Statistical Mechanics and Algorithms on Sparse and Random Graphs

Andrea Montanari[1]

October 8, 2014

[1]A. Montanari is with the Departments of Electrical Engineering and Statistics, Stanford University

Contents

Chapter 1

Main definitions and examples

The main object of these lectures is the study of specific classes of high-dimensional probability distributions. To be concrete, consider a random binary vector $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n)$ taking values in $\{0,1\}^n$. Throughout these lectures, we use boldface fonts to indicate random variables, and ordinary font for the deterministic values they can take.

Specifying the probability distribution of $\boldsymbol{\sigma}$ requires $2^n - 1$ real numbers, corresponding to the probabilities $\mathbb{P}(\boldsymbol{\sigma} = \sigma)$ for each $\sigma \in \{0,1\}^n$. In these lectures, we will be interested in the subset of probability distributions that can be written as products of pairwise factors, i.e. such that

$$\mathbb{P}(\boldsymbol{\sigma} = \sigma) = \prod_{i=1}^{n} \prod_{j=i+1}^{n} f_{ij}(\sigma_i, \sigma_j)\,, \tag{1.0.1}$$

for some functions $f_{ij} : \{0,1\} \times \{0,1\} \to \mathbb{R}_{\geq 0}$, and all $\sigma \in \{0,1\}^n$ Note that a probability distribution in this class is entirely specified by $O(n^2)$ numbers.

Indeed, we shall largely focus on an even more specific class of models. In particular, for most of these lectures we will assume that the factorization (1.0.1) with most of the functions $f_{ij}(\,\cdot\,,\,\cdot\,)$ being identically equal to one. It is convenient to associate a graph G to the factorization (1.0.1). The graph has vertex set $V = [n] \equiv \{1, 2, \dots, n\}$, and edge set E consisting of all the pairs (i,j) such that $f_{ij}(\,\cdot\,,\,\cdot\,) \neq 1$ (i.e. the function $f_{ij}(\,\cdot\,,\,\cdot\,)$ is not identically equal to one). We will be interested in probability distributions whose associated graph is *sparse*, and more precisely has only $O(n)$ edges. (But see Chapter 6 for examples beyond this setting.)

There are several reasons to study this specific class of high-dimensional probability distributions:

(a) They are *natural*. For the sake of comparison, consider proability distributions that can be written as products of singleton terms

$$\mathbb{P}(\boldsymbol{\sigma} = \sigma) = \prod_{i=1}^{n} f_i(\sigma_i)\,. \tag{1.0.2}$$

Under such a probability distribution, the random variables $\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_n$ are mutually independent. A factorized distribution as in (1.0.1) can be regarded as a natural generalization beyond complete independence. Indeed, classical theory [Lau96] elegantly connects the structure of the graph G to the conditional independence properties of $\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_n$.

(*b*) They form a *low-dimensional* family of probability distributions. As mentioned above, only $O(n^2)$ parameters are required to specify a probability distribution of the form (1.0.1). This simplifies modeling, and facilitates estimating the model parameters from data. As a consequence 'graphical models' of this type have been intensely used and studied in machine learning and statistics [KF09, HTF09]. They are both rich enough to cover a wide array of applications, and structured enough to simplify statistical inference.

(*c*) Probability distributions of the form (1.0.1) naturally arise in many applications. Historically, an important source of applications is provided by statistical physics. In many circumstances, physical systems can be modeled through discrete degrees of freedom (for instance atoms' magnetic moments) interacting locally. Physical laws often lead to pairwise interactions, and Boltzmann distribution leads to a joint distribution of the form (1.0.1).

Statistical physics is also at the origin of many important concepts, and examples in this field. More recently, theoretical computer science and information theory have provided a large number of interesting applications [MM09].

The rest of the chapter is organized as follows. In Section 1.1 we introduce a class of models that admit a pairwise factorization of the form (1.0.1). From a mathematical point of view, this class of models is rich enough to include many fascinating phenomena, and a host of challenging open problems. From the point of view of applications, the class of models defined is somewhat restrictive. We will overcome this problem by mentioning several generalizations in Section 1.4, and developing some of them in Chapter 6.

