SPIN SYSTEMS ON RANDOM GRAPHS AND RANDOM CONSTRAINT SATISFACTION PROBLEMS

For the second half of the course we will consider models of spin systems on random graphs and their fascinating random of thresholds and phase transitions. Examples of this class of models include colourings of random graphs and the random K-SAT model. Throughout we will focus on the example of the maximal independent set of a random regular graph. Ideas from statistical physics describe a series of phase transitions these models go through as the density of constraints increases.

1. Spin systems

A *constraint satisfaction problem* or CSP consists of a set of variables subject to a collection of constraints. A simple example of this is a proper q-colouring of a graph. The variables are the colours of the vertices and the edges are the constraints, that neighbouring vertices must have different colours. A *random constraint satisfaction problem* is a CSP where the collection of constraints are chosen randomly from some distribution. The analogue would be colourings of a random graph. We will be interested in the question of how many constraints a random CSP can have while still being satisfiable (the *satisfiability threshold*), how many satisfying assignments of the variables there are, the geometry of how the solutions are arranged and the algorithmic question of finding solutions efficiently.

Another variant of CSPs is to find the satisfying assignment that maximizes some objective function. A natural example of this is the maximal independent set question. An *independent set* of a graph $G = (V, E)$ is a subset $I \subset V$ such that now two vertices of I share an edge. We will build up towards an understanding of this problem for random d-regular graphs when d is large.

In this theory it is often best to consider a random solution of the CSP which leads us to the study of *spin systems*. These models, also called *graphical models* or *Markov random fields* in some areas, are a broad class of stochastic processes on networks giving a probability distribution on $\mathcal{X}^V$ for some (usually) discrete set $\mathcal{X}$, satisfying a local Markov property. Let us take the example of a random uniformly chosen independent set which consider as an element $\sigma \in \{0,1\}^V$ where σ_u is the indicator that u is in the independent set. We can write this as

$$\mathbb{P}[\sigma] = \frac{1}{Z} \prod_{u \sim v} I(\sigma_u \sigma_v = 0)$$

where the normalizing constant Z, called the partition function, is in the case simply the number of independent sets. Of course we will be interested in larger independent sets so it is useful to give larger weight to larger independent sets. The *Hardcore Model* is a distribution over independent sets given by

$$\mathbb{P}[\sigma] = \frac{1}{Z_\lambda} \lambda^{\sum_u \sigma_u} \prod_{u \sim v} I(\sigma_u \sigma_v = 0)$$

This is a special case of a more general collection of models called *spin systems*. These are distributions over $\mathcal{X}^V$ of the form

$$\mu(\sigma) = \frac{1}{Z} \prod_{u \in V} \psi_u(\sigma_u) \prod_{(u,v) \in E} \psi_{u,v}(\sigma_u, \sigma_v) \tag{1.1}$$

where $\{\psi_u\}_{u \in V}$ and $\{\psi_{u,v}\}_{(u,v) \in E}$ are non-negative functions on $\mathcal{X}$ and $\mathcal{X}^2$ respectively. In most cases we will consider homogeneous spin systems where the ψ_u (resp. $\psi_{u,v}$) do not depend on the vertex u (resp. edge (u,v)). In each case the partition function Z is the normalizing constant to make the distribution a probability measure. The value σ_u at the vertex u is called the *spin* at u. We give some examples.

Example 1. *The* Ising model *is a distribution over* $\{-1,1\}^V$ *given by*

$$\mu(\sigma) = \frac{1}{Z} \exp\left(\beta \sum_{u \sim v} \sigma_u \sigma_v + h \sum_u \sigma_u\right).$$

The parameter β *is called the* inverse temperature *and* h *is called the* external field. *These terms come from statistical mechanics where the Ising model originated as a model of magnetic systems such as a piece of iron. It is called* ferromagnetic *when* $\beta > 0$ *and* anti-ferromagnetic *when* $\beta < 0$. *In the ferromagnetic case spin configurations with neighbouring spins agreeing tend to get more weight and the strength of these interactions are governed by the magnitude of* β. *As* β *is the inverse temperature,* high temperature *refers to small values of* β *and weaker interaction while* low temperatures *refers to large* β *and strong interactions.*

Example 2. *A generalization of the Ising model to more spins is the* q*-state* Potts model *which takes values in* $[q]^V$ *(where* $[k]$ *denotes* $\{1,\ldots,q\}$*) given by*

$$\mu(\sigma) = \frac{1}{Z} \exp\left(\beta \sum_{u \sim v} I(\sigma_u = \sigma_v)\right).$$

When $q = 2$ this is the Ising distribution but with a factor of 2 difference in β. A special case of the anti-ferromagnetic Potts model with $\beta = -\infty$ (referred to as zero temperature*) corresponds to a uniform q-colouring of the graph*

$$\mu(\sigma) = \frac{1}{Z} \prod_{u \sim v} I(\sigma_u \neq \sigma_v).$$

In this case Z is the number of proper q-colourings.

1.1. **Markov Random Field Property.** Spin systems share a spatial Markov property. If $A \subset V$ we write the exterior boundary of A as

$$\partial A := \{v \in V \setminus A : \exists u \in A, (u, v) \in E\}.$$

The following lemma says that to understand the conditional distribution of σ on A given the rest of the conriguration it is enough to know σ on ∂A.

Lemma 3. *For a Markov random Field σ and $A \subset V$ and any $x \in \mathcal{X}^V$,*

$$\mathbb{P}[\sigma_A = x_A \mid \sigma_{A^c} = x_{A^c}] = \mathbb{P}[\sigma_A = x_A \mid \sigma_{\partial A} = x_{\partial A}].$$

Proof. Let x'_A be another configuration on A and let

$$y_u = \begin{cases} x'_u & u \in A \\ x_u & u \in A^c. \end{cases}$$

Then

$$\begin{aligned} \frac{\mathbb{P}[\sigma_A = x'_A \mid \sigma_{A^c} = x_{A^c}]}{\mathbb{P}[\sigma_A = x_A \mid \sigma_{A^c} = x_{A^c}]} &= \frac{\frac{1}{Z}\prod_{u \in V} \psi_u(y_u) \prod_{(u,v)\in E} \psi_{u,v}(y_u, y_v)}{\frac{1}{Z}\prod_{u \in V} \psi_u(x_u) \prod_{(u,v)\in E} \psi_{u,v}(x_u, x_v)} \\ &= \frac{\prod_{u \in A} \psi_u(x'_u) \prod_{(u,v)\in E(A\cup\partial A)} \psi_{u,v}(x'_u, x'_v)}{\prod_{u \in A} \psi_u(x_u) \prod_{(u,v)\in E(A\cup\partial A)} \psi_{u,v}(x_u, x_v)}. \end{aligned}$$

The lemma follows from the fact that the formula does not depend on x_{A^c} expect for $x_{\partial A}$. ■

The distribution of σ_A given $\sigma_{\partial A} = x_{\partial A}$ is simply the spin system on $A \cup \partial A$ with the spins on ∂A fixed to $x_{\partial A}$. It may be the case, for example for colourings, that not every boundary condition gives to configurations with positive probability. We call a spin system *permissive* if for every boundary condition on ∂A, the is a configuration of A with positive probability. The hardcore model is permissive because the empty independent set always has positive probability.

1.2. **Infinite Graphs.** So far we have only defined the model on finite graphs and in general (1.1) only makes sense on finite graphs. But we will want to take limits of graphs so defining spin systems on infinite graphs is important. This is done via the conditional distribution property from Lemma 3. On an infinite graph V with a set of weights $\{\psi_u\}, \{\psi_{u,v}\}$ a measure μ on $\mathcal{X}^V$ is infinite a *Gibbs measure* for the spin system if for any finite $A \subset V$,

$$\mathbb{P}[\sigma_A = x_A \mid \sigma_{A^c} = x_{A^c}] = \mathbb{P}[\sigma_A = x_A \mid \sigma_{\partial A} = x_{\partial A}]$$

holds and the conditional distribution is given by (1.1) on the finite graph $A \cup \partial A$. This is called the DLR property for Dobrushin–Lanford–Ruelle.

