Gibbs Measures and Phase Transitions on Sparse Random Graphs

Amir Dembo and Andrea Montanari

Stanford University

July 16, 2008

Contents

Chapter 1

Introduction

A stochastic process is a collection of random variables $\{X_i : i \in V\}$. For our purposes, a stochastic process on a graph G is a collection of random variables indexed by the vertices of the graph, whose distribution is 'structured' according to the the graph itself.

This is of course vague, but we shall provide a more precise definition below. We shall start with the simplest possible example, the Curie-Weiss model and use it to illustrate the point of view we will take on stochastic processes on graphs. Then we shall move on to a list of more interesting (and motivating) examples.

1.1 The Curie-Weiss model

The Curie-Weiss model is deceivingly simple, but is a good pretext for explaining what are we talking about.

1.1.1 A story about opinion formation

At time 0, each of N individuals takes one of two opinions $x_i \in \{+1, -1\}$ independently and uniformly at random for $i \in \{1, \dots, N\}$. At each subsequent time, one individual i chosen uniformly at random considers whether changing its opinion. If the majority of the other individuals disagree with her, then she changes her opinion. This is just *conformism*.

In the other case, she takes an *anti-conformist* behavior with probability that becomes larger when the aggregate opinion is close to neutral. To be definite, she computes the opinion imbalance

$$M \equiv \sum_{j=1}^{N} x_j \,. \tag{1.1}$$

and $M^{(i)} \equiv M - x_i$. Then she changes her opinion with probability

$$p_{\text{flip}}(x) = \exp\left\{-2\beta |M^{(i)}|/N\right\} \,. \tag{1.2}$$

Her choice depends on the history so far only through the current value of $M^{(i)}$.

Let $M(t)$ be the opinion imbalance after t steps. Despite its simplicity, this model raises several interesting questions.

(a) How long does is take for the process to become approximately stationary?

(b) How often do individuals change opinion in stationary state?

(c) Is the typical opinion pattern strongly polarized (*herding*)?

(d) If this is the case, how often does the popular opinion change?

These lectures are not concerned with question 1, but with (some version of) questions 2 to 4.

To be more precise, notice that the above dynamics is an aperiodic irreducible Markov chain whose (unique) stationary state is

$$\mu_{N,\beta}(x) = \frac{1}{Z_N(\beta)} \exp\left\{ \frac{\beta}{N} \sum_{(i,j)} x_i x_j \right\} . \tag{1.3}$$

To prove this it is sufficient to check that the above dynamics is *reversible* with respect to the measure $\mu_{N,\beta}$, i.e. that $\mu_{N,\beta}(x)\mathbb{P}(x \to x') = \mu_{N,\beta}(x')\mathbb{P}(x' \to x)$ for any two configurations x, x' (where $\mathbb{P}(x \to x')$ is the one-step transition probability).

In writing $\mu_{N,\beta}$, we emphasized the dependence of this distribution on N (population size) and β (interaction strength). We will be particularly interested in the large-N behavior, and its dependence on β.

Then we can ask the following 'static' version of the above questions

2'. What is the distribution of $p_{\text{flip}}(x)$ when x has distribution $\mu_{N,\beta}(\,\cdot\,)$.

3'. What is the distribution of the opinion imbalance M. Is it concentrated near 0 (evenly spread opinions), or far from 0 (herding)?

4'. In the herding case: how unlikely are balanced ($M \approx 0$) configurations?

Exercise 1: Which measures replaces $\mu_{N,\beta}(\,\cdot\,)$ of (1.3) if the flipping probability is a general function of $M^{(i)}$ and possibly x_i but nothing else?

[You are not required to solve all the exercises in these notes. Some of them are pretty difficult and represent suggestions for those students who are interested in a deeper understanding of the material.]

1.1.2 Stochastic processes on graphs

Recall that a graph $G = (V, E)$ is defined by a set of vertices V and of edges E (an edge being an unordered pair of vertices.) We shall always assume G to be finite with $|V| = N$ and often make the identification $V = [N]$.

In these lectures a **stochastic process on graph** G is a process, i.e. a collection of random variables, indexed by the vertices of G, $X = \{X_i : i \in V\}$. We shall further assume that the joint distribution $\mu(x) = \mathbb{P}\{X = x\}$ of such variables factorizes according to G. By this we mean that there exist non-negative weights ψ_{ij} such that

$$\mu(x) = \frac{1}{Z} \prod_{(i,j)\in E} \psi_{ij}(x_i, x_j)\,. \tag{1.4}$$

Finally we shall restrict our attention to the case of variables X_i taking value in a finite alphabet $\mathcal{X}$. It is not hard to see that the distribution (1.3) takes this form (with G being the complete graph over N vertices.)

Rather than studying stochastic processes on graphs in this generality, we shall mostly focus on a few concepts/tools that have been object of recent research effort.

Coexistence. Roughly speaking, we will say that a odel shows chexistence if the measure μ decomposes into the convex combination of well separated lumps. In order to formalize this notion, we consider sequences of measures μ_N on graphs $G_N = ([N], E_N)$, and say that a coexistence occurs if, for each N, there exists a partition of the configuration space into subsets $\Omega_{1,N}, \dots, \Omega_{r,N}$ such that the following happens

(a) The measure of any of the subsets in the partition is bounded away from 1:

$$\max\{\mu_N(\Omega_{1,N}), \dots, \mu_N(\Omega_{r,N})\} \le 1 - \delta\,. \tag{1.5}$$

(b) The subsets are separated by 'bottlenecks.' More precisely, for $\Omega \subseteq \mathcal{X}^N$, define its ϵ-boundary as

$$\partial_\epsilon \Omega \equiv \{x \in \mathcal{X}^N : 1 \le d(x, \Omega) \le N\epsilon\}\,. \tag{1.6}$$

where d is the Hamming distance[1]. Then we require

$$\frac{\mu_N(\partial_\epsilon \Omega_{s,N})}{\mu_N(\Omega_{s,N})(1 - \mu_N(\Omega_{s,N}))} \to 0\,, \tag{1.7}$$

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

for some $\epsilon > 0$ and all $s \in \{1, \dots, r\}$. The normalization by $\mu_N(\Omega_{s,N})$ is introduced to avoid false bottlenecks due to small $\Omega_{s,N}$. The term $1 - \mu_N(\Omega_{s,N})$ is there just for the sake of symmetry.

Depending on the circumstances, one can require the above limit to be approached at some specific rate.

Often we shall consider phamilies 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 transitio can separate values of the parameter such that coexistence does occur from values such that it does not.

Mean field models. Again roughly speaking, mean field models are models that lack any (finite-dimensional) geometrical structure. For instance, models on the complete graphs or on standard random graphs are mean field. On the other hand, models on (finite portions) of finite dimensional grids are not.

A particular class of mean field models is defined by the requirement that $\mu(x_1, \dots, x_N)$ is exchangeable.

A wider class is obtained by considering *random* distributions[2] μ. Given μ, consider k iid configurations $X^{(1)}, \dots, X^{(k)}$ each having distribution μ. These are called 'replicas' in statistical physics. The unconditional, joint distribution of these k-copies is

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

which we view as a distribution over $(\mathcal{X}^k)^N$. For the model to be mean field we require $\mu^{(k)}$ to be exchangeable with respect to permutations of the vertices indexes in $[N]$.

While such requirement is sufficient in 'natural' examples, there are examples of models that intuitively are not mean-field and yet meet the requirement. For instance, given a non-random measure μ, and a uniformly random permutation π, define $\mu_\pi(x_1, \dots, x_N) \equiv \mu_\pi(x_{\pi(1)}, \dots, x_{\pi(N)})$. Then μ_π meets the requirement. A satisfactory formalization of the intuitive notion of 'mean field model,' is an open problem.

Mean field models are a pretty restrictive class, but a rich array of phenomena can be studied in detail.

Mean field equations. The problem with the model (1.4) is that distinct variables can be correlated in very subtle ways. Nevertheless, mean field models

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

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 close system of equations for the local marginals that hold in the large size limit. Such equations allow to determine the local marginals up to (eventually) a finite multiplicity. Finding a good formalization of this notion is an open problem. We shall instead provide specific examples throughout the course.

1.1.3 The Curie-Weiss model: Phase transition

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 some materials different magnetic moments like to point in the same direction (as people like to have similar opinions). Physicists want to understand whether this interaction might lead to a macroscopic magnetization (imbalance), or not.

In order to study the model, it is convenient to generalize it slightly by introducing a linear term in the exponent ('magnetic field')

$$\mu_{N,\beta}(x) = \frac{1}{Z_N(\beta)} \exp\left\{ \frac{\beta}{N} \sum_{(i,j)} x_i x_j + h \sum_{i=1}^{N} x_i \right\} . \tag{1.9}$$

In this context $1/\beta$ is referred to as the 'temperature.' We shall always assume $\beta \geq 0$ (positive interaction) and (without loss of generality) $h \geq 0$.

The good question to ask for understanding the Curie-Weiss model (1.9) is: what is the distribution of the magnetization?

Lemma 1.1.1. *For $m \in [-1, +1]$, define*

$$\psi_\beta(m) = hm + \frac{1}{2}\beta m^2 + H\left(\frac{1+m}{2}\right) , \tag{1.10}$$

where $H(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function.

Let $M \in \{-N, -N+2, \dots, N-2, N\}$, and $X = (X_1, \dots, X_N)$ be a random configuration from the Curie-Weiss model. Then,

$$\frac{e^{-\beta/2}}{N+1} \frac{1}{Z_N(\beta)} e^{N\psi_\beta(M/N)} \leq \mathbb{P}\left\{ \sum_{i=1}^{N} X_i = M \right\} \leq \frac{1}{Z_N(\beta)} e^{N\psi_\beta(M/N)} . \tag{1.11}$$

[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.

Proof. It is immediate to see that

$$\mathbb{P}\left\{\sum_{i=1}^{N} X_i = M\right\} = \frac{1}{Z_N(\beta)} \binom{N}{(N+M)/2} \exp\left\{hM + \frac{\beta M^2}{2N} - \frac{1}{2}\beta\right\} . \quad (1.12)$$

Our thesis then follows by Stirling approximation of thre binomial coefficient, cf. [1], Theorem 12.1.3. □

A major role in this course is played by the free-entropy density (the term 'density' refers here to the fact that we are dividing by the number of variables),

$$\phi_N(\beta) = \frac{1}{N} \log Z_N(\beta) . \quad (1.13)$$

Lemma 1.1.2. *Let*

$$\phi_*(\beta) \equiv \sup\{\psi_\beta(m) : m \in [-1, 1]\} , \quad (1.14)$$

and $\phi_N(\beta)$ be the free entropy density of the Curie-Weiss model. Then, for all N large enough

$$\phi_*(\beta) - \frac{\beta}{2N} - \frac{1}{N}\log\{N(N+1)\} \le \phi_N(\beta) \le \phi_*(\beta) + \frac{1}{N}\log(N+1) . \quad (1.15)$$

Proof. The upper bound follows upon summing the upper bound in Eq. (1.11) over M. From the lower bound in the same equation, we get

$$\phi_N(\beta) \ge \max\left\{\psi_\beta(m) : m \in S_N\right\} - \frac{\beta}{2N} - \frac{1}{N}\log(N+1) . \quad (1.16)$$

where $S_N \equiv \{-1, -1+2/N, \dots, 1-2/N, 1\}$. A little calculus shows that maximum of $\psi_\beta(m)$ over the finite set S_N is not smaller that the maximum over the interval $[-1, +1]$ minus $(\log N)/N$, for all N large enough. □

Consider the optimization problem in Eq. (1.14). Since $\psi_\beta(m)$ is continuous in the interval $[-1, 1]$ and differentiable in its interior, with $\psi'_\beta(m) \to \pm\infty$ as $m \to \mp 1$, the maximum is achieved at points $m \in (-1, 1)$ such that $\psi'_\beta(m) = 0$. A direct calculation shows that this condition is equivalent to

$$m = \tanh(\beta m + h) . \quad (1.17)$$

It is not hard to study the solutions of this equation. Here we limit ourself to presenting the results.

For $\beta \le 1$, the equation admits a unique solution $m_*(\beta, h)$ increasing in h with $m_*(\beta, h) \downarrow 0$ as $h \downarrow 0$. Obviously $m_*(\beta, h)$ maximizes $\psi_\beta(m)$.

For $\beta \ge 1$ there exists $h_*(\beta) > 0$ continuously increasing in β with $\lim_{\beta\to 1} h_*(\beta) = 0$ such that the following happens. For $0 \le h < h_*(\beta)$, Eq. (1.17) admits three

distinct solutions $m_-(\beta,h) < m_0(\beta,h) \le 0 \le m_+(\beta,h)$. For $h = h_*(\beta)$ two of these solutions coincide $m_-(\beta,h) = m_0(\beta,h)$ and for $h > h_*(\beta)$ only the positive one $m_+(\beta,h)$ survives.

Further, for any $h > 0$, $m_\pm(\beta,h)$ correspond to local maxima of $\psi_\beta(m)$, while $m_0(\beta,h)$ is a local minimum. The global maximum coincides with $m_+(\beta,h)$, that we shall henceforth denote as $m_*(\beta,h)$.

At $h = 0$ (and always $\beta > 1$), $\psi_\beta(m)$ is an even function of m. As a consequence $m_0(\beta,0) = 0$ and $m_\pm(\beta,0) = \pm m_*(\beta,0)$.

The theorem below answers question 3' in Section 1.1.1 of these notes.

Theorem 1.1.3. *Let $m_*(\beta,h)$ be defined as above, X be a random configuration of the Curie-Weiss model and $\overline{X} \equiv N^{-1}\sum_{i=1}^N X_i$. For $h > 0$ or $h = 0$ and $\beta \le 1$, and for any $\varepsilon > 0$, there exists $C(\varepsilon) > 0$ such that, for all N large enough*

$$\mathbb{P}\left\{\left|\overline{X} - m_*(\beta,h)\right| \le \varepsilon\right\} \ge 1 - e^{-NC(\varepsilon)}\,. \tag{1.18}$$

For $h = 0$ and $\beta > 1$, and for any $\varepsilon > 0$, there exists $C(\varepsilon) > 0$ such that, for all N large enough

$$\mathbb{P}\left\{\left|\overline{X} - m_*(\beta,0)\right| \le \varepsilon\right\} = \mathbb{P}\left\{\left|\overline{X} + m_*(\beta,0)\right| \le \varepsilon\right\} \ge \frac{1}{2} - e^{-NC(\varepsilon)}\,. \tag{1.19}$$

Proof. Consider first the case $\beta \le 1$ or $h > 0$. Under this assumption, $\psi_\beta(m)$ has a non-degenerate maximum at $m = m_*(\beta,h)$. Then, by Lemma 1.1.1

$$\mathbb{P}\left\{|\overline{X} - m_*(\beta,h)| \ge \varepsilon\right\} \le \frac{1}{Z_N(\beta)}\,(N{+}1)\,\exp\Big\{N\max[\psi_\beta(m) :\, |m{-}m_*(\beta,h)| \ge \varepsilon]\Big\}\,.$$

Using Lemma 1.1.2 we get

$$\mathbb{P}\left\{|\overline{X} - m_*(\beta,h)| \ge \varepsilon\right\} \le (N{+}1)^3 e^{-\beta/2}\,\exp\Big\{N\max[\psi_\beta(m){-}\phi_*(\beta) :\, |m{-}m_*(\beta,h)| \ge \varepsilon]\Big\}\,.$$

whence Eq. (1.18) follows.

Equation (1.19) is proved analogously, using the symmetry of the model for $h = 0$. □

We just encountered our first example of phase transition.

Theorem 1.1.4. *The Curie-Weiss model shows coexistence if and only if $h = 0$ and $\beta > 1$.*

Proof. We will limit ourselves to the 'if' part of this statement: for $h = 0$, $\beta > 1$, the Curie-Weiss model shows coexistence. Consider the partition of the configuration space given by $\{+1,-1\}^V = \Omega_+ \cup \Omega_-$, whereby $\Omega_+ \equiv \{x : \sum_i x_i \ge 0\}$ and $\Omega_- \equiv \{x : \sum_i x_i < 0\}$. We have to check that such partition satisfies the conditions in Section 1.1.2.

It follows immediately from Eq. (1.19) that, choosing $\epsilon < m_*(\beta, 0)/2$, we have

$$\mu_{\beta,N}(\Omega_\pm) \geq \frac{1}{2} - e^{-CN}, \qquad \mu_{\beta,N}(\partial_\epsilon \Omega_\pm) \leq e^{-CN}\,, \tag{1.20}$$

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 the first example of a mean field model, the first example of phase transition, and also the first example of *mean field equation*, namely Eq. (1.17). In the present Section, we will rederive this equation, using a somewhat more general type of argument.

Before doing this, it is worth trying to 'interpret' Eq. (1.17) and verify that indeed it matches the general definition of a mean field equation in Section 1.1.2. Throughout this section we will assume not to be on the coexistence line $h = 0$, $\beta > 1$. It then follows from Theorem 1.1.3 that $\mathbb{E}\, X_i = \mathbb{E}\, \overline{X} \approx m_*(\beta, h)$ (we will use $\approx$ whenever we do not want to get into a precise mathematical definition.) Therefore, Eq. (1.17) can be rewritten as

$$\mathbb{E}\, X_i \approx \tanh\left\{ h + \frac{\beta}{N} \sum_{j \in V} \mathbb{E}\, X_j \right\}. \tag{1.21}$$

In agreement with our general description of mean field equations, this is a closed form relation between the local marginals under the measure μ.

In fact, (1.21) follows just from the fact that $\overline{X}$ concentrated in probability, and does not require such a fine control as in Theorem 1.1.3. This is interesting because in more complicate models, some bound on fluctuations of $\overline{X}$ might be available without the analogous of Theorem 1.1.3. We start by proving an auxiliary result.

Lemma 1.1.5. *Denote by $\mathbb{E}_{N,\beta}$ expectation with respect to the Curie-Weiss model with N variables at inverse temperature β (and magnetic field h). Let $\overline{X} = N^{-1}\sum_{i=1}^N X_i$ and $\beta' = \beta(1 + 1/N)$. Then, for any $i \in [N]$:*

$$|\mathbb{E}_{N+1,\beta'} X_i - \mathbb{E}_{N,\beta} X_i| \leq \beta \sinh(\beta + h) \mathrm{Var}_{N,\beta}(\overline{X}) \tag{1.22}$$

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

$$\mathbb{E}_{N+1,\beta'}\{F(X)\} = \frac{\mathbb{E}_{N,\beta}\{F(X)\cosh(h + \beta\overline{X})\}}{\mathbb{E}_{N,\beta}\{\cosh(h + \beta\overline{X})\}}\,. \tag{1.23}$$

Therefore

$$\begin{aligned} |\mathbb{E}_{N+1,\beta'}\{F(X)\} - \mathbb{E}_{N,\beta}\{F(X)\}| &\leq ||F||_\infty \sqrt{\mathrm{Var}(\cosh(h + \beta\overline{X}))} \\ &\leq ||F||_\infty \beta \sinh(h + \beta) \sqrt{\mathrm{Var}(\overline{X})}\,. \end{aligned}$$

Here the first inequality follows from $\cosh a \geq 1$ and Cauchy-Schwarz, and the second from the Lipschitz behavior of $x \mapsto \cosh(h + \beta x)$ together with $|\overline{X}| \leq 1$. □

The following Theorem implies (a formal version of) Eq. (1.21) for $\beta \leq 1$ or $h > 0$.

Theorem 1.1.6. *There exists a constant $C(\beta, h)$ such that*

$$\left| \mathbb{E} X_i - \tanh\left\{ h + \frac{\beta}{N} \sum_{j \in V} \mathbb{E}\, X_j \right\} \right| \leq C(\beta, h) \sqrt{\mathrm{Var}(\overline{X})}\,. \tag{1.24}$$

Proof. Let $\mathbb{E}_{N,\beta}$, and $\mathbb{E}_{N+1,\beta'}$ be defined as in Lemma 5 and $\overline{X} = N^{-1} \sum_{i=1}^{N} X_i$. By direct computation

$$\mathbb{E}_{N+1,\beta'}\{X_{N+1}\} = \frac{\mathbb{E}_{N,\beta} \sinh(h + \beta \overline{X})}{\mathbb{E}_{N,\beta} \cosh(h + \beta \overline{X})}\,. \tag{1.25}$$

Notice that (by Lipschitz property of $\cosh(h + \beta x)$ and $\sinh(h + \beta x)$ together with $|\overline{X}| \leq 1$)

$$|\mathbb{E}_{N,\beta} \cosh(h + \beta \overline{X}) - \cosh(h + \beta \mathbb{E}_{N,\beta} \overline{X})| \leq \sinh(\beta + h) \sqrt{\mathrm{Var}(\overline{X})}\,, \tag{1.26}$$

$$|\mathbb{E}_{N,\beta} \sinh(h + \beta \overline{X}) - \sinh(h + \beta \mathbb{E}_{N,\beta} \overline{X})| \leq \cosh(\beta + h) \sqrt{\mathrm{Var}(\overline{X})}\,. \tag{1.27}$$

Using the inequality $|a_1/b_1 - a_2/b_2| \leq |a_1 - a_2|/b_1 + a_2|b_1 - b_2|/b_1 b_2$ (for $a_i, b_i \geq 0$), the bound $|\overline{X}| \leq 1$, and the fact that $\mathbb{E}_{N+1,\beta'} X_i$ is independent of i, this implies

$$\left| \mathbb{E}_{N+1,\beta'}\{X_i\} - \tanh\left\{ h + \frac{\beta}{N} \sum_{j \in V} \mathbb{E}_{N,\beta}\, X_j \right\} \right| \leq C(\beta, h) \sqrt{\mathrm{Var}(\overline{X})}\,. \tag{1.28}$$

The thesis then follows by applying Lemma 1.1.5. □

Exercise 2: Repeat the derivations in this lecture for the model

$$\mu(x) = \frac{1}{Z} \exp\left\{ \frac{\beta}{N^{p-1}} \sum_{(i_1 \dots i_p)} x_{i(1)} \cdots x_{i(p)} + h \sum_{i=1}^{N} x_i \right\}, \tag{1.29}$$

where $p \geq 2$ is a fixed integer and the first sum runs over all the p-uples of distinct indices (we solved the case $p = 2$).

Exercise 3: Consider the model

$$\mu_z(x) = \frac{1}{Z} \exp\left\{ \frac{\beta}{N} \sum_{(i,j)} x_i x_j + h \sum_{i=1}^{N} z_i x_i \right\}, \tag{1.30}$$

where z_i's are iid standard gaussian random variables. What is the value of Z for a 'typical' realization of the z_i's? What about the the expectations $\mathbb{E}\{x_i\}$ for a typical realization of the z_i's?

1.2 Models

In this Section we'll list a number of examples of stochastic processes on graphs, coming from different domains. In all of these examples $G = (V, E)$ is a graph.

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 $x_i \in \{+1, -1\}$ and

$$\mu_{N,\beta}(\underline{x}) = \frac{1}{Z_N(\beta)} \exp\left\{ \beta \sum_{(i,j)\in E} x_i x_j + h \sum_{i\in V} x_i \right\}, \tag{1.31}$$

with $\beta \geq 0$. Interaction between vertices i, j connected by an edge pushes the variable x_i, x_j towards taking the same value. It is expected that this leads to a global alignment the variables (spins) at low temperature, for a large family of graphs. This transition should be analogous to the one of the Curie-Weiss model, but remarkably little is known about Ising models on *general graphs*. In the next chapter we'll consider the case of random sparse graphs.

Antiferromagnetic Ising model. The model takes the same form (1.31), but with $\beta < 0$ (although normaly one introduces explicitly a minus sign to keep β positive). If the graph is bipartite, and $h = 0$, the model is completely equivalent to the ferromagnetic one. On non-bipartite graphs the antiferromagnetic model is much more complicate than the ferromagnetic one, and indeed even determining the most likely (lowest energy) configuration is a difficult matter. For $h = 0$, the latter is indeed the celebrated max-cut problem.

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

$$\mu_{N,\beta}(\underline{x}) = \frac{1}{Z_N(\beta)} \exp\left\{ \beta \sum_{(i,j)\in E} J_{ij} x_i x_j + h \sum_{i\in V} x_i \right\}. \tag{1.32}$$

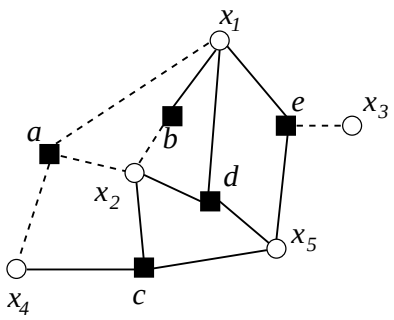

Figure 1.1: Factor graph representation of the satisfiability formula $(\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_4) \wedge (x_1 \vee \bar{x}_2) \wedge (x_2 \vee x_4 \vee x_5) \wedge (x_1 \vee x_2 \vee x_5) \wedge (x_1 \vee \bar{x}_2 \vee x_5)$. Edges are continuous or dashed depending whether the corresponding variable is directed or negated in the clause.

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 $N \to \infty$ of $\mu_{\beta,N}(\,\cdot\,)$ for typical realizations of the couplings.

1.2.2 Random constraint satisfaction problems

A constraint satisfaction problem (CSP) is defined by a finite variable domain $\mathcal{X}$, and by a class $\mathcal{C}$ of constraints over variables in $\mathcal{X}$. An instance of this problem is defined by an integer N (number of variables) and a set of M constraints from the class $\mathcal{C}$ over variables $x_1, \dots, x_N \in \mathcal{X}$. A solution of this instances is an assignment of the variables that satisfies all constraints.

Several questions are of interest within computer science:

(a) *Decision problem.* Does the system have a solution?

(b) *Optimization problem.* Maximize the number of satisfied constraints?

(c) *Counting problem.* Count the number of solutions.

There are many ways of associating a stochastic process on a graph to an instance of constraint satisfaction problem. If the instance admits a solution, a very simple idea is to consider the uniform measure over all such solutions. Let us see how does this work in specific examples.

Coloring. A proper q-coloring of a graph G is an assignment of colors in $\{1, \dots, q\}$ to the vertices of G in such a way that no edges has both endpoints of the same color. This defines a CSP, where the variable domain is $\mathcal{X} = \{1, \dots, q\}$ and the possible constraints are indexed by pair of indices $(i, j) \in V \times V$ and

include the two variables x_i and x_j. The constraint (i,j) isv satisfied if and only if $x_i \neq x_j$.

Assuming that graph G admits a proper coloring, the uniform measure reads

$$\mu_G(\underline{x}) = \frac{1}{Z_G} \prod_{(i,j)\in E} \mathbb{I}(x_i \neq x_j)\,. \tag{1.33}$$

Notice that Z_G is the number of proper colorings of G.

k-Satisfiability. In this case variables are binary $x_i \in \{0,1\}$, and possible constraints are indexed by k-uples of indices in $\{1,\dots,N\}$, $(i(1),\dots,i(k))$. Each constraint takes the form $(x_{i(1)},\dots,x_{i(k)}) \neq (x^*_{i(1)},\dots,x^*_{i(k)})$ for some k-uple $(x^*_{i(1)},\dots,x^*_{i(k)})$. In this context constraints are often referred to as 'clauses' and can be written as the disjuncton (logical OR) of k variables or their negations. The uniform measure over solutions is

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{a=1}^{M} \mathbb{I}\Big((x_{i_a(1)},\dots,x_{i_a(k)}) \neq (x^a_{i_a(1)},\dots,x^a_{i_a(k)})\Big)\,. \tag{1.34}$$

An instance can be associated to a *factor graph*, cf. Fig. 1.1. This is a bipartite graph including two types of nodes: variable nodes $V \equiv \{1,\dots,N\}$ for variables and function (or factor) nodes $F \equiv \{1,\dots,M\}$ for constraints. Variable node i and function node a are connected by an edge if and only if variable x_i appears in the a-th clause.

It is not hard to realize that the last construction can be generalized to arbitrary CSPs: one can represent an arbitrary CSP instance using a factor graph.

1.2.3 Communications, estimation, detection

A number of problems from mathematical engineering can be phrased in terms models on graphs. We will describe a canonical model for these systems: several specific applications can in fact be phrased in terms of this model.

Let $X_1,\dots,X_n$ be a collection of 'hidden' random variables that we shall assume iid with a common distribution $p_0(\,\cdot\,)$ over a finite alphabet $\mathcal{X}$. We are interested in estimating these variables from a collection of observations $Y_1,\dots,Y_M$. The a-th observation (for $a \in \{1,\dots,M\}$) is a random function of the X_i's with $i \in \partial a = \{i_a(1),\dots,i_a(k)\}$. By this we mean that Y_a is conditionally independent of all the other variables given $\{X_i : i \in \partial a\}$. We'll write

$$\mathbb{P}\left\{Y_a \in A | \underline{X}_{\partial a} = \underline{x}_{\partial a}\right\} = Q_a(A|\underline{x}_{\partial a})\,. \tag{1.35}$$

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

The *a posteriori* distribution of the hidden variables given the observations reads

$$\mu(\underline{x}|\underline{y}) = \frac{1}{Z(\underline{y})} \prod_{a=1}^{M} Q_a(y_a|\underline{x}_{\partial a}) \prod_{i=1}^{N} p_0(x_i)\,. \tag{1.36}$$

Exercise 4: Describe the (factor) graph associated with the distribution (1.36).

