,, \tag{1.22}$$

with Z_G counting the number of proper q-colorings of G.

k-SAT. In case of k-satisfiability (in short, k-SAT), the variables are binary $x_i \in \mathcal{X} = \{0,1\}$ and each constraint is of the form $(x_{i(1)}, \ldots, x_{i(k)}) \neq (x^*_{i(1)}, \ldots, x^*_{i(k)})$ for some prescribed k-tuple $(i(1), \ldots, i(k))$ of indices in $V = [n]$ and their prescribed values $(x^*_{i(1)}, \ldots, x^*_{i(k)})$. In this context constraints are often referred to as 'clauses' and can be written as the disjunction (logical OR) of k variables or their negations. The uniform measure over solutions of an instance of this problem, if such solutions exist, is then

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{a=1}^{m} \mathbb{I}\Big((x_{i_a(1)}, \ldots, x_{i_a(k)}) \neq (x^*_{i_a(1)}, \ldots, x^*_{i_a(k)})\Big)\,,$$

with Z counting the number of solutions. An instance can be associated to a *factor graph*, cf. Fig. 1. This is a bipartite graph having two types of nodes: variable nodes in $V = [n]$ denoting the unknowns $x_1, \ldots, x_n$ and function (or factor) nodes in $F = [m]$ denoting the specified constraints. Variable node i and function node a are connected by an edge in the factor graph if and only if variable x_i appears in the a-th clause, so $\partial a = \{i_a(1), \ldots, i_a(k)\}$ and ∂i corresponds to the set of clauses in which i appears.

In general, such a construction associates to arbitrary CSP instance a factor graph $G = (V, F, E)$. The uniform measure over solutions of such an instance is then of the form

$$\mu_{G,\underline{\psi}}(\underline{x}) = \frac{1}{Z(G, \underline{\psi})} \prod_{a \in F} \psi_a(\underline{x}_{\partial a})\,, \tag{1.23}$$

for a suitable choice of $\underline{\psi} \equiv \{\psi_a(\cdot) : a \in F\}$. Such measures can also be viewed as the zero temperature limit of certain Boltzmann distributions. We note in passing that the probability measure of Eq. (1.4) corresponds to the special case where all function nodes are of degree two.

1.2.3. Communications, estimation, detection

We describe next a canonical way of phrasing problems from mathematical engineering in terms of graphical models. Though we do not detail it here, this approach applies to many specific cases of interest.

Let $X_1, \ldots, X_n$ be a collection of i.i.d. 'hidden' random variables with a common distribution $p_0(\,\cdot\,)$ over a finite alphabet $\mathcal{X}$. We want to estimate these variables from a given collection of observations $Y_1, \ldots, Y_m$. The a-th observation (for $a \in [m]$) is a random function of the X_i's for which $i \in \partial a =$

$\{i_a(1), \ldots, i_a(k)\}$. By this we mean that Y_a is conditionally independent of all the other variables given $\{X_i : i \in \partial a\}$ and we write

$$\mathbb{P}\{Y_a \in A | \underline{X}_{\partial a} = \underline{x}_{\partial a}\} = Q_a(A|\underline{x}_{\partial a})\,. \tag{1.24}$$

for some probability kernel $Q_a(\,\cdot\,|\,\cdot\,)$.

The *a posteriori* distribution of the hidden variables given the observations is thus

$$\mu(\underline{x}|\underline{y}) = \frac{1}{Z(\underline{y})} \prod_{a=1}^{m} Q_a(y_a|\underline{x}_{\partial a}) \prod_{i=1}^{n} p_0(x_i)\,. \tag{1.25}$$

1.2.4. Graph and graph ensembles

The structure of the underlying graph G is of much relevance for the general measures $\mu_{G,\underline{\psi}}$ of (1.4). The same applies in the specific examples we have outlined in Section 1.2.

As already hinted, we focus here on (random) graphs that lack finite dimensional Euclidean structure. A few well known ensembles of such graphs (c.f. [54]) are:

I. *Random graphs with a given degree distribution.* Given a probability distribution $\{P_l\}_{l\geq 0}$ over the non-negative integers, for each value of n one draws the graph G_n uniformly at random from the collection of all graphs with n vertices of which precisely $\lfloor nP_k \rfloor$ are of degree $k \geq 1$ (moving one vertex from degree k to $k+1$ if needed for an even sum of degrees). We will denote this ensemble by $\mathbb{G}(P, n)$.

II. The ensemble of *random k-regular graphs* corresponds to $P_k = 1$ (with kn even). Equivalently, this is defined by the set of all graphs G_n over n vertices with degree k, endowed with the uniform measure. With a slight abuse of notation, we will denote it by $\mathbb{G}(k, n)$.

III. *Erdös-Renyi graphs.* This is the ensemble of all graphs G_n with n vertices and $m = \lfloor n\alpha \rfloor$ edges endowed with the uniform measure. A slightly modified ensemble is the one in which each edge (i, j) is present independently with probability $n\alpha/\binom{n}{2}$. We will denote it as $\mathbb{G}(\alpha, n)$.

As further shown in Section 2.1, an important property of these graph ensembles is that they converge locally to trees. Namely, for any integer ℓ, the depth-ℓ neighborhood $\mathsf{B}_i(\ell)$ of a uniformly chosen random vertex i converges in distribution as $n \to \infty$ to a certain random tree of depth (at most) ℓ.

1.3. Detour: The Ising model on the integer lattice

In statistical physics it is most natural to consider models with local interactions on a finite dimensional integer lattice $\mathbb{Z}^d$, where $d = 2$ and $d = 3$ are often the physically relevant ones. While such models are of course non-mean field type, taking a short detour we next present a classical result about ferromagnetic Ising models on finite subsets of $\mathbb{Z}^2$.

Theorem 1.8. *Let $\mathbb{E}_{n,\beta}$ denote expectations with respect to the ferromagnetic Ising measure (1.20) at zero magnetic field, in case $G = (V, E)$ is a square grid of side $\sqrt{n}$. Then, for large n the average magnetization $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ concentrates around zero for high temperature but not for low temperature. More precisely, for some $\beta_o > 0$,*

$$\lim_{\beta\to\infty} \inf_n \mathbb{E}_{n,\beta}\{\,|\overline{X}|\,\} = 1\,, \tag{1.26}$$

$$\lim_{n\to\infty} \mathbb{E}_{n,\beta}\{\,|\overline{X}|^2\,\} = 0 \qquad\qquad \forall \beta < \beta_o\,. \tag{1.27}$$

While this theorem and its proof refer to $\mathbb{Z}^2$, the techniques we use are more general.

Low temperature: Peierls argument. The proof of (1.26) is taken from [47] and based on the Peierls contour representation for the two dimensional Ising model. We start off by reviewing this representation. First, given a square grid $G = (V, E)$ of side $\sqrt{n}$ in $\mathbb{Z}^2$, for each $(i, j) \in E$ draw a perpendicular edge of length one, centered at the midpoint of (i, j). Let E^* denote the collection of all these perpendicular edges and V^* the collection of their end points, viewed as a finite subset of $\mathbb{R}^2$. A *contour* is a simple path on the 'dual' graph $G^* = (V^*, E^*)$, either closed or with both ends at boundary (i.e. degree one) vertices. A closed contour C divides V to two subsets, the *inside* of C and the *outside* of C. We further call as 'inside' the smaller of the two subsets into which a non-closed contour divides V (an arbitrary convention can be used in case the latter two sets are of equal size). A Peierls contours configuration $(\mathcal{C}, s)$ consists of a sign $s \in \{+1, -1\}$ and an edge-disjoint finite collection $\mathcal{C}$ of non-crossing contours (that is, whenever two contours share a vertex, each of them bends there). Starting at an Ising configuration $\underline{x} \in \Omega \equiv \{+1, -1\}^V$ note that the set $V_+(\underline{x}) = \{v \in V : x_v = +1\}$ is separated from $V_-(\underline{x}) = \{v \in V : x_v = -1\}$ by an edge-disjoint finite collection $\mathcal{C} = \mathcal{C}(\underline{x})$ of non-crossing contours. Further, it is not hard to check that the non-empty set $U(\underline{x}) = \{v \in V : v$ not inside any contour from $\mathcal{C}\}$ is either contained in $V_+(\underline{x})$, in which case $s(\underline{x}) = +1$ or in $V_-(\underline{x})$, in which case $s(\underline{x}) = -1$, partitioning Ω to $\Omega_+ = \{\underline{x} : s(\underline{x}) = +1\}$ and

$\Omega_- = \{\underline{x} : s(\underline{x}) = -1\}$. In the reverse direction, the Ising configuration is read off a Peierls contours configuration $(\mathcal{C}, s)$ by setting $x_v = s$ when the number of contours $C \in \mathcal{C}$ such that $v \in V$ lies in the inside of C is even while $x_v = -s$ when it is odd. The mapping $\underline{x} \mapsto -\underline{x}$ exchanges Ω_+ with Ω_- so

$$\mathbb{E}_{n,\beta}[|\overline{X}|] \geq 2\mathbb{E}_{n,\beta}[\overline{X}\mathbb{I}(\underline{X} \in \Omega_+)] = 1 - \frac{4}{n}\mathbb{E}_{n,\beta}[|V_-(\underline{X})|\mathbb{I}(\underline{X} \in \Omega_+)]\,. \quad (1.28)$$

If $\underline{x}$ is in Ω_+ then $|V_-(\underline{x})|$ is bounded by the total number of vertices of V inside contours of $\mathcal{C}$, which by isoperimetric considerations is at most $\sum_{C\in\mathcal{C}} |C|^2$ (where $|C|$ denotes the length of contour C). Further, our one-to-one correspondence between Ising and Peierls contours configurations maps the Ising measure at $\beta > 0$ to uniform $s \in \{+1, -1\}$ independent of $\mathcal{C}$ whose distribution is the Peierls measure

$$\mu_*(\mathcal{C}) = \frac{1}{Z_*(\beta)} \prod_{C\in\mathcal{C}} e^{-2\beta|C|}\,.$$

Recall that if a given contour C is in some edge-disjoint finite collection $\mathcal{C}$ of non-crossing contours, then $\mathcal{C}' = \mathcal{C}\backslash C$ is another such collection, with $\mathcal{C} \mapsto \mathcal{C}'$ injective, from which we easily deduce that $\mu_*(C \in \mathcal{C}) \leq \exp(-2\beta|C|)$ for any fixed contour C. Consequently,

$$\begin{aligned}\mathbb{E}_{n,\beta}[|V_-(\underline{X})|\mathbb{I}(\underline{X} \in \Omega_+)] &\leq \sum_C |C|^2 \mu_*(C \in \mathcal{C}) \\ &\leq \sum_{\ell\geq 2} \ell^2 N_c(n,\ell) e^{-2\beta\ell}\,, \end{aligned} \quad (1.29)$$

where $N_c(n,\ell)$ denotes the number of contours of length ℓ for the square grid of side $\sqrt{n}$. Each such contour is a length ℓ path of a non-reversing nearest neighbor walk in $\mathbb{Z}^2$ starting at some point in V^*. Hence, $N_c(n,\ell) \leq |V^*|3^\ell \leq n3^{\ell+1}$. Combining this bound with (1.28) and (1.29) we conclude that for all n,

$$\mathbb{E}_{n,\beta}[|\overline{X}|] \geq 1 - \frac{4}{n}\sum_{\ell\geq 2} \ell^2 N_c(n,\ell) e^{-2\beta\ell} \geq 1 - 12\sum_{\ell\geq 2} \ell^2 3^\ell e^{-2\beta\ell}\,.$$

We are thus done, as this lower bound converges to one for $\beta \to \infty$.

High-temperature expansion. The proof of (1.27), taken from [39], is by the method of high-temperature expansion which serves us again when dealing with the unfrustrated XORSAT model in Section 6.1. As in the

low-temperature case, the first step consists of finding an appropriate 'geometrical' representation. To this end, given a subset $U \subseteq V$ of vertices, let

$$Z_U(\beta) = \sum_{\underline{x}} x_U \exp\{\beta \sum_{(i,j)\in E} x_i x_j\}$$

and denote by $\mathcal{G}(U)$ the set of subgraphs of G having an odd-degree at each vertex in U and an even degree at all other vertices. Then, with $\theta \equiv \tanh(\beta)$ and $F \subseteq E$ denoting both a subgraph of G and its set of edges, we claim that

$$Z_U(\beta) = 2^{|V|}(\cosh\beta)^{|E|} \sum_{F\in\mathcal{G}(U)} \theta^{|F|}\,. \tag{1.30}$$

Indeed, $e^{\beta y} = \cosh(\beta)[1 + y\theta]$ for $y \in \{+1, -1\}$, so by definition

$$\begin{aligned} Z_U(\beta) &= (\cosh\beta)^{|E|} \sum_{\underline{x}} x_U \prod_{(i,j)\in E} [1 + x_i x_j \theta] \\ &= (\cosh\beta)^{|E|} \sum_{F\subseteq E} \theta^{|F|} \sum_{\underline{x}} x_U \prod_{(i,j)\in F} x_i x_j\,. \end{aligned}$$

By symmetry $\sum_{\underline{x}} x_R$ is zero unless each $v \in V$ appears in the set R an even number of times, in which case the sum is $2^{|V|}$. In particular, the latter applies for $x_R = x_U \prod_{(i,j)\in F} x_i x_j$ if and only if $F \in \mathcal{G}(U)$ from which our stated high-temperature expansion (1.30) follows.

We next use this expansion to get a uniform in n decay of correlations at all $\beta < \beta_o \equiv \operatorname{atanh}(1/3)$, with an exponential rate with respect to the graph distance $d(i,j)$. More precisely, we claim that for any such β, n and $i, j \in V$

$$\mathbb{E}_{n,\beta}\{X_i X_j\} \le (1 - 3\theta)^{-1}(3\theta)^{d(i,j)}\,. \tag{1.31}$$

Indeed, from (1.30) we know that

$$\mathbb{E}_{n,\beta}\{X_i X_j\} = \frac{Z_{(i,j)}(\beta)}{Z_\emptyset(\beta)} = \frac{\sum_{F\in\mathcal{G}(\{i,j\})} \theta^{|F|}}{\sum_{F'\in\mathcal{G}(\emptyset)} \theta^{|F'|}}\,.$$

Let $\mathcal{F}(i,j)$ denote the collection of all simple paths from i to j in $\mathbb{Z}^2$ and for each such path $F_{i,j}$, denote by $\mathcal{G}(\emptyset, F_{i,j})$ the sub-collection of graphs in $\mathcal{G}(\emptyset)$ that have no edge in common with $F_{i,j}$. The sum of vertex degrees in a connected component of a graph F is even, hence any $F \in \mathcal{G}(\{i,j\})$ contains some path $F_{i,j} \in \mathcal{F}(i,j)$. Further, F is the edge-disjoint union of

$F_{i,j}$ and $F' = F \setminus F_{i,j}$ with F' having an even degree at each vertex. As $F' \in \mathcal{G}(\emptyset, F_{i,j})$ we thus deduce that

$$\mathbb{E}_{n,\beta}\{X_i X_j\} \leq \sum_{F_{i,j} \in \mathcal{F}(i,j)} \theta^{|F_{i,j}|} \frac{\sum_{F' \in \mathcal{G}(\emptyset, F_{i,j})} \theta^{|F'|}}{\sum_{F' \in \mathcal{G}(\emptyset)} \theta^{|F'|}} \leq \sum_{F_{i,j} \in \mathcal{F}(i,j)} \theta^{|F_{i,j}|} .$$

The number of paths in $\mathcal{F}(i,j)$ of length ℓ is at most 3^ℓ and their minimal length is $d(i,j)$. Plugging this in the preceding bound establishes our correlation decay bound (1.31).

We are done now, for there are at most $8d$ vertices in $\mathbb{Z}^2$ at distance d from each $i \in \mathbb{Z}^2$. Hence,

$$\begin{aligned} \mathbb{E}_{n,\beta}\{ |\overline{X}|^2 \} &= \frac{1}{n^2} \sum_{i,j \in V} \mathbb{E}_{n,\beta}\{X_i X_j\} \\ &\leq \frac{1}{n^2(1-3\theta)} \sum_{i,j \in V} (3\theta)^{d(i,j)} \leq \frac{1}{n(1-3\theta)} \sum_{d=0}^{\infty} 8d(3\theta)^d , \end{aligned}$$

which for $\theta < 1/3$ decays to zero as $n \to \infty$.

2. Ising models on locally tree-like graphs

A ferromagnetic *Ising model on the finite graph* G (with vertex set V, and edge set E) is defined by the Boltzmann distribution $\mu_{\beta,B}(\underline{x})$ of (1.20) with $\beta \geq 0$. In the following it is understood that, unless specified otherwise, the model is ferromagnetic, and we will call it 'Ising model on G.'

For sequences of graphs $G_n = (V_n, E_n)$ of diverging size n, non-rigorous statistical mechanics techniques, such as the 'replica' and 'cavity methods,' make a number of predictions on this model when the graph G 'lacks any finite-dimensional structure.' The most basic quantity in this context is the asymptotic *free entropy density*, cf. Eq. (1.12),

$$\phi(\beta, B) \equiv \lim_{n \to \infty} \frac{1}{n} \log Z_n(\beta, B) . \tag{2.1}$$

The Curie-Weiss model, cf. Section 1.1, corresponds to the complete graph $G_n = K_n$. Predictions exist for a much wider class of models and graphs, most notably, sparse random graphs with bounded average degree that arise in a number of problems from combinatorics and theoretical computer science (c.f. the examples of Section 1.2.2). An important new feature of sparse graphs is that one can introduce a notion of distance between vertices as the

length of shortest path connecting them. Consequently, phase transitions and coexistence can be studied with respect to the correlation decay properties of the underlying measure. It turns out that this approach is particularly fruitful and allows to characterize these phenomena in terms of appropriate features of Gibbs measures on infinite trees. This direction is pursued in [58] in the case of random constraint satisfaction problems.

Statistical mechanics also provides methods for approximating the local marginals of the Boltzmann measure of (1.20). Of particular interest is the algorithm known in artificial intelligence and computer science under the name of *belief propagation*. Loosely speaking, this procedure consists of solving by iteration certain mean field (cavity) equations. Belief propagation is shown in [29] to converge exponentially fast for an Ising model on any graph (even in a low-temperature regime lacking uniform decorrelation), with resulting asymptotically tight estimates for large locally tree-like graphs (see Section 2.3).

2.1. Locally tree-like graphs and conditionally independent trees

We follow here [29], where the asymptotic free entropy density (2.1) is determined rigorously for certain sparse graph sequences $\{G_n\}$ that converge locally to trees. In order to make this notion more precise, we denote by $\mathsf{B}_i(t)$ the subgraph induced by vertices of G_n whose distance from i is at most t. Further, given two rooted trees T_1 and T_2 of the same size, we write $T_1 \simeq T_2$ if T_1 and T_2 are identical upon labeling their vertices in a breadth first fashion following lexicographic order among siblings.

Definition 2.1. *Let $\mathbb{P}_n$ denote the law of the ball $\mathsf{B}_i(t)$ when $i \in V_n$ is a uniformly chosen random vertex. We say that $\{G_n\}$* converges locally *to the random rooted tree T if, for any finite t and any rooted tree T of depth at most t,*

$$\lim_{n\to\infty} \mathbb{P}_n\{\mathsf{B}_i(t) \simeq T\} = \mathbb{P}\{\mathsf{T}(t) \simeq T\}\,, \tag{2.2}$$

where $\mathsf{T}(t)$ denotes the subtree of first t generations of T.

We also say that $\{G_n\}$ is uniformly sparse *if*

$$\lim_{l\to\infty} \limsup_{n\to\infty} \frac{1}{|V_n|} \sum_{i\in V_n} |\partial i|\, \mathbb{I}(|\partial i| \geq l) = 0\,, \tag{2.3}$$

where $|\partial i|$ denotes the size of the set ∂i of neighbors of $i \in V_n$ (i.e. the degree of i).

The proof that for locally tree-like graphs $\phi_n(\beta, B) = \frac{1}{n} \log Z_n(\beta, B)$ converges to (an explicit) limit $\phi(\beta, B)$ consists of two steps

(a). Reduce the computation of $\phi_n(\beta, B)$ to computing expectations of local (in G_n) quantities with respect to the Boltzmann measure (1.20). This is achieved by noting that the derivative of $\phi_n(\beta, B)$ with respect to β is a sum of such expectations.

(b). Show that under the Boltzmann measure (1.20) on G_n expectations of local quantities are, for t and n large, well approximated by the same expectations with respect to an Ising model on the associated random tree $\mathsf{T}(t)$ (a philosophy related to that of [9]).

The key is of course step (b), and the challenge is to carry it out when the parameter β is large and we no longer have uniqueness of the Gibbs measure on the limiting tree T. Indeed, this is done in [29] for the following collection of trees of conditionally independent (and of bounded average) offspring numbers.

Definition 2.2. *An infinite labeled tree T rooted at the vertex $\emptyset$ is called* conditionally independent *if for each integer $k \geq 0$, conditional on the subtree $\mathsf{T}(k)$ of the first k generations of T, the number of offspring Δ_j for $j \in \partial\mathsf{T}(k)$ are independent of each other, where $\partial\mathsf{T}(k)$ denotes the set of vertices at generation k. We further assume that the (conditional on $\mathsf{T}(k)$) first moments of Δ_j are uniformly bounded by a given non-random finite constant Δ and say that an unlabeled rooted tree T is conditionally independent if $\mathsf{T} \simeq \mathsf{T}'$ for some conditionally independent labeled rooted tree T'.*

As shown in [29, Section 4] (see also Theorem 2.10), on such a tree, local expectations are insensitive to boundary conditions that stochastically dominate the free boundary condition. Our program then follows by monotonicity arguments. An example of the monotonicity properties enjoyed by the Ising model is provided by Lemma 2.12.

We next provide a few examples of well known random graph ensembles that are uniformly sparse and converge locally to conditionally independent trees. To this end, let $P = \{P_k : k \geq 0\}$ be a probability distribution over the non-negative integers, with finite, positive first moment $\overline{P}$, set $\rho_k = (k+1)P_{k+1}/\overline{P}$ and denote its mean as $\overline{\rho}$. We denote by $\mathsf{T}(\rho, t)$ the rooted Galton-Watson tree of $t \geq 0$ generations, i.e. the random tree such that each node has offspring distribution $\{\rho_k\}$, and the offspring numbers at different nodes are independent. Further, $\mathsf{T}(P, \rho, t)$ denotes the modified ensemble where only the offspring distribution at the root is changed to

P. In particular, $\mathsf{T}(P,\rho,\infty)$ is clearly conditionally independent. Other examples of conditionally independent trees include: (a) deterministic trees with bounded degree; (b) percolation clusters on such trees; (c) multi-type branching processes.

When working with random graph ensembles, it is often convenient to work with the *configuration models* [17] defined as follows. In the case of the Erdös-Renyi random graph, one draws m i.i.d. edges by choosing their endpoints i_a, j_a independently and uniformly at random for $a = 1,\ldots,m$. For a graph with given degree distribution $\{P_k\}$, one first partitions the vertex sets into subsets V_0, of $\lfloor nP_0 \rfloor$ vertices, V_1 of $\lfloor nP_1 \rfloor$ vertices, V_2 of $\lfloor nP_2 \rfloor$ vertices, etc. Then associate k half-edges to the vertices in V_k for each k (eventually adding one half edge to the last node, to make their total number even). Finally, recursively match two uniformly random half edges until there is no unmatched one. Whenever we need to make the distinction we denote by $\mathbb{P}_*(\,\cdot\,)$ probabilities under the corresponding configuration model.

The following simple observation transfers results from configuration models to the associated uniform models.

Lemma 2.3. *Let A_n be a sequence of events, such that, under the configuration model*

$$\sum_n \mathbb{P}_*(G_n \notin A_n) < \infty\,. \tag{2.4}$$

Further, assume $m = \lfloor \alpha n \rfloor$ with α fixed (for Erdös-Renyi random graphs), or $\{P_k\}$ fixed, with bounded first moment (for general degree distribution). Then, almost surely under the uniform model, property A_n holds for all n large enough.

Proof. The point is that, the graph chosen under the configuration model is distributed uniformly when further conditional on the property L_n that it has neither self-loops nor double edges (see [54]). Consequently,

$$\mathbb{P}(G_n \notin A_n) = \mathbb{P}_*(G_n \notin A_n | L_n) \le \mathbb{P}_*(G_n \notin A_n)/\mathbb{P}_*(L_n)\,.$$

The thesis follows by recalling that $\mathbb{P}_*(L_n)$ is bounded away from 0 uniformly in n for the models described here (c.f. [54]), and applying the Borel-Cantelli lemma. □

Our next lemma ensures that we only need to check the local (weak) convergence in expectation with respect to the configuration model.

Lemma 2.4. *Given a finite rooted tree T of at most t generations, assume that*

$$\lim_{n\to\infty} \mathbb{P}_*\{\mathsf{B}_i(t) \simeq T\} = \mathbb{Q}_T\,, \tag{2.5}$$

for a uniformly random vertex $i \in G_n$. Then, under both the configuration and the uniform models of Lemma 2.3, $\mathbb{P}_n\{\mathsf{B}_i(t) \simeq T\} \to \mathbb{Q}_T$ almost surely.

Proof. Per given value of n consider the random variable $Z \equiv \mathbb{P}_n\{\mathsf{B}_i(t) \simeq T\}$. In view of Lemma 2.3 and the assumption (2.5) that $\mathbb{E}_*[Z] = \mathbb{P}_*\{\mathsf{B}_i(t) \simeq T\}$ converges to $\mathbb{Q}_T$, it suffices to show that $\mathbb{P}_*\{|Z - \mathbb{E}_*[Z]| \geq \delta\}$ is summable (in n), for any fixed $\delta > 0$. To this end, let r denote the maximal degree of T. The presence of an edge (j,k) in the resulting multi-graph G_n affects the event $\{\mathsf{B}_i(t) \simeq T\}$ only if there exists a path of length at most t in G_n between i and $\{j,k\}$, the maximal degree along which is at most r. Per given choice of (j,k) there are at most $u = u(r,t) \equiv 2\sum_{l=0}^{t} r^l$ such values of $i \in [n]$, hence the Lipschitz norm of Z as a function of the location of the m edges of G_n is bounded by $2u/n$. Let $G_n(t)$ denote the graph formed by the first t edges (so $G_n(m) = G_n$), and introduce the martingale $Z(t) = \mathbb{E}_*[Z|G_n(t)]$, so $Z(m) = Z$ and $Z(0) = \mathbb{E}_*[Z]$. A standard argument (c.f. [10, 81]), shows that the conditional laws $\mathbb{P}_*(\,\cdot\,|G_n(t))$ and $\mathbb{P}_*(\,\cdot\,|G_n(t+1))$ of G_n can be coupled in such a way that the resulting two (conditional) realizations of G_n differ by at most two edges. Consequently, applying Azuma-Hoeffding inequality we deduce that for any T, M and $\delta > 0$, some $c_0 = c_0(\delta, M, u)$ positive and all $m \leq nM$,

$$\mathbb{P}_*(\big|Z - \mathbb{E}_*[Z]\,\big| \geq \delta) = \mathbb{P}_*(\big|Z_m - Z_0\,\big| \geq \delta) \leq 2e^{-c_0 n}\,, \tag{2.6}$$

which is more than enough for completing the proof. □

Proposition 2.5. *Given a distribution $\{P_l\}_{l\geq 0}$ of finite mean, let $\{G_n\}_{n\geq 1}$ be a sequence of graphs whereby G_n is distributed according to the ensemble $\mathbb{G}(P,n)$ with degree distribution P. Then the sequence $\{G_n\}$ is almost surely uniformly sparse and converges locally to $\mathsf{T}(P,\rho,\infty)$.*

Proof. Note that for any random graph G_n of degree distribution P,

$$E_n(l) \equiv \sum_{i\in V_n} |\partial i|\,\mathbb{I}(|\partial i| \geq l) \leq 1 + n\sum_{k\geq l} kP_k \equiv 1 + n\overline{P}_l\,. \tag{2.7}$$

Our assumption that $\overline{P} = \sum_k kP_k$ is finite implies that $\overline{P}_l \to 0$ as $l \to \infty$, so any such sequence of graphs $\{G_n\}$ is uniformly sparse.

As the collection of finite rooted trees of *finite depth* is countable, by Lemma 2.4 we have the almost sure local convergence of $\{G_n\}$ to $\mathsf{T}(P,\rho,\infty)$ once we show that $\mathbb{P}_*(\mathsf{B}_i(t) \simeq T) \to \mathbb{P}(\mathsf{T}(P,\rho,t) \simeq T)$ as $n \to \infty$, where $i \in G_n$ is a uniformly random vertex and T is any fixed finite, rooted tree of at most t generations.

To this end, we opt to describe the distribution of $\mathsf{B}_i(t)$ under the configuration model as follows. First fix a non-random partition of $[n]$ to subsets V_k with $|V_k| = \lfloor nP_k \rfloor$, and assign k half-edges to each vertex in V_k. Then, draw a uniformly random vertex $i \in [n]$. Assume it is in V_k, i.e. has k half-edges. Declare these half-edges 'active'. Recursively sample k unpaired (possibly active) half-edges, and pair the active half-edges to them. Repeat this procedure for the vertices thus connected to i and proceed in a breadth first fashion for t generations (i.e. until all edges of $\mathsf{B}_i(t)$ are determined). Consider now the modified procedure in which, each time an half-edge is selected, the corresponding vertex is put in a separate list, and replaced by a new one with the same number of half-edges, in the graph. Half-edges in the separate list are active, but they are not among the candidates in the sampling part. This modification yields $\mathsf{B}_i(t)$ which is a random tree, specifically, an instance of $\mathsf{T}(\widetilde{P}^{(n)}, \widetilde{\rho}^{(n)}, t)$, where $\widetilde{P}^{(n)}_k = \lfloor n\, P_k \rfloor / \sum_l \lfloor n\, P_l \rfloor$. Clearly, $\mathsf{T}(\widetilde{P}^{(n)}, \widetilde{\rho}^{(n)}, t)$ converges in distribution as $n \to \infty$ to $\mathsf{T}(P,\rho,t)$. The proof is thus complete by providing a coupling in which the probability that either $\mathsf{B}_i(t) \simeq T$ under the modified procedure and $\mathsf{B}_i(t) \not\simeq T$ under the original procedure (i.e. the configurational model), or vice versa, is at most $4|T|^2/n$. Indeed, after ℓ steps, a new vertex j is sampled by the pairing with probability $p_j \propto k_j(\ell)$ in the original procedure and $p'_j \propto k_j(0)$ in the modified one, where $k_j(\ell)$ is the number of free half-edges associated to vertex j at step ℓ. Having to consider at most $|T|$ steps and stopping once the original and modified samples differ, we get the stated coupling upon noting that $||p - p'||_{\mathrm{TV}} \leq 2|T|/n$ (as both samples must then be subsets of the given tree T). □

Proposition 2.6. *Let $\{G_n\}_{n\geq 1}$ be a sequence of Erdös-Renyi random graphs, i.e. of graphs drawn either from the ensemble $\mathbb{G}(\alpha, n)$ or from the uniform model with $m = m(n)$ edges, where $m(n)/n \to \alpha$. Then, the sequence $\{G_n\}$ is almost surely uniformly sparse and converges locally to the Galton-Watson tree $\mathsf{T}(P,\rho,\infty)$ with Poisson(2α) offspring distribution P (in which case $\rho_k = P_k$).*

Proof. We denote by $\mathbb{P}^{\langle m\rangle}(\cdot)$ and $\mathbb{E}^{\langle m\rangle}(\cdot)$ the probabilities and expectations with respect to a random graph G_n chosen uniformly from the ensemble of

all graphs of m edges, with $\mathbb{P}_*^{\langle m\rangle}(\cdot)$ and $\mathbb{E}_*^{\langle m\rangle}(\cdot)$ in use for the corresponding configuration model.

We start by proving the almost sure uniform sparsity for graphs G_n from the uniform ensemble of $m = m(n)$ edges provided $m(n)/n \le M$ for all n and some finite M. To this end, by Lemma 2.3 it suffices to prove this property for the corresponding configuration model. Setting $Z \equiv n^{-1}E_n(l)$ for $E_n(l)$ of (2.7) and $P^{\langle m\rangle}$ to be the Binomial$(2m, 1/n)$ distribution of the degree of each vertex of G_n in this configuration model, note that $\mathbb{E}_*^{\langle m\rangle}[Z] = \overline{P}_l^{\langle m\rangle} \le \overline{P}_l$ for $\overline{P}_l \equiv \sum_{k\ge l} kP_k$ of the Poisson$(4M)$ degree distribution P, any $n \ge 2$ and $m \le nM$. Since $\sum_k kP_k$ is finite, necessarily $\overline{P}_l \to 0$ as $l \to \infty$ and the claimed almost sure uniform sparsity follows from the summability in n, per fixed l and $\delta > 0$ of $\mathbb{P}_*^{\langle m\rangle}\{Z - \mathbb{E}_*^{\langle m\rangle}[Z] \ge \delta\}$, uniformly in $m \le nM$. Recall that the presence of an edge (j, k) in the resulting multi-graph G_n changes the value of $E_n(l)$ by at most $2l$, hence the Lipschitz norm of Z as a function of the location of the m edges of G_n is bounded by $2l/n$. Thus, applying the Azuma-Hoeffding inequality along the lines of the proof of Lemma 2.4 we get here a uniform in $m \le nM$ and summable in n bound of the form of (2.6).

As argued in proving Proposition 2.5, by Lemma 2.4 we further have the claimed almost sure local convergence of graphs from the uniform ensembles of $m = m(n)$ edges, once we verify that (2.5) holds for $\mathbb{P}_*^{\langle m\rangle}(\cdot)$ and $\mathbb{Q}_T = \mathbb{P}\{\mathsf{T}(P, \rho, t) \simeq T\}$ with the Poisson(2α) offspring distribution P. To this end, fix a finite rooted tree T of depth at most t and order its vertices from 1 (for $\emptyset$) to $|T|$ in a breadth first fashion following lexicographic order among siblings. Let Δ_v denote the number of offspring of $v \in T$ with $T(t-1)$ the sub-tree of vertices within distance $t-1$ from the root of T (so $\Delta_v = 0$ for $v \notin T(t-1)$), and denoting by $b \equiv \sum_{v\le T(t-1)} \Delta_v = |T| - 1$ the number of edges of T. Under our equivalence relation between trees there are

$$\prod_{v=1}^{b} \frac{n-v}{\Delta_v!}$$

distinct embeddings of T in $[n]$ for which the root of T is mapped to 1. Fixing such an embedding, the event $\{\mathsf{B}_1(t) \simeq T\}$ specifies the b edges in the restriction of E_n to the vertices of T and further forbids having any edge in E_n between $T(t-1)$ and a vertex outside T. Thus, under the configuration model $\mathbb{P}_*^{\langle m\rangle}(\cdot)$ with m edges chosen with replacement uniformly among the $n_2 \equiv \binom{n}{2}$ possible edges, the event $\{\mathsf{B}_1(t) \simeq T\}$ occurs per such an embedding for precisely $(n_2 - a - b)^{m-b} m!/(m-b)!$ of the n_2^m possible edge selections,

where $a = (n - |T|)|T(t-1)| + \binom{b}{2}$. With $\mathbb{P}_*^{\langle m \rangle}(\mathsf{B}_i(t) \simeq T)$ independent of $i \in [n]$, it follows that

$$\mathbb{P}_*^{\langle m \rangle}(\mathsf{B}_i(t) \simeq T) = \frac{2^b m!}{n^b (m-b)!} \Big(1 - \frac{a+b}{n_2}\Big)^{m-b} \prod_{v=1}^{b} \frac{n-v}{(n-1)\Delta_v!}\,.$$

Since b is independent of n and $a = n|T(t-1)| + O(1)$, it is easy to verify that for $n \to \infty$ and $m/n \to \alpha$ the latter expression converges to

$$\mathbb{Q}_T \equiv (2\alpha)^b e^{-2\alpha|T(t-1)|} \prod_{v=1}^{b} \frac{1}{\Delta_v!} = \prod_{v=1}^{|T(t-1)|} P_{\Delta_v} = \mathbb{P}\{\mathsf{T}(P,\rho,t) \simeq T\}$$

(where $P_k = (2\alpha)^k e^{-2\alpha}/k!$, hence $\rho_k = P_k$ for all k). Further, fixing $\gamma < 1$ and denoting by I_n the interval of width $2n^\gamma$ around αn, it is not hard to check that $\mathbb{P}_*^{\langle m \rangle}(\mathsf{B}_i(t) \simeq T) \to \mathbb{Q}_T$ uniformly over $m \in I_n$.

Let $\mathbb{P}^{(n)}(\cdot)$ and $\mathbb{E}^{(n)}(\cdot)$ denote the corresponding laws and expectations with respect to random graphs G_n from the ensembles $\mathbb{G}(\alpha, n)$, i.e. where each edge is chosen independently with probability $q_n = 2\alpha/(n-1)$. The preceding almost sure local convergence and uniform sparseness extend to these graphs since each law $\mathbb{P}^{(n)}(\cdot)$ is a mixture of the laws $\{\mathbb{P}^{\langle m \rangle}(\cdot), m = 1, 2, \ldots\}$ with mixture coefficients $\mathbb{P}^{(n)}(|E_n| = m)$ that are concentrated on $m \in I_n$. Indeed, by the same argument as in the proof of Lemma 2.3, for any sequence of events A_n,

$$\mathbb{P}^{(n)}(G_n \notin A_n) \le \mathbb{P}^{(n)}(|E_n| \notin I_n) + \eta^{-1} \sup_{m \in I_n} \mathbb{P}_*^{\langle m \rangle}(G_n \notin A_n)\,, \tag{2.8}$$

where

$$\eta = \liminf_{n \to \infty} \inf_{m \in I_n} \mathbb{P}_*^{\langle m \rangle}(L_n)\,,$$

is strictly positive (c.f. [54]). Under $\mathbb{P}^{(n)}(\cdot)$ the random variable $|E_n|$ has the Binomial$(n(n-1)/2, q_n)$ distribution (of mean αn). Hence, upon applying Markov's inequality, we find that for some finite $c_1 = c_1(\alpha)$ and all n,

$$\mathbb{P}^{(n)}(|E_n| \notin I_n) \le n^{-4\gamma}\, \mathbb{E}^{(n)}[(|E_n| - \alpha n)^4] \le c_1 n^{2-4\gamma}\,,$$

so taking $\gamma > 3/4$ guarantees the summability (in n), of $\mathbb{P}^{(n)}(|E_n| \notin I_n)$. For given $\delta > 0$ we already proved the summability in n of $\sup_{m \in I_n} \mathbb{P}_*^{\langle m \rangle}(G_n \notin A_n)$ both for $A_n = \{n^{-1} E_n(l) < \overline{P}_l + \delta\}$ and for $A_n = \{|\mathbb{P}_n(\mathsf{B}_i(t) \simeq T) - \mathbb{Q}_T| < 2\delta\}$. In view of this, considering (2.8) for the former choice of A_n yields the almost sure uniform sparsity of Erdös-Renyi random graphs

from $\mathbb{G}(\alpha, n)$, while the latter choice of A_n yields the almost sure local convergence of these random graphs to the Galton-Watson tree $\mathsf{T}(P, \rho, \infty)$ with Poisson(2α) offspring distribution. □

Remark 2.7. *As a special case of Proposition 2.5, almost every sequence of uniformly random k-regular graphs of n vertices converges locally to the (non-random) rooted k-regular infinite tree $T_k(\infty)$.*

Let $T_k(\ell)$ denote the tree induced by the first ℓ generations of $T_k(\infty)$, i.e. $T_k(0) = \{\emptyset\}$ and for $\ell \geq 1$ the tree $T_k(\ell)$ has k offspring at $\emptyset$ and $(k-1)$ offspring for each vertex at generations 1 to $\ell - 1$. It is easy to check that for any $k \geq 3$, the sequence of finite trees $\{T_k(\ell)\}_{\ell \geq 0}$ *does not converge locally to* $T_k(\infty)$. Instead, it converges to the following random k-canopy tree (c.f. [7] for a closely related definition).