Section 1.2 introduces two basic questions that we will investigate in the following chapters. These questions are wide open, except for specific cases. A few concrete example are introduced 1.3. and generalizations are discussed in Section 1.4. Finally, as mentioned above, models of the type (1.0.1) have been studied across multiple research areas. In Section 1.5 we briefly discuss the relation between definitions developed by different communities.

1.1 Factor models

A *simple undirected graph* is a pair $G = (V, E)$ with V a set that is finite or countable (the *vertex set*) and E a set of unordered pairs in V (the *edge set*). Edges will be denoted by $e \equiv (i, j)$, with $i, j \in V$. All graphs will be assumed to be *locally finite*, i.e. each vertex will have finite degree (number of neighbors).

In some circumstances, it is technically convenient to consider *multigraphs* i.e. graphs that are also allowed to contain self-loops (i.e. edges of the form (i, i)) and multiple edges. As we will see (cf. for instance Section 2.1) differences are minimal, and we will only mention them when necessary.

Given a finite set $\mathcal{X}$, we let $\mathcal{X}^V$ be the set of vectors with entries in $\mathcal{X}$, indexed by V or equivalently the set of functions from V to $\mathcal{X}$. One such vector is denoted as $\sigma = (\sigma_i)_{i\in V}$. Thoughout these notes, a *specification*[1] is a pair of functions $\psi = (\psi_{\rm e}, \psi_{\rm v})$:

$$\psi_{\rm e} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}\,, \qquad \psi_{\rm v} : \mathcal{X} \to \mathbb{R}_{>0}\,, \tag{1.1.1}$$

[1]The term 'specification' is borrowed from the general theory of Gibbs measures [Geo11]. Here we use it in a more restrictive setting.

with $\psi_{\rm e}$ symmetric (i.e. $\psi_{\rm e}(\sigma,\tau) = \psi_{\rm e}(\tau,\sigma)$ for all $\sigma,\tau \in \mathcal{X}$). Given a pair (G,ψ) with V finite, we are interested in the probability measure $\mu_{G,\psi}$ on $\mathcal{X}^V$ defined by letting, for $\sigma \in \mathcal{X}^V$,

$$\mu_{G,\psi}(\sigma) = \frac{1}{Z(G,\psi)} \prod_{(i,j)\in E} \psi_{\rm e}(\sigma_i,\sigma_j) \prod_{i\in V} \psi_{\rm v}(\sigma_i)\,, \tag{1.1.2}$$

where the *partition function* is defined by the normalizing condition

$$Z(G,\psi) = \sum_{\sigma\in\mathcal{X}^V} \prod_{(i,j)\in E} \psi_{\rm e}(\sigma_i,\sigma_j) \prod_{i\in V} \psi_{\rm v}(\sigma_i)\,. \tag{1.1.3}$$

Notice that there is no loss of generality in assuming $\psi_{\rm v}(\tau) > 0$ for all $\tau \in \mathcal{X}$, because otherwise we can replace $\mathcal{X}$ by $\mathcal{X}' \equiv \{\tau \in \mathcal{X} :\ \psi_{\rm v}(\tau) > 0\}$. On the other hand, the probability distribution (1.1.2) is not as general as the factorized form (1.0.1).

We will refer to a probability distribution of the form (1.1.2) as to a *factor model*. Several alternative names can be found in different research domains, including: Markov random field, graphical model, Gibbs measure. Following the statistical mechanics terminology, we will sometimes call the variables σ_i spins, or spin values. We will discuss in Section 1.5 the difference in emphasis across different fields.

1.2 Large graph properties

In these lectures we are mainly interested in the properties of the probability measure $\mu_{G,\psi}$ when G is a large sparse graph. Roughly speaking, this means that the number of vertices $|V|$ is large and the number of edges $|E|$ is of the same order as $|V|$. (For a finite set S, we let $|S|$ denote its cardinality.) Classical statistical mechanics focuses on the case in which G is a d-dimensional grid, or a portion of the grid itself. Such graphs are sparse since they have bounded degree. By contrast, we shall focus on graphs G that are 'locally tree-like.'