1.2.1. *Existence.* For all permissive spin systems, Gibbs measures exist. They can be constructed from taking limits of measures on finite spin systems. Let $\{D_i\}_{i\geq 1}$ be an increasing sequence of finite sets with $D_i \uparrow V$, let $x^{(i)}$ be a sequence of boundary conditions on $V \setminus D_i$. This define a sequence μ_i of measures on $\mathcal{X}^{D_i}$ given by

$$\mathbb{P}[\sigma_{D_i} \mid \sigma_{\partial D_i} = d_{\partial D_i}].$$

We say the sequence converges if its pushforward onto $\mathcal{X}^A$ converges for all finite A. Note that A will be s subset of D_i for all sufficiently large i. The sequence may not converge for all choices of boundary conditions but by compactness some subsequence will converge. This gives rise to a measure μ on $\mathcal{X}^V$. As the Markov random field property holds for all finite μ_i, the limit measure μ will also satisfy it and thus is a Gibbs measure.

1.2.2. *Uniqueness.* Given the existence of Gibbs measures, it is natural to ask if they are unique in which case we say the spin system has *uniqueness.* This, as we will see, depends on the spin system. For the ferromagnetic Ising model on $\mathbb{Z}^d$ with $d \geq 2$ and with no external field $h = 0$, there is a critical β_c such that there is uniqueness for $\beta \leq \beta_c$ and non-uniqueness for $\beta > \beta_c$. We will consider the case of large and small β but leave the behaviour around the critical temperature (see [**?**] for more details).

Large β**:** At low temperature the Ising model can be shown to have multiple Gibbs measures. We will consider the case $V = \mathbb{Z}^2$ where D_i are boxes or radius i and the boundary conditions are all plus. First we will derive some monotonicity properties of the Ising model.

We say that an event B on a partially ordered set is *increasing* if $x \in B$ implies $x' \in B$ for all $x' \geq x$. A measure μ *stochastically dominates* a measure μ' if for all increasing sets B,

$$\mu(B) \geq \mu'(B)$$

which we denote by $\mu \succeq \mu'$. If μ stochastically dominates μ' then there exists a coupling of $\sigma \sim \mu$ and $\sigma' \sim \mu'$ such that $\sigma \geq \sigma'$.

Lemma 4. *For the ferromagnetic Ising model, if A is a set and $x \leq x'$ are two boundary conditions on ∂A, then*

$$\mathbb{P}[\sigma_A \in \cdot \mid \sigma_{\partial A} = x_{\partial A}] \succeq \mathbb{P}[\sigma_A \in \cdot \mid \sigma_{\partial A} = x'_{\partial A}].$$

To prove this we will introduce the *Glauber dynamics* Markov chain. This is a Markov chain X_t taking values in $\mathcal{X}^V$ which can be defined for any spin system μ and is reversible with respect to μ, giving a way to simulate the distribution using *Markov Chain Monte Carlo* (MCMC) which essentially means running the Markov chain long enough until the chain is close to its stationary disribution. The transitions of X_t are as follows:

- Choose $v \in V$ uniformly at randomly.
- Set $X_{t+1}(u) = X_t(u)$ for $u \neq v$.
- Pick a_{t+1} according to the measure $\mathbb{P}[\sigma_v \in \cdot \mid v, \sigma_{\partial v} = X_t(\partial v)]$ and set $X_{t+1}(v) = a_{t+1}$.

A variant of the Glauber dynamics, which can be defined on infinite graphs, is its continuous time version where each spin is updated at times given by independent rate 1 Poisson clocks. In the case of the Ising model, the spin would be set to + with probability

$$\frac{e^{h+\beta \sum_{v \in \partial A} X_t(v)}}{e^{h+\beta \sum_{v \in \partial A} X_t(v)} + e^{-h-\beta \sum_{v \in \partial A} X_t(v)}} \tag{1.2}$$

which is an increasing function for the ferromagnetic Ising model.

Lemma 4. Let X_t and X'_t be the Ising model on A with boundary conditions x, x' respectively on ∂A and $X_0 \geq X'_0$. By equation (1.2) the probability of updating a vertex to a plus is increasing in X_t so we can couple X_t and X'_t so that the same vertex is chosen to be update and that $X_t \geq X'_t$ for all t. Let $B \subset \mathcal{X}^A$ be an increasing set, then

$$\mathbb{P}[X_t \in B] \geq \mathbb{P}[X'_t \in B]$$
$$X_t \xrightarrow{d} \mathbb{P}[\sigma_A \in \cdot \mid \sigma_{\partial A} = x_{\partial A}]$$
$$X'_t \xrightarrow{d} \mathbb{P}[\sigma_A \in \cdot \mid \sigma_{\partial A} = x'_{\partial A}]$$

and so

$$\mathbb{P}[\sigma_A \in \cdot \mid \sigma_{\partial A} = x_{\partial A}] \geq \mathbb{P}[\sigma_A \in \cdot \mid \sigma_{\partial A} = x'_{\partial A}]$$

which completes the proof. ■

It is similarly true that conditioning on more spins to be + stochastically increases the measure so if $D \subset D'$ and if $x_{V\setminus D'} \geq x'_{V\setminus D'}$ then

$$\mathbb{P}[\sigma_D \in \cdot \mid \sigma_{D'\setminus D} = +, \sigma_{V\setminus D'} = x_{V\setminus D'}] \succeq \mathbb{P}[\sigma_D \in \cdot \mid \sigma_{V\setminus D'} = x'_{V\setminus D'}].$$

Now consider a sequence of measures μ_i on D_i with all plus boundary conditions, which we can interpret as conditioned to be all plus outside of D_i. Then μ_i is a stochastically decreasing set of measures and so converges to a limit μ_+, called the*plus measure*. The plus measure stochastically dominates all Gibbs measures. We can construct μ_- in the same way and for all Gibbs measures

$$\mu_- \preceq \mu \preceq \mu_+.$$

So it will be enough to check if $\mu_- \neq \mu_+$. We begin with the case of β small.

The dual lattice has edge set

$$\mathbb{Z}^2_* = \{(x + \frac{1}{2}, y + \frac{1}{2}) : (u, v) \in \mathbb{Z}^2\}.$$

and edges between vertices at distance 1. Let $\mathcal{C}$ be the set of closed self-avoiding paths in $\mathbb{Z}^2_*$ and let $\mathcal{C}_0$ be the set of such contours whose interior contains the origin. Let B_γ be the event that σ is all plus on the exterior boundary of γ and all minus on the interior boundary of γ.

Suppose that $\sigma_0 = -$ under μ_i. Let $A = A(\sigma)$ be the set connected component of minuses in $\mathbb{Z}^2$ that contains the origin and let γ_A be the contour in $\mathcal{C}_0$ that surrounds it. Then B_{γ_A} holds. Now for a configuration $x \in \mathcal{X}^V$ construct the configuration x^γ where all the signs in the interior of γ are flipped.

$$x^\gamma_u = \begin{cases} -x_u & \text{if u is in the interior of } \gamma, \\ x_u & \text{otherwise.} \end{cases}$$

for neighbouring vertices $u \sim v$ the sign of $\sigma_u \sigma_v$ only changes when one vertex is in the interior and the other is in the exterior. When B_γ holds all are switched from minus to plus. Thus

$$\mu_i(x) = e^{-2\beta|\gamma|} \mu_i(x^\gamma)$$

for $x \in B_\gamma$. Hence

$$\mu_i(B_\gamma) = \sum_{x \in B_\gamma} \mu_i(x) = e^{-2\beta|\gamma|} \sum_{x \in B_\gamma} \mu_i(x^\gamma) \leq e^{-2\beta|\gamma|}$$

since the x^γ are distinct and the sum of their probabilities is bounded by 1. Since B_{γ_A} holds,

$$\mu_i(\sigma_0 = -) \leq \sum_{\gamma \in \mathcal{C}_0} \mu_i(B_\gamma) \leq \sum_{\gamma \in \mathcal{C}_0} e^{-2\beta|\gamma|}.$$

Any such contour is at least length 4. To count the number of contours of length ℓ in $\mathcal{C}_0$ note that they must cross the positive x-axis within distance ℓ of the origin and after that edge at each step there are at most 3 choices giving a total of at most $\ell 3^\ell$. Thus

$$\mu_i(\sigma_0 = -) \leq \sum_{\ell \geq 4} \ell 3^\ell e^{-2\beta\ell}$$

and so when $\beta \geq 1$ we have that $\mu_i(\sigma_0 = -) \leq \frac{1}{3}$ and so $\mu_+(\sigma_0 = -) \leq \frac{1}{3}$. By symmetry $\mu_+(\sigma_0 = -) \geq \frac{2}{3}$ so $\mu_- \neq \mu_+$ and we have non-uniqueness. This is called the *Peierls Argument*.