Lemma 2.8. *For any $k \geq 3$, the sequence of finite trees $\{T_k(\ell)\}_{\ell \geq 0}$ converges locally to the* k-canopy tree. *This random infinite tree, denoted CT_k, is formed by the union of the infinite ray $\vec{R} \equiv \{(r, r+1), r \geq 0\}$ and additional finite trees $\{T_{k-1}(r), r \geq 0\}$ such that $T_{k-1}(r)$ is rooted at the r-th vertex along $\vec{R}$. The root of CT_k is on $\vec{R}$ with $\mathbb{P}(\mathsf{CT}_k$ rooted at $r) = (k-2)/(k-1)^{r+1}$ for $r \geq 0$.*

Proof. This local convergence is immediate upon noting that there are exactly $n_r = k(k-1)^{r-1}$ vertices at generation $r \geq 1$ of $T_k(\ell)$, hence $|T_k(\ell)| = [k(k-1)^\ell - 2]/(k-2)$ and $n_{\ell - r}/|T_k(\ell)| \to \mathbb{P}(\mathsf{CT}_k$ rooted at $r)$ as $\ell \to \infty$, for each fixed $r \geq 0$ and $k \geq 3$ (and $\mathsf{B}_i(\ell)$ matches for each i of generation $\ell - r$ in $T_k(\ell)$ the ball $\mathsf{B}_r(\ell)$ of the k-canopy tree). □

Remark 2.9. *Note that the k-canopy tree is not conditionally independent.*

2.2. Ising models on conditionally independent trees

Following [29] it is convenient to extend the model (1.20) by allowing for vertex-dependent magnetic fields B_i, i.e. to consider

$$\mu(\underline{x}) = \frac{1}{Z(\beta, \underline{B})} \exp\Big\{\beta \sum_{(i,j) \in E} x_i x_j + \sum_{i \in V} B_i x_i\Big\}. \tag{2.9}$$

In this general context, it is possible to prove correlation decay results for Ising models on conditionally independent trees. Beyond their independent interest, such results play a crucial role in our analysis of models on sparse graph sequences.

To state these results denote by $\mu^{\ell,0}$ the Ising model (2.9) on $\mathsf{T}(\ell)$ with magnetic fields $\{B_i\}$ (also called free boundary conditions), and by $\mu^{\ell,+}$ the modified Ising model corresponding to the limit $B_i \uparrow +\infty$ for all $i \in \partial\mathsf{T}(\ell)$ (also called plus boundary conditions), using μ^{ℓ} for statements that apply to both free and plus boundary conditions.

Theorem 2.10. *Suppose* T *is a conditionally independent infinite tree of average offspring numbers bounded by* Δ*, as in Definition 2.2. Let* $\langle\,\cdot\,\rangle_i^{(r)}$ *denote the expectation with respect to the Ising distribution on the subtree of* i *and all its descendants in* $\mathsf{T}(r)$ *and* $\langle x;y\rangle \equiv \langle xy\rangle - \langle x\rangle\langle y\rangle$ *denotes the centered two point correlation function. There exist* A *finite and* λ *positive, depending only on* $0 < B_{\min} \le B_{\max}$*,* $\beta_{\max}$ *and* Δ *finite, such that if* $B_i \le B_{\max}$ *for all* $i \in \mathsf{T}(r-1)$ *and* $B_i \ge B_{\min}$ *for all* $i \in \mathsf{T}(\ell)$*, then for any* $r \le \ell$ *and* $\beta \le \beta_{\max}$*,*

$$\mathbb{E}\Big\{ \sum_{i\in\partial\mathsf{T}(r)} \langle x_\emptyset; x_i\rangle_\emptyset^{(\ell)} \Big\} \le A\, e^{-\lambda r}\,. \tag{2.10}$$

If in addition $B_i \le B_{\max}$ *for all* $i \in \mathsf{T}(\ell-1)$ *then for some* $C = C(\beta_{\max}, B_{\max})$ *finite*

$$\mathbb{E}\,||\mu^{\ell,+}_{\mathsf{T}(r)} - \mu^{\ell,0}_{\mathsf{T}(r)}||_{\mathrm{TV}} \le A e^{-\lambda(\ell-r)}\,\mathbb{E}\{C^{|\mathsf{T}(r)|}\}\,. \tag{2.11}$$

The proof of this theorem, given in [29, Section 4], relies on monotonicity properties of the Ising measure, and in particular on the following classical inequality.

Proposition 2.11 (Griffiths inequalities)**.** *Given a finite set* V *and parameters* $\underline{J} = (J_R, R \subseteq V)$ *with* $J_R \ge 0$*, consider the extended ferromagnetic Ising measure*

$$\mu_{\underline{J}}(\underline{x}) = \frac{1}{Z(\underline{J})} \exp\Big\{ \sum_{R\subseteq V} J_R x_R \Big\}\,, \tag{2.12}$$

where $\underline{x} \in \{+1,-1\}^V$ *and* $x_R \equiv \prod_{u\in R} x_u$*. Then, for* $\underline{X}$ *of law* $\mu_{\underline{J}}$ *and any* $A, B \subseteq V$*,*

$$\mathbb{E}_{\underline{J}}[X_A] = \frac{1}{Z(\underline{J})} \sum_{\underline{x}} x_A \exp\Big\{ \sum_{R\subseteq V} J_R x_R \Big\} \ge 0\,, \tag{2.13}$$

$$\frac{\partial}{\partial J_B}\mathbb{E}_{\underline{J}}[X_A] = \mathrm{Cov}_{\underline{J}}(X_A, X_B) \ge 0\,. \tag{2.14}$$

Proof. See [61, Theorem IV.1.21] (and consult [44] for generalizations of this result).

Note that the measure $\mu(\cdot)$ of (2.9) is a special case of $\mu_{\underline{J}}$ (taking $J_{\{i\}} = B_i$, $J_{\{i,j\}} = \beta$ for all $(i,j) \in E$ and $J_R = 0$ for all other subsets of V). Thus, Griffiths inequalities allow us to compare certain marginals of the latter measure for a graph G and non-negative β, B_i with those for other choices of G, β and B_i. To demonstrate this, we state (and prove) the following well known general comparison results.

Lemma 2.12. *Fixing $\beta \geq 0$ and $B_i \geq 0$, for any finite graph $G = (V, E)$ and $A \subseteq V$ let $\langle x_A \rangle_G = \mu(x_A = 1) - \mu(x_A = -1)$ denote the mean of x_A under the corresponding Ising measure on G. Similarly, for $U \subseteq V$ let $\langle x_A \rangle_U^0$ and $\langle x_A \rangle_U^+$ denote the magnetization induced by the Ising measure subject to free (i.e. $x_u = 0$) and plus (i.e. $x_u = +1$) boundary conditions, respectively, at all $u \notin U$. Then, $\langle x_A \rangle_U^0 \leq \langle x_A \rangle_G \leq \langle x_A \rangle_U^+$ for any $A \subseteq U$. Further, $U \mapsto \langle x_A \rangle_U^0$ is monotone non-decreasing and $U \mapsto \langle x_A \rangle_U^+$ is monotone non-increasing, both with respect to set inclusion (among sets U that contain A).*

Proof. From Griffiths inequalities we know that $\underline{J} \mapsto \mathbb{E}_{\underline{J}}[X_A]$ is monotone non-decreasing (where $\underline{J} \geq \underline{\widehat{J}}$ if and only if $J_R \geq \widehat{J}_R$ for all $R \subseteq V$). Further, $\langle x_A \rangle_G = \mathbb{E}_{\underline{J}^0}[X_A]$ where $J^0_{\{i\}} = B_i$, $J^0_{\{i,j\}} = \beta$ when $(i,j) \in E$ and all other values of $\underline{J}^0$ are zero. Considering

$$J_R^{\eta,U} = J_R^0 + \eta \mathbb{I}(R \subseteq U^c, |R| = 1)\,,$$

with $\eta \mapsto \underline{J}^{\eta,U}$ non-decreasing, so is $\eta \mapsto \mathbb{E}_{\underline{J}^{\eta,U}}[X_A]$. In addition, $\mu_{\underline{J}^{\eta,U}}(x_u = -1) \leq C\, e^{-2\eta}$ whenever $u \notin U$. Hence, as $\eta \uparrow \infty$ the measure $\mu_{\underline{J}^{\eta,U}}$ converges to $\mu_{\underline{J}}$ subject to plus boundary conditions $x_u = +1$ for $u \notin U$. Consequently,

$$\langle x_A \rangle_G \leq \mathbb{E}_{\underline{J}^{\eta,U}}[X_A] \uparrow \langle x_A \rangle_U^+\,.$$

Similarly, let $J_R^U = J_R^0 \mathbb{I}(R \subseteq U)$ noting that under $\mu_{\underline{J}^U}$ the random vector $\underline{x}_U$ is distributed according to the Ising measure μ restricted to G_U (alternatively, having free boundary conditions $x_u = 0$ for $u \notin U$). With $A \subseteq U$ we thus deduce that

$$\langle x_A \rangle_U^0 = \mathbb{E}_{\underline{J}^U}[X_A] \leq \mathbb{E}_{\underline{J}^0}[X_A] = \langle x_A \rangle_G\,.$$

Finally, the stated monotonicity of $U \mapsto \langle x_A \rangle_U^0$ and $U \mapsto \langle x_A \rangle_U^+$ are in view of Griffiths inequalities the direct consequence of the monotonicity (with respect to set inclusions) of $U \mapsto \underline{J}^U$ and $U \mapsto \underline{J}^{\eta,U}$, respectively. □

In addition to Griffiths inequalities, the proof of Theorem 2.10 uses also the GHS inequality [48] which regards the effect of a magnetic field $\underline{B}$ on the local magnetizations at various vertices. It further uses an extension of Simon's inequality (about the centered two point correlation functions in ferromagnetic Ising models with zero magnetic field, see [82, Theorem 2.1]), to arbitrary magnetic field, in the case of Ising models on trees. Namely, [29, Lemma 4.3] states that if edge (i,j) is on the unique path from $\emptyset$ to $k \in \mathsf{T}(\ell)$, with j a descendant of $i \in \partial\mathsf{T}(t)$, $t \geq 0$, then

$$\langle x_\emptyset; x_k \rangle_\emptyset^{(\ell)} \leq \cosh^2(2\beta + B_i)\, \langle x_\emptyset; x_i \rangle_\emptyset^{(t)} \langle x_j; x_k \rangle_j^{(\ell)} \,. \tag{2.15}$$

2.3. Algorithmic implications: belief propagation

The 'belief propagation' (BP) algorithm consists of solving by iterations a collection of Bethe-Peierls (or cavity) mean field equations. More precisely, for the Ising model (1.20) we associate to each directed edge in the graph $i \to j$, with $(i,j) \in G$, a distribution (or 'message') $\nu_{i\to j}(x_i)$ over $x_i \in \{+1,-1\}$, using then the following update rule

$$\nu_{i\to j}^{(t+1)}(x_i) = \frac{1}{z_{i\to j}^{(t)}}\, e^{Bx_i} \prod_{l\in\partial i\setminus j} \sum_{x_l} e^{\beta x_i x_l} \nu_{l\to i}^{(t)}(x_l) \tag{2.16}$$

starting at a *positive* initial condition, namely where $\nu_{i\to j}^{(0)}(+1) \geq \nu_{i\to j}^{(0)}(-1)$ at each directed edge.

Applying Theorem 2.10 we establish in [29, Section 5] the uniform exponential convergence of the BP iteration to the same fixed point of (2.16), irrespective of its positive initial condition. As we further show there, for tree-like graphs the limit of the BP iteration accurately approximates local marginals of the Boltzmann measure (1.20).

Theorem 2.13. *Assume $\beta \geq 0$, $B > 0$ and G is a graph of finite maximal degree Δ. Then, there exists $A = A(\beta,B,\Delta)$ and $c = c(\beta,B,\Delta)$ finite, $\lambda = \lambda(\beta,B,\Delta) > 0$ and a fixed point $\{\nu_{i\to j}^*\}$ of the BP iteration (2.16) such that for any positive initial condition $\{\nu_{l\to k}^{(0)}\}$ and all $t \geq 0$,*

$$\sup_{(i,j)\in E} \|\nu_{i\to j}^{(t)} - \nu_{i\to j}^*\|_{\mathrm{TV}} \leq A \exp(-\lambda t)\,. \tag{2.17}$$

Further, for any $i_o \in V$, if $\mathsf{B}_{i_o}(t)$ is a tree then for $U \equiv \mathsf{B}_{i_o}(r)$

$$||\mu_U - \nu_U||_{\mathrm{TV}} \leq \exp\left\{c^{r+1} - \lambda(t-r)\right\}, \tag{2.18}$$

where $\mu_U(\cdot)$ is the law of $\underline{x}_U \equiv \{x_i : i \in U\}$ under the Ising model (1.20) and ν_U the probability distribution

$$\nu_U(\underline{x}_U) = \frac{1}{z_U} \exp\Big\{\beta \sum_{(i,j)\in E_U} x_i x_j + B \sum_{i\in U\setminus\partial U} x_i\Big\} \prod_{i\in\partial U} \nu^*_{i\to j(i)}(x_i)\,, \quad (2.19)$$

with E_U the edge set of U whose border is ∂U (i.e. the set of its vertices at distance r from i_o), and $j(i)$ is any fixed neighbor in U of i.

2.4. Free entropy density, from trees to graphs

Bethe-Peierls approximation (we refer to Section 3.1 for a general introduction), allows us to predict the asymptotic free entropy density for sequences of graphs that converge locally to conditionally independent trees. We start by explaining this prediction in a general setting, then state a rigorous result which verifies it for a specific family of graph sequences.

To be definite, assume that $B > 0$. Given a graph sequence $\{G_n\}$ that converges to a conditionally independent tree T with bounded average offspring number, let $L = \Delta_\emptyset$ be the degree of its root. Define the 'cavity fields' $\{h_1, \dots, h_L\}$ by letting $h_j = \lim_{t\to\infty} h_j^{(t)}$ with $h_j^{(t)} \equiv \operatorname{atanh}[\langle x_j\rangle_j^{(t)}]$, where $\langle\,\cdot\,\rangle_j^{(t)}$ denotes expectation with respect to the Ising distribution on the subtree induced by $j \in \partial\emptyset$ and all its descendants in $\mathsf{T}(t)$ (with free boundary conditions). We note in passing that $t \mapsto h_j^{(t)}$ is stochastically monotone (and hence has a limit in law) by Lemma 2.12. Further $\{h_1, \dots, h_L\}$ are conditionally independent given L. Finally, define $\theta = \tanh(\beta)$ and

$$h_{-j} = B + \sum_{k=1,k\neq j}^{L} \operatorname{atanh}[\theta \tanh(h_k)]\,. \quad (2.20)$$

The Bethe-Peierls free energy density is given by

$$\varphi(\beta, B) \equiv \frac{1}{2}\mathbb{E}\{L\}\gamma(\theta) - \frac{1}{2}\,\mathbb{E}\Big\{\sum_{j=1}^{L} \log[1 + \theta\tanh(h_{-j})\tanh(h_j)]\Big\}$$
$$+ \mathbb{E}\log\big\{e^B \prod_{j=1}^{L}[1 + \theta\tanh(h_j)] + e^{-B}\prod_{j=1}^{L}[1 - \theta\tanh(h_j)]\big\}\,, \quad (2.21)$$

for $\gamma(u) = -\frac{1}{2}\log(1-u^2)$. We refer to Section 3.3 where this formula is obtained as a special case of the general expression for a Bethe-Peierls free

energy. The prediction is extended to $B < 0$ by letting $\varphi(\beta, B) = \varphi(\beta, -B)$, and to $B = 0$ by letting $\varphi(\beta, 0)$ be the limit of $\varphi(\beta, B)$ as $B \to 0$.

As shown in [29, Lemma 2.2], when $\mathsf{T} = \mathsf{T}(P, \rho, \infty)$ is a Galton-Watson tree, the random variables $\{h_j\}$ have a more explicit characterization in terms of the following fixed point distribution.

Lemma 2.14. *In case $\mathsf{T} = \mathsf{T}(P, \rho, \infty)$ consider the random variables $\{h^{(t)}\}$ where $h^{(0)} \equiv 0$ and for $t \geq 0$,*

$$h^{(t+1)} \stackrel{\mathrm{d}}{=} B + \sum_{i=1}^{K} \operatorname{atanh}[\theta \tanh(h_i^{(t)})]\,, \tag{2.22}$$

with $h_i^{(t)}$ i.i.d. copies of $h^{(t)}$ that are independent of the variable K of distribution ρ. If $B > 0$ and $\overline{\rho} < \infty$ then $t \mapsto h^{(t)}$ is stochastically monotone (i.e. there exists a coupling under which $\mathbb{P}(h^{(t)} \leq h^{(t+1)}) = 1$ for all t), and converges in law to the unique fixed point h^* of (2.22) that is supported on $[0, \infty)$. In this case, h_j of (2.21) are i.i.d. copies of h^* that are independent of L.

The main result of [29] confirms the statistical physics prediction for the free entropy density.

Theorem 2.15. *If $\overline{\rho}$ is finite then for any $B \in \mathbb{R}$, $\beta \geq 0$ and sequence $\{G_n\}_{n \in \mathbb{N}}$ of uniformly sparse graphs that converges locally to $\mathsf{T}(P, \rho, \infty)$,*

$$\lim_{n \to \infty} \frac{1}{n} \log Z_n(\beta, B) = \varphi(\beta, B)\,. \tag{2.23}$$

We proceed to sketch the outline of the proof of Theorem 2.15. For uniformly sparse graphs that converge locally to $\mathsf{T}(P, \rho, \infty)$ the model (1.20) has a line of first order phase transitions for $B = 0$ and $\beta > \beta_{\mathrm{c}}$ (that is, where the continuous function $B \mapsto \varphi(\beta, B)$ exhibits a discontinuous derivative). Thus, the main idea is to utilize the magnetic field B to explicitly break the $+/-$ symmetry, and to carefully exploit the monotonicity properties of the ferromagnetic Ising model in order to establish the result even at $\beta > \beta_{\mathrm{c}}$.

Indeed, since $\phi_n(\beta, B) \equiv \frac{1}{n} \log Z_n(\beta, B)$ is invariant under $B \to -B$ and is uniformly (in n) Lipschitz continuous in B with Lipschitz constant one, for proving the theorem it suffices to fix $B > 0$ and show that $\phi_n(\beta, B)$ converges as $n \to \infty$ to the predicted expression $\varphi(\beta, B)$ of (2.21). This is obviously true for $\beta = 0$ since $\phi_n(0, B) = \log(2 \cosh B) = \varphi(0, B)$. Next, denoting by $\langle \cdot \rangle_n$ the expectation with respect to the Ising measure on G_n

(at parameters β and B), it is easy to see that

$$\partial_\beta \phi_n(\beta, B) = \frac{1}{n} \sum_{(i,j)\in E_n} \langle x_i x_j \rangle_n = \frac{1}{2} \mathbb{E}_n \Big[\sum_{j \in \partial i} \langle x_i x_j \rangle_n \Big] . \tag{2.24}$$

With $|\partial_\beta \phi_n(\beta, B)| \le |E_n|/n$ bounded by the assumed uniform sparsity, it is thus enough to show that the expression in (2.24) converges to the partial derivative of $\varphi(\beta, B)$ with respect to β. Turning to compute the latter derivative, after a bit of real analysis we find that the dependence of $\{h_j, h_{-j}\}$ on β can be ignored (c.f. [29, Corollary 6.3] for the proof of this fact in case $\mathsf{T} = \mathsf{T}(P, \rho, \infty)$). That is, hereafter we simply compute the partial derivative in β of the expression (2.21) while considering the law of $\{h_j\}$ and $\{h_{-j}\}$ to be independent of β. To this end, setting $z_j = \tanh(h_j)$ and $y_j = \tanh(h_{-j})$, the relation (2.20) amounts to

$$y_j = \frac{e^B \prod_{k\neq j}(1+\theta z_k) - e^{-B} \prod_{k\neq j}(1-\theta z_k)}{e^B \prod_{k\neq j}(1+\theta z_k) + e^{-B} \prod_{k\neq j}(1-\theta z_k)}$$

for which it follows that

$$\begin{aligned} \frac{\partial}{\partial\theta} \Big\{ \sum_{j=1}^{l} \log(1+\theta z_j y_j) \Big\} &= \frac{\partial}{\partial\theta} \log \Big\{ e^B \prod_{j=1}^{l} (1+\theta z_j) + e^{-B} \prod_{j=1}^{l} (1-\theta z_j) \Big\} \\ &= \sum_{j=1}^{l} \frac{z_j y_j}{1+\theta z_j y_j} \end{aligned}$$

and hence a direct computation of the derivative in (2.21) leads to

$$\partial_\beta \, \varphi(\beta, B) = \frac{1}{2} \mathbb{E} \Big[\sum_{j \in \partial \emptyset} \langle x_\emptyset x_j \rangle_{\mathsf{T}} \Big] , \tag{2.25}$$

where $\langle \cdot \rangle_{\mathsf{T}}$ denotes the expectation with respect to the Ising model

$$\mu_{\mathsf{T}}(x_\emptyset, x_1, \ldots, x_L) = \frac{1}{z} \exp \Big\{ \beta \sum_{j=1}^{L} x_\emptyset x_j + B x_\emptyset + \sum_{j=1}^{L} h_j x_j \Big\} , \tag{2.26}$$

on the 'star' $\mathsf{T}(1)$ rooted at $\emptyset$ and the random cavity fields h_j of (2.21).

In comparison, fixing a positive integer t and considering Lemma 2.12 for $A \equiv \{i, j\}$ and $U \equiv \mathsf{B}_i(t)$, we find that the correlation $\langle x_i x_j \rangle_n$ lies between the correlations $\langle x_i x_j \rangle^0_{\mathsf{B}_i(t)}$ and $\langle x_i x_j \rangle^+_{\mathsf{B}_i(t)}$ for the Ising model on

the subgraph $\mathsf{B}_i(t)$ with free and plus, respectively, boundary conditions at $\partial\mathsf{B}_i(t)$. Thus, in view of (2.24)

$$\frac{1}{2}\mathbb{E}_n\{F_0(\mathsf{B}_i(t))\} \le \partial_\beta \phi_n(\beta, B) \le \frac{1}{2}\mathbb{E}_n\{F_+(\mathsf{B}_i(t))\}\,,$$

where $F_{0/+}(\mathsf{B}_i(t)) \equiv \sum_{j\in\partial i}\langle x_i x_j\rangle^{0/+}_{\mathsf{B}_i(t)}$.

Next, taking $n\to\infty$ we rely on the following consequence of the local convergence of a uniformly sparse graph sequence $\{G_n\}$ (c.f. [29, Lemma 6.4] for the derivation of a similar result).

Lemma 2.16. *Suppose a uniformly sparse graph sequence $\{G_n\}$ converges locally to the random tree T. Fix an integer $t\ge 0$ and a function $F(\cdot)$ on the collection of all possible subgraphs that may occur as $\mathsf{B}_i(t)$, such that $F(\mathsf{B}_i(t))/(|\partial i|+1)$ is uniformly bounded and $F(T_1)=F(T_2)$ whenever $T_1\simeq T_2$. Then,*

$$\lim_{n\to\infty}\mathbb{E}_n\{F(\mathsf{B}_i(t))\} = \mathbb{E}\{F(\mathsf{T}(t))\}\,. \tag{2.27}$$

Indeed, applying this lemma for the functions $F_0(\cdot)$ and $F_+(\cdot)$ we find that

$$\frac{1}{2}\mathbb{E}\{F_0(\mathsf{T}(t))\} \le \liminf_{n\to\infty}\partial_\beta\phi_n(\beta,B) \le \limsup_{n\to\infty}\partial_\beta\phi_n(\beta,B) \le \frac{1}{2}\mathbb{E}\{F_+(\mathsf{T}(t))\}\,.$$

To compute $F_{0/+}(\mathsf{T}(t))$ we first sum over the values of x_k for $k\in\mathsf{T}(t)\setminus\mathsf{T}(1)$. This has the effect of reducing $F_{0/+}(\mathsf{T}(t))$ to the form of $\sum_{j\in\partial\emptyset}\langle x_\emptyset x_j\rangle_{\mathsf{T}}$ and the cavity fields are taken as $h_j^{(t),0/+}\equiv \operatorname{atanh}[\langle x_j\rangle_j^{(t),0/+}]$. Further, from (2.11) we deduce that as $t\to\infty$ both sets of cavity fields converge in law to the same limit $\{h_j\}$. Since $\mathbb{E}[\langle x_\emptyset x_j\rangle_{\mathsf{T}}]$ are continuous with respect to such convergence in law, we get by (2.25) that

$$\lim_{t\to\infty}\frac{1}{2}\mathbb{E}\{F_{0/+}(\mathsf{T}(t))\} = \partial_\beta\varphi(\beta,B)\,,$$

which completes the proof of the theorem.

2.5. Coexistence at low temperature

We focus here on the ferromagnetic Ising model on a random k-regular graph with $k\ge 3$ and zero magnetic field. In order to simplify derivations, it is convenient to use the so-called *configuration model* for random regular graphs [17]. A graph from this ensemble is generated by associating k half

edges to each $i \in [n]$ (with kn even) and pairing them uniformly at random. In other words, the collection E_n of $m \equiv kn/2$ edges is obtained by pairing the kn half-edges. Notice that the resulting object is in fact a *multi-graph* i.e. it might include double edges and self-loops. However the number of such 'defects' is $O(1)$ as $n \to \infty$ and hence the resulting random graph model shares many properties with random k-regular graphs. With a slight abuse of notation, we keep denoting by $\mathbb{G}(k,n)$ the multi-graph ensemble.

For $\beta \geq 0$ we consider the distribution

$$\mu_{n,\beta,k}(\underline{x}) = \frac{1}{Z(G_n)} \exp\Big\{\beta \sum_{(i,j)\in E_n} x_i x_j\Big\}\,. \tag{2.28}$$

Recall Remark 2.7 that any sequence of random graphs G_n from the ensembles $\mathbb{G}(k,n)$ is almost surely uniformly sparse and converges locally to the infinite k-regular tree. Thus, considering the function

$$\begin{aligned}\varphi_k(\beta,h) \equiv & \frac{k}{2}\{\gamma(\theta) - \log[1+\theta\tanh^2(h)]\} \\ & + \log\left\{[1+\theta\tanh(h)]^k + [1-\theta\tanh(h)]^k\right\},\end{aligned}$$

of $h \in \mathbb{R}$ and $\theta = \tanh(\beta)$, we have from Theorem 2.15 that

$$\lim_{n\to\infty} \frac{1}{n} \log Z(G_n) = \varphi_k(\beta, h^*)\,, \tag{2.29}$$

where the cavity field h^* is the largest solution of

$$g(h) \equiv (k-1)\mathrm{atanh}[\theta\tanh(h)] - h = 0\,. \tag{2.30}$$

Indeed, the expression for $\varphi_k(\beta,h^*)$ is taken from (2.21), noting that here $L = K+1 = k$ is non-random, hence so are $h_{-j} = h_j = h^*$. It is not hard to check by calculus that the limit as $B \downarrow 0$ of the unique positive solution of $g(h) = -B$ is strictly positive if and only if $\beta > \beta_{\rm c} \equiv \mathrm{atanh}(1/(k-1))$, in which case $g(-h) = -g(h)$ is zero if and only if $h \in \{0, \pm h^*\}$ with $g'(0) > 0$ and $g'(h^*) < 0$ (c.f. [64]).

We expect coexistence in this model if and only if $\beta > \beta_{\rm c}$ (where we have a line of first order phase transitions for the asymptotic free entropy at $B = 0$), and shall next prove the 'if' part.

Theorem 2.17. *With probability one, the ferromagnetic Ising measures $\mu_{n,\beta,k}$ on uniformly random k-regular multi-graphs from the ensemble $\mathbb{G}(k,n)$ exhibit coexistence if $(k-1)\tanh(\beta) > 1$.*

Proof. As in the proof of Theorem 1.5 (for the Curie-Weiss model), we again consider the partition of $\mathcal{X}^n$ to $\Omega_+ \equiv \{\underline{x} : \sum_i x_i \geq 0\}$ and $\Omega_- \equiv \{\underline{x} : \sum_i x_i < 0\}$. From the invariance of $\mu_{n,\beta,k}$ with respect to the sign change $\underline{x} \mapsto -\underline{x}$ it follows that $\mu_{n,\beta,k}(\Omega_+) = \mu_{n,\beta,k}(\Omega_0) + \mu_{n,\beta,k}(\Omega_-)$ where $\Omega_r \equiv \{\underline{x} : \sum_i x_i = r\}$. Hence, to prove coexistence it suffices to show that for $\epsilon > 0$ small enough, with probability one

$$\limsup_{n\to\infty} \frac{1}{n} \log \Big\{ \sum_{|r|\leq n\epsilon} \mu_{n,\beta,k}(\Omega_r) \Big\} < 0 \,. \tag{2.31}$$

To this end, note that $\mu_{n,\beta,k}(\Omega_r) = Z_r(G_n)/Z(G_n)$ for the *restricted partition function*

$$Z_r(G_n) \equiv \sum_{\underline{x}\in\Omega_r} \exp\Big\{\beta \sum_{(i,j)\in E_n} x_i x_j\Big\} \,. \tag{2.32}$$

Further, recall that by Markov's inequality and the Borel-Cantelli lemma, for any positive random variables Y_n, with probability one

$$\limsup_{n\to\infty} \frac{1}{n} \log Y_n \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}(Y_n) \,.$$

Thus, combining (2.29) with the latter inequality for $Y_n = \sum_{|r|\leq n\epsilon} Z_r(G_n)$ we arrive at the inequality (2.31) upon proving the following lemma (c.f. [43, Section 5]).

Lemma 2.18. *Considering even values of n and assuming $\beta > \beta_{\rm c}$, we have that*

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \Big\{ \sum_{|r|\leq n\epsilon} \mathbb{E} Z_r(G_n) \Big\} = \varphi_k(\beta, 0) < \varphi_k(\beta, h^*) \,. \tag{2.33}$$

Proof. First, following the calculus preceding (2.25) we get after some algebraic manipulations that

$$\partial_h \varphi_k(\beta, h) = \frac{k\theta}{\cosh^2(h)} [f(\tanh(u), c) - f(\tanh(h), c)]$$

for $c = c(h) \equiv \theta \tanh(h)$ and $u = u(h) \equiv (k-1)\text{atanh}(c)$, where for $c \geq 0$ the function $f(x, c) = x/(1 + cx)$ is monotone increasing in $x \geq 0$. With $\beta > \beta_{\rm c}$ we know already that $g(h) > 0$ (for $g(\cdot)$ of (2.30)), hence $u(h) > h$ for any $h \in (0, h^*)$. From the preceding expression for $\partial_h \varphi_k(\beta, h)$ and the monotonicity of $f(\cdot, c)$ we thus deduce that $\varphi_k(\beta, h^*) > \varphi_k(\beta, 0)$.

Next, since $Z_r(G) = Z_{-r}(G)$ we shall consider hereafter only $r \geq 0$, setting $s \equiv (n-r)/2$. Further, let $\Delta_G(\underline{x})$ denote the number of edges $(i,j) \in E$ such that $x_i \neq x_j$ and $Z_r(G,\Delta)$ be the number of configurations $\underline{x} \in \Omega_r$ such that $\Delta_G(\underline{x}) = \Delta$. Since $|E| = m$ it follows that $\sum_{(i,j)\in E} x_i x_j = m - 2\Delta_G(\underline{x})$ and hence

$$Z_r(G) = e^{\beta m} \sum_{\Delta=0}^{m} Z_r(G,\Delta)\, e^{-2\beta\Delta}\,.$$

By the linearity of the expectation and since the distribution of G_n (chosen uniformly from $\mathbb{G}(k,n)$) is invariant under any permutation of the vertices, we have that

$$\begin{aligned}\mathbb{E}\{\, Z_r(G_n,\Delta)\} &= \sum_{\underline{x}\in\Omega_r} \mathbb{P}\{\Delta_{G_n}(\underline{x}) = \Delta\} = \binom{n}{s}\mathbb{P}\{\Delta_{G_n}(\underline{x}^*) = \Delta\} \\ &= \binom{n}{s} \frac{|\{G \in \mathbb{G}(k,n) \text{ and } \Delta_G(\underline{x}^*) = \Delta\}|}{|\mathbb{G}(k,n)|}\,,\end{aligned}$$

where $x_i^* = -1$ for $i \leq s$ and $x_i^* = 1$ for $s < i \leq n$.

The size of the ensemble $\mathbb{G}(k,n)$ is precisely the number of pairings of nk objects, i.e.

$$|\mathbb{G}(k,n)| = \mathfrak{P}(nk) \equiv \frac{(nk)!}{(nk/2)!2^{nk/2}}\,.$$

Similarly, the number of such pairings with exactly Δ edges of unequal endpoints is

$$|\{G \in \mathbb{G}(k,n) \text{ and } \Delta_G(\underline{x}^*) = \Delta\}| = \binom{ks}{\Delta}\binom{\hat{n}}{\Delta}\,\Delta!\,\mathfrak{P}(ks-\Delta)\mathfrak{P}(\hat{n}-\Delta)\,,$$

where $\hat{n} \equiv k(n-s)$. Putting everything together we get that

$$\mathbb{E}\{Z_r(G_n)\} = \frac{e^{\beta m}}{\mathfrak{P}(2m)}\binom{n}{s}\sum_{\Delta=0}^{ks}\binom{ks}{\Delta}\binom{\hat{n}}{\Delta}\,\Delta!\,\mathfrak{P}(ks-\Delta)\mathfrak{P}(\hat{n}-\Delta)e^{-2\beta\Delta}\,. \quad (2.34)$$

Recall that for any $q \in [0,1]$

$$n^{-1}\log\binom{n}{nq} = H(q) + o(1)\,, \qquad n^{-1}\log\mathfrak{P}(n) = \frac{1}{2}\log\left(\frac{n}{e}\right) + o(1)\,,$$

where $H(x) \equiv -x\log x - (1-x)\log(1-x)$ denotes the binary entropy function.

Setting $\Delta = \delta k n$, $s = un$ and

$$\psi_\beta(u,\delta) \equiv (u-\delta)\log(u-\delta) + (1-u-\delta)\log(1-u-\delta) + 2\delta\log\delta + 4\beta\delta\,,$$

we find upon substituting these estimates in the expression (2.34) that

$$n^{-1}\log\mathbb{E}\{Z_r(G_n)\} = \frac{\beta k}{2} + (1-k)H(u) - \frac{k}{2}\inf_{\delta\in[0,u]}\psi_\beta(u,\delta) + o(1)\,.$$

Differentiating $\psi_\beta(u,\delta)$ in δ we deduce that the infimum in the preceding expression is achieved for the positive solution $\delta = \delta_*(\beta,u)$ of $(u-\delta)(1-u-\delta) = \delta^2 e^{4\beta}$. Using this value of δ we get that $n^{-1}\log\mathbb{E}\{Z_r(G_n)\} = \eta_k(\beta,u) + o(1)$, where

$$\eta_k(\beta,u) \equiv \frac{\beta k}{2} + (1-k)H(u) - \frac{k}{2}\big\{u\log(u-\delta_*(\beta,u)) + (1-u)\log(1-u-\delta_*(\beta,u))\big\}\,.$$

Next, note that $\delta_*(\beta,1/2) = 1/[2(1+e^{2\beta})]$ from which we obtain after some elementary algebraic manipulations that $\eta_k(\beta,1/2) = \varphi_k(\beta,0)$. Further, as $\eta_k(\beta,u)$ is continuous in $u \geq 0$, we conclude that

$$\limsup_{n\to\infty} n^{-1}\log\Big\{\sum_{|r|\leq\epsilon n}\mathbb{E}Z_r(G_n)\Big\} = \sup_{|2u-1|\leq\epsilon}\eta_k(\beta,u)\,,$$

which for $\epsilon\to 0$ converges to $\eta_k(\beta,1/2) = \varphi_k(\beta,0)$, as claimed. □

3. The Bethe-Peierls approximation

Bethe-Peierls approximation reduces the problem of computing partition functions and expectation values to the one of solving a set of non-linear equations. While in general this 'reduction' involves an uncontrolled error, for mean-field models it is expected to be asymptotically exact in the large system limit. In fact, in Section 2 we saw such a result for the ferromagnetic Ising model on sparse tree-like graphs.

Bethe states, namely those distributions that are well approximated within the Bethe-Peierls scheme play for mean-field models the role that pure Gibbs states do on infinite lattices (for the latter see [42]). For example, it is conjectured by physicists that a large class of models, including for instance the examples in Section 1, decompose into convex combinations of Bethe states.