Exercise 5: Write the explicit form of the distribution (1.36) when the *a priori* distribution p_0 of X_i is uniform in $\{+1, -1\}$ and the observatios are defined by

$$\underline{Y} = A\,\underline{X} + \underline{W}\,, \tag{1.37}$$

with A a sparse $N \times M$ matrix (i.e. with few non-vanishing entries) and $\underline{W} = (W_1, \dots, W_N)$ a vector of N iid centered Gaussian variables.

1.2.4 Graph and graph ensembles

The properties of the models introduced so far depend on the strcture of the underlying graph $G = (V, E)$. Let us first summarize a few general notations. We shall denote vertices as $i, j, k, \dots \in V$, and edges as (i, j) with $i, j \in V$. Given a vertex $i \in V$, we let $\partial i = \{j \in V : (i, j) \in E\}$ be the set of its neighbors.

These lectures focus on (random) graph models without euclidean structure. Here are a few ensemble of such graphs.

1. *Random-regular graphs of degree k.* This is the ensemble of graph over N vertices with homogeneous degree k. The ensemble is endowed with the uniform measure.

2. *Random graphs.* This is the ensemble of graph with N vertices and $N\gamma$ edges, again endowed with the uniform measure. A slightly modified ensemble is he one in which each edge (i, j) is present independently with probability $N\gamma/\binom{N}{2}$.

3. *Radom graphs with given degree distribution.* Given $N \in \mathbb{N}$ and a distribution over the non-negative integers $\{P_l\}_{l\geq 0}$, one considers graphs with N vertices of which NP_0 have degree 0, NP_1 degree 1, NP_2 have degree 2, etc (assuming NP_l to be an integer for each l). A graph from this ensemble is again drawn uniformly at random.

Exercise 6: What is expected fraction of vertices of degree k in the random graph model? How does it behave for large k? What can you say about the *actual* fraction of vertices of degree k? Is it close to its expectation?

Among the important properties of these graphs, is that they converge locally to trees. Namely, for any integer ℓ, let $\mathsf{B}(i, \ell)$ denote the depth-ℓ neighborhood of

a uniformly random vertex i. Then $\mathsf{B}(i, \ell)$ converges in distribution to a random tree of depth ℓ.

Chapter 2

Ising models on locally tree-like graphs

In this Chapter we study ferromagnetic Ising models on graph sequences that converge locally to trees. The last section is a technical appendix. It contains a brief exposition of Griffiths inequalities which are an important tool in the study of ferromagnetic Ising models.

2.1 Introduction

An *Ising model on the finite graph* G (with vertex set V, and edge set E) has the Boltzmann distribution

$$\mu(\underline{x}) = \frac{1}{Z(\beta, B)} \exp\Big\{\beta \sum_{(i,j)\in E} x_i x_j + B \sum_{i\in V} x_i\Big\}, \tag{2.1}$$

over $\underline{x} = \{x_i : i \in V\}$, with $x_i \in \{+1, -1\}$, parametrized by the 'magnetic field' B and 'inverse temperature' $\beta \geq 0$, where the partition function $Z(\beta, B)$ is fixed by the normalization condition $\sum_{\underline{x}} \mu(\underline{x}) = 1$.

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* (also called free energy or pressure)

$$\phi(\beta, B) \equiv \lim_{n\to\infty} \frac{1}{n} \log Z_n(\beta, B)\,. \tag{2.2}$$

It is characterized in great detail [3] for the complete graph $G_n = K_n$ (scaling β by $1/n$ to get a non-trivial limit). Predictions exist for a much wider class of graphs. Most notably, sparse random graphs with bounded average degree that arise in a number of problems from combinatorics and theoretical computer science

Examples include random satisfiability, coloring of random graphs, and graph partitioning [2], where the uniform measure over solutions can be regarded as the Boltzmann distribution for multi-spin interactions. Such problems have been successfully attacked using non rigorous statistical mechanics techniques whose mathematical foundation is still lacking and of much interest. In sparse graphs, with the distance between vertices as the length of shortest path connecting them, one also has a new characterization of the measure (2.1) in terms of correlation decay, related to Gibbs measures on infinite trees [59].

The asymptotic free entropy density (2.2) was determined rigorously only in a few cases of sparse graphs. In [17], this was accomplished for random regular graphs and in [66] for random graphs with independent edges, but only at high or zero temperature (and with vanishing magnetic field). We generalize these results by considering generic graph sequences that converge locally to trees and control the free entropy density by proving that the Boltzmann measure (2.1) converges locally to the Boltzmann measure of a model on the appropriate infinite random tree (a philosophy related to that of [49]). Our results also have algorithmic interpretations, providing an efficient procedure for approximating the local marginals of the Boltzmann measure. The essence of this procedure consists in solving by iteration certain mean field (cavity) equations. Such an algorithm is known in artificial intelligence and computer science under the name of *belief propagation*.

Our emphasis is on the low-temperature regime where although uniform decorrelation does not hold, we show that belief propagation converges exponentially fast on any graph, and that the resulting estimates are asymptotically exact for large locally tree-like graphs. The main idea is to introduce a magnetic field to break explicitly the $+/-$ symmetry, and to carefully exploit the monotonicity properties of the model.

2.2 Locally tree-like graphs and free entropy

Let $P = \{P_k : k \geq 0\}$ be a probability distribution over the non-negative integers, with finite, positive first moment $\overline{P}$, and set $\rho_k = (k+1)P_{k+1}/\overline{P}$ of mean $\overline{\rho}$. We use $\mathbb{P}_\rho\{\,\cdot\,\}$ for the law of the random rooted Galton-Watson tree $\mathsf{T}(\rho, t)$ of $t \geq 0$ generations where independently of each other each node has offspring distribution $\{\rho_k\}$ and denote by $\mathsf{T}(P, \rho, t)$ the modified ensemble where the offspring distribution at the root is changed to P.

Definition 2.2.1. *Let $\mathbb{P}_n$ denote the law of the ball $\mathsf{B}_i(t)$ (i.e. the subgraph induced by vertices of G_n whose distance from i is at most t), centered at a uniformly chosen random vertex $i \in V_n$. We say that $\{G_n\}$* converges locally *to the random tree $\mathsf{T}(P, \rho, \infty)$ if, for any t, and any rooted tree T with t generations*

$$\lim_{n\to\infty} \mathbb{P}_n\{\mathsf{B}_i(t) \simeq T\} = \mathbb{P}_\rho\{\mathsf{T}(t) \simeq T\} \tag{2.3}$$

(where $T_1 \simeq T_2$ if two trees T_1 and T_2 of same fize are identical upon labeling their vertices in a breadth first fashion following lexicographic order among siblings). We also say that $\{G_n\}$ is uniformly sparse *if*

$$\lim_{l\to\infty} \sup_n \frac{1}{|V_n|} \sum_{i\in V_n} |\partial i|\, \mathbb{I}(|\partial i| \geq l) = 0\,, \tag{2.4}$$

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

The model (2.1) has a line of first order phase transitions for $B = 0$ and $\beta > \beta_{\rm c}$ (that is, where the continuous function $B \mapsto \phi(\beta, B)$ exhibits a discontinuous derivative). The critical temperature is determined by the condition $\overline{\rho}\,(\tanh\beta_{\rm c}) = 1$. The asymptotic free-entropy density is given in terms of the following fixed point distribution.

Lemma 2.2.2. *Consider the random variables $\{h^{(t)}\}$ where $h^{(0)} \equiv 0$ and for $t \geq 0$,*

$$h^{(t+1)} \stackrel{\rm d}{=} B + \sum_{i=1}^{K} \operatorname{atanh}[\tanh(\beta)\tanh(h_i^{(t)})]\,, \tag{2.5}$$

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 and converges in law to the unique fixed point h^* of (2.5) that is supported on $[0, \infty)$.

Our main result confirms the statistical physics prediction for the free entropy density.

Theorem 2.2.3. *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 $\mathsf{T}(P, \rho, \infty)$,*

$$\lim_{n\to\infty} \frac{1}{n} \log Z_n(\beta, B) = \phi(\beta, B)\,, \tag{2.6}$$

where taking L of distribution P independently of the 'cavity fields' h_i that are i.i.d. copies of the fixed point h^* of Lemma 2.2.2, $\phi(\beta, B) = \phi(\beta, -B)$ is given for $B > 0$ by

$$\begin{aligned}\phi(\beta, B) \equiv \frac{\overline{P}}{2} \log\cosh(\beta) - \frac{\overline{P}}{2}\, \mathbb{E} \log[1 + \tanh(\beta)\tanh(h_1)\tanh(h_2)] \\ + \mathbb{E}\, \log\Big\{ e^B \prod_{i=1}^{L} [1 + \tanh(\beta)\tanh(h_i)] + e^{-B} \prod_{i=1}^{L} [1 - \tanh(\beta)\tanh(h_i)] \Big\}\,,\end{aligned} \tag{2.7}$$

and $\phi(\beta, 0)$ is the limit of $\phi(\beta, B)$ as $B \to 0$.

The proof of Theorem 2.2.3 is based on two steps

(a) Reduce the computation of $\phi_n(\beta, B) = \frac{1}{n} \log Z_n(\beta, B)$ to computing expectations of local (in G_n) quantities with respect to the Boltzmann measure (2.1). This is achieved by noticing that the derivative of $\phi_n(\beta, B)$ with respect to β is a sum of such expectations.

(b) Show that expectations of local quantities on G_n are well approximated by the same expectations with respect to an Ising model on the associated tree $\mathsf{T}(P, \rho, t)$ (for t and n large.) This is proved by showing that, on such a tree, local expectations are insensitive to boundary conditions that dominate stochastically free boundaries. The thesis then follows by monotonicity arguments.

The key step is of course the last one with the challenge to carry it out above β_c, when we no longer have uniqueness of the Gibbs measure on $\mathsf{T}(P, \rho, \infty)$. Indeed, insensitivity to positive boundary conditions is proved for a large family of 'conditionally independent' trees. Beyond the random tree $\mathsf{T}(P, \rho, \infty)$, these include deterministic trees with bounded degrees and multi-type branching processes (so it allows for generalizing Theorem 2.2.3 to other graph sequences, that converge locally to random trees different from $\mathsf{T}(P, \rho, \infty)$).

2.3 Algorithmic implications

The free entropy density is not the only quantity that can be characterized for Ising models on locally tree-like graphs. Indeed, local marginals can be efficiently computed with good accuracy. The basic idea is to solve a set of mean field equations iteratively. These are known as Bethe-Peierls or cavity equations and the corresponding algorithm is referred to as 'belief propagation' (BP).

More precisely, 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

$$\nu^{(t+1)}_{i\to j}(x_i) = \frac{1}{z^{(t)}_{i\to j}} e^{Bx_i} \prod_{l\in\partial i\setminus j} \sum_{x_l} e^{\beta x_i x_l} \nu^{(t)}_{l\to i}(x_l) \tag{2.8}$$

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

We establish uniform exponential convergence of the BP iteration to the same fixed point of (2.8), irrespective of its positive initial condition.

Theorem 2.3.1. *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 (2.8) such that for*

any positive initial condition $\{\nu^{(0)}_{l\to k}\}$ *and all* $t \geq 0$,

$$\sup_{(i,j)\in E} \|\nu^{(t)}_{i\to j} - \nu^{*}_{i\to j}\|_{\mathrm{TV}} \leq A\exp(-\lambda t)\,. \tag{2.9}$$

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

$$||\mu_U - \nu_U||_{\mathrm{TV}} \leq \exp\Big\{c^{r+1} - \lambda(t-r)\Big\}\,, \tag{2.10}$$

where $\mu_U(\,\cdot\,)$ *is the law of* $\underline{x}_U \equiv \{x_i : i \in U\}$ *under the Ising model (2.1) and* ν_U *the probability distribution*

$$\nu_U(\underline{x}_U) = \frac{1}{z_U}\exp\Big\{\beta\sum_{(i,j)\in E_U} x_i x_j + B\sum_{i\in U\setminus\partial U} x_i\Big\}\prod_{i\in\partial U}\nu^{*}_{i\to j(i)}(x_i)\,, \tag{2.11}$$

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

2.3.1 Examples

Many common random graph ensembles naturally fit our framework.

Random regular graphs. Let G_n be a uniformly random graph with degree k. As $n \to \infty$, the sequence $\{G_n\}$ is obviously uniformly sparse, and converges locally almost surely to the random rooted Cayley tree of degree k. Therefore, in this case Theorem 2.2.3 applies with $P_k = 1$. The distributional recursion (2.5) then evolves with a deterministic sequence $h^{(t)}$ recovering the result of [17].

Erdös-Renyi graphs. Let G_n be a uniformly random graph with $m = n\gamma$ edges over n vertices. The sequence $\{G_n\}$ converges locally almost surely to a Galton-Watson tree with Poisson offspring distribution of mean 2γ. This corresponds to taking $P_k = (2\gamma)^k e^{-2\gamma}/k!$. The same happens to classical variants of this ensemble. For instance, one can add an edge independently for each pair (i,j) with probability $2\gamma/n$, or consider a multi-graph with Poisson$(2\gamma/n)$ edges between each pair (i,j). In all these cases $\{G_n\}$ is almost surely uniformly sparse. In particular, Theorem 2.2.3 extends the results of [66] to arbitrary non-zero temperature and magnetic field.

Arbitrary degree distribution. Let P be a distribution with finite second moment and G_n a uniformly random graph with degree distribution P (the number of vertices of degree k is obtained by rounding nP_k). Then $\{G_n\}$ is almost surely uniformly sparse and converges locally to $\mathsf{T}(P,\rho,\infty)$. The same happens if G_n is drawn according to the so-called configuration model.

2.4 Ising models on trees

We extend (2.1) by allowing for vertex-dependent magnetic fields B_i, namely, we consider

$$\mu(\underline{x}) = \frac{1}{Z(\beta, \underline{B})} \exp\Big\{\beta \sum_{(i,j)\in E} x_i x_j + \sum_{i\in V} B_i x_i\Big\} \tag{2.12}$$

and derive correlation decay results for Ising models on trees, which are of independent interest. More precisely, let T denote a *conditionally independent* infinite tree rooted at the vertex $\emptyset$. That is, 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. In addition to $\mathsf{T} = \mathsf{T}(P, \rho, \infty)$ this flexible framework accommodates for example random bipartite trees, deterministic trees of bounded degree and percolation clusters on them.

For each $\ell \geq 1$, we denote by $\mu^{\ell,0}$ the Ising model (2.12) 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 boudary conditions), using μ^{ℓ} for statements that apply to both free and plus boundary conditions.

Theorem 2.4.1. *Suppose T is a conditionally independent infinite tree of average offspring numbers bounded by Δ. 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 $0 < B_{\min} \leq B_{\max}$, $\beta_{\max}$ and Δ finite, such that if $B_i \leq B_{\max}$ for all $i \in \mathsf{T}(r-1)$ and $B_i \geq B_{\min}$ for all $i \in \mathsf{T}(\ell)$, then for any $r \leq \ell$ and $\beta \leq \beta_{\max}$,*

$$\mathbb{E}\Big\{\sum_{i\in\partial\mathsf{T}(r)} \langle x_\emptyset; x_i\rangle_\emptyset^{(\ell)}\Big\} \leq A\, e^{-\lambda r}\,. \tag{2.13}$$

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

$$\mathbb{E}\,||\mu^{\ell,+}_{\mathsf{T}(r)} - \mu^{\ell,0}_{\mathsf{T}(r)}||_{\text{TV}} \leq A e^{-\lambda(\ell-r)}\,\mathbb{E}\{C^{|\mathsf{T}(r)|}\}\,. \tag{2.14}$$

In deriving this theorem we rely on monotonicity properties of the Ising measure, such as Griffiths inequality (see [41, Theorem IV.1.21]), and the GHS inequality (see [18]) about the effect of the magnetic field $\underline{B}$ on the local magnetizations at various vertices. We also extend Simon's inequality about the (centered) two point correlation functions in ferromagnetic Ising models with

zero magnetic field (see [67, Theorem 2.1]), to arbitrary magnetic field, in the case of Ising models on trees. That is, we show 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

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

2.5 From trees to graphs: proof of Theorem 2.2.3.

We have the following consequence of the local convergence of the graph sequence $\{G_n\}$.

Lemma 2.5.1. *Suppose a uniformly sparse graph sequence $\{G_n\}$ converges locally to the random tree $\mathsf{T}(P,\rho,\infty)$. Fixing a non-negative integer t, for each $(i,j) \in E_n$ denote the subgraph of G_n induced by vertices at distance at most t from (i,j) by $\mathsf{B}_{ij}(t)$. Let $F(\cdot)$ be a fixed, bounded function on the collection of all possible subgraphs that may occur as $\mathsf{B}_{ij}(t)$, such that $F(T_1) = F(T_2)$ whenever $T_1 \simeq T_2$. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{(i,j)\in E_n} F(\mathsf{B}_{ij}(t)) = \frac{\overline{P}}{2}\, \mathbb{E}\{F(\overline{\mathsf{T}}(\rho,t))\} \,, \tag{2.16}$$

for the random tree $\overline{\mathsf{T}}(\rho,t)$ obtained by 'gluing' two independent copies of $\mathsf{T}(\rho,t)$ through an extra edge e between their roots which we match to the center (i,j) of $\mathsf{B}_{ij}(t)$.

Let $h \mapsto \varphi_h$ denotes the functional that, given a random variable h, evaluates the right side of (2.7). 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, it suffices to fix $B > 0$ and show that $\phi_n(\beta,B)$ converges as $n \to \infty$ to the predicted $\varphi_{h^*}(\beta,B)$ of (2.7), whereby $h^* = h^*_\beta$ is the unique fixed point of the recursion (2.5) that is supported on $[0,\infty)$ (see Lemma 2.2.2).

This is obviously true for $\beta = 0$ since $\phi_n(0,B) = \log(2\cosh B) = \varphi_h(0,B)$. 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

$$\partial_\beta \phi_n(\beta,B) = \frac{1}{n} \sum_{(i,j)\in E_n} \langle x_i x_j \rangle_n \,. \tag{2.17}$$

As $|\partial_\beta \phi_n(\beta,B)| \leq |E_n|/n$ is bounded by the assumed uniform sparsity, it is enough to show that the expression in (2.17) converges to the partial derivative of $\varphi_{h^*_\beta}(\beta,B)$ with respect to β. Turning to compute the latter derivative, a bit of real analysis shows that we can ignore the dependence of h^*_β on β. That

is, we simply compute the partial derivative in β of the expression (2.7) while considering the law of h_i to be fixed. Indeed, by the exchangeability of $\tanh(h_i)$ whose law has the fixed point property for (2.5), a direct computation leads to

$$\partial_\beta \varphi(\beta, B) = \frac{\overline{P}}{2}\mathbb{E}\Big[\langle x_i x_j\rangle_{\overline{\mathsf{T}}}\Big]\,, \tag{2.18}$$

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

$$\mu_{\overline{\mathsf{T}}}(x_i, x_j) = \frac{1}{z_{ij}}\,\exp\Big\{\beta x_i x_j + H_i x_i + H_j x_j\Big\}\,,$$

on one edge (ij) and random magnetic fields H_i and H_j that are independent copies of h^*_β.

In comparison, fixing a positive integer t, by Griffiths inequality the correlation $\langle x_i x_j\rangle_n$ lies between the correlations $F_0(\mathsf{B}_{ij}(t)) \equiv \langle x_i x_j\rangle^0_{\mathsf{B}_{ij}(t)}$ and $F_+(\mathsf{B}_{ij}(t)) \equiv \langle x_i x_j\rangle^+_{\mathsf{B}_{ij}(t)}$ for the Ising model on the subgraph $\mathsf{B}_{ij}(t)$ with free and plus, respectively, boundary conditions at $\partial\mathsf{B}_{ij}(t)$. Thus, in view of (2.17)

$$\frac{1}{n}\sum_{(i,j)\in E_n} F_0(\mathsf{B}_{ij}(t)) \le \partial_\beta\,\phi_n(\beta, B) \le \frac{1}{n}\sum_{(i,j)\in E_n} F_+(\mathsf{B}_{ij}(t))\,,$$

and taking $n\to\infty$ we get by Lemma 2.5.1 that

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

To compute $F_{0/+}(\overline{\mathsf{T}}(\rho,t))$ we first sum over the values of x_k for $k\in\overline{\mathsf{T}}(\rho,t)\setminus\{i,j\}$. This has the effect of reducing $F_{0/+}(\overline{\mathsf{T}}(\rho,t))$ to a form of $\langle x_i x_j\rangle_{\overline{\mathsf{T}}}$. Further, as in the proof of Lemma 2.2.2, we get $F_{0/+}(\overline{\mathsf{T}}(\rho,t))$ by setting for H_i and H_j two independent copies of the variables $h^{(t)}$ and $h^{(t)}_+$, respectively, which converge in law to h^*_β when $t\to\infty$. Since $\mathbb{E}[\langle x_i x_j\rangle_{\overline{\mathsf{T}}}]$ is continuous with respect to convergence in law of H_i, by (2.18)

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

which completes the proof of the theorem.

Appendix: Griffiths inequalities and the case of regular trees

Griffiths inequalities allow us to compare certain marginals of ferromagnetic Ising measures for one graph G and non-negative parameters β, h with certain other

choices for G, β and h. To this end, we consider the *extended* ferromagnetic Ising measure

$$\mu_J(x) = \frac{1}{Z(J)} \exp\left\{\widehat{H}_J(x)\right\} = \frac{1}{Z(J)} \exp\Big\{ \sum_{R\subseteq V} J_R x_R \Big\}\,, \tag{2.19}$$

for a finite set V and parameters $J_R \geq 0$, where hereafter $x_R = \prod_{u\in R} x_u$ and $x = (x_u, u \in V)$ for spin variables $x_u \in \mathcal{X} = \{-1, 1\}$. We note in passing that the Ising measure $\mu_{G,\beta,h}$ of (2.1) is merely μ_J in case $J_{\{i\}} = h$ for all $i \in V$, $J_{\{i,j\}} = \beta$ for all $(i, j) \in E$ and $J_R = 0$ for all other subsets of V.

In this context Griffiths inequalities are [1]

Proposition 2.5.2 (Griffiths inequalities)**.** *For $A, B \subseteq V$ and any $J = (J_R, R \subseteq V)$ with $J_R \geq 0$,*

$$\mathbf{E}_J[x_A] = \frac{1}{Z(J)} \sum_x x_A \exp\left\{\widehat{H}_J(x)\right\} \geq 0\,, \tag{2.20}$$

$$\frac{d}{dJ_B}\mathbf{E}_J[x_A] = \mathrm{Cov}_J(x_A, x_B) \geq 0\,. \tag{2.21}$$

Proof. Fixing $A \subseteq V$ we start with (2.20), where for V finite,

$$\begin{aligned}\sum_x x_A \exp\left\{\widehat{H}_J(x)\right\} &= \sum_x x_A \sum_{n=0}^{\infty} \frac{1}{n!} \widehat{H}_J(x)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_x x_A (\sum_R J_R x_R)^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{R_1,\ldots,R_n} \prod_{\ell=1}^{n} J_{R_\ell} \sum_x x_A \prod_{\ell=1}^{n} x_{R_\ell}\,.\end{aligned}$$

Since $x_u^2 = 1$ for all u we have that $x_A \prod_{\ell=1}^n x_{R_\ell} = x_C$ for $C = \{u \in V : u \text{ in an odd number of sets among } A, R_1, \ldots, R_n\}$. Further, with $\sum_x x_u = 0$ it follows that $\sum_x x_C = 0$ if C is non-empty (and $\sum_x x_C = 2^{|V|} > 0$ for $C = \emptyset$). Thus $\sum_x x_A \prod_{\ell=1}^n x_{R_\ell} \geq 0$ for all $A, R_1, \ldots, R_n$, and with $J_R \geq 0$ for all R, we have established (2.20). Turning to deal with (2.21) we fix $A, B \subseteq V$ and check that

$$\begin{aligned}\frac{d}{dJ_B}\mathbf{E}_J[x_A] &= \frac{d}{dJ_B} \frac{\sum_x x_A \exp\{\sum_R J_R x_R\}}{\sum_x \exp\{\sum_R J_R x_R\}} \\ &= Z(J)^{-2} \sum_{x,y} (x_A x_B - x_A y_B) \exp\{\sum_R J_R (x_R + y_R)\}\,,\end{aligned}$$

which is precisely the covariance of x_A and x_B under $\mu_J(x)$. We shall use (2.20) to verify that this quantity is non-negative. To this end, let $z_u = x_u y_u \in \mathcal{X}$

[1] Our source for both statement and proof is [41, Theorem IV.1.21], see also [47] for more general results in this direction.

noting that $y_R = x_R z_R$ for any $R \subseteq V$ (as $x_R^2 = 1$), and as before $x_A x_B = x_C$ for the symmetric difference C between A and B. Consequently,

$$\sum_{x,y}(x_A x_B - x_A y_B)\exp\{\sum_R J_R(x_R + y_R)\} = \sum_z (1 - z_B)\sum_x x_C \exp\{\sum_R J_R(z)x_R\}$$

where $J_R(z) = J_R(1 + z_R) \geq 0$. From (2.20) we thus have that

$$\sum_x x_C \exp\{\sum_R J_R(z)x_R\} \geq 0$$

for each $z \in \mathcal{X}^V$, and with $1 - z_B \geq 0$ we complete the proof of (2.21). □

Fixing $\beta > 0$ and $h \geq 0$, for any finite graph G let $m_v(G) = \mu_{G,\beta,h}(x_v = 1) - \mu_{G,\beta,h}(x_v = -1)$ denote the magnetization at $v \in V$ induced by the corresponding (ferromagnetic) Ising measure. For $S \subseteq V$ we similarly define $m_v(S; b)$ as the magnetization at v induced by the same Ising measure subject to fixed boundary conditions $x_u = b_u$ for $u \notin S$. Of particular interest to us are $m_v(S; +)$ and $m_v(S; f)$ corresponding to $b_u = 1$, respectively $b_u = 0$, for all $u \notin S$. The latter are called *free boundary conditions* since subject to $b_u = 0$, $u \notin S$, the restriction of the Ising measure $\mu_{G,\beta,h}$ to $(x_u, u \in S)$ coincides with the Ising measure $\mu_{G|_S,\beta,h}$ for the restriction $G\,|_S$ of G to S (i.e. with S as its vertices and $\{(i,j) \in E : i \in S, j \in S\}$ as its edges). We then get by Griffiths inequalities the following comparison results

Lemma 2.5.3. *If $v \in S \subseteq V$ then $m_v(S; f) \leq m_v(G) \leq m_v(S; +)$. Further, $S \mapsto m_v(S; f)$ is monotone non-decreasing and $S \mapsto m_v(S; +)$ is monotone non-increasing, both with respect to set inclusion (among sets S that contain v).*

Proof. From Griffiths inequalities we know that $J \mapsto \mathbf{E}_J[x_v]$ is monotone non-decreasing (where $J \geq \hat{J}$ if and only if $J_R \geq \hat{J}_R$ for all $R \subseteq V$).

Recall further that $m_v(G) = \mathbf{E}_{J^0}[x_v]$ where $J^0_{\{i\}} = h$, $J^0_{\{i,j\}} = \beta$ when $(i,j) \in E$ and all other values of J^0 are zero. Considering

$$J_R^{\eta,S} = J_R^0 + \eta \mathbf{1}_{R \subseteq S^c, |R|=1}\,,$$

with $\eta \mapsto J^{\eta,S}$ non-decreasing, so is $\eta \mapsto \mathbf{E}_{J^{\eta,S}}[x_v]$. In addition, $\mu_{J^{\eta,S}}(x_i = -1) \leq ce^{-2\eta}$ when $i \notin S$, hence as $\eta \uparrow \infty$ the measure $\mu_{J^{\eta,S}}$ converges to μ_J subject to the fixed boundary conditions $x_u = 1$ for $u \notin S$. Consequently, $m_v(G) \leq \mathbf{E}_{J^{\eta,S}}[x_v] \uparrow m_v(S; +)$.

Similarly, let $J_R^S = J_R^0 \mathbf{1}_{R \subseteq S}$ noting that under μ_{J^S} the random vector $(x_u, u \in S)$ is distributed according to the Ising measure $\mu_{G|_S,\beta,h}$. With $v \in S$ we thus deduce that $m_v(S; f) = \mathbf{E}_{J^S}[x_v] \leq \mathbf{E}_{J^0}[x_v] = m_v(G)$.

Finally, the stated monotonicity of $S \mapsto m_v(S; f)$ and $S \mapsto m_v(S; +)$ are in view of Griffiths inequalities the direct consequence of the monotonicity (with respect to set inclusions) of $S \mapsto J^S$ and $S \mapsto J^{\eta,S}$, respectively. □

Chapter 3

XORSAT on random regular graphs

XORSAT is arguably the simplest constraint satisfaction problem. It was introduced by Nadia Creignou and Herve Daude in 1999, and studied in a number of papers since then.