Specifically, we consider sequences of graphs $\{G_n = (V_n, E_n)\}_{n\in\mathbb{N}}$ indexed by the number of vertices $n = |V_n|$. Informally, a graph sequence $\{G_n\}_{n\geq 0}$ is 'locally tree-like' if the neighborhoods of most vertices is a tree, and the empirical distribution of these neighborhoods has a well defined limit as $n \to \infty$.

In order to formalize these notions, given a graph G, we let $d_G : V \times V \to \mathbb{N}$ denote the *graph distance* in G, i.e. $d_G(i,j)$ is the length of the shortest path connecting i and j in G (with $d_G(i,j) = \infty$ if no such path exists). We denote by $B_t(i;G)$ the *ball of radius t centered at i*, i.e. the graph *induced*[2] by the subset of vertices j so that $d_G(i,j) \leq t$. We always regard $B_t(i;G)$ as a *rooted graph*, i.e. a graph with a distinguished vertices that is the center i. In the following we will often drop the argument G, whenever clear from the context.

The intuition of 'locally-tree like' graph sequence is made precise by the definition of local weak convergence, first introduced by Benjamini and Schramm in [BS01]. (As mentioned above, we will use boldface to denote random variables, whenever possible.)

Definition 1.2.1. *Let $\{G_n = (V_n, E_n)\}_{n\in\mathbb{N}}$ be a sequence of graphs with $|V_n| = n$ and, for each n, let $\boldsymbol{I}_n$ a uniformly random vertex in V_n. Let $\boldsymbol{T}$ be a random rooted tree and, for each $t \in \mathbb{N}$ denote by $\boldsymbol{T}_t$ the subtree formed by its first t generations.*

[2]Given a graph $G = (V,E)$, the graph induced by the subset of vertices $U \subseteq V$ is the graph formed by the vertices in U and the edges in E whose endpoints are both in U.

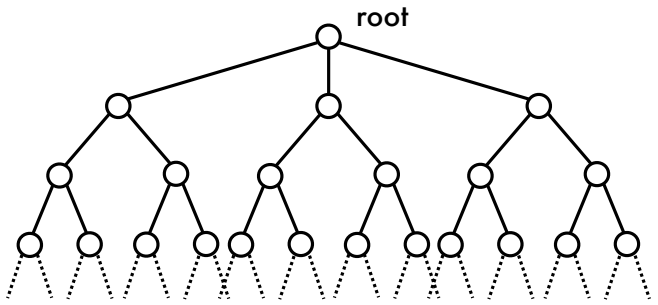

Figure 1.1: The regular rooted tree $T^{\mathrm{reg},k}$ of degree $k = 3$. Due to space limitations, only the first three generations are shown by solid lines.

We say that the sequence $\{G_n\}$ converges locally (weakly) *to* $\boldsymbol{T}$ *(and write* $G_n \xrightarrow{\mathrm{loc}} \boldsymbol{T}$*) if, for every rooted tree* T*,*

$$\lim_{n\to\infty} \mathbb{P}\big\{B_t(\boldsymbol{I}_n; G_n) \simeq T\big\} = \mathbb{P}\big\{\boldsymbol{T}_t \simeq T\big\}\,, \tag{1.2.1}$$

where $\simeq$ *denotes equivalence under root-preserving graph isomorphism.*

For completeness, we recall that, given two graphs $G_1 = (V_1, E_1)$ and $G_1 = (V_2, E_2)$, an isomorphism is a bijection $f : V_1 \to V_2$ such that $(i, j) \in E_1$ if and only if $(f(i), f(j)) \in E_2$. The isomorphism is *root preserving'* if it maps the root of G_1 to the root of G_2. It is clear that root-preserving isomorphisms define an equivalence relation. We refer to Chapter 2 for further background on graph convergence, random trees, and various examples of local weak convergence.

Let us emphasize that –in the definition 1.2.1– the graph sequence $\{G_n\}_{n\in\mathbb{N}}$ in non-random, and the only randomness is given by the root $\boldsymbol{I}_n$.