In the context of mean field spin glasses, the Bethe-Peierls method was significantly extended by Mézard, Parisi and Virasoro to deal with proliferation of pure states [70]. In the spin glass jargon, this phenomenon is referred to as 'replica symmetry breaking,' and the whole approach is known as the 'cavity method'. A closely related approach is provided by the so-called TAP (Thouless-Anderson-Palmer) equations [70].

Section 3.1 outlines the rationale behind the Bethe-Peierls approximation of local marginals, based on the Bethe mean field equations (and the belief propagation algorithm for iteratively solving them). Complementing it, Section 3.2 introduces the Bethe free entropy. In Section 3.3 we explain how these ideas apply to the ferromagnetic Ising, the Curie-Weiss model, the Sherrington-Kirkpatrick model and the independent set model. Finally, in Section 3.4 we define a notion of correlation decay which generalizes the so called 'extremality condition' in trees. We show that if the graphical model associated with a permissive graph-specification pair $(G, \underline{\psi})$ satisfies such correlation decay condition then it is a Bethe state. Subject to a slightly stronger condition, [30] validates also the Bethe-Peierls approximation for its free entropy.

While in general extremality on the graph G does not coincide with extremality on the associated tree model, in Section 5 we shall provide a sufficient condition for this to happen for models on random graphs.

3.1. Messages, belief propagation and Bethe equations

Given a variable domain $\mathcal{X}$ and a simple finite graph $G \equiv (V, E)$ without double edges or self loops, let $\vec{E} \equiv \{i \to j : (i,j) \in E\}$ denote the induced set of directed edges. The Bethe-Peierls method provides an approximation for the marginal on $U \subset V$ of the probability measure $\mu \equiv \mu_{G,\underline{\psi}}$ cf. Eq. (1.4). The basic idea is to describe the influence of the factors outside U via factorized boundary conditions. Such a boundary law is fully specified by a collection of distributions on $\mathcal{X}$ indexed by the directed edges on the 'internal' boundary $\partial U = \{i \in U : \partial i \not\subseteq U\}$ of U (where as usual ∂i is the set of neighbors of $i \in V$). More precisely, this is described by appropriately choosing a *set of messages*.

Definition 3.1. *A* set of messages *is a collection* $\{\nu_{i\to j}(\,\cdot\,) : i \to j \in \vec{E}\}$ *of probability distributions over* $\mathcal{X}$ *indexed by the directed edges in* G.

A set of messages is permissive *for a permissive graph-specification pair* $(G, \underline{\psi})$ *if* $\nu_{i\to j}(x_i^{\mathrm{p}})$ *are positive and further* $\nu_{i\to j}(\,\cdot\,) = \psi_i(\,\cdot\,)/z_i$ *whenever* $\partial i = \{j\}$.

As we shall soon see, in this context the natural candidate for the Bethe-Peierls approximation is the following *standard message set*.

Definition 3.2. *The* standard message set *for the canonical probability measure μ associated to a permissive graph-specification pair $(G, \underline{\psi})$ is $\nu^*_{i\to j}(x_i) \equiv \mu^{(ij)}_i(x_i)$, that is, the marginal on i of the probability measure on $\mathcal{X}^V$*

$$\mu^{(ij)}(\underline{x}) = \frac{1}{Z_{ij}} \prod_{(k,l)\in E\setminus(i,j)} \psi_{kl}(x_k, x_l) \prod_{k\in V} \psi_k(x_k)\,, \tag{3.1}$$

obtained from equation (1.4) upon 'taking out' the contribution $\psi_{ij}(\cdot,\cdot)$ of edge (i,j) (and with Z_{ij} an appropriate normalization constant).

Remark 3.3. *Since $\underline{\psi}$ is permissive, the measure $\mu^{(ij)}(\cdot)$ is well defined and strictly positive at $\underline{x} = (x^{\rm p}_1, \ldots, x^{\rm p}_n)$. Further, the marginal on i of $\mu^{(ij)}(\cdot)$ is precisely $\psi_i(x_i)/\sum_x \psi_i(x)$ whenever $\partial i = \{j\}$, so the collection $\{\nu^*_{i\to j}(\cdot)\}$ is indeed a permissive set of messages (per Definition 3.1).*

In order to justify the Bethe-Peierls method let $\mu^{(i)}(\,\cdot\,)$ denote the probability measure obtained from the canonical measure of (1.4) when the vertex $i \in V$ and all edges incident on i are removed from G. That is,

$$\mu^{(i)}(\underline{x}) \equiv \frac{1}{Z_i} \prod_{(k,l)\in E,\, i\notin(k,l)} \psi_{kl}(x_k, x_l) \prod_{k\in V, k\neq i} \psi_k(x_k)\,. \tag{3.2}$$

For any $U \subseteq V$ we let μ_U (respectively, $\mu^{(ij)}_U$, $\mu^{(i)}_U$), denote the marginal distribution of $\underline{x}_U \equiv \{x_i : i \in U\}$ when $\underline{x}$ is distributed according to μ (respectively $\mu^{(ij)}$, $\mu^{(i)}$).

Clearly, finding good approximations to the marginals of the modified models $\mu^{(ij)}$, $\mu^{(i)}$ is essentially equivalent to finding good approximations for the original model μ. Our first step consists of deriving an identity between certain marginals of $\mu^{(ij)}$ in terms of marginals of $\mu^{(i)}$. Hereafter, we write $f(\cdot) \cong g(\cdot)$ whenever two non-negative functions f and g on the same domain differ only by a positive normalization constant. By definition we then have that

$$\mu^{(ij)}_{ij}(x_i, x_j) \cong \psi_i(x_i) \sum_{\underline{x}_{\partial i\setminus j}} \mu^{(i)}_{\partial i}(\underline{x}_{\partial i}) \prod_{l\in\partial i\setminus j} \psi_{il}(x_i, x_l)\,. \tag{3.3}$$

To proceed, we let $\nu^*_{i\to j}(x_i) \equiv \mu^{(ij)}_i(x_i)$ and make the crucial approximate

independence assumptions

$$\mu_{ij}^{(ij)}(x_i, x_j) = \nu_{i\to j}^*(x_i)\nu_{j\to i}^*(x_j) + \mathsf{ERR}\,, \tag{3.4}$$

$$\mu_{\partial i}^{(i)}(\underline{x}_{\partial i}) = \prod_{l\in\partial i} \nu_{l\to i}^*(x_l) + \mathsf{ERR}\,, \tag{3.5}$$

where the error terms ERR are assumed to be small. Indeed, upon neglecting the error terms, plugging these expressions in equation (3.3), setting $x_j = x_j^{\rm p}$ and dividing by the positive common factor $\nu_{j\to i}^*(x_j)$, we get the following *Bethe equations*.

Definition 3.4. *Let $\mathcal{M}(\mathcal{X})$ denote the space of probability measures over $\mathcal{X}$ and consider the Bethe (or* belief propagation, BP*) mapping T of the space $\mathcal{M}(\mathcal{X})^{\vec{E}}$ of possible message sets to itself, whose value at ν is*

$$(\mathsf{T}\nu)_{i\to j}(x_i) \equiv \frac{\psi_i(x_i)}{z_{i\to j}} \prod_{l\in\partial i\setminus j} \Big[\sum_{x_l\in\mathcal{X}} \psi_{il}(x_i, x_l)\nu_{l\to i}(x_l) \Big]\,, \tag{3.6}$$

where $z_{i\to j}$ is determined by the normalization condition $\sum_{x\in\mathcal{X}}(\mathsf{T}\nu)_{i\to j}(x) = 1$. The Bethe equations *characterize fixed points of the BP mapping. That is,*

$$\nu_{i\to j}(x_i) \equiv (\mathsf{T}\nu)_{i\to j}(x_i)\,. \tag{3.7}$$

Remark 3.5. *The BP mapping T is well defined when the specification $\underline{\psi}$ is permissive. Indeed, in such a case there exists for each $i \to j \in \vec{E}$ and any message set ν, a positive constant $z_{i\to j} \geq \psi_{\min}^{|\partial i|}$ for which $(\mathsf{T}\nu)_{i\to j} \in \mathcal{M}(\mathcal{X})$.*

Moreover, in this case by definition $(\mathsf{T}\nu)_{i\to j}(x)$ is positive at $x = x_i^{\rm p}$ and further, equals $\psi_i(x)/z_i$ whenever $\partial i = \{j\}$. In particular, any solution of the Bethe equations is a permissive set of messages.

These equations characterize the set of messages $\{\nu_{i\to j}^*(\cdot)\}$ to be used in the approximation. Bethe-Peierls method estimates marginals of the graphical model $\mu_{G,\underline{\psi}}$ in a manner similar to that expressed by (3.4) and (3.5). For instance, $\mu_i(\cdot)$ is then approximated by

$$\mu_i(x_i) \cong \psi_i(x_i) \prod_{j\in\partial i} \sum_{x_j} \psi_{ij}(x_i, x_j)\nu_{j\to i}^*(x_j)\,. \tag{3.8}$$

A more general expression will be provided in Section 3.4.

At this point the reader can verify that if G is a (finite) tree then the error terms in equations (3.4) and (3.5) vanish, hence in this case the Bethe

equations have a unique solution, which is precisely the standard message set for the canonical measure μ. More generally, it is not hard to verify that in the framework of a (permissive) specification $\underline{\psi}$ for a factor graph $G = (V, F, E)$ the Bethe equations are then

$$\begin{aligned}\nu_{a\to i}(x_i) &\cong \sum_{\underline{x}_{\partial a\setminus i}} \psi_a(\underline{x}_{\partial a}) \prod_{l\in\partial a\setminus i} \nu_{l\to a}(x_l)\,,\\ \nu_{i\to a}(x_i) &\cong \prod_{b\in\partial i\setminus a} \nu_{b\to i}(x_i)\end{aligned}$$

and that when the factor graph is a (finite) tree these equations have a unique solution which is precisely the standard message set for the (canonical) measure $\mu_{G,\underline{\psi}}(\cdot)$ of (1.23). That is, $\nu_{i\to a}(\cdot)$ and $\nu_{a\to i}(\cdot)$ are then the marginals on variable i for factor graphs in which factor a and all factors in $\partial i \setminus a$ are removed, respectively.

In view of the preceding, we expect such an approximation to be tight as soon as G lacks short cycles or for a sequence of graphs that converges locally to a tree.

3.2. The Bethe free entropy

Within the Bethe approximation all marginals are expressed in terms of the permissive messages $\{\nu_{i\to j}\}$ that solve the Bethe equations (3.7). Not surprisingly, the free entropy $\log Z(G,\underline{\psi})$ can also be approximated in terms as the *Bethe free entropy* at this message set.

Definition 3.6. *The real valued function on the space of permissive message sets*

$$\begin{aligned}\Phi_{G,\underline{\psi}}(\nu) &= -\sum_{(i,j)\in E} \log\Big\{\sum_{x_i,x_j} \psi_{ij}(x_i,x_j)\nu_{i\to j}(x_i)\nu_{j\to i}(x_j)\Big\}\\ &\quad+\sum_{i\in V}\log\Big\{\sum_{x_i}\psi_i(x_i)\prod_{j\in\partial i}\sum_{x_j}\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\Big\}\,, \quad (3.9)\end{aligned}$$

is called the Bethe free entropy *associated with the given permissive graph-specification pair $(G,\underline{\psi})$. In the following we shall often drop the subscripts and write $\Phi(\nu)$ for the Bethe free entropy.*

In the spirit of the observations made at the end of Section 3.1, this approximation is exact whenever G is a tree and the Bethe messages are used.

Proposition 3.7. *Suppose G is a tree and let ν^* denote the unique solution of the Bethe equations (3.7). Then,* $\log Z(G,\underline{\psi}) = \Phi_{G,\underline{\psi}}(\nu^*)$.

Proof. We progressively disconnect the tree G in a recursive fashion. In doing so, note that if $f(x) = f_1(x)f_2(x)/f_3(x)$ and $f_a(x) \cong p(x)$ for $a \in \{1,2,3\}$ and some probability distribution p, then

$$\log\{\sum_x f(x)\} = \log\{\sum_x f_1(x)\} + \log\{\sum_x f_2(x)\} - \log\{\sum_x f_3(x)\} \quad (3.10)$$

(adopting hereafter the convention that $0/0 = 0$).

Proceeding to describe the first step of the recursion, fix an edge $(i,j) \in E$. Without this edge the tree G breaks into disjoint subtrees $G^{(i)}$ and $G^{(j)}$ such that $i \in G^{(i)}$ and $j \in G^{(j)}$. Consequently, the measure $\mu^{(ij)}(\cdot)$ of (3.1) is then the product of two canonical measures, corresponding to the restriction of the specification $\underline{\psi}$ to $G^{(i)}$ and to $G^{(j)}$, respectively. Let $Z_{i\to j}(x)$ denote the constrained partition function for the specification $\underline{\psi}$ restricted to the subtree $G^{(i)}$ whereby we force the variable x_i to take the value x. With $Z_{j\to i}(x)$ defined similarly for the subtree $G^{(j)}$, we obviously have that

$$Z(G,\underline{\psi}) = \sum_{x_i,x_j} Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j)\,.$$

Further, recall our earlier observation that for a tree G the unique solution $\{\nu^*_{i\to j}(\cdot)\}$ of (3.7) is $\{\mu_i^{(ij)}(\cdot)\}$. Hence, in this case $Z_{i\to j}(x_i) \cong \nu^*_{i\to j}(x_i)$, $Z_{j\to i}(x_j) \cong \nu^*_{j\to i}(x_j)$ and $\nu^*_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu^*_{j\to i}(x_j) \cong \mu_{ij}(x_i,x_j)$. Setting $\psi^*_{i\to j}(x_i,x_j) \equiv \nu^*_{i\to j}(x_i)\psi_{ij}(x_i,x_j)$ we next apply the identity (3.10) for $x = (x_i,x_j)$, $f_1(x) = Z_{i\to j}(x_i)\psi^*_{j\to i}(x_i,x_j)$, $f_2(x) = \psi^*_{i\to j}(x_i,x_j)Z_{j\to i}(x_j)$ and $f_3(x) = \nu^*_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu^*_{j\to i}(x_j)$ to get that

$$\log Z(G,\underline{\psi}) = \log Z(G^{(i\to j)},\underline{\psi}^{(i\to j)}) + \log Z(G^{(j\to i)},\underline{\psi}^{(j\to i)}) - \log\varphi(i,j)\,,$$

where for each edge $(i,j) \in E$,

$$\varphi(i,j) \equiv \sum_{x_i,x_j} \nu^*_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu^*_{j\to i}(x_j)$$

and the term

$$Z(G^{(i\to j)},\underline{\psi}^{(i\to j)}) \equiv \sum_{x_i,x_j} Z_{i\to j}(x_i)\psi^*_{j\to i}(x_i,x_j)\,,$$

is the partition function for the (reduced size) subtree $G^{(i\to j)}$ obtained when adding to $(G^{(i)},\underline{\psi}^{(i)})$ the edge (i,j) and the vertex j whose specification is

now $\psi^*_j \equiv \nu^*_{j\to i}$. We have the analogous representation for

$$Z(G^{(j\to i)}, \underline{\psi}^{(j\to i)}) \equiv \sum_{x_i,x_j} \psi^*_{i\to j}(x_i,x_j) Z_{j\to i}(x_i)\,.$$

It is not hard to verify that the unique solution of the Bethe equations (3.7) for the graph-specification $(G^{(i\to j)}, \underline{\psi}^{(i\to j)})$ coincides with $\nu^*(\cdot)$ at all directed edges of $G^{(i\to j)}$. Likewise, the unique solution of the Bethe equations (3.7) for the graph-specification $(G^{(j\to i)}, \underline{\psi}^{(j\to i)})$ coincides with $\nu^*(\cdot)$ at all directed edges of $G^{(j\to i)}$. Thus, recursively repeating this operation until we have dealt once with each edge of G, we find a contribution $-\log\varphi(k,l)$ from each $(k,l)\in E$, the sum of which is precisely the first term in (3.9), evaluated at the permissive set of messages $\nu^*_{i\to j}(\cdot)$. The residual graph remaining at this stage consists of disconnected 'stars' centered at vertices of G, with specification $\nu^*_{l\to k}(x_l)$ at vertices $l\in\partial k$ for the 'star' centered at $k\in V$ (and original specification at vertex k and the edges $(k,l)\in E$). The log-partition function for such star is $\log\{\sum_{x_k}\psi_k(x_k)\prod_{l\in\partial k}\sum_{x_l}\psi^*_{l\to k}(x_l,x_k)\}$ so the aggregate of these contributions over all vertices of G is precisely the second term in (3.9), evaluated at $\nu^*_{i\to j}(\cdot)$. □

Lemma 3.8. *Solutions of the Bethe equations (3.7) for a given permissive graph-specification pair $(G,\underline{\psi})$ are stationary points of the corresponding Bethe free entropy $\Phi_{G,\underline{\psi}}(\cdot)$. The converse holds when the $|\mathcal{X}|$-dimensional matrices $\{\psi_{ij}(x,y)\}$ are invertible for all $(i,j)\in E$.*

Proof. From the formula (3.9) and our definition (3.6) we find that for any $j\to i\in\vec{E}$ and $x_j\in\mathcal{X}$,

$$\frac{\partial\Phi(\nu)}{\partial\nu_{j\to i}(x_j)} = - \frac{\sum_{x_i}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)}{\sum_{x'_i,x'_j}\nu_{i\to j}(x'_i)\nu_{j\to i}(x'_j)\psi_{ij}(x'_i,x'_j)} + \frac{\sum_{x_i}(\mathsf{T}\nu)_{i\to j}(x_i)\psi_{ij}(x_i,x_j)}{\sum_{x'_i,x'_j}(\mathsf{T}\nu)_{i\to j}(x'_i)\nu_{j\to i}(x'_j)\psi_{ij}(x'_i,x'_j)}\,.$$

Hence, if $\{\nu_{i\to j}(\cdot)\}$ satisfies the Bethe equations (3.7), then $\partial\Phi(\nu)/\partial\nu_{j\to i}(x)=0$ for all $x\in\mathcal{X}$ and any $j\to i\in\vec{E}$, as claimed.

Conversely, given a permissive specification, if a permissive set of messages ν is a stationary point of $\Phi(\cdot)$, then by the preceding we have that for any $i\to j\in\vec{E}$, some positive $c_{i\to j}$ and all $y\in\mathcal{X}$,

$$\sum_x\big[(\mathsf{T}\nu)_{i\to j}(x) - c_{i\to j}\nu_{i\to j}(x)\big]\psi_{ij}(x,y)=0\,.$$

By assumption the matrices $\{\psi_{ij}(x,y)\}$ are invertible, hence $\nu_{i\to j}(x) \cong (\mathsf{T}\nu)_{i\to j}(x)$ for any $i \to j \in \vec{E}$. The probability measures $\nu_{i\to j}$ and $(\mathsf{T}\nu)_{i\to j}$ are thus identical, for each directed edge $i \to j$. That is, the set of messages ν satisfies the Bethe equations for the given specification. □

3.3. Examples: Bethe equations and free entropy

In most of this section we consider the extension of the Ising measure (2.9) on $\{+1,-1\}^V$, of the form

$$\mu_{\beta,\underline{B},\underline{J}}(\underline{x}) = \frac{1}{Z(\beta,\underline{B},\underline{J})} \exp\Big\{\beta \sum_{(i,j)\in E} J_{ij}x_i x_j + \sum_{i\in V} B_i x_i\Big\}, \tag{3.11}$$

where $\underline{J} = \{J_{ij}, (i,j) \in E\}$ for generic 'coupling constants' $J_{ij} \in \mathbb{R}$ as in the spin-glass example of (1.21). This model corresponds to the permissive specification $\psi_{ij}(x_i,x_j) = \exp(\beta J_{ij}x_i x_j)$ and $\psi_i(x_i) = \exp(B_i x_i)$. Since $\mathcal{X} = \{+1,-1\}$, any set of messages $\{\nu_{i\to j}\}$ is effectively encoded through the 'cavity fields'

$$h_{i\to j} \equiv \frac{1}{2}\log\frac{\nu_{i\to j}(+1)}{\nu_{i\to j}(-1)}. \tag{3.12}$$

Using these cavity fields, we find the following formulas.

Proposition 3.9. *The Bethe equations for the cavity fields and the measure $\mu_{\beta,\underline{B},\underline{J}}(\cdot)$ are*

$$h_{i\to j} = B_i + \sum_{l\in\partial i\setminus j} \operatorname{atanh}\{\theta_{il}\tanh(h_{l\to i})\}, \tag{3.13}$$

where $\theta_{il} \equiv \tanh(\beta J_{il})$. The expected magnetization $\langle x_i\rangle$ for this measure is approximated (in terms of the Bethe cavity fields $h^*_{i\to j}$), as

$$\langle x_i\rangle = \tanh\Big\{B_i + \sum_{l\in\partial i} \operatorname{atanh}\{\theta_{il}\tanh(h^*_{l\to i})\}\Big\}, \tag{3.14}$$

and the Bethe free entropy of any permissive cavity field $\underline{h} = \{h_{i\to j}\}$ is

$$\Phi_{G,\beta,\underline{B},\underline{J}}(\underline{h}) = \frac{1}{2}\sum_{i\in V}\sum_{j\in\partial i}\Big\{\gamma(\theta_{ij}) - \log\big[1+\theta_{ij}\tanh(h_{i\to j})\tanh(h_{j\to i})\big]\Big\} \tag{3.15}$$

$$+\sum_{i\in V}\log\Big\{e^{B_i}\prod_{j\in\partial i}[1+\theta_{ij}\tanh(h_{j\to i})] + e^{-B_i}\prod_{j\in\partial i}[1-\theta_{ij}\tanh(h_{j\to i})]\Big\},$$

where $\gamma(u) \equiv -\frac{1}{2}\log(1-u^2)$.

Proof. Expressing the BP mapping for the Ising measure $\mu(\underline{x}) \equiv \mu_{\beta,\underline{B},\underline{J}}(\underline{x})$ in terms of cavity fields we find that

$$(\mathsf{T}h)_{i\to j}(x_i) \equiv \frac{e^{B_i x_i}}{\tilde{z}_{i\to j}} \prod_{l\in\partial i\setminus j} \cosh(h_{l\to i} + \beta J_{il} x_i)$$

for some positive normalization constants $\tilde{z}_{i\to j}$. Thus, the identity

$$\frac{1}{2}\,\log\frac{\cosh(a+b)}{\cosh(a-b)} = \operatorname{atanh}(\tanh(a)\tanh(b))\,,$$

leads to the formula (3.13) for the Bethe equations. The approximation (3.8) of local marginals then results with $\mu_i(x_i) \cong e^{B_i x_i}\prod_{l\in\partial i}\cosh(h^*_{l\to i}+\beta J_{il}x_i)$, out of which we get the formula (3.14) for $\langle x_i\rangle = \mu_i(+1) - \mu_i(-1)$ by the identity $\frac{1}{2}\log(a) = \operatorname{atanh}(\frac{a-1}{a+1})$. Next note that if $u = \tanh(b)$ then $\gamma(u) = \log\cosh(b)$ and recall that by definition, for any $(i,j)\in E$ and $x\in\mathcal{X}$,

$$\nu_{j\to i}(x) = \frac{\exp(x h_{j\to i})}{2\cosh(h_{j\to i})}\,. \tag{3.16}$$

Hence, using the identity

$$\frac{1}{4}\sum_{x,y\in\{+1,-1\}} \frac{e^{axy}e^{bx}e^{cy}}{\cosh(a)\cosh(b)\cosh(c)} = 1 + \tanh(a)\tanh(b)\tanh(c)\,,$$

the first term in the formula (3.9) of the Bethe free entropy $\Phi(\cdot)$ is in this case

$$-\sum_{(i,j)\in E}\gamma(\theta_{ij}) - \sum_{(i,j)\in E}\log\left[1+\theta_{ij}\tanh(h_{i\to j})\tanh(h_{j\to i})\right].$$

Similarly, using (3.16) and the identity

$$\frac{1}{2}\sum_{x\in\{+1,-1\}} \frac{e^{axy}e^{bx}}{\cosh(a)\cosh(b)} = 1 + y\tanh(a)\tanh(b)\,,$$

for $y = x_i \in \{+1,-1\}$, we find that the second term in the formula (3.9) is in our case

$$\sum_{i\in V}\sum_{j\in\partial i}\gamma(\theta_{ij}) + \sum_{i\in V}\log\Big\{e^{B_i}\prod_{j\in\partial i}[1+\theta_{ij}\tanh(h_{j\to i})] + e^{-B_i}\prod_{j\in\partial i}[1-\theta_{ij}\tanh(h_{j\to i})]\Big\}\,.$$

Combining the preceding expressions for the two terms of (3.9) we arrive at the formula of (3.15). □

We proceed with a few special models of interest.

The Curie-Weiss model. This model, which we already considered in Section 1.1, corresponds to $G = K_n$ (the complete graph of n vertices), with $B_i = B$ and $J_{ij} = 1/n$ for all $1 \le i \ne j \le n$. Since this graph-specification pair is invariant under re-labeling of the vertices, the corresponding Bethe equations (3.13) admit at least one constant solution $h^*_{i\to j} = h^*(n)$, possibly dependent on n, such that

$$h^*(n) = B + (n-1)\mathrm{atanh}\{\tanh(\beta/n)\tanh(h^*(n))\}\,.$$

These cavity fields converge as $n \to \infty$ to solutions of the (limiting) equation $h^* = B + \beta\tanh(h^*)$. Further, the Bethe approximations (3.14) for the magnetization $m(n) = \langle x_i\rangle$ are of the form $m(n) = \tanh(h^*(n)) + O(1/n)$ and thus converge as $n \to \infty$ to solutions of the (limiting) equation $m = \tanh(B+\beta m)$. Indeed, we have already seen in Theorem 1.4 that the Curie-Weiss magnetization (per spin) concentrates for large n around the relevant solutions of the latter equation.

Ising models on random k-regular graphs. By the same reasoning as for the Curie-Weiss model, in case of a k-regular graph of n vertices with $J_{ij} = +1$, and $B_i = B$, the Bethe equations admit a constant solution $h_{i\to j} = h^*$ such that

$$h^* = B + (k-1)\mathrm{atanh}\{\theta\tanh(h^*)\}\,,$$

for $\theta \equiv \tanh(\beta)$, with the corresponding magnetization approximation $m = \tanh\left(B + k\mathrm{atanh}\{\theta\tanh(h^*)\}\right)$ and Bethe free entropy

$$\begin{aligned} n^{-1}\Phi_n(h^*) = \frac{k}{2}\Big\{\gamma(\theta) - \log\left[1+\theta\tanh^2(h^*)\right]\Big\} \\ + \log\Big\{e^B[1+\theta\tanh(h^*)]^k + e^{-B}[1-\theta\tanh(h^*)]^k\Big\}\,. \end{aligned}$$

Ising models on k-regular trees. It is instructive to contrast the above free entropy with the analogous result for rooted k-regular trees $T_k(\ell)$. From Proposition 3.7 we know that the free entropy $\log Z_\ell(B,\beta)$ for the Ising measure on the finite tree $T_k(\ell)$ is precisely the Bethe free entropy of (3.15) for the unique solution of the Bethe equations (3.13) with $J_{ij} = +1$ and $B_i = B$.

We denote by n_t the number of vertices at generation $t \in \{0, \ldots, \ell\}$ (thus $n_0 = 1$ and $n_t = k(k-1)^{t-1}$ for $t \geq 1$), and by

$$n(\ell) = |T_k(\ell)| = k((k-1)^\ell - 1)/(k-2)\,,$$

the total number of vertices in $T_k(\ell)$. Due to symmetry of $T_k(\ell)$, the Bethe cavity field assumes the same value h_r on all directed edges leading from a vertex at generation $\ell - r$ to one at generation $\ell - r - 1$ of $T_k(\ell)$. Thus, we have

$$h_r = B + (k-1)\mathrm{atanh}(\theta \tanh h_{r-1})\,, \tag{3.17}$$

with initial condition $h_{-1} = 0$. Similarly, we denote by h_r^ℓ of the Bethe cavity field on the $n_{\ell-r}$ directed edges leading from a vertex at generation $\ell - r - 1$ to one at generation $\ell - r$. We then have

$$h_r^\ell = B + (k-2)\mathrm{atanh}(\theta \tanh h_r) + \mathrm{atanh}(\theta \tanh h_{r+1}^\ell)\,,$$

for $r = \ell - 1, \ell - 2, \ldots, 0$, with initial condition $h_\ell^\ell = h_{\ell-1}$. The (Bethe) free entropy is in this case

$$\begin{aligned}\log Z_\ell(B,\beta) = (n(\ell)-1)\gamma(\theta) &- \sum_{r=0}^{\ell-1} n_{\ell-r} \log\left[1 + \theta \tanh h_r \tanh h_r^\ell\right] \\ &+ \sum_{r=0}^{\ell} n_{\ell-r} \log\Big\{e^B[1+\theta\tanh h_{r-1}]^{k-1}[1+\theta\tanh h_r^\ell] \\ &\qquad + e^{-B}[1-\theta\tanh h_{r-1}]^{k-1}[1-\theta\tanh h_r^\ell]\Big\}\,.\end{aligned}$$

Using the relation (3.17) you can verify that the preceding formula simplifies to

$$\begin{aligned}\log Z_\ell(B,\beta) &= (n(\ell)-1)\gamma(\theta) \\ &+ \log\left\{e^B[1+\theta\tanh h_{\ell-1}]^k + e^{-B}[1-\theta\tanh h_{\ell-1}]^k\right\} \\ &+ \sum_{r=0}^{\ell-1} n_{\ell-r} \log\left\{e^B[1+\theta\tanh h_{r-1}]^{k-1} + e^{-B}[1-\theta\tanh h_{r-1}]^{k-1}\right\}.\end{aligned}$$

The $\ell \to \infty$ limit can then be expressed in terms of the k-canopy tree CT_k (c.f. Lemma 2.8). If R denotes the random location of the root of CT_k, then we get

$$\begin{aligned}&\lim_{\ell\to\infty} \frac{1}{n(\ell)} \log Z_\ell(B,\beta) = \\ &\qquad \gamma(\theta) + \mathbb{E}\log\left\{e^B[1+\theta\tanh h_{R-1}]^{k-1} + e^{-B}[1-\theta\tanh h_{R-1}]^{k-1}\right\}.\end{aligned}$$

Locally tree-like graphs. Recall Remark 2.7, that k-regular graphs converge locally to the Galton-Watson tree $\mathsf{T}(P,\rho,\infty)$ with $P_k = 1$. More generally, consider the ferromagnetic Ising model $\mu_{\beta,B}(\underline{x})$ of (1.20), namely, with $J_{ij} = +1$ and $B_i = B$, for a uniformly sparse graph sequence $\{G_n\}$ that converges locally to the random rooted tree T. Then, for any n and cavity field $\underline{h} = \{h_{i\to j}\}$ we have from (3.15) that

$$n^{-1}\Phi_n(\underline{h}) = \frac{1}{2}\mathbb{E}_n\Big[\sum_{j\in\partial i}\{\gamma(\theta) - \log[1+\theta\tanh(h_{i\to j})\tanh(h_{j\to i})]\}\Big] + \mathbb{E}_n\Big[\log\{e^B\prod_{j\in\partial i}[1+\theta\tanh(h_{j\to i})] + e^{-B}\prod_{j\in\partial i}[1-\theta\tanh(h_{j\to i})]\}\Big],$$

where $\mathbb{E}_n$ corresponds to expectations with respect to a uniformly chosen $i \in V_n$. For $n\to\infty$, as shown in Lemma 2.16 we have by local convergence and uniform sparsity that these expectations converge to the corresponding expectations on the tree T rooted at $\emptyset$. Consequently, we expect to have

$$\lim_{n\to\infty} n^{-1}\Phi_n(\underline{h}^*_n) = \frac{1}{2}\mathbb{E}\Big[\sum_{j=1}^{L}\{\gamma(\theta) - \log[1+\theta\tanh(h^*_{\emptyset\to j})\tanh(h^*_{j\to\emptyset})]\}\Big] + \mathbb{E}\Big[\log\{e^B\prod_{j=1}^{L}[1+\theta\tanh(h^*_{j\to\emptyset})] + e^{-B}\prod_{j=1}^{L}[1-\theta\tanh(h^*_{j\to\emptyset})]\}\Big],$$

where $L = |\partial\emptyset|$, the variables $\{\tanh(h^*_{j\to\emptyset})\}$ are the limit as $t\to\infty$ of the Ising magnetizations $\langle x_j\rangle_j^{(t)}$ on the sub-trees of $j\in\partial\emptyset$ and all its descendants (in $\mathsf{T}(t)$, either with free or plus boundary conditions), and for $j = 1,\ldots,L$,

$$h^*_{\emptyset\to j} = B + \sum_{k=1,k\neq j}^{L}\operatorname{atanh}\{\theta\tanh(h^*_{k\to\emptyset})\}\,.$$

Indeed, this is precisely the prediction (2.21) for the free entropy density of ferromagnetic Ising models on such graphs (which is proved in [29] to hold in case T is a Galton-Watson tree).

The Sherrington-Kirkpatrick model. The Sherrington-Kirkpatrick spin-glass model corresponds to the complete graph $G_n = K_n$ with the scaling $\beta\to\beta/\sqrt{n}$, constant $B_i = B$ and J_{ij} which are i.i.d. standard normal random variables. Expanding the corresponding Bethe equations (3.13), we find that for large n and any i,j,

$$h_{i\to j} = B + \frac{\beta}{\sqrt{n}}\sum_{l=1,l\neq i,j}^{n} J_{il}\tanh(h_{l\to i}) + o(\frac{1}{\sqrt{n}})\,. \tag{3.18}$$

Similarly, expanding the formula (3.14), we get for the local magnetizations $m_i \equiv \langle x_i \rangle$ and large n that

$$\operatorname{atanh}(m_i) = h_{i\to j} + \frac{\beta J_{ij}}{\sqrt{n}} \tanh(h_{j\to i}) + o(\frac{1}{\sqrt{n}}) = h_{i\to j} + \frac{\beta J_{ij}}{\sqrt{n}}\, m_j + o(\frac{1}{\sqrt{n}})\,.$$

Substituting this in both sides of equation (3.18), and neglecting terms of $O(n^{-1/2})$ yields the so-called TAP equations

$$\operatorname{atanh}(m_i) = B + \frac{\beta}{\sqrt{n}} \sum_{l=1, l\neq i}^{n} J_{il} m_l - m_i \frac{\beta^2}{n} \sum_{l=1, l\neq i}^{n} J_{il}^2 (1 - m_l^2)\,. \qquad (3.19)$$

The independent set model. In this model, which is not within the framework of (3.11), we consider the measure

$$\mu_{G,\lambda}(\underline{x}) = \frac{1}{Z(G,\lambda)}\, \lambda^{|\underline{x}|} \prod_{(i,j)\in E} \mathbb{I}((x_i, x_j) \neq (1,1))\,, \qquad (3.20)$$

where $|\underline{x}|$ denotes the number of non-zero entries in the vector $\underline{x} \in \{0,1\}^V$. It corresponds to the permissive specification $\psi_{ij}(x,y) = \mathbb{I}\big((x,y) \neq (1,1)\big)$, and $\psi_i(x) = \lambda^x$, having $x_i^{\mathrm{p}} = 0$ for all $i \in V$. In this case the Bethe equations are

$$\nu_{i\to j} = \frac{1}{1 + \lambda \prod_{l\in\partial i\setminus j} \nu_{l\to i}}\,,$$

for $\nu_{i\to j} \equiv \nu_{i\to j}(0)$ and their solution $\{\nu^*_{i\to j}\}$ provides the approximate densities

$$\mu(x_i = 1) = \frac{\lambda \prod_{j\in\partial i} \nu^*_{j\to i}}{1 + \lambda \prod_{j\in\partial i} \nu^*_{j\to i}}\,,$$

and the approximate free entropy

$$\Phi(\nu^*) = \sum_{i\in V} \log\Big\{1 + \lambda \prod_{j\in\partial i} \nu^*_{j\to i}\Big\} - \sum_{(i,j)\in E} \log[\nu^*_{i\to j} + \nu^*_{j\to i} - \nu^*_{i\to j}\nu^*_{j\to i}]\,.$$

3.4. Extremality, Bethe states and Bethe-Peierls approximation

Following upon Section 3.1 we next define the Bethe-Peierls approximation of local marginals in terms of a given set of messages. To this end, recall that each subset $U \subseteq V$ has a (possibly infinite) diameter $\operatorname{diam}(U) = \max\{d(i,j) : i,j \in U\}$ (where $d(i,j)$ is the number of edges traversed in the shortest path on G from $i \in V$ to $j \in V$), and it induces the subgraph $G_U = (U, E_U)$ such that $E_U = \{(i,j) \in E : i,j \in U\}$.

Definition 3.10. *Let $\mathcal{U}$ denote the collection of $U \subseteq V$ for which $G_U = (U, E_U)$ is a tree and each $i \in \partial U$ is a leaf of G_U (i.e. $|\partial i \cap U| = 1$ whenever $i \in \partial U$). A set of messages $\{\nu_{i\to j}\}$ induces on each $U \in \mathcal{U}$ the probability measure*

$$\nu_U(\underline{x}_U) = \frac{1}{Z_U} \prod_{i\in U} \psi_i^*(x_i) \prod_{(ij)\in E_U} \psi_{ij}(x_i, x_j)\,, \tag{3.21}$$

where $\psi_i^*(\cdot) = \psi_i(\cdot)$ except for $i \in \partial U$ in which case $\psi_i(\cdot) = \nu_{i\to u(i)}(\cdot)$ with $\{u(i)\} = \partial i \cap U$.

A probability measure $\rho(\underline{x})$ on $\mathcal{X}^V$ is (ε, r)-Bethe approximated *by a set of messages $\{\nu_{i\to j}\}$ if*

$$\sup_{U\in\mathcal{U},\mathrm{diam}(U)\le 2r} ||\rho_U - \nu_U||_{\mathrm{TV}} \le \varepsilon\,, \tag{3.22}$$

where $\rho_U(\,\cdot\,)$ denotes the marginal distribution of $\underline{x}_U$ under $\rho(\cdot)$. We call any such $\rho(\cdot)$ an (ε, r)-Bethe state *for the graph-specification pair $(G, \underline{\psi})$.*

Remark 3.11. *Note that if $i \notin \partial U$ is a leaf of an induced tree G_U then $\partial i = \{u(i)\}$ and if $\{\nu_{i\to j}\}$ is a permissive set of messages then $\nu_{i\to u(i)}(\cdot) \cong \psi_i(\cdot)$. Consequently, in (3.21) we may and shall not distinguish between ∂U and the collection of all leaves of G_U.*

We phrase our error terms and correlation properties in terms of valid rate functions, and consider graphs that are *locally tree-like*. Namely,

Definition 3.12. *A* valid rate function *is a monotonically non-increasing function $\delta : \mathbb{N} \to [0, 1]$ that decays to zero as $r \to \infty$. By (eventually) increasing $\delta(r)$, we assume, without loss of generality, that $\delta(r+1) \ge \delta_*\delta(r)$ for some positive δ_* and all $r \in \mathbb{N}$.*

Given an integer $R \ge 0$ we say that G is R-tree like *if its girth exceeds $2R + 1$ (i.e. $\mathsf{B}_i(R)$ is a tree for every $i \in V$).*

We show in the sequel that the Bethe approximation holds when the canonical measure on a tree like graph satisfies the following correlation decay hypotheses.