3.0.1 k-XORSAT

A k-XORSAT instance is described by a factor graph $G = (V, F, E)$ where $|V| = N$, $|F| = M$ and the factor nodes have degree k. We will denote the set of variable nodes adjacent to factor node a as $\partial a = \{i_1(a), \dots, i_k(a)\}$. Further, for any $a \in F$, we need to specify $J_a \in \{+1, -1\}$.

We will consider the following distribution over $x \in \{+1, -1\}^N$

$$\mu_G(x) = \frac{1}{Z_N(\beta)} \exp\left\{\beta \sum_{a \in F} J_a x_{i_1(a)} \cdots x_{i_k(a)}\right\}. \tag{3.1}$$

In the following we shall be mostly interested in the case in which there exists at least one solution x^* with $J_a x^*_{i_1(a)} \cdots x^*_{i_k(a)} = 1$ for all $a \in F$. In this case, it can be shows that (under a change of variable) the above distribution is 'essentially equivalent' to

$$\mu_G(x) = \frac{1}{Z_N(\beta)} \exp\left\{\beta \sum_{a \in F} x_{i_1(a)} \cdots x_{i_k(a)}\right\}. \tag{3.2}$$

When it will be necessary to specify, we will refer to this as to the *unfrustrated XORSAT* model.

Exercise: What does it mean 'essentially equivalent' the above statement? Describe the change of variables we are referring to.

Basic information on XORSAT, focusing on the zero-temperature case is provided by the Chapter 17 of the book with Marc Mézard, posted online.

In this note we shall focus on ensembles of random (l, k)-*regular graphs*. 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.

Throughout, the *adjacency matrix* $\mathbb{H}$ of G will be the binary matrix whose rows correspond to function nodes in F and columns to variable nodes in V. Its entry H_{ai}, $a \in F$, $i \in V$ is just the parity of the multiplicity of edge (a, i) in G.

3.0.2 Free energy

Theorem 3.0.4. *Let G be a random regular (l, k) factor graph, with $k > l \geq 2$. Then, with high probability*

$$Z_N(\beta) = 2^N (\cosh \beta)^{lN/k} \,. \tag{3.3}$$

In particular, the number of solutions is, with high probability $2^{N(1-l/k)}$.

For proving this Theorem, it is convenient to first derive an exact expression for the free energy. In order to do this, we introduce the notion of *hyperloop*. Given a factor graph $G = (V, F, E)$, an hyperloop is a subset F' of the factor node, such that the induced subgraph G' has even degree.

Lemma 3.0.5. *Let $G = (V, F, E)$ be a factor graph, $Z_N(\beta)$ the partition function of the associated unfrustrated XORSAT model, and $n_G(\ell)$ denote the number of hyperloops of size ℓ in G. Then, for any β,*

$$Z_N(\beta) = 2^N (\cosh \beta)^M \sum_{\ell=0}^{M} n_G(\ell) \, (\tanh \beta)^\ell \,. \tag{3.4}$$

Proof. By high-temperature expansion. □

We also need a result on the solutions of of random regular linear systems.

Theorem 3.0.6. *Let G be a random regular (l, k) factor graph, with $l > k \geq 2$, and $\mathbb{H}$ denote the corresponding adjacency matrix. Then the linear system $\mathbb{H}x = 0, \mod 2$ has, with high probability, the unique solution $x = 0$.*

Proof. Let $Z_{\mathbb{H}}(w)$ denote the number of solutions of $\mathbb{H}x = 0$ with w non-zero entries. Compute $\mathbb{E}\, Z_{\mathbb{H}}(w)$ and show that

$$\lim_{N \to \infty} \sum_{w=1}^{N} \mathbb{E}\, Z_{\mathbb{H}}(w) = 0 \,. \tag{3.5}$$

Some details of the computation are in Chapter 11 of the book online. □

Theorem 3.0.4. In view of the previous lemma, it is sufficient to show that, with high probability, $n_G(\ell) = 0$ for all $\ell \geq 1$, since $n_G(0) = 1$. Let $\mathbb{H}$ be the adjacency matrix of G (with rows corresponding to nodes in F and columns to nodes in V). Then our claim is equivalent to the following: The linear system $\mathbb{H}^T x = 0 \mod 2$ admits the unique solution $x = 0$. This follows from Theorem 3.0.6 once we notice that $\mathbb{H}^T$ is the adjacency matrix of a (k, l) regular factor graph. $\square$

Chapter 4

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. In general, this 'reduction' involves an uncontrolled error. We are interested here in mean-field models, for which Bethe-Peierls is expected to be asymptotically exact in the large system limit. In fact we already saw an example of this type, namely the ferromagnetic Ising model on random regular graphs.

In the context of mean field spin glasses, the method was further refined by Mézard, Parisi and Virasoro to include 'replica symmetry breaking' effects. In such applications, this is referred to as the 'cavity method.' A closely related approach is provided by the so-called TAP (Thouless-Anderson-Palmer) equations [6].

Section 4.1 provides a syntethic non-rigorous presentation of the basic ideas. A formalization of these ideas is developed in Section 4.2.

4.1 An informal introduction

4.1.1 Bethe equations

Consider a general model on the graph $G = (V, E)$ with variables $x_i \in \mathcal{X}$, and a distribution

$$\mu(x) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(x_i, x_j)\,. \tag{4.1}$$

Given $(i, j) \in E$, define $\mu^{(ij)}(\,\cdot\,)$ to be the modified distribution whereby the contribution of edge (i, j) has been 'taken out.' Explicitly

$$\mu^{(ij)}(x) \equiv \frac{1}{Z^{(ij)}} \prod_{(kl)\in E\setminus(ij)} \psi_{kl}(x_k, x_l)\,. \tag{4.2}$$

Further, for $i \in V$ we let $\mu^{(i)}(\,\cdot\,)$ be the modified measure in which all the edges incident on i have been removed:

$$\mu^{(i)}(x) \equiv \frac{1}{Z^{(i)}} \prod_{(kl)\in E,\, i\notin(kl)} \psi_{kl}(x_k, x_l)\,. \tag{4.3}$$

Given $U \subseteq V$, I shall denote by μ_U (respectively, by $\mu_U^{(ij)}$, $\mu_U^{(i)}$) the marginal distribution of $x_U \equiv \{x_i : i \in U\}$ when x is distributed according to μ (respectively $\mu^{(ij)}$, $\mu^{(i)}$).

The first step consists in deriving a set of exact equations relating the marginals of $\mu^{(i)}$ to the ones of $\mu^{(ij)}$. In order to write such equations, it is convenient to adopt the following notation. Whenever f and g are two non-negative functions on the same domain, we shall write $f(x) \cong g(x)$ if they differ by an overall normalization. We then have

$$\mu_{ij}^{(ij)}(x_i, x_j) \cong \sum_{x_{\partial i\setminus i,j}} \mu_{\partial i}^{(i)}(x_{\partial i}) \prod_{l\in\partial i\setminus j} \psi_{il}(x_i, x_l)\,. \tag{4.4}$$

Of course this does not solve the problem because we have more variables than unknowns. The Bethe-Peierls method consists in writing a set of mean field equations for the 'modified marginals'

$$\nu_{i\to j}(x_i) \equiv \mu_i^{(ij)}(x_i)\,. \tag{4.5}$$

The crucial approximation consists in assuming that

$$\mu_{ij}^{(ij)}(x_i, x_j) \;=\; \nu_{i\to j}(x_i)\nu_{j\to i}(x_j) + \mathsf{ERR}\,, \tag{4.6}$$

$$\mu^{(i)}(x_{\partial i}) = \prod_{l\in\partial i} \nu_{l\to i}(x_l) + \mathsf{ERR}\,. \tag{4.7}$$

where ERR is an error term that is assumed to be small.

Bethe-Peierls equations are obtained by plugging the last expressions in Eq. (4.4) and neglecting the error term. If $\nu_{j\to i}(x_j) > 0$, we get

$$\nu_{i\to j}(x_i) \cong \prod_{l\in\partial i\setminus j} \sum_{x_l\in\mathcal{X}} \psi_{il}(x_i, x_l)\nu_{l\to i}(x_l)\,. \tag{4.8}$$

Bethe-Peierls method consists in solving these equations for the 'messages' (or 'cavity fields') $\{\nu_{i\to j}\}$ and then using them for estimating the marginals of μ. For instance

$$\mu_i(x_i) \cong \prod_{j\in\partial i} \sum_{x_j} \psi_{ij}(x_i, x_j)\nu_{j\to i}(x_j)\,. \tag{4.9}$$

Exercise 1: Assume G to be a tree. Prove that the error terms in Eqs. (4.6), (4.7) vanish in this case, and that Bethe-Peierls equations have a unique solution corresponding to the actual marginals of μ.

4.1.2 Bethe free entropy

Since within Bethe approximation all marginals can be expressed in terms of the messages $\{\nu_{i\to j}\}$ that solve Bethe equations (4.8), it is perhaps not surprising that the free entropy can be expressed in terms of the same messages as well. It is more surprising that a simple expression exists

$$\Phi\{\nu\} = -\sum_{(ij)\in G}\log\Big\{\sum_{x_i,x_j}\psi_{ij}(x_i,x_j)\nu_{i\to j}(x_i)\nu_{j\to i}(x_j)\Big\} + \sum_{i\in V}\log\Big\{\sum_{x_i}\prod_{j\in\partial i}\sum_{x_j}\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\Big\}\,. \tag{4.10}$$

In general we shall regard this as a function from the set of possible messages to reals $\Phi : \{\nu_{i\to j}\} \mapsto \phi\{\nu_{i\to j}\} \in \mathbb{R}$. This function is called the *Bethe free entropy*.

It is easy to prove that this expression is exact if G is a tree. More precisely, if G is a tree and ν^* is the unique solution of the Bethe equations (4.8), then $\log Z = \Phi\{\nu^*\}$. There are many ways of proving this fact. A simple one consists in progressively disconnecting G in a recursive fashion.

We start form a simple remark. If $f_a(x) \cong p(x)$, for $a \in \{1,2,3\}$ and some probability distribution p, then

$$\log\left\{\sum_x \frac{f_1(x)f_2(x)}{f_3(x)}\right\} = \log\left\{\sum_x f_1(x)\right\} + \log\left\{\sum_x f_2(x)\right\} - \log\left\{\sum_x f_3(x)\right\} \tag{4.11}$$

Let us then describe the first step of the recursion. Let $(ij) \in E$ be an edge in G. Denote by $Z_{i\to j}(x_i)$ the constrained partition function for the subtree rooted at i and not including j, whereby we force x_i to take the value in argument. Then we obviously have

$$Z = \sum_{x_i,x_j} Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j) \tag{4.12}$$

$$= \sum_{x_i,x_j} \frac{\big\{Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\big\}\big\{\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j)\big\}}{\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)} \tag{4.13}$$

It is easy to see that, if G is a tree, $Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j) \cong \mu_{ij}(x_i,x_j)$, $\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j) \cong \mu_{ij}(x_i,x_j)$, $\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j) \cong \mu_{ij}(x_i,x_j)$. By the remark (4.11)

$$\log Z = \log\Big\{\sum_{x_i,x_j} Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\Big\} + \log\Big\{\sum_{x_i,x_j}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j)\Big\} - \log\Big\{\sum_{x_i,x_j}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\Big\}\,. \tag{4.14}$$

At this point we can interpret the first two terms as log partition functions for models of reduced size. The associated graphs are subtrees rooted at i and j.

If we repeat recursively this operation, we end up with a term of the form $-\log\{\sum_{x_i,x_j}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\}$ for each edge. The residual system is composed of disconnected 'stars' centered at the vertices of G. The corresponding log-partition function is given by the second sum in Eq. (4.10).

One important property of the Bethe free entropy is that its stationary points are solutions of the Bethe equations (4.8). This is proved by simple calculus

$$\begin{aligned}\frac{\partial\Phi\{\nu\}}{\partial\nu_{j\to i}(x_j)} = &-\frac{\sum_{x_i}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)}{\sum_{x_i',x_j'}\nu_{i\to j}(x_i')\nu_{j\to i}(x_j')\psi_{ij}(x_i',x_j')}\\ &+\frac{\sum_{x_i}\mathsf{T}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)}{\sum_{x_i',x_j'}\mathsf{T}\nu_{i\to j}(x_i')\nu_{j\to i}(x_j')\psi_{ij}(x_i',x_j')}\end{aligned}$$

where we defined

$$\mathsf{T}\nu_{i\to j}(x_i) \equiv \frac{1}{z_{i\to j}}\prod_{l\in\partial i\setminus j}\sum_{x_l}\psi_{il}(x_i,x_l)\nu_{l\to i}(x_l)\,. \tag{4.15}$$

Exercise 2: Prove that this implies that Bethe equations have at least one solution.

4.1.3 Example: Ising models

As an illustration, we consider again the Ising model, that gives a distribution over $x\in\{+1,-1\}^V$, of the form

$$\mu(x) = \frac{1}{Z}\exp\Big\{\beta\sum_{(ij)\in E}J_{ij}x_ix_j\Big\}\,. \tag{4.16}$$

Here we don't assume any more ferromagnetic interactions. The 'couplings' $\{J_{ij}\}$ are generic real numbers. The messages $\nu_{i\to j}$ can then be encoded effectively through the following log-likelihood ratios ('cavity fields')

$$h_{i\to j} = \frac{1}{2}\log\frac{\nu_{i\to j}(+1)}{\nu_{i\to j}(-1)}\,. \tag{4.17}$$

The Bethe equations then reduce to the ones below

$$h_{i\to j} = \sum_{l\in\partial i\setminus j}\operatorname{atanh}\left\{\tanh\beta J_{il}\tanh h_{l\to i}\right\}\,. \tag{4.18}$$

In terms of the cavity fields $\{h_{i\to j}\}$, we can compute the local magnetizations $m_i=\mathbb{E}\{x_i\}$ using Eq. (4.9):

$$m_i = \tanh\left\{\sum_{l\in\partial i}\operatorname{atanh}\left\{\tanh\beta J_{il}\tanh h_{l\to i}\right\}\right\}\,. \tag{4.19}$$

As a particular case, let us reconsider a regular graph with $J_{ij} = +1$ for each edge $(ij) \in E$. Then these equations admit a solution of the form $h_{i\to j} = h$, where h solves

$$h = (k-1)\text{atanh}\left\{\tanh\beta\tanh h\right\}\,. \tag{4.20}$$

For the ferromagnetic Ising model on a random regular graph, we indeed proved that the marginals can be computed by solving this equation.

Exercise 3: Write the Bethe free entropy (4.10) in the case of Ising models. Show that, in the case of regular graphs, under the assumption $h_{i\to j} = h$, it reduces to the form we proved a in Chapter 2.
[Indeed this is true for general graph sequences that converge locally to trees.]

Exercise 4: Exhibit an example of regular graph G, with degree $k \geq 2$, such that the associated ferromagnetic Ising model does not have phase transitions. What happens within Bethe-Peierls approximation?

4.1.4 Fully connected limit

When G is the complete graph, Bethe equations can often be simplified. Here we'll consider two such examples: the Curie-Weiss model (that we already considered at the beginning of the course) and the Sherrington-Kirkpatrick model.

Curie-Weiss model

We already encountered the Curie-Weiss model. This is defined by the general form (4.16) whereby G is the complete graph and $J_{ij} = 1/N$ for all (i,j). It follows from Eq. (4.18) that

$$h_{i\to j} = \frac{\beta}{N}\sum_{l\in V\setminus i,j}\tanh h_{l\to i} + O(1/N)\,. \tag{4.21}$$

Therefore $h_{i\to j} = h + O(1/N)$ where h solves

$$h = \beta\tanh h\,. \tag{4.22}$$

By Eq. (4.9), the local magnetization is given by $m = \tanh h$ and thus solves the equation

$$m = \tanh\beta m\,, \tag{4.23}$$

which we proved at the beginning of the course.

Sherrington-Kirkpatrick model

In the Sherrington-Kirkpatrick model, G is the complete graph and the couplings are $J_{ij} = J'_{ij}/\sqrt{N}$ with J'_{ij} iid normal random variables of mean 0 and variance 1. Hereafter I'll drop the prime.

By expanding again Eq. (4.18), we get

$$h_{i\to j} = \sum_{l\in V\setminus i,j} \frac{\beta J_{il}}{\sqrt{N}} \tanh h_{l\to i} + O(N^{-1/2})\,. \tag{4.24}$$

The relation to local magnetizations is not as simple as for the Curie-Weiss model. Expanding Eq. (4.9), we get

$$\begin{aligned} \mathrm{atanh} m_i &= \sum_{l\in V\setminus i,j} \frac{\beta J_{il}}{\sqrt{N}} \tanh h_{l\to i} + O(N^{-1/2}) = & (4.25)\\ &= h_{i\to j} + \frac{\beta J_{ij}}{\sqrt{N}}\, m_j + O(N^{-1/2})\,. & (4.26) \end{aligned}$$

Substituting in (both sides of!) Eq. (4.24), and neglecting terms of order $N^{-1/2}$, we get the so-called TAP equations

$$\mathrm{atanh} m_i = \frac{\beta}{\sqrt{N}} \sum_{l\in V\setminus i} J_{il} m_l - m_i \sum_{l\in V\setminus i} \frac{\beta^2 J_{il}^2}{N}(1-m_l^2)\,. \tag{4.27}$$

4.2 Bethe-Peierls approximation: a formal treatment

This is a first attempt at formalizing the ideas of the previous section in a general setting. We define Bethe states as distributions that are well approximated within Bethe-Peierls scheme. These play the same role as pure Gibbs states on infinite lattices [16]. It is conjectured by physicists that a large class of models, including for instance the examples in Chapter 1, decompose into convex combinations of Bethe states. As a running example it is convenent to keep in mind the independent set model, whose configurations are of the form $\underline{x} = (x_1, \dots, x_N)$ with $x_i \in \{0, 1\}$. The model is defined by the joint distribution

$$\mu_G(\underline{x}) = \frac{1}{Z_G}\, \lambda^{|\underline{x}|} \prod_{(i,j)\in E} \mathbb{I}(x_i \neq x_j)\,, \tag{4.28}$$

where $|\underline{x}|$ denotes the number of non-zero entries in the vector $\underline{x}$.

We shall further define a notion of correlation decay (weaker than uniqueness) that generalizes extremality in trees, and show that it implies Bethe property. A

slightly stronger condition allows us to validate the Bethe-Peierls approximation for the log partition function.

While in general extremality on the graph G does not coincide with extremality on the associated tree model, in Chapter 6 provides a sufficient condition for this to happen.

4.2.1 Notations

We consider models defined on simple finite graphs $G \equiv (V, E)$, without double edges or self loops, and denote by $\vec{E} \equiv \{i \to j : \ (i,j) \in E\}$ the induced set of directed edges. The *distance* $d(i,j)$ between two vertices $i, j \in V$ is the number of edges traversed in the shortest path on G from i to j with $d(A,B) = \min\{d(i,j) : i \in A, j \in B\}$ the corresponding distance between $A \subseteq V$ and $B \subseteq V$. The set of neighbors of $i \in V$ is denoted by ∂i (so $\partial i = \{j \in V : \ (i,j) \in E\}$), with $\Delta = \max_i |\partial i|$ the maximal degree of G (which we assume hereafter to be at least two). To each subset $U \subseteq V$ of vertices corresponds a diameter $\mathrm{diam}(U) = \max\{d(i,j) : i,j \in U\}$, the 'internal' boundary $\partial U = \{i \in U : \partial i \not\subseteq U\}$ and induced subgraph $G_U = (U, E_U)$ such that $E_U = \{(i,j) \in E : i,j \in U\}$. Given a vertex $i \in V$ and a non-negative integer r, the *ball of radius r centered at i*, $\mathsf{B}(i,r)$ is the subgraph induced by the vertices $j \in V$ such that $d(i,j) \le r$ and for $i \to j \in \vec{E}$ the *directed neighborhood* (or *ball*) $\mathsf{B}(i \to j, r)$ of radius r around i is the ball of radius r centered at i in $G_{V \setminus \{j\}}$ (so in particular, $\mathsf{B}(i \to j, r) \subseteq \mathsf{B}(i,r)$).

Definition 4.2.1. *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 $i \in V$.*

With $\mathcal{X}$ a finite set, 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 4.2.2. *A* bounded specification $\psi \equiv \{\psi_{ij} : \ (ij) \in E\}$ *for G 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 $(ij) \in E$). A specification 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*

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

It is not hard to realize that the example (4.28) of independent sets is indeed of this type with $x_i^{\mathrm{p}} = 0$ where one defines the specification in this case by letting $\psi_{ij}(x_i, x_j) = \mathbb{I}\big((x_i, x_j) \neq (1,1)\big) \, \lambda^{\frac{x_i}{|\partial i|} + \frac{x_j}{|\partial j|}}$.

Error terms and correlation properties will be often phrased in terms of valid rate functions, namely,

Definition 4.2.3. *A* valid rate function *is a monotonically non-increasing function $\delta : \mathbb{N} \to [0,1]$ that decay to zero as $r \to \infty$. By (eventually) increasing $\delta(r)$, we can and will assume, without loss of generality, that $\delta(r+1) \geq \delta_* \delta(r)$ for some positive δ_* and all $r \in \mathbb{N}$.*

Whenever we mention 'constants' we refer in fact to quantities that might depend on $|\mathcal{X}|$, Δ, κ and δ_*, but neither on the graph G nor on the permissive bounded specification ψ. We shall generically denote such constants as γ, C (and, if necessary, $C_1, C_2, \ldots$), and agree that redefinitions are not explicitly mentioned.

4.2.2 Measures, messages and correlation decay

Let $\mathsf{M}(\mathcal{X})$ denote the space of probability measures over $\mathcal{X}$. The *canonical probability measure* associated to the graph-specification pair (G, ψ) is

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(x_i, x_j)\,. \tag{4.29}$$

Given a probability measure $\rho(\,\cdot\,)$ on $\mathcal{X}^V$, we denote by $\rho_U(\,\cdot\,)$ the marginal distribution of $\underline{x}_U$ under ρ. The Bethe-Peierls approximation as defined below provides a way to express local marginals in terms of a set of messages.

Definition 4.2.4. *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 such that if $\partial i = \{j\}$ then $\nu_{i\to j}(\,\cdot\,)$ is the uniform measure on $\mathcal{X}$.*

Let $\mathcal{U}$ denote the collection of $U \subseteq V$ for which $G_U = (U, E_U)$ is a tree and for any $i \in U$, either $\partial i \subseteq U$ or $|\partial i \cap U| = 1$, so each vertex $i \in \partial U$ is a leaf of G_U (i.e. $i \in U$ and $|\partial i \cap U| = 1$). The probability measure on $\mathcal{X}^U$ induced by the messages $\{\nu_{i\to j}\}$ on each $U \in \mathcal{U}$ is then

$$\nu_U(\underline{x}_U) = \frac{1}{Z_U} \prod_{i\in\partial U} \nu_{i\to u(i)}(x_i) \prod_{(ij)\in E_U} \psi_{ij}(x_i, x_j)\,, \tag{4.30}$$

where $\{u(i)\} = \partial i \cap U$ for i a leaf of G_U.

We say that 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*

$$\sup_{U\in\mathcal{U}, \mathrm{diam}(U)\leq 2r} ||\rho_U - \nu_U||_{\mathrm{TV}} \leq \varepsilon\,, \tag{4.31}$$

and call such $\rho(\cdot)$ an (ε, r) Bethe state *for the graph-specification pair (G, ψ).*

Remark 4.2.5. *Note that if $i \notin \partial U$ is a leaf of an induced tree G_U then $\partial i = \{u(i)\}$ and $\nu_{i\to u(i)}$ must be the uniform measure on $\mathcal{X}$. Consequently, in (4.30) we may and shall not distinguish between ∂U and the collection of all leaves of G_U.*

We shall prove that Bethe approximation holds under the following correlation decay hypotheses on μ.

Definition 4.2.6. *A probability measure ρ on $\mathcal{X}^V$ is* strongly extremal *for G with valid rate function $\delta(\,\cdot\,)$ if for any $A, B, C \subseteq V$*

$$\sum_{\underline{x}_C} \rho_C(\underline{x}_C) ||\rho_{A,B|C}(\,\cdot\,,\,\cdot\,|\underline{x}_C) - \rho_{A|C}(\,\cdot\,|\underline{x}_C)\rho_{B|C}(\,\cdot\,|\underline{x}_C)||_{\mathrm{TV}} \le \delta(d(A,B))\,. \quad (4.32)$$

It is extremal *for G with valid rate function $\delta(\cdot)$ as soon as the preceding applies for $C = \emptyset$, namely, for any $A, B \subseteq V$,*

$$||\rho_{A,B}(\,\cdot\,,\,\cdot\,) - \rho_A(\,\cdot\,)\rho_B(\,\cdot\,)||_{\mathrm{TV}} \le \delta(d(A,B))\,. \quad (4.33)$$

The *standard message set* which we define next, is the natural candidate for Bethe approximation of a canonical probability measure.

Definition 4.2.7. *The* standard message set *for the canonical measure μ associated to a graph-specification pair (G, ψ) is $\nu_{i\to j}(x_i) \equiv \mu_i^{(ij)}(x_i)$, that is, the marginal on i of the probability measure on $\mathcal{X}^V$*

$$\mu^{(ij)}(\underline{x}) = \frac{1}{Z_{ij}} \prod_{(kl)\in E\setminus(ij)} \psi_{kl}(x_k, x_l)\,, \quad (4.34)$$

obtained from Eq. (4.29) upon omitting the factor $\psi_{ij}(\cdot,\cdot)$. Indeed, note that the marginal on i of $\mu^{(ij)}(\cdot)$ is a uniform measure on $\mathcal{X}$ whenever $\partial i = \{j\}$.

A special role is played in our analysis by the Bethe free energy and the Bethe (or belief propagation) equations. Both are functions on the space of possible messages sets $\mathsf{M}(\mathcal{X})^{\vec{E}}$, where $\mathsf{M}(\mathcal{X})$ denotes the space of probability measures over $\mathcal{X}$.

Definition 4.2.8. *Given a graph-specification pair (G, ψ), the* Bethe free energy *of a message set $\nu = \{\nu_{i\to j}\}$ is*

$$\begin{aligned}\Phi(\nu) = &- \sum_{(ij)\in E} \log\Big\{ \sum_{x_i, x_j} \psi_{ij}(x_i, x_j)\nu_{i\to j}(x_i)\nu_{j\to i}(x_j)\Big\} \\ &+ \sum_{i\in V} \log\Big\{ \sum_{x_i} \prod_{j\in\partial i} \sum_{x_j} \psi_{ij}(x_i, x_j)\nu_{j\to i}(x_j)\Big\}\,,\end{aligned}$$

and the value at ν of the Bethe (or belief propagation, BP*) mapping T of $\mathsf{M}(\mathcal{X})^{\vec{E}}$ to itself is*

$$(\mathsf{T}\nu)_{i\to j}(x_i) \equiv \frac{1}{z_{i\to j}} \prod_{l\in\partial i\setminus j} \Big[\sum_{x_l\in\mathcal{X}} \psi_{il}(x_i, x_l)\nu_{l\to i}(x_l)\Big]\,, \quad (4.35)$$

where $z_{i\to j}$ is determined by the normalization condition $\sum_{x\in\mathcal{X}}(\mathsf{T}\nu)_{i\to j}(x) = 1$. It is easy to check that T is well defined for any permissive specification ψ. That is, in this case there exists for each $i \to j \in \vec{E}$ a positive constant $z_{i\to j} \ge \psi_{\min}^{|\partial i|-1}$ for which $(\mathsf{T}\nu)_{i\to j} \in \mathsf{M}(\mathcal{X})$.

4.2.3 Main results

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

Here we can prove the simplest case of this broad conjecture, namely when the canonical measure $\mu(\,\cdot\,)$ is itself extremal.

Theorem 4.2.9. *Let G be an R-tree like graph, ψ a permissive bounded specification, and $\delta(\,\cdot\,)$ a valid rate function. Suppose that the corresponding canonical measure $\mu(\,\cdot\,)$ is extremal with rate $\delta(\,\cdot\,)$.*

Then, $\mu(\,\cdot\,)$ is (ε, r) Bethe approximated by its standard message set for $\varepsilon = \exp(C_0^r)\delta(R-r)$ and all $r < R-1$. In particular, any such $\mu(\,\cdot\,)$ is an (ε, r) Bethe state for (G, ψ).

Further in this case we get a sharp approximation of the log partition function.