We limit ourselves to two simple examples.

Example 1.2.1. For $k \geq 2$, let $\{G_n\}$ be a sequence of simple graphs with uniform degree k, and diverging girth[3] $\lim_{n\to\infty} \mathrm{girth}(G_n) = \infty$. It is *a priori* not obvious that such a sequence exists. A non-constructive argument follows –for instance– from Theorem 2.12 in [Wor99]. Explicit constructions are discussed, among others, in [Big98].

For any vertex $i \in V_n$, and any $t < \lfloor \mathrm{girth}(G_n)/2 \rfloor$, $B_t(i; G_n)$ is a regular rooted tree and therefore $\{G_n\}$ converges locally in the sense of our definition. Namely, let $T^{\mathrm{reg},k}$ be the infinite deterministic rooted tree with uniform degree k, see Fig. 1.1, and $\boldsymbol{T}$ be the random tree such that $\boldsymbol{T} = T^{\mathrm{reg},k}$ with probabilty one. Then $G_n \xrightarrow{\mathrm{loc}} \boldsymbol{T}$.

Hereafter, we will write $G_n \xrightarrow{\mathrm{loc}} T$ when the limit is a deterministic tree T (e.g. in the present example $G_n \xrightarrow{\mathrm{loc}} T^{\mathrm{reg},k}$).

Example 1.2.2. An important motivation for studying locally tree-like graph is that the vast majority of sparse graphs are locally tree-like. Here is a simple example of this general remark (we refer to Chapter 2 for other examples with the same flavor).

Let $\boldsymbol{G}_n$ be a uniformly random regular graph with degree k. By this, we mean the following: there is a finite number of regular graphs of degree k (i.e. graphs with all

[3]The *girth* of a graph G is the length of the shortest loop in G.

vertices of the same degree) over n vertices. For each n, let $\boldsymbol{G}_n$ be independent random graph drawn uniformly at random in this set. (We refer to [Wor99] for a survey on random regular graphs.)

It is not hard to check that $\boldsymbol{G}_n$ contains a few short loop. For instance, $\boldsymbol{G}_n$ contains a triangle with probability bounded away from zero as $n \to \infty$ [Bol80, Wor81]. Nevertheless, with high probability, the neighborhoods of most vertices are trees. We indeed have that $\boldsymbol{G}_n \xrightarrow{\text{loc}} T_*^{\text{reg},k}$, almost surely with respect to the law of $\{\boldsymbol{G}_n\}$. (See [DM10, Section 2.1] for a detailed proof in a more general context.)

Given locally-tree like graph sequence $\{G_n\}$ and a specification ψ, we can consider the sequence of probability measures $\mu_{G_n,\psi}$ on $\mathcal{X}^{V_n}$ through the general equation (1.1.2). We are interested in the limit properties of these measures. There are various possible interpretations of this general research question, but here are two that we will consider extensively in the following chapters.

Convergence of the free energy. The *free energy density* for the pair (G_n, ψ) is defined as

$$\phi(G_n, \psi) \equiv \frac{1}{n} \log Z(G_n, \psi)\,. \tag{1.2.2}$$

We will often write ϕ_n for $\phi(G_n, \psi)$ when the arguments are clear from the context. The general problem is therefore:

> *Let G_n be a sequence of graphs that converge locally and ψ a specification. Does the free energy density have a limit?*
>
> $$\lim_{n\to\infty} \phi(G_n, \psi) \stackrel{?}{=} \phi\,. \tag{1.2.3}$$
>
> *Can we characterize or compute this limit?*

Local weak convergence. The next question is, roughly speaking, 'how does $\mu_{G_n,\psi}$ look locally?' This can be formalized –once again– through the notion of local weak convergence.