Definition 3.13. *A probability measure ρ on $\mathcal{X}^V$ is* extremal *for G with valid rate function $\delta(\cdot)$ if for any $A, B \subseteq V$,*

$$||\rho_{A,B}(\,\cdot\,,\,\cdot\,) - \rho_A(\,\cdot\,)\rho_B(\,\cdot\,)||_{\mathrm{TV}} \le \delta(d(A, B))\,, \tag{3.23}$$

where $d(A, B) = \min\{d(i, j) : i \in A, j \in B\}$ is the length of the shortest path in G between $A \subseteq V$ and $B \subseteq V$.

We consider the notions of Bethe measure and extremality for general probability distributions over $\mathcal{X}^V$ (and not only for the canonical measure $\mu_{G,\underline{\psi}}(\,\cdot\,)$). The key (unproven) assumption of statistical physics approaches is that the canonical measure (which is ultimately, the object of interest), can be decomposed as a unique convex combination of extremal measures, up to small error terms. This motivates the name 'extremal'. Further, supposedly each element of this decomposition can then be treated accurately within its Bethe approximation.

Here is the first step in verifying this broad conjecture, dealing with the case where the canonical measure $\mu_{G,\underline{\psi}}(\,\cdot\,)$ is itself extremal.

Theorem 3.14. *Let $\underline{\psi}$ be a permissive specification for an R-tree like graph G and $\delta(\,\cdot\,)$ a valid rate function. If $\mu_{G,\underline{\psi}}(\cdot)$ is extremal with rate $\delta(\,\cdot\,)$ then it is (ε, r)-Bethe approximated by its standard message set for $\varepsilon = \exp(c^r)\delta(R-r)$ and all $r < R-1$, where the (universal) constant c depends only on $|\mathcal{X}|$, δ_*, κ and the maximal degree $\Delta \geq 2$ of G. In particular, $\mu_{G,\underline{\psi}}(\,\cdot\,)$ is then an (ε, r)-Bethe state for this graph-specification pair.*

To prove the theorem, recall first that for any probability measures ρ_a on a discrete set $\mathcal{Z}$ and $f : \mathcal{Z} \mapsto [0, f_{\max}]$ we have the elementary bound

$$||\widehat{\rho}_1 - \widehat{\rho}_2||_{\text{TV}} \leq \frac{3 f_{\max}}{2\langle \rho_1, f\rangle} ||\rho_1 - \rho_2||_{\text{TV}}\,, \tag{3.24}$$

where $\widehat{\rho}_a(z) \equiv \rho_a(z) f(z)/\langle \rho_a, f\rangle$ and $\langle \rho_a, f\rangle \equiv \sum_{z\in\mathcal{Z}} \rho_a(z) f(z)$ (c.f. [29, Lemma 3.3]). Further, it is easy to check that if $\mu(\cdot) = \mu_{G,\underline{\psi}}(\cdot)$ and $(G, \underline{\psi})$ is a permissive graph-specification pair, then for any $C \subseteq V$,

$$\mu_C(\underline{x}_C^{\text{p}}) \quad \geq \quad \mathcal{X}^{-|C|} \kappa^{\Delta|C|}\,. \tag{3.25}$$

In addition, as shown in [30, Section 3], for such $\mu(\cdot)$, if $G_{U'}$ is a tree, $(i,j) \in E_{U'}$ and $j \notin A \supseteq \partial U'$, then

$$||\mu^{(ij)}_{i|A}(\,\cdot\,|\underline{x}_A) - \mu^{(ij)}_{i|A}(\,\cdot\,|\underline{y}_A)||_{\text{TV}} \leq b||\mu_{ij|A}(\,\cdot\,|\underline{x}_A) - \mu_{ij|A}(\,\cdot\,|\underline{y}_A)||_{\text{TV}}\,, \tag{3.26}$$

for $b \equiv 2|\mathcal{X}|\kappa^{-(\Delta+1)}$ and all $\underline{x}, \underline{y} \in \mathcal{X}^V$. Finally, the following lemma is also needed for our proof of the theorem.

Lemma 3.15. *If the canonical measure μ for 2-tree like graph and a permissive specification is extremal of valid rate function $\delta(\cdot)$ then for some finite $K = K(|\mathcal{X}|, \kappa, \Delta)$ and any $A \subseteq V$*

$$||\mu^{(ij)}_A - \mu_A||_{\text{TV}} \leq K\delta\big(d(\{i,j\}, A)\big)\,.$$

Proof. Set $B = \partial i \cup \partial j \setminus \{i,j\}$ and $C = \bigcup_{l\in B} \partial l$ noting that $|B| \le 2(\Delta-1)$, $|C| \le 2\Delta(\Delta-1)$ and since G is 2-tree like, necessarily the induced subgraph G_B has no edges. Hence,

$$\begin{aligned}\mu_B(\underline{x}_B) \ge& \mu_C(\underline{x}^{\rm p}_C)\mu_{B|C}(\underline{x}_B|\underline{x}^{\rm p}_C)\\ \ge& \mu_C(\underline{x}^{\rm p}_C)\prod_{l\in B}\Big(\frac{\psi_l(x_l)\prod_{k\in\partial l}\psi_{lk}(x_l,x^{\rm p}_k)}{\sum_{x'_l}\psi_l(x'_l)\prod_{k\in\partial l}\psi_{lk}(x'_l,x^{\rm p}_k)}\Big) \ge \mu_C(\underline{x}^{\rm p}_C)\,\kappa^{2(\Delta^2-1)}\end{aligned}$$

so by the bound (3.25) we deduce that $\mu_B(\underline{x}_B) \ge c_0$ for all $\underline{x}_B$ and some positive $c_0 = c_0(|\mathcal{X}|,\kappa,\Delta)$. Next assume, without loss of generality, that $A \cap B = \emptyset$. Then

$$\begin{aligned}||\mu^{(ij)}_A - \mu_A||_{\rm TV} =& \frac{1}{2}\sum_{\underline{x}_A}\Big|\sum_{\underline{x}_B}\mu^{(ij)}_B(\underline{x}_B)\mu_{A|B}(\underline{x}_A|\underline{x}_B) - \sum_{\underline{x}'_B}\mu_B(\underline{x}'_B)\mu_{A|B}(\underline{x}_A|\underline{x}'_B)\Big|\\ \le& \sup_{\underline{x}_B,\underline{x}'_B}||\mu_{A|B}(\,\cdot\,|\underline{x}_B) - \mu_{A|B}(\,\cdot\,|\underline{x}'_B)||_{\rm TV}\\ \le& \frac{1}{c_0^2}\mathbb{E}\Big\{||\mu_{A|B}(\,\cdot\,|\underline{X}^{(1)}_B) - \mu_{A|B}(\,\cdot\,|\underline{X}^{(2)}_B)||_{\rm TV}\Big\},\end{aligned}$$

where $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent random configurations, each of distribution μ. Next, from the extremality of $\mu(\cdot)$ we deduce that

$$\mathbb{E}\{||\mu_{A|B}(\,\cdot\,|\underline{X}^{(1)}_B) - \mu_{A|B}(\,\cdot\,|\underline{X}^{(2)}_B)||_{\rm TV}\} \le 2\delta\big(d(A,B)\big)\,,$$

so taking $K = 2/c_0^2$ we arrive at our thesis. □

Proof of Theorem 3.14. Fixing $r < R-1$, a permissive graph-specification pair $(G,\underline{\psi})$ that is extremal for R-tree like graph G with valid rate function $\delta(\,\cdot\,)$ and $U \in \mathcal{U}$ with $\mathrm{diam}(U) \le 2r$, let $\overline{U}_{R'} = \{k \in V : d(k,U) \ge R'\}$ for $R' = R - r > 1$. Note that

$$\begin{aligned}||\mu_U(\cdot) - \nu_U(\cdot)||_{\rm TV} \le& \mathbb{E}||\mu_U(\,\cdot\,) - \mu_{U|\overline{U}_{R'}}(\,\cdot\,|\underline{\widetilde{X}}_{\overline{U}_{R'}})||_{\rm TV}\\ &+ \mathbb{E}||\mu_{U|\overline{U}_{R'}}(\,\cdot\,|\underline{\widetilde{X}}_{\overline{U}_{R'}}) - \nu_U(\,\cdot\,)||_{\rm TV}\,,\end{aligned} \tag{3.27}$$

where ν_U corresponds to the standard message set (i.e. $\nu_{i\to j} = \mu^{(ij)}_i$ for the measure $\mu^{(ij)}(\cdot)$ of (3.1)), and the expectation is with respect to the random configuration $\underline{\widetilde{X}}$ of distribution μ. The first term on the right side is precisely $||\mu_{U,\overline{U}_{R'}}(\,\cdot\,,\,\cdot\,) - \mu_U(\,\cdot\,)\mu_{\overline{U}_{R'}}(\,\cdot\,)||_{\rm TV}$ which for $\mu(\cdot)$ extremal of valid

rate function $\delta(\cdot)$ is bounded by $\delta(d(U,\overline{U}_{R'})) = \delta(R-r)$. Turning to the second term, consider the permissive set of messages

$$\widetilde{\nu}_{i\to j}(x_i) = \mu^{(ij)}_{i|\overline{\mathsf{B}}_i(R')}(x_i|\underline{\widetilde{X}}_{\overline{\mathsf{B}}_i(R')})\,,$$

where $\overline{\mathsf{B}}_i(t)$ denotes the collection of vertices of distance at least t from i. Since $\mathrm{diam}(U) \le 2r$ there exists $i_o \in V$ such that $U \subseteq \mathsf{B}_{i_o}(r)$ and as $\mathsf{B}_{i_o}(R)$ is a tree, the canonical measure for $\mathsf{B}_{i_o}(R) \setminus G_U$ is the product of the corresponding measures for the subtrees rooted at $i \in \partial U$. Noting that $V \setminus \mathsf{B}_{i_o}(R) \subseteq \overline{U}_{R'}$, it is thus not hard to verify that we have the representation

$$\mu_{U|\overline{U}_{R'}}(\underline{x}_U|\underline{\widetilde{X}}_{\overline{U}_{R'}}) = \frac{1}{\widetilde{Z}_U}\prod_{i\in U}\widetilde{\psi}^*_i(x_i)\prod_{(ij)\in E_U}\psi_{ij}(x_i,x_j)\,, \tag{3.28}$$

as in (3.21), corresponding to the messages $\{\widetilde{\nu}_{i\to j}\}$ (i.e. with $\widetilde{\psi}^*_i(\cdot) = \psi_i(\cdot)$ except for $i \in \partial U$ in which case $\widetilde{\psi}_i(\cdot) = \widetilde{\nu}_{i\to u(i)}(\cdot)$). Consequently, we proceed to bound $||\widetilde{\nu}_U - \nu_U||_{\mathrm{TV}}$ by applying the inequality (3.24) for the function

$$f(\underline{x}_U) = \prod_{i\in U\setminus\partial U}\psi_i(x_i)\prod_{(ij)\in E_U}\psi_{ij}(x_i,x_j)$$

on $\mathcal{Z} = \mathcal{X}^U$ and probability measures ρ_a that are uniform on $\mathcal{X}^{U\setminus\partial U}$ with $\rho_1(\underline{x}_{\partial U}) = \prod_{i\in\partial U}\nu_{i\to u(i)}(x_i)$ and $\rho_2(\underline{x}_{\partial U}) = \prod_{i\in\partial U}\widetilde{\nu}_{i\to u(i)}(x_i)$. To this end, recall that $f(\underline{x}_U) \le f_{\max} = \psi^M_{\max}$ for $M = |U| - |\partial U| + |E_U|$. Further, since G_U is a tree (hence $|E_U| \le |U|$), and $\underline{\psi}$ is a permissive specification (also when (i,j) is removed from E), upon applying (3.25) for $|C| = 1$, we have that

$$\begin{aligned}\langle\rho_1, f\rangle &\ge \prod_{i\in U\setminus\partial U}\frac{\psi_i(x^{\mathrm{p}}_i)}{|\mathcal{X}|}\prod_{(ij)\in E_U}\psi_{ij}(x^{\mathrm{p}}_i,x^{\mathrm{p}}_j)\prod_{i\in\partial U}\nu_{i\to u(i)}(x^{\mathrm{p}}_i)\\ &\ge f_{\max}|\mathcal{X}|^{-|U|}\kappa^{M+\Delta|\partial U|} \ge f_{\max}c_1^{-|U|}\,,\end{aligned}$$

where $c_1 = |\mathcal{X}|\kappa^{-(\Delta+1)}$ is a finite constant. Consequently, we deduce upon applying (3.24) that

$$\begin{aligned}||\mu_{U|\overline{U}_R}(\,\cdot\,|\underline{\widetilde{X}}_{\overline{U}_R}) - \nu_U(\,\cdot\,)||_{\mathrm{TV}} &= ||\widehat{\rho}_2 - \widehat{\rho}_1||_{\mathrm{TV}} \le 2c_1^{|U|}||\rho_1 - \rho_2||_{\mathrm{TV}}\\ &\le 2c_1^{|U|}\sum_{i\in\partial U}||\nu_{i\to u(i)} - \widetilde{\nu}_{i\to u(i)}||_{\mathrm{TV}}\,.\end{aligned} \tag{3.29}$$

Following [30] we show in the sequel that

$$\mathbb{E}\{||\nu_{i\to u(i)} - \widetilde{\nu}_{i\to u(i)}||_{\mathrm{TV}}\} \le c_2\delta(R-r)\,, \tag{3.30}$$

for some finite $c_2 = c_2(|\mathcal{X}|, \Delta, \kappa, \delta_*)$ and all $i \in \partial U$. As $|\partial U| \le |U| \le |\mathsf{B}_{i_o}(r)| \le \Delta^{r+1}$, we can choose $c = c(|\mathcal{X}|, \Delta, \kappa, \delta_*)$ finite such that $1 + 2c_1^{|U|}|\partial U|c_2 \le \exp(c^r)$. Then, combining the inequalities (3.27), (3.29) and (3.30) results with

$$||\mu_U - \nu_U||_{\rm TV} \le \exp(c^r)\, \delta(R-r)\,,$$

for every $U \in \mathcal{U}$ of $\mathrm{diam}(U) \le 2r$ and $r < R-1$, which is the thesis of Theorem 3.14.

As for the proof of (3.30), fixing $i \in \partial U$ let $A = \overline{\mathsf{B}}_i(R')$ and $\nu'_{i\to j} = \mu^{(ij)}_{i|A}(\cdot|\underline{X}'_A)$ where $\underline{X}'$ of distribution $\mu^{(ij)}$ is independent of $\underline{\widetilde{X}}$. Then,

$$\begin{aligned}\mathbb{E}\{||\nu_{i\to j} - \widetilde{\nu}_{i\to j}||_{\rm TV}\} &= \mathbb{E}\{||\,\mathbb{E}\nu'_{i\to j} - \widetilde{\nu}_{i\to j}||_{\rm TV}\}\\ &\le \mathbb{E}\{||\nu'_{i\to j} - \widetilde{\nu}_{i\to j}||_{\rm TV}\}\,. \end{aligned} \tag{3.31}$$

Further, setting $U' = \mathsf{B}_i(R')$ note that $G_{U'}$ is a tree (since G is R-tree like), such that $\partial U' \subseteq A$ (while ∂i and A are disjoint). Thus, from (3.26) we have that for any $j \in \partial i$,

$$\begin{aligned}||\nu'_{i\to j} - \widetilde{\nu}_{i\to j}||_{\rm TV} &= ||\mu^{(ij)}_{i|A}(\,\cdot\,|\underline{X}'_A) - \mu^{(ij)}_{i|A}(\,\cdot\,|\underline{\widetilde{X}}_A)||_{\rm TV}\\ &\le b\,||\mu_{ij|A}(\,\cdot\,|\underline{X}'_A) - \mu_{ij|A}(\,\cdot\,|\underline{\widetilde{X}}_A)||_{\rm TV}\,. \end{aligned} \tag{3.32}$$

Taking the expectation with respect to the independent random configurations $\underline{X}'$ (of law $\mu^{(ij)}$) and $\underline{\widetilde{X}}$ (of law μ), leads to

$$\begin{aligned}&\mathbb{E}\{||\mu_{ij|A}(\,\cdot\,|\underline{X}'_A) - \mu_{ij|A}(\,\cdot\,|\underline{\widetilde{X}}_A)||_{\rm TV}\}\\ &\qquad\le 2||\mu_{\{ij\},A} - \mu_{\{ij\}}\mu_A||_{\rm TV} + ||\mu^{(ij)}_A - \mu_A||_{\rm TV}\,.\end{aligned}$$

For μ extremal of valid rate function $\delta(\cdot)$ the latter expression is, due to Lemma 3.15, bounded by $(2+K)\delta(R'-1) \le (2+K)\delta(R-r)/\delta_*$, which together with (3.31) and (3.32) results with (3.30). □

4. Colorings of random graphs

Given a graph $G = (V, E)$, recall that a proper q-coloring of G is an assignment of colors to the vertices of G such that no edge has both end-points of the same color. Deciding whether a graph is q-colorable is a classical NP-complete constraint satisfaction problem. Here we shall study this problem

when G is sparse and random. More precisely, we shall consider the uniform measure $\mu_G(\cdot)$ over proper q-colorings of G, with $q \geq 3$.

As the average degree of G increases, the measure $\mu_G(\cdot)$ undergoes several phase transitions and exhibits coexistence when the average degree is within a certain interval. Eventually, for any q, if the average degree is large enough, a random graph becomes, with high probability, non q-colorable. Statistical physicists have put forward a series of exact conjectures on these phase transitions [57, 58, 78], but as of now most of it can not be rigorously verified (c.f. [1, 4, 5] for what has been proved so far).

We begin in Section 4.1 with an overview of the various phase transitions as they emerge from the statistical mechanics picture. Some bounds on the q-colorability of a random graph are proved in Section 4.2. Finally, Section 4.3 explores the nature of the coexistence threshold for q-coloring, in particular, connecting it with the question of information reconstruction, to which Section 5 is devoted.

4.1. The phase diagram: a broad picture

Let $\underline{x} = \{x_i : i \in V\}$ denote a q-coloring of the graph $G = (V, E)$ (i.e. for each vertex i, let $x_i \in \{1, \ldots, q\} \equiv \mathcal{X}_q$). Assuming that the graph G admits a proper q-coloring, the uniform measure over the set of proper q-colorings of G is

$$\mu_G(\underline{x}) = \frac{1}{Z_G} \prod_{(i,j)\in E} \mathbb{I}(x_i \neq x_j)\,, \tag{4.1}$$

with Z_G denoting the number of proper q-colorings of G. We shall consider the following two examples of a random graph $G = G_n$ over the vertex set $V = [n]$:

(a). $G = G_{n,\alpha}$ is uniformly chosen from the Erdös-Renyi ensemble $\mathbb{G}(\alpha, n)$ of graphs of $m = \lfloor n\alpha \rfloor$ edges (hence of average degree 2α).
(b). $G = G_{n,k}$ is a uniformly chosen random k-regular graph.

Heuristic statistical mechanics studies suggest a rich phase transition structure for the measure $\mu_G(\cdot)$. For any $q \geq 4$, different regimes are separated by three distinct critical values of the average degree: $0 < \alpha_{\rm d}(q) < \alpha_{\rm c}(q) < \alpha_{\rm s}(q)$ (the case $q = 3$ is special in that $\alpha_{\rm d}(q) = \alpha_{\rm c}(q)$, whereas $q = 2$ is rather trivial, as 2-colorability is equivalent to having no odd cycles, in which case each connected component of G admits two proper colorings, independently of the coloring of the rest of G). In order to characterize such

phase transitions we will use two notions (apart from colorability), namely coexistence and *sphericity*. To define the latter notion we recall that the joint type of two color assignments $\underline{x} = \{x_i : i \in V\}$ and $\underline{y} = \{y_i : i \in V\}$ is a $q \times q$ matrix whose x, y entry (for $x, y \in \{1, \ldots, q\}$) is the fraction of vertices with color x in the first assignment and color y in the second.

Definition 4.1. *Let $\nu = \{\nu(x, y)\}_{x,y\in[q]}$ be the joint type of two independent color assignments, each distributed according to $\mu_G(\,\cdot\,)$, with $\overline{\nu}(x, y) = 1/q^2$ denoting the uniform joint type. We say that μ_G is* (ε, δ)-spherical *if $||\nu - \overline{\nu}||_2 \le \varepsilon$ with probability at least $1 - \delta$.*

The various regimes of $\mu_G(\cdot)$ are characterized as follows (where all statements are to hold with respect to the uniform choice of $G \in \mathbb{G}(\alpha, n)$ with probability approaching one as $n \to \infty$):

I. For $\alpha < \alpha_{\rm d}(q)$ the set of proper q-colorings forms a unique compact lump: there is no coexistence. Further, $\mu_G(\,\cdot\,)$ is with high probability (ε, δ)-spherical for any $\varepsilon, \delta > 0$.

II. For $\alpha_{\rm d}(q) < \alpha < \alpha_{\rm c}(q)$ the measure μ_G exhibits coexistence in the sense of Section 1.1.2. More precisely, there exist $\epsilon > 0$, $C > 0$ and for each n a partition of the space of configurations $\mathcal{X}_q^n$ into $\mathcal{N} = \mathcal{N}_n$ sets $\{\Omega_{\ell,n}\}$ such that for any n and $1 \le \ell \le \mathcal{N}$,

$$\frac{\mu_G(\partial_\epsilon \Omega_{\ell,n})}{\mu_G(\Omega_{\ell,n})} \le e^{-C\,n}\,.$$

Furthermore, there exists $\Sigma = \Sigma(\alpha) > 0$, called *complexity* or *configurational entropy* and a subfamily $\mathsf{Typ} = \mathsf{Typ}_n$ of the partition $\{\Omega_{\ell,n}\}_{\ell\in\mathsf{Typ}}$ such that

$$\sum_{\ell\in\mathsf{Typ}} \mu_G(\Omega_{\ell,n}) \ge 1 - e^{-C'n}\,,$$

for some $C' > 0$ independent of n and

$$e^{-n\Sigma-o(n)} \le \inf_{\ell\in\mathsf{Typ}} \mu_G(\Omega_{\ell,n}) \le \sup_{\ell\in\mathsf{Typ}} \mu_G(\Omega_{\ell,n}) \le e^{-n\Sigma+o(n)}$$

so in particular, $|\mathsf{Typ}_n| = e^{n\Sigma+o(n)}$.

III. For $\alpha_{\rm c}(q) < \alpha < \alpha_{\rm s}(q)$ the situation is analogous to the last one, but now $\mathcal{N}_n$ is sub-exponential in n. More precisely, for any $\delta > 0$, a fraction $1 - \delta$ of the measure μ_G is comprised of $\mathcal{N}(\delta)$ elements of the partition, whereby $\mathcal{N}(\delta)$ converges as $n \to \infty$ to a finite random variable. Furthermore, $\mu_G(\,\cdot\,)$ is no longer spherical.

IV. For $\alpha_{\rm s}(q) < \alpha$ the random graph G_n is, with high probability, uncolorable (i.e. non q-colorable).

Statistical mechanics methods provide semi-explicit expressions for the threshold values $\alpha_{\rm d}(q)$, $\alpha_{\rm c}(q)$ and $\alpha_{\rm s}(q)$ in terms of the solution of a certain identity whose argument is a probability measure on the $(q-1)$-dimensional simplex.

4.2. The COL-UNCOL transition

Though the existence of a colorable-uncolorable transition is not yet established, q-colorability is a monotone graph property (i.e. if G is q-colorable, so is any subgraph of G). As such, Friedgut's theory [2, 3] provides the first step in this direction. Namely,

Theorem 4.2. *Suppose the random graph $G_{n,\alpha}$ is uniformly chosen from the Erdös-Renyi graph ensemble $\mathbb{G}(\alpha, n)$. Then, for any $q \geq 3$ there exists $\alpha_{\rm s}(q; n)$ such that for any $\delta > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\{G_{n,\alpha_{\rm s}(q;n)(1-\delta)} \text{ is } q\text{-colorable}\} = 1\,, \tag{4.2}$$

$$\lim_{n\to\infty} \mathbb{P}\{G_{n,\alpha_{\rm s}(q;n)(1+\delta)} \text{ is } q\text{-colorable}\} = 0\,. \tag{4.3}$$

We start with a simple upper bound on the COL-UNCOL transition threshold.

Proposition 4.3. *The COL-UNCOL threshold is upper bounded as*

$$\alpha_{\rm s}(q; n) \leq \overline{\alpha}_{\rm s}(q) \equiv \frac{\log q}{\log(1 - 1/q)}\,. \tag{4.4}$$

Proof. A q-coloring is a partition of the vertex set $[n]$ into q subsets of sizes n_x, $x \in \mathcal{X}_q$. Given a q-coloring, the probability that a uniformly chosen edge has both end-points of the same color is

$$\sum_{x\in\mathcal{X}_q} \binom{n_x}{2} \Big/ \binom{n}{2} \geq \frac{1}{q} - \frac{2}{n-1}\,.$$

Consequently, choosing first the q-coloring and then choosing uniformly the m edges to be included in $G = G_{n,\alpha}$ we find that the expected number of proper q-colorings for our graph ensemble is bounded by

$$\mathbb{E}\{Z_G\} \leq q^n \Big(\frac{n+1}{n-1} - \frac{1}{q}\Big)^m\,.$$

Since $\mathbb{E}\{Z_G\} \to 0$ for $\alpha > \overline{\alpha}_{\rm s}(q)$ our thesis follows from Markov's inequality. □

Notice that $\overline{\alpha}_{\rm s}(q) = q \log q[1 + o(1)]$ as $q \to \infty$. This asymptotic behavior is known to be tight, for it is shown in [4] that

Theorem 4.4. *The COL-UNCOL threshold is lower bounded as*

$$\alpha_{\rm s}(q; n) \geq \underline{\alpha}_{\rm s}(q) \equiv (q-1)\log(q-1)\,. \tag{4.5}$$

Sketch of proof. Let Z denote the number of *balanced* q-colorings, namely q-colorings having exactly n/q vertices of each color. A computation similar to the one we used when proving Proposition 4.3 yields the value of $\mathbb{E}Z$. It captures enough of $\mathbb{E}Z_G$ to potentially yield a tight lower bound on $\alpha_{\rm s}(q)$ by the second moment method, namely, using the bound $\mathbb{P}(Z_G > 0) \geq \mathbb{P}(Z > 0) \geq (\mathbb{E}Z)^2/\mathbb{E}Z^2$. The crux of the matter is of course to control the second moment of Z, for which we defer to [4]. □

The proof of Theorem 4.4 is non-constructive. In particular, it does not suggest a way of efficiently finding a q-coloring when α is near $\alpha_{\rm s}(q; n)$ (and as of now, it is not even clear if this is possible). In contrast, we provide next a simple, 'algorithmic' (though sub-optimal), lower bound on $\alpha_{\rm s}(q; n)$. To this end, recall that the k-core of a graph G is the largest induced subgraph of G having minimal degree at least k.

Proposition 4.5. *If G does not have a non-empty q-core then it is q-colorable.*

Proof. Given a graph G and a vertex i, denote by $G \setminus \{i\}$ the graph obtained by removing vertex i and all edges incident to it. If G does not contain a q-core, then we can sequentially remove vertices of degree less than q (and the edges incident to them), one at a time, until we have decimated the whole graph. This simple 'peeling algorithm' provides an ordering $i(1), i(2), \ldots, i(n)$ of the vertices, such that setting $G_0 = G$ and $G_t = G_{t-1} \setminus \{i(t)\}$, we have that for any $t \leq n$, the degree of $i(t)$ in G_{t-1} is smaller than q. Our thesis follows from the observation that if $G \setminus \{i\}$ is q-colorable, and i has degree smaller than q, then G is q-colorable as well. □

As mentioned before, this proof outlines an efficient algorithm for constructing a q-coloring for any graph G whose q-core is empty, and in principle, also for enumerating in this case the number of q-colorings of G. The threshold for the appearance of a q-core in a random Erdös-Renyi graph chosen uniformly from $\mathbb{G}(\alpha, n)$ was first determined in [80].

Proposition 4.6. *Let* $h_\alpha(u) = \mathbb{P}\{\mathsf{Poisson}(2\alpha u) \geq q-1\}$*, and define (for* $q \geq 3$*)*

$$\alpha_{\rm core}(q) = \sup\{\alpha \geq 0 :\, h_\alpha(u) \leq u \quad \forall u \in [0,1]\}\,. \tag{4.6}$$

Then, with high probability, a uniformly random graph G *from* $\mathbb{G}(\alpha, n)$ *has a* q*-core if* $\alpha > \alpha_{\rm core}(q)$*, and does not have one if* $\alpha < \alpha_{\rm core}(q)$*.*

Sketch of proof. Starting the peeling algorithm at such graph $G_0 = G_{n,\alpha}$ yields an inhomogeneous Markov chain $t \mapsto G_t$ which is well approximated by a chain of reduced state space $\mathbb{Z}_+^q$ and smooth transition kernel. The asymptotic behavior of such chains is in turn governed by the solution of a corresponding ODE, out of which we thus deduce the stated asymptotic of the probability that a uniformly random graph G from $\mathbb{G}(\alpha, n)$ has a q-core. We shall not detail this approach here, as we do so in Section 6.4 for the closely related problem of finding the threshold for the appearance of a 2-core in a uniformly random hypergraph. □

We note in passing that the value of $\alpha_{\rm core}(q)$ can be a-priori predicted by the following elegant heuristic 'cavity' argument. For a vertex $i \in V$ we call 'q-core induced by i' the largest induced subgraph having minimum degree at least q except possibly at i. We denote by u the probability that for a uniformly chosen random edge (i,j), its end-point i belongs to the q-core induced by j. Recall that for large n the degree Δ of the uniformly chosen vertex i of $G_{n,\alpha}$, excluding the distinguished edge (i,j), is approximately a $\mathsf{Poisson}(2\alpha)$ random variable. We expect each of these Δ edges to connect i to a vertex from the q-core induced by j with probability u and following the Bethe ansatz, these events should be approximately independent of each other. Hence, under these assumptions the vertex i is in the q-core induced by j with probability $h_\alpha(u)$, leading to the self-consistency equation $u = h_\alpha(u)$. The threshold $\alpha_{\rm core}(q)$ then corresponds to the appearance of a positive solution of this equation.

4.3. Coexistence and clustering: the physicist's approach

For $\alpha < \alpha_{\rm s}(q)$, the measure $\mu_G(\,\cdot\,)$ is well defined but can have a highly non-trivial structure, as discussed in Section 4.1. We describe next the physicists conjecture for the corresponding threshold $\alpha_{\rm d}(q)$ and the associated complexity function $\Sigma(\alpha)$. For the sake of simplicity, we shall write the explicit formulae in case of random $(k+1)$-regular ensembles instead of the Erdös-Renyi ensembles $\mathbb{G}(\alpha, n)$ we use in our overview.

4.3.1. Clustering and reconstruction thresholds: a conjecture

Following [66], the conjectured value for $\alpha_{\rm d}(q)$ has a particularly elegant interpretation in terms of a phase transition for a model on the rooted Galton-Watson tree $\mathsf{T} = \mathsf{T}(P, \infty)$ with offspring distribution $P = \mathsf{Poisson}(2\alpha)$. With an abuse of notation, let μ also denote the free boundary Gibbs measure over proper q-colorings of T (recall that every tree is 2-colorable). More explicitly, a proper q-coloring $\underline{x} = \{x_i \in \mathcal{X}_q : i \in \mathsf{T}\}$ is sampled from μ as follows. First sample the root color uniformly at random. Then, recursively, for each colored node i, sample the colors of its offspring uniformly at random among the colors that are different from x_i.

We denote by $\emptyset$ the root of T and by $\overline{\mathsf{B}}_\emptyset(t)$ the set of vertices of T whose distance from the root is at least t. Finally, for any subset of vertices U, we let $\mu_U(\,\cdot\,)$ be the marginal law of the corresponding color assignments.

For small α the color at the root de-correlates from colors in $\overline{\mathsf{B}}_\emptyset(t)$ when t is large, whereas at large α they remain correlated at any distance t. The 'reconstruction threshold' separates these two regimes.

Definition 4.7. *The reconstruction threshold $\alpha_{\rm r}(q)$ is the maximal value of α such that*

$$\lim_{t\to\infty} \mathbb{E}\{\, ||\mu_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu_\emptyset \times \mu_{\overline{\mathsf{B}}_\emptyset(t)}||_{\rm TV}\,\} = 0 \tag{4.7}$$

(where the expectation is over the random tree T). If the limit on the left-hand side is positive, we say that the reconstruction problem is solvable.

It is conjectured that the coexistence threshold $\alpha_{\rm d}(q)$ for locally tree like random graphs coincides with the reconstruction threshold $\alpha_{\rm r}(q)$ for the corresponding random trees. We next present a statistical physics argument in favor of this conjecture. There are various non-equivalent versions of this argument, all predicting the same location for the threshold. The argument that we will reproduce was first developed in [21, 40, 69], to explore the physics of glasses and spin glasses.

Note that the major difficulty in trying to identify the existence of 'lumps' is that we do not know, a priori, where these lumps are in the space of configurations. However, if $\underline{X}^*$ is a configuration sampled from $\mu(\,\cdot\,)$, it will fall inside one such lump so the idea is to study how a second configuration $\underline{x}$ behaves when tilted towards the first one. Specifically, fix $\underline{x}^* = \{x_i^* \in \mathcal{X}_q : i \in V\}$ and consider the tilted measures

$$\mu^*_{G,\underline{x}^*,\epsilon}(\underline{x}) = \frac{1}{Z_\epsilon} \prod_{(i,j)\in E} \mathbb{I}(x_i \neq x_j) \prod_{i\in V} \psi_\epsilon(x_i^*, x_i)\,,$$

where $\psi_\epsilon(x,y)$ is a tilting function depending continuously on ϵ, such that $\psi_0(x,y)=1$ (so μ_0^* reduces to the uniform measure over proper colorings), and which favors $x=y$ when $\epsilon>0$. For instance, we might take

$$\psi_\epsilon(x,y)=\exp\left\{\epsilon\,\mathbb{I}(x=y)\right\}.$$

While the study of the measure $\mu^*_{G,\underline{x}^*,\epsilon}$ is beyond our current means, we gain valuable insight from examining its Bethe approximation. Specifically, in this setting messages depend in addition to the graph also on $\underline{x}^*$ and ϵ, and the Bethe equations of Definition 3.4 are

$$\nu_{i\to j}(x_i)=z_{i\to j}^{-1}\psi_\epsilon(x_i^*,x_i)\prod_{l\in\partial i\setminus j}\left(1-\nu_{l\to i}(x_i)\right), \tag{4.8}$$

with $z_{i\to j}$ a normalization constant. In shorthand we write this equation as

$$\nu_{i\to j}=\mathsf{F}_\epsilon\{\nu_{l\to i}:\,l\in\partial i\setminus j\}.$$

Let us now assume that G is a regular graph of degree $k+1$ and that $\underline{X}^*$ is a uniformly random proper q-coloring of G. Then, the message $\nu_{i\to j}$ is itself a random variable, taking values in the $(q-1)$-dimensional probability simplex $\mathcal{M}(\mathcal{X}_q)$. For each $x\in\mathcal{X}_q$ we denote by Q_x (which also depends on ϵ), the conditional law of $\nu_{i\to j}$ given that $X_i^*=x$. In formulae, for any Borel measurable subset A of $\mathcal{M}(\mathcal{X}_q)$, we have

$$Q_x(A)\equiv\mathbb{P}\left\{\nu_{i\to j}(\,\cdot\,)\in A\big|X_i^*=x\right\}.$$

Assume that, conditionally on the reference coloring $\underline{X}^*$, the messages $\nu_{l\to i}$ for $l\in\partial i\setminus j$ are asymptotically independent, and have the laws $Q_{X_i^*}$. We then obtain the following recursion for $\{Q_x\}$,

$$Q_x(A)=\sum_{x_1\ldots x_k}\mu(x_1,\ldots,x_k|x)\int\mathbb{I}(\mathsf{F}_\epsilon(\nu_1,\ldots,\nu_k)\in A)\prod_{i=1}^k Q_{x_i}(\mathrm{d}\nu_i),$$

where $(x_1,\ldots,x_k)$ denote the values of $(X_l^*,\,l\in\partial i\setminus j)$ and $\mu(x_1,\ldots,x_k|x)$ the corresponding conditional marginal of $\mu=\mu_0^*$ given $X_i^*=x$. Assuming further that for a random regular graph $G=G_{n,k+1}$ the measure $\mu(x_1,\ldots,x_k|x)$ converges as $n\to\infty$ to the analogous conditional law for the regular k-ary tree, we obtain the fixed point equation

$$Q_x(A)=\frac{1}{(q-1)^k}\sum_{x_1\ldots x_k\neq x}\int\mathbb{I}(\mathsf{F}_\epsilon(\nu_1,\ldots,\nu_k)\in A)\prod_{i=1}^k Q_{x_i}(\mathrm{d}\nu_i). \tag{4.9}$$

In the limit $\epsilon = 0$ this equation admits a trivial degenerate solution, whereby $Q_x = \delta_{\overline{\nu}}$ is concentrated on one point, the uniform vector $\overline{\nu}(x) = 1/q$ for all $x \in \mathcal{X}_q$. The interpretation of this solution is that, as $\epsilon \downarrow 0$, a random coloring from the tilted measure $\mu^*_{G,\underline{X}^*,\epsilon}$, becomes uncorrelated from the reference coloring $\underline{X}^*$.