Theorem 4.2.10. *If ν is the standard message set corresponding to a canonical measure $\mu(\,\cdot\,)$ for a graph-specification pair (G, ψ) that is strongly extremal for R-tree like graph G with valid rate function $\delta(\,\cdot\,)$, then $|V|^{-1}|\log Z - \Phi(\nu)| \to 0$ as $R \to \infty$.*

4.2.4 Groundwork

Lemma 4.2.11. *For any probability measures ρ_a on a discrete set $\mathcal{Z}$ and positive constants β_a, $a = 1, 2$,*

$$||\beta_1\rho_1 - \beta_2\rho_2||_{\text{TV}} \geq \frac{1}{2}\min(\beta_1, \beta_2)||\rho_1 - \rho_2||_{\text{TV}}\,. \tag{4.36}$$

Given a bounded function $f : \mathcal{Z} \mapsto [0, f_{\max}]$ such that $\langle \rho_a, f\rangle \equiv \sum_{z\in\mathcal{Z}} \rho_a(z)f(z)$ is positive for $a = 1, 2$ consider the probability measures $\widehat{\rho}_a(z) \equiv \rho_a(z)f(z)/\langle\rho_a, f\rangle$ on $\mathcal{Z}$. Then,

$$||\widehat{\rho}_1 - \widehat{\rho}_2||_{\text{TV}} \leq \frac{3f_{\max}}{2\langle\rho_1, f\rangle}||\rho_1 - \rho_2||_{\text{TV}} \tag{4.37}$$

so if $f(\cdot)$ is also bounded below by $f_{\min} > 0$, then $||\widehat{\rho}_1 - \widehat{\rho}_2||_{\text{TV}} \leq (3f_{\max}/(2f_{\min}))||\rho_1 - \rho_2||_{\text{TV}}$.

Similarly, if $f(\cdot)$ is bounded below by $f_{\min} > 0$, then

$$||\rho_1 - \rho_2||_{\text{TV}} \leq \frac{3\langle\rho_1, f\rangle}{2f_{\min}}||\widehat{\rho}_1 - \widehat{\rho}_2||_{\text{TV}} \tag{4.38}$$

and if it is also bounded above by $f_{\max}$, then $||\rho_1 - \rho_2||_{\text{TV}} \le (3f_{\max}/(2f_{\min}))||\widehat{\rho}_1 - \widehat{\rho}_2||_{\text{TV}}$.

Finally, in case ρ_a are two product measures on $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ we further have that for all $y \in \mathcal{Y}$,

$$\left|\frac{\widehat{\rho}_1(y)}{\rho_1(y)} - \frac{\widehat{\rho}_2(y)}{\rho_2(y)}\right| \le \frac{2f_{\max}^2}{\langle \rho_1, f\rangle\langle \rho_2, f\rangle}||\rho_1 - \rho_2||_{\text{TV}}\,. \tag{4.39}$$

Proof. Starting with (4.36) we can assume that $\lambda = \beta_2/\beta_1 > 1$. Setting $A = A_\lambda = \{z \in \mathcal{Z} : \rho_1(z) \ge \lambda\rho_2(z)\}$ note that

$$||\beta_1\rho_1 - \beta_2\rho_2||_{\text{TV}} = \frac{\beta_1}{2}[\rho_1(A) - \lambda\rho_2(A) + \rho_1(A) - \rho_2(A) + (\lambda - 1)(1 - \rho_2(A))]\,.$$

Since A_1 is the disjoint union of $A = A_\lambda$ and $B = \{z \in \mathcal{Z} : \lambda\rho_2(z) > \rho_1(z) \ge \rho_2(z)\}$, we further have

$$\frac{\beta_1}{2}||\rho_1 - \rho_2||_{\text{TV}} = \frac{\beta_1}{2}[\rho_1(A) - \rho_2(A) + \rho_1(B) - \rho_2(B)]\,.$$

The difference between the right side of these two identities is

$$\frac{\beta_1}{2}[\rho_1(A) - \lambda\rho_2(A) + (\lambda - 1)(1 - \rho_2(A) - \rho_2(B)) + \lambda\rho_2(B) - \rho_1(B)]\,.$$

We thus get (4.36) upon noting that the preceding expression is non-negative (since $\lambda \ge 1$ and $\rho_2(A)+\rho_2(B) = \rho_2(A_1) \le 1$, while by definition $\rho_1(A) \ge \lambda\rho_2(A)$ and $\lambda\rho_2(B) \ge \rho_1(B)$).

The bound (4.37) is proved in [19, Lemma 3.3]. The inequality (4.38) follows upon noting that $\rho_a(z) = \widehat{\rho}_a(z)g(z)/\langle\widehat{\rho}_a, g\rangle$ for $g(z) = 1/f(z)$ bounded above by $g_{\max} = 1/f_{\min}$ and that $\langle\widehat{\rho}_1, g\rangle = 1/\langle\rho_1, f\rangle$. Finally note that for $\rho_a(x, y) = \rho_a(x)\rho_a(y)$ the left side of (4.39) is merely

$$\begin{aligned}
&\left|\frac{\sum_x \rho_1(x)f(x,y)}{\langle\rho_1, f\rangle} - \frac{\sum_x \rho_2(x)f(x,y)}{\langle\rho_2, f\rangle}\right| \\
\le\ & \frac{|\langle\rho_1, f\rangle - \langle\rho_2, f\rangle|}{\langle\rho_1, f\rangle\langle\rho_2, f\rangle}\sum_x \rho_2(x)f(x,y) + \frac{1}{\langle\rho_1, f\rangle}\left|\sum_x \rho_1(x)f(x,y) - \sum_x \rho_2(x)f(x,y)\right| \\
\le\ & \frac{f_{\max}^2||\rho_1 - \rho_2||_{\text{TV}}}{\langle\rho_1, f\rangle\langle\rho_2, f\rangle} + \frac{f_{\max}||\rho_1 - \rho_2||_{\text{TV}}}{\langle\rho_1, f\rangle}
\end{aligned}$$

and we arrive at (4.39) upon noting that $f_{\max}/\langle\rho_2, f\rangle \ge 1$. □

Here are some useful consequences of having a permissive bounded specification.

Lemma 4.2.12. *Suppose μ is the canonical measure associated with a permissive bounded specification ψ for G. Then, for any partial assignment $\underline{x}_U$, any $i \notin U$ and $(i,j) \in E$,*

$$\mu_{i|U}(x_i^{\rm p}|\underline{x}_U) \;\geq\; |\mathcal{X}|^{-1}\kappa^{\Delta}\,, \tag{4.40}$$

$$\mu_{i|U}^{(ij)}(x_i^{\rm p}|\underline{x}_U) \;\geq\; |\mathcal{X}|^{-1}\kappa^{\Delta-1}\,. \tag{4.41}$$

Further, for any message set $\{\nu_{i\to j}\}$ and any $i \to j \in \vec{E}$

$$(\mathsf{T}\nu)_{i\to j}(x_i^{\rm p}) \geq |\mathcal{X}|^{-1}\kappa^{\Delta-1}\,. \tag{4.42}$$

Proof. Let $\mu^{(i)}(\,\cdot\,)$ denote the canonical measure for the graph G without the vertex i, that is, the measure obtained upon removing the factors $\psi_{ik}(\,\cdot\,,\,\cdot\,)$ for $k \in \partial i$. Since $i \notin U$ we have that

$$\mu_{i|U}(x_i|\underline{x}_U) = \frac{\sum_{\underline{x}_{\partial i}} \prod_{k\in\partial i}\psi_{ik}(x_i,x_k)\,\mu^{(i)}(\underline{x}_{\partial i}|\underline{x}_U)}{\sum_{x',\underline{x}_{\partial i}} \prod_{k\in\partial i}\psi_{ik}(x',x_k)\,\mu^{(i)}(\underline{x}_{\partial i}|\underline{x}_U)}\,.$$

With ψ bounded and permissive, the inequality (4.40) follows upon using the upper bound $\psi_{ik}(x',x_k) \leq \psi_{\max}$ in the denominator and the lower bound $\psi_{ik}(x_i^{\rm p},x_k) \geq \kappa\psi_{\max}$ in the numerator. Further, with j removed from ∂i the same representation applies for $\mu_{i|U}^{(ij)}(x_i|\underline{x}_U)$ and from (4.35) we deduce that such representation applies also for $(\mathsf{T}\nu)_{i\to j}(x_i)$ apart from replacing $\mu^{(i)}(\underline{x}_{\partial i}|\underline{x}_U)$ with $\prod_{k\in\partial i\setminus j}\nu_{k\to i}(x_k)$. Consequently, we get both (4.41) and (4.42) by the preceding argument. □

The bound (4.39) allows us to control the partial derivatives of the Bethe free energy for a permissive bounded specification in terms of the difference between the message set and its image under the Bethe mapping T.

Proposition 4.2.13. *Suppose ψ is a permissive bounded specification for G and recall the mapping $\nu \to \Phi(\nu)$ of a set of messages to its Bethe free energy. If $\max[\nu_{i\to j}(x_i^{\rm p}),\nu_{j\to i}(x_j^{\rm p})] \geq K > 0$, then for the constant $C = 2|\mathcal{X}|\kappa^{-(\Delta+1)}$ and any $y \in \mathcal{X}$,*

$$\left|\frac{\partial\Phi(\nu)}{\partial\nu_{j\to i}(y)}\right| \leq \frac{C}{K}\;||(\mathsf{T}\nu)_{i\to j} - \nu_{i\to j}||_{\rm TV}\,. \tag{4.43}$$

Proof. By direct computation we find that

$$\frac{\partial\Phi(\nu)}{\partial\nu_{j\to i}(y)} = -\frac{\sum_x \nu_{i\to j}(x)\psi_{ij}(x,y)}{\sum_{x',y'}\nu_{i\to j}(x')\nu_{j\to i}(y')\psi_{ij}(x',y')} + \frac{\sum_x(\mathsf{T}\nu)_{i\to j}(x)\psi_{ij}(x,y)}{\sum_{x',y'}(\mathsf{T}\nu)_{i\to j}(x')\nu_{j\to i}(y')\psi_{ij}(x',y')}$$

which is precisely the term in the left side of (4.39) for $\mathcal{Y} = \mathcal{X}$, the function $f(x,y) = \psi_{ij}(x,y)$ and the product measures $\rho_1(x,y) = (\mathsf{T}\nu)_{i\to j}(x)\nu_{j\to i}(y)$ and $\rho_2(x,y) = \nu_{i\to j}(x)\nu_{j\to i}(y)$.

With ψ permissive and bounded we have that $f_{\max} \leq \psi_{\max}$ and by (4.42) also

$$\langle \rho_1, f \rangle \geq (\mathsf{T}\nu)_{i\to j}(x_i^{\rm p}) \sum_y \psi_{ij}(x_i^{\rm p}, y)\nu_{j\to i}(y) \geq \kappa\psi_{\max}(\mathsf{T}\nu)_{i\to j}(x_i^{\rm p}) \geq \psi_{\max}\kappa^{\Delta}|\mathcal{X}|^{-1}\,.$$

Similarly, $\langle \rho_2, f\rangle \geq \psi_{\min}\nu_{i\to j}(x_i^{\rm p})$ and further

$$\langle \rho_2, f\rangle \geq \sum_x \nu_{i\to j}(x)\psi_{ij}(x, x_j^{\rm p})\nu_{j\to i}(x_j^{\rm p}) \geq \psi_{\min}\nu_{j\to i}(x_j^{\rm p})\,,$$

so by our assumption $\langle \rho_2, f\rangle \geq \psi_{\min}K$. Plugging these bounds in the right side of (4.39) results with our thesis (since $||\rho_1 - \rho_2||_{\rm TV} = ||(\mathsf{T}\nu)_{i\to j} - \nu_{i\to j}||_{\rm TV}$). $\square$

Our next lemma provides a uniform bound on the total variation between the message set at a directed edge $i \to j$ and its corresponding image under the Bethe mapping T.

Lemma 4.2.14. *Suppose a message set $\{\nu_{i\to j}\}$ is an $(\varepsilon, 2)$ Bethe approximation of some probability measure $\rho(\,\cdot\,)$ for a permissive bounded specification ψ and a 2-tree like graph G. Then, for any directed edge $i \to j \in \vec{E}$,*

$$||(\mathsf{T}\nu)_{i\to j} - \nu_{i\to j}||_{\rm TV} \leq 3|\mathcal{X}|\kappa^{-\Delta}\,\varepsilon\,. \tag{4.44}$$

Proof. Fixing $i \to j \in \vec{E}$ it is easy to check that if $\partial i = \{j\}$ then $(\mathsf{T}\nu)_{i\to j} = \nu_{i\to j}$ (is the uniform measure on $\mathcal{X}$). Thus, we assume hereafter that $|\partial i| > 1$ and note that since G has girth exceeding five, $G_{U'}$ is a subtree of the tree G_U for $U' = \mathsf{B}(j, 1)$ and $U = \partial i \cup \partial j$ of diameter at most three with leaves $\partial U' = \partial j$ and $\partial U = U \setminus \{i, j\}$ (in case $|\partial j| = 1$ we ignore the leaf j of G_U and $G_{U'}$ since $\nu_{j\to i}$ is then the uniform measure on $\mathcal{X}$). Since the message set ν is a $(\varepsilon, 2)$ Bethe approximation of $\rho(\,\cdot\,)$ we have by (4.31) that

$$||\rho_{U'} - \nu_{U'}||_{\rm TV} + ||\rho_U - \nu_U||_{\rm TV} \leq 2\varepsilon\,.$$

Further, since $i \in U'$ and $\rho_{U'}$, ρ_U are marginals of same probability measure, it follows that $||(\nu_U)_i - (\nu_{U'})_i||_{\rm TV} \leq 2\varepsilon$ for the marginals $(\nu_U)_i$ and $(\nu_{U'})_i$ on i of ν_U and $\nu_{U'}$, respectively. Recall (4.30) that here

$$\begin{aligned}
\nu_U(\underline{x}_U) &= \frac{1}{Z_U}\psi_{ij}(x_i, x_j) \prod_{l\in\partial i\setminus j} \psi_{il}(x_i, x_l)\nu_{l\to i}(x_l) \prod_{k\in\partial j\setminus i} \psi_{jk}(x_j, x_k)\nu_{k\to j}(x_k)\,,\\
\nu_{U'}(\underline{x}_{U'}) &= \frac{1}{Z_{U'}}\psi_{ij}(x_i, x_j)\nu_{i\to j}(x_i) \prod_{k\in\partial j\setminus i} \psi_{jk}(x_j, x_k)\nu_{k\to j}(x_k)\,.
\end{aligned}$$

Summing the first expression over $\{x_l : l \in U, l \neq i\}$ and the second over $\{x_l : l \in U', l \neq i\}$ we get in terms of the Bethe mapping T of (4.35) that

$$\begin{aligned}
(\nu_U)_i(x_i) &= (\mathsf{T}\nu)_{i\to j}(x_i)f(x_i)/\langle (\mathsf{T}\nu)_{i\to j}, f\rangle\,,\\
(\nu_{U'})_i(x_i) &= \nu_{i\to j}(x_i)f(x_i)/\langle \nu_{i\to j}, f\rangle\,,
\end{aligned}$$

for the function $f(x) = \sum_y \psi_{ij}(x_i, y)(\mathsf{T}\nu)_{j\to i}(y)$ on $\mathcal{X}$. With ψ bounded and permissive, $f(x)$ is bounded above by $f_{\max} = \psi_{\max}$ and bounded below by $\psi_{\min}(\mathsf{T}\nu)_{j\to i}(x_j^{\rm p})$ which by (4.42) is further bounded below by $f_{\min} = \psi_{\max}\kappa^{\Delta}/|\mathcal{X}| > 0$. Thus, by (4.38) the bound $||(\nu_U)_i - (\nu_{U'})_i||_{\rm TV} \le 2\varepsilon$ translates to

$$||(\mathsf{T}\nu)_{i\to j} - \nu_{i\to j}||_{\rm TV} \le 3\varepsilon f_{\max}/f_{\min}$$

which is precisely the thesis of the lemma. □

Combining Proposition 4.2.13 and Lemma 4.2.14 we obtain the following.

Corollary 4.2.15. *Suppose a message set $\{\nu_{i\to j}\}$ is an $(\varepsilon, 2)$ Bethe approximation of some probability measure for a permissive bounded specification ψ, a 2-tree like graph G and $\varepsilon \le 1/c$ (where $c = 18|\mathcal{X}|^3\kappa^{-3\Delta}$ is a finite constant). Then, for any directed edge $i \to j \in \vec{E}$ and $y \in \mathcal{X}$,*

$$\Big|\frac{\partial\Phi(\nu)}{\partial\nu_{j\to i}(y)}\Big| \le c\varepsilon\,.$$

Proof. Recall (4.42) that $(\mathsf{T}\nu)_{i\to j}(x_i^{\rm p}) \ge c_1$ for the positive constant $c_1 = |\mathcal{X}|^{-1}\kappa^{\Delta-1}$ so in view of (4.44) we have that $\nu_{i\to j}(x_i^{\rm p}) \ge (\mathsf{T}\nu)_{i\to j}(x_i^{\rm p}) - 2c_2\varepsilon \ge c_1 - 2c_2\varepsilon$ for the finite constant $c_2 = 3|\mathcal{X}|\kappa^{-\Delta}$. Thus, with C of Proposition 4.2.13 bounded below by one, if $\varepsilon \le 1/c \equiv c_1/(3c_2C)$ then $\nu_{i\to j}(x_i^{\rm p}) \ge K = c_1/3$ in which case combining the bounds of Proposition 4.2.13 and Lemma 4.2.14 we deduce that

$$\Big|\frac{\partial\Phi(\nu)}{\partial\nu_{j\to i}(y)}\Big| \le \frac{C}{K}||(\mathsf{T}\nu)_{i\to j} - \nu_{i\to j}||_{\rm TV} \le c\varepsilon\,,$$

as stated. □

Lemma 4.2.16. *Suppose μ is the canonical measure for a permissive bounded specification. If $i, j \notin \partial U'$ and the edge (i, j) is in the tree $G_{U'}$, then for any configurations $\underline{x}^1$, $\underline{x}^2 \in \mathcal{X}^V$*

$$||\mu^{(ij)}_{i|\partial U'}(\,\cdot\,|\underline{x}^1_{\partial U'}) - \mu^{(ij)}_{i|\partial U'}(\,\cdot\,|\underline{x}^2_{\partial U'})||_{\rm TV} \le 2|\mathcal{X}|\kappa^{-\Delta}\,||\mu_{ij|\partial U'}(\,\cdot\,|\underline{x}^1_{\partial U'}) - \mu_{ij|\partial U'}(\,\cdot\,|\underline{x}^2_{\partial U'})||_{\rm TV}\,.$$

Proof. Since $G_{U'}$ is a tree, it is the disjoint union of the edge (i, j) and the two disjoint subtrees formed when removing this edge from $G_{U'}$. Denoting the intersections of $\partial U'$ with these two subtrees by $\partial U(i)$ and $\partial U(j)$ the structure of $\mu(\cdot)$ is such that for $a = 1, 2$,

$$\mu_{ij|\partial U'}(x, y|\underline{x}^a_{\partial U'}) = \frac{1}{z_a}\psi_{ij}(x, y)\mu^{(ij)}_{i|\partial U(i)}(x|\underline{x}^a_{\partial U(i)})\mu^{(ij)}_{j|\partial U(j)}(y|\underline{x}^a_{\partial U(j)})\,,$$

with normalizing constants $z_a = \sum_{x,y}\psi_{ij}(x, y)\mu^{(ij)}_{i|\partial U(i)}(x|\underline{x}^a_{\partial U(i)})\mu^{(ij)}_{j|\partial U(j)}(y|\underline{x}^a_{\partial U(j)})$. Since ψ is a permissive specification, it follows that

$$||\mu_{ij|\partial U'}(\,\cdot\,|\underline{x}^1_{\partial U'}) - \mu_{ij|\partial U'}(\,\cdot\,|\underline{x}^2_{\partial U'})||_{\rm TV} \ge \kappa\psi_{\max}||\beta_1\mu^{(ij)}_{i|\partial U(i)}(\,\cdot\,|\underline{x}^1_{\partial U(i)}) - \beta_2\mu^{(ij)}_{i|\partial U(i)}(\,\cdot\,|\underline{x}^2_{\partial U(i)})||_{\rm TV}\,,$$

where $\beta_a = \mu^{(ij)}_{j|\partial U(j)}(x^{\rm p}_j|\underline{x}^a_{\partial U(j)})/z_a$ for $a = 1, 2$. Clearly z_a is bounded above by $\psi_{\max}$ and as $j \notin \partial U(j)$ we get from (4.41) that $\beta_a \geq \psi^{-1}_{\max}|\mathcal{X}|^{-1}\kappa^{\Delta-1}$. The thesis thus follows by applying the bound (4.36). □

Lemma 4.2.17. *Suppose the canonical measure μ for 2-tree like graph and a permissive bounded specification is extremal of valid rate function $\delta(\cdot)$. Then, some finite constant K and any $A \subseteq V$*

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

Proof. Let $B = \partial i \cup \partial j \setminus \{i,j\}$, and assume, without loss of generality, that $A \cap B = \emptyset$. Then

$$\begin{aligned} ||\mu^{(ij)}_A - \mu_A||_{\rm TV} &= \sum_{\underline{x}_A} \Big| \sum_{\underline{x}_B} \mu^{(ij)}_B(\underline{x}_B)\mu_{A|B}(\underline{x}_A|\underline{x}_B) - \sum_{\underline{x}_B} \mu_B(\underline{x}_B)\mu_{A|B}(\underline{x}_A|\underline{x}_B)\Big| \\ &\leq \sup_{\underline{x}_B, \underline{x}'_B} ||\mu_{A|B}(\,\cdot\,|\underline{x}_B) - \mu_{A|B}(\,\cdot\,|\underline{x}'_B)||_{\rm TV}\,. \end{aligned}$$

We claim that there exists a constant C such that $\mu_B(\underline{x}_B) \geq C$ for any $\underline{x}_B$. This implies the thesis since

$$||\mu_{A|B}(\,\cdot\,|\underline{x}_B) - \mu_{A|B}(\,\cdot\,|\underline{x}'_B)||_{\rm TV} \leq \frac{1}{C^2}\mathbb{E}||\mu_{A|B}(\,\cdot\,|\underline{x}_B) - \mu_{A|B}(\,\cdot\,|\underline{x}'_B)||_{\rm TV} \leq \frac{1}{C^2}\delta\big(d(A,B)\big)\,.$$

In order to prove our claim notice that the vertices in B are not joined by any edge (the induced subgraph has empty vertex set). Let C be the set of vertices k such that $d(k, B) = 1$. Then

$$\begin{aligned} \mu_B(\underline{x}_B) &= \sum_{\underline{x}_C} \mu_{B|C}(\underline{x}_B|\underline{x}_C)\mu_C(\underline{x}_C) \geq \mu_C(\underline{x}^{\rm p}_C)\mu_{B|C}(\underline{x}_B|\underline{x}^{\rm p}_C) \qquad (4.45) \\ &\geq \mu_C(\underline{x}^{\rm p}_C)\prod_{l\in B}\left(\frac{\prod_{k\in\partial l}\psi_{lk}(x_l, x^{\rm p}_k)}{\sum_{x'_l}\prod_{k\in\partial l}\psi_{lk}(x_l, x^{\rm p}_k)}\right) \geq \mu_C(\underline{x}^{\rm p}_C)\,\kappa^{\Delta^2}\,. \end{aligned}$$

Finally $\mu_C(\underline{x}^{\rm p}_C) \geq \kappa^{\Delta|C|}$ by the same argument as in the proof of Lemma 4.2.12. □

4.2.5 Proof of Theorem 4.2.9

Proof. Fixing $r < R - 1$, the permissive bounded graph-specification pair (G, ψ) that is extremal for R-tree like graph G with valid rate function $\delta(\,\cdot\,)$ and $U \in \mathcal{U}$ with $\mathrm{diam}(U) \leq 2r$, let $\overline{U}_{R'} = \{k \in V : d(k, U) \geq R'\}$ for $R' = R - r > 1$. Note that

$$\begin{aligned} ||\mu_U(\cdot) - \nu_U(\cdot)||_{\rm TV} \leq\ & \mathbb{E}||\mu_U(\,\cdot\,) - \mu_{U|\overline{U}_{R'}}(\,\cdot\,|\underline{x}^*_{\overline{U}_{R'}})||_{\rm TV} \\ & + \mathbb{E}||\mu_{U|\overline{U}_{R'}}(\,\cdot\,|\underline{x}^*_{\overline{U}_{R'}}) - \nu_U(\,\cdot\,)||_{\rm TV}\,, \qquad (4.46) \end{aligned}$$

where ν_U corresponds to the standard message set and the expectation is with respect to the random configuration $\underline{x}^*$ 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 bound the second term, for each directed edge $i\to j\in\vec{E}$ let

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

where $\overline{\mathsf{B}}(i,R')$ is the collection of vertices of distance R' from i and $\mu^{(ij)}(\underline{x})$ is the measure of (4.34). 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 $i\in\partial U$. Consequently, it is not hard to verify that we have the representation

$$\mu_{U|\overline{U}_{R'}}(\underline{x}_U|\underline{x}^*_{\overline{U}_{R'}})=\frac{1}{Z'_U}\prod_{(ij)\in E_U}\psi_{ij}(x_i,x_j)\prod_{i\in\partial U}\nu^*_{i\to u(i)}(x_i)\,,\tag{4.47}$$

where $u(i)=\partial i\cap U$ for each $i\in\partial U$. The expression on the right side of (4.47) is of the same form as ν_U except for using $\nu^*_{i\to u(i)}$ in the former and $\nu_{i\to u(i)}=\mu^{(ij)}_i$ in the latter. With this in mind, we apply the bound of (4.37) for the function $f(\underline{x}_U)=\prod_{(ij)\in E_U}\psi_{ij}(x_i,x_j)$ that is bounded above by $f_{\max}=\psi_{\max}^{|U|-1}$ and 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 $\rho_2(\underline{x}_{\partial U})=\prod_{i\in\partial U}\nu^*_{i\to u(i)}(x_i)$. In doing so, recall that since ψ is a permissive specification, by (4.41),

$$\langle\rho_1,f\rangle\ge|\mathcal{X}|^{-|U|}\prod_{(ij)\in E_U}\psi_{ij}(x^{\rm p}_i,x^{\rm p}_j)\prod_{i\in\partial U}\left[|\mathcal{X}|\,\nu_{i\to u(i)}(x^{\rm p}_i)\right]\ge f_{\max}c^{-|U|}\,,$$

for the finite constant $c=|\mathcal{X}|\kappa^{-\Delta}$, so we get from (4.37) that

$$\begin{aligned}||\mu_{U|\overline{U}_R}(\,\cdot\,|\underline{x}^*_{\overline{U}_R})-\nu_U(\,\cdot\,)||_{\rm TV}&=||\widehat{\rho}_1-\widehat{\rho}_2||_{\rm TV}\le 2c^{|U|}||\rho_1-\rho_2||_{\rm TV}\\&\le 2c^{|U|}\sum_{i\in\partial U}||\nu_{i\to u(i)}-\nu^*_{i\to u(i)}||_{\rm TV}\,.\end{aligned}\tag{4.48}$$

Let $\nu'_{i\to j}=\mu^{(ij)}_{i|\overline{\mathsf{B}}(i,R')}(\cdot|\underline{x}'_{\overline{\mathsf{B}}(i,R')})$ and $\underline{x}'$ of distribution $\mu^{(ij)}$, independent of $\underline{x}^*$. Then,

$$\mathbb{E}||\nu_{i\to u(i)}-\nu^*_{i\to u(i)}||_{\rm TV}=\mathbb{E}||\,\mathbb{E}\nu'_{i\to u(i)}-\nu^*_{i\to u(i)}||_{\rm TV}\le\mathbb{E}||\nu'_{i\to u(i)}-\nu^*_{i\to u(i)}||_{\rm TV}\tag{4.49}$$

Now, by Lemma 4.2.16 and the fact that G is R-tree like, we have that

$$\begin{aligned}||\nu'_{i\to j}-\nu^*_{i\to j}||_{\rm TV}&=||\mu^{(ij)}_{i|\partial U'}(\,\cdot\,|\underline{x}'_{\partial U'})-\mu^{(ij)}_{i|\partial U'}(\,\cdot\,|\underline{x}^*_{\partial U'})||_{\rm TV}\\&\le 2|\mathcal{X}|\kappa^{-\Delta}\,||\mu_{ij|\partial U'}(\,\cdot\,|\underline{x}'_{\partial U'})-\mu_{ij|\partial U'}(\,\cdot\,|\underline{x}^*_{\partial U'})||_{\rm TV}\,.\end{aligned}\tag{4.50}$$

for $j = u(i)$, $i \in \partial U$, $U' = \mathsf{B}(i, R')$ and $\partial U' = \overline{\mathsf{B}}(i, R')$. Further, taking the expectation with respect to the independent random configurations $\underline{x}'$ and $\underline{x}^*$ leads to

$$\mathbb{E}||\mu_{ij}(\,\cdot\,|\underline{x}'_{\overline{\mathsf{B}}(i,R')}) - \mu_{ij}(\,\cdot\,|\underline{x}^*_{\overline{\mathsf{B}}(i,R')})||_{\text{TV}} \le 2||\mu_{\{ij\},\overline{\mathsf{B}}(i,R')} - \mu_{\{ij\}}\mu_{\overline{\mathsf{B}}(i,R')}||_{\text{TV}} + ||\mu^{(ij)}_{\overline{\mathsf{B}}(i,R')} - \mu_{\overline{\mathsf{B}}(i,R')}||_{\text{TV}}\,,$$

which for μ extremal of valid rate function $\delta(\cdot)$ is, due to Lemma 4.2.17, bounded by $(2+K)\delta(R'-1) \le c_1\delta(R-r)$ with $c_1 = (2+K)/\delta_*$ a finite constant. Combining this with (4.49) and (4.50) we have that for any $i \in \partial U$,

$$\mathbb{E}||\nu_{i\to u(i)} - \nu^*_{i\to u(i)}||_{\text{TV}} \le 2c_1|\mathcal{X}|\kappa^{-\Delta}\delta(R-r)\,,$$

and consequently, by (4.48) also

$$\mathbb{E}||\mu_{U|\overline{U}_R}(\,\cdot\,|\underline{x}_{\overline{U}_R}) - \nu_U(\,\cdot\,)||_{\text{TV}} \le 4c_1|\mathcal{X}|\kappa^{-\Delta}|\partial U|c^{|U|}\delta(R-r)\,.$$

Since $|\partial U| \le |U| \le |\mathsf{B}(i_o, r)| \le \Delta^{r+1}$ we have that $4c_1|\mathcal{X}|\kappa^{-\Delta}|\partial U|c^{|U|} \le \exp(C_0^r) - 1$ for some constant $C_0 = C_0(|\mathcal{X}|, \Delta, \kappa, \delta_*)$ and all U such that $\text{diam}(U) \le 2r$, which by (4.46) thus yields the bound

$$||\mu_U - \nu_U||_{\text{TV}} \le \exp(C_0^r)\,\delta(R-r)\,,$$

for every $U \in \mathcal{U}$ of $\text{diam}(U) \le 2r$ and $r < R-1$, as claimed. □

Chapter 5

q-Colorings of Random Graphs

In this chapter we shall consider the uniform measure $\mu_G(\,\cdot\,)$ over proper q-colorings of a graph $G = (V, E)$, with $q \geq 3$. Deciding whether a graph is q-colorable is a classical NP-complete constraint satisfaction problem. Here we study $\mu_G(\,\cdot\,)$ when G is sparse and random. Despite many efforts, and some fascinating conjectures from statistical physics, very little is known rigorously on the behavior of this random measure.