These two Gibbs measures are translation invariant and have different densities of pluses and minuses. In dimension 2 the only Gibbs measures are mixtures of μ_+ and μ_- but in higher dimensions there are more exotic boundary conditions called *Dobrushin states* where the configuration is predominantly plus on one sides of the hyperplane $y \geq 0$ and predominantly minus on the other.

Small β: We'll give two proofs that there is a unique Gibbs measure when β is small. For the first we introduce the *Fortuin-Kasteleyn model*, sometimes known as the *random cluster model* or the FK-model. The q-state FK model is a probability distribution over percolation configurations $\xi \in \{0,1\}^E$ given by

$$\mathbb{P}[\xi = w] = \frac{1}{Z} p^{\sum_{e\in E} w_e} (1-p)^{|E| - \sum_{e\in E} w_e} q^{C(w)}.$$

where $C(w)$ is the number of connected components of the configuration w. Note that the case of $q = 1$ corresponds to independent percolation. We will write $u \leftrightarrow v$ to mean that there is a path of open edges in ξ from u to v.

Lemma 5. *The FK-model with $q > 1$ is stochastically dominated by bond percolation with probability p.*

Proof. For an edge $e = (u, v)$ we calculate

$$\mathbb{P}[\xi_e = 1 \mid \xi_{E\setminus e}] = \begin{cases} p, & \text{if } u \leftrightarrow v in \xi_{E\setminus e} \\ \frac{p}{p+2(1-p)}, & \text{otherwise.} \end{cases}$$

In each case $\mathbb{P}[\xi_e = 1 \mid \xi_{E\setminus e}] \leq 1$ and thus if we reveal the edges of ξ one by one we can stochastically dominate it by Bernoulli percolation with probability p. ■

The *Edwards-Sokal coupling* is a joint distribution of the Ising (or Potts) model and the FK-model. Given an FK-configuration ξ one chooses a Potts configuration $[q]^V$ as follows. Choose an independent uniform spin s_A from $[q]$ for each component A of the percolation cluster ξ. Assign $\sigma_u = s_A$ for all

$u \in A$. A pair (σ, ξ) is admissible if the spins in each percolation component are constant and let $\mathcal{A}$ denote the set of admissible pairs. We can calculate

$$\mathbb{P}[(\sigma, \xi) = (x, w)] = \mathbb{P}[\xi = w] q^{-C(w)} = \frac{1}{Z} p^{\sum_{e \in E} w_e} (1-p)^{|E| - \sum_{e \in E} w_e}.$$

Then with $\beta = -\log(1-p)$,

$$\begin{aligned} \mathbb{P}[\sigma = x] &= \sum_{w:(x,w) \in \mathcal{A}} \mathbb{P}[(\sigma, \xi) = (x, w)] \\ &= \frac{1}{Z} \sum_{w:(x,w) \in \mathcal{A}} p^{\sum_{e \in E} w_e} (1-p)^{|E| - \sum_{e \in E} w_e} \\ &= \frac{1}{Z} \prod_{u \sim v} (1-p)^I (\sigma_u \neq \sigma_v) \\ &= \frac{1}{Z'} \prod_{u \sim v} \exp(-\beta I(\sigma_u \neq \sigma_v)) = \frac{1}{Z''} \exp(\beta \sum_{u \sim v} I(\sigma_u = \sigma_v)) \end{aligned}$$

using the fact that admissability means that $w_{(u,v)}$ must be 0 if $\sigma_u \neq \sigma_v$ and can be either 0 or 1 if $\sigma_u = \sigma_v$. Thus the marginal σ is distributed according to the q-state Potts model.

Now let us construct the measure μ_i in terms of the FK model. The all plus boundary condition on D_i corresponds to the wired boundary condition on D_i in the FK-model meaning that all the vertices of ∂D_i are connected together and the spins will be set to $+$. Thus

$$\mu_i(\sigma_0 = +) = \mu_i(0 \leftrightarrow \partial D_i) + \frac{1}{2}(1 - \mu_i(0 \leftrightarrow \partial D_i)) = \frac{1}{2} + \frac{1}{2}\mu_i(0 \leftrightarrow \partial D_i)$$

since if $0 \not\leftrightarrow \partial D_i$ then σ_0 is set to $+$ with probability $\frac{1}{2}$. By Lemma 5 $\mu_i(0 \leftrightarrow \partial D_i)$ is bounded above by the probability of a connection from 0 to ∂D_i in bond percolation. This means that there is an open path of length at least i starting at 0 each of which has probability p^i. There are at most 4^i such paths so if $p < \frac{1}{4}$,

$$\mu_i(0 \leftrightarrow \partial D_i) \leq (4p)^i \to 0.$$

Hence for $p < \frac{1}{4}$ which corresponds to $\beta < \log(\frac{4}{3})$ we have that $\mu_i(\sigma_0 = +) \to \frac{1}{2}$ and so $\mu_+(\sigma_0 = +) = \frac{1}{2}$. Similarly $\mu_-(\sigma_0 = +) = \frac{1}{2}$. By translation invariance this holds for al vertices in $\mathbb{Z}^2$. Since μ_+ stochastically dominates μ_- we can find couple $\sigma^\pm \sim \mu_\pm$ so that $\sigma^+ \geq \sigma^-$. So

$$\mathbb{P}[\sigma_v^+ \neq \sigma_v^-] = \mathbb{P}[\sigma_v^+ = 1, \sigma_v^- = -1] = \mathbb{P}[\sigma_v^+ = 1]\mathbb{P}[\sigma_v^- = -1] = 0.$$

Thus $\sigma^+ \equiv \sigma^-$ and so $\mu_+ = \mu_-$ and there is a unique Gibbs measure.

In this argument we specifically used the monotonicity of the Ising model. Many spin systems are not monotone but nonetheless if they have weak interactions there will still be a unique Gibbs measure. Another argument can be given in terms of the Glauber dynamics. On the lattice Z^d suppose that

$$\max_{v,x_{\partial v},x'_{\partial v}} d_{TV}(\mathbb{P}[\sigma_v \in \cdot \mid \sigma_{\partial v} = x_{\partial v}], \mathbb{P}[\sigma_v \in \cdot \mid \sigma'_{\partial v} = x'_{\partial v}]) \le \frac{1-\epsilon}{2d}. \quad (1.3)$$

Note this condition can be strengthened, stronger versions include the Dobrushin and the Dobrushin–Shlosman conditions. Let μ and μ' be two Gibbs measures of the spin system and X_t, X'_t be continuous time Glauber dynamics initialized with $X_0 \sim \mu$ and $X'_0 \sim \mu'$. The Glauber dynamics is reversible with respect to any Gibbs measure so $X_t \sim \mu$ and $X'_t \sim \mu'$ for all t. We couple the two chains so that vertices are updated at the same time, that is the same Poisson clocks are used. Also for each update we couple the choice of a new spin to minimize the probability that the disagree. Condition (1.3) ensures that when a vertex v is updated the probability of a disagreement is never more than $\frac{1-\epsilon}{2d}$ and if $X_t(\partial v) = X'_t(\partial v)$ then we can update both chains with the same spin. We choose α satisfying $1-\epsilon < \alpha < 1$ and set

$$D_t = \sum_{v \in V} \alpha^{d(v,0)} I(X_t(v) \neq X'_t(v))$$

To measure the amount of disagreement between X_t and X'_t. When a vertex v with a disagreement is updated its disagreement is (possibly temporarily) removed and D_t decreases by $\alpha^{d(v,0)}$. But v has a disagreement and its neighbour is updated with some probability, bounded by $\frac{1-\epsilon}{2d}$, the neighbour may become disagreeing. Thus with $\mathcal{F}_t$ denoting the filtration generated by the chains,

$$\frac{d}{dt}\mathbb{E}[D_t \mid \mathcal{F}_t] \le \sum_{v \in V} -\alpha^{d(v,0)} + \frac{1-\epsilon}{2d} \sum_{u \sim v} \alpha^{d(v,0)} \le D_t((1-\epsilon)\alpha^{-1} - 1)$$

Since $((1-\epsilon)\alpha^{-1} - 1)$ and $D_0 \le \sum_{v \in V} \alpha^{d(v,0)} < \infty$, we have that $\mathbb{E} D_t \to 0$ exponentially quickly. Hence for any finite $A \subset \mathbb{Z}^2$ we have that $\mathbb{P}[X_t(A) \neq X_t(A)] \to 0$. But X_t and X'_t are stationary so their stationary measures must be identical. Hence there is a unique Gibbs measure.