It is not hard to verify that this is the only degenerate solution (namely, where each measure Q_x is supported on one point), of (4.9) at $\epsilon = 0$. A second scenario is however possible. It might be that, as $\epsilon \downarrow 0$ (and, in particular, for $\epsilon = 0$), Eq. (4.9) admits also a non-trivial solution, whereby at least one of the measures Q_x is not supported on the uniform vector $\overline{\nu}$. This is interpreted by physicists as implying coexistence: the coloring sampled from the tilted measure $\mu^*_{G,\underline{X}^*,\epsilon}$ remains trapped in the same 'state' (i.e. in the same subset of configurations $\Omega_{\ell,n}$), as $\underline{X}^*$.

Let us summarize the statistical physics conjecture: the uniform measure $\mu_G(\,\cdot\,)$ over proper q-colorings of a random $(k+1)$-regular graph exhibits coexistence if and only if Eq. (4.9) admits a non-trivial solution for $\epsilon = 0$. In the next subsection we show that this happens if and only if $k \geq k_{\rm r}(q)$, with $k_{\rm r}(q)$ the reconstructibility threshold on k-ary trees (which is defined analogously to the Poisson tree threshold $\alpha_{\rm r}(q)$, see Definition 4.7).

4.3.2. The reconstruction threshold for k-ary trees

We say that a probability measure on $\mathcal{M}(\mathcal{X}_q)$ is *color-symmetric* if it is invariant under the action of color permutations on its argument $\nu \in \mathcal{M}(\mathcal{X}_q)$. Following [66, Proposition 1] we proceed to show that the existence of certain non-trivial solutions $\{Q_x\}$ of (4.9) at $\epsilon = 0$ is equivalent to solvability of the corresponding reconstruction problem for k-ary trees.

Proposition 4.8. *The reconstruction problem is solvable on k-ary trees if and only if Eq. (4.9) admits at $\epsilon = 0$ a solution $\{Q_x, x \in \mathcal{X}_q\}$ such that each Q_x has the Radon-Nikodym density $q\nu(x)$ with respect to the same color-symmetric, non-degenerate probability measure Q.*

Proof. First notice that $\{Q_x, x \in \mathcal{X}_q\}$ is a solution of (4.9) at $\epsilon = 0$ if and only if for any $x \in \mathcal{X}_q$ and bounded Borel function g,

$$\int g(\nu) q^{-1} Q_x(\mathrm{d}\nu) = c_{q,k} \int g(\mathsf{F}_0(\nu_1, \ldots, \nu_k)) \prod_{i=1}^{k} [Q_* - q^{-1} Q_x](\mathrm{d}\nu_i)\,, \quad (4.10)$$

where $Q_* = q^{-1} \sum_{x=1}^{q} Q_x$ and $c_{q,k} = q^{k-1}(q-1)^{-k}$. If this solution is of the stated form, then $Q_* = Q$ and upon plugging $Q_x(\mathrm{d}\nu) = q\nu(x)Q(\mathrm{d}\nu)$ in the

identity (4.10) we see that for any bounded Borel function h,

$$\int h(\nu)Q(\mathrm{d}\nu) = \int \left[\frac{z(\nu_1,\ldots,\nu_k)}{z(\overline{\nu},\ldots,\overline{\nu})}\right] h(\mathsf{F}_0(\nu_1,\ldots,\nu_k)) \prod_{i=1}^{k} Q(\mathrm{d}\nu_i)\,, \quad (4.11)$$

where $z(\nu_1,\ldots,\nu_k) = \sum_{x=1}^{q}\prod_{i=1}^{k}(1-\nu_i(x))$ is the normalization constant of the mapping $\mathsf{F}_0(\cdot)$ (so $c_{q,k} = 1/z(\overline{\nu},\ldots,\overline{\nu})$). Conversely, for any color-symmetric probability measure Q on $\mathcal{M}(\mathcal{X}_q)$ the value of $\int \nu(x)Q(\mathrm{d}\nu)$ is independent of $x \in \mathcal{X}_q$, hence $Q_x(\mathrm{d}\nu) = q\nu(x)Q(\mathrm{d}\nu)$ are then also probability measures on $\mathcal{M}(\mathcal{X}_q)$ and such that $Q = Q_*$. Further, recall that for any $x \in \mathcal{X}_q$ and $\nu_i \in \mathcal{M}(\mathcal{X}_q)$,

$$z(\nu_1,\ldots,\nu_k)\mathsf{F}_0(\nu_1,\ldots,\nu_k)(x) = \prod_{i=1}^{k}(1-\nu_i(x))\,,$$

so if such Q satisfies (4.11), then considering there $h(\nu) = g(\nu)\nu(x)$ leads to $\{Q_x\}$ satisfying (4.10).

If a solution Q of (4.11) is degenerate, i.e. supported on one point ν, then $\nu = \mathsf{F}_0(\nu,\ldots,\nu)$, hence $\nu = \overline{\nu}$. That is, any non-trivial solution $Q \neq \delta_{\overline{\nu}}$ is also non-degenerate. We thus proceed to show that solvability of the reconstruction problem on k-ary trees is equivalent to having color-symmetric solution $Q \neq \delta_{\overline{\nu}}$ of (4.11). To this end, consider a proper q-coloring $X = \{X_v : v \in \mathsf{T}\}$ of the k-ary tree, sampled at random according to the free boundary Gibbs measure μ. Let $\nu^{(t)}$ denote the marginal distribution of the root color given the colors at generation t. In formulae, this is the $\mathcal{M}(\mathcal{X}_q)$-valued random variable such that for $x \in \{1,\ldots,q\}$,

$$\nu^{(t)}(x) = \mu_{\emptyset|\overline{\mathsf{B}}_\emptyset(t)}(x|X_{\overline{\mathsf{B}}_\emptyset(t)}) = \mathbb{P}\{X_\emptyset = x|X_{\overline{\mathsf{B}}_\emptyset(t)}\}\,.$$

Denote by $Q_x^{(t)}$ the conditional law of $\nu^{(t)}$ given the root value $X_\emptyset = x$. The k-ary tree of $(t+1)$ generations is the merging at the root of k disjoint k-ary trees, each of which has t generations. Thus, conditioning on the colors $x_1,\ldots,x_k$ of the root's offspring, one finds that the probability measures $\{Q_x^{(t)}\}$ satisfy for any x and any bounded Borel function $h(\cdot)$ the recursion

$$\int h(\nu)Q_x^{(t+1)}(\mathrm{d}\nu) = \frac{1}{(q-1)^k}\sum_{x_1,\ldots,x_k\neq x}\int h(\mathsf{F}_0(\nu_1,\ldots,\nu_k)) \prod_{i=1}^{k} Q_{x_i}^{(t)}(\mathrm{d}\nu_i)\,,$$

starting at $Q_x^{(0)} = \delta_{\nu_x}$, where ν_x denotes the probability vector that puts weight one on the color x.

Let $Q^{(t)}$ denote the unconditional law of $\nu^{(t)}$. That is, $Q^{(t)} = q^{-1}\sum_{x=1}^q Q_x^{(t)}$. By the tower property of the conditional expectation, for any $x \in \mathcal{X}_q$ and bounded measurable function h on $\mathcal{M}(\mathcal{X}_q)$,

$$\begin{aligned}\int h(\nu)Q_x^{(t)}(\mathrm{d}\nu) &= q\mathbb{E}[h(\nu^{(t)})\mathbb{I}(X_\emptyset = x)] \\ &= q\mathbb{E}[h(\nu^{(t)})\nu^{(t)}(x)] = q\int \nu(x)h(\nu)Q^{(t)}(\mathrm{d}\nu)\,.\end{aligned}$$

Consequently, $Q_x^{(t)}$ has the Radon-Nikodym derivative $q\nu(x)$ with respect to $Q^{(t)}$. Plugging this into the recursion for $Q_x^{(t)}$ we find that $Q^{(t)}$ satisfies the recursion relation

$$\int h(\nu)Q^{(t+1)}(\mathrm{d}\nu) = \int \left[\frac{z(\nu_1,\ldots,\nu_k)}{z(\overline{\nu},\ldots,\overline{\nu})}\right] h(\mathsf{F}_0(\nu_1,\ldots,\nu_k))\prod_{i=1}^k Q^{(t)}(\mathrm{d}\nu_i)\,, \tag{4.12}$$

starting at $Q^{(0)} = q^{-1}\sum_{x=1}^q \delta_{\nu_x}$.

Note that for each $x \in \mathcal{X}_q$, the sequence $\{\nu^{(t)}(x)\}$ is a reversed martingale with respect to the filtration $\mathcal{F}_{-t} = \sigma(X_{\overline{\mathsf{B}}_\emptyset(t)})$, $t \geq 0$, hence by Lévy's downward theorem, it has an almost sure limit. Consequently, the probability measures $\{Q^{(t)}\}$ converge weakly to a limit $Q^{(\infty)}$.

As $Q^{(0)}$ is color-symmetric and the recursion (4.12) transfers the color-symmetry of $Q^{(t)}$ to that of $Q^{(t+1)}$, we deduce that $Q^{(\infty)}$ is also color-symmetric. Further, with the function $\mathsf{F}_0 : \mathcal{M}(\mathcal{X}_q)^k \to \mathcal{M}(\mathcal{X}_q)$ continuous at any point $(\nu_1,\ldots,\nu_k)$ for which $z(\nu_1,\ldots,\nu_k) > 0$, it follows from the recursion (4.12) that $Q^{(\infty)}$ satisfies (4.11) for any continuous h, hence for any bounded Borel function h. By definition,

$$||\mu_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu_\emptyset \times \mu_{\overline{\mathsf{B}}_\emptyset(t)}||_{\mathrm{TV}} = \int ||\nu - \overline{\nu}||_{\mathrm{TV}}\, Q^{(t)}(\mathrm{d}\nu)\,,$$

and with $\mathcal{X}_q$ finite, the function $\nu \mapsto ||\nu - \overline{\nu}||_{\mathrm{TV}}$ is continuous. Hence, the reconstruction problem is solvable if and only if $Q^{(\infty)} \neq \delta_{\overline{\nu}}$. That is, as claimed, solvability implies the existence of a non-trivial color-symmetric solution $Q^{(\infty)}$ of (4.11).

To prove the converse assume there exists a color-symmetric solution $Q \neq \delta_{\overline{\nu}}$ of Eq. (4.11). Recall that in this case $Q_x(\mathrm{d}\nu) = q\nu(x)Q(\mathrm{d}\nu)$ are probability measures such that $Q = q^{-1}\sum_{x=1}^q Q_x$. Further, if a random variable $Y(t)$ is conditionally independent of $X_\emptyset$ given $X_{\overline{\mathsf{B}}_\emptyset(t)}$ then

$$\begin{aligned}||\mu_{\emptyset,Y(t)} - \mu_\emptyset \times \mu_{Y(t)}||_{\mathrm{TV}} &\leq ||\mu_{\emptyset,Y(t),\overline{\mathsf{B}}_{\emptyset(t)}} - \mu_\emptyset \times \mu_{Y(t),\overline{\mathsf{B}}_{\emptyset(t)}}||_{\mathrm{TV}} \\ &= ||\mu_{\emptyset,\overline{\mathsf{B}}_{\emptyset(t)}} - \mu_\emptyset \times \mu_{\overline{\mathsf{B}}_{\emptyset(t)}}||_{\mathrm{TV}}\end{aligned}$$

(where $\mu_{\emptyset,Y(t)}$ denotes the joint law of $X_\emptyset$ and $Y(t)$). Turning to construct such a random variable $Y(t) \in \mathcal{M}(\mathcal{X}_q)$, let $\partial \mathsf{B}_\emptyset(t)$ denote the vertices of the tree at distance t from $\emptyset$ and set $\nu_i \in \mathcal{M}(\mathcal{X}_q)$ for $i \in \partial \mathsf{B}_\emptyset(t)\}$ to be conditionally independent given $X_{\overline{\mathsf{B}}_\emptyset(t)}$, with ν_i distributed according to the random measure $Q_{X_i}(\,\cdot\,)$. Then, define recursively $\nu_v \equiv \mathsf{F}_0(\nu_{u_1}, \ldots, \nu_{u_k})$ for $v \in \partial \mathsf{B}_\emptyset(s)$, $s = t-1, t-2, \ldots, 0$, where $u_1, \ldots, u_k$ denote the offspring of v in T. Finally, set $Y(t) = \nu_\emptyset$.

Under this construction, the law $P_{v,x}$ of ν_v conditional upon $X_v = x$ is Q_{X_v}, for any $v \in \partial \mathsf{B}_\emptyset(s)$, $s = t, \ldots, 0$. Indeed, clearly this is the case for $s = t$ and proceeding recursively, assume it applies at levels $t, \ldots, s+1$. Then, as $\{Q_x, x \in \mathcal{X}_q\}$ satisfy (4.10), we see that for $v \in \partial \mathsf{B}_\emptyset(s)$ of offspring $u_1, \ldots, u_k$, any $x \in \mathcal{X}_q$ and bounded Borel function $g(\cdot)$,

$$\begin{aligned}
\int g(\nu) P_{v,x}(\mathrm{d}\nu) &= \mathbb{E}[\int g(\mathsf{F}_0(\nu_1, \ldots, \nu_k)) \prod_{i=1}^k Q_{X_{u_i}}(\mathrm{d}\nu_i) | X_v = x] \\
&= q c_{q,k} \int g(\mathsf{F}_0(\nu_1, \ldots, \nu_k)) \prod_{i=1}^k [Q - q^{-1} Q_x](\mathrm{d}\nu_i) \\
&= \int g(\nu) Q_x(\mathrm{d}\nu)\,.
\end{aligned}$$

That is, $P_{v,x} = Q_x$, as claimed. In particular, $\mu_{Y(t)|\emptyset} = Q_{X_\emptyset}$, $\mu_{Y(t)} = Q$ and with $Q_x(\mathrm{d}\nu) = q\nu(x)Q(\mathrm{d}\nu)$, it follows that

$$||\mu_{\emptyset,Y(t)} - \mu_\emptyset \times \mu_{Y(t)}||_{\mathrm{TV}} = \frac{1}{q} \sum_{x=1}^q ||Q_x - Q||_{\mathrm{TV}} = \int ||\nu - \overline{\nu}||_{\mathrm{TV}}\, Q(\mathrm{d}\nu)\,,$$

which is independent of t and strictly positive (since $Q \neq \delta_{\overline{\nu}}$). By the preceding inequality, this is a sufficient condition for reconstructibility. □

4.3.3. Complexity: exponential growth of the number of clusters

We provide next a heuristic derivation of the predicted value of the complexity parameter $\Sigma = \Sigma(k)$ for proper q-colorings of a uniformly chosen random regular graph $G = G_{n,k+1}$, as defined in Section 4.1, regime II, namely, when $k_d(q) < k < k_c(q)$. This parameter is interpreted as the exponential growth rate of the number of 'typical' lumps or 'clusters' to which the uniform measure $\mu_G(\,\cdot\,)$ decomposes. Remarkably, we obtain an expression for $\Sigma(k)$ in terms of the non-degenerate solution of (4.9) at $\epsilon = 0$.

Recall Definition 3.6 that the Bethe free entropy for proper q-colorings of G and a given (permissive) message set $\{\nu_{i\to j}\}$ is

$$\Phi\{\nu_{i\to j}\} = -\sum_{(i,j)\in E}\log\{1-\sum_{x=1}^{q}\nu_{i\to j}(x)\nu_{j\to i}(x)\} + \sum_{i\in V}\log\Big\{\sum_{x=1}^{q}\prod_{j\in\partial i}(1-\nu_{j\to i}(x))\Big\}. \tag{4.13}$$

According to the Bethe-Peierls approximation, the logarithm of the number Z_n of proper q-colorings for $G = G_{n,k+1}$ is approximated for large n by the value of $\Phi\{\nu_{i\to j}\}$ for a message set $\{\nu_{i\to j}\}$ which solves the Bethe-Peierls equations (4.8) at $\epsilon = 0$. One trivial solution of these equations is $\nu_{i\to j} = \overline{\nu}$ (the uniform distribution over $\{1,\ldots,q\}$), and for $G = G_{n,k+1}$ the corresponding Bethe free entropy is

$$\begin{aligned}\Phi(\overline{\nu}) &= n\Big\{-\frac{k+1}{2}\log\{1-\sum_{x=1}^{q}\overline{\nu}(x)^2\}+\log\{\sum_{x=1}^{q}(1-\overline{\nu}(x))^{k+1}\}\Big\}\\ &= n[\log q+\frac{k+1}{2}\log(1-1/q)]\,.\end{aligned} \tag{4.14}$$

As explained before, when $k_{\rm d}(q) < k < k_{\rm s}(q)$, upon fixing n large enough, a regular graph G_n of degree $k+1$ and a reference proper q-coloring $\underline{x}^*$ of its vertices, we expect Eq. (4.8) to admit a second solution $\{\nu^*_{i\to j}\}$ for all $\epsilon > 0$ small enough. In the limit $\epsilon \downarrow 0$, this solution is conjectured to describe the uniform measure over proper q-colorings in the cluster $\Omega_{\ell,n}$ containing $\underline{x}^*$. In other words, the restricted measure

$$\mu_{\ell,n}(\underline{x}) = \mu_{G_n}(\underline{x}|\Omega_{\ell,n}) = \frac{1}{Z_{\ell,n}}\prod_{(i,j)\in E}\mathbb{I}(x_i\neq x_j)\,\mathbb{I}(\underline{x}\in\Omega_{\ell,n})\,, \tag{4.15}$$

is conjectured to be Bethe approximated by such message set $\{\nu^*_{i\to j}\}$. One naturally expects the corresponding free entropy approximation to hold as well. That is, to have

$$\log Z_{\ell,n} = \Phi\{\nu^*_{i\to j}\} + o(n)\,.$$

As discussed in Section 4.1, in regime II, namely, for $k_{\rm d}(q) < k < k_{\rm c}(q)$, it is conjectured that for uniformly chosen proper q-coloring $\underline{X}^*$, the value of $n^{-1}\log Z_{\ell,n}$ (for the cluster $\Omega_{\ell,n}$ containing $\underline{X}^*$), concentrates in probability as $n\to\infty$, around a non-random value. Recall that $\log Z_n = \Phi(\overline{\nu}) + o(n)$, so

with most of the Z_n proper q-colorings of G_n comprised within the $e^{n\Sigma+o(n)}$ 'typical' clusters $\Omega_{\ell,n}$, $\ell \in \mathsf{Typ}_n$, each having $Z_{\ell,n}$ proper q-colorings, we conclude that

$$\begin{aligned}\Phi(\overline{\nu}) = \log Z_n + o(n) &= \log\Big\{\sum_{\ell=1}^{|\mathsf{Typ}_n|} Z_{\ell,n}\Big\} + o(n)\\ &= n\Sigma + \mathbb{E}[\Phi\{\nu^*_{i\to j}\}] + o(n)\,, \end{aligned} \tag{4.16}$$

where the latter expectation is with respect to both the random graph G_n and the reference configuration $\underline{X}^*$ (which together determine the message set $\{\nu^*_{i\to j}\}$).

This argument provides a way to compute the exponential growth rate $\Sigma(k)$ of the number of clusters, as

$$\Sigma(k) = \lim_{n\to\infty} n^{-1}\Big\{\Phi(\overline{\nu}) - \mathbb{E}[\Phi\{\nu^*_{i\to j}\}]\Big\}\,.$$

For a uniformly chosen random proper q-coloring $\underline{X}^*$, the distribution of $\{\nu^*_{i\to j}\}$ can be expressed in the $n\to\infty$ limit in terms of the corresponding solution $\{Q_x\}$ of the fixed point equation (4.9) at $\epsilon = 0$. Specifically, following the Bethe ansatz, we expect that for uniformly chosen $i \in [n]$, the law of $\{\nu^*_{j\to i}, j \in \partial i\}$ conditional on $\{X^*_i, X^*_j, j \in \partial i\}$ converges as $n\to\infty$ to the product measure $\prod_{j=1}^{k+1} Q_{X^*_j}$ and the law of $\{\nu^*_{i\to j}, \nu^*_{j\to i}\}$ conditional on X^*_i and X^*_j converges to the product measure $Q_{X^*_i}\times Q_{X^*_j}$. By the invariance of the uniform measure over proper q-colorings to permutations of the colors, for any edge (i,j) of $G_{n,k+1}$, the pair (X^*_i, X^*_j) is uniformly distributed over the $q(q-1)$ choices of $x_i \neq x_j$ in $\mathcal{X}_q^2$. Moreover, for large n the Bethe approximation predicts that $\{X^*_i, X^*_j, j\in\partial i\}$ is nearly uniformly distributed over the $q(q-1)^{k+1}$ choices of $x_j \in \mathcal{X}_q$, all of which are different from $x_i \in \mathcal{X}_q$. We thus conclude that

$$\begin{aligned}\Sigma(k) = &-\frac{k+1}{2}\frac{1}{q(q-1)}\sum_{x_1\neq x_2}\int W_{\mathrm{e}}(\nu_1,\nu_2)\,Q_{x_1}(\mathrm{d}\nu_1)\,Q_{x_2}(\mathrm{d}\nu_2)\\ &+\frac{1}{q(q-1)^{k+1}}\sum_{x=1}^{q}\sum_{x_j\neq x}\int W_{\mathrm{v}}(\nu_1,\ldots,\nu_{k+1})\prod_{j=1}^{k+1}Q_{x_j}(\mathrm{d}\nu_j)\,,\end{aligned} \tag{4.17}$$

where

$$W_{\mathrm{e}}(\nu_1,\nu_2) = \log\Big\{\frac{1-\sum_{x=1}^q \nu_1(x)\nu_2(x)}{1-1/q}\Big\}\,, \tag{4.18}$$

$$W_{\mathrm{v}}(\nu_1,\ldots,\nu_{k+1}) = \log\Big\{\frac{1}{q}\sum_{x=1}^{q}\prod_{j=1}^{k+1}\frac{1-\nu_j(x)}{1-1/q}\Big\}\,. \tag{4.19}$$

5. Reconstruction and extremality

As shown in Section 3.4, the Bethe-Peierls approximation applies for permissive graph-specification pairs $(G, \underline{\psi})$ such that:

(a). The graph $G = (V, E)$ has large girth (and it often suffices for G to merely have a large girth in the neighborhood of most vertices).
(b). The dependence between the random vectors $\underline{x}_A$ and $\underline{x}_B$ is weak for subsets A and B which are far apart on G (indeed, we argued there that 'extremality' is the appropriate notion for this property).

While these conditions suffice for Bethe-Peierls approximation to hold on general graphs with bounded degree, one wishes to verify them for specific models on sparse random graphs. For condition (a) this can be done by standard random graph techniques (c.f. Section 2.1), but checking condition (b) is quite an intricate task. Thus, largely based on [43], we explore here the extremality condition in the context of random sparse graphs.

Beyond the relevance of extremality for the Bethe-Peierls approximation, it is interesting *per se* and can be rephrased in terms of the *reconstruction problem*. In Section 4.3.1 we considered the latter in case of proper q-colorings, where it amounts to estimating the color of a distinguished (root) vertex $\emptyset \in V$ for a uniformly chosen proper coloring $\underline{X}$ of the given graph $G = (V, E)$, when the colors $\{X_j, j \in U\}$ on a subset U of vertices are revealed. In particular, we want to understand whether revealing the colors at large distance t from the root, induces a non-negligible bias on the distribution of $X_\emptyset$.

It turns out that, for a random Erdös-Renyi graph chosen uniformly from the ensemble $\mathbb{G}(\alpha, n)$, there exists a critical value $\alpha_{\rm r}(q)$, such that reconstruction is possible (in the sense of Definition 4.7), when the number of edges per vertex $\alpha > \alpha_{\rm r}(q)$, and impossible when $\alpha < \alpha_{\rm r}(q)$. Recall from Section 4.3, that the reconstruction threshold $\alpha_{\rm r}(q)$ is conjectured to coincide with the so-called 'clustering' threshold $\alpha_{\rm d}(q)$. That is, the uniform measure over proper q-colorings of these random graphs should exhibit coexistence if and only if $\alpha_{\rm r}(q) = \alpha_{\rm d}(q) \le \alpha < \alpha_{\rm c}(q)$. As we will show, this relation provides a precise determination of the clustering threshold.

More generally, consider a graph-specification pair $(G, \underline{\psi})$, with a distinguished marked vertex $\emptyset \in V$ (which we call hereafter the 'root' of G), and a sample $\underline{X}$ from the associated graphical model $\mu_{G,\underline{\psi}}(\underline{x})$ of (1.4). The reconstructibility question asks whether 'far away' variables $\underline{X}_{\overline{\mathsf{B}}_\emptyset(t)}$ provide non-negligible information about $X_\emptyset$ (here $\overline{\mathsf{B}}_\emptyset(t)$ denotes the subset of vertices $i \in V$ at distance $d(\emptyset, i) \ge t$ from the root). This is quantified by the

following definition, where as usual, for $U \subseteq V$ we denote the corresponding marginal distribution of $\underline{X}_U = \{X_j : j \in U\}$ by $\mu_U(\underline{x}_U)$.

Definition 5.1. *The reconstruction problem is* (t,ε)-solvable *(also called,* (t,ε)-reconstructible*), for the graphical model associated with* $(G,\underline{\psi})$ *and rooted at* $\emptyset \in V$*, if*

$$\|\mu_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu_\emptyset \times \mu_{\overline{\mathsf{B}}_\emptyset(t)}\|_{\mathrm{TV}} \geq \varepsilon\,. \tag{5.1}$$

We say that the reconstruction problem is solvable *(*reconstructible*), for a given sequence* $\{G_n\}$ *of random graphs (and specified joint distributions of the graph* G_n*, the specification* $\underline{\psi}$ *on it, and the choice of* $\emptyset \in V_n$*), if for some* $\varepsilon > 0$ *and all* $t \geq 0$*, the events* $A_n(t)$ *that the reconstruction problem is* (t,ε)*-solvable on* G_n*, occur with positive probability. That is, when* $\inf_t \limsup_{n\to\infty} \mathbb{P}\{A_n(t)\} > 0$.

Remark 5.2. *The inequality (5.1) fails when the connected component of* $\emptyset$ *in* G*, has diameter less than* t*. Hence, for sparse random graphs* G_n*, the sequence* $n \mapsto \mathbb{P}\{A_n(t)\}$ *is often bounded away from one (on account of* $\emptyset$ *possibly being in a small connected component).*

The rationale for this definition is that the total variation distance on the left hand side of Eq. (5.1) measures the information about $X_\emptyset$ that the variables in $\overline{\mathsf{B}}_\emptyset(t)$ provide. For instance, it is proportional to the difference between the probability of correctly guessing $X_\emptyset$ when knowing $X_{\overline{\mathsf{B}}_\emptyset(t)}$, and the a-priori probability of doing so without knowing $X_{\overline{\mathsf{B}}_\emptyset(t)}$.

Note that non-reconstructibility is slightly weaker than the extremality condition of Section 3.4. Indeed, we require here a decay of the correlations between a vertex $\emptyset$ and an arbitrary subset of vertices at distance t from it, whereas in Definition 3.13, we require such decay for *arbitrary* subsets of vertices A and B of distance t apart. However, it is not hard to check that when proving Theorem 3.14 we only consider the extremality condition in cases where the size of the subset B does not grow with R (or with the size of G) and for graph sequences that converge locally to trees, this is in turn implied by a non-reconstructibility type condition, where $B = \{\emptyset\}$ is a single vertex.

Recall Section 2.1 that for a uniformly chosen root $\emptyset$ and locally tree-like sparse random graphs G_n, for any $t \geq 0$ fixed, the finite neighborhood $\mathsf{B}_\emptyset(t)$ converges in distribution to a (typically random) tree. We expect that with high probability the vertices on the corresponding boundary set $\partial\mathsf{B}_\emptyset(t) = \{i \in \mathsf{B}_\emptyset(t) : \partial i \not\subseteq \mathsf{B}_\emptyset(t)\}$, are 'far apart' from each other in the complementary subgraph $\overline{\mathsf{B}}_\emptyset(t)$. This suggests that for the graphical model

on $\overline{\mathsf{B}}_\emptyset(t)$, the variables $\{X_j, j \in \partial\mathsf{B}_\emptyset(t)\}$ are then weakly dependent, and so approximating G_n by its limiting tree structure might be a good way to resolve the reconstruction problem. In other words, one should expect reconstructibility on G_n to be determined by reconstructibility on the associated limiting random tree.

Beware that the preceding argument is circular, for we assumed that variables on 'far apart' vertices (with respect to the residual graph $\overline{\mathsf{B}}_\emptyset(t)$), are weakly dependent, in order to deduce the same for variables on vertices that are 'far apart' in G_n. Indeed, its conclusion fails for many graphical models. For example, [43] shows that the tree and graph reconstruction thresholds do not coincide in the simplest example one can think of, namely, ferromagnetic Ising models.

On the positive side, we show in the sequel that the tree and graph reconstruction problems are equivalent under the sphericity condition of Definition 4.1 (we phrased this definition in terms proper colorings, but it applies verbatim to general graphical models). More precisely, if for any $\varepsilon, \delta > 0$, the canonical measure $\mu(\,\cdot\,)$ is (ε, δ)-spherical with high probability (with respect to the graph distribution), then the graph and tree reconstructions do coincide. It can indeed be shown that, under the sphericity condition, sampling $\underline{X}$ according to the graphical model on the residual graph $\overline{\mathsf{B}}_\emptyset(t)$, results with $\{X_j, j \in \partial\mathsf{B}_\emptyset(t)\}$ which are approximately independent.

This sufficient condition was applied in [43] to the Ising spin glass (where sphericity can be shown to hold as a consequence of a recent result by Guerra and Toninelli [51]). More recently, [74] deals with proper colorings of random graphs (building on the work of Achlioptas and Naor, in [4]). For a family of graphical models parametrized by their average degree, it is natural to expect reconstructibility to hold at large average degrees (as the graph is 'more connected'), but not at small average degrees (since the graph 'falls' apart into disconnected components). We are indeed able to establish a threshold behavior (i.e. a critical degree value above which reconstruction is solvable) both for spin glasses and for proper colorings.

5.1. Applications and related work

Beyond its relation with the Bethe-Peierls approximation, the reconstruction problem is connected to a number of other interesting problems, two of which we briefly survey next.

Markov Chain Monte Carlo (MCMC) algorithms provide a well established way of approximating marginals of the distribution $\mu = \mu_{G,\underline{\psi}}$ of (1.4). The idea is to define an (irreducible and aperiodic) Markov chain whose

unique stationary distribution is $\mu(\cdot)$, so if this chain converges rapidly to its stationary state (i.e., its mixing time is small), then it can be effectively used to generate a sample $\underline{X}$ from $\mu(\cdot)$.

In many interesting cases, the chain is reversible and consists of local updates (i.e. consecutive states differ in only few variables, with transition probabilities determined by the restriction of the state to a neighborhood in G of the latter set). Under these conditions, the mixing time is known to be related to the correlation decay properties of the stationary distribution $\mu(\,\cdot\,)$ (see, [35, 52]). With

$$\Delta(t;\underline{x}) \equiv \|\mu_{\emptyset|\overline{\mathsf{B}}_\emptyset(t)}(\,\cdot\,|\underline{x}_{\overline{\mathsf{B}}_\emptyset(t)}) - \mu_\emptyset(\,\cdot\,)\|_{\mathrm{TV}}\,, \tag{5.2}$$

one usually requires in this context that the dependence between $x_\emptyset$ and $\underline{x}_{\overline{\mathsf{B}}_\emptyset(t)}$ decays uniformly, i.e. $\sup_{\underline{x}} \Delta(t;\underline{x}) \to 0$ as $t \to \infty$. On graphs with sub-exponential growth, a fast enough (uniform) decay is necessary and sufficient for fast mixing. However, for more general graphs, this *uniform* decay is often a too strong requirement, which one might opt to replace by the weaker assumption of non-reconstructibility (indeed, the inequality (5.1) can be re-written as $\mathbb{E}[\Delta(t;\underline{X})] \geq \varepsilon$, where the expectation is with respect to the random sample $\underline{X}$).

In this direction, it was shown in [13] that non-reconstructibility is a necessary condition for fast mixing. Though the converse may in general fail, non-reconstructibility is sufficient for rapid decay of the variance of local functions (which in physics is often regarded as the criterion for fast dynamics, see [75]). Further, for certain graphical models on trees, [13] shows that non-reconstructibility is equivalent to polynomial spectral gap, a result that is sharpened in [64] to the equivalence between non-reconstructibility and fast mixing (for these models on trees).

Random constraint satisfaction problems. Given an instance of a constraint satisfaction problem (CSP), consider the uniform distribution over its solutions. As we have seen in Section 1.2.2, it takes the form (1.23), which is an immediate generalization of (1.4).

Computing the marginal $\mu_\emptyset(x_\emptyset)$ is useful both for finding a solution and the number of solutions of such a CSP. Suppose we can generate *only one* uniformly random solution $\underline{X}$. In general this is not enough for approximating the law of $X_\emptyset$ in a meaningful way, but one can try the following: First, fix all variables 'far from $\emptyset$' to take the same value as in the sampled configuration, namely $X_{\overline{\mathsf{B}}_\emptyset(t)}$. Then, compute the conditional distribution at $\emptyset$ (which for locally tree-like graphs can be done efficiently via dynamic programming). While the resulting distribution is in general not a good approx-

imation of $\mu_\emptyset(\cdot)$, non-reconstructibility implies that it is, with high probability within total variation distance ε of $\mu_\emptyset(\cdot)$. That is, non-reconstructibility yields a good approximation of $\mu_\emptyset(x_\emptyset)$ based on a single sample (namely, a single uniformly random solution $\underline{X}$). The situation is even simpler under the assumptions of our main theorem (Theorem 5.4), where the boundary condition $\underline{X}_{\overline{\mathsf{B}}_\emptyset(t)}$ may be replaced by an i.i.d. uniform boundary condition.

We have explained in Section 4 why for a typical sparse random graph of large average degree one should expect the set of proper colorings to form well-separated 'clusters'. The same rationale should apply, at high constraint density, for the solutions of a typical instance of a CSP based on large, sparse random graphs (c.f. [6, 67, 71]). This in turn increases the computational complexity of sampling even one uniformly random solution.

Suppose the set of solutions partitions into clusters and any two solutions that differ on at most $n\varepsilon$ vertices, are in the same cluster. Then, knowing the value of all 'far away' variables $\underline{X}_{\overline{\mathsf{B}}_\emptyset(t)}$ determines the cluster to which the sample $\underline{X}$ belongs, which in turn provides some information on $X_\emptyset$. The preceding heuristic argument connects reconstructibility to the appearance of well-separated solution clusters, a connection that has been studied for example in [58, 65].

Reconstruction problems also emerge in a variety of other contexts: (*i*) Phylogeny (where given some evolved genomes, one aims at reconstructing the genome of their common ancestor, c.f. [28]); (*ii*) Network tomography (where given end-to-end delays in a computer network, one aims to infer the link delays in its interior, c.f. [14]); (*iii*) Gibbs measures theory (c.f. [16, 42]).

Reconstruction on trees: A brief survey.

The reconstruction problem is relatively well understood in case the graph is a tree (see [76]). The fundamental reason for this is that then the canonical measure $\mu(\underline{x})$ admits a simple description. More precisely, to sample $\underline{X}$ from $\mu(\cdot)$, first sample the value of $X_\emptyset$ according to the marginal law $\mu_\emptyset(x_\emptyset)$, then recursively for each node j, sample its children $\{X_\ell\}$ *independently* conditional on their parent value.

Because of this Markov structure, one can derive a recursive distributional equation for the conditional marginal at the root $\nu^{(t)}(\,\cdot\,) \equiv \mu_{\emptyset|\overline{\mathsf{B}}_\emptyset(t)}(\,\cdot\,|\underline{X}_{\overline{\mathsf{B}}_\emptyset(t)})$ given the variable values at generation t (just as we have done in the course of proving Proposition 4.8). Note that $\nu^{(t)}(\,\cdot\,)$ is a random quantity even for a deterministic graph G_n (because $\underline{X}_{\overline{\mathsf{B}}_\emptyset(t)}$ is itself drawn randomly from the distribution $\mu(\,\cdot\,)$). Further, it contains all the information in the boundary about $X_\emptyset$ (i.e. it is a 'sufficient statistic'), so the standard approach to tree reconstruction is to study the asymptotic behavior of the distributional

recursion for $\nu^{(t)}(\,\cdot\,)$.

Indeed, following this approach, reconstructibility has been thoroughly characterized for zero magnetic field Ising models on generic trees (c.f. [16, 19, 38]). More precisely, for such model on an infinite tree T of branching number $\mathrm{br}(\mathsf{T})$, the reconstruction problem is solvable if and only if $\mathrm{br}(\mathsf{T})(\tanh\beta)^2 > 1$. For the cases we treat in the sequel, $\mathrm{br}(\mathsf{T})$ coincides with the mean offspring number of any vertex, hence this result establishes a sharp reconstruction threshold in terms of the average degree (or in terms of the inverse temperature parameter β), that we shall generalize here to random graphs.

Reconstruction on general graphs poses new challenges, since it lacks such recursive description of sampling from the measure $\mu(\,\cdot\,)$. The result of [13] allows for deducing non-reconstructibility from fast mixing of certain reversible Markov chains with local updates. However, proving such fast mixing is far from being an easy task, and in general the converse does not hold (i.e. one can have slow mixing and non-reconstructibility).

A threshold λ_{r} for fast mixing has been established in [77] for the independent set model of (3.20), in case G_n are random bipartite graphs. Arguing as in [43], it can be shown that this is also the graph reconstruction threshold. An analogous result was proved in [43] for the ferromagnetic Ising model and random regular graphs (and it extends also to Poisson random graphs, see [31]). In all of these cases, the graph reconstruction threshold does not coincide with the tree reconstruction threshold, but coincides instead with the tree 'uniqueness threshold' (i.e. the critical parameter such that the uniform decorrelation condition $\sup_{\underline{x}} \Delta(t;\underline{x}) \to 0$ holds).

5.2. Reconstruction on graphs: sphericity and tree-solvability

For the sake of clarity, we focus hereafter on *Poisson* graphical models. Specifying such an ensemble requires an alphabet $\mathcal{X}$, a *density parameter* $\gamma \geq 0$, a finite collection of non-negative, symmetric functionals $\psi_a(\,\cdot\,,\,\cdot\,)$ on $\mathcal{X}\times\mathcal{X}$, indexed by $a \in \mathcal{C}$, and a probability distribution $\{p(a) : a \in \mathcal{C}\}$ on $\mathcal{C}$. In the random multi-graph G_n the multiplicities of edges between pairs of vertices $i \neq j \in [n]$ are independent Poisson$(2\gamma/n)$ random variables, and G_n has additional independent Poisson(γ/n) self-loops at each vertex $i \in [n]$. For each occurrence of an edge $e = \{e_1, e_2\}$ in G_n (including its self-loops), we draw an independent random variable $A_e \in \mathcal{C}$ according to the distribution $\{p(\,\cdot\,)\}$ and consider the graphical model of specification $\underline{\psi} \equiv \{\psi_{A_e}(x_{e_1}, x_{e_2}) : e \in G_n\}$. Finally, the root $\emptyset$ is uniformly chosen in $[n]$, independently of the graph-specification pair $(G_n, \underline{\psi})$.