5.0.6 Definitions and broad picture

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. We shall consider the uniform distribution over proper q-colorings. If $x = \{x_i : in \in V\}$ denotes a q-coloring (here $x_i \in \{1, \dots, q\}$), this takes the form

$$\mu_G(x) = \frac{1}{Z_G} \prod_{(ij)\in E} \mathbb{I}(x_i \neq x_j)\,. \tag{5.1}$$

The graph will be random over the vertex set $V = [N]$. We have two examples in mind (and will switch between them):

(a) G is uniformly random with $M = N\alpha$ edges (and therefore has average degree 2α).

(b) G is a random regular graph with degree $(k + 1)$.

Heuristic statistical mechanics studies suggest a rich phase transition structure for the measure $\mu_G(\,\cdot\,)$. For any[1] $q \geq 4$, different regimes are separated by three distinct critical values of the average degree: $0 < \alpha_{\rm d}(q) < \alpha_{\rm c}(q) < \alpha_{\rm s}(q)$. Such regimes can be characterized as follows (all the statements below are understood to hold with high probability with respect to the graph choice):

[1]The case $q = 3$ is special in that $\alpha_{\rm d}(q) = \alpha_{\rm c}(q)$. The reader is invited to discuss the case $q = 2$.

I. $\alpha < \alpha_{\rm d}(q)$. The model does not undergo a phase transition. Roughly speaking, the set of proper q colorings forms a unique compact lump.

II. $\alpha_{\rm d}(q) < \alpha < \alpha_{\rm c}(q)$. The model is undergoes a phase transition. More precisely, there exists a partition of the space of configurations $\{1,\dots,q\}^V$ into $\mathcal{N}$ sets $\{\Omega_\alpha\}$, such that for any α

$$\frac{\mu(\partial_\epsilon \Omega_\alpha)}{\mu(\Omega_\alpha)[1-\mu(\Omega_\alpha)]} \le e^{-C\,N}\,. \tag{5.2}$$

Further $\mathcal{N} \doteq e^{N\Sigma}$.

III. $\alpha_{\rm c}(q) < \alpha < \alpha_{\rm s}(q)$. The situation is analogous to the previous one, but $\mathcal{N}$ is sub-exponential. More precisely, for any $\delta > 0$, a fraction $1-\delta$ of the measure μ is comprised in $\mathcal{N}(\delta)$ 'lumps', whereby $\mathcal{N}(\delta)$ converges to a finite random variable.

IV. $\alpha_{\rm s}(q) < \alpha$. A random graph is, with high probability, uncolorable.

Statistical mechanics methods provide semi-explicit expressions for the values $\alpha_{\rm d}(q)$, $\alpha_{\rm c}(q)$, $\alpha_{\rm s}(q)$. Such expressions involve solving an equation for a probability distribution over the $(q-1)$-dimensional simplex. The thresholds values are given in terms of the solution of this an equation.

5.0.7 The COL-UNCOL transition: A few simple results

Although the existence of a colorable-uncolorable transition is not established, Friedgut theory allows to make a first step in this direction.

Theorem 5.0.18. *Denote by $G_{N,\alpha}$ a uniformly random graph with N vertices and $N\alpha$ edges. For any $q \ge 3$ there exists $\alpha_{\rm s}(q;N)$ such that*

$$\lim_{N\to\infty} \mathbb{P}\{G_{N,\alpha_{\rm s}(q;N)(1-\delta)} \textit{ is } q\textit{-colorable}\} = 1\,, \tag{5.3}$$

$$\lim_{N\to\infty} \mathbb{P}\{G_{N,\alpha_{\rm s}(q;N)(1+\delta)} \textit{ is } q\textit{-colorable}\} = 0\,, \tag{5.4}$$

In the following we shall drop the N dependence from $\alpha_{\rm s}(q;N)$.

Proposition 5.0.19. *The COL-UNCOL threshold is upper bounded as*

$$\alpha_{\rm s}(q) \le \overline{\alpha}_{\rm s}(q) \equiv \frac{\log q}{\log(1-1/q)}\,. \tag{5.5}$$

Proof. The expected number of proper q-colorings is

$$\mathbb{E}\{Z_G\} = q^N \left(1-\frac{1}{q}\right)^M\,, \tag{5.6}$$

and $\mathbb{E}\{Z_G\} \to 0$ for $\alpha > \overline{\alpha}_{\rm s}(q)$. The thesis follows from Markov 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.

Recall that the k-core of a graph G is the maximal induced subgraph with minimal degree not smaller than k. A simple 'algorithmic' lower bound follows from the next remark.

Proposition 5.0.20. *If G does not contain a 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 of the adjacent edges. If G does not contain a q-core, then there exists an ordering $i(1), i(2), \dots, i(N)$ of the vertices, such the following is true. If we define G_t by $G_0 = G$ and $G_t = G_{t-1} \setminus i(t)$, then, for any $t \le N$, $i(t)$ has degree smaller than q in G_{t-1}.

The 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. □

The ordinary differential equations method can be used to establish a threshold for the appearance of a q-core in a random graph G.

Proposition 5.0.21. *Let $F(\rho; \alpha) = \mathbb{P}\{\text{Poisson}(2\alpha\rho) \ge q - 1\}$, and define (for $q \ge 3$)*

$$\alpha_{\rm core}(q) = \sup\{\alpha : \, F(\rho; \alpha) \le \rho \, \forall \rho \in [0, 1]\} \, . \tag{5.7}$$

Then, with high probability, a uniformly random graph G with $M = N\alpha$ edges over N vertices has a q-core if $\alpha > \alpha_{\rm core}(q)$, and does not have one if $\alpha < \alpha_{\rm core}(q)$.

We omit the proof of this statement (due to Wormald, Spencer and Pittel) since it loosely follows the one we explained for the 2-core of a random hypergraph. The value of $\alpha_{\rm core}(q)$ can be derived by a simple heuristic argument. For a vertex $i \in V$ we call 'q-core induced by i' the largest induced subgraph that has minimum degree not smaller than q except (possibly) at i. Given a random edge (i, j), we denote by ρ the probability that i belongs to the q-core induced by j. It is then natural to write the following equation for ρ:

$$\rho = \mathbb{P}\{\text{Poisson}(2\alpha\rho) \ge q - 1\} \, . \tag{5.8}$$

The threshold $\alpha_{\rm core}(q)$ corresponds to the appearance of a positive solution of this equation.

5.0.8 The clustering transition: The conjecture

The conjectured value for $\alpha_{\rm d}(q)$ has a particularly elegant interpretation in terms of phase transition on a tree. Let T be a Galton-Watson tree with Poisson

offspring distribution of mean 2α. With an abuse of notation, let μ denote the free boundary measure over proper colorings of T. More explicitly, a proper coloring $x = \{x_i : i \in V\}$ is sampled from μ as follows. First sample the root color uniformly are random. Then recursively, for each colored node i, sample the colors of its direct descendants uniformly at random among the ones different from x_i.

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

At small α the color at the root decorrelates from the far away ones in $\overline{\mathsf{B}}(r,t)$. At large α they remain correlated at any distance t. The 'reconstruction threshold' separates these two regimes.

Definition 5.0.22. *The reconstruction threshold $\alpha_{\rm r}(q)$ is the supremum value of α such that*

$$\lim_{t\to\infty} \mathbb{E}||\mu_{r,\overline{\mathsf{B}}(r,t)}(\,\cdot\,,\,\cdot\,) - \mu_r(\,\cdot\,)\mu_{\overline{\mathsf{B}}(r,t)}(\,\cdot\,)||_{\rm TV} = 0\,. \tag{5.9}$$

It is conjectured that $\alpha_{\rm d}(q)$ (the clustering threshold on random graphs) and $\alpha_{\rm r}(q)$ (the reconstruction threshold on random trees), do indeed coincide. We shall try to argue in favor of this conjecture in the following.

5.0.9 The clustering transition: A physicist's approach

We now present a statistical physics argument to derive the location of the clustering threshold. There are various versions of this argument and not all of them are necessarily equivalent. However, they all predict the same location for the threshold.

In trying to identify the existence of 'lumps,' the major difficulty is that we do not know, a priori, where the lumps are. However if x^* is a configuration sampled from $\mu(\,\cdot\,)$, it will fall inside one such lumps. The idea is to study how a second configuration x behaves when tilted towards the first one.

In practice we fix $x^* = \{x_i^* : i \in V\}$ and study the tilted measure $\mu(\,\cdot\,) = \mu_{G,x^*,\epsilon}(\,\cdot\,)$ defined by

$$\mu^*(x) = \frac{1}{Z} \prod_{(i,j)\in E} \mathbb{I}(x_i \neq x_j) \prod_{i\in V} \psi_\epsilon(x_i^*, x_i)\,. \tag{5.10}$$

Here ψ_ϵ is a tilting function depending on the continuous parameter ϵ. We assume $\psi_0(x,y) = 1$ identically (therefore μ reduces to the uniform measure over colorings in this case) and $\epsilon > 0$ to favor $x = y$. For instance we might take

$$\psi_\epsilon(x,y) = \exp\Big\{\epsilon\,\mathbb{I}(x=y)\Big\}\,. \tag{5.11}$$

Studying the above distribution is not an easy task, but we can hope Bethe approximation to work in this case. Messages will depend on the graph but also on x^* and ϵ. Bethe equations read

$$\nu_{i\to j}(x_i) \cong \psi_\epsilon(x_i^*, x_i) \prod_{l\in\partial i\setminus j} \sum_{x_l\neq x_i} \nu_{l\to i}(x_l)\,. \tag{5.12}$$

Introducing a shorthand for the right-hand side, we will write this equation as

$$\nu_{i\to j} = \mathsf{F}_\epsilon\{\nu_{l\to i} : l\in\partial i\setminus j\}\,. \tag{5.13}$$

Let us now assume that G is a regular graph of degree $k+1$ and x^* a uniformly random proper coloring of G. Then the message $\nu_{i\to j}$ is itself a random variable, taking values in the $(q-1)$-dimensional simplex. For $x\in\{1,\dots,q\}$, let us denote by Q_x the conditional distribution of $\nu_{i\to j}$ given that $x_i^* = x$. In formulae, for a subset A of the simplex, we have

$$Q_x(A) \equiv \mathbb{P}\left\{\nu_{i\to j}(\,\cdot\,)\in A \middle| x_i^* = x\right\}\,. \tag{5.14}$$

It is then easy to write a recursion for Q, namely

$$Q_x(A) = \sum_{x_1\dots x_k} \mu(x_1,\dots,x_k|x) \int \mathbb{I}(\mathsf{F}_\epsilon(\nu_1,\dots,\nu_k)\in A) \prod_{i=1}^k Q_{x_i}(\mathrm{d}\nu_i)\,. \tag{5.15}$$

If w no consider a random regular graph, we might hope that $\mu(x_1,\dots,x_k|x)$ converge to the analogous conditional distribution on a tree. If this is the case we obtain a fixed point equation for Q

$$Q_x(A) = \frac{1}{(q-1)^k} \sum_{x_1\dots x_k\neq x} \int \mathbb{I}(\mathsf{F}_\epsilon(\nu_1,\dots,\nu_k)\in A) \prod_{i=1}^k Q_{x_i}(\mathrm{d}\nu_i)\,. \tag{5.16}$$

It is generally believed (at least by physicists!) that the measure μ undergoes a phase transition if and only if this equation admits a non-degenerate solution.

5.0.10 Complexity

We will now introduce the statistical physics formula for the complexity of the uniform measure over clusters. We will finish discussing the relation between tree reconstruction and phase transitions in q-colorings of random graphs. Subsequently we show that proper colorings on a random regular graph with $k \ll q$ does not undergo a phase transition.

For the sake of simplicity we shall restrict ourselves to regular graphs of degree $(k+1)$.

Recall that, within Bethe-Peierls approximation, the free entropy of proper colorings is given by

$$\Phi\{\nu_{i\to j}\} = -\sum_{(i,j)\in E}\log\left\{\sum_{x_i\neq x_j}\nu_{i\to j}(x_i)\nu_{j\to i}(x_j)\right\} + \sum_{i\in V}\log\left\{\sum_x\prod_{j\in\partial i}\sum_{x_j\neq x}\nu_{j\to i}(x_j)\right\},$$

where the messages $\{\nu_{i\to j}\}$ must solve (within some accuracy) the Bethe-Peierls equations. Let $\overline{\nu}$ is the uniform distribution over $\{1,\dots,q\}$.

Consider the ϵ-tilted model in the limit $\epsilon\downarrow 0$. One solution of the BP equations is given by $\nu_{i\to j}=\overline{\nu}$. This corresponds the system being completely uncorrelated with the reference configuration. The corresponding free-entropy on a regular graph of degree $(k+1)$ is

$$\begin{aligned}\Phi(\overline{\nu}) &= N\left\{-\frac{k+1}{2}\log\left\{\sum_{x_1\neq x_2}\overline{\nu}(x_1)\overline{\nu}(x_2)\right\} + \log\left\{\sum_x\prod_{j=1}^{k+1}\sum_{x_j\neq x}\overline{\nu}(x_j)\right\}\right\}\\ &= N\log\left\{q\left[1-\frac{1}{q}\right]^{(k+1)/2}\right\}.\end{aligned}$$

However there is another 'non-trivial' solution that corresponds to the free entropy of a single partition Ω_α. The exponential growth rate of the number of pure states Σ is obtained by taking the difference of these two free entropies and dividing by the number of nodes

$$\Sigma = \lim_{N\to\infty} N^{-1}\left\{\Phi(\overline{\nu}) - \Phi\{\nu_{i\to j}\}\right\}. \tag{5.17}$$

When the reference configuration is random, the messages appearing in each term of $\Phi\{\nu_{i\to j}\}$ are themselves random, with distribution that can be expressed in terms of Q_x. By taking expectation of the above, we get

$$\begin{aligned}\Sigma &= -\frac{k+1}{2}\frac{1}{q(q-1)}\sum_{x_1\neq x_2}\int W_{\rm e}(\nu_1,\nu_2)\,Q_{x_1}(\mathrm{d}\nu_1)\,Q_{x_2}(\mathrm{d}\nu_2) + \qquad (5.18)\\ &\quad + \frac{1}{q(q-1)^{k+1}}\sum_x\sum_{x_1,\dots,x_k\neq x}\int W_{\rm v}(\nu_1,\dots,\nu_{k+1})\prod_{i=1}^{k+1}Q_{x_i}(\mathrm{d}\nu_i),\end{aligned}$$

where

$$W_{\rm e}(\nu_1,\nu_2) = \log\left\{\frac{\sum_{x_1\neq x_2}\nu_1(x_1)\nu_2(x_2)}{\sum_{x_1\neq x_2}\overline{\nu}(x_1)\overline{\nu}(x_2)}\right\}, \tag{5.19}$$

$$W_{\rm v}(\nu_1,\dots,\nu_{k+1}) = \log\left\{\frac{\sum_x\prod_{j=1}^{k+1}\sum_{x_j\neq x}\nu_j(x_j)}{\sum_x\prod_{j=1}^{k+1}\sum_{x_j\neq x}\overline{\nu}(x_j)}\right\}. \tag{5.20}$$

5.0.11 Reconstruction on trees and clustering threshold

Consider a random regular graph of degree $(k+1)$. We argued heuristically that a phase transition occurs if there exists a family of non-degenerate distributions $\{Q_x\}$ over the $(q-1)$-dimensional simplex that satisfy the equation

$$Q_x(A) = \frac{1}{(q-1)^k} \sum_{x_1,\dots,x_k \neq x} \int \mathbb{I}(\mathsf{F}(\nu_1,\dots,\nu_k) \in A) \prod_{i=1}^{k} Q_{x_i}(\mathrm{d}\nu_i)\,. \qquad (5.21)$$

The relation with the reconstruction problem on regular trees is shown in the following Proposition.

Proposition 5.0.23. *The reconstruction problem is solvable on k-regular trees if and only if Eq. (5.21) admits a non-degenerate solution.*

Proof. Consider a random (according to the free boundary Gibbs measure) proper coloring of the regular tree. Let $\nu^{(t)}$ be the marginal distribution of the root color given the colors at generation t. In formulae, for $x \in \{1,\dots,q\}$, we define

$$\nu^{(t)}(x) = \mu_{r,\overline{\mathsf{B}}(r,t)}(x|X_{\overline{\mathsf{B}}(r,t)}) = \mathbb{P}\{X_r = x|X_{\overline{\mathsf{B}}(r,t)}\}\,. \qquad (5.22)$$

Denote by $Q_x^{(t)}$ the distribution of $\nu^{(t)}$ given the root value $X_r = x$. It is immediate to show that this sequence of distributions satisfies the recursion

$$Q_x^{(t+1)}(A) = \frac{1}{(q-1)^k} \sum_{x_1,\dots,x_k \neq x} \int \mathbb{I}(\mathsf{F}(\nu_1,\dots,\nu_k) \in A) \prod_{i=1}^{k} Q^{(t)}(\mathrm{d}\nu_i)\,. \qquad (5.23)$$

The initial condition is $Q_x^{(0)} = \delta_{\nu_x}$, where ν_x is the singleton over the color x.

The sequence of distributions $\{Q^{(t)}\}$ converges weakly to a limit $Q^{(\infty)}$ because $\{\nu^{(t)}\}$ is a backward martingale. Further we have

$$||\mu_{r,\overline{\mathsf{B}}(r,t)}(\,\cdot\,,\,\cdot\,) - \mu_r(\,\cdot\,)\mu_{\overline{\mathsf{B}}(r,t)}(\,\cdot\,)||_{\mathrm{TV}} = \frac{1}{q}\sum_{x=1}^{q} \int ||\nu - \overline{\nu}||_{\mathrm{TV}}\, Q_x(\mathrm{d}\nu)\,. \qquad (5.24)$$

Finally, it follows from Bayes theorem that $Q_x^{(t)}(\mathrm{d}\nu) = q\nu(x)Q^{(t)}(\mathrm{d}\nu)$, where $Q^{(t)}$ is the unconditional distribution of $\nu^{(t)}$. As a consequence of these remarks, the reconstruction problem is solvable if and only if the distributions $Q_x^{(\infty)}$ are non-degenerate.

Therefore solvability implies the existence of a non-trivial solution of Eq. (5.21), namely $Q_x^{(\infty)}$.

To prove the converse, first notice that a sufficient condition of reconstructibility is that

$$||\mu_{r,Y(t)}(\,\cdot\,,\,\cdot\,) - \mu_r(\,\cdot\,)\mu_{Y(t)}(\,\cdot\,)||_{\mathrm{TV}} \not\to 0 \qquad (5.25)$$

for some random variable $Y(t)$ that is conditionally independent of X_r given $X_{\overline{\mathsf{B}}(r,t)}$. We then take $Y(t) = \dots$. □

Chapter 6

Reconstruction and extremality

Extremality was introduced in Chapter 4 as a sufficient condition (together with a large girth condition) for the Bethe-Peierls approximation to be accurate. Unhappily it s a difficult conditon to check. In the present Chapter we discuss a relation between extremality on random graphs and extremality on associated models trees. This is a first step towards understanding extremality on random graphs.

This paper is largely based on [17] which we refer to for further details.

6.1 Introduction and outline

Let $G = (V, E)$ be a graph, and $X = \{X_i : i \in V\}$ a proper coloring of its vertices sampled uniformly at random. The *reconstruction problem* amounts to estimating the color of a distinguished (root) vertex $r \in V$, when the colors $\{X_j = x_j : j \in U\}$ of subset of vertices are revealed. In particular, we want to understand whether the revealed values induce a substantial bias on the distribution of X_i.

As in the previous chapters, we shall consider the more general setting of *graphical models*. Such a model is defined by a graph $G = (V, E)$, and a set of weights $\psi = \{\psi_{ij} : (ij) \in E\}$, $\psi_{ij} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$. Given a graph-weights pair (G, ψ), we let

$$\mathbb{P}\big\{\underline{X} = \underline{x}\big|(G, \psi)\big\} \equiv \mu(\underline{x}) \equiv \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(x_i, x_j)\,, \tag{6.1}$$

where we assume $\psi_{ij}(x, y) = \psi_{ij}(y, x)$. The example of proper colorings is recovered by letting $\mathcal{X} = \{1, \dots, q\}$ (q being the number of colors) and $\psi_{ij}(x, y) = 1$ if $x \neq y$ and $= 0$ otherwise. Ising models from statistical mechanics provide another interesting class, whereby $\mathcal{X} = \{+1, -1\}$. In the 'ferromagnetic' case the weights are $\psi_{ij}(+, +) = \psi_{ij}(-, -) = 1 - \epsilon$ and $\psi_{ij}(+, -) = \psi_{ij}(-, +) = \epsilon$ for some $\epsilon \in [0, 1/2]$.

For economy of notation, we shall often write $\mathbb{P}\{\,\cdot\,|G\}$ as a shorthand for $\mathbb{P}\{\,\cdot\,|(G,\psi)\}$, and 'the graph G' for 'the graph-weights pair (G,ψ).' It is understood that, whenever G is given, the weights ψ are given as well. Further, for $U \subseteq V_N$, we let $\underline{X}_U = \{X_j : j \in U\}$ and $\mathbb{P}_U\{\underline{x}_U|G\} = \mathbb{P}\{\underline{X}_U = \underline{x}_U|G\}$ be its marginal distribution that can be obtained by marginalizing Eq. (6.1).

For $i, j \in V$, let $d(i,j)$ be their graph theoretic distance. Further for any $t \geq 0$, we let $\overline{\mathsf{B}}(i,t)$ be the set of vertices j such that $d(i,j) \geq t$, (and, by abuse of notation, the induced subgraph). The reconstructibility question asks whether the 'far away' variables $\underline{X}_{\overline{\mathsf{B}}(r,t)}$ provide significant information about X_r. This is captured by the following definition (recall that, given two distributions p, q on the same space $\mathcal{S}$, their total variation distance is $||p-q||_{\text{TV}} \equiv (1/2)\sum_{x\in\mathcal{S}}|p_x - q_x|$).

Definition 6.1.1. *The reconstruction problem is* (t,ε)-solvable *(*reconstructible*) for the graphical model (G,ψ) rooted at $r \in V$ if*

$$\|\mathbb{P}_{r,\overline{\mathsf{B}}(r,t)}\{\,\cdot\,,\,\cdot\,|G\} - \mathbb{P}_r\{\,\cdot\,|G\}\mathbb{P}_{\overline{\mathsf{B}}(r,t)}\{\,\cdot\,|G\}\|_{\text{TV}} \geq \varepsilon\,.$$

In the following we will consider graphs G that are themselves random. By this we mean that we will specify a joint distribution of the graph $G_N = (V_N = [N], E_N)$, of the weights $\{\psi_{ij}\}$, and of the root vertex r whose variable we are interested in reconstructing. Equation (6.1) then specifies the conditional distribution of $\underline{X}$, given the random structure (G_N,ψ) (again, we'll drop reference to ψ).

Definition 6.1.2. *The reconstruction problem is* solvable *(*reconstructible*) for the sequence of random graphical models $\{G_N\}$ if there exists $\varepsilon > 0$ such that, for all $t \geq 0$ it is (t,ε)-solvable with positive probability, i.e. if*

$$\|\mathbb{P}_{r,\overline{\mathsf{B}}(r,t)}\{\,\cdot\,,\,\cdot\,|G_N\} - \mathbb{P}_r\{\,\cdot\,|G_N\}\mathbb{P}_{\overline{\mathsf{B}}(r,t)}\{\,\cdot\,|G_N\}\|_{\text{TV}} \geq \varepsilon\,. \tag{6.2}$$

with positive probability[1]*.*

To be specific, we shall assume G_N to be a sparse random graph. In this case, any finite neighborhood of r converges in distribution to a tree [49]. Further, imagine to mark the boundary vertices of such a neighborhood, and then take the neighborhood out of G_N (thus obtaining the subgraph denoted above as $\overline{\mathsf{B}}(r,t)$). The marked vertices will be (with high probability) 'far apart' from each other in $\overline{\mathsf{B}}(r,t)$. This suggests that the corresponding random variables $\{X_j\}$ will be approximately independent when the tree-like neighborhood is taken out.

[1]Here and below, we say that the sequence of events $\{A_N\}$ holds with positive probability (wpp) if there exists $\delta > 0$ and an infinite sequence $\mathcal{N} \subseteq \mathbb{N}$, such that $\mathbb{P}\{A_N\} \geq \delta$ for any $N \in \mathcal{N}$. Notice that, in a random graph, r might be in a small connected component. Therefore Eq. (6.2) cannot be required to hold with high probability.

Hence, approximating G_N by its local tree structure might be a good way to determine correlations between X_r and the boundary variables $\{X_j : d(r,j) = t\}$. In other words, one would expect reconstructibility on G_N to be determined by reconstructibility on the associated random tree.

Of course the above conclusion does not hold in general, as it is based on a circular argument. We assumed that 'far apart' variables (with respect to the residual graph $\overline{\mathsf{B}}(r,t)$) are weakly correlated, to understand whether 'far apart' variable (in G_N) are. In fact, we will prove that tree and graph reconstruction do not coincide in the simplest example one can think of, namely the Ising ferromagnet (binary variables with attractive interactions).

On the positive side, we prove a general sufficient condition for the tree and graph behaviors to coincide. The condition has a suggestive geometrical interpretation, as it requires two independent random configurations $X^{(1)}$ and $X^{(2)}$ to be, with high probability, at an approximately fixed 'distance' from each other. In the example of coloring, we require two uniformly random independent colorings of the same graph to take the same value on about $1/q$ of the vertices. The set of 'typical configurations' looks like a sphere when regarded from any typical configuration. Under such a condition, the above argument can be put on firmer basis. We show that, once the the neighborhood of the root r is taken out, boundary variables become roughly independent. This in turns implies that graph and tree reconstruction do coincide.

We apply this sufficient condition to the Ising spin glass (where the condition can be shown to hold as a consequence of a recent result by Guerra and Toninelli [50]), and to antiferromagnetic colorings of random graphs (building on the work of Achlioptas and Naor [51]). In both cases we will introduce a family of graphical models parametrized by their average degree. It is natural to expect reconstructibility to hold at large degrees (as the graph is 'more connected') and not to hold at small average degrees (since the graph 'falls' apart into disconnected components). In the spin glass case we are indeed able to estabilish a threshold behavior (i.e. a critical degree value above which reconstruction is solvable). While we didn't achieve the same for colorings, we essentially reduced the problem to establishing a threshold for the tree model.

6.1.1 Applications and related work

Let us discuss a selection of related problems that are relevant to our work.

Markov Chain Monte Carlo (MCMC) algorithms provide a well established way of approximating marginals of the distribution (6.1). If the chain is reversible and has local updates, the mixing time is known to be related to the correlation decay properties of the stationary distribution $\mathbb{P}\{\,\cdot\,|G_N\}$ [52, 53]. In this context, correlations between X_r and $\underline{X}_{\overline{\mathsf{B}}(r,t)}$ are usually characterized by measures of the type $\Delta(t) \equiv \sup_{\underline{x}} \|\mathbb{P}_{r|\overline{\mathsf{B}}(r,t)}\{\,\cdot\,|\underline{x}_{\overline{\mathsf{B}}(r,t)}, G_N\} - \mathbb{P}_r\{\,\cdot\,|G_N\}\|_{\text{TV}}$. The 'uniqueness ' condition requires $\Delta(t)$ to decay at large t, and is easily shown to

imply non-reconstructibility. On graphs with sub-exponential growth, a fast enough decay is a necessary and sufficient condition for fast mixing. On the other hand, in more general cases this is too strong a criterion, and one might want to replace it with the non-reconstructibility one.