2. Spin systems on Trees

As tree are the local weak limits of random graphs, spin systems on trees play a major role in understanding spin systems on random graphs. The Markov property is particularly useful here as conditional on the spin at one vertex, it's neighbours are conditionally independent. This conditional independence makes the models amenable to recursive calculations of marginals.

[Add Figure]

Let T be a finite tree rooted at ρ with children $u_1, \ldots, u_d$ and let T' be the graph with the edges at ρ removed. We will denote $\mathbb{P}_G$ denote the Gibbs measure on a graph G. We will relate the marginal of ρ in T in terms of the marginal of the u_i in the T' which will give a method for recursively calculating marginal distributions.

Let T_i be the subtree rooted at u_i and let $m_{u_i\to\rho}(x_i) = \mathbb{P}_{T_i}[\sigma_{u_i} = x_i]$ denote the marginal of u_i in T_i. In the graph T' subtrees T_i are disconnected and so their spins are independent. Then summing over all the spin configurations on ρ and the u_i,

$$\begin{aligned} Z_T &= \sum_{x\in\mathcal{X}} \psi(x) \sum_{\{x_i\}\in\mathcal{X}^d} \prod_{i=1}^{d} \psi(x,x_i) m_{u_i\to\rho}(x_i) Z_{T_i} \\ &= \sum_{x\in\mathcal{X}} \psi(x) \prod_{i=1}^{d} \left(\sum_{x_i\in\mathcal{X}} \psi(x,x_i) m_{u_i\to\rho}(x_i) Z_{T_i} \right) \end{aligned}$$

and

$$\mathbb{P}_T[\sigma_\rho = x_\rho] = \frac{\psi(x_\rho)\prod_{i=1}^d \left(\sum_{x_i\in\mathcal{X}} \psi(x_\rho,x_i) m_{u_i\to\rho}(x_i)\right)}{\sum_{x'\in\mathcal{X}} \psi(x')\prod_{i=1}^d \left(\sum_{x_i\in\mathcal{X}} \psi(x',x_i) m_{u_i\to\rho}(x_i)\right)} \tag{2.1}$$

Building up from the leaves this gives an algorithm to determine the marginal at the root. We can formalize this in terms of what are called message passing algorithms. Let $\overrightarrow{E}$ be the set of directed edges in the graph. For $(u,v)\in\overrightarrow{E}$ we write

$$m_{u\to v}(x_u) := \mathbb{P}_{T\backslash(u,v)}[\sigma_u = x_u]$$

where $T\backslash(u,v)$ is the tree T with the edge (u,v) removed. This is interpreted as the message from u telling v what it believes its marginal is removing the effect of v. For u a leaf of the tree,

$$m_{u\to v}(x_u) = \frac{\psi(x_u)}{\sum_{x'\in\mathcal{X}}\psi(x')}.$$

The beliefs on internal edges can be calculated by applying the same reasoning as in equation (2.1),

$$\begin{aligned} m_{u\to v}(x_u) &= \frac{\psi(x_u)\prod_{w\in\partial u\backslash v}\left(\sum_{x_w\in\mathcal{X}}\psi(x_u,x_w)m_{w\to u}(x_i)\right)}{\sum_{x'\in\mathcal{X}}\psi(x')\prod_{w\in\partial u\backslash v}\left(\sum_{x_w\in\mathcal{X}}\psi(x',x_w)m_{w\to u}(x_i)\right)} \\ &=: BP[\{m_{w\to u}\}_{w\in\partial u\backslash v}](x_u). \end{aligned}$$

The function BP is called the *belief propagation* function. The marginal distribution at a vertex is simply

$$\mathbb{P}_T[\sigma_v = x_v] = BP[\{m_{u\to v}\}_{u\in\partial v}](x_u)$$

A set of messages $\{r_{u\to v}\}_{(u,v)\in\overrightarrow{E}}$ is a BP-fixed point if

$$r_{u\to v} = BP[\{r_{w\to u}\}_{w\in\partial u\setminus v}] \tag{2.2}$$

for all $(u,v)\in\overrightarrow{E}$. Our construction show that on a tree this is unique. However, on an infinite tree or a graph with cycles there may be multiple BP-fixed points

On an infinite graph each set of BP messages $\{r_{u\to v}\}$ defines a Gibbs measure. The distribution of σ_A for a finite connected set $A\subset V$ is given by

$$\mu_r[\sigma_A = x_A] = \frac{1}{Z}\sum_{x_{\partial A}\in\mathcal{X}^{\partial A}}\prod_{\substack{u\sim v\\ u\in\partial A, v\in A}} r_{u\to v}(x_u)\psi(x_u,x_v)\prod_{v\in A}\psi(x_v)\prod_{\substack{v\sim v'\\ v,v'\in A}}\psi(v,v')$$

which can be verified to satisfy the properties of a Gibbs measure. The effect of the measure outside of $A\cup\partial A$ is to change ψ_u with $r_{u\to v}(x_u)$ on ∂A. We will let μ_r generated by the Gibbs measure from the BP-fixed point r.

A special case of a BP-fixed point in a d-regular tree is when all the messages in the tree are constant and

$$m = BP[m,\dots,m] \tag{2.3}$$

for $d-1$ copies of m. We will call this a *translation invariant BP-fixed point* or TIFP. For any homogeneous spin system there is always a TIFP by the Brouwer fixed-point theorem and gives rise to a translation invariant Gibbs measure.

Ising model: We calculate the TIFPs in the case of the Ferromagnetic Ising model on a $d+1$ regular tree. We parameterise the marginals by $y = m_{u\to v}(+) - \frac{1}{2}$ and equation (2.2) gives

$$\begin{aligned} y &= BP[m,\dots,m] - \frac{1}{2}\\ &= \frac{((\frac{1}{2}+y)e^{\beta}+(\frac{1}{2}-y)e^{-\beta})^d}{((\frac{1}{2}+y)e^{\beta}+(\frac{1}{2}-y)e^{-\beta})^d+((\frac{1}{2}+y)e^{\beta}+(\frac{1}{2}-y)e^{-\beta})^d} - \frac{1}{2}\\ &= \frac{\frac{1}{2}(1+2y\tanh\beta)^d - \frac{1}{2}(1-2y\tanh\beta)^d}{(1+2y\tanh\beta)^d+(1-2y\tanh\beta)^d} =: f(y). \end{aligned}$$

[Figure of $f(y)$]

The TIFP for the Ising model correspond to solutions of $y = f(y)$. We always have $f(0) = 0$ as a solution which corresponds to $m_{u\to v}(\pm) = \frac{1}{2}$, the Gibbs measure which is symmetric between plus and minus. It is called the *free measure* which corresponds to a limit of finite trees with free boundary conditions, another way of saying no boundary conditions. Let y_+ be the largest root of $y = f(y)$ and y_- the smallest root.

By considering a tree with plus boundary conditions at depth ℓ from u, we have

$$m_{u\to v}(+) - \frac{1}{2} = f(f(\dots f(\frac{1}{2})\dots))$$

iterated ℓ times. Note that since f is an increasing function, this is a decreasing sequence which converges to y_+, the largest root of $y = f(y)$. Now the plus measure μ_+ is a limit of these boundary conditions as ℓ goes to infinity and so is a TIFP whose messages are $\frac{1}{2} + y_+$. Similarly μ_- is a TIFP with messages $\frac{1}{2} + y_-$.