For example, the uniform measure over proper q-colorings fits this framework (simply take $\mathcal{X} = \mathcal{X}_q$ and $|\mathcal{C}| = 1$ with $\psi(x, y) = \mathbb{I}(x \neq y)$).

It is easy to couple the multi-graph G_n of the Poisson model and the Erdös-Renyi random graph from the ensemble $\mathbb{G}(\gamma, n)$ such that the two graphs differ in at most $\Delta_n = \sum_{1 \le i \le j \le n} Y_{\{i,j\}}$ edges, where the independent variables $Y_{\{i,j\}}$ have the Poisson(γ/n) distribution when $i = j$ and that of (Poisson$(2\gamma/n)$-1)$_+$ when $i \neq j$. It is not hard to check that $\Delta_n/(\log n)$ is almost surely uniformly bounded, and hence by Proposition 2.6, almost surely the Poisson multi-graphs $\{G_n\}$ are uniformly sparse and converge locally to the rooted at $\emptyset$, Galton-Watson tree T of Poisson(2γ) offspring distribution. Let $\mathsf{T}(\ell)$, $\ell \ge 0$ denote the graph-specification pair on the first ℓ generations of T, where each edge carries the specification $\psi_a(\,\cdot\,,\,\cdot\,)$ with probability $p(a)$, independently of all other edges and of the realization of T.

It is then natural to ask whether reconstructibility of the original graphical models is related to reconstructibility of the graphical models $\mu^{\mathsf{T}(\ell)}(\underline{x})$ per Eq. (1.4) for $G = \mathsf{T}(\ell)$ and the same specification $\underline{\psi}$.

Definition 5.3. *Consider a sequence of random graphical models $\{G_n\}$ converging locally to the random rooted tree T. We say that the reconstruction problem is* tree-solvable *for the sequence $\{G_n\}$ if it is solvable for $\{\mathsf{T}(\ell)\}$. That is, there exists $\varepsilon > 0$ such that, as $\ell \to \infty$, for any $t \ge 0$,*

$$\|\mu^{\mathsf{T}(\ell)}_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu^{\mathsf{T}(\ell)}_{\emptyset} \times \mu^{\mathsf{T}(\ell)}_{\overline{\mathsf{B}}_\emptyset(t)}\|_{\mathrm{TV}} \ge \varepsilon\,, \tag{5.3}$$

with positive probability.

This definition could have been expressed directly in terms of the free boundary Gibbs measure μ^{T} on the infinite rooted tree T. Indeed, the reconstruction problem is tree-solvable if and only if with positive probability

$$\liminf_{t\to\infty} \|\mu^{\mathsf{T}}_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu^{\mathsf{T}}_{\emptyset} \times \mu^{\mathsf{T}}_{\overline{\mathsf{B}}_\emptyset(t)}\|_{\mathrm{TV}} > 0\,.$$

While Eqs. (5.3) and (5.1) are similar, as explained before, passing from the original graph to the tree is a significant simplification (due to the recursive description of sampling from $\mu^{\mathsf{T}(\ell)}(\,\cdot\,)$).

We proceed with a sufficient condition for graph-reconstruction to be equivalent to tree reconstruction. To this end, we introduce the concept of 'two-replicas type' as follows. Consider a graphical model G and two i.i.d. samples $\underline{X}^{(1)}$, $\underline{X}^{(2)}$ from the corresponding canonical measure $\mu(\,\cdot\,) = \mu_{(G,\underline{\psi})}(\cdot)$ (we will call them *replicas* following the spin glass terminology).

The *two replica type* is a matrix $\{\nu(x,y) : x,y \in \mathcal{X}\}$ where $\nu(x,y)$ counts the fraction of vertices j such that $X_j^{(1)} = x$ and $X_j^{(2)} = y$. We denote by $\mathcal{R}$ the set of distributions ν on $\mathcal{X}\times\mathcal{X}$ and by $\mathcal{R}_n$ the subset of *valid two-replicas types*, that is, distributions ν with $n\nu(x,y) \in \mathbb{N}$ for all $x,y \in \mathcal{X}$.

The matrix $\nu = \nu_n$ is a random variable, because the graph G_n is random, and the two replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$ are i.i.d. conditional on G_n. If $\mu(\,\cdot\,)$ was the uniform distribution, then ν_n would concentrate (for large n), around $\overline{\nu}(x,y) \equiv 1/|\mathcal{X}|^2$. Our sufficient condition requires this to be approximately true.

Theorem 5.4. *Consider a sequence of random Poisson graphical models $\{G_n\}$. Let $\nu_n(\,\cdot\,,\,\cdot\,)$ be the type of two i.i.d. replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$, and $\Delta\nu_n(x,y) \equiv \nu_n(x,y) - \overline{\nu}(x,y)$. Assume that, for any $x \in \mathcal{X}$,*

$$\lim_{n\to\infty} \mathbb{E}\Big\{ \big[\Delta\nu_n(x,x) - 2|\mathcal{X}|^{-1}\sum_{x'} \Delta\nu_n(x,x')\big]^2 \Big\} = 0\,. \tag{5.4}$$

Then, the reconstruction problem for $\{G_n\}$ is solvable if and only if it is tree-solvable.

Remark 5.5. *The expectation in Eq. (5.4) is with respect to the two replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$ (which the type $\nu_n(\,\cdot\,,\,\cdot\,)$ is a function of), conditional on G_n, as well as with respect to G_n. Explicitly,*

$$\mathbb{E}\{F(\underline{X}^{(1)},\underline{X}^{(2)})\} = \mathbb{E}\Big\{ \sum_{\underline{x}^{(1)},\underline{x}^{(2)}} \mu_{G_n}(\underline{x}^{(1)})\mu_{G_n}(\underline{x}^{(2)})\, F(\underline{x}^{(1)},\underline{x}^{(2)}) \Big\}\,. \tag{5.5}$$

Remark 5.6. *It is easy to see that the sphericity condition of Definition 4.1 implies Eq. (5.4). That is, (5.4) holds if μ_{G_n} are (ε,δ_n)-spherical for any $\varepsilon > 0$ and some $\delta_n(\varepsilon) \to 0$.*

Remark 5.7. *In fact, as is hinted by the proof, condition (5.4) can be weakened, e.g. $\overline{\nu}(\,\cdot\,\cdot\,)$ can be chosen more generally than the uniform matrix. Such a generalization amounts to assuming that 'replica symmetry is not broken' (in the spin glass terminology, see [65]). For the sake of simplicity we omit such generalizations.*

Condition (5.4) emerges naturally in a variety of contexts, a notable one being second moment method applied to random constraint satisfaction problems. As an example, consider proper colorings of random graphs, cf. Section 4. The second moment method was used in [5] to bound from below the colorability threshold. The reconstruction threshold on trees was estimated in [15, 83]. Building on these results, and as outlined at the end of Section 5.3 the following statement is obtained in [74].

Theorem 5.8. *For proper q-colorings of a Poisson random graph of density γ, the reconstruction problem is solvable if and only if $\gamma > \gamma_{\rm r}(q)$, where for large q,*

$$\gamma_{\rm r}(q) = \frac{1}{2}\, q\, [\log q + \log\log q + o(1)]\,. \tag{5.6}$$

In general the graph and tree reconstruction thresholds do not coincide. For example, as mentioned before, zero magnetic field ferromagnetic Ising models on the Galton-Watson tree $\mathsf{T}(P, \rho, \infty)$ (of Section 2), are solvable if and only if $\overline{\rho}(\tanh(\beta))^2 > 1$. The situation changes dramatically for graphs, as shown in [31, 43].

Theorem 5.9. *For both Poisson random graphs and random regular graphs, reconstruction is solvable for zero magnetic field, ferromagnetic Ising models, if and only if $\overline{\rho}\tanh(\beta) > 1$.*

In physicists' language, the ferromagnetic phase transition occurring at $\overline{\rho}\tanh(\beta) = 1$, cf. Section 2, 'drives' the reconstruction threshold. The proof of reconstructibility for $\overline{\rho}\tanh(\beta) > 1$ essentially amounts to finding a bottleneck in Glauber dynamics. As a consequence it immediately implies that the mixing time is exponential in this regime. We expect this to be a tight estimate of the threshold for exponential mixing.

On the other hand, for a zero magnetic field, Ising spin-glass, the tree and graph thresholds do coincide. In fact, for such a model on a Galton-Watson tree with Poisson(2γ) offspring distribution, reconstruction is solvable if and only if $2\gamma(\tanh(\beta))^2 > 1$ (see, [38]). The corresponding graph result is:

Theorem 5.10. *Reconstruction is solvable for Ising spin-glasses of zero magnetic field, on Poisson random graph of density parameter γ, provided $2\gamma(\tanh(\beta))^2 > 1$, and it is unsolvable if $2\gamma(\tanh(\beta))^2 < 1$.*

5.3. Proof of main results

Hereafter, let $\mathsf{B}_i(t) = \{j \in [n] : d(i,j) \le t\}$, $\overline{\mathsf{B}}_i(t) = \{j \in [n] : d(i,j) \ge t\}$ and $\mathsf{D}_i(t) \equiv \mathsf{B}_i(t) \cap \overline{\mathsf{B}}_i(t)$ (i.e. the set of vertices of distance t from i). Further, partition the edges of G_n between the subgraphs $\mathsf{B}_i(t)$ and $\overline{\mathsf{B}}_i(t)$ so edges between two vertices from $\mathsf{D}_i(t)$ are all in $\overline{\mathsf{B}}_i(t)$, and excluded from $\mathsf{B}_i(t)$.

Beyond the almost sure convergence of the law of $\mathsf{B}_\emptyset(t)$ to the corresponding Galton-Watson tree of depth-t, rooted at $\emptyset$ (which as explained before, is a consequence of Proposition 2.6), the proof of Theorem 5.4 relies on the following form of independence between $\mathsf{B}_\emptyset(t)$ and $\overline{\mathsf{B}}_\emptyset(t)$ for Poisson random graphs.

Proposition 5.11. *Let G_n be a Poisson random graph on vertex set $[n]$ and density parameter γ. Then, conditional on $\mathsf{B}_\emptyset(t)$, $\overline{\mathsf{B}}_\emptyset(t)$ is a Poisson random graph on vertex set $[n] \setminus \mathsf{B}_\emptyset(t-1)$ with same edge distribution as G_n.*

Proof. Condition on $\mathsf{B}_\emptyset(t) = \mathsf{G}(t)$, and let $\mathsf{G}(t-1) = \mathsf{B}_\emptyset(t-1)$ (notice that this is uniquely determined from $\mathsf{G}(t)$). This is equivalent to conditioning on a given edge realization between the vertices k, l such that $k \in \mathsf{G}(t-1)$ and $l \in \mathsf{G}(t)$.

The graph $\overline{\mathsf{B}}_\emptyset(t)$ has as vertices the set $[n] \setminus \mathsf{G}(t)$ and its edges are those $(k,l) \in G_n$ such that $k, l \notin \mathsf{G}(t-1)$. Since the latter set of edges is disjoint from the one we are conditioning upon, the claim follows by the independence of the choice of edges taken into G_n. □

We also need to bound the tail of the distribution of the number of vertices in the depth-t neighborhood of $\emptyset$. This can be done by comparison with a Galton-Watson process.

Proposition 5.12. *Let $\|\mathsf{B}_\emptyset(t)\|$ denote the number of edges (counting their multiplicities), in depth-t neighborhood of the root in a Poisson random graph G_n of density γ. Then, for any $\lambda > 0$ there exists finite $g_t(\lambda, \gamma)$ such that, for any n, $M \geq 0$*

$$\mathbb{P}\{\|\mathsf{B}_\emptyset(t)\| \geq M\} \leq g_t(\lambda,\gamma)\, \lambda^{-M}\,. \tag{5.7}$$

Proof. Notice that, because of the symmetry of the graph distribution under permutation of the vertices, we can and shall fix $\emptyset$ to be a deterministic vertex. Starting at $\emptyset$ we explore G_n in breadth-first fashion and consider the sequence of random variables $E_t = \|\mathsf{B}_\emptyset(t)\|$. Then, for each $t \geq 0$, the value of $E_{t+1} - E_t$ is, conditional on $\mathsf{B}_\emptyset(t)$, upper bounded by the sum of $|\mathsf{D}_\emptyset(t)| \times |\overline{\mathsf{B}}_\emptyset(t)|$ i.i.d. Poisson$(2\gamma/n)$ random variables. Since $|\overline{\mathsf{B}}_\emptyset(t)| \leq n$ and $|\mathsf{D}_\emptyset(t)| \leq E_t - E_{t-1}$ for $t \geq 1$ (with $|\mathsf{D}_\emptyset(0)| = 1$), it follows that E_t is stochastically dominated by $|\mathsf{T}(t)|$, where $\mathsf{T}(t)$ is a depth-t Galton-Watson tree with Poisson(2γ) offspring distribution. By Markov's inequality,

$$\mathbb{P}\{\|\mathsf{B}_\emptyset(t)\| \geq M\} \leq \mathbb{E}\{\lambda^{|\mathsf{T}(t)|}\}\, \lambda^{-M}\,.$$

To complete the proof, recall that $g_t(\lambda,\gamma) \equiv \mathbb{E}\{\lambda^{|\mathsf{T}(t)|}\}$ is the finite solution of the recursion $g_{t+1}(\lambda,\gamma) = \lambda\xi(g_t(\lambda,\gamma),\gamma)$ for $\xi(\lambda,\gamma) = e^{2\gamma(\lambda-1)}$ and $g_0(\lambda,\gamma) = \lambda$. □

In order to prove Theorem 5.4 we will first establish that, under condition (5.4), any (fixed) subset of the variables $\{X_1, \ldots, X_n\}$ is (approximately)

uniformly distributed. This is, at first sight, a surprising fact. Indeed, the condition (5.4) only provides direct control on two-variables correlations. It turns out that two-variables correlations control k-variable correlations for any bounded k because of the symmetry among $X_1, \ldots, X_n$. To clarify this point, it is convenient to take a more general point of view.

Definition 5.13. *For any distribution $\mu(\cdot)$ over $\mathcal{X}^n$ (where $\mathcal{X}$ is a generic measure space), and any permutation π over the set $\{1 \ldots, n\}$ let $\mu^\pi(\cdot)$ denote the distribution obtained acting with π on $\mathcal{X} \times \cdots \times \mathcal{X}$.*

Let $\mu(\cdot)$ be a random probability distribution over $\mathcal{X} \times \cdots \times \mathcal{X}$. We say that μ is stochastically exchangeable *if μ is distributed as μ^π for any permutation π.*

Proposition 5.14. *Suppose (5.4) holds for a finite set $\mathcal{X}$ and the type ν_n of two i.i.d. replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$ from a sequence of stochastically exchangeable random measures $\mu^{(n)}$ on $\mathcal{X}^n$. Then, for any fixed set of vertices $i(1), \ldots, i(k) \subseteq [n]$ and any $\xi_1, \ldots, \xi_k \in \mathcal{X}$, as $n \to \infty$,*

$$\mathbb{E}\Big\{ \big| \mu^{(n)}_{i(1),\ldots,i(k)}(\xi_1, \ldots, \xi_k) - |\mathcal{X}|^{-k} \big|^2 \Big\} \to 0 \,. \tag{5.8}$$

Proof. Per given replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$, we define, for any $\xi \in \mathcal{X}$ and $i \in [n]$,

$$\mathsf{Q}_i(\xi) = \Big\{ \mathbb{I}(X_i^{(1)} = \xi) - \frac{1}{|\mathcal{X}|} \Big\} \Big\{ \mathbb{I}(X_i^{(2)} = \xi) - \frac{1}{|\mathcal{X}|} \Big\}$$

and let $\mathsf{Q}(\xi) = n^{-1} \sum_{i=1}^n \mathsf{Q}_i(\xi)$ denote the average of $\mathsf{Q}_i(\xi)$ over a uniformly random $i \in [n]$. Since

$$\mathsf{Q}(\xi) = \Delta\nu_n(\xi, \xi) - |\mathcal{X}|^{-1} \sum_{x'} \Delta\nu_n(\xi, x') - |\mathcal{X}|^{-1} \sum_{x'} \Delta\nu_n(x', \xi),$$

it follows from (5.4) and the triangle inequality, that $\mathbb{E}\{\mathsf{Q}(\xi)^2\} \to 0$ as $n \to \infty$. Further, $|Q(\xi)| \le 1$, so by the Cauchy-Schwarz inequality we deduce that for any fixed, non-empty $U \subseteq [n]$, $b \in U$ and $\xi_a \in \mathcal{X}$,

$$\Big| \mathbb{E}\{ \prod_{a \in U} \mathsf{Q}(\xi_a) \} \Big| \le \mathbb{E}|\mathsf{Q}(\xi_b)| \to 0 \,.$$

Next, fixing $i(1), i(2), \ldots, i(k)$ and $U \subseteq [k]$, let

$$Y_U \equiv \mathbb{E}\Big\{ \prod_{a \in U} \left(\mathbb{I}(X_{i(a)} = \xi_a) - |\mathcal{X}|^{-1} \right) | \mu \Big\} \,,$$

where $\mathbb{E}\{\cdot\,|\mu\}$ denotes the expectation with respect to the measure $\mu(\cdot)$ of the replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$, i.e. at fixed realization of $\mu = \mu^{(n)}$. Note that by the stochastic exchangeability of μ, and since $\sup_\xi |\mathsf{Q}(\xi)| \le 1$, we have that for any non-empty $U \subseteq [k]$,

$$\mathbb{E}\{Y_U^2\} = \mathbb{E}\{\prod_{a\in U} \mathsf{Q}_{i(a)}(\xi_a)\} = \mathbb{E}\{\prod_{a\in U} \mathsf{Q}(\xi_a)\} + \Delta_{U,n}\,,$$

where $|\Delta_{U,n}|$ is upper bounded by the probability that $|U| \le k$ independent uniform in $[n]$ random variables are not distinct, which is $O(1/n)$. Thus, $\mathbb{E}\{Y_U^2\} \to 0$ as $n \to \infty$, for any fixed, non-empty $U \subseteq [k]$.

The proof of the proposition is completed by noting that $Y_\emptyset = 1$ and

$$\mu_{i(1),\ldots,i(k)}(\xi_1,\ldots,\xi_k) = \sum_{U\subseteq[k]} |\mathcal{X}|^{|U|-k} Y_U\,,$$

hence by the Cauchy-Schwarz inequality,

$$\mathbb{E}\Big\{\big|\mu_{i(1),\ldots,i(k)}(\xi_1,\ldots,\xi_k) - |\mathcal{X}|^{-k}\big|^2\Big\} \le \sum_{\emptyset\neq U,V\subseteq[k]} \mathbb{E}|Y_U Y_V| \le 2^k \sum_{\emptyset\neq U\subseteq[k]} \mathbb{E}\{Y_U^2\}$$

goes to zero as $n \to \infty$. □

The following lemma is the key for relating the solvability of the reconstruction problem to its tree-solvability.

Lemma 5.15. *For any graphical model $\mu = \mu_{G_n,\underline{\psi}}$, any vertex $\emptyset \in [n]$, and all $t \le \ell$,*

$$\Big|\|\mu_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu_\emptyset \times \mu_{\overline{\mathsf{B}}_\emptyset(t)}\|_{\rm TV} - \|\mu^{<}_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu^{<}_\emptyset \times \mu^{\le}_{\overline{\mathsf{B}}_\emptyset(t)}\|_{\rm TV}\Big| \le 5|\mathcal{X}|^{|\mathsf{B}_\emptyset(\ell)|}\,\|\mu^{>}_{\mathsf{D}_\emptyset(\ell)} - \rho_{\mathsf{D}_\emptyset(\ell)}\|_{\rm TV}\,, \tag{5.9}$$

where for any $U \subseteq [n]$, we let $\rho_U(\underline{x}_U) = 1/|\mathcal{X}|^{|U|}$ denote the uniform distribution of $\underline{x}_U$, with $\mu^{<}_U$ denoting the marginal law of $\underline{x}_U$ for the graphical model in which the edges of $\overline{\mathsf{B}}_\emptyset(\ell)$ are omitted, whereas $\mu^{>}_U$ denotes such marginal law in case all edges of $\mathsf{B}_\emptyset(\ell)$ are omitted.

Proof. Adopting hereafter the shorthands $\mathsf{B}(t)$, $\overline{\mathsf{B}}(t)$ and $\mathsf{D}(t)$ for $\mathsf{B}_\emptyset(t)$, $\overline{\mathsf{B}}_\emptyset(t)$ and $\mathsf{D}_\emptyset(t)$, respectively, recall that by the definition of these sets there are no edges in G_n between $\mathsf{B}(t)$ and $\overline{\mathsf{B}}(t) \setminus \mathsf{D}(t)$. Hence, $\Delta(t,\underline{x})$ of Eqn. (5.2) depends only on $\underline{x}_{\mathsf{D}(t)}$ and consequently,

$$\begin{aligned}\|\mu_{\emptyset,\overline{\mathsf{B}}(t)} - \mu_\emptyset \times \mu_{\overline{\mathsf{B}}(t)}\|_{\rm TV} &= \sum_{\underline{x}} \mu_{\overline{\mathsf{B}}(t)}(\underline{x}_{\overline{\mathsf{B}}(t)})\Delta(t,\underline{x}) \\ &= \sum_{\underline{x}} \mu_{\mathsf{D}(t)}(\underline{x}_{\mathsf{D}(t)})\|\mu_{\emptyset|\mathsf{D}(t)}(\,\cdot\,|\underline{x}_{\mathsf{D}(t)}) - \mu_\emptyset(\,\cdot\,)\|_{\rm TV}\,.\end{aligned}$$

By the same reasoning also

$$\|\mu^{<}_{\emptyset,\overline{\mathsf{B}}(t)} - \mu^{<}_{\emptyset} \times \mu^{\leq}_{\overline{\mathsf{B}}(t)}\|_{\mathrm{TV}} = \sum_{\underline{x}} \mu^{<}_{\mathsf{D}(t)}(\underline{x}_{\mathsf{D}(t)}) \|\mu^{<}_{\emptyset|\mathsf{D}(t)}(\,\cdot\,|\underline{x}_{\mathsf{D}(t)}) - \mu^{<}_{\emptyset}(\,\cdot\,)\|_{\mathrm{TV}}\,.$$

Since $\emptyset \in \mathsf{B}(t) \subseteq \mathsf{B}(\ell)$, the conditional law of $x_\emptyset$ given $\underline{x}_{\mathsf{D}(t)}$ is the same under the graphical model for G_n and the one in which all edges of $\overline{\mathsf{B}}(\ell)$ are omitted. Further, by definition of the total variation distance, the value of $\|\mu_U - \mu^{<}_U\|_{\mathrm{TV}}$ is non-decreasing in $U \subseteq \mathsf{B}(\ell)$. With the total variation distance bounded by one, it thus follows from the preceding identities and the triangle inequality that the left hand side of Eq. (5.9) is bounded above by

$$\|\mu_\emptyset - \mu^{<}_{\emptyset}\|_{\mathrm{TV}} + 2\|\mu_{\mathsf{D}(t)} - \mu^{<}_{\mathsf{D}(t)}\|_{\mathrm{TV}} \leq 3\|\mu_{\mathsf{B}(\ell)} - \mu^{\leq}_{\overline{\mathsf{B}}(\ell)}\|_{\mathrm{TV}}\,.$$

Next, considering the distribution $\mu_{\mathsf{B}(\ell)}(z)$ on the discrete set $\mathcal{Z} = \mathcal{X}^{\mathsf{B}(\ell)}$, notice that, as a consequence of Eq. (1.4) and of the fact that $\mathsf{B}(\ell)$ and $\overline{\mathsf{B}}(\ell)$ are edge disjoint,

$$\mu_{\mathsf{B}(\ell)}(z) = \frac{f(z)\rho_2(z)}{\sum_{z' \in \mathcal{Z}} f(z')\rho_2(z')}\,, \tag{5.10}$$

for the $[0,1]$-valued function $f = \mu^{<}_{\overline{\mathsf{B}}(\ell)}$ on $\mathcal{Z}$, and the distribution $\rho_2 = \mu^{>}_{\overline{\mathsf{B}}(\ell)}$ on this set. Clearly, replacing ρ_2 in the right hand side of (5.10) by the uniform distribution $\rho_1 = \rho$ on $\mathcal{Z}$, results with $\sum_{z' \in \mathcal{Z}} f(z')\rho_1(z') = 1/|\mathcal{Z}|$ and in the notations of (3.24), also with $\widehat{\rho}_1 = f$. We thus deduce from the latter bound that

$$\|\mu_{\mathsf{B}(\ell)} - \mu^{<}_{\overline{\mathsf{B}}(\ell)}\|_{\mathrm{TV}} \leq \frac{3}{2}|\mathcal{Z}|\|\mu^{>}_{\overline{\mathsf{B}}(\ell)} - \rho_{\mathsf{B}(\ell)}\|_{\mathrm{TV}}\,,$$

and the proof of the lemma is complete upon noting that $\mu^{>}_{\overline{\mathsf{B}}(\ell)}$ deviates from the uniform distribution only in terms of its marginal on $\mathsf{D}(\ell)$. □

Proof of Theorem 5.4. Fixing $t \leq \ell$, let Δ_n denote the left hand side of Eq. (5.9). We claim that its expectation with respect to the Poisson random model G_n vanishes as $n \to \infty$. First, with $\Delta_n \leq 1$ and $\sup_n \mathbb{P}(\|\mathsf{B}_\emptyset(\ell)\| \geq M) \to 0$ as $M \to \infty$, see Proposition 5.12, it suffices to prove that for any finite M, as $n \to \infty$,

$$\mathbb{E}\{\Delta_n \mathbb{I}(\|\mathsf{B}_\emptyset(\ell)\| < M)\} \to 0\,.$$

Decomposing this expectation according to the finitely many events $\{\mathsf{B}_\emptyset(\ell) = \mathsf{H}\}$, indexed by rooted, connected, multi-graphs H of less than M edges (counting multiplicities), we have by (5.9) that

$$\mathbb{E}\{\Delta_n \mathbb{I}(\|\mathsf{B}_\emptyset(\ell)\| < M)\} \le 5|\mathcal{X}|^M \sum_{\|\mathsf{H}\|<M} \mathbb{E}\{||\mu^{>}_{\mathsf{D}_\emptyset(\ell)} - \rho_{\mathsf{D}_\emptyset(\ell)}||_{\mathrm{TV}} \,|\, \mathsf{B}_\emptyset(\ell) = \mathsf{H}\}\,,$$

and it is enough to show that each of the terms on the right hand side vanishes as $n \to \infty$.

Recall Proposition 5.11 that each term in the sum is the expectation, with respect to a Poisson graphical model of density γ over the collection $[n] \setminus \mathsf{B}_\emptyset(\ell-1)$ of at least $n-M$ vertices. The event $\{\mathsf{B}_\emptyset(\ell) = \mathsf{H}\}$ fixes the set $\mathsf{D} = \mathsf{D}_\emptyset(\ell)$ whose finite size depends only on the rooted multi-graph H. By Proposition 5.14 we thus deduce that conditional on this event, the expected value of

$$\|\mu^{>}_{\mathsf{D}} - \rho_{\mathsf{D}}\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\underline{x}_{\mathsf{D}}} \left| \mu^{>}_{\mathsf{D}}(\underline{x}_{\mathsf{D}}) - |\mathcal{X}|^{-|\mathsf{D}|} \right|\,,$$

vanishes as $n \to \infty$. To recap, we have shown that for any $t \le \ell$, the expected value of the left hand side of Eq. (5.9) vanishes as $n \to \infty$.

In view of Definition 5.1, this implies that the reconstruction problem is solvable for $\{G_n\}$ if and only if $\inf_t \limsup_{\ell\to\infty} \limsup_{n\to\infty} \mathbb{P}\{A_n(t,\ell,\varepsilon)\} > 0$ for some $\varepsilon > 0$, where $A_n(t,\ell,\varepsilon)$ denotes the event

$$\|\mu^{<}_{\emptyset,\overline{\mathsf{B}}_\emptyset(t)} - \mu^{<}_{\emptyset} \times \mu^{\le}_{\overline{\mathsf{B}}_\emptyset(t)}\|_{\mathrm{TV}} \ge \varepsilon\,.$$

Recall that $\mu^{<}(\,\cdot\,)$ is the canonical measure for the edge-independent random specification on the random graph $\mathsf{B}_\emptyset(\ell)$ and that almost surely the uniformly sparse Poisson random graphs $\{G_n\}$ converge locally to the Galton-Watson tree T of Poisson(2γ) offspring distribution. Applying Lemma 2.16 for the uniformly bounded function $\mathbb{I}(A_n(t,\ell,\varepsilon))$ of $\mathsf{B}_\emptyset(\ell)$ and averaging first under our uniform choice of $\emptyset$ in $[n]$, we deduce that $\mathbb{P}\{A_n(t,\ell,\varepsilon)\} \to \mathbb{P}\{A_\infty(t,\ell,\varepsilon)\}$, where $A_\infty(t,\ell,\varepsilon)$ denotes the event on the left hand side of (5.3). That is, $\{G_n\}$ is solvable if and only if $\inf_t \limsup_{\ell\to\infty} \mathbb{P}\{A_\infty(t,\ell,\varepsilon)\} > 0$ for some $\varepsilon > 0$, which is precisely the definition of tree-solvability. □

Proof of Theorem 5.8. Following [74], this proof consists of four steps:

(1) It is shown in [83] that for regular trees of degree 2γ the reconstruction threshold $\gamma_{\mathrm{r,tree}}(q)$ for proper q-colorings grows with $q \to \infty$ as in (5.6). In the large γ limit considered here, a Poisson(2γ) random variable is tightly

concentrated around its mean. Hence, as noted in [83], the result (5.6) extends straightforwardly to the case of random Galton-Watson trees with offspring distribution Poisson(2γ).

(2) Given two balanced proper q-colorings $\underline{x}^{(1)}$, $\underline{x}^{(2)}$ of G_n (a q-coloring is balanced if it has exactly n/q vertices of each color), recall that their *joint type* is the q-dimensional matrix $\nu(\cdot,\cdot)$ such that $\nu(x,y)$ counts the fraction of vertices $i \in [n]$ with $x_i^{(1)} = x$ and $x_i^{(2)} = y$. Let $Z_{\rm b}(\nu)$ denote the number of balanced pairs of proper q-colorings $\underline{x}^{(1)}$, $\underline{x}^{(2)}$ of G_n with the given joint type ν. For $\gamma \le q\log q - O(1)$, while proving Theorem 4.4 it is shown in [4] that $\mathbb{E}\, Z_{\rm b}(\nu)/\mathbb{E}\, Z_{\rm b}(\overline{\nu}) \to 0$ exponentially in n (where $\overline{\nu}(x,y) = 1/q^2$ denotes the uniform joint type).

(3) The preceding result implies that, for any $\varepsilon > 0$ and some non-random $\delta_n(\varepsilon) \to 0$, the uniform measure over proper q-colorings of an instance of the random Poisson multi-graph G_n is with high probability (ε, δ_n)-spherical (see Definition 4.1). Notice that this implication is not straightforward as it requires bounding the expected ratio of $Z_{\rm b}(\nu)$ to the total number of pairs of proper q-colorings. We refer to [74] for this part of the argument.

(4) As mentioned in Remark 5.6, by Theorem 5.4 the latter sphericity condition yields that with high probability the q-colorings reconstruction problem is solvable if and only if it is tree-solvable. Therefore, the result of step (1) about the tree-reconstruction threshold $\gamma_{\rm r,tree}(q)$ completes the proof. □

6. XORSAT and finite-size scaling

XORSAT is a special constraint satisfaction problem (CSP) first introduced in [25]. An instance of XORSAT is defined by a pair $\mathcal{F} = (\mathbb{H}, \underline{b})$, where $\mathbb{H}$ is an $m \times n$ binary matrix and $\underline{b}$ is a binary vector of length m. A solution of this instance is just a solution of the linear system

$$\mathbb{H}\,\underline{x} = \underline{b} \qquad \mod 2\,. \tag{6.1}$$

In this section we shall focus on the l-XORSAT problem that is defined by requiring that $\mathbb{H}$ has l non zero entries per row. Throughout this section we assume $l \ge 3$.

It is quite natural to associate to an instance $\mathcal{F} = (\mathbb{H}, \underline{b})$ the uniform measure over its solutions, $\mu(\underline{x})$. If the positions of the non-vanishing entries of the a-th row of $\mathbb{H}$ are denoted by $i_1(a), \dots, i_l(a)$, the latter takes the form

$$\mu(\underline{x}) = \frac{1}{Z_{\mathbb{H},\underline{b}}} \prod_{a=1}^{m} \mathbb{I}(x_{i_1(a)} \oplus \cdots \oplus x_{i_l(a)} = b_a)\,, \tag{6.2}$$

where $\oplus$ denotes sum modulo 2, and $Z_{\mathbb{H},\underline{b}}$ is the number of solutions of the linear system. A factor graph representation is naturally associated to this measure analogously to what we did in Section 1.2 for the satisfiability problem.

In the following we study the set of solutions (equivalently, the uniform measure $\mu(\underline{x})$ over such solutions), of random l-XORSAT instances, distributed according to various ensembles. The relevant control parameter is the number of equations per variable $\alpha = m/n$. For the ensembles discussed here there exists a critical value $\alpha_{\rm s}(l)$ such that, if $\alpha < \alpha_{\rm s}(l)$, a random instance has solutions with high probability. Vice-versa, if $\alpha > \alpha_{\rm s}(l)$, a random instance typically does not have solutions.

In the regime in which random instances have solutions, the structure of the solution set changes dramatically as α crosses a critical value $\alpha_{\rm d}(l)$. The two regimes are characterized as follows (where all statements should be understood as holding with high probability).

I. $\alpha < \alpha_{\rm d}(l)$. The set of solutions of the linear system $\mathbb{H}\underline{x} = \underline{b}$ forms a 'well connected lump.' More precisely, there exist $c = c(\epsilon) < \infty$ such that if a set $\Omega \subseteq \{0,1\}^n$ contains at least one solution and at most half of the solutions, then for all $\epsilon > 0$ and n,

$$\frac{\mu(\partial_\epsilon \Omega)}{\mu(\Omega)} \geq n^{-c}\,. \tag{6.3}$$

II. $\alpha_{\rm d}(l) < \alpha < \alpha_{\rm s}(l)$. The set of solutions is 'clustered.' There exists a partition of the hypercube $\{0,1\}^n$ into sets $\Omega_1, \dots, \Omega_{\mathcal{N}}$ such that $\mu(\Omega_\ell) > 0$ and

$$\mu(\partial_\epsilon \Omega_\ell) = 0\,, \tag{6.4}$$

for some $\epsilon > 0$ and all ℓ. Further $\mathcal{N} = e^{n\Sigma + o(n)}$ for some $\Sigma > 0$, and each subset Ω_ℓ contains the same number of solutions $|\{\underline{x} \in \Omega_\ell : \mathbb{H}\underline{x} = \underline{b}\}| = e^{ns + o(n)}$, for some $s \geq 0$. Finally, the uniform measure over solutions in Ω_ℓ, namely $\mu_\ell(\,\cdot\,) = \mu(\,\cdot\,|\Omega_\ell)$, satisfies the condition (6.3).

The fact that the set of solutions of XORSAT forms an affine space makes it a much simpler problem, both computationally and in terms of analysis. Nevertheless, XORSAT shares many common features with other CSPs. For example, regimes I and II are analogous to the first two regimes introduced in Section 4.1 for the coloring problem. In particular, the measure $\mu(\,\cdot\,)$ exhibits coexistence in the second regime, but not in the first. This phenomena is

seen in many random CSPs ensembles following the framework of Section 1.2.2, well beyond the cases of coloring and XORSAT (see [65] for further details and references). Some rigorous results in this direction are derived in [6, 68], but the precise parameter range for the various regimes remains a conjecture, yet to be proved. Even for XORSAT, where the critical values $\alpha_{\rm d}(l)$, $\alpha_{\rm s}(l)$ have been determined rigorously (see [23, 72]), the picture we described is not yet completely proved. Indeed, as of now, we neither have a proof of absence of coexistence for $\alpha < \alpha_{\rm d}(l)$, i.e. the estimate Eq. (6.3), nor a proof of its analog for the uniform measure $\mu_\ell(\,\cdot\,)$ when $\alpha > \alpha_{\rm d}(l)$.

In Section 6.1 we focus on the case of random regular graphs, moving in Section 6.2 to uniformly random ensembles and their 2-cores, for which Sections 6.3 to 6.6 explore the dynamical (or 'clustering') phase transition and derive its precise behavior at moderate values of n.

6.1. XORSAT on random regular graphs

We begin with some general facts about the XORSAT problem. Given a XORSAT instance $\mathcal{F} = (\mathbb{H}, \underline{b})$, we denote by $r(\mathbb{H})$ the rank of $\mathbb{H}$ over the finite field GF(2), and by $Z_{\mathbb{H}} \geq 1$ the number of solutions $\underline{x}$ of $\mathbb{H}\underline{x} = \underline{0}$ mod 2. From linear algebra we know that $\mathcal{F}$ is satisfiable (i.e. there exists a solution for the system $\mathbb{H}\underline{x} = \underline{b} \mod 2$) if and only if $\underline{b}$ is in the image of $\mathbb{H}$. This occurs for precisely $2^{r(\mathbb{H})}$ of the 2^m possible binary vectors $\underline{b}$ and in particular it occurs for $\underline{b} = \underline{0}$. If $\mathcal{F}$ is satisfiable then the set of all solutions is an affine space of dimension $n - r(\mathbb{H})$ over GF(2), hence of size $Z_{\mathbb{H}} = 2^{n-r(\mathbb{H})}$. Further, $r(\mathbb{H}) \leq m$ with $r(\mathbb{H}) = m$ if and only if the rows of $\mathbb{H}$ are linearly independent (over GF(2)), or equivalently iff $Z_{\mathbb{H}^T} = 1$ (where $\mathbb{H}^T$ denotes the transpose of the matrix $\mathbb{H}$).