In [54] it was proved that non-reconstructibility is equivalent to polynomial spectral gap for a class of models on trees. The equivalence was sharpened in [55], showing that non-reconstructibility is equivalent to fast mixing in the same models. Further, [54] proved that non-reconstructibility is a necessary condition for fast mixing on general graphs. While a converse does not hold in general, non-reconstructibility is sufficient for rapid decay of the variance of local functions (which is often regarded as the criterion for fast dynamics in physics) [56].

Random constraint satisfaction problems. Given an instance of a constraint satisfaction problem (CSP), consider the uniform distribution over its solutions. This takes the form (6.1), where ψ_{ij} is the indicator function over the constraint involving variables x_i, x_j being satisfied (Eq. (6.1) is trivially generalized to k-variables constraints). For instance, in coloring it is the indicator function on $x_i \neq x_j$.

Computing the marginal $\mathbb{P}_r\{\,\cdot\,|G_N\}$ can be useful both for finding and for counting the solutions of such a CSP. Assume to be able to generate *one* uniformly random solution $\underline{X}$. In general, this is not sufficient to approximate the marginal of X_i in any meaningful way. However one can try the following: fix all the variables 'far from r' to take the same value as in the sampled configuration, namely $X_{\overline{\mathsf{B}}(r,t)}$, and compute the conditional distribution at the root. If the graph is locally tree-like, the conditional distribution of X_r can be computed through an efficient dynamic programming procedure. The result of this computation needs not to be near the actual marginal. However, non-reconstructibility implies the result to be with high probability within ε (in total variation distance) from the marginal.

As a consequence, a single sample (a single random solution $\underline{x}$) is sufficient to approximate the marginal $\mathbb{P}_r\{\,\cdot\,|G_N\}$. The situation is even simpler under the sufficient condition in our main theorem (Theorem 6.1.4). In fact this implies that the boundary condition $\underline{x}_{\overline{\mathsf{B}}(r,t)}$ can be replaced by an iid uniform boundary.

For random CSP's, G_N becomes a sparse random graph. Statistical mechanics studies [8] suggest that, for typical instances the set of solutions decomposes into 'clusters' at sufficiently large constraint density [57, 58]. This leads to the speculation that sampling from the uniform measure $\mathbb{P}\{\,\cdot\,|G_N\}$ becomes harder in this regime.

The decomposition in clusters is related to reconstructibility, as per the following heuristic argument. Assume the set of solutions to be splitted into clusters, and that two solutions whose Hamming distance is smaller than $N\varepsilon$ belong to the same cluster. Then knowing the far away variables $\underline{x}_{\overline{\mathsf{B}}(r,t)}$ (i.e. all but a bounded number of variables) does determine the cluster. This in turns provides some

information on X_r.

In fact, it was conjectured in [2] that tree and graph reconstruction thresholds should coincide for 'frustrated' models on random graphs. Both should coincide with the clustering phase transition in the set of solutions [59].

Statistical inference and message passing. Graphical models of the form (6.1) are used in a number of contexts, from image processing to artificial intelligence, etc. Statistical inference requires to compute marginals of such a distribution and message passing algorithms (in particular, belief propagation, BP) are the methods of choice for accomplishing this task.

The (unproven) assumption in such algorithms is that, if a tree neighborhood of vertex i is cut away from G_N, then the variables $\{X_j\}$ on the boundary of this tree are approximately independent. Assuming the marginals of the boundary variables to be known, the marginal of X_i can be computed through dynamic programming. Of course the marginals to start from are unknown. However, the dynamic programming procedure defines an mapping on the marginals themselves. In BP this mapping is iterated recursively over all the nodes, without convergence guarantees.

Lemma 6.3.2 shows that, under the stated conditions, the required independence condition does indeed hold. As stressed above, this is instrumental in proving equivalence of graph and tree reconstructibility in Theorem 6.1.4.

Reconstruction problems also emerge in a variety of other contexts: (*i*) Phylogeny [39] (given some evolved genomes, one aims at reconstructing the genome of their common ancestor); (*ii*) Network tomography [40] (given end-to-end delays in a computer network, infer the link delays in its interior); (*iii*) Statistical mechanics [16, 60] (reconstruction being related to the extremality of Gibbs measures).

6.1.2 Previous results

If the graph G_N is a tree, the reconstruction problem is relatively well understood [11]. The fundamental reason is that the distribution $\mathbb{P}\{\underline{X} = \underline{x}|G_N\}$ admits a simple description. First sample the root variable X_r from its marginal $\mathbb{P}\{X_r = x_r|G_N\}$. Then recursively for each node j, sample its children $\{X_l\}$ independently conditional on their parent value.

Because of this Markov structure, one can prove a recursive distributional equation for the conditional marginal at the root $\mathbb{P}_{r|\overline{\mathsf{B}}(r,t)}\{\,\cdot\,|\underline{X}_{\overline{\mathsf{B}}(r,t)}, G_N\} \equiv \eta_t(\,\cdot\,)$ given the variable values at generation t. Notice that this is a random quantity even for a deterministic graph G_N, because $\underline{X}_{\overline{\mathsf{B}}(r,t)}$ is itself drawn randomly from the distribution $\mathbb{P}\{\,\cdot\,|G_N\}$. Further, it contains all the information (it is a 'sufficient statistic') in the boundary about the root variable X_r. In fact asymptotic behavior of $\eta_t(\,\cdot\,)$ as $t \to \infty$ then determines the solvability of the reconstruction problem. Studying the asymptotic behavior of the sequence $\eta_t(\,\cdot\,)$ (which satisfies

a recursive distributional equation) is the standard approach to tree reconstruction.

Among the other results, reconstructibility has been thoroughly characterized for Ising models on generic trees [60, 61, 62]. For an infinite tree T the reconstruction problem is solvable if and only if $\mathrm{br}(\mathsf{T})(1-2\epsilon)^2 > 1$, whereby (for the cases treated below) $\mathrm{br}(\mathsf{T})$ coincides with the mean descendant number of any vertex. This result establishes a sharp threshold in the tree average degree (or in the parameter ϵ), that we shall generalize to random graphs below. However, as we will see, the behavior is richer than in the tree case.

Reconstruction on general graphs poses new challenges, since the above recursive description of the measure $\mathbb{P}\{\,\cdot\,|G_N\}$ is lacking. The result of [54] allows to deduce non-reconstructibility from fast mixing of reversible MCMC with local updates. However, proving fast mixing is far from an easy task. Further, the converse does not usually hold (one can have slow mixing and non-reconstructibility).

An exception is provided by the recent paper by Mossel, Weitz and Wormald [63] that establishes a threshold for fast mixing for weighted independent sets on random bipartite graphs (the threshold being in the weight parameter λ). Arguing as in Section [17], it can be shown that this is also the graph reconstruction threshold. This result is analogous to ours for the ferromagnetic Ising model: it provides an example in which the graph reconstruction threshold does not coincide with the tree reconstruction threshold. In both cases the graph reconstruction threshold coincides instead with the tree 'uniqueness threshold' (i.e. the critical parameter for the uniqueness condition mentioned above to hold).

6.1.3 Basic definitions

We consider two families of random graphical models: *regular* and *Poisson* models. In both cases the root $r \in V$ is uniformly random and independent of G_N. A *regular ensemble* is specified by assigning an alphabet $\mathcal{X}$ (the variable range), a degree $(k+1)$ and and edge weight $\psi : \mathcal{X}\times\mathcal{X} \to \mathbb{R}_+$. For any $N > 0$, a random model is defined by letting G_N be a uniformly random regular graph of degree $(k+1)$ over vertex set $V = [N]$. The joint distribution of $(X_1, \dots, X_N)$ is given by Eq. (6.1), with $\psi_{ij}(\,\cdot\,,\,\cdot\,) = \psi(\,\cdot\,,\,\cdot\,)$. A variation of this ensemble is obtained by letting G be a random regular *multi-graph* according to the configuration model [64] (notice that our definitions make sense for multigraphs as well). Indeed in the following we assume this model when working with regular graphs.

As an example, the *random regular Ising ferromagnet* is obtained by letting $\mathcal{X} = \{+1, -1\}$ and, for some $\epsilon \le 1/2$, $\psi(x_1, x_2) = 1-\epsilon$ if $x_1 = x_2$ and $\psi(x_1, x_2) = \epsilon$ otherwise.

Specifying a *Poisson ensemble* requires an alphabet $\mathcal{X}$, a *density* $\gamma \in \mathbb{R}_+$, a finite collection of weights $\{\psi_a(\,\cdot\,,\,\cdot\,) : a \in \mathcal{C}\}$, and a probability distribution $\{p(a) : a \in \mathcal{C}\}$ over the weights. In this case G is a multigraph where the number edges among any pair of vertices i and j is an independent Poisson random

variable of parameter $2\gamma/n$. Each loop (i,i) is present with multiplicity which is Poisson of mean[2] γ/n. Finally, for each edge in the multi-graph, we draw an independent random variable a with distribution $p(\,\cdot\,)$ and set $\psi_{ij}(\,\cdot\,,\,\cdot\,) = \psi_a(\,\cdot\,,\,\cdot\,)$.

Two examples of Poisson ensembles to be treated below are the *Ising spin glass*, and *antiferromagnetic colorings* (aka 'antiferromagnetic Potts model'). In the first case $\mathcal{X} = \{+1,-1\}$ and two type of weights appear with equal probability (i.e. $\mathcal{C} = \{+,-\}$ and $p(+) = p(-) = 1/2$): $\psi_+(x_1,x_2) = 1-\epsilon$ for $x_1 = x_2$, $\psi_+(x_1,x_2) = \epsilon$ for $x_1 \neq x_2$, while $\psi_-(x_1,x_2) = \epsilon$ for $x_1 = x_2$, $\psi_-(x_1,x_2) = 1-\epsilon$ for $x_1 \neq x_2$. For proper colorings $\mathcal{X} = \{1,\dots,q\}$, and $|\mathcal{C}| = 1$ with $\psi(x_1,x_2) = 1$ if $x_1 \neq x_2$, and $\psi(x_1,x_2) = \epsilon < 1$ otherwise (for $\epsilon = 0$ one recovers the uniform measure over proper colorings of G).

Both graphical model ensembles defined above converge locally to trees. In the case of regular models, the corresponding tree model is an infinite rooted tree of uniform degree $(k+1)$, each edge being associated the same weight $\psi(\,\cdot\,,\,\cdot\,)$. For Poisson models, the relevant tree is a rooted Galton-Watson tree with Poisson distributed degrees of mean 2γ. Each edge carries the weight $\psi_a(\,\cdot\,,\,\cdot\,)$ independently with probability $p(a)$.

Given such infinite weighted trees, let T_ℓ, $\ell \geq 0$ be the weighted subgraph obtained by truncating it at depth ℓ. One can introduce random variables $\underline{X} = \{X_j : j \in \mathsf{T}_\ell\}$, by defining $\mathbb{P}\{\underline{X} = \underline{x}|\mathsf{T}_\ell\}$ as in Eq. (6.1) (with G replaced by T_ℓ). With an abuse of notation we shall call r the root of T_ℓ. It is natural to ask whether reconstruction on the original graphical models and on the corresponding trees are related.

Definition 6.1.3. *Consider a sequence of random graphical models $\{G_N\}$ (distributed according either to the regular or to the Poisson ensemble), and let $\{\mathsf{T}_\ell\}$ be the corresponding sequence of tree graphical models. We say that the reconstruction problem is* tree-solvable *for the sequence $\{G_N\}$ if there exists $\varepsilon > 0$ such that, for any $t \geq 0$*

$$\|\mathbb{P}_{r,\overline{\mathsf{B}}(r,t)}\{\,\cdot\,,\,\cdot\,|\mathsf{T}_\ell\} - \mathbb{P}_r\{\,\cdot\,|\mathsf{T}_\ell\}\mathbb{P}_{\overline{\mathsf{B}}(r,t)}\{\,\cdot\,|\mathsf{T}_\ell\}\|_{\text{TV}} > \varepsilon\,, \tag{6.3}$$

with positive probability (as $\ell \to \infty$)

Notice that tree-reconstruction is actually a question on the sequence of tree graphical models $\{\mathsf{T}_\ell\}$ indexed by ℓ. The only role of the original random graphs sequence $\{G_N\}$ is to determine the distribution of T_ℓ.

Despite the similarity of Eqs. (6.3) and (6.2), passing from the original graph to the tree is a huge simplification because $\mathbb{P}\{\,\cdot\,|\mathsf{T}_\ell\}$ has a simple description as mentioned above. For instance, in the case of a ferromagnetic Ising model, one can sample the variables X_j on the tree through a 'broadcast' process. First,

[2]Notice that in a typical realization there will be only a few loops and non-simple edges.

generate the root value X_r uniformly at random in $\{+1, -1\}$. Then recursively, for each node j, generate the values of its children $\{l\}$ conditional on $X_j = x_j$ by letting $X_l = x_j$ independently with probability $1 - \epsilon$, and $X_l = -x_j$ otherwise. Analogous descriptions exist for the spin-glass and colorings models.

6.1.4 Main results

Our first result is a sufficient condition for graph-reconstruction to be equivalent to tree reconstruction. In order to phrase it, we need to define the 'two-replicas type.' Consider a graphical model G_N and two two iid assignments of the variables $\underline{X}^{(1)}$, $\underline{X}^{(2)}$ with common distribution $\mathbb{P}\{\,\cdot\,|G_N\}$ (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 $X_j^{(2)} = y$. (Conversely, the set of distributions ν on $\mathcal{X} \times \mathcal{X}$ such that $N\nu(x,y) \in \mathbb{N}$ will be called the set of *valid two-replicas types* $\mathcal{R}_N$. When we drop the constraint $N\nu(x,y) \in \mathbb{N}$, we shall use $\mathcal{R}$.)

The matrix ν is random. If $\mathbb{P}\{\,\cdot\,|G_N\}$ were the uniform distribution, then ν would concentrate around $\overline{\nu}(x,y) \equiv 1/|\mathcal{X}|^2$. Our sufficient condition requires this to be approximately true.

Theorem 6.1.4. *Consider a sequence of random Poisson graphical models $\{G_N\}$, and let $\nu(\,\cdot\,,\,\cdot\,)$ be the type of two iid replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$, and $\Delta\nu(x,y) \equiv \nu(x,y) - \overline{\nu}(x,y)$. Assume that, for any $x \in \mathcal{X}$,*

$$\mathbb{E}\left\{\left[\Delta\nu(x,x) - 2|\mathcal{X}|^{-1}\sum_{x'}\Delta\nu(x,x')\right]^2\right\} \stackrel{N}{\to} 0\,. \tag{6.4}$$

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

Remark 1: Notice that the expectation in Eq. (6.4) is both over the two replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$ (which the type $\nu(\,\cdot\,,\,\cdot\,)$ is a function of) conditional on G_N, and over G_N. Explicitly $\mathbb{E}\{\cdots\} = \mathbb{E}\{\mathbb{E}[\cdots|G_N]\}$. *Remark 2:* In fact, as is hinted by the proof, the condition (6.4) can be weakened, e.g. $\overline{\nu}(\,\cdot\,\cdot\,)$ can be chosen more generally than the uniform matrix. This will be treated in a longer publication.

The condition (6.4) emerges naturally in a variety of contexts, a notable one being second moment method applied to random constraint satisfaction problems [65]. As an example, consider proper colorings of random graphs. In bounding on the colorability threshold, one computes the second moment of the number of colorings, and, as an intermediate step, an upper bound on the large deviations of the type ν. Oversimplifying, one might interpret Theorem 6.1.4 by saying that, when second moment method works, then tree and graph reconstruction are equivalent. Building on [51] we can thus establish the following.

Theorem 6.1.5. *Consider antiferromagnetic q-colorings of a Poisson random graph and let* $\gamma_q \equiv (q-1)\log(q-1)$. *If* $\gamma \in [0, \gamma_q)$ *and* $\epsilon \in [0,1]$, *then the reconstruction problem is solvable if and only if it is tree solvable.*

A proof for $\epsilon > 0$ and $\gamma \in [0, \gamma_q] \setminus \Gamma$ with Γ countable was provided in [17]. The complete proof was obtained only recently (A. M., R. Restrepo and P. Tetali, in preparation).

The above theorems might suggests that graph and tree reconstruction do generally coincide. This expectation is falsified by the simplest possible example: the Ising model. This has been studied in depth for trees [60, 61, 62]. If the tree is regular with degree $(k+1)$, the problem is solvable if and only if $k(1-2\epsilon)^2 > 1$. The situation changes dramatically for graphs.

Theorem 6.1.6. *Reconstruction is solvable for random regular Ising ferromagnets if and only if* $k(1-2\epsilon) > 1$.

This result possibly generalizes to Ising ferromagnets on other graphs that converge locally to trees. The proof of reconstructibility for $k(1-2\epsilon) > 1$ 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 fast mixing.

On the other hand, for an Ising spin-glass, the tree and graph thresholds do coincide. In fact, for an Ising model on a Galton-Watson tree with Poisson(2γ) offspring distribution, reconstruction is solvable if and only if $2\gamma(1-2\epsilon)^2 > 1$ [61]. The corresponding graph result is established below.

Theorem 6.1.7. *Reconstruction is solvable for Poisson Ising spin-glasses if* $2\gamma(1-2\epsilon)^2 > 1$, *and it is unsolvable if* $2\gamma(1-2\epsilon)^2 < 1$.

6.2 Random graph preliminaries

Let us start with a few more notations. Given $i \in V$, and $t \in \mathbb{N}$, $\mathsf{B}(i,t)$ is the set of vertices j such that $d(i,j) \le t$ (as well as the subgraph formed by those vertices and by edges that are not in $\overline{\mathsf{B}}(i,t)$). Further we introduce the set of vertices $\mathsf{D}(i,t) \equiv \mathsf{B}(i,t) \cap \overline{\mathsf{B}}(i,t)$.

The proof of Theorem 6.1.4 relies on two remarkable properties of Poisson graphical models: the local convergence of $\mathsf{B}(r,t)$ to the corresponding Galton-Watson tree of depth t (whose straightforward proof we omit), and a form of independence of $\overline{\mathsf{B}}(r,t)$ relatively to $\mathsf{B}(r,t)$. Notice that, because of the symmetry of the graph distribution under permutation of the vertices, we can fix r to be a deterministic vertex (say, $r = 1$).

Proposition 6.2.1. *Let* $\mathsf{B}(r,t)$ *be the depth-t neighborhood of the root in a Poisson random graph* G_N, *and* T_t *a Galton-Watson tree of depth t and offspring*

distribution Poisson(2γ). *Given any (labeled) tree* T_*, *we write* $\mathsf{B}(r,t) \simeq \mathsf{T}_*$ *if* T_* *is obtained by the depth-first relabeling of* $\mathsf{B}(r,t)$ *following a pre-established convention*[3]. *Then* $\mathbb{P}\{\mathsf{B}(r,t) \simeq \mathsf{T}_*\}$ *converges to* $\mathbb{P}\{\mathsf{T}_t \simeq \mathsf{T}_*\}$ *as* $N \to \infty$.

Proposition 6.2.2. *Let* $\mathsf{B}(r,t)$ *be the depth-t neighborhood of the root in a Poisson random graph* G_N. *Then, for any* $\lambda > 0$ *there exists* $C(\lambda,t)$ *such that, for any* N, $M \geq 0$

$$\mathbb{P}\{|\mathsf{B}(r,t)| \geq M\} \leq C(\lambda,t)\,\lambda^{-M}\,. \tag{6.5}$$

Proof. Imagine to explore $\mathsf{B}(r,t)$ in breadth-first fashion. For each t, $|\mathsf{B}(r,t+1)| - |\mathsf{B}(r,t)|$ is upper bounded by the sum of $|\mathsf{D}(r,t)|$ iid binomial random variables with parameters $N - |\mathsf{B}(r,t)|$ and $1 - e^{-2\gamma/N} \leq 2\gamma/N$ (the number of neighbors of each node in $\mathsf{D}(r,t)$). Therefore $|\mathsf{B}(r,t)|$ is stochastically dominated by $\sum_{s=0}^{t} Z_N(s)$, where $\{Z_N(t)\}$ is a Galton-Watson process with offspring distribution Binom($N, 2\gamma/N$). By Markov inequality $\mathbb{P}\{|\mathsf{B}(r,t)| \geq M\} \leq \mathbb{E}\{\lambda^{\sum_{s=0}^{t} Z_N(s)}\}\,\lambda^{-M}$. By elementary branching processes theory $g_t^N(\lambda) \equiv \mathbb{E}\{\lambda^{\sum_{s=0}^{t} Z_N(s)}\}$ satisfies the recursion $g_{t+1}^N(\lambda) = \lambda\xi_N(g_t^N(\lambda))$, $g_0^N(\lambda) = \lambda$, with $\xi_N(\lambda) = \lambda(1 + 2\gamma(\lambda-1)/N)^N$. The thesis follows by $g_t^N(\lambda) \leq g_t(\lambda)$, the latter being obtained by replacing $\xi_N(\lambda)$ with $\xi(\lambda) = e^{2\gamma(\lambda-1)} \geq \xi_N(\lambda)$. □

Proposition 6.2.3. *Let* G_N *be a Poisson random graph on vertex set* $[N]$ *and edge probability* $p = 2\gamma/N$. *Then, conditional on* $\mathsf{B}(r,t)$, $\overline{\mathsf{B}}(r,t)$ *is a Poisson random graph on vertex set* $[N] \setminus \mathsf{B}(r,t-1)$ *and same edge probability.*

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

On the other hand, $\overline{\mathsf{B}}(r,t)$ is the graph with vertices set $[N] \setminus \mathsf{G}(t)$ and edge set $(k,l) \in G_N$ such that $k, l \notin \mathsf{G}(t-1)$. Since this set of vertices couples is disjoint from the one we are conditioning upon, and by independence of edges in G_N, the claim follows. □

6.3 Proof of Theorem 6.1.4

We start from a simple technical result.

Lemma 6.3.1. *Let* p, q *be probability distribution over a finite set* $\mathcal{S}$, *and denote by* $q_0(x) = 1/|\mathcal{S}|$ *the uniform distribution over the same set. Define* $\widehat{p}(x) \equiv p(x)q(x)/z$, *where* $z \equiv \sum_x p(x)q(x)$. *Then* $||\widehat{p} - p||_{\text{TV}} \leq 3|\mathcal{S}|^2\,||q - q_0||_{\text{TV}}$.

[3]For instance one might agree to preserve the original lexicographic order among siblings

Proof. Since $||\widehat{p} - p||_{\text{TV}} \leq 1$ we can assume, without loss of generality, that $||q - q_0||_{\text{TV}} \leq (2|\mathcal{S}|)^{-1}$. If we write $p(x) = p(x)q_0(x)/z_0$, with $z_0 = 1/|\mathcal{S}|$, then $|z - z_0| \leq |\sum_x [p(x)q(x) - p(x)q_0(x)]| \leq ||q - q_0||_{\text{TV}}$ and in particular $z \geq z_0/2$. From triangular inequality we have on the other hand

$$||\widehat{p} - p||_{\text{TV}} \leq \frac{1}{2} \left| z^{-1} - z_0^{-1} \right| + \frac{1}{2z_0} \sum_x p(x)|q(x) - q_0(x)| \, .$$

Using $|x^{-1} - y^{-1}| \leq |x - y|/\min(x, y)^2$, the first term is bounded by $2|z - z_0|/z_0^2 \leq 2|\mathcal{S}|^2 ||q - q_0||_{\text{TV}}$. The second is at most $|\mathcal{S}| \, ||q - q_0||_{\text{TV}}$ which proves the thesis. □

In order to prove Theorem 6.1.4 we first establish that, under the condition (6.4), any (fixed) subset of the variables $\{X_1, \dots, X_N\}$ is (approximately) uniformly distributed.

Proposition 6.3.2. *Let $i(1), \dots, i(k) \subseteq [N]$ be any (fixed) set of vertices, and $\xi_1, \dots, \xi_k \in \mathcal{X}$. Then, under the hypotheses of Theorem 6.1.4, for any $\varepsilon > 0$*

$$\left| \mathbb{P}_{i(1),\dots,i(k)}\{\xi_1, \dots, \xi_k | G_N\} - \frac{1}{|\mathcal{X}|^k} \right| \leq \varepsilon \, , \tag{6.6}$$

with high probability.

Proof. Given two replicas $\underline{X}^{(1)}$, $\underline{X}^{(2)}$, define, for $\xi \in \mathcal{X}$ (with $\mathbb{I}_{\dots}$ the indicator function)

$$\mathsf{Q}(\xi) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \mathbb{I}_{X_i^{(1)} = \xi} - \frac{1}{|\mathcal{X}|} \right\} \left\{ \mathbb{I}_{X_i^{(2)} = \xi} - \frac{1}{|\mathcal{X}|} \right\} .$$

Notice that $\mathsf{Q}(\xi) = \Delta\nu(\xi, \xi) - (2/|\mathcal{X}|) \sum_x \Delta\nu(\xi, x)$ is the quantity in Eq. (6.4). Therefore, under the hypothesis of Theorem 6.1.4, $\mathbb{E}\{\mathsf{Q}(\xi)^2\} \stackrel{N}{\to} 0$. Further, since $|Q(\xi)| \leq 1$ and using Cauchy-Schwarz, for any $\xi_1, \dots, \xi_k \in \mathcal{X}$

$$|\mathbb{E}\left\{\mathsf{Q}(\xi_1) \cdots \mathsf{Q}(\xi_k)\right\}| \leq \mathbb{E}|\mathsf{Q}(\xi_1)| \stackrel{N}{\to} 0 \, .$$

If we denote by $\mathsf{Q}_i(\xi)$ the quantity on the right hand side of the sum in Eq. (6.7) then $\mathsf{Q}(\xi)$ is the uniform average of $\mathsf{Q}_i(\xi)$ over a uniformly random $i \in [N]$. By symmetry of the graph distribution with respect to permutation of the vertices in $[N]$, and since $|\mathsf{Q}(\xi)| \leq 1$ we get

$$\begin{aligned} \mathbb{E}\left\{\mathsf{Q}(\xi_1) \cdots \mathsf{Q}(\xi_k)\right\} &= \mathbb{E}\left\{\mathsf{Q}_{i(1)}(\xi_1) \cdots \mathsf{Q}_{i(k)}(\xi_k)\right\} + \varepsilon_{k,N} \\ &= \mathbb{E}\{\mathbb{E}\{\prod_{a=1}^{k} (\mathbb{I}_{X_{i(a)} = \xi_a} - |\mathcal{X}|^{-1}) | G_N\}^2\} + \varepsilon_{k,N} \, , \end{aligned}$$

where $|\varepsilon_{k,N}|$ is upper bounded by the probability that k random variable uniform in $[N]$ are not distinct (which is $O(1/N)$). Therefore the expectation on right hand side vanishes as $N \to \infty$ as well, which implies (since the quantity below is, again, bounded by 1)

$$\left| \mathbb{E}\left\{ \prod_{a=1}^{k} (\mathbb{I}_{X_i(a)=\xi_a} - |\mathcal{X}|^{-1}) \middle| G_N \right\} \right| \leq \varepsilon \tag{6.7}$$

with high probability for any $\varepsilon > 0$. The proof is completed by noting that the left hand side of Eq. (6.6) can be written as

$$\left| \sum_{\emptyset \neq U \subseteq [k]} \mathbb{E}\left\{ \prod_{a \in U} (\mathbb{I}_{X_{i(a)}=\xi_{i(a)}} - |\mathcal{X}|^{-1}) \middle| G_N \right\} \right| \leq 2^k \varepsilon \,,$$

where the last bound holds whp thanks to Eq. (6.7) and ε can eventually be rescaled. □

In order to write the proof Theorem 6.1.4 we need to introduce a few shorthands. Given a graphical model G_N, and $U \subseteq [N]$, we let $\mu_U(\underline{x}_U) \equiv \mathbb{P}\{\underline{X}_U = \underline{x}_U | G_N\}$ (omitting subscripts if $U = V$). If r is its root, $\ell \in \mathbb{N}$ and $U \subseteq \mathsf{B}(r,\ell)$, we define $\mu^{<}_U(\underline{x}_U) \equiv \mathbb{P}\{\underline{X}_U = \underline{x}_U | \mathsf{B}(r,\ell)\}$ (i.e. $\mu^{<}$ is the distribution obtained by restricting the product in (6.1) to edges $(i,j) \in \mathsf{B}(r,\ell)$). Analogously $\mu^{>}_U(\underline{x}_U) \equiv \mathbb{P}\left\{\underline{X}_U = \underline{x}_U | \overline{\mathsf{B}}(r,\ell)\right\}$. Finally for $U \subseteq [N]$, we let $\rho_U(\underline{x}_U) = 1/|\mathcal{X}|^{|U|}$ be the uniform distribution on $\mathcal{X}^U$.