Note that f is concave for $y > 0$ and convex for $y < 0$ so whether the are multiple solutions is determined by the derivative at 0 which is

$$f'(0) = d \tanh \beta.$$

There is a unique solution when $0 \le d \tanh \beta \le 1$ in which case $y_- = 0 = y_+$ and so $\mu_- = \mu_+$ and the Gibbs measure is unique. When $d \tanh \beta > 1$ there are three distinct solutions $y_- < 0 < y_+$ and three TIFPs corresponding to the free measures, μ_+ and μ_-. These are not the only Gibbs measures, for instance taking plus boundary conditions in some subtrees of the tree and minus in others can produce a wide range of BP-fixed points which are close to the plus measure in some parts of the tree and close the minus measure in others.

In the case of the anti-ferromagnetic Ising model when $\beta < 0$ there is only ever a single solution and so only one TIFP. But there are multiple Gibbs measures when $d \tanh \beta < -1$. In this case there are multiple solutions to $f(f(y))$ and they correspond to semi-translation invariant Gibbs measures where even sites are more likely to be plus and odd sites are more likely to be minus. Another way to see this is the mapping $\sigma_u^* = (-1)^{d(u,\rho)}\sigma_u$ which flips the spins of the vertices at an odd distance to the root. It can be verified that σ^* is distributed according to a ferromagnetic Gibbs measure with inverse temperature $-\beta > 0$.

3. Reconstruction Threshold

The *reconstruction problem* asks whether distant spins provide information about the state at the root or in its neighbourhood under some Gibbs measure μ. We let B_ℓ be the set of vertices at distance at most ℓ from the root ρ and $S_\ell = \{u : d(u,\rho) = \ell\}$ denote the vertices at distance exactly ℓ. For a finite set $A \subset V$, we write

$$n_{A,\ell}(x_A) = \mu(\sigma_A = x_A \mid \sigma_{B_\ell^c}).$$

Since $\sigma_{B_\ell^c}$ is a random variable, so is $n_{A,\ell}$. Note that large enough ℓ we have that $A \subset B_\ell$ and so by the Markov Random Field Property $n_{A,\ell}(x_A) =$

$\mu(\sigma_A = x_A \mid \sigma_{S_\ell})$. Furthermore, $n_{A,\ell}$ is a bounded backwards martingale so converges almost surely,

$$n_{A,\ell}(x_A) \overset{a.s}{\to} n_A(x_A).$$

Taking expected values,

$$\mathbb{E}n_A(x_A) = \mathbb{E}n_{A,\ell}(x_A) = \lim_{\ell} n_{A,\ell}(x_A) = \mu(\sigma_A = x_A).$$

If $n_A(x_A)$ were always identically equal to $\mu(\sigma_A = x_A)$ then distant spins provide no information asymptotically about σ_A. If this holds for all A then we say that the reconstruction problem is *non-solvable* for μ. It is also equivalent to tail-triviality of the Gibbs measure which we describe as follows. Conversely if for some x_A,

$$\mathbb{P}[n_A(x_A) \neq \mu(\sigma_A = x_A)] > 0$$

then we say that the reconstruction problem is *solvable*. At the same time we may construct a random BP-fixed point derived from σ as follows. We set

$$m^{(0)}_{u\to v}(x) = I(\sigma_u = x)$$

and

$$m^{(t)}_{u\to v}(x) = BP[\{m^{(t-1)}_{u'\to u}(x)\}_{u'\in\partial u\setminus v}],$$

that is $m^{(t)}_{u\to v}$ is the conditional distribution of u in the tree $T\setminus(u,v)$ with a boundary condition on at depth t from u given by σ. The joint distribution of an edge given the spins at distance t is

$$\frac{1}{Z} m^{(t)}_{u\to v}(x_u) m^{(t)}_{v\to u}(x_v)\psi(x_u,x_v)$$

Since this converges to $n_{(u,v)}(x_u,x_v)$ but $m^{(t)}_{u\to v}$ and $m^{(t)}_{v\to u}$ as $t\to\infty$. We denote this limit $m_{u\to v} = m^{\sigma}_{u\to v}$ which by construction must be a BP-fixed point. The probability n_A can be constructed as

$$n_A(X_A) = \frac{1}{Z}\sum_{x_{\partial A}}\prod_{u\in A}\psi(x_u)\prod_{\substack{u\sim v\\ u,v\in A}}\psi(x_u,x_v)\prod_{\substack{u\sim v\\ u\in\partial A, v\in A}} m_{u\to v}(x_u)\psi(x_u,x_v).$$

An equivalent formulation of solvability of the reconstruction problem is extremality of Gibbs measures. Recall that any convex combination of Gibbs measures is also a Gibbs measures. We say a Gibbs measure is *extremal* or *pure* if it cannot be written as a non-trivial convex combination of other Gibes measures. This decomposition will be important as we consider the clustering of solutions of random constraint satisfaction problems. The next lemma shows the equivalence of these two ideas.

Lemma 6. *A Gibbs measure μ is extremal if and only if the reconstruction problem is solvable. The set of measures*

Proof. Let ν be the measure over BP-fixed points m^σ induced by μ. The reconstruction problem being non-solvable is equivalent to the n being deterministic which is equivalent to m being deterministic and ν being a point mass. If ν is not a point mass then $\mu = \int \mu_m \nu(dm)$ is a non-trivial convex combination of Gibbs measures so μ is not extremal. Conversely if μ can be written as a non-trivial convex combination of Gibbs measures $\mu = p\mu' + (1-p)\mu''$ then we can write $\nu = p\nu' + (1-p)\nu''$ and so ν cannot be a point mass since $\nu' \neq \nu''$. ■

Clearly in the case of a unique Gibbs measure, it is extremal. But they are not equivalent, we will see that reconstruction is a strictly stronger property than non-uniqueness. Let us consider the Ising model and when its TIFP's are extremal. The measures μ_+ and μ_- are extremal as they are the maximal and minimal measures respectively. This can be seen as follows, suppose that $\mu_+ = p\mu' + (1-p)\mu''$. For any increasing set B,

$$\mu_+(B) = p\mu'(B) + (1-p)\mu''(B), \quad \mu_+(B) \geq \mu'(B), \mu''(B)$$

and so $\mu_+(B) = \mu'(B) = \mu''(B)$. As any $I(\sigma_A = x_A)$ can be written as a linear combination of indicators of increasing functions it follows that $\mu = \mu' = \mu''$ and so μ_+ is extremal and similarly μ_-.

Now consider the case of the symmetric TIFP where $m \equiv \frac{1}{2}$. This is the called the *free measure* as it arises as a limit for free boundary conditions meaning no spins or interactions on ∂A.

A *broadcast model* is a random spin configuration on the tree given by the following Markov model. Let M be a $\mathcal{X} \times \mathcal{X}$-Markov transition matrix M, reversible with respect to π. Then the broadcast model on T given by M chooses the state of the root σ_ρ according to π and then the state of children is assigned as $\mathbb{P}[\sigma_u = y \mid \sigma_{u^+} = x] = M_{xy}$ where u^+ is the parent of u. Thus along each path in the tree, the states are given by the Markov chain with transition M in stationarity. Reversibility means that the distribution does not depend on the location of the root.

All TIFP are given by broadcast models. This follows from the fact that if T' is the component of $T \setminus (u, u^+)$ containing u^+ then by the Markov random field property,

$$\mathbb{P}[\sigma_u = y \mid \sigma_T] = \mathbb{P}[\sigma_u = y \mid \sigma_{u^+}]$$

and the transition probability is

$$M_{xy} = \mathbb{P}[\sigma_u = y \mid \sigma_{u^+} = x] = \frac{\psi(x,y) m_{u\to u^+}(y)}{\sum_{y' \in \mathcal{X}} \psi(x,y') m_{u \to u^+}(y')}.$$

In the case of the free measure

$$M = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}\theta & \frac{1}{2} - \frac{1}{2}\theta \\ \frac{1}{2} - \frac{1}{2}\theta & \frac{1}{2} + \frac{1}{2}\theta \end{pmatrix}$$

where $\theta = \tanh \beta$. The eigenvalues of M are 1 and θ. An alternative description of the Markov transition is that with probability θ the child has the same spin as the parent and with probability $1-\theta$ it has an independent spin. Then for $u, v \in T$,

$$\mathbb{E}\sigma_u\sigma_v = \mathbb{P}[\sigma_u = \sigma_v] - \mathbb{P}[\sigma_u \neq \sigma_v] = 2\mathbb{P}[\sigma_u = \sigma_v] - 1,$$

and

$$\mathbb{P}[\sigma_u = \sigma_v] = \theta^{d(u,v)} + \frac{1}{2}(1 - \theta^{d(u,v)}) = \frac{1}{2} + \frac{1}{2}\theta^{d(u,v)}$$

taking them as equal if in each step along the path the child copied the parent and otherwise the states are independent. Together we have that

$$\mathbb{E}\sigma_u\sigma_v = \mathrm{Cov}(\sigma_u, \sigma_v) = \theta^{d(u,v)}.$$

3.1. **Reconstruction and the Kesten-Stigum Bound.** If θ is the second eigenvalue of the broadcast channel the reconstruction problem is always solvable on the $(d+1)$-regular tree or d-ary tree when $d\theta^2 > 1$. This was established by Kesten and Stigum in the context of multi-type branching processes. Reconstruction is done simply by using only the information on the number of spins of each type in level ℓ without using the information of how those spins are arranged. We will prove this for the free Ising model on the d-ary tree.