Let us now consider a generic distribution over instances $\mathcal{F} = (\mathbb{H}, \underline{b})$ such that $\underline{b}$ is chosen uniformly from $\{0,1\}^m$ independently of the matrix $\mathbb{H}$. It follows from the preceding that

$$\mathbb{P}(\mathcal{F} \text{ is satisfiable}) = 2^{n-m}\mathbb{E}[1/Z_{\mathbb{H}}] \geq \mathbb{P}(Z_{\mathbb{H}^T} = 1)\,. \tag{6.5}$$

By these considerations also

$$\mathbb{P}(\mathcal{F} \text{ is satisfiable}) \leq \frac{1}{2} + \frac{1}{2}\mathbb{P}(Z_{\mathbb{H}^T} = 1)\,.$$

As mentioned already, the probability that a random instance is satisfiable, $\mathbb{P}(\mathcal{F}$ is satisfiable), abruptly drops from near one to near zero when the number of equations per variable crosses a threshold $\alpha_{\rm s}(l)$. As a consequence

of our bounds, $\mathbb{P}(Z_{\mathbb{H}^T} = 1)$ abruptly drops from near one to near zero at the same threshold.

Suppose that $\mathbb{H}$ is the *parity matrix* of a factor graph $G = (V, F, E)$ which may have multiple edges. That is, each entry H_{ai} of the binary matrix $\mathbb{H}$ is just the parity of the multiplicity of edge (a, i) in G. We associate to the instance $\mathcal{F}$ an energy function $\mathcal{E}_{\mathbb{H},\underline{b}}(\underline{x})$ given by the number of unsatisfied equations in the linear system $\mathbb{H}\underline{x} = \underline{b}$, with the corresponding partition function

$$Z_{\mathbb{H},\underline{b}}(\beta) = \sum_{\underline{x}} \exp\{-2\beta \mathcal{E}_{\mathbb{H},\underline{b}}(\underline{x})\}\,. \tag{6.6}$$

In particular, $Z_{\mathbb{H}} = \lim_{\beta\to\infty} Z_{\mathbb{H},\underline{b}}(\beta)$ whenever the instance is satisfiable. Moreover, it is easy to show that $Z_{\mathbb{H},\underline{b}}(\beta)$ is independent of $\underline{b}$ whenever the instance is satisfiable.

We thus proceed to apply the general approach of high-temperature expansion on $Z_{\mathbb{H},\underline{0}}(\beta)$, which here yields important implications for all $\beta > 0$ and in particular, also on $Z_{\mathbb{H}}$. For doing so, it is convenient to map the variable domain $\{0, 1\}$ to $\{+1, -1\}$ and rewrite $Z_{\mathbb{H},\underline{0}}(\beta)$ as the partition function of a generalized (ferromagnetic) Ising model of the form (2.12). That is,

$$Z_{\mathbb{H},\underline{0}}(\beta) = e^{-\beta|F|} \sum_{\underline{x}\in\{+1,-1\}^V} \exp\Big\{\beta \sum_{a\in F} x_a\Big\} \equiv e^{-\beta|F|} Z_G(\beta)\,, \tag{6.7}$$

where $x_a \equiv \prod_{i\in\partial a} x_i$ for each $a \in F$. We also introduce the notion of a *hyper-loop* in a factor graph $G = (V, F, E)$, which is a subset F' of function nodes such that every variable node $i \in V$ has an even degree in the induced subgraph $G' = (V, F', E')$.

Lemma 6.1. *Set $N_G(0) \equiv 1$ and $N_G(\ell)$ denote the number of hyper-loops of size $\ell \geq 1$ in a factor graph $G = (V, F, E)$. Then, for any $\beta \in \mathbb{R}$,*

$$Z_G(\beta) = 2^{|V|}(\cosh\beta)^{|F|} \sum_{\ell=0}^{|F|} N_G(\ell)\,(\tanh\beta)^{\ell}\,. \tag{6.8}$$

Further, if $Z_{\mathbb{H}^T} = 1$ with $\mathbb{H}$ the parity matrix of G then

$$Z_G(\beta) = 2^{|V|}(\cosh\beta)^{|F|}\,.$$

Proof. Observe that $e^{\beta x_a} = \cosh(\beta)[1 + x_a(\tanh\beta)]$ for any function node $a \in F$ and any $\underline{x} \in \{+1, -1\}^V$. Thus, setting $F = [m]$, we have the following

'high-temperature expansion' of $Z_G(\beta)$ as a polynomial in $(\tanh\beta)$,

$$Z_G(\beta) = (\cosh\beta)^m \sum_{\underline{x}} \prod_{a\in F} [1 + x_a(\tanh\beta)]$$
$$= (\cosh\beta)^m \sum_{F'\subseteq[m]} (\tanh\beta)^{|F'|} \sum_{\underline{x}} \prod_{a\in F'} x_a \,.$$

Since $x_i^2 = 1$ for each $i \in V$ we see that $\prod_{a\in F'} x_a$ is simply the product of x_i over all variable nodes i that have an odd degree in the induced subgraph G'. The sum of this quantity over all $\underline{x} \in \{+1,-1\}^V$ is by symmetry zero, unless F' is either an empty set or a hyper-loop of G, in which case this sum is just the number of such binary vectors $\underline{x}$, that is $2^{|V|}$. Therefore, upon collecting together all hyper-loops F' in G of the same size we get the stated formula (6.8). To complete the proof of the lemma note that the sum of columns of the transpose $\mathbb{H}^T$ of the parity matrix of G corresponding to the function nodes in a hyper-loop F' in G must be the zero vector over the field GF(2). Hence, the existence of a hyper-loop in G provides a non-zero solution of $\mathbb{H}^T\underline{y} = \underline{0} \mod 2$ (in addition to the trivial zero solution). Consequently, if $Z_{\mathbb{H}^T} = 1$ then $N_G(\ell) = 0$ for all $\ell \geq 1$, yielding the stated explicit expression for $Z_G(\beta)$. □

We now consider the k-XORSAT for ensembles $\mathcal{G}_{l,k}(n,m)$ of random (l,k)-*regular graphs* drawn from the corresponding configuration model. Such an ensemble is defined whenever $nl = mk$ as follows. Attach l half-edges to each variable node $i \in V$, and k half-edges to each function node $a \in F$. Draw a uniformly random permutation over nl elements, and connect edges on the two sides accordingly. We then have the following result about uniqueness of solutions of random regular linear systems.

Theorem 6.2. *Let $\mathbb{H}$ denote the $m \times n$ parity matrix of a random (l,k)-regular factor graph from $\mathcal{G}_{l,k}(n,m)$, with $l > k \geq 2$. Then, the linear system $\mathbb{H}\underline{x} = \underline{0} \mod 2$ has, with high probability as $n \to \infty$, the unique solution $\underline{x} = \underline{0}$.*

Proof. Let $Z_{\mathbb{H}}(w)$ denote the number of solutions of $\mathbb{H}\underline{x} = \underline{0}$ with w non-zero entries. Such solution corresponds to a coloring, by say red, w vertices of the multi-graph G, and by, say blue, the remaining $n - w$ vertices, while having an even number of red half-edges at each function node. A convenient way to compute $\mathbb{E}\, Z_{\mathbb{H}}(w)$ is thus to divide the number of possible graph colorings with this property by the total size $(nl)!$ of the ensemble $\mathcal{G}_{l,k}(n,m)$. Indeed,

per integers $(m_r, r = 0, \ldots, k)$ that sum to m there are

$$\binom{m}{m_0, \ldots, m_k} \prod_{r=0}^{k} \binom{k}{r}^{m_r}$$

ways of coloring the mk half-edges of function nodes such that m_r nodes have r red half-edges, for $r = 0, \ldots, k$. There are $\binom{n}{w}$ ways of selecting red vertices and $(wl)!(nl - wl)!$ color consistent ways of matching half-edges of factor nodes with those of vertices, provided one has the same number of red half-edges in both collections, that is $\sum_r r m_r = wl$. The value of $\mathbb{E}Z_{\mathbb{H}}(w)$ is thus obtained by putting all this together, and summing over all choices of $(m_0, \ldots, m_k)$ such that $m_r = 0$ for odd values of r. Setting $w = n\omega$ and using Stirling's formula one finds that $\mathbb{E}Z_{\mathbb{H}}(w) = \exp(n\phi(\omega) + o(n))$ for any fixed $\omega \in (0, 1)$, with an explicit expression for $\phi(\cdot)$ (c.f. [65, Section 11.2.1] and the references therein). For $l > k \geq 2$ the only local maximum of $\phi(\omega)$ is at $\omega = 1/2$, with $\phi(0) = \phi(1) = 0$ and $\phi(1/2) < 0$. Hence, in this case $\phi(\omega) < 0$ for all $\omega \in (0, 1)$. Further, from the formula for $\mathbb{E}Z_{\mathbb{H}}(w)$ one can show that for $\kappa > 0$ small enough the sum of $\mathbb{E}Z_{\mathbb{H}}(w)$ over $1 \leq w \leq \kappa n$ and $n - w \leq \kappa n$ decays to zero with n. Therefore,

$$\lim_{n\to\infty} \sum_{w=1}^{n} \mathbb{E}\, Z_{\mathbb{H}}(w) = 0\,,$$

which clearly implies our thesis. □

We have the following consequence about satisfiability of XORSAT for random (l, k)-regular factor graphs.

Corollary 6.3. *Choosing a random (l, k)-regular factor graph G from the ensemble $\mathcal{G}_{l,k}(n, m)$, with $k > l \geq 2$. Then, the probability that $Z_G(\beta) = 2^n(\cosh\beta)^{ln/k}$ goes to one as $n \to \infty$ and so does the probability that k-XORSAT with the corresponding parity matrix $\mathbb{H}$ has 2^{sn} solutions for $s = 1 - l/k$.*

Proof. Let $\mathbb{H}$ be the parity matrix of a randomly chosen (l, k)-regular factor graph G from ensemble $\mathcal{G}_{l,k}(n, m)$. Then, $\mathbb{H}^T$ has the law of the parity matrix for a random (k, l)-regular factor graph from ensemble $\mathcal{G}_{k,l}(m, n)$. Thus, from Theorem 6.2 we know that $\mathbb{P}(Z_{\mathbb{H}^T} = 1) \to 1$ as $n \to \infty$ and by Lemma 6.1 with same probabilities also $Z_G(\beta) = 2^n(\cosh\beta)^{ln/k}$ (as here $|V| = n$ and $|F| = nl/k$). We complete the proof upon recalling that if $Z_{\mathbb{H}^T} = 1$ then $r(\mathbb{H}) = m = |F|$ and there are 2^{n-m} solutions $\underline{x}$ of the corresponding XORSAT problem (for any choice of $\underline{b}$). □

See [65, Chapter 18] for more information on XORSAT models, focusing on the zero-temperature case.

6.2. Hyper-loops, hypergraphs, cores and a peeling algorithm

As we saw in the previous section, if the bipartite graph G associated to the matrix $\mathbb{H}$ does not contain hyper-loops, then the linear system $\mathbb{H}\underline{x} = \underline{b}$ is solvable for any $\underline{b}$. This is what happens for $\alpha < \alpha_{\rm s}(d)$: a random matrix $\mathbb{H}$ with l non-zero elements per row is, with high probability, free from hyper-loops. Vice versa, when $\alpha > \alpha_{\rm s}(d)$ the matrix $\mathbb{H}$ contains, with high probability, $\Theta(n)$ hyper-loops. Consequently, the linear system $\mathbb{H}\underline{x} = \underline{b}$ is solvable only for $2^{m-\Theta(n)}$ of the vectors $\underline{b}$. If $\underline{b}$ is chosen uniformly at random, this implies that $\mathbb{P}\{\mathcal{F} \text{ is satisfiable}\} \to 0$ as $n \to \infty$.

Remarkably, the clustering threshold $\alpha_{\rm d}(l)$ coincides with the threshold for appearance of a specific subgraph of the bipartite graph G, called the *core* of G. The definition of the core is more conveniently given in the language of hypergraphs. This is an equivalent description of factor graphs, where the hypergraph corresponding to $G = (V, F, E)$ is formed by associating with each factor node $a \in F$ the hyper-edge (i.e. a subset of vertices in V), ∂a consisting of all vertices $i \in V$ such that $(i, a) \in E$. The same applies for factor multi-graphs, in which case a vertex $i \in V$ may appear with multiplicity larger than one in some hyper-edges.

Definition 6.4. *The* r-core *of hyper-graph G is the unique subgraph obtained by recursively removing all vertices of degree less than r (when counting multiplicities of vertices in hyper-edges). In particular, the* 2*-core, hereafter called the* core *of G, is the maximal collection of hyper-edges having no vertex appearing in only one of them (and we use the same term for the induced subgraph).*

Obviously, if G contains a non-empty hyper-loop, it also contains a non-empty core. It turns out that the probability that a random hypergraph contains a non-empty core grows sharply from near zero to near one as the number of hyper-edges crosses a threshold which coincides with the clustering threshold $\alpha_{\rm d}(l)$ of XORSAT.

Beyond XORSAT, the core of a hyper-graph plays an important role in the analysis of many combinatorial problems.

For example, Karp and Sipser [55] consider the problem of finding the largest possible matching (i.e. vertex disjoint set of edges) in a graph G. They propose a simple *peeling algorithm* that recursively selects an edge $e = (i, j) \in G$ for which the vertex i has degree one, as long as such an

edge exists, and upon including e in the matching, the algorithm removes it from G together with all edges incident on j (that can no longer belong to the matching). Whenever the algorithm successfully matches all vertices, the resulting matching can be shown to have maximal size. Note that this happens if an only if the core of the hyper-graph $\widetilde{G}$ is empty, where $\widetilde{G}$ has a c-node $\widetilde{e}$ per edge e of G and a v-node $\widetilde{i}$ per vertex i of degree two or more in G that is incident on $\widetilde{e}$ in $\widetilde{G}$ if and only if e is incident on i in G. Consequently, the performance of the Karp-Sipser algorithm for a randomly selected graph has to do with the probability of non-empty core in the corresponding graph ensemble. For example, [55] analyze the asymptotics of this probability for a uniformly chosen random graph of N vertices and $M = \lfloor Nc/2 \rfloor$ edges, as $N \to \infty$ (c.f. [12, 34] for recent contributions).

A second example deals with the decoding of a noisy message when communicating over the binary erasure channel with a low-density parity-check code ensemble. This amounts to finding the *unique* solution of a linear system over GF(2) (the solution exists by construction, but is not necessarily unique, in which case decoding fails). If the linear system includes an equation with only one variable, we thus determine the value of this variable, and substitute it throughout the system. Repeated recursively, this procedure either determines all the variables, thus yielding the unique solution of the system, or halts on a linear sub-system each of whose equations involves at least two variables. While such an algorithm is not optimal (when it halts, the resulting linear sub-system might still have a unique solution), it is the simplest instance of the widely used belief propagation decoding strategy, that has proved extremely successful. For example, on properly optimized code ensembles, this algorithm has been shown to achieve the theoretical limits for reliable communication, i.e., Shannon's channel capacity (see [62]). Here a hyper-edge of the hyper-graph G is associated to each variable, and a vertex is associated to each equation, or parity check, and the preceding decoding scheme successfully finds the unique solution if and only if the core of G is empty.

In coding theory one refers to each variable as a v-node of the corresponding bipartite factor graph representation of G and to each parity-check as a c-node of this factor graph. This coding theory setting is dual to the one considered in XORSAT. Indeed, as we have already seen, satisfiability of an instance $(\mathbb{H}, \underline{b})$ for most choices of $\underline{b}$ is equivalent to the uniqueness of the solution of $\mathbb{H}^T \underline{x} = 0$. Hereafter we adopt this 'dual' but equivalent coding theory language, considering a hyper-graph G chosen uniformly from an ensemble $\mathcal{G}_l(n, m)$ with n hyper-edges (or v-nodes), each of whom is a collection of $l \geq 3$ vertices (or c-nodes), from the vertex set $[m]$. Note that

as we passed to the dual formulation, we also inverted the roles of n and m. More precisely, each hyper-graph in $\mathcal{G} = \mathcal{G}_l(n,m)$ is described by an ordered list of edges, i.e. couples (i,a) with $i \in [n]$ and $a \in [m]$

$$E = [(1,a_1),(1,a_2),\ldots,(1,a_l);(2,a_{l+1}),\ldots;(n,a_{(n-1)l+1}),\ldots,(n,a_{nl})]\,,$$

where a couple (i,a) appears *before* (j,b) whenever $i<j$ and each v-node i appears *exactly* l times in the list, with $l \geq 3$ a fixed integer parameter. In this configuration model the *degree* of a v-node i (or c-node a), refers to the number of edges (i,b) (respectively (j,a)) in E to which it belongs (which corresponds to counting hyper-edges and vertices *with their multiplicity*).

To sample G from the uniform distribution over $\mathcal{G}$ consider the v-nodes in order, $i = 1,\ldots,n$, choosing for each v-node and $j = 1,\ldots,l$, independently and uniformly at random a c-node $a = a_{(i-1)l+j} \in [m]$ and adding the couple (i,a) to the list E. Alternatively, to sample from this distribution first attribute sockets $(i-1)l+1,\ldots,il$ to the i-th v-node, $i = 1,\ldots,n$, then attribute k_a sockets to each c-node a, where k_a's are mutually independent Poisson(ζ) random variables, conditioned upon their sum being nl (these sockets are ordered using any pre-established convention). Finally, connect the v-node sockets to the c-node sockets according to a permutation σ of $\{1,\ldots,nl\}$ that is chosen uniformly at random and independently of the choice of k_a's.

In the sequel we take $m = \lfloor n\rho \rfloor$ for $\rho = l/\gamma > 0$ bounded away from 0 and ∞ and study the large n asymptotics of the probability

$$P_l(n,\rho) \equiv \mathbb{P}\{G \in \mathcal{G}_l(n,m) \text{ has a non-empty core}\} \tag{6.9}$$

that a hyper-graph G of this distribution has a non-empty core. Setting $\mathbb{H}^T$ as the parity matrix of a uniformly chosen G from $\mathcal{G}_l(n,m)$ corresponds to a binary matrix $\mathbb{H}$ chosen uniformly (according to a configuration model), among all $n \times m$ matrices with l non-zero entries per row. That is, the parameter ρ corresponds to $1/\alpha$ in the l-XORSAT.

6.3. The approximation by a smooth Markov kernel

Our approach to $P_l(n,\rho)$ is by analyzing whether the process of sequentially peeling, or decimating, c-nodes of degree one, corresponding to the decoding scheme mentioned before, ends with an empty graph, or not. That is, consider the inhomogeneous Markov chain of graphs $\{G(\tau),\ \tau \geq 0\}$, where $G(0)$ is a uniformly random element of $\mathcal{G}_l(n,m)$ and for each $\tau = 0,1,\ldots,$ if there is a non-empty set of c-nodes of degree 1, choose one of them (let's say

a) uniformly at random, deleting the corresponding edge (i,a) together with all the edges incident to the v-node i. The graph thus obtained is $G(\tau+1)$. In the opposite case, where there are no c-nodes of degree 1 in $G(\tau)$, we set $G(\tau+1)=G(\tau)$.

Reducing the state space to $\mathbb{Z}_+^2$. We define furthermore the process $\{\vec{z}(\tau)=(z_1(\tau),z_2(\tau)),\,\tau\geq 0\}$ on $\mathbb{Z}_+^2$, where $z_1(\tau)$ and $z_2(\tau)$ are, respectively, the number of c-nodes in $G(\tau)$, having degree one or larger than one. Necessarily, $(n-\widehat{\tau})l\geq z_1(\tau)+2z_2(\tau)$, with equality if $z_2(\tau)=0$, where $\widehat{\tau}\equiv\min(\tau,\inf\{\tau'\geq 0: z_1(\tau')=0\})$, i.e. $\widehat{\tau}=\tau$ till the first τ' such that $z_1(\tau')=0$, after which $\widehat{\tau}$ is frozen (as the algorithm stops).

Fixing $l\geq 3$, m and n, set $\vec{z}\equiv(z_1,z_2)\in\mathbb{Z}_+^2$ and $\mathcal{G}(\vec{z},\tau)$ denote the ensemble of possible bipartite graphs with z_1 c-nodes of degree one and z_2 c-nodes of degree at least two, after exactly τ removal steps of this process. Then, $\mathcal{G}(\vec{z},\tau)$ is non-empty only if $z_1+2z_2\leq(n-\tau)l$ with equality whenever $z_2=0$. Indeed, each element of $\mathcal{G}(\vec{z},\tau)$ is a bipartite graph $G=(U,V;R,S,T;E)$ where U,V are disjoint subsets of $[n]$ with $U\cup V=[n]$ and R,S,T are disjoint subsets of $[m]$ with $R\cup\ S\cup T=[m]$, having the cardinalities $|U|=\tau$, $|V|=n-\tau$, $|R|=m-z_1-z_2$, $|S|=z_1$, $|T|=z_2$ and the ordered list E of $(n-\tau)l$ edges (i,a) with i a v-node and a a c-node such that each $i\in V$ appears as the first coordinate of exactly l edges in E, while each $j\in U$ does not appear in any of the couples in E. Similarly, each $c\in R$ does not appear in E, each $b\in S$ appears as the second coordinate of exactly one edge in E, and each $a\in T$ appears in some $k_a\geq 2$ such edges.

The following observation allows us to focus on the much simpler process $\vec{z}(\tau)$ on $\mathbb{Z}_+^2$ instead of the graph process $G(\tau)\in\mathcal{G}(\vec{z},\widehat{\tau})$.

Lemma 6.5. *Conditional on $\{\vec{z}(\tau'),0\leq\tau'\leq\tau\}$, the graph $G(\tau)$ is uniformly distributed over $\mathcal{G}(\vec{z},\widehat{\tau})$. Consequently, the process $\{\vec{z}(\tau)\,\tau\geq 0\}$ is an inhomogeneous Markov process.*

Proof outline: Fixing τ, $\vec{z}=\vec{z}(\tau)$ such that $z_1>0$, $\vec{z}'=\vec{z}(\tau+1)$ and $G'\in\mathcal{G}(\vec{z}',\tau+1)$, let $N(G'|\vec{z},\tau)$ count the pairs of graphs $G\in\mathcal{G}(\vec{z},\tau)$ and choices of the deleted c-node from S that result with G' upon applying a single step of our algorithm. Obviously, G and G' must be such that $R\subset R'$, $S\subseteq R'\cup S'$ and $T'\subseteq T$. With $q_0\equiv|R'\cap S|$, $p_0\equiv|R'\cap T|$, $q_1\equiv|S'\cap T|$ and q_2 denoting the number of c-nodes $a\in T'$ for which $k_a>k'_a$, it is shown in [32, proof of Lemma 3.1] that $(p_0,\,q_0,\,q_1,\,q_2)$ belongs to the subset $\mathcal{D}$ of $\mathbb{Z}_+^4$ where both the relations

$$\left\{\begin{array}{rcl} z_0 &=& z'_0-q_0-p_0\,,\\ z_1 &=& z'_1+q_0-q_1\,,\\ z_2 &=& z'_2+p_0+q_1\,,\end{array}\right. \tag{6.10}$$

for $z_0 = m - z_1 - z_2$, $z_0' = m - z_1' - z_2'$, and the inequalities $(n-\tau)l - (z_1 + 2z_2) \geq l - (2p_0 + q_0 + q_1) \geq q_2$, $q_0 + p_0 \leq z_0'$, $q_1 \leq z_1'$ (equivalently, $q_0 \leq z_1$), $q_2 \leq z_2'$ (equivalently, $p_0 + q_1 + q_2 \leq z_2$) hold. In particular $|\mathcal{D}| \leq (l+1)^4$. It is further shown there that

$$N(G'|\vec{z},\tau) = (\tau+1)\, l! \sum_{\mathcal{D}} \binom{m - z_1' - z_2'}{q_0, p_0, \cdot} \binom{z_1'}{q_1} \binom{z_2'}{q_2} c_l(q_0, p_0, q_1, q_2)\,, \quad (6.11)$$

depends on G' only via $\vec{z}'$, where

$$c_l(q_0, p_0, q_1, q_2) = q_0 \mathsf{coeff}[(e^{\mathbf{x}} - 1 - \mathbf{x})^{p_0} (e^{\mathbf{x}} - 1)^{q_1+q_2}, \mathbf{x}^{l-q_0}]\,.$$

We start at $\tau = 0$ with a uniform distribution of $G(0)$ within each possible ensemble $\mathcal{G}(\vec{z}(0), 0)$. As $N(G'|\vec{\omega}, \tau)$ depends on G' only via $\vec{\omega}'$ it follows by induction on $\tau = 1, 2, \ldots$ that conditional on $\{\vec{z}(\tau'), 0 \leq \tau' \leq \tau\}$, the graph $G(\tau)$ is uniformly distributed over $\mathcal{G}(\vec{z}, \widehat{\tau})$ as long as $\widehat{\tau} = \tau$. Indeed, if $z_1(\tau) > 0$, then with $h(\vec{z}, \tau)$ denoting the number of graphs in $\mathcal{G}(\vec{z}, \tau)$,

$$\mathbb{P}\left\{G(\tau+1) = G' | \{\vec{z}(\tau'), 0 \leq \tau' \leq \tau\}\right\} = \frac{1}{z_1} \frac{N(G'|\vec{z}(\tau), \tau)}{h(\vec{z}(\tau), \tau)}\,,$$

is the same for all $G' \in \mathcal{G}(\vec{z}', \tau+1)$. Moreover, noting that $G(\tau) = G(\widehat{\tau})$ and $\vec{z}(\tau) = \vec{z}(\widehat{\tau})$ we deduce that this property extends to the case of $\widehat{\tau} < \tau$ (i.e. $z_1(\tau) = 0$). Finally, since there are exactly $h(\vec{z}', \tau+1)$ graphs in the ensemble $\mathcal{G}(\vec{z}', \tau+1)$ the preceding implies that $\{\vec{z}(\tau),\, \tau \geq 0\}$ is an inhomogeneous Markov process whose transition probabilities

$$W_\tau^+(\Delta\vec{z}|\vec{z}) \equiv \mathbb{P}\{\vec{z}(\tau+1) = \vec{z} + \Delta\vec{z}\,|\,\vec{z}(\tau) = \vec{z}\}\,,$$

for $\Delta\vec{z} \equiv (\Delta z_1, \Delta z_2)$ and $z_1' = z_1 + \Delta z_1$, $z_2' = z_2 + \Delta z_2$ are such that $W_\tau^+(\Delta\vec{z}|\vec{z}) = \mathbb{I}(\Delta\vec{z} = 0)$ in case $z_1 = 0$, whereas $W_\tau^+(\Delta\vec{z}|\vec{z}) = h(\vec{z}', \tau+1) N(G'|\vec{z}, \tau)/(z_1 h(\vec{z}, \tau))$ when $z_1 > 0$. □

To sample from the uniform distribution on $\mathcal{G}(\vec{z}, \tau)$ first partition $[n]$ into U and V uniformly at random under the constraints $|U| = \tau$ and $|V| = (n-\tau)$ (there are $\binom{n}{\tau}$ ways of doing this), and independently partition $[m]$ to $R \cup S \cup T$ uniformly at random under the constraints $|R| = m - z_1 - z_2$, $|S| = z_1$ and $|T| = z_2$ (of which there are $\binom{m}{z_1, z_2, \cdot}$ possibilities). Then, attribute l v-sockets to each $i \in V$ and number them from 1 to $(n-\tau)l$ according to some pre-established convention. Attribute one c-socket to each $a \in S$ and k_a c-sockets to each $a \in T$, where k_a are mutually independent Poisson(ζ) random variables conditioned upon $k_a \geq 2$, and further conditioned upon

$\sum_{a\in T} k_a$ being $(n-\tau)l - z_1$. Finally, connect the v-sockets and c-sockets according to a uniformly random permutation on $(n-\tau)l$ objects, chosen independently of the k_a's. Consequently,

$$h(\vec{z},\tau) = \binom{m}{z_1, z_2, \cdot} \binom{n}{\tau} \mathsf{coeff}[(e^{\mathbf{x}} - 1 - \mathbf{x})^{z_2}, \mathbf{x}^{(n-\tau)l-z_1}]((n-\tau)l)! \quad (6.12)$$

Approximation by a smooth Markov transition kernel. Though the transition kernel $W_\tau^+(\cdot|\vec{z})$ of the process $\vec{z}(\cdot)$ is given explicitly via (6.11) and (6.12), it is hard to get any insight from these formulas, or to use them directly for finding the probability of this process hitting the line $z_1(\tau) = 0$ at some $\tau < n$ (i.e. of the graph $G(0)$ having a non-empty core). Instead, we analyze the simpler transition probability kernel

$$\widehat{W}_\theta(\Delta\vec{z}|\vec{x}) \equiv \binom{l-1}{q_0 - 1, q_1, q_2} \mathfrak{p}_0^{q_0-1} \mathfrak{p}_1^{q_1} \mathfrak{p}_2^{q_2}\,, \quad (6.13)$$

with $q_0 = -\Delta z_1 - \Delta z_2 \geq 1$, $q_1 = -\Delta z_2 \geq 0$ and $q_2 = l + \Delta z_1 + 2\Delta z_2 \geq 0$, where

$$\mathfrak{p}_0 = \frac{x_1}{l(1-\theta)}\,, \quad \mathfrak{p}_1 = \frac{x_2\lambda^2 e^{-\lambda}}{l(1-\theta)(1-e^{-\lambda}-\lambda e^{-\lambda})}\,, \quad \mathfrak{p}_2 = 1 - \mathfrak{p}_0 - \mathfrak{p}_1, \quad (6.14)$$

for each $\theta \in [0,1)$ and $\vec{x} \in \mathbb{R}_+^2$ such that $x_1 + 2x_2 \leq l(1-\theta)$. In case $x_2 > 0$ we set $\lambda = \lambda(\vec{x},\theta)$ as the unique positive solution of

$$\frac{\lambda(1-e^{-\lambda})}{1-e^{-\lambda}-\lambda e^{-\lambda})} = \frac{l(1-\theta)-x_1}{x_2} \quad (6.15)$$

while for $x_2 = 0$ we set by continuity $\mathfrak{p}_1 = 0$ (corresponding to $\lambda \to \infty$).

Intuitively, $(\mathfrak{p}_0, \mathfrak{p}_1, \mathfrak{p}_2)$ are the probabilities that each of the remaining $l-1$ edges emanating from the v-node to be deleted at the $\tau = n\theta$ step of the algorithm is connected to a c-node of degree 1, 2 and at least 3, respectively. Indeed, of the $nl(1-\theta)$ v-sockets at that time, precisely $z_1 = nx_1$ are connected to c-nodes of degree one, hence the formula for $\mathfrak{p}_0$. Our formula for $\mathfrak{p}_1$ corresponds to postulating that the $z_2 = nx_2$ c-nodes of degree at least two in the collection T follow a $\mathsf{Poisson}(\lambda)$ degree distribution, conditioned on having degree at least two, setting $\lambda > 0$ to match the expected number of c-sockets per c-node in T which is given by the right side of (6.15). To justify this assumption, note that

$$\mathsf{coeff}[(e^{\mathbf{x}} - 1 - \mathbf{x})^t, \mathbf{x}^s]\lambda^s(e^\lambda - 1 - \lambda)^{-t} = \mathbb{P}(\sum_{i=1}^t N_i = s)\,, \quad (6.16)$$

for i.i.d. random variables N_i, each having the law of a $\mathsf{Poisson}(\lambda)$ random variable conditioned to be at least two. We thus get from (6.11) and (6.12), upon applying the local CLT for such partial sums, that the tight approximation

$$\left|W_\tau^+(\Delta\vec{z}|\vec{z}) - \widehat{W}_{\tau/n}(\Delta\vec{z}|n^{-1}\vec{z})\right| \leq \frac{C(l,\epsilon)}{n}$$

applies for $(\vec{z},\tau) \in \mathcal{Q}_+(\epsilon)$, $\Delta z_1 \in \{-l,\ldots,l-2\}$, $\Delta z_2 \in \{-(l-1),\ldots,0\}$, with

$$\mathcal{Q}_+(\epsilon) \equiv \big\{(\vec{z},\tau) \,:\, 1 \leq z_1\,;\; n\epsilon \leq z_2\,; 0 \leq \tau \leq n(1-\epsilon)\,; \\ n\epsilon \leq (n-\tau)l - z_1 - 2z_2\big\}\,,$$

approaching (as $\epsilon \downarrow 0$) the set $\mathcal{Q}_+(0) \subset \mathbb{Z}^3$ in which the trajectory $(\vec{z}(\tau),\tau)$ evolves till hitting one of its absorbing states $\{(\vec{z},\tau) : z_1(\tau) = 0, \tau \leq n\}$ (c.f. [32, Lemma 4.5] for the proof, where the restriction to $\mathcal{Q}_+(\epsilon)$ guarantees that the relevant values of t in (6.16) are of order n).

The initial distribution. Considering $m = \lfloor n\rho \rfloor$, for $\rho = l/\gamma \in [\epsilon, 1/\epsilon]$ and large n, recall that

$$\mathbb{P}(\vec{z}(0) = \vec{z}) = \frac{h(\vec{z},0)}{m^{nl}} = \frac{\mathbb{P}_\gamma\left\{\vec{S}_m = (z_1,z_2,nl)\right\}}{\mathbb{P}_\gamma\left\{S_m^{(3)} = nl\right\}}$$

where $\vec{S}_m = \sum_{i=1}^m \vec{X}_i$ for $\vec{X}_i = (\mathbb{I}_{N_i=1}, \mathbb{I}_{N_i\geq 2}, N_i) \in \mathbb{Z}_+^3$ and N_i that are i.i.d. $\mathsf{Poisson}(\gamma)$ random variables (so $\mathbb{E}S_m^{(3)} = nl$ up to the quantization error of at most γ). Hence, using sharp local CLT estimates for $\vec{S}_m$ we find that the law of $\vec{z}(0)$ is well approximated by the multivariate Gaussian law $\mathsf{G}_2(\cdot|n\vec{y}(0); n\mathbb{Q}(0))$ whose mean $n\vec{y}(0) \equiv n\vec{y}(\theta;\rho)$ consists of the first two coordinates of $n\rho\mathbb{E}\vec{X}_1$, that is,

$$\vec{y}(0;\rho) = \rho(\gamma e^{-\gamma}, 1 - e^{-\gamma} - \gamma e^{-\gamma})\,, \tag{6.17}$$

and its positive definite covariance matrix $n\mathbb{Q}(0;\rho)$ equals $n\rho$ times the conditional covariance of the first two coordinates of $\vec{X}_1$ given its third coordinates. That is,

$$\left\{\begin{array}{lcl} Q_{11}(0) &=& \frac{l}{\gamma}\,\gamma\, e^{-2\gamma}(e^\gamma - 1 + \gamma - \gamma^2)\,, \\ Q_{12}(0) &=& -\frac{l}{\gamma}\,\gamma\, e^{-2\gamma}(e^\gamma - 1 - \gamma^2)\,, \\ Q_{22}(0) &=& \frac{l}{\gamma}\, e^{-2\gamma}\left[(e^\gamma - 1) + \gamma(e^\gamma - 2) - \gamma^2(1+\gamma)\right]. \end{array}\right. \tag{6.18}$$

More precisely, as shown for example in [32, Lemma 4.4], for all n, r and $\rho \in [\epsilon, 1/\epsilon]$,

$$\sup_{\vec{u}\in\mathbb{R}^2} \sup_{x\in\mathbb{R}} |\mathbb{P}\{\vec{u}\cdot\vec{z} \le x\} - \mathsf{G}_2(\vec{u}\cdot\vec{z} \le x | n\vec{y}(0); n\mathbb{Q}(0))| \le \kappa(\epsilon) n^{-1/2} . \quad (6.19)$$

Absence of small cores. A considerable simplification comes from the observation that a typical large random hyper-graph does not have a non-empty core of size below a certain threshold. Indeed, a subset of v-nodes of a hyper-graph is called a *stopping set* if the restriction of the hyper-graph to this subset has no c-node of degree one. With $N(s,r)$ counting the number of stopping sets in our random hyper-graph which involve exactly s v-nodes and r c-nodes, observe that necessarily $r \le \lfloor ls/2 \rfloor$. Further, adapting a result of [79] (and its proof) to our graph ensemble, it is shown in [32, Lemma 4.7] that for $l \ge 3$ and any $\epsilon > 0$ there exist $\kappa = \kappa(l,\epsilon) > 0$ and $C = C(l,\epsilon)$ finite, such that for any $m \ge \epsilon n$

$$\mathbb{E}\Big[\sum_{s=1}^{m\kappa} \sum_{r=1}^{\lfloor ls/2 \rfloor} N(s,r)\Big] \le C m^{1-l/2} .$$

Since the core is the stopping set including the maximal number of v-nodes, this implies that a random hyper-graph from the ensemble $\mathcal{G}_l(n,m)$ has a non-empty core of less than $m\kappa$ v-nodes with probability that is at most $Cm^{1-l/2}$ (alternatively, the probability of having a non-empty core with less than $n\kappa$ v-nodes is at most $C\, n^{1-l/2}$).

6.4. The ODE method and the critical value

In view of the approximations of Section 6.3 the asymptotics of $P_l(n,\rho)$ reduces to determining the probability $\widehat{\mathbb{P}}_{n,\rho}(z_1(\tau) = 0 \text{ for some } \tau < n)$ that the inhomogeneous Markov chain on $\mathbb{Z}_+^2$ with the transition kernel $\widehat{W}_{\tau/n}(\Delta\vec{z}|n^{-1}\vec{z})$ of (6.13) and the initial distribution $\mathsf{G}_2(\cdot|n\vec{y}(0); n\mathbb{Q}(0))$, hits the line $z_1(\tau) = 0$ for some $\tau < n$.