Lemma 6.3.3. *Let G_N be a graphical model rooted at r, and $\ell \in \mathbb{N}$. Then for any $t \leq \ell$,*

$$\begin{aligned} \left| ||\mu_{r,\overline{\mathsf{B}}(r,t)} - \mu_r \mu_{\overline{\mathsf{B}}(r,t)}||_{\mathrm{TV}} - ||\mu^{<}_{r,\overline{\mathsf{B}}(r,t)} - \mu^{<}_r \mu^{<}_{\overline{\mathsf{B}}(r,t)}||_{\mathrm{TV}} \right| \leq \\ \leq 9|\mathcal{X}|^{2|\mathsf{B}(r,\ell)|} \, ||\mu^{>}_{\mathsf{D}(r,\ell)} - \rho_{\mathsf{D}(r,\ell)}||_{\mathrm{TV}} \,. \end{aligned} \tag{6.8}$$

Proof. First notice that, by elementary properties of the total variation distance, $||\mu_U - \mu^{<}_U||_{\mathrm{TV}} \leq ||\mu_{\overline{\mathsf{B}}(r,\ell)} - \mu^{<}_{\overline{\mathsf{B}}(r,\ell)}||_{\mathrm{TV}}$ for any $U \subseteq \mathsf{B}(r,\ell)$. Applying this remark and triangular inequality, the left hand side of Eq. (6.8) can be upper bounded by $3\,||\mu_{\overline{\mathsf{B}}(r,\ell)} - \mu^{<}_{\overline{\mathsf{B}}(r,\ell)}||_{\mathrm{TV}}$. Next notice that, as a consequence of Eq. (6.1) and of the fact that $\mathsf{B}(r,\ell)$ and $\overline{\mathsf{B}}(r,\ell)$ are edge disjoint (and using the shorthands $\mathsf{B}(\ell)$ and $\mathsf{D}(\ell)$ for $\mathsf{B}(r,\ell)$ and $\mathsf{D}(r,\ell)$)

$$\mu_{\mathsf{B}(\ell)}(\underline{x}_{\mathsf{B}(\ell)}) = \frac{\mu^{<}_{\mathsf{B}(\ell)}(\underline{x}_{\mathsf{B}(\ell)}) \mu^{>}_{\mathsf{D}(\ell)}(\underline{x}_{\mathsf{D}(\ell)})}{\sum_{\underline{x}'_{\mathsf{B}(\ell)}} \mu^{<}_{\mathsf{B}(\ell)}(\underline{x}'_{\mathsf{B}(\ell)}) \mu^{>}_{\mathsf{D}(\ell)}(\underline{x}'_{\mathsf{D}(\ell)})} \,.$$

We can therefore apply Lemma 6.3.1 whereby p is $\mu^{<}_{\mathsf{B}(\ell)}$, $\widehat{p}$ is $\mu_{\mathsf{B}(\ell)}$, q is $\mu^{>}_{\mathsf{D}(\ell)}$, and $\mathcal{S} = \mathcal{X}^{\mathsf{B}(\ell)}$. This proves the thesis. □

Proof of Theorem 6.1.4. Let Δ_N denote the left hand side of Eq. (6.8). We claim that its expectation (with respect to a random graph G_N) vanishes as $N \to \infty$. Since the probability that $\mathsf{B}(r,\ell) \geq M$ can be made arbitrarily small by letting M large enough, cf. Lemma 6.2.2, and using the fact that the left hand side of Eq. (6.8) is bounded by 1, it is sufficient to prove that

$$\sum_{|\mathsf{G}|\leq M} \mathbb{P}\{\mathsf{B}(r,\ell) = \mathsf{G}\}\mathbb{E}\{\Delta_N | \mathsf{B}(r,\ell) = \mathsf{G}\} \leq$$
$$\leq K^{M+1} \sum_{|\mathsf{G}|\leq M} \mathbb{E}\{||\mu^{>}_{\mathsf{D}(r,\ell)} - \rho_{\mathsf{D}(r,\ell)}||_{\mathrm{TV}} | \mathsf{B}(r,\ell) = \mathsf{G}\}\,,$$

vanishes as $N \to \infty$. Each term in the sum is the expectation, with respect to a random graph over $N - |\mathsf{G}| \geq N - M$ vertices of the total variation distance between the joint distribution of a fixed set of vertices, and the uniform distribution (for $\mathsf{D} = \mathsf{D}(r,\ell)$):

$$||\mu^{>}_{\mathsf{D}} - \rho_{\mathsf{D}}||_{\mathrm{TV}} = \frac{1}{2}\sum_{\underline{x}_{\mathsf{D}}} \left|\mathbb{P}_{\mathsf{D}}\{\underline{x}_{\mathsf{D}} | \overline{\mathsf{B}}(i,\ell)\} - |\mathcal{X}|^{-|\mathsf{D}|}\right|\,.$$

This vanishes by Lemma 6.3.2, thus proving the above claim.

This implies that there exists $\varepsilon > 0$ such that $||\mu_{r,\overline{\mathsf{B}}(r,t)} - \mu_r \mu_{\overline{\mathsf{B}}(r,t)}||_{\mathrm{TV}} \geq \varepsilon$ with positive probability, if and only if there exists $\varepsilon' > 0$ such that $||\mu^{<}_{r,\overline{\mathsf{B}}(r,t)} - \mu^{<}_r \mu^{\leq}_{\overline{\mathsf{B}}(r,t)}||_{\mathrm{TV}} \geq \varepsilon'$ with positive probability. In other words, since $\mu(\,\cdot\,) \equiv \mathbb{P}\{\cdots | G_N\}$, reconstruction is solvable if and only if $||\mu^{<}_{r,\overline{\mathsf{B}}(r,t)} - \mu^{<}_r \mu^{\leq}_{\overline{\mathsf{B}}(r,t)}||_{\mathrm{TV}} \geq \varepsilon'$ with positive probability.

Finally, recall that $\mu^{<}(\,\cdot\,) \equiv \mathbb{P}\{\,\cdot\,|\mathsf{B}(r,\ell)\}$ and that $\mathsf{B}(r,\ell)$ converges in distribution to $\mathsf{T}(\ell)$, by Lemma 6.2.1. Since $||\mu^{<}_{r,\overline{\mathsf{B}}(i,t)} - \mu^{<}_r \mu^{\leq}_{\overline{\mathsf{B}}(r,t)}||_{\mathrm{TV}}$ is a bounded function of $\overline{\mathsf{B}}(r,t)$ (and using as above Lemma 6.2.2 to reduce to a finite set of graphs), it converges in distribution to $\|\mathbb{P}_{r,\overline{\mathsf{B}}(r,t)}\{\,\cdot\,,\,\cdot\,|\mathsf{T}_\ell\} - \mathbb{P}_r\{\,\cdot\,|\mathsf{T}_\ell\}\mathbb{P}_{\overline{\mathsf{B}}(r,t)}\{\,\cdot\,|\mathsf{T}_\ell\}\|_{\mathrm{TV}}$. We conclude that $||\mu^{<}_{r,\overline{\mathsf{B}}(i,t)} - \mu^{<}_r \mu^{\leq}_{\overline{\mathsf{B}}(r,t)}||_{\mathrm{TV}} \geq \varepsilon'$ with positive probability if and only if reconstruction is tree solvable, thus proving the thesis. □

Chapter 7

Peeling algorithms and finite-size scaling

7.1 Random hypergraphs, their core and a peeling algorithm

A hyper-graph G, with n hyper-edges, each of whom is a collection of vertices, is identified with a bipartite, factor graph, having two types of nodes: v-nodes, corresponding to hyper-edges, and c-nodes to vertices. For example, a hyper-graph in the ensemble $\mathcal{G}_l(n, m)$ consists of a set $V \equiv [n]$ of v-nodes, each corresponding to an l-tuple of vertices from the set $C \equiv [m]$ of c-nodes, and an ordered list of edges, i.e. couples (i, a) with $i \in V$ and $a \in C$

$$E = [(1, a_1), (1, a_2), \ldots, (1, a_l); (2, a_{l+1}), \ldots; (n, a_{(n-1)l+1}), \ldots, (n, a_{nl})] \,,$$

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*).

Definition 7.1.1. *The* k-core *of a non-directed graph G is the unique subgraph obtained by recursively removing all vertices of degree less than k. In particular, the* 2*-core, hereafter called the* core *of G, is the maximal collection of edges having no vertex appearing in only one of them (and we use the same term for the induced subgraph). Similarly, the core of an hyper-graph is the maximal collection of hyper-edges (i.e. v-nodes) within which no vertex (i.e. c-node) appears only once (when counting their multiplicities).*

The core of a hyper-graph plays an important role in the analysis of many combinatorial problems.

For example, Karp and Sipser [36] 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, [36] 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. [21, 27] 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 halt 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 [38]). Here a hyper-edge of G is associated to each variable and a vertex to each equation, so the successful decoding finds the unique solution if and only if the core of G is empty.

The 'dual' problem is XOR-SAT, where given a linear system over m binary variables, composed of n equations modulo 2, each involving exactly $l \geq 3$ variables, the 'leaf removal' algorithm of [26, 44] recursively selects a variable that appears in a single equation, and eliminates the corresponding equation from the system (since this equation can always be satisfied by properly setting the selected variable). If all the equations are removed by this procedure, a solution can be constructed by running the process backward and fixing along the way the selected variables. Associate a hyper-graph G of m vertices and n hyper-edges to the linear system, where hyper-edge e is incident on vertex i if and only if the corresponding equation involves the i-th variable with a non-zero coefficient, noting that the leaf removal algorithm is successful if and only if the core of the

corresponding hyper-graph G is empty.

We focus hereafter on the relevant ensemble $\mathcal{G} = \mathcal{G}_l(n, m)$ and uniformly select a hyper-graph G from among the m^{nl} hyper-graphs in $\mathcal{G}$. Indeed, one samples from this distribution by considering 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 $a = a_{(i-1)l+j} \in C$ 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 this context, setting $m = \lfloor n\rho \rfloor$ for $\rho = l/\gamma > 0$ fixed, we are thus interested in the large n asymptotics of the probability $P_l(n, \rho)$ that a hyper-graph G of this distribution has a non-empty core.

7.2 Smooth Markov kernel on a reduced state space

Reducing the state space to $\mathbb{Z}_+^2$. 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 $G(\tau + 1) = G(\tau)$.

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 7.2.1. *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. 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. We start at $\tau = 0$ with a uniform distribution of $G(0)$ within each possible ensemble $\mathcal{G}(\vec{z}(0), 0)$. If $N(G'|\vec{\omega}, \tau)$ 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 $\widehat{\tau} = \tau$, since 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 $G' \in \mathcal{G}(\vec{z}', \tau + 1)$ and 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. $z_1(\tau) = 0$).

Obviously, G and G' must be such that $R \subset R'$, $S \subseteq R' \cup S'$ and $T' \subseteq T$. So, let $q_0 \geq 0$ denote the size of $R' \cap S$, $p_0 \geq 0$ the size of $R' \cap T$, and $q_1 \geq 0$ the size of $S' \cap T$. We have $q_0 + p_0 \leq m - z_1' - z_2'$, $q_1 \leq z_1'$ and

$$\left\{ \begin{array}{rcl} z_0 & = & z_0' - q_0 - p_0\,, \\ z_1 & = & z_1' + q_0 - q_1\,, \\ z_2 & = & z_2' + p_0 + q_1\,, \end{array} \right. \tag{7.1}$$

where $z_0 = m - z_1 - z_2$ and $z_0' = m - z_1' - z_2'$. Let T^* denote the set of c-nodes $a \in T'$ for which $k_a > k_a'$, and denote the size of T^* by $q_2 \leq z_2'$. Observe that of the l edges of the v-node i deleted by the algorithm in the move from G to G', exactly one edge hits each of the nodes in $R' \cap S$, at least one edge hits each of the nodes in $S' \cap T$, and each of the nodes in T^*, while at least two edges hit each of the notes in $R' \cap T$. Consequently, $2p_0 + q_0 + q_1 + q_2 \leq l$. Since $z_1 > 0$ we know that $\widehat{\tau} = \tau$ and further, $(n - \tau - 1)l \geq z_1' + 2z_2'$, which in view of (7.1) is equivalent to $(n - \tau)l - (z_1 + 2z_2) \geq l - (2p_0 + q_0 + q_1) \geq q_2$.

To recap, we see that (p_0, q_0, q_1, q_2) belongs to the subset $\mathcal{D}$ of $\mathbb{Z}_+^4$ for which both the relations (7.1) 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 (and in particular $|\mathcal{D}| \leq (l+1)^4$).

To count $N(G'|\vec{z},\tau)$ we first select the v-node i to add to G' from among the $\tau+1$ elements of U', and the order (permutation) of the l sockets of i that we use when connecting it to the c-nodes for creating $G \in \mathcal{G}(\vec{z},\tau)$. Summing over the set $\mathcal{D}$ of allowed values of p_0, q_0, q_1, q_2, for each such value we have $\binom{m-z_1'-z_2'}{q_0,p_0,\cdot}$ ways to subdivide the nodes of R' among S, T and R, then $\binom{z_1'}{q_1}$ ways to select the nodes of S' that are assigned to T and $\binom{z_2'}{q_2}$ ways to select those of T' that are assigned to T^*. We further have $\mathsf{coeff}[(e^{\mathbf{x}}-1-\mathbf{x})^{p_0}(e^{\mathbf{x}}-1)^{q_1+q_2}, \mathbf{x}^{l-q_0}]$ ways to select the precise number of edges (≥ 2) from i that we are to connect to each of the p_0 nodes in $R' \cap T$, and the precise number of edges (≥ 1) from i that we are to connect to each of the q_1 nodes in $S' \cap T$ and to each of the q_2 nodes in T^*, while allocating in this manner exactly $l-q_0$ edges out of i (the remaining q_0 are then used to connect to nodes in $R' \cap S$). Noting that for each of the graphs G thus created we have exactly q_0 ways to choose the deleted node from S while still resulting with the graph G', we conclude that

$$N(G'|\vec{z},\tau) = \tag{7.2}$$
$$(\tau+1)\, l! \sum_{\mathcal{D}} \binom{m-z_1'-z_2'}{q_0,p_0,\cdot} \binom{z_1'}{q_1} \binom{z_2'}{q_2} q_0 \mathsf{coeff}[(e^{\mathbf{x}}-1-\mathbf{x})^{p_0}(e^{\mathbf{x}}-1)^{q_1+q_2}, \mathbf{x}^{l-q_0}]\,,$$

depends on G' only via $\vec{z}'$, as claimed.

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,

$$h(\vec{z},\tau) = \binom{m}{z_1, z_2, .} \binom{n}{\tau} \mathsf{coeff}[(e^{\mathbf{x}} - 1 - \mathbf{x})^{z_2}, \mathbf{x}^{(n-\tau)l - z_1}]((n-\tau)l)! \ . \tag{7.3}$$

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 (7.2) and (7.3), 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

$$\widehat{W}_\theta(\Delta\vec{z}|\vec{x}) \equiv \binom{l-1}{q_0 - 1, q_1, q_2} \mathfrak{p}_0^{q_0 - 1} \mathfrak{p}_1^{q_1} \mathfrak{p}_2^{q_2} \,, \tag{7.4}$$

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

$$\mathfrak{p}_0 = \frac{x_1}{l(1-\theta)} \,, \qquad \mathfrak{p}_1 = \frac{x_2 \lambda^2 e^{-\lambda}}{l(1-\theta)(1 - e^{-\lambda} - \lambda e^{-\lambda})} \,, \qquad \mathfrak{p}_2 = 1 - \mathfrak{p}_0 - \mathfrak{p}_1 \,, \tag{7.5}$$

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

$$\frac{\lambda(1 - e^{-\lambda})}{1 - e^{-\lambda} - \lambda e^{-\lambda})} = \frac{l(1-\theta) - x_1}{x_2} \tag{7.6}$$

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 (7.6). To justify this assumption, note that

$$\mathsf{coeff}[(e^{\mathbf{x}} - 1 - \mathbf{x})^t, \mathbf{x}^s] \lambda^s (e^\lambda - 1 - \lambda)^{-t} = \mathbb{P}(\sum_{i=1}^t N_i = s) \,,$$

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 (7.2) and (7.3), 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

$$\mathcal{Q}_+(\epsilon) \equiv \{(\vec{z},\tau) \,:\, 1 \le z_1 \,;\, n\epsilon \le z_2 \,;\, 0 \le \tau \le n(1-\epsilon) \,;\, n\epsilon \le (n-\tau)l - z_1 - 2z_2\} \,,$$

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 $\{(\vec{z},\tau) : z_1(\tau) = 0, \tau \le n\}$ (c.f. [30, Lemma 4.5] for the proof, where the restriction to $\mathcal{Q}_+(\epsilon)$ guarantees that the relevant values of t 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 $\vec{X}_i = (\mathbb{I}_{N_i=1}, \mathbb{I}_{N_i\ge 2}, N_i) \in \mathbb{Z}_+^3$ 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,

$$\vec{y}(0;\rho) = \rho(\gamma e^{-\gamma}, 1 - e^{-\gamma} - \gamma e^{-\gamma}) \,, \tag{7.7}$$

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,

$$\left\{\begin{array}{rcl} Q_{11}(0) &=& \frac{l}{\gamma}\,\gamma\, e^{-2\gamma}(e^\gamma - 1 + \gamma - \gamma^2)\,, \\ Q_{12}(0) &=& -\frac{l}{\gamma}\,\gamma\, e^{-2\gamma}(e^\gamma - 1 - \gamma^2)\,, \\ Q_{22}(0) &=& \frac{l}{\gamma}\, e^{-2\gamma}\left[(e^\gamma - 1) + \gamma(e^\gamma - 2) - \gamma^2(1+\gamma)\right]. \end{array}\right. \tag{7.8}$$

More precisely, as shown for example in [30, Lemma 4.4], for all n, r and $\rho \in [\epsilon, 1/\epsilon]$,

$$\sup_{\vec{u}\in\mathbb{R}^2}\sup_{x\in\mathbb{R}} |\mathbb{P}\{\vec{u}\cdot\vec{z} \le x\} - \mathsf{G}_2(\vec{u}\cdot\vec{z} \le x | n\vec{y}(0); n\mathbb{Q}(0))| \;\le\; \kappa(\epsilon) n^{-1/2} \,. \tag{7.9}$$

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, adapting a result of [45] (and its proof) to the context of our graph ensemble, one finds that

Lemma 7.2.2. *For $l \ge 3$ and any $\epsilon > 0$ there exist $\kappa(l,\epsilon) > 0$ and $C(l,\epsilon)$ finite such that for any $m \ge \epsilon n$ the probability that a random hyper-graph from the ensemble $\mathcal{G}_l(n,m)$ has a non-empty core of less than $m\kappa(l,\epsilon)$ v-nodes is at most $C(l,\epsilon)\, m^{1-l/2}$ (alternatively, the probability of having a non-empty core with less than $n\kappa$ v-nodes is at most $C\, n^{1-l/2}$).*

Proof. 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. Since the core is the stopping set including the maximal number of v-nodes, it suffices to bound well the number $N(s,r)$ of stopping sets in our random hyper-graph which involve exactly s v-nodes and r c-nodes. To this end, note that necessarily $r \leq \lfloor ls/2 \rfloor$ and

$$\mathbb{E}\, N(s,r) = \binom{n}{s}\binom{m}{r}\frac{1}{m^{sl}}\,\mathsf{coeff}[(e^{\mathbf{x}}-1-\mathbf{x})^r, \mathbf{x}^{sl}](sl)!$$

(multiply the number of sets of s v-nodes and r c-nodes by the probability that such a set forms a stopping set, with $\mathsf{coeff}[(e^{\mathbf{x}}-1-\mathbf{x})^r, \mathbf{x}^{sl}](sl)!$ counting the number of ways of connecting the s v-nodes to these r c-nodes so as to form a stopping set, while m^{sl} is the total number of ways of connecting the s v-nodes in our graph ensemble). It is easy to see that for any integers $r, t \geq 1$,

$$\mathsf{coeff}[(e^{\mathbf{x}}-1-\mathbf{x})^r, \mathbf{x}^t] \leq (e^{\mathbf{x}}-1-\mathbf{x})^r \mid_{\mathbf{x}=1} \leq 1\,.$$

Hence, for some $\zeta = \zeta(l,\epsilon)$ finite, any $m \geq \epsilon n$, $sl \leq m$ and $r \leq \lfloor ls/2 \rfloor$,

$$\mathbb{E}\, N(s,r) \leq \binom{n}{s}\binom{m}{r}\frac{(sl)!}{m^{sl}} \leq \frac{n^s}{s!}\,\frac{m^{\lfloor sl/2 \rfloor}}{\lfloor sl/2 \rfloor!}\,\frac{(sl)!}{m^{sl}} \leq \frac{n^s}{s!}\left(\frac{sl}{m}\right)^{\lceil sl/2 \rceil} \leq \left[\zeta\left(\frac{s}{m}\right)^{l/2-1}\right]^s .$$

Thus, fixing $0 < \kappa < 1/l$ (so $sl \leq m$ whenever $s \leq \kappa m$), for $l \geq 3$, the probability that a random hyper-graph from the ensemble $\mathcal{G}_l(n,m)$ has a stopping set of size at most $m\kappa$ is bounded above by

$$\mathbb{E}\left[\sum_{s=1}^{m\kappa}\sum_{r=1}^{\lfloor ls/2 \rfloor} N(s,r)\right] \leq \zeta m^{1-l/2}\sum_{s=1}^{\infty} sl(\zeta\kappa^{l/2-1})^{s-1} \leq 4\zeta l m^{1-l/2}\,,$$

provided $\zeta\kappa^{l/2-1} \leq 1/2$. □

7.3 The ODE method and the critical value

In view of the approximations of Section 7.2 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 (7.4) 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

$$\widehat{q}_+(\epsilon) \;\equiv\; \{(\vec{x},\theta) \,:\, 0 \leq x_1\,;\; 0 \leq x_2\,;\; \theta \in [0, 1-\epsilon]\,;\; 0 \leq (1-\theta)l - x_1 - 2x_2\}\,,$$

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 (7.4) 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 $\vec{x}, \vec{x}\,' \in \mathbb{R}^2$

$$\left|\left|\widehat{W}_{\theta'}(\,\cdot\,|\vec{x}\,') - \widehat{W}_{\theta}(\,\cdot\,|\vec{x})\right|\right|_{\mathrm{TV}} \leq L\,\left(||\vec{x}\,' - \vec{x}|| + |\theta' - \theta|\right)$$

(with $||\,\cdot\,||_{\mathrm{TV}}$ 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

$$\frac{\mathrm{d}\vec{y}}{\mathrm{d}\theta}(\theta) \;\;=\;\; \vec{F}(\vec{y}(\theta), \theta)\,, \tag{7.10}$$

starting at $\vec{y}(0)$ of (7.7), 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 (7.4). This is shown for instance in [38, 44, 26].

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 [37], and such degenerate (or zero-one) fluid-limits have been established for many other problems. For example, this was done in [36] for the largest possible matching and in [46] for the size of k-core of random graphs (c.f. [43] 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 (7.10) admits a unique solution $\vec{y}(\theta; \rho)$ subject to the initial condition (7.7) 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*

$$\rho_{\mathrm{c}} \equiv \inf\{\rho > 0 :\; h_\rho(u) > 0 \quad \forall u \in (0, 1]\}\,,$$

then $y_1(\theta; \rho)$ is strictly positive for all $\theta \in [0, 1)$, while for any $\rho \leq \rho_{\mathrm{c}}$ the solution $\vec{y}(\theta; \rho)$ first hits the line $y_1 = 0$ at some $\theta_*(\rho) < 1$.

Returning to the XOR-SAT problem, [26, 44] 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 $\rho > \rho_{\mathrm{c}}$ and fails with high probability if $\rho < \rho_{\mathrm{c}}$. See Fig. 7.1 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]), [38] show that with high probability this algorithm successfully decimates the whole hyper-graph without ever running out of degree one vertices if $\rho > \rho_{\mathrm{c}}$. Vice versa, for $\rho < \rho_{\mathrm{c}}$, 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

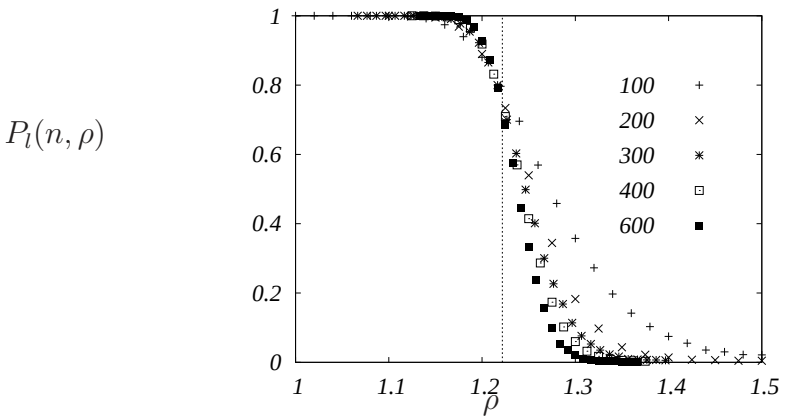

Figure 7.1: Probability that a random $l = 3$-hyper-graph with m vertices and $n = m/\rho$ hyper-edges has a non-empty 2-core estimated numerically for $m = 100, \ldots, 600$. The vertical line corresponds to the asymptotic threshold $\rho_c \approx 1.2218$.

into noise level in this communication application, so [38] in essence explicitly characterize 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 $\rho = \rho_c$ where it abruptly changes from an empty core for $\rho > \rho_c$ to a core whose size is a positive fraction of n for $\rho < \rho_c$, has other important implications. For example, as shown in [26, 44] the structure of the set of solutions of the linear system changes dramatically at ρ_c, exhibiting a 'clustering effect' when $\rho < \rho_c$. 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 $O(n)$. 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.

Such 'solution clustering' phenomenon has been conjectured for a variety of random combinatorial decision problems, on the basis of non-rigorous statistical mechanics calculations. The most studied among these problems is the random K-satisfiability, for which some indication of clustering is rigorously proved in [42, 23]. Several authors suggest that the solution clustering phenomenon is related to the poor performance of search algorithms on properly chosen ensembles of random instances. Still within random K-satisfiability, the performance of certain standard solution heuristics (such as the 'pure-literal' rule), is also related to the

appearance of properly defined cores (see [43]).

We conclude this subsection with a 'cavity type' direct prediction of the value of $\rho_{\rm c}$ 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 $\rho\le\rho_{\rm c}$.

7.4 Diffusion approximation and scaling window size

As mentioned before, the ODE asymptotics as in [38] is of limited value for decoding with code block-length that are relevant in practice. For this reason, [22] 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 $\rho_n\to\rho_{\rm c}$).

Indeed, fixing $\rho>0$ 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:

$$\frac{\mathrm{d}\mathbb{Q}(\theta)}{\mathrm{d}\theta}=\mathbb{G}(\vec{y}(\theta),\theta)+\mathbb{A}(\vec{y}(\theta),\theta)\mathbb{Q}(\theta)+\mathbb{Q}(\theta)\mathbb{A}(\vec{y}(\theta),\theta)^\dagger \tag{7.11}$$

(c.f. [22]). 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 (7.10) 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 (7.4). That is, the non-negative definite symmetric matrix with entries

$$\left\{\begin{array}{rcl} G_{11}(\vec{x},\theta) &=& (l-1)[\mathfrak{p}_0+\mathfrak{p}_1-(\mathfrak{p}_0-\mathfrak{p}_1)^2]\,,\\ G_{12}(\vec{x},\theta) &=& -(l-1)[\mathfrak{p}_0\mathfrak{p}_1+\mathfrak{p}_1(1-\mathfrak{p}_1)]\,,\\ G_{22}(\vec{x},\theta) &=& (l-1)\mathfrak{p}_1(1-\mathfrak{p}_1)\end{array}\right. \tag{7.12}$$

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 (7.8) for the ODE (7.11) as well as the terms $\vec{y}(\theta) = \vec{y}(\theta;\rho)$ that appear in its right side.

Focusing hereafter on the critical case $\rho = \rho_c$, there exists then a unique *critical time* $\theta_c \equiv \theta_*(\rho_c)$ in $(0,1)$ with $y_1(\theta_c) = y_1'(\theta_c) = 0$ and $y_1''(\theta_c) > 0$, while the smooth solution $\theta \mapsto y_1(\theta;\rho_c)$ is positive when $\theta \neq \theta_c$ and $\theta \neq 1$ (for more on $\vec{y}(\cdot;\cdot)$ see [30, Proposition 4.2]).

For $\rho_n = \rho_c + rn^{-1/2}$ the leading contribution to $P_l(n,\rho_n)$ is the probability $\widehat{\mathbb{P}}_{n,\rho_n}(z_1(n\theta_c) \leq 0)$ 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 (7.4) 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 $y_1(\theta_c;\rho_c) = 0$, hence

$$y_1(\theta_c;\rho_n) = rn^{-1/2}[\frac{\partial y_1}{\partial \rho}(\theta_c;\rho_c) + o(1)]\,.$$

Thus, setting $\alpha_l \equiv \sqrt{Q_{11}}/\frac{\partial y_1}{\partial \rho}$, both evaluated at $\theta = \theta_c$ and $\rho = \rho_c$, by the preceding Gaussian approximation

$$P_l(n,\rho_n) = \widehat{\mathbb{P}}_{n,\rho_n}(z_1(n\theta_c) \leq 0) + o(1) = \mathsf{G}_1(-r/\alpha_l) + o(1)\,, \tag{7.13}$$

as shown in [22]. In particular, the phase transition scaling window around $\rho = \rho_c$ is of size $\Theta(n^{-1/2})$.