A *$\mathcal{X}$-marked graph* (or for simplicity, a *marked graph*) is a pair (H, σ) whose first element is a graph H with vertex set U, and whose second element is a vector $\sigma \in \mathcal{X}^U$ indexed by the vertices in U. Equivalently, σ is a function $\sigma : U \to \mathcal{X}$. We will sometimes write (H, σ_U) to emphasize the domain of σ. These definitions extend strightforwardly when H is a rooted graph. A random marked graph is a random variable $(\boldsymbol{H}, \boldsymbol{\sigma})$ taking values in the space of marked graphs. (We defer measure-theoretic technicalities to Chapter 2.)

Finally, if $\sigma \in \mathcal{X}^V$ and $U \subseteq V$, we let $\sigma_U = (\sigma_i)_{i\in U}$ be the subvector indexed by the vertices in U. With an abuse of notation, we write $\boldsymbol{\sigma}_{G_U}$ for the same vector, whereby G_U is the subgraph induced by U in G.

Definition 1.2.2. *Let $(\boldsymbol{T}, \boldsymbol{\sigma})$ be a random marked graph, with $\boldsymbol{T}$ a random rooted tree. Let $\{G_n\}$ be a sequence of graphs that converges locally to $\boldsymbol{T}$. Let ψ be a specification and for any $n \in \mathbb{N}$, let $\boldsymbol{\sigma}^n \sim \mu_{G_n,\psi}$. Finally, for each n, let $\boldsymbol{I}_n$ be uniformly random in V_n.*

We say that $(G_n, \mu_{G_n,\psi})$ converges locally *to $(\boldsymbol{T}, \boldsymbol{\sigma})$ if, for each (T, σ_T) with T a rooted tree*

$$\lim_{n\to\infty} \mathbb{P}\Big\{(B_t(\boldsymbol{I}_n; G_n), \boldsymbol{\sigma}^n_{B_t(\boldsymbol{I}_n;G_n)}) \simeq (T, \sigma_T)\Big\} = \mathbb{P}\Big\{(\boldsymbol{T}_t, \boldsymbol{\sigma}_{\boldsymbol{T}_t}) \simeq (T, \sigma_T)\Big\}\,. \tag{1.2.4}$$

Remark 1.2.3. Since we assume that G_n converges locally to $\boldsymbol{T}$, the definition above is a statement about the conditional distribution of $\boldsymbol{\sigma}^n_{B_t(i;G_n)})$ given $B_t(i;G_n)$.

Also, the limit $(\boldsymbol{T}, \boldsymbol{\sigma})$ allows to define the conditonal law of $\boldsymbol{\sigma}$ given $\boldsymbol{T}$ (as a regular conditional probability distribution). Denoting this conditional distribution by $\nu_{\boldsymbol{T}} = \nu_{\boldsymbol{\sigma}|\boldsymbol{T}}$, we will sometimes say that $\mu_{G_n,\psi}$ converges locally to $\nu_{\boldsymbol{T}}$ (and write $\mu_{G_n,\psi} \xrightarrow{\text{loc}} \nu_{\boldsymbol{T}}$) to express the fact that $(G_n, \mu_{G_n,\psi})$ converges locally to $(\boldsymbol{T}, \boldsymbol{\sigma})$

With this definition, we can formulate the second main problem that we will consider.

> *Let G_n be a sequence of graphs that converge locally and ψ a specification. Do the measures $\mu_{G_n,\psi}$ converge locally? Can we characterize or compute this limit?*

Remark 1.2.4. Both definitions 1.2.1 and 1.2.2 are special cases of the more general formulation of local weak convergence developed in [BS01, AL07]. In Chapter 2 revisit these notions from a more abstract point of view.

Remark 1.2.5. In plain english, Definition 1.2.2 corresponds to the following phenomenon. Draw a random configuration $\boldsymbol{\sigma}^n \sim \mu_{G_n,\psi}$ and a random vertex $\boldsymbol{I}_n$. The distribution of the resulting marked graph $(B_t(\boldsymbol{I}_n; G_n), \boldsymbol{\sigma}^n_{B_t(\boldsymbol{I}_n;G_n)})$ is then well approximated by the limiting formula.