Let

$$Y_\ell = (d\theta)^{-\ell} \sum_{u \in S_\ell} \sigma_u.$$

Then

$$\mathbb{E}[Y_\ell \mid \sigma_\rho = 1] = \mathbb{E}[Y_\ell \sigma_\rho] = (d\theta)^{-\ell} d^\ell \theta^\ell = 1.$$

Computing its variance using the fact that each leaf of the tree has $(d-1)d^{k-1}$ other leaves at distance $2k$,

$$\begin{aligned} \mathrm{Var} Y_\ell &= (d\theta)^{-2\ell} \sum_{u,v \in S_\ell} \mathrm{Cov}(\sigma_u, \sigma_v) \\ &= (d\theta)^{-2\ell} d^\ell \left(1 + \sum_{k=1}^{\ell} (d-1) d^{k-1} \theta^{2k}\right) \\ &= d^\ell \theta^{-2\ell} \left(1 + (d-1)\theta^2 \frac{(d\theta^2)^{\ell-1} - 1}{d\theta^2 - 1}\right). \\ &= O(1). \end{aligned}$$

where the last equality used the fact that $d\theta^2 > 1$. Since $\mathbb{E}[\sigma_u \mid \mathcal{F}_\ell] = \theta\sigma_{u^+}$,

$$\mathbb{E}[Y_{\ell+1} \mid \mathcal{F}_\ell] = (d\theta)^{-\ell-1} \sum_{u\in S_{\ell+1}} \mathbb{E}[\sigma_u \mid \mathcal{F}_\ell] = (d\theta)^{-\ell-1} \sum_{u\in S_\ell} d\theta\sigma_u = Y_\ell$$

and so Y_ℓ is an L^2-bounded martingale. Hence Y_ℓ converges in L^2 to Y such that

$$\mathbb{E}Y, \quad \mathbb{E}[Y\sigma_\rho = 1] = 1, \quad \mathrm{Var}(Y) = O(1).$$

Since Y is measurably with respect to the tail event σ-algebra but is correlated with σ_ρ the reconstruction problem is solvable. For any broadcast process there is a generalization of this argument proving reconstruction.

3.2. **Non-reconstruction for the Ising model.** Since there is non-reconstruction when uniqueness holds we are left with the case of $\frac{1}{d} < \theta \leq \frac{1}{\sqrt{d}}$.

Theorem 7. *When $0 \leq \tanh\beta \leq \frac{1}{\sqrt{d}}$ the reconstruction problem is non-solvable on the d-ary tree.*

Proof. Let ρ be the root of the tree with children $u_1,\dots,u_d$. Let $m_{\rho,n} = \mathbb{P}[\sigma_\rho = + \mid \sigma_{S_\ell}]$ and let $m_{u_i,n} = m^{(n)}_{u_i\to\rho}(+)$. We will write $\mathbb{E}_+[\cdot]$ for $\mathbb{E}[\cdot \mid \sigma_\rho = +]$. Let

$$x_n = \mathbb{E}_+[m_{\rho,n}] - \frac{1}{2}.$$

We will show that x_n controls the variance of $m_{\rho,n}$.

Claim 8. *The conditional probability satisfies*

$$\mathbb{E}[(m_{\rho,n} - \frac{1}{2})^2] = \mathbb{E}_+[(m_{\rho,n} - \frac{1}{2})^2] = \frac{1}{2}x_n$$

Proof. The first equality holds by symmetry in plus and minus. Then

$$\begin{aligned}
x_n + \frac{1}{2} &= \mathbb{E}[m_{\rho,n} \mid \sigma_\rho = +]\\
&= \sum_{B\in\mathcal{X}^{S_\ell}} \mathbb{P}[\sigma_\rho = + \mid \sigma_{S_\ell} = B]\mathbb{P}[\sigma_{S_\ell} = B \mid \sigma_\rho = +]\\
&= \sum_{B\in\mathcal{X}^{S_\ell}} \mathbb{P}[\sigma_\rho = + \mid \sigma_{S_\ell} = B]\frac{\mathbb{P}[\sigma_\rho = + \mid \sigma_{S_\ell} = B]\mathbb{P}[\sigma_{S_\ell} = B]}{\mathbb{P}[\sigma_\rho = +]}\\
&= 2\sum_{B\in\mathcal{X}^{S_\ell}} \mathbb{P}[\sigma_\rho = + \mid \sigma_{S_\ell} = B]^2\mathbb{P}[\sigma_{S_\ell} = B]\\
&= 2\mathbb{E}[(m_{\rho,n})^2] = 2\mathbb{E}[(m_{\rho,n} - \frac{1}{2})^2] + \frac{1}{2}.
\end{aligned}$$

∎

Thus showing that $x_n \to 0$ will imply that $m_{\rho,n}$ converges in probability to $\frac{1}{2}$ and there is non-reconstruction. We proceed by analysing x_n recursively.

Conditional on the spin at the root, the $m_{u_i,n}$ are conditionally independent and so

$$\begin{aligned}\mathbb{E}_+[m_{u_i,n} - \frac{1}{2}] &= (\frac{1}{2} + \theta)\mathbb{E}[m_{u_i,n} - \frac{1}{2} \mid \sigma_{u_i} = +] + (\frac{1}{2} - \theta)\mathbb{E}[m_{u_i,n} - \frac{1}{2} \mid \sigma_{u_i} = -]\\ &= \theta\mathbb{E}[m_{u_i,n} - \frac{1}{2} \mid \sigma_{u_i} = +] = \theta x_{n-1}.\end{aligned}$$

We also have that

$$\mathbb{E}_+[(m_{u_i,n} - \frac{1}{2})^2] = (1+\theta)\mathbb{E}[(m_{u_i,n} - \frac{1}{2})^2 \mid \theta_{u_i} = +] + (1-\theta)\mathbb{E}[(m_{u_i,n} - \frac{1}{2})^2 \mid \theta_{u_i} = -] = \frac{x_{n-1}}{2}$$

Then the BP equation gives that

$$m_{\rho,n} = \frac{\prod_{i=1}^d \left(1 + 2\theta(m_{u_i,n} - \frac{1}{2})\right)}{\prod_{i=1}^d \left(1 + 2\theta(m_{u_i,n} - \frac{1}{2})\right) + \prod_{i=1}^d \left(1 - 2\theta(m_{u_i,n} - \frac{1}{2})\right)} = \frac{Z_+}{Z_+ + Z_-}.$$

where $Z_\pm = \prod_{i=1}^d \left(1 \pm 2\theta(m_{u_i,n} - \frac{1}{2})\right)$. We now evaluate the moments of $Z_\pm$. First

$$\mathbb{E}_+[Z_\pm] = (1 \pm 2\theta\mathbb{E}_+[m_{u_i,n} - \frac{1}{2}])^d = (1 \pm 2\theta x_{n-1})^d.$$

For the second moments

$$\begin{aligned}\mathbb{E}_+[Z_+^2] &= \left(\mathbb{E}_+[1 + 2\theta(m_{u_i,n} - \frac{1}{2})]^2\right)^d\\ &= \left(1 + 4\theta\mathbb{E}_+[m_{u_i,n} - \frac{1}{2}] + 4\theta^2\mathbb{E}_+[(m_{u_i,n} - \frac{1}{2})^2]\right)^d\\ &= \left(1 + 4\theta^2 x_{n-1} + 4\theta^2 \cdot x_{n-1}/2\right)^d = \left(1 + 6\theta^2 x_{n-1}\right)^d.\end{aligned}$$