In a related work, [29] determine the asymptotic core size for a random hypergraph from an ensemble which is the 'dual' of $\mathcal{G}_l(n,m)$. In their model the hyperedges (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 [28, 32, 33].

7.5 Finite size scaling: corrections to ρ_c

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 [25, 48]. 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

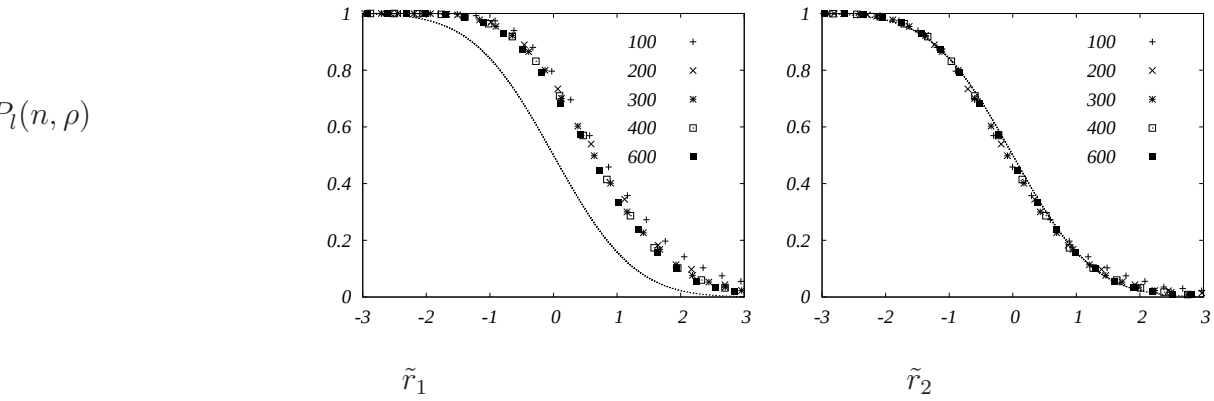

Figure 7.2: The numerical estimates for the core probabilities in Fig. 7.1, plotted versus scaling variables $\tilde{r}_1$, $\tilde{r}_2$. On the left: $\tilde{r}_1 = \sqrt{n}(\rho - \rho_c)/\alpha_l$. On the right: $\tilde{r}_2 = \sqrt{n}(\rho - \rho_c - \delta_l n^{-2/3})/\alpha_l$ where $\delta_l = \alpha_l \beta_l \Omega$. According to Theorem 7.5.3, corrections to the asymptotic curve $\mathsf{G}_1(-\tilde{r})$ (dashed) are $\Theta(n^{-1/6})$ on the left, and $O(n^{-5/26+\epsilon})$ on the right.

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 $\Theta(n^{-1/3})$ [24]. 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.

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 [35] and references therein), for $\mathcal{G}_l(n,m)$ and $\rho_n = \rho_c + rn^{-1/2}$ inside the scaling window, it is conjectured in [22] and proved in [30] that the leading correction to the diffusion approximation for $P_l(n,\rho_n)$ is of order $\Theta(n^{-1/6})$ (and admits a sharp characterization in terms of the distribution of a Brownian motion with quadratic shift, from which it inherits the scaling with n). Comparing this finite size scaling expression with numerical simulations, as illustrated in Figure 7.2, 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. [31]). 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-side of (7.13) is provided in [30, Section 5].

Proposition 7.5.1. *Let $\beta \in (3/4, 1)$, $J_n = [n\theta_c - n^\beta, n\theta_c + n^\beta]$ and $|\rho - \rho_c| \le n^{\beta'-1}$ with $\beta' < 2\beta - 1$. Then, for $\varepsilon_n = A \log n$ and $\delta_n = D\, n^{-1/2} (\log n)^2$,*

$$\widehat{\mathbb{P}}_{n,\rho}\Big\{ \inf_{\tau \in J_n} z_1(\tau) \le -\varepsilon_n \Big\} - \delta_n \le P_l(n,\rho) \le \widehat{\mathbb{P}}_{n,\rho}\Big\{ \inf_{\tau \in J_n} z_1(\tau) \le \varepsilon_n \Big\} + \delta_n \,. \quad (7.14)$$

At the critical point (i.e. for $\rho = \rho_c$ and $\theta = \theta_c$) the solution of the ODE (7.10) 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 θ_c we have $y_1(\theta) \simeq \frac{1}{2}\widetilde{F}(\theta - \theta_c)^2$, for the positive constant

$$\widetilde{F} \equiv \frac{\mathrm{d}^2 y_1}{\mathrm{d}\theta^2}(\theta_c; \rho_c) = \frac{\mathrm{d}F_1}{\mathrm{d}\theta}(\vec{y}(\theta_c;\rho_c), \theta_c) = \frac{\partial F_1}{\partial \theta} + \frac{\partial F_1}{\partial y_2} F_2 \quad (7.15)$$

(omitting hereafter arguments that refer to the critical point). In the same neighborhood, the contribution of fluctuations to $z_1(n\theta) - z_1(n\theta_c)$ is approximately $\sqrt{\widetilde{G} n |\theta - \theta_c|}$, with $\widetilde{G} = G_{11}(\vec{y}(\theta_c;\rho_c), \theta_c) > 0$. Comparing these two contributions we see that the relevant scaling is $X_n(t) = n^{-1/3}[z_1(n\theta_c + n^{2/3}t) - z_1(n\theta_c)]$, which as shown in [30, 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 7.5.2. *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 $\theta = \theta_c$ and $\rho = \rho_c$), which is independent of $W(t)$.*

For some $\beta \in (3/4, 1)$, any $\eta < 5/26$, all $A > 0$, $r \in \mathbb{R}$ and n large enough, if $\rho_n = \rho_c + r\, n^{-1/2}$ and $\varepsilon_n = A \log n$, then

$$\Big| \widehat{\mathbb{P}}_{n,\rho_n}\big\{ \inf_{\tau \in J_n} z_1(\tau) \le \pm\varepsilon_n \big\} - \mathbb{P}\big\{ n^{1/6}\xi + \inf_t X(t) \le 0 \big\} \Big| \le n^{-\eta} \,. \quad (7.16)$$

We note in passing that within the scope of weak convergence [20] pioneered the use of Brownian motion with quadratic drift (ala $X(t)$ of Proposition 7.5.2), to examine the near-critical behavior of the giant component in Erdös-Rényi random graphs, and his method was extended in [33] to the giant set of identifiable vertices in Poisson random hyper-graph models.

Combining Propositions 7.5.1 and 7.5.2 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 [34], yielding the following conclusion.

Theorem 7.5.3. *For $l \geq 3$ set $\alpha_l = \sqrt{Q_{11}}/\frac{\partial y_1}{\partial \rho}$, $\beta_l = \frac{1}{\sqrt{Q_{11}}}\widetilde{G}^{2/3}\,\widetilde{F}^{-1/3}$ and $\rho_n = \rho_c + r\,n^{-1/2}$. Then, for any $\eta < 5/26$*

$$P_l(n,\rho_n) = \mathsf{G}_1(-r/\alpha_l) + \beta_l \Omega\ \mathsf{G}_1'(-r/\alpha_l)\ n^{-1/6} + O(n^{-\eta})\,, \tag{7.17}$$

for $\Omega \equiv \int_0^\infty [1-\mathcal{K}(z)^2]\,\mathrm{d}z$ and an explicit function $\mathcal{K}(\cdot)$ (see [30, equation (2.17)]).

Remark 7.5.4. *The simulations in Figure 7.2 suggest that the approximation of $P_l(n,\rho_n)$ we provide in (7.17) 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 $\Phi(-r/\alpha_l + Cn^{-1/6})$, see (7.18). We believe this is indeed the order of the next term in the expansion (7.17). Determining its form is an open problem.*

Proof. Putting together Propositions 7.5.1 and 7.5.2, we get that

$$P_l(n,\rho_n) = \mathbb{P}\left\{n^{1/6}\xi + \inf_t X(t) \leq 0\right\} + O(n^{-\eta})\,.$$

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

$$\begin{aligned} P_l(n,\rho_n) &= \mathbb{P}\left\{n^{1/6}\left(\frac{\partial y_1}{\partial \rho}\right) r + n^{1/6}\sqrt{Q_{11}}Y + \widetilde{F}^{-1/3}\widetilde{G}^{2/3}Z \leq 0\right\} + O(n^{-\eta}) \\ &= \mathbb{E}\Big\{\mathsf{G}_1\big(-\frac{r}{\alpha_l} - \beta_l n^{-1/6} Z\big)\Big\} + O(n^{-\eta})\,. \end{aligned} \tag{7.18}$$

The proof of the theorem is thus completed by a first order Taylor expansion of $\mathsf{G}_1(\,\cdot\,)$ around $-r/\alpha_l$, as soon as we show that $\mathbb{E}\,Z = -\Omega$, and $\mathbb{E}|Z|^2$ is finite. To this end, from [34, Theorem 3.1], we easily deduce that Z has the continuous distribution function $F_Z(z) = 1 - \mathcal{K}(-z)^2$, for $z < 0$, while $F_Z(z) = 1$ for $z \geq 0$, resulting after integration by parts with the explicit formula (2.16) of [30] for Ω. We note in passing that taking $c = 1/2$ and $s = 0$ in [34, (5.2)] provides the explicit expression of [30, formula (2.17)] for $\mathcal{K}(x)$, en-route to which [34] also proves the the finiteness of the relevant integral. Further, [34, Corollary 3.4] shows that the probability that the minimum of $\widetilde{X}(t)$ is achieved as some $t \notin [-T,T]$ is at most $A_0^{-1}\exp(-A_0T^3)$ for a positive constant A_0. With $\widetilde{X}(t) \geq \widetilde{W}(t)$ we therefore have that

$$F_Z(z) \equiv \mathbb{P}\{Z \leq z\} \leq \mathbb{P}\left\{\inf_{t\in[-T,T]} \widetilde{X}(t) \leq z\right\} + A_0^{-1}\,e^{-A_0T^3} \leq e^{-z^2/2T} + A_0^{-1}\,e^{-A_0T^3}\,,$$

Taking $T = \sqrt{z}$ we deduce that if $z < 0$, then $F_Z(z) < C^{-1}\exp(-C|z|^{3/2})$ for some $C > 0$, which yields the stated finiteness of each moment of Z (and in particular, of $\mathbb{E}|Z|^2$ and Ω). $\square$

Consider the (time) evolution of the core for the hyper-graph process in which one hyper-edge is added uniformly at random at each time step. In other words, 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 $n_{\rm c} \equiv \min\{n : S(n) \geq 1\}$ the onset of a non-empty core. From Lemma 7.2.2 we have that for any $\rho > 0$ there exist $\kappa > 0$ and $C < \infty$ such that $\{S(n) : 0 \leq n \leq m/\rho\}$ intersects $[1, m\kappa]$ with probability at most $Cm^{1-l/2}$. Further, fixing $\rho < \rho_{\rm c}$, the probability of an empty core, i.e. $S(m/\rho) = 0$, decays (exponentially) in m. We thus deduce that for large m most of the trajectories $\{S(n)\}$ jump from having no core to a linear (at least $m\kappa$) core size at the well defined (random) critical edge number $n_{\rm c}$. By the monotonicity of $S(n)$ we also know that $\mathbb{P}_m\{n_{\rm c} \leq m/\rho\} = P_l(\rho, m/\rho)$. Therefore, Theorem 7.5.3 allows us to determine the asymptotic distribution of $n_{\rm c}$. Indeed, expressing n in terms of m in Eq. (7.17) we get that for each fixed $x \in \mathbb{R}$,

$$\mathbb{P}\left\{n_{\rm c} \leq m\rho_{\rm c}^{-1} + m^{1/2}\rho_{\rm c}^{-3/2}\alpha_l\, x\right\} = \Phi(x) + \beta_l\Omega\rho_{\rm c}^{1/6}\Phi'(x)\, m^{-1/6} + O(m^{-\eta})\,,$$

whence we read off that $\widehat{n}_{\rm c} \equiv (n_{\rm c} - m/\rho_{\rm c})/(\sqrt{m}\rho_{\rm c}^{-3/2}\alpha_l) + \beta_l\Omega\rho_{\rm c}^{1/6}m^{-1/6}$ converge in distribution to the standard normal law (and the corresponding distribution functions converge point-wise faster than $m^{-\eta}$ for any $\eta < 5/26$).

Remark 7.5.5. *Our techniques are applicable to many other properties of the core in the 'scaling regime' $\rho_n = \rho_{\rm c} + r\, n^{-1/2}$. For example, the distribution of the number of hyper-edges S in the core can be derived from the approximation of the trajectory of the decimation algorithm. Namely, as shown in [30, Section 6] for such ρ_n, near the critical time $z_1(t) \simeq \sqrt{n}\xi(r) + X_n(t)$ for $\xi(r)$ and $X_n(t) \equiv n^{1/3}X(n^{-2/3}(t - n\theta_{\rm c}))$ as in Proposition 7.5.2. With $\mathbb{E}X_n(t) = \frac{\widetilde{F}}{2n}(t - n\theta_{\rm c})^2$, upon noting that $n - S = \min\{t : z_1(t) = 0\}$, we obtain that, conditional to the existence of a non-empty core, $(S - n(1 - \theta_{\rm c}))/n^{3/4}$ converges in distribution to $(4Q_{11}/\widetilde{F}^2)^{1/4}\, Z$ with Z a non-degenerate random variable. Indeed, at the relevant time window $n\theta_{\rm c} \pm O(n^{3/4})$ the contribution of $X_n(\cdot) - \mathbb{E}X_n(\cdot)$ to the fluctuations of S is negligible in comparison with that of $\sqrt{n}\xi(r)$. So, more precisely, based on the explicit distribution of $\xi(r)$ we have that $Z \stackrel{\rm d}{=} \sqrt{U - rb}$ for $b \equiv Q_{11}^{-1/2}\frac{\partial y_1}{\partial \rho}$ and U a standard normal random variable conditioned to $U \geq rb$. In formulae, Z is supported on $\mathbb{R}_+$ and admits there the probability density*

$$p_Z(z) = \frac{2z\, e^{-\frac{1}{2}(rb + z^2)^2}}{\sqrt{2\pi}\,[1 - \Phi(rb)]}\,.$$

Naively one expects the core size to have $\Theta(n^{1/2})$ fluctuations. This is indeed the asymptotic behavior for a fixed $\rho < \rho_{\rm c}$, but as usual in phase transitions, fluctuations are enhanced near the critical point.

Appendix A

A combinatorial calculation

The objective of this lecture is to spell out a single calculation from the course in all of its painful details. The choice of which calculation was made on the basis of its level of complication (moderate) rather than on it intrinsic interest.

We shall deal with the ferromagnetic Ising model on random regular graphs, which we treated a couple of weeks ago. Recall that this is a model over variables $x_i \in \{+1, -1\}$, $i \in \{1, \dots, N\}$ with distribution

$$\mu(x) = \frac{1}{Z_G} \exp\left\{ \beta \sum_{(i,j)\in E} x_i x_j \right\} . \tag{A.1}$$

Here E is the edge set of a random graph $G = (V = [N], E)$ that we shall take to be a random graph with degree k. More precisely G is generated as follows (*configuration model*). Associate to each vertex $i \in V$ k 'half edges,' sample a uniformly random pairing over kN objects, and pair all the half-edges accordingly. Two distinct pairings are considered as distinct graphs.

We will be interested in the *restricted partition function* defined by constraining the usual sum to configurations of vanishing magnetization:

$$Z_G^* \equiv \sum_{x:\, \sum_i x_i = 0} \exp\left\{ \beta \sum_{(i,j)\in E} x_i x_j \right\} . \tag{A.2}$$

A few lectures ago, we made the following claim, that we will now prove.

Lemma A.0.6. *Assume N to be even. Then the expectation of the restricted partition function Z_G^* is, to the leading exponential order,*

$$\mathbb{E}\{Z_G^*\} \doteq 2^N (\cosh\beta)^{kN/2} . \tag{A.3}$$

Proof. Throughout the proof $M = Nk/2$ will denote the number of edges in G, and $\mathbb{G}_{N,k}$ the graph ensemble. Finally, for a set $\mathcal{S}$, we shall denote by $|\mathcal{S}|$ its cardinality.

Let $\Delta_G(x)$ denote the number of edges $(i,j) \in E$ such that $x_i \neq x_j$, and $Z_G^*(\Delta)$ be the number of configurations x such that $\Delta_G(x) = \Delta$. Then we clearly have $\sum_{(i,j)\in E} x_i x_j = M - 2\Delta_G(x)$. As a consequence

$$Z_G^* = e^{\beta M} \sum_{\Delta=0}^{M} Z_G^*(\Delta)\, e^{-2\beta\Delta}\,. \tag{A.4}$$

By linearity of expectation, and since the graph distribution is invariant under vertices permutations, we have

$$\begin{aligned} \mathbb{E}\, Z_G^*(\Delta) &= \sum_{\sum_i x_i = 0} \mathbb{P}\{\Delta_G(x) = \Delta\} = \binom{N}{N/2} \mathbb{P}\{\Delta_G(x_*) = \Delta\} = & \text{(A.5)} \\ &= \binom{N}{N/2} \frac{|\{G \in \mathbb{G}_{N,k} \text{ st } \Delta_G(x_*) = \Delta\}|}{|\mathbb{G}_{N,k}|}\,. & \text{(A.6)} \end{aligned}$$

Here x_* denote the configuration consisting in $N/2$ $+1$'s followed by $N/2$ -1's.

The number of graph in the ensembles is just the number of pairings of Nk objects

$$|\mathbb{G}_{N,k}| = \mathfrak{P}(Nk) \equiv \frac{(Nk)!}{(Nk/2)! 2^{Nk/2}}\,. \tag{A.7}$$

On the other hand it is not too hard to compute the number of such pairings for which the number edges with unequal end-points is Δ:

$$|\{G \in \mathbb{G}_{N,k} \text{ st } \Delta_G(x_*) = \Delta\}| = \binom{M}{\Delta}^2 \Delta!\, \mathfrak{P}(M-\Delta)^2 \tag{A.8}$$

Putting everything together we get

$$\begin{aligned} \mathbb{E}\{Z_G^*\} &= e^{\beta M} \binom{N}{N/2} \frac{1}{\mathfrak{P}(2M)} \sum_{\Delta=0}^{M} \binom{M}{\Delta}^2 \Delta!\, \mathfrak{P}(M-\Delta)^2 e^{-2\beta\Delta} & \text{(A.9)} \\ &\doteq e^{\beta M} \binom{N}{N/2} \frac{1}{\mathfrak{P}(2M)} \sup_{\Delta\in\{0,\dots,M\}} \binom{M}{\Delta}^2 \Delta!\, \mathfrak{P}(M-\Delta)^2 e^{-2\beta\Delta} & \text{(A.10)} \end{aligned}$$

Let us recall the exponential behaviors (for $q \in [0,1]$)

$$\binom{N}{Nq} \doteq e^{NH(q)}\,, \qquad \mathfrak{P}(N) \doteq \left(\frac{N}{e}\right)^{N/2}\,, \tag{A.11}$$

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

Substituting in the expression for $\mathbb{E}\{Z_G^*\}$ we get

$$\begin{aligned}\mathbb{E}\{Z_G^*\} &\doteq e^{\beta M} 2^N \left(\frac{2M}{e}\right)^{-M} \sup_{\delta\in[0,1]} \left\{ e^{2MH(\delta)} \left(\frac{M\delta}{e}\right)^{M\delta} \left(\frac{M(1-\delta)}{e}\right)^{M(1-\delta)} e^{-2M\beta\delta} \right\} \\ &= 2^N\, e^{\beta M}\, 2^{-M} \exp\left\{ M \sup_{\delta\in[0,1]} [H(\delta) - 2\beta\delta] \right\}\end{aligned}$$

The sup is achieved when $H'(\delta) = 2\beta$, which implies $\delta = \delta_*(\beta) \equiv (1+e^{2\beta})^{-1}$. At this point we have $H(\delta_*) - 2\beta\delta_* = \log(1+e^{-2\beta})$, which yields

$$\mathbb{E}\{Z_G^*\} \doteq 2^N\, e^{\beta M}\, 2^{-M} \exp\left\{M \log(1+e^{-2\beta})\right\} = 2^N (\cosh\beta)^M\,. \quad \text{(A.12)}$$

□

Bibliography

[1] T. Cover and J. A. Thomas, "Elements of Information Theory," Wiley Interscience, New York, 1991

[2] Mézard, M. and Montanari, A., "Information, Physics and Computation", Oxford University Press, Oxford, 2008.

[3] R. S. Ellis and C. M. Newman, *The statistics of Curie-Weiss models*, J. Stat. Phys. 19 (1978) 149-161

[4] M. Leone, A. Vázquez, A. Vespignani and R. Zecchina, *Ferromagnetic ordering in graphs with arbitrary degree distribution*, Eur. Phys. J. B 28 (2002) 191-197

[5] S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, *Ising model on networks with an arbitrary distribution of connections* Phys. Rev. E 66 (2002) 016104

[6] M. Mézard, G. Parisi and M. A. Virasoro, "Spin Glass Theory and Beyond," World Scientific, 1987

[7] S. Janson, T. Luczak and A. Ruciński, "Random graphs," John Wiley and sons, 2000

[8] Mézard, M., Parisi, G. and Zecchina, R. *Analytic and Algorithmic Solution of Random Satisfiability Problems,* Science **297** (2002), 812-815.

[9] Weigt, M. and Hartmann, A. K., *Minimal vertex covers on finite-connectivity random graphs: A hard-sphere lattice-gas picture* Phys. Rev. E 63, 056127 (2001)

[10] Mézard, M. and Montanari, A., *Reconstruction on Trees and Spin Glass Transition,* J. Stat. Phys. **124** (2006), 1317-1350.

[11] Mossel, E. and Peres, Y., *Information Flow on Trees*, Ann. Appl. Probab. **13** (2003), 817-844.

[12] 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

[13] Martin, J. B., *Reconstruction thresholds on regular trees*, Discr. Math. Theor. Comp. Sci. AC (2003), 191-204

[14] Kelly, F. P., *Stochastic models of computer communication systems*, J. R. Stat. Soc. Ser. B **47** (1985), 379-395

[15] Tatikonda, S. and Jordan, M. I., *Loopy belief propagation and Gibbs measures*, Uncertainty in Artificial Intelligence, 2002.

[16] Georgii, H.-O., "Gibbs Measures and Phase Transition," de Gruyter, Berlin, 1988.

[17] Gerschenfeld, A., and Montanari, A., *Reconstruction for models on random graphs*, 48nd Annual Symposium on Foundations of Computer Science, Providence, RI, 2007.

[18] 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

[19] Dembo, A. and Montanari, A. *Ising models on locally tree-like graphs*, preprint (2008).

[20] Aldous, D., *Brownian excursions, critical random graphs and the multiplicative coalescent*, Ann. Prob. **25** (1997), 812–854.

[21] 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.

[22] 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`

[23] 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.

[24] Bollobas, B. and 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.

[25] 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.

[26] 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.

[27] 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.

[28] Darling, R.W.R., Levin, D., and Norris, J. *Continuous and discontinuous phase transitions in hypergraph processes*, Random Structures Algorithms, **24** (2004), 397–419.

[29] Darling, R.W.R., and Norris, J. *Structure of large random hypergraphs*, Ann. Appl. Probab. **15** (2005), 125–152.

[30] Dembo, A., and Montanari, A. *Finite size scaling for the core of large random hypergraphs*, Ann. Appl. Probab. **18** (2008), to appear.

[31] 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.

[32] Goldschmidt, C., and Norris, J., *Essential edges in Poisson random hypergraphs*, Random Structures and Algorithms **24**, (2004), 381–396.

[33] Goldschmidt, C., *Critical random hypergraphs: the emergence of a giant set of identifiable vertices*, Ann. Prob. **33** (2005), 1573–1600.

[34] Groeneboom, P., *Brownian motion with a parabolic drift and Airy functions*, Probab. Th. Rel. Fields, **81** (1989), 79–109.

[35] Janson, S., Knuth, D. E., Luczak, T., and Pittel, B., *The birth of the giant component*, Rand. Struct. and Alg., **4** (1993), 231–358.

[36] Karp, R.M., and Sipser, M., *Maximum matching in sparse random graphs*, Proc. 22nd Annual IEEE Symposium on Foundations of Computing (1981), 364–375.

[37] Kurtz, T.G., *Solutions of ordinary differential equations as limits of pure Markov jump processes*, J. Appl. Probab. **7** (1970), 49–58.

[38] Luby, M., Mitzenmacher, M., Shokrollahi, A., and Spielman, D. *Efficient erasure correcting codes*, IEEE Trans. Inform. Theory, **47** (2001), 569–584.

[39] C. Daskalakis, E. Mossel, and S. Roch, 'Optimal Phylogenetic Reconstruction,' Proc. of ACM STOC 2006.

[40] S. Bhamidi, R. Rajagopal and S. Roch 'Network Delay Inference from Additive Metrics,' preprint, 2007

[41] T. M. Liggett "Interacting particle systems", Springer Verlag, New York, 1985.

[42] 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`

[43] Molloy, M., *Cores in random hypergraphs and Boolean formulas*, Rand. Struct. and Alg., **27** (2005), 124–135.

[44] 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.

[45] Orlitsky, A., Viswanathan, K., and Zhang, J., *Stopping set distribution of LDPC code ensembles*, IEEE Trans. Inform. Theory, **51** (2005), 929–953.

[46] 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.

[47] J. Ginibre, *General formulation of Griffiths' inequalities*, Comm. Math. Phys., 16 (1970), 310-328

[48] Wilson, D. B., *On the critical exponents of random k-SAT.* Rand. Struct. and Alg. **21** (2002), 182–195.

[49] D. Aldous, J. M. Steele, 'The Objective Method: Probabilistic Combinatorial Optimization and Local Weak Convergence,' in *Probability on discrete structures*, H. Kesten (ed.), New York, 2004

[50] F. Guerra and F. L. Toninelli, 'The high temperature region of the Viana-Bray diluted spin glass model,' *J. Stat. Phys.* **115** (2004), 531-555.

[51] D. Achlioptas and A. Naor, 'The two possible values of the chromatic number of a random graph,' Proc. of ACM STOC 2004

[52] A. Guionnet and B. Zegarlinski, 'Lectures on Logarithmic Sobolev Inequalities,' Séminaire de Probabilites de Strasbourg, **36** (2002) 1–134

[53] M. Dyer, A. Sinclair, E. Vigoda and D. Weitz, 'Mixing in time and space for lattice spin systems: A combinatorial view,' Rand. Struc. and Alg. 24, 461-479 (2004)

[54] N. Berger, C. Kenyon, E. Mossel and Y. Peres, 'Glauber dynamics on trees and hyperbolic graphs,' *Prob. Theory Rel. Fields* **131**, 311 (2005)

[55] F. Martinelli, A. Sinclair, and D. Weitz, 'The Ising model on trees: boundary conditions and mixing time,' Proc. of IEEE FOCS 2003

[56] A. Montanari and G. Semerjian, 'Rigorous Inequalities Between Length and Time Scales in Glassy Systems,' *J. Stat. Phys.* 125 (2006), 23-54

[57] M. Mézard, T. Mora and R. Zecchina, 'Clustering of Solutions in the Random Satisfiability Problem,' *Phys. Rev. Lett.* **94** (2005) 197205.

[58] D. Achlioptas and F. Ricci-Tersenghi, 'On the solution-space geometry of random constraint satisfaction problems,' Proc. of ACM STOC 2006

[59] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian and L. Zdeborova, 'Gibbs States and the Set of Solutions of Random Constraint Satisfaction Problems,' Proc. Natl. Acad. Sci. 1004, 10318 (2007)

[60] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, 'On the purity of limiting Gibbs state for the Ising model on the Bethe lattice,' *J. Stat. Phys.*, **79** (1995), 473–482.

[61] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman. 'Broadcasting on trees and the Ising model,' *Ann. Appl. Probab.*, **10** (2000) 410–433.

[62] C. Borgs, J. Chayes, E. Mossel and S. Roch, 'The Kesten-Stigum Reconstruction Bound Is Tight for Roughly Symmetric Binary Channels,' Proc. of IEEE FOCS 2006

[63] E. Mossel, D. Weitz, and N. Wormald, 'On the hardness of sampling independent sets beyond the tree threshold,' `arXiv:math/0701471`

[64] B. Bóllobas, 'A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs,' Eur. J. Combin. **1** (1980), 311.

[65] D. Achlioptas, A. Naor, and Y. Peres, 'Rigorous location of phase transitions in hard optimization problems,' Nature **435** (2005), 759

[66] L. De Sanctis, F. Guerra, *Mean field dilute ferromagnet I. High temperature and zero temperature behavior*, `arXiv:0801.4940`

[67] B. Simon, *Correlation inequalities and the decay of correlations in ferromagnets*, Commun. Math Phys. **77** (1980) 111-126.