This definition does not distinguish between the randomness produced by the random choice of $\boldsymbol{I}_n$, and the one due to the random choice of $\boldsymbol{\sigma}^n$. Refined definitions that capure this difference have been used in [MMS12, BD12]. Definition 1.2.2 corresponds to the notion of local convergence 'on average' in the terminology of [MMS12, BD12].

1.3 Examples and open problems

It is useful to keep in mind a few concrete examples of the general framework introduced in the previous sections. As mentioned above, randomized constructions provide simple way to define sequences of graphs that converge locally to trees.

Two special random graph models occur quite frequently because of their simplicity. The first one is the random k-regular graph, introduced in Example 1.2.2. In this case $\boldsymbol{G}_n$ is uniformly random among all graphs with n vertices and regular degree k. The second is the Erdös-Renyi random graph model, with average degree d. In this case every edge (i, j) is present independently with probability d/n. As we will see in Chapter 2, Erdös-Renyi random graphs converge locally to a rather simple tree model.

1.3.1 Ising models

In this case $\mathcal{X} = \{+1, -1\}$ and the specification is given by

$$\psi_{\mathrm{e}}(\sigma_1, \sigma_2) = e^{\beta\sigma_1\sigma_2}\,, \qquad \psi_{\mathrm{v}}(\sigma) = e^{B\sigma}\,, \tag{1.3.1}$$

with $\beta, B \in \mathbb{R}$ two parameters. The probability measure $\mu_{G_n,\psi} \equiv \mu_{G_n,\beta,B}$ can be written in exponential form as

$$\mu_{G_n,\beta,B}(\sigma) = \frac{1}{Z(G_n,\beta,B)} \exp\Big\{\beta \sum_{(i,j)\in E_n} \sigma_i\sigma_j + B\sum_{i\in V_n}\sigma_i\Big\}\,. \tag{1.3.2}$$

Ising models were introduced at the beginning of the twentieth century [Isi25] as models for magnetic materials. Following the statistical physics terminology, the parameter B is called 'magnetic field' and the parameter β 'inverse temperature.'

By symmetry, we can always restrict ourself to the case $B \geq 0$. The sign of β does instead play a crucial role. For $\beta > 0$, the measure $\mu_{G_n,\beta,B}$ promotes configurations σ such that $\sigma_i = \sigma_j$ for $(i,j) \in E$. This is known as the *ferromagnetic* Ising model. The opposite happens for $\beta < 0$, which promotes $\sigma_i = -\sigma_j$. This is known as the *antiferromagnetic Ising model*.

The difference between ferromagnetic and antiferromagnetic models has far-reaching consequences. For instance, it was proved in [DMS11] that, for essentially any sequence of graphs $\{G_n\}$ that converge locally to trees, the free energy density of ferromagnetic Ising models does converge. An expression for the asymptotic free energy density can be given in terms of the so-called 'Bethe free energy.'

The situation is dramatically different for antiferromagnetic models $\beta < 0$. First of all, the asymptotic behavior of the free energy density does not depend uniquely on the local weak convergence of the graph sequence $\{G_n\}$.

Proposition 1.3.1. *Consider the antiferromagnetic Ising model (1.3.2) with $\beta < 0$, and let $\phi(G_n, \beta, B) = (1/n) \log Z(G_n, \beta, B)$ be the corresponding free energy density.*

Then there exists $k_0 \in \mathbb{N}$ and, for each $k \geq k_0$ $\beta_*(k) < \infty$ such that the following happens. There exists a sequence of graphs $\{G_n\}$ such that $G_n \xrightarrow{\text{loc}} T_*^{\text{reg},k}$ but $\lim_{n\to\infty} \phi(G_n, \beta, 0)$ does not exist for any $k \geq k_0$, $\beta > \beta_*(k)$.

The proof of this statement uses the connection with optimization and can be found in Section 1.4.3.

This motivates the following general question.

Open Problem 1.3.1. Assume that the sequence $\{G_n\}$ converges locally to a random rooted tree $\boldsymbol{T}$. What are necessary and sufficient conditions on $\{G_n\}$ under which $\lim_{n\to\infty} \phi(G_n, \beta, B)$ exists for $\beta < 0$?