Similarly

$$\mathbb{E}_+[Z_-^2] = \left(1 - 2\theta^2 x_{n-1}\right)^d, \qquad \mathbb{E}_+[Z_+ Z_-] = \left(1 - 2\theta^2 x_{n-1}\right)^d$$

Using the equality

$$\frac{1}{a+b} = \frac{1}{a} - \frac{b}{a^2} + \frac{b^2}{a^2}\frac{1}{a+b},$$

we have that

$$\begin{aligned}m_{\rho,n} - \frac{1}{2} &= = \frac{Z_+}{Z_+ + Z_-} - \frac{1}{2}\\ &= \frac{Z_+}{2} - \frac{Z_+(Z_+ + Z_- - 2)}{4} + \frac{(Z_+ + Z_- - 2)^2}{4} \cdot \frac{Z_+}{Z_+ + Z_-} - \frac{1}{2}\\ &= \frac{Z_+}{2} - \frac{Z_+(Z_+ + Z_- - 2)}{4} + \frac{(Z_+ + Z_- - 2)^2}{4} - \frac{1}{2}\\ &= \frac{1}{2} + \frac{Z_+ Z_- + Z_-^2 - 4Z_-}{4}.\end{aligned}$$

Taking expected values and using the fact that $(1-x)^d \geq 1 - xd$,

$$\begin{aligned} x_n &\leq \mathbb{E}_+\left[\frac{1}{2} + \frac{Z_+Z_- + Z_-^2 - 4Z_-}{4}\right] \\ &= \frac{1}{2} - \frac{1}{2}(1-2\theta^2 x_{n-1})^d \\ &\leq \frac{1}{2} - \frac{1}{2}(1-2\theta^2 d x_{n-1}) \\ &= d\theta^2 x_{n-1}. \end{aligned}$$

Thus if $d\theta^2 < 1$ then $x_n \to 0$. In the case $d\theta^2 = 1$ a more careful analysis of the above equation, going to a higher order Taylor Series expansion gives $x_n \leq x_{n-1} - cx_{n-1}^2$ and so again $x_n \to 0$. Thus $m_{\rho,n}$ converges in probability to $\frac{1}{2}$ and there the reconstruction problem is non-solvable. ■

Thus for the Ising model the Kesten-Stigum bound is in fact tight.

3.3. **Freezing and reconstruction for the colouring model.** The Kesten-Stigum bound is not tight for all spin systems as we will prove in the case of the q-state colouring model, at least for large values of q. In this case the transition matrix of the broadcast model is $M_{xy} = \frac{1}{q-1} I(x \neq y)$ which has second eigenvalue $\theta = -\frac{1}{(q-1)^2}$. So the Kesten-Stigum bound implies reconstruction when $d > (q-1)^2$.

In the colouring model it is sometimes the case that the boundary condition exactly determines the value at the root which is called *freezing*. At depth one this would correspond to all the other colours appearing among the children of the root leaving the colour of the root as the only possibility. We write p_ℓ for the depth ℓ freezing probability,

$$p_n := \mathbb{P}[m_{\rho\to\rho^+}(x) = 1 \mid \sigma_\rho = x],$$

with $p_0 = 1$. Given a boundary condition at depth ℓ, then $\mathrm{Bin}(d, p_{n-1})$ of the children of ρ can be determined exactly.

Having all colours appear among the children is an instance of the coupon collector problem. Let $f(n,m)$ denote the probability that after m IID samples drawn uniformly from n possible 'coupons' that all n have appeared at least once. The well known theorem for the coupon collector problem is that for any $\delta > 0$,

$$f(n, (1+\delta)n\log n) \to 1, \qquad f(n, (1-\delta)n\log n) \to 0$$

as $n \to \infty$. If m of the children of the root are known exactly, then the root is known with probability $f(q-1, m)$. Thus

$$p_n = \mathbb{E}[f\left(q-1, \mathrm{Bin}(d, p_{n-1})\right)]$$

If $d \geq (1+\epsilon) q \log q$ then for large enough q and $\delta > 0$ sufficiently small,

$$\mathbb{P}[\text{Bin}(d, 1-\delta) \geq (1+\epsilon/2) q \log q] \geq 1 - \delta/2,$$

and

$$f\left(q-1, (1+\epsilon/2) q \log q\right) \geq \frac{1-\delta}{1-\delta/2}.$$

Thus if $p_{n-1} \geq 1-\delta$ then

$$p_n \geq \mathbb{E}[f\left(q-1, \text{Bin}(d, 1-\delta)\right)] \geq 1-\delta,$$

and $\inf p_n \geq 1-\delta$ which implies the reconstruction problem is solvable since the root is frozen with probability bounded away from 0. Conversely if $d \leq q \log q$ it can be shown that there is non-reconstruction for large q so $q \log q$ gives the correct asymptotics, much smaller that $(q-1)^2$ from the Kesten-Stigum bound.

4. Free Energy and the Ising model on Random Graphs

A key quantity in the analysis of spin systems and random CSPs is the *free energy*, the normalized log-partition funciton

$$\Phi = \lim_{n \to \infty} \frac{1}{n} \log Z_n$$

assuming the limit exists. We will consider three methods for evaluating the free energy, the cavity method, interpolation and moments. We will begin with the cavity method which is a non-rigourous method from statistical mechanics but is can be made rigorous in some cases.

A sequence of graphs G_n is *locally treelike* if for all ℓ,

$$\frac{1}{n} \#\{v : B_\ell(v) \text{ is a tree}\} \to 1.$$

We will consider methods for evaluating the free energy of a d-regular locally treelike sequence of graphs which have the infinite d-regular tree as the local weak limit. The classic example is of such a sequence are the random d-regular graphs. Consider the following operation for removing two vertices of the graph. Start with the graph $G = G_n$, pick two vertices ρ, ρ' uniformly at random and remove them to form the graph G^- with $n-2$ vertices, $2d$ of which have degree $d-1$ (note we are removing two vertices in case d is odd). Choosing a uniform perfect matching of these vertices we add d edges to form G_{n-2}, a d-regular graph on $n-2$ vertices. If G_n is chosen from the configuration model, the so if G_{n-2}. If we only assume that G_n is locally treelike then the following lemma show that after many iterations of this operation, it remains locally treelike.

Lemma 9. *For all $\delta, \ell > 0$ and $0 < \alpha < 1$, there exists $\delta', \ell' > 0$ such*

$$\frac{1}{n}\#\{v \in G_n : B_{\ell'}(v) \text{ is a tree in } G_n\} \geq 1 - \delta'.$$

then for all $k \in [0, \frac{1-\alpha}{2}n]$ with G_{n-2k}

$$\frac{1}{n}\#\{v \in G_{n-2k} : B_{\ell}(v) \text{ is a tree in } G_{n-2k}\} \geq 1 - \delta.$$

Now suppose that the local weak limit of the graph G_n and a configuration σ from the spin system ψ was a Gibbs measure μ on the infinite d-regular tree given by an extremal TIFP with message m. If $u_1, \ldots, u_d$ are the neighbours of ρ then for large ℓ, then extremality means that the distribution of $\sigma_{\rho\cup\partial\rho}$ is almost independent of $\sigma_{S_\ell(\rho)}$ and

$$\mathbb{P}[\sigma_{\rho\cup\partial\rho} = x_{\rho\cup\partial\rho} \mid S_\ell(\rho)] \approx \frac{\psi(x_\rho)\prod_{i=1}^d (\psi(x_\rho, x_i)m(x_{u_i}))}{\sum_{x'\in\mathcal{X}}\psi(x'_\rho)\prod_{i=1}^d \left(\sum_{x'_{u_i}\in\mathcal{X}}\psi(x'_\rho, x_i)m(x_{u'_i})\right)}.$$