The functions $(\vec{x},\theta) \mapsto \mathfrak{p}_a(\vec{x},\theta)$, $a = 0,1,2$ are of Lipschitz continuous partial derivatives on each of the compact subsets

$$\widehat{q}_+(\epsilon) \equiv \{(\vec{x},\theta) \,:\, 0 \le x_1 \,;\, 0 \le x_2 \,;\, \theta \in [0, 1-\epsilon] \,;\, 0 \le (1-\theta)l - x_1 - 2x_2\} ,$$

of $\mathbb{R}^2 \times \mathbb{R}_+$ where the rescaled (macroscopic) state and time variables $\vec{x} \equiv n^{-1}\vec{z}$ and $\theta \equiv \tau/n$ are whenever $(\vec{z},\tau) \in \mathcal{Q}_+(\epsilon)$. As a result, the transition

kernels of (6.13) can be extended to any $\vec{x} \in \mathbb{R}^2$ such that for some $L = L(l,\epsilon)$ finite, any $\theta, \theta' \in [0, 1-\epsilon]$ and $\vec{x}, \vec{x}\,' \in \mathbb{R}^2$

$$\left|\left|\widehat{W}_{\theta'}(\,\cdot\,|\vec{x}\,') - \widehat{W}_{\theta}(\,\cdot\,|\vec{x})\right|\right|_{\rm TV} \leq L\ (||\vec{x}\,' - \vec{x}|| + |\theta' - \theta|)$$

(with $||\cdot||_{\rm TV}$ denoting the total variation norm and $||\cdot||$ the Euclidean norm in $\mathbb{R}^2$).

So, with the approximating chain of kernel $\widehat{W}_\theta(\Delta\vec{z}|\vec{x})$ having bounded increments ($= \Delta\vec{z}$), and its transition probabilities depending smoothly on $(\vec{x}, \theta)$, the scaled process $n^{-1}\vec{z}(\theta n)$ concentrates around the solution of the ODE

$$\frac{{\rm d}\vec{y}}{{\rm d}\theta}(\theta) \;\; = \;\; \vec{F}(\vec{y}(\theta), \theta)\,, \tag{6.20}$$

starting at $\vec{y}(0)$ of (6.17), where $\vec{F}(\vec{x}, \theta) = (-1 + (l-1)(\mathfrak{p}_1 - \mathfrak{p}_0), -(l-1)\mathfrak{p}_1)$ is the mean of $\Delta\vec{z}$ under the transitions of (6.13). This is shown for instance in [23, 62, 72].

We note in passing that this approach of using a deterministic ODE as an asymptotic approximation for slowly varying random processes goes back at least to [59], and such degenerate (or zero-one) fluid-limits have been established for many other problems. For example, this was done in [55] for the largest possible matching and in [80] for the size of r-core of random graphs (c.f. [73] for a general approach for deriving such results without recourse to ODE approximations).

Setting $h_\rho(u) \equiv u - 1 + \exp(-\gamma u^{l-1})$, with a bit of real analysis one verifies that for $\gamma = l/\rho$ finite, the ODE (6.20) admits a unique solution $\vec{y}(\theta;\rho)$ subject to the initial condition (6.17) such that $y_1(\theta;\rho) = l u^{l-1} h_\rho(u)$ for $u(\theta) \equiv (1-\theta)^{1/l}$, as long as $h_\rho(u(\theta)) \geq 0$. Thus, if ρ exceeds the finite and positive *critical density*

$$\rho_{\rm d} \equiv \inf\{\rho > 0 :\ h_\rho(u) > 0 \quad \forall u \in (0,1]\}\,,$$

then $y_1(\theta;\rho)$ is strictly positive for all $\theta \in [0,1)$, while for any $\rho \leq \rho_{\rm d}$ the solution $\vec{y}(\theta;\rho)$ first hits the line $y_1 = 0$ at some $\theta_*(\rho) < 1$.

Returning to the XORSAT problem, [23, 72] prove that for a uniformly chosen linear system with n equations and $m = \rho n$ variables the leaf removal algorithm is successful with high probability if $\rho > \rho_{\rm d}$ and fails with high probability if $\rho < \rho_{\rm d}$. See [32, Figure 1] for an illustration of this phenomenon. Similarly, in the context of decoding of a noisy message over the binary erasure channel (i.e. uniqueness of the solution for a given linear system over GF(2)), [62] show that with high probability this algorithm successfully decimates the whole hyper-graph without ever running out of degree

one vertices if $\rho > \rho_{\rm d}$. Vice versa, for $\rho < \rho_{\rm d}$, the solution $\vec{y}(\theta;\rho)$ crosses the $y_1 = 0$ plane near which point the algorithm stops with high probability and returns a core of size $O(n)$. The value of ρ translates into noise level in this communication application, so [62] in essence explicitly characterizes the critical noise value, for a variety of codes (i.e. random hyper-graph ensembles). Though this result has been successfully used for code design, it is often a poor approximation for the moderate code block-length (say, $n = 10^2$ to 10^5) that are relevant in practice.

The first order phase transition in the size of the core at $\rho = \rho_{\rm d}$ where it abruptly changes from an empty core for $\rho > \rho_{\rm d}$ to a core whose size is a positive fraction of n for $\rho < \rho_{\rm d}$, has other important implications. For example, as shown in [23, 72] and explained before, the structure of the set of solutions of the linear system changes dramatically at $\rho_{\rm d}$, exhibiting a 'clustering effect' when $\rho < \rho_{\rm d}$. More precisely, a typical instance of our ensemble has a core that corresponds to $n(1-\theta_*(\rho)) + o(n)$ equations in $ny_2(\theta_*(\rho)) + o(n)$ variables. The approximately 2^{m-n} solutions of the original linear system partition to about $2^{n\xi(\rho)}$ clusters according to their projection on the core, such that the distance between each pair of clusters is $O(n)$. This analysis also determines the location $\rho_{\rm s}$ of the satisfiability phase transition. That is, as long as $\xi(\rho) = y_2(\theta_*(\rho)) - (1-\theta_*(\rho))$ is positive, with high probability the original system is solvable (i.e the problem is satisfiable), whereas when $\xi(\rho) < 0$ it is non-solvable with high probability.

We conclude this subsection with a 'cavity type' direct prediction of the value of $\rho_{\rm d}$ without reference to a peeling algorithm (or any other stochastic dynamic). To this end, we set u to denote the probability that a typical c-node of $\mathcal{G}_l(n,m)$, say a, is part of the core. If this is the case, then an hyper-edge i incident to a is also part of the core iff all other $l-1$ sockets of i are connected to c-nodes from the core. Using the Bethe ansatz we consider the latter to be the intersection of $l-1$ independent events, each of probability u. So, with probability u^{l-1} an hyper-edge i incident to a from the core, is also in the core. As already seen, a typical c-node in our graph ensemble has $\mathsf{Poisson}(\gamma)$ hyper-edges incident to it, hence $\mathsf{Poisson}(\gamma u^{l-1})$ of them shall be from the core. Recall that a c-node belongs to the core iff at least one hyper-edge incident to it is in the core. By self-consistency, this yields the identity $u = 1 - \exp(-\gamma u^{l-1})$, or alternatively, $h_\rho(u) = 0$. As we have already seen, the existence of $u \in (0,1]$ for which $h_\rho(u) = 0$ is equivalent to $\rho \le \rho_{\rm d}$.

6.5. Diffusion approximation and scaling window

As mentioned before, the ODE asymptotics as in [62] is of limited value for decoding with code block-length that are relevant in practice. For this reason, [11] go one step further and using a diffusion approximation, provide the probability of successful decoding in the double limit of large size n and noise level approaching the critical value (i.e. taking $\rho_n \to \rho_{\rm d}$). The resulting asymptotic characterization is of finite-size scaling type.

Finite-size scaling has been the object of several investigations in statistical physics and in combinatorics. Most of these studies estimate the size of the corresponding scaling window. That is, fixing a small value of $\varepsilon > 0$, they find the amount of change in some control parameter which moves the probability of a relevant event from ε to $1-\varepsilon$. A remarkably general result in this direction is the rigorous formulation of a 'Harris criterion' in [22, 86]. Under mild assumptions, this implies that the scaling window has to be at least $\Omega(n^{-1/2})$ for a properly defined control parameter (for instance, the ratio ρ of the number of nodes to hyper-edges in our problem). A more precise result has recently been obtained for the satisfiable-unsatisfiable phase transition for the random 2-SAT problem, yielding a window of size $\Theta(n^{-1/3})$ [18]. Note however that statistical physics arguments suggest that the phase transition we consider here is not from the same universality class as the satisfiable-unsatisfiable transition for random 2-SAT problem.

If we fix $\rho > 0$, the fluctuations of $\vec{z}(n\theta)$ around $n\vec{y}(\theta)$ are accumulated in $n\theta$ stochastic steps, hence are of order $\sqrt{n}$. Further, applying the classical Stroock-Varadhan martingale characterization technique, one finds that the rescaled variable $(\vec{z}(n\theta) - n\vec{y}(\theta))/\sqrt{n}$ converges in law as $n \to \infty$ to a Gaussian random variable whose covariance matrix $\mathbb{Q}(\theta;\rho) = \{Q_{ab}(\theta;\rho); 1 \le a,b \le 2\}$ is the symmetric positive definite solution of the ODE:

$$\frac{{\rm d}\mathbb{Q}(\theta)}{{\rm d}\theta} = \mathbb{G}(\vec{y}(\theta),\theta) + \mathbb{A}(\vec{y}(\theta),\theta)\mathbb{Q}(\theta) + \mathbb{Q}(\theta)\mathbb{A}(\vec{y}(\theta),\theta)^T \tag{6.21}$$

(c.f. [11]). Here $\mathbb{A}(\vec{x},\theta) \equiv \{A_{ab}(\vec{x},\theta) = \partial_{x_b} F_a(\vec{x},\theta)\,;\ 1 \le a,b \le 2\}$ is the matrix of derivatives of the drift term for the mean ODE (6.20) and $\mathbb{G}(\vec{x},\theta) = \{G_{ab}(\vec{x},\theta) : a,b \in \{1,2\}\}$ is the covariance of $\Delta\vec{z}$ at $(\vec{x},\theta)$ under the transition kernel (6.13). That is, the non-negative definite symmetric matrix with entries

$$\left\{\begin{array}{lcl} G_{11}(\vec{x},\theta) &=& (l-1)[\mathfrak{p}_0 + \mathfrak{p}_1 - (\mathfrak{p}_0 - \mathfrak{p}_1)^2]\,, \\ G_{12}(\vec{x},\theta) &=& -(l-1)[\mathfrak{p}_0\mathfrak{p}_1 + \mathfrak{p}_1(1-\mathfrak{p}_1)]\,, \\ G_{22}(\vec{x},\theta) &=& (l-1)\mathfrak{p}_1(1-\mathfrak{p}_1) \end{array}\right. \tag{6.22}$$

The dependence of $\mathbb{Q}(\theta) \equiv \mathbb{Q}(\theta;\rho)$ on ρ is via the positive definite initial condition $\mathbb{Q}(0;\rho)$ of (6.18) for the ODE (6.21) as well as the terms $\vec{y}(\theta) = \vec{y}(\theta;\rho)$ that appear in its right side.

Focusing hereafter on the critical case $\rho = \rho_{\rm d}$, there exists then a unique *critical time* $\theta_{\rm d} \equiv \theta_*(\rho_{\rm d})$ in $(0,1)$ with $y_1(\theta_{\rm d}) = y_1'(\theta_{\rm d}) = 0$ and $y_1''(\theta_{\rm d}) > 0$, while the smooth solution $\theta \mapsto y_1(\theta;\rho_{\rm d})$ is positive when $\theta \neq \theta_{\rm d}$ and $\theta \neq 1$ (for more on $\vec{y}(\cdot;\cdot)$ see [32, Proposition 4.2]).

For $\rho_n = \rho_{\rm d} + rn^{-1/2}$ the leading contribution to $P_l(n,\rho_n)$ is the probability $\widehat{\mathbb{P}}_{n,\rho_n}(z_1(n\theta_{\rm d}) \leq 0)$ for the inhomogeneous Markov chain $\vec{z}(\tau)$ on $\mathbb{Z}_+^2$ with transition kernel $\widehat{W}_{\tau/n}(\Delta\vec{z}|n^{-1}\vec{z})$ of (6.13) and the initial distribution $\mathsf{G}_2(\cdot|n\vec{y}(0);n\mathbb{Q}(0))$ at $\rho = \rho_n$. To estimate this contribution, note that $y_1(\theta_{\rm d};\rho_{\rm d}) = 0$, hence

$$y_1(\theta_{\rm d};\rho_n) = rn^{-1/2}[\frac{\partial y_1}{\partial \rho}(\theta_{\rm d};\rho_{\rm d}) + o(1)]\,.$$

Thus, setting $a_l \equiv \frac{\partial y_1}{\partial \rho}/\sqrt{Q_{11}}$, both evaluated at $\theta = \theta_{\rm d}$ and $\rho = \rho_{\rm d}$, by the preceding Gaussian approximation

$$P_l(n,\rho_n) = \widehat{\mathbb{P}}_{n,\rho_n}(z_1(n\theta_{\rm d}) \leq 0) + o(1) = \mathsf{G}_1(-ra_l) + o(1)\,, \tag{6.23}$$

as shown in [11]. In particular, the phase transition scaling window around $\rho = \rho_{\rm d}$ is of size $\Theta(n^{-1/2})$.

In a related work, [27] determine the asymptotic core size for a random hyper-graph from an ensemble which is the 'dual' of $\mathcal{G}_l(n,m)$. In their model the hyper-edges (i.e. v-nodes) are of random, Poisson distributed sizes, which allows for a particularly simple Markovian description of the peeling algorithm that constructs the core. Dealing with random hyper-graphs at the critical point, where the asymptotic core size exhibits a discontinuity, they describe the fluctuations around the deterministic limit via a certain linear SDE. In doing so, they heavily rely on the powerful theory of weak convergence, in particular in the context of convergence of Markov processes. For further results that are derived along this line of reasoning, see [26, 45, 46].

6.6. Finite size scaling correction to the critical value

In contrast with the preceding and closer in level of precision to that for the scaling behavior in the emergence of the giant component in Erdös-Rényi random graphs (see [53] and references therein), for $\mathcal{G}_l(n,m)$ and $\rho_n = \rho_{\rm d} + rn^{-1/2}$ inside the scaling window, it is conjectured in [11] and proved in [32] that the leading correction to the diffusion approximation for

$P_l(n,\rho_n)$ is of order $\Theta(n^{-1/6})$. Comparing this finite size scaling expression with numerical simulations, as illustrated in [32, Figure 2], we see that it is very accurate even at $n \approx 100$.

Such finite size scaling result is beyond the scope of weak convergence theory, and while its proof involve delicate coupling arguments, expanding and keeping track of the rate of decay of approximation errors (in terms of n), similar results are expected for other phase transitions within the same class, such as k-core percolation on random graphs (with $k \geq 3$), or the pure literal rule threshold in random k-SAT (with $k \geq 3$, c.f. [41]). In a different direction, the same approach provides rates of convergence (in the sup-norm) as n grows, for distributions of many inhomogeneous Markov chains on $\mathbb{R}^d$ whose transition kernels $W_{t,n}(x_{t+1} - x_t = y | x_t = x)$ are approximately (in n) linear in x, and "strongly-elliptic" of uniformly bounded support with respect to y.

As a first step in proving the finite size scaling, the following refinement of the left hand side of (6.23) is provided in [32, Section 5].

Proposition 6.6. *Let $w \in (3/4, 1)$, $J_n = [n\theta_{\rm d} - n^w, n\theta_{\rm d} + n^w]$ and $|\rho - \rho_{\rm d}| \leq n^{w'-1}$ with $w' < 2w - 1$. Then, for $\varepsilon_n = A \log n$ and $\delta_n = D\, n^{-1/2} (\log n)^2$,*

$$\widehat{\mathbb{P}}_{n,\rho}\Big\{ \inf_{\tau \in J_n} z_1(\tau) \leq -\varepsilon_n \Big\} - \delta_n \leq P_l(n,\rho) \leq \widehat{\mathbb{P}}_{n,\rho}\Big\{ \inf_{\tau \in J_n} z_1(\tau) \leq \varepsilon_n \Big\} + \delta_n \,. \tag{6.24}$$

At the critical point (i.e. for $\rho = \rho_{\rm d}$ and $\theta = \theta_{\rm d}$) the solution of the ODE (6.20) is tangent to the $y_1 = 0$ plane and fluctuations in the y_1 direction determine whether a non-empty (hence, large), core exists or not. Further, in a neighborhood of $\theta_{\rm d}$ we have $y_1(\theta) \simeq \frac{1}{2}\widetilde{F}(\theta - \theta_{\rm d})^2$, for the positive constant

$$\widetilde{F} \equiv \frac{{\rm d}^2 y_1}{{\rm d}\theta^2}(\theta_{\rm d}; \rho_{\rm d}) = \frac{{\rm d}F_1}{{\rm d}\theta}(\vec{y}(\theta_{\rm d}; \rho_{\rm d}), \theta_{\rm d}) = \frac{\partial F_1}{\partial \theta} + \frac{\partial F_1}{\partial y_2} F_2 \tag{6.25}$$

(omitting hereafter arguments that refer to the critical point). In the same neighborhood, the contribution of fluctuations to $z_1(n\theta) - z_1(n\theta_{\rm d})$ is approximately $\sqrt{\widetilde{G} n |\theta - \theta_{\rm d}|}$, with $\widetilde{G} = G_{11}(\vec{y}(\theta_{\rm d}; \rho_{\rm d}), \theta_{\rm d}) > 0$. Comparing these two contributions we see that the relevant scaling is $X_n(t) = n^{-1/3}[z_1(n\theta_{\rm d} + n^{2/3}t) - z_1(n\theta_{\rm d})]$, which as shown in [32, Section 6] converges for large n, by strong approximation, to $X(t) = \frac{1}{2}\widetilde{F} t^2 + \sqrt{\widetilde{G}} W(t)$, for a standard two-sided Brownian motion $W(t)$ (with $W(0) = 0$). That is,

Proposition 6.7. *Let $\xi(r)$ be a normal random variable of mean $\left(\frac{\partial y_1}{\partial \rho}\right) r$ and variance Q_{11} (both evaluated at $\theta = \theta_{\rm d}$ and $\rho = \rho_{\rm d}$), which is independent of $W(t)$.*

For some $w \in (3/4, 1)$, any $\eta < 5/26$, all $A > 0$, $r \in \mathbb{R}$ and n large enough, if $\rho_n = \rho_{\rm d} + r\, n^{-1/2}$ and $\varepsilon_n = A \log n$, then

$$\left| \widehat{\mathbb{P}}_{n,\rho_n}\{ \inf_{\tau \in J_n} z_1(\tau) \le \pm \varepsilon_n\} - \mathbb{P}\{n^{1/6}\xi + \inf_t X(t) \le 0\} \right| \le n^{-\eta}\,. \qquad (6.26)$$

We note in passing that within the scope of weak convergence Aldous pioneered in [8] the use of Brownian motion with quadratic drift (ala $X(t)$ of Proposition 6.7), to examine the near-critical behavior of the giant component in Erdös-Rényi random graphs, and his method was extended in [46] to the giant set of identifiable vertices in Poisson random hyper-graph models.

Combining Propositions 6.6 and 6.7 we estimate $P_l(n, \rho_n)$ in terms of the distribution of the global minimum of the process $\{X(t)\}$. The latter has been determined already in [50], yielding the following conclusion.

Theorem 6.8. *For $l \ge 3$ set $a_l = \frac{\partial y_1}{\partial \rho}/\sqrt{Q_{11}}$, $b_l = \frac{1}{\sqrt{Q_{11}}} \widetilde{G}^{2/3}\, \widetilde{F}^{-1/3}$ and $\rho_n = \rho_{\rm d} + r\, n^{-1/2}$. Then, for any $\eta < 5/26$*

$$P_l(n, \rho_n) = \mathsf{G}_1(-r a_l) + b_l \kappa\, \mathsf{G}_1'(-r a_l)\, n^{-1/6} + O(n^{-\eta})\,, \qquad (6.27)$$

for $\kappa \equiv \int_0^\infty [1 - \mathcal{K}(z)^2]\, {\rm d}z$ and an explicit function $\mathcal{K}(\cdot)$ (see [32, equation (2.17)]).

Proof outline. Putting together Propositions 6.6 and 6.7, we get that

$$P_l(n, \rho_n) = \mathbb{P}\left\{ n^{1/6}\xi + \inf_t X(t) \le 0 \right\} + O(n^{-\eta})\,.$$

By Brownian scaling, $X(t) = \widetilde{F}^{-1/3} \widetilde{G}^{2/3} \widetilde{X}(\widetilde{F}^{2/3} \widetilde{G}^{-1/3} t)$, where $\widetilde{X}(t) = \frac{1}{2}t^2 + \widetilde{W}(t)$ and $\widetilde{W}(t)$ is also a two sided standard Brownian motion. With $Z = \inf_t \widetilde{X}(t)$, and Y a standard normal random variable which is independent of $\widetilde{X}(t)$, we clearly have that

$$\begin{aligned} P_l(n, \rho_n) &= \mathbb{P}\left\{ n^{1/6} \left(\frac{\partial y_1}{\partial \rho}\right) r + n^{1/6} \sqrt{Q_{11}} Y + \widetilde{F}^{-1/3} \widetilde{G}^{2/3} Z \le 0 \right\} + O(n^{-\eta}) \\ &= \mathbb{E}\Big\{ \mathsf{G}_1\big(- r a_l - b_l n^{-1/6} Z\big) \Big\} + O(n^{-\eta})\,. \end{aligned} \qquad (6.28)$$

From [50, Theorem 3.1] we deduce that Z has the continuous distribution function $F_Z(z) = 1 - \mathcal{K}(-z)^2 \mathbb{I}(z < 0)$, resulting after integration by parts

with the explicit formula for $\kappa = -\mathbb{E}\, Z$ (and where [50, (5.2)] provides the explicit expression of [32, formula (2.17)] for $\mathcal{K}(x)$). Further, as shown in [32, proof of Theorem 2.3] all moments of Z are finite and the proof is thus completed by a first order Taylor expansion of $\mathsf{G}_1(\,\cdot\,)$ in (6.28) around $-ra_l$. $\square$

Remark 6.9. *The simulations in [32, Figure 2] suggest that the approximation of $P_l(n,\rho_n)$ we provide in (6.27) is more accurate than the $O(n^{-5/26+\epsilon})$ correction term suggests. Our proof shows that one cannot hope for a better error estimate than $\Theta(n^{-1/3})$ as we neglect the second order term in expanding $\Phi(-ra_l + Cn^{-1/6})$, see (6.28). We believe this is indeed the order of the next term in the expansion (6.27). Determining its form is an open problem.*

Remark 6.10. *Consider the (time) evolution of the core for the hyper-graph process where one hyper-edge is added uniformly at random at each time step. That is, n increases with time, while the number of vertices m is kept fixed. Let $S(n)$ be the corresponding (random) number of hyper-edges in the core of the hyper-graph at time n and $n_{\rm d} \equiv \min\{n : S(n) \geq 1\}$ the onset of a non-empty core. Recall that small cores are absent for a typical large random hyper-graph, whereas fixing $\rho < \rho_{\rm d}$ the probability of an empty core, i.e. $S(m/\rho) = 0$, decays in m. Thus, for large m most trajectories $\{S(n)\}$ abruptly jump from having no core for $n < n_{\rm d}$ to a linear in m core size at the random critical edge number $n_{\rm d}$. By the monotonicity of $S(n)$ we further see that $\mathbb{P}_m\{n_{\rm d} \leq m/\rho\} = P_l(\rho, m/\rho)$, hence Theorem 6.8 determines the asymptotic distribution of $n_{\rm d}$. Indeed, as detailed in [32, Remark 2.5], upon expressing n in terms of m in equation (6.27) we find that the distributions of $\widehat{n}_{\rm d} \equiv a_l(\rho_{\rm d} n_{\rm d} - m)/\sqrt{m/\rho_{\rm d}} + b_l \kappa \rho_{\rm d}^{1/6} m^{-1/6}$ converge point-wise to the standard normal law at a rate which is faster than $m^{-5/26+\epsilon}$.*

Remark 6.11. *The same techniques are applicable for other properties of the core in the 'scaling regime' $\rho_n = \rho_{\rm d} + r\, n^{-1/2}$. For example, as shown in [32, Remark 2.6], for $m = n\rho_n$ and conditional to the existence of a non-empty core, $(S(n) - n(1-\theta_{\rm d}))/n^{3/4}$ converges in distribution as $n \to \infty$ to $(4Q_{11}/\widetilde{F}^2)^{1/4}\, Z_r$ where Z_r is a non-degenerate random variable (whose density is explicitly provided there). In particular, the $\Theta(n^{1/2})$ fluctuations of the core size at fixed $\rho < \rho_{\rm d}$ are enhanced to $O(n^{3/4})$ fluctuations near the critical point.*

References

[1] Achlioptas, D. and Coja-Oghlan, A., *Algorithmic barriers from phase transitions*, 49th Annual Symposium on Foundations of Computer Science, Philadelphia, PA, 2008.

[2] Achlioptas, D. and Friedgut, E., *A sharp threshold for k-Colorability*, Rand. Struct. Alg. **14** (1999), 63–70.

[3] Achlioptas, D. and Friedgut, E., *Sharp thresholds of graph properties and the k-SAT Problem*, with an appendix by J. Bourgain, J. Amer. Math. Soc. **12** (1999), 1017–1054.

[4] Achlioptas D. and Naor, A., *The two possible values of the chromatic number of a random graph*, Proc. of ACM STOC, 2004.

[5] Achlioptas, D., Naor, A. and Peres, Y., *Rigorous location of phase transitions in hard optimization problems*, Nature **435** (2005), 759.

[6] Achlioptas, D. and Ricci-Tersenghi, F., *On the solution-space geometry of random constraint satisfaction problems*, Proc. 38th ACM Symposium on Theory Of Computing, Seattle (USA) 2006, 130–139.

[7] Aizenman, M. and Warzel, S., *The canopy graph and level statistics for random operators on tree*, Math. Phys. Anal. and Geom., **9** (2007), 291–333.

[8] Aldous, D., *Brownian excursions, critical random graphs and the multiplicative coalescent*, Ann. Prob. **25** (1997), 812–854.

[9] Aldous, D. and Steele, J. M., *The objective method: probabilistic combinatorial optimization and local weak convergence*, in *Probability on discrete structures*, H. Kesten ed., New York, 2004.

[10] Alon, N. and Spencer, J., *The probabilistic method*, Wiley, New York, 2000.

[11] Amraoui, A., Montanari, A., Richardson, T. and Urbanke, R., *Finite-length scaling for iteratively decoded LDPC ensembles*, submitted to IEEE Trans. on Information Theory (2004). Available on-line at `http://arxiv.org/abs/cs.IT/0406050`

[12] Aronson, J., Frieze, A. and Pittel, B.G., *Maximum matchings in sparse random graphs: Karp-Sipser re-visited*, Random Structures and Algorithms, **12** (1998), 111–177.

[13] Berger, N., Kenyon, C., Mossel, E. and Peres, Y., *Glauber dynamics on trees and hyperbolic graphs*, Prob. Theory Rel. Fields **131**, (2005), 311.

[14] Bhamidi, S., Rajagopal R. and Roch, S, *Network delay inference from additive metrics*, To appear in Random Structures Algorithms **35**, (2009).

[15] Bhatnagar, N., Vera, J., Vigoda, E. and Weitz, D., *Reconstruction for*

colorings on trees. To appear in SIAM J. of Disc. Math.

[16] Bleher, P. M., Ruiz, J. and Zagrebnov, V. A., *On the purity of limiting Gibbs state for the Ising model on the Bethe lattice*, J. Stat. Phys. **79** (1995), 473–482.

[17] Bollobás, B., *A probabilistic proof of an asymptotic formula for the Number of labeled regular graphs*, Eur. J. Combinatorics, **1** (1980), 311–316.

[18] Bollobás, B., Borgs, C., Chayes, J. T., Kim, J. H. and Wilson D. B, *The scaling window of the 2-SAT transition*, Rand. Struct. and Alg. **18** (2001), 201–256.

[19] Borgs, C., Chayes, J., Mossel, E. and Roch, S., *The Kesten-Stigum reconstruction bound is tight for roughly symmetric binary channels*, Proc. of IEEE FOCS, 2006.

[20] Brightwell, G. R. and Winkler, P., *A second threshold for the hard-core model on a Bethe lattice*, Rand. Struct. and Alg. **24** (2004), 303–314.

[21] Caracciolo, S., Parisi, G., Patarnello, S. and Sourlas, N., *3d Ising spin glass in a magnetic field and mean-field theory*, Europhys. Lett. **11** (1990), 783.

[22] Chayes, J. T., Chayes, L., Fisher, D. S. and Spencer, T. *Finite-size scaling and correlation lengths for disordered systems*, Phys. Rev. Lett. **57** (1986), 2999–3002.

[23] Cocco, S., Dubois, O., Mandler, J. and Monasson, R., *Rigorous decimation-based construction of ground pure states for spin glass models on random lattices*, Phys. Rev. Lett. **90** (2003), 47205.

[24] Cover, T. and Thomas, J. A., *Elements of information theory*, Wiley Interscience, New York, 1991.

[25] Creignou, N. and Daudé, H., *Satisfiability threshold for random XOR-CNF formulas*, Discrete Applied Math. **96-97** (1999), 41–53.

[26] Darling, R.W.R., Levin, D. and Norris, J., *Continuous and discontinuous phase transitions in hypergraph processes*, Random Structures Algorithms, **24** (2004), 397–419.

[27] Darling, R.W.R. and Norris, J., *Structure of large random hypergraphs*, Ann. Appl. Probab. **15** (2005), 125–152.

[28] Daskalakis, C., Mossel, E. and Roch, S., *Optimal phylogenetic reconstruction*, Proc. of ACM STOC, 2006.

[29] Dembo, A. and Montanari, A. *Ising models on locally tree-like graphs*, to appear in Ann. Appl. Probab. (2010).

[30] Dembo, A. and Montanari, A. *Bethe states for graphical models*, preprint (2008).

[31] Dembo, A. and Montanari, A., *unpublished*, (2008).

[32] Dembo, A. and Montanari, A., *Finite size scaling for the core of large random hypergraphs*, Ann. Appl. Probab. **18** (2008), to appear.
[33] Dorogovtsev, S. N., Goltsev A. V. and Mendes J. F. F., *Ising model on networks with an arbitrary distribution of connections*, Phys. Rev. E **66** (2002), 016104.
[34] Dyer, M.E., Frieze, A.M. and Pittel, B., *On the average performance of the greedy algorithm for finding a matching in a graph*, Ann. Appl. Probab. **3** (1993), 526–552.
[35] Dyer, M., Sinclair, A., Vigoda, E. and Weitz, D., *Mixing in time and space for lattice spin systems: A combinatorial view*, Rand. Struc. and Alg. **24** (2004), 461–479.
[36] Dyson, F. J. *Existence of a phase-transition in a one-dimensional Ising ferromagnet*, Commun. Math. Phys. **12** (1969), 91–107.
[37] Ellis, R. S. and Newman, C. M., *The statistics of Curie-Weiss models*, J. Stat. Phys. **19** (1978), 149–161.
[38] Evans, W., Kenyon, C., Peres, Y. and Schulman, L. J., *Broadcasting on trees and the Ising model*, Ann. Appl. Probab. **10** (2000), 410–433.
[39] Fisher, M. E. *Critical temperatures of aniostropic Ising lattices. II. General upper bounds*, Phys. Rev. **162** (1967), 480–485.
[40] Franz, S. and Parisi, G., *Recipes for metastable states in spin glasses*, J. Physique I **5** (1995), 1401.
[41] Frieze, A.M. and Suen, S., *Analysis of two simple heuristics on a random instance of k-SAT*, Journal of Algorithms **20** (1996), 312–355.
[42] Georgii, H.-O., *Gibbs measures and phase transition*, de Gruyter, Berlin, 1988.
[43] Gerschenfeld, A. and Montanari, A., *Reconstruction for models on random graphs*, 48nd Annual Symposium on Foundations of Computer Science, Providence, RI, 2007.
[44] Ginibre, J., *General formulation of Griffiths' inequalities*, Comm. Math. Phys., **16** (1970), 310–328.
[45] Goldschmidt, C. and Norris, J., *Essential edges in Poisson random hypergraphs*, Random Structures and Algorithms **24**, (2004), 381–396.
[46] Goldschmidt, C., *Critical random hypergraphs: the emergence of a giant set of identifiable vertices*, Ann. Prob. **33** (2005), 1573–1600.
[47] Griffiths, R. B., *Peierls proof of spontaneous magnetization in a two-dimensional Ising ferromagnet*, Phys. Rev. A **136** (1964), 437–439.
[48] Griffiths, R. B., Hurst, C. A. and Sherman, S., *Concavity of magnetization of an Ising ferromagnet in a positive external field*, J. Math. Phys. **11** (1970), 790–795.
[49] Grimmett, G. *Percolation*, Springer, New York, 1999.

[50] Groeneboom, P., *Brownian motion with a parabolic drift and Airy functions*, Probab. Th. Rel. Fields, **81** (1989), 79–109.

[51] Guerra, F. and Toninelli, F. L., *The high temperature region of the Viana-Bray diluted spin glass model*, J. Stat. Phys. **115** (2004), 531–555.

[52] Guionnet, A. and Zegarlinski, B., *Lectures on logarithmic Sobolev inequalities*, Séminaire de Probabilites de Strasbourg, **36** (2002), 1–134.

[53] Janson, S., Knuth, D. E., Luczak, T. and Pittel, B., *The birth of the giant component*, Rand. Struct. and Alg., **4** (1993), 231–358.

[54] Janson, S., Luczak, T. and Ruciński, A., *Random graphs*, John Wiley and sons, 2000.

[55] Karp, R.M. and Sipser, M., *Maximum matching in sparse random graphs*, Proc. 22nd Annual IEEE Symposium on Foundations of Computing (1981), 364–375.

[56] Kelly, F. P., *Stochastic models of computer communication systems*, J. R. Stat. Soc. Ser. B **47** (1985), 379–395.

[57] Krzakala, F. and Zdeborova, L., *Phase transitions in the coloring of random graphs*, Phys. Rev. **E 76** (2007), 031131.

[58] Krzakala, F., Montanari, A., Ricci-Tersenghi, F., Semerjian, G. and Zdeborova, L., *Gibbs states and the set of solutions of random constraint satisfaction problems*, Proc. Natl. Acad. Sci. **1004** (2007), 10318.

[59] Kurtz, T.G., *Solutions of ordinary differential equations as limits of pure Markov jump processes*, J. Appl. Probab. **7** (1970), 49–58.

[60] Leone, M., Vázquez, A., Vespignani A. and Zecchina, R., *Ferromagnetic ordering in graphs with arbitrary degree distribution*, Eur. Phys. J. B **28** (2002), 191–197.

[61] Liggett, T., *Interacting particle systems*, Springer Verlag, New York, 1985.

[62] Luby, M., Mitzenmacher, M., Shokrollahi, A. and Spielman, D., *Efficient erasure correcting codes*, IEEE Trans. Inform. Theory, **47** (2001), 569–584.

[63] Martin, J. B., *Reconstruction thresholds on regular trees*. Discr. Math. Theor. Comp. Sci. AC (2003), 191–204.

[64] Martinelli, F., Sinclair, A. and Weitz, D., *The Ising model on trees: boundary conditions and mixing time*, Proc. of IEEE FOCS, 2003.

[65] Mézard, M. and Montanari, A., *Information, physics and computation*, Oxford University Press, Oxford, 2009.

[66] Mézard, M. and Montanari, A., *Reconstruction on trees and spin glass transition*, J. Stat. Phys. **124** (2006), 1317–1350.

[67] Mézard, M., Mora, T. and Zecchina, R., *Clustering of solutions in the*

random satisfiability problem, Phys. Rev. Lett. **94** (2005), 197205.

[68] Mézard, M., Mora, T. and Zecchina, R., *Pairs of SAT assignments and clustering in random Boolean formulae*, submitted to Theoretical Computer Science (2005), available online at `http://arxiv.org/abs/cond-mat/0506053`.

[69] Mézard, M. and Parisi, G., *Thermodynamics of glasses: a first principles computation*, Phys. Rev. Lett. **82** (1999), 747–751.

[70] Mézard, M., Parisi, G. and Virasoro, M. A., *Spin glass theory and beyond*, World Scientific, 1987.

[71] Mézard, M., Parisi, G. and Zecchina, R., *Analytic and Algorithmic Solution of Random Satisfiability Problems*, Science **297** (2002), 812–815.

[72] Mézard, M., Ricci-Tersenghi, F. and Zecchina, R., *Two solutions to diluted p-spin models and XORSAT problems*, J. Stat. Phys. **111** (2003), 505–533.

[73] Molloy, M., *Cores in random hypergraphs and Boolean formulas*, Rand. Struct. and Alg., **27** (2005), 124–135.

[74] Montanari, A., Restrepo, R. and Tetali, P., *Reconstruction and clustering in random constraint satisfaction problems*, `arXiv:0904.2751`, submitted to SIAM J. of Discr. Math.

[75] Montanari, A. and Semerjian, G., it Rigorous inequalities between length and time scales in glassy systems, J. Stat. Phys. **125** (2006), 23–54.

[76] Mossel, E. and Peres, Y., *Information flow on trees*, Ann. Appl. Probab. **13** (2003), 817–844.

[77] Mossel, E., Weitz, D. and Wormald, N., *On the hardness of sampling independent sets beyond the tree threshold*, Prob. Theory Rel. Fields **142** (2008), 401-439

[78] Mulet, R., Pagnani, A., Weigt, M. and Zecchina, R., *Coloring random graphs* Phys. Rev. Lett. **89** (2002), 268701.

[79] Orlitsky, A., Viswanathan, K. and Zhang, J., *Stopping set distribution of LDPC code ensembles*, IEEE Trans. Inform. Theory, **51** (2005), 929–953.

[80] Pittel, B., Spencer, J. and Wormald, N., *Sudden emergence of a giant k-core in a random graph*, J. of Comb. Theory, B **67** (1996), 111–151.

[81] Richardson, T. and Urbanke, R., *Modern coding theory*, Cambridge University Press, Cambridge, 2008.

[82] Simon, B., *Correlation inequalities and the decay of correlations in ferromagnets*, Commun. Math Phys. **77** (1980), 111–126.

[83] Sly, A., *Reconstruction of random colourings*, Commun. in Math. Phys. **288** (2008), 943–961

[84] Tatikonda, S. and Jordan, M. I., *Loopy belief propagation and Gibbs measures*, Uncertainty in Artificial Intelligence, 2002.

[85] Weigt, M. and Hartmann, A. K., *Minimal vertex covers on finite-connectivity random graphs: A hard-sphere lattice-gas picture*, Phys. Rev. E **63**, (2001), 056127.

[86] Wilson, D. B., *On the critical exponents of random k-SAT*, Rand. Struct. and Alg. **21** (2002), 182–195.