It is natural to expect that the limit exists if the graphs G_n are sufficiently 'homogeneous.' Indeed, it was proven in [BGT10] that, if $\boldsymbol{G}_n$ is a random regular graph of degree k, or an Erdös-Renyi random graph of average degree d, independent of n, then, for $\beta \leq 0$, $\phi(\boldsymbol{G}_n, \beta, B)$ has almost surely a deterministic limit as $n \to \infty$. Unfortunately, the proof does not provide the value of this limit.

Open Problem 1.3.2. Let $\{\boldsymbol{G}_n\}$ be a sequence of uniformly random k-regular graphs, or random Erdös-Renyi graphs with average degree d. Provide a characterization of $\lim_{n\to\infty} \phi(\boldsymbol{G}_n, \beta, B)$ for $\beta < 0$.

The question might appear a bit vague at this point, but there exists an exact conjecture for the limit value, based on the notion of 'replica symmetry breaking' (see Section 3.3).

The existence of a local limit for $\mu_{G_n,\beta,B}$ is an open problem as well for $\beta < 0$.

1.3.2 Independent sets

Given a graph $G = (V, E)$, an *independent set* of G is a subset of the vertices $S \subseteq V$ such that there is no edge $(i,j) \in E$ for which both $i \in S$, and $j \in S$.

Apart from being useful graph-theoretic objects, independent sets play an important role in theoretical computer science. The problem of computing the size of the largest

indepedent set S in a graph G is a prototypical *NP-complete problem* and has attracted considerable amount of work [Kar72, GJ79]. It is beyond the scope of these lectures to provide a background in complexity theory. In a nutshell, the fact that maximum independent set is NP-complete means the following. It is widely conjectured (the P $\neq$ NP conjecture) that no algorithm can do this computation in a time that is bounded by a polynomial in $|V|$. Indeed no polynomial algorithm is expected to exists to compute even a rough approximation of this quantity [Has96, DS05].

A number of problems related to independent sets inherit the same property, i.e. cannot be solved in polynomial time unless P = NP. In particular the problem of counting the number of independent sets of a graph. This is naturally generalized to the problem of computing the partition function

$$Z_{\rm IS}(G;\lambda) \equiv \sum_{S\in\mathcal{IS}(G)} \lambda^{|S|}\,, \tag{1.3.3}$$

where $\lambda > 0$ is a parameter and $\mathcal{IS}(G)$ is the collection of independent sets of G. For $\lambda = 1$, the quantity $Z_{\rm IS}(G;\lambda = 1)$ is the number of independent sets, while, as λ gets large, large independent dominate the sum.

It is easy to recognize that $Z_{\rm IS}(G;\lambda)$ is the partition function of a factor model. Indeed consider $\mathcal{X} = \{0,1\}$ and

$$\psi(\sigma_1,\sigma_2) = \mathbb{I}\big((\sigma_1,\sigma_2)\neq(1,1)\big)\,, \qquad \psi(\sigma) = \lambda^{\sigma}\,. \tag{1.3.4}$$

Then $Z(G,\psi) = Z_{\rm IS}(G,\lambda)$ (we will drop the subscript IS in the following). The resulting probability measure reads

$$\mu_{G,\lambda}(\sigma) = \frac{1}{Z(G,\lambda)}\prod_{(i,j)}\mathbb{I}\big((\sigma_i,\sigma_j)\neq(1,1)\big)\,\lambda^{\sum_{i\in V}\sigma_i}\,. \tag{1.3.5}$$

This can be interpreted as a probability distribution over independent sets $S = \{i\in V : \sigma_i = 1\}$, assigning probability proportional to $\lambda^{|S|}$ to independent set S.

In statistical physics, this is also known as the *hard core model*. It can be thought of as a model for a gas of particles whereby $\sigma_i = 1$ means that the site i is occupied by a particle. The term $\mathbb{I}\big((\sigma_i,\sigma_j)\neq(1,1)\big)$ conveys the fact that particles repel at short distance. The parameter λ is known as the gas 'fugacity' and gauges the particle density. As λ gets larger, a typical configuration $\boldsymbol{\sigma}\sim\mu_{G,\lambda}$ becomes denser (i.e. $\sum_{i\in V}\boldsymbol{\sigma}_i$ gets larger).