Then provided $d(\rho, \rho') > 2\ell + 2$,

$$\begin{aligned}
&\mathbb{P}[\sigma_{\rho\cup\partial\rho} = x_{\rho\cup\partial\rho}, \sigma_{\rho'\cup\partial\rho'} = x'_{\rho'\cup\partial\rho'}] \\
&\quad = \mathbb{E}\mathbb{P}[\sigma_{\rho\cup\partial\rho} = x_{\rho\cup\partial\rho}, \sigma_{\rho'\cup\partial\rho'} = x'_{\rho'\cup\partial\rho'} \mid S_\ell(\rho), S_\ell(\rho')] \\
&\quad = \mathbb{E}\mathbb{P}[\sigma_{\rho\cup\partial\rho} = x_{\rho\cup\partial\rho} \mid S_\ell(\rho)]\mathbb{P}[\sigma_{\rho'\cup\partial\rho'} = x'_{\rho'\cup\partial\rho'} \mid S_\ell(\rho')] \\
&\quad \approx \mathbb{P}[\sigma_{\rho\cup\partial\rho} = x_{\rho\cup\partial\rho}]\mathbb{P}[\sigma_{\rho'\cup\partial\rho'} = x'_{\rho'\cup\partial\rho'}]
\end{aligned}$$

where the second inequality used the Markov Random Field property. Thus the distribution in the local negibourhoods of ρ and ρ' are asymptotically independent and we can calculate the effect of the change in the partition function from G_n to G_{n-2}. First note that since

$$\mathbb{P}_{G_n}[\sigma_{\rho\cup\partial\rho} = x_{\rho\cup\partial\rho}] \propto \psi(x_\rho)\prod_{i=1}^d (\psi(x_\rho, x_i)m(x_{u_i}))$$

we have that

$$\mathbb{P}_{G^-}[\sigma_{\partial\rho} = x_{\partial\rho}] = \prod_{i=1}^d m(x_{u_i}).$$

Then we define the effect on the partition function of removing a vertex from a tree with messages m as

$$\Phi_{vertex}(m) := \log\sum_{x\in\mathcal{X}}\psi(x_\rho)\prod_{i=1}^d\left(\sum_{x_{u_i}\in\mathcal{X}}\psi(x_\rho, x_i)m(x_{u_i})\right)$$

Applying this to the neighbourhood of ρ and ρ' and the fact thatt they are approximately independent we have that

$$\log\frac{G_n}{Z^-} = 2\Phi_{vertex}(m) + o(1)$$

Similarly the effect of removing an edge,

$$\Phi_{edge}(m) := \log \sum_{x,x' \in \mathcal{X}} \psi(x,x')m(x)m(x'),$$

and so removing d edges from G_{n-2} to form G^- we have that

$$\log \frac{G_n}{Z^-} = d\Phi_{edge}(m) + o(1).$$

Together we get that

$$\log Z_{G_n} - \log Z_{G_{n-2}} = 2\Phi_{vertex} - d\Phi_{edge}.$$

Writing this as a telescoping sum, and using the fact that G_{n-2k} remain locally treelike,

$$\begin{aligned}\log Z_{G_n} &= \log Z_{\alpha n} + \sum_{k=1}^{\frac{1-\alpha}{2}n} \log Z_{G_{n-2(k-1)}} - \log Z_{G_{n-2k}} \\ &= \log Z_{\alpha n} + \frac{1-\alpha}{2} n \left(2\Phi_{vertex} - d\Phi_{edge}\right) + o(n).\end{aligned}$$

The partition function of $\log Z_{\alpha n}$ is $O(\alpha n)$ as it is a sum of $|\mathcal{X}|^{\alpha n}$ terms each of which has magnitude $e^{O(\alpha n)}$. Thus taking $\alpha \to 1$ the free energy is given by

$$\Phi = \lim_n \frac{1}{n} \log Z_{G_n} = \Phi_{vertex}(m) - \frac{d}{2}\Phi_{edge}(m). \tag{4.1}$$

When there is a unique Gibbs measure this formula determines the free energy. But what if there are multiple Gibbs measures, which one is the right one? In general there is no recipe to decide but in some cases such as the Ferromagnetic Ising model it is known.

4.1. **Ferromagnetic Ising Model Free Energy.** On the d-regular tree we saw that there was a unique Gibbs measure when $0 \le (d-1)\tanh\beta \le 1$ in which case the TIFP has message $m(x) \equiv \frac{1}{2}$. Then

$$\Phi_{vertex}(m) = \log \sum_{x \in \{-1,1\}} \prod_{i=1}^{d} \left(\sum_{x_{u_i} \in \{-1,1\}} \exp(\beta x_\rho x_i) \frac{1}{2} \right) = 2\cosh^d \beta,$$

and

$$\Phi_{edge}(m) = \log \sum_{x,x' \in \{-1,1\}} \exp(\beta x x') \frac{1}{2} = \cosh \beta,$$

the free energy is

$$\Phi = \log 2 + \frac{d}{2} \log \cosh \beta.$$

One can interpret this as $\log 2$ for the free energy if there were no interaction plus $\log \cosh \beta$ for each of the $\frac{dn}{2}$ edges of the graph.

We will consider the Ising model with a positive external field h as in this case we will be able to determine the local weak limit of σ directly. To evaluate the free energy when $h = 0$ we will then take a limit as $h \to 0$. An equivalent way to add an external field to a grpah is to add an addition vertex v^*, set a boundary condition of $+$ on v^* and connect each edge to v^* with inverse temperature $\beta_{u,v^*} = h$. We can then use the Edwards-Sokal coupling with the cluster of v^* set to plus.

Like the Ising mode, the FK-model satisfies monotonicity in its parameters so if $p, p' \in [0,1]^E$ and $p \leq p'$ then

$$\mu_p(\xi \in \cdot) \preceq \mu_{p'}(\xi \in \cdot),$$

which can be verified by the Glauber dynamics coupling argument similarly to Lemma 4. For some large ℓ, construct $\hat{G}$ by setting $p_{u,v^*} = 1$ for $v \in S_\ell$ which is equivalent to placing a plus boundary condition on S_ℓ. Also construct $\tilde(G)$ by removing the edges between S_ℓ and $S_{\ell+1}$ or equivalently setting p to 0 on these edges. Then

$$\mu_{\tilde{G}}(\xi \in \cdot) \preceq \mu_G(\xi \in \cdot) \preceq \mu_{\hat{G}}(\xi \in \cdot)$$

and so in effect we have sandwiched the measure on G between B_ℓ with free and plus boundary conditions. The latter will converge to the plus measure. If there were no external field the former would converge to the free measure but with a positive external field we will check that it instead converges to the plus measure. Let m^t be a BP message at a vertex t levels above the boundary to its parent. Then by the BP equations $m^t = f(m^{t-1})$ where

$$\begin{aligned} f(m) &= \frac{e^h(me^\beta + (1-m)e^{-\beta})^{d-1}}{e^h(me^\beta + (1-m)e^{-\beta})^{d-1} + e^{-h}(me^{-\beta} + (1-m)e^\beta)^{d-1}} \\ &= \frac{e^h(1 + 2(m-\frac{1}{2})\tanh\beta)^{d-1}}{e^h(1 + 2(m-\frac{1}{2})\tanh\beta)^{d-1} + e^{-h}(1 - 2(m-\frac{1}{2})\tanh\beta)^{d-1}} \end{aligned}$$

Starting from plus boundary condition corresponds to $m^0 = 1$ we will have $m^t \downarrow m+$ the largest fixed point of $f(m) = m$. Starting from free boundary conditions, $m^0 = \frac{1}{2}$, we have $m^t \uparrow m+$. So both $\mu_{\tilde{G}}$ and $\mu_{\hat{G}}$ converge to the plus measure around ρ. Since G is sandwiched between $\tilde{G}$ and $\hat{G}$ we have that

$$\mu_G(\sigma_\rho = +) = \frac{1}{2} + \frac{1}{2}\mu_G(\rho \leftrightarrow v^*) = \mu_+(\sigma_\rho = +) + o(1)$$

as $\ell \to \infty$. It follows that the local weak limit of μ_G is μ_+. As this is an extremal measure we can apply (4.1) and get that

$$\begin{aligned} \Phi_{\beta,h} = \log &\left((m_+e^\beta + (1-m_+)e^{-\beta})^d + (m_+e^{-\beta} + (1-m_+)e^\beta)^d \right) \\ &- \frac{d}{2}\log\left((m_+^2 + (1-m_+)^2)e^\beta + 2m_+(1-m_+)e^{-\beta} \right). \end{aligned}$$