The hard core (or independent set) model shares many features of the antiferromagnetic Ising model, and indeed can be thought as a special limit of the latter. Considering for instance a graph with regular degree k, we set $\beta = -(2B+\log\lambda)/(2k)$ and let $B\to\infty$ in Eq. (1.3.2). Under the identification $0\leftrightarrow(+1)$ and $1\leftrightarrow(-1)$, the resulting model coincides indeed with the hard core model. This correspondence can be generalized to arbitrary graphs by allowng the magnetic field B_i of the Ising model to depend on the vertex degree.

In particular, the vaue of the limit free energy density is an open problem or the hard core model as well.

Open Problem 1.3.3. Let $\{\boldsymbol{G}_n\}$ be a sequence of uniformly random k-regular graphs, or random Erdös-Renyi graphs with average degree d. Let $\phi(\boldsymbol{G}_n,\lambda)$ be the free energy density of the hard-core (independent sets) model on G_n.

Prove a characterization (an explicit formula) for $\lim_{n\to\infty}\phi(\boldsymbol{G}_n,\lambda)$.

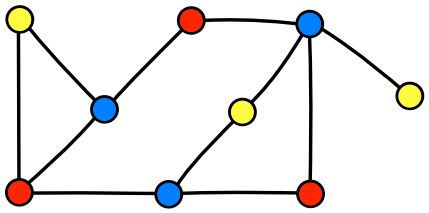

Figure 1.2: A finite graph G and a proper 3-coloring G.

1.3.3 Proper colorings

Given a graph $G = (V, E)$, and an integer $q \geq 2$, a proper q-coloring of G is an assigment q of colors to the vertices of G so that no edge has both endpoints of the same color, see Figure 1.2. Formally, a proper color assigment is given by $\sigma \in \mathcal{X}^V$, with $\mathcal{X} = [q] \equiv \{1, 2, \dots, q\}$, such that $\sigma_i \neq \sigma_j$ for all $(i, j) \in E$.

Given a graph G and $q \geq 3$, the problem of deciding whether G is q-colorable is NP-complete [Kar72, GJ79]. The case $q = 2$ is instead special: checking whether G is two-colorable amounts to checking whether G is bipartite, which can be done in time linear in the number of vertices (fix arbitrarily the color of one vertex, and propagate it).

The smallest integer q such that G is q colorable is known as the chromatic number of G, denoted by $\chi(G)$, and is a quantity of fundamental interest in graph theory. Another quantity of interest is the number of proper q-colorings of G, $Z(G; q)$ (which is of course non-zero only if $q \geq \chi(G)$). Viewed as a function of q, this is a polynomial of degree $|V|$, known as the chromatic polynomial of G. Chromatic polynomials have been studied for over a century [Big93].

Once again, $Z(G; q)$ can be viewed as the partition function of a factor model on the graph G with $\mathcal{X} = [q] \equiv \{1, 2, \dots, q\}$. Namely, the model is defined by letting

$$\psi_{\mathrm{e}}(\sigma_1, \sigma_2) = \mathbb{I}\Big(\sigma_1 \neq \sigma_2\Big)\,, \qquad \psi_{\mathrm{v}}(\sigma) = 1\,. \tag{1.3.6}$$

The free energy density $\lim_{n\to\infty} \phi(G_n, q)$ is, in this case, the exponential growth rate of the number of q-colorings of the sequence of graphs $\{G_n\}$. Except in special random graph models [BGT10], it is unknown whether this growth rate has a limit. Its value is known only for random regular graphs under special conditions on the degree k and the number of colors q [AN04]

1.4 Some generalizations

The class of models defined by Eq. (1.1.2) leaves out, by choice, many generalizations that are important, in particular for applications. We discuss some of these generalizations in this section and will revisit them in Chapter 6.

There are well-understood relations between the various classes of models that we will introduce. In particular, any model discussed in this section can be expressed as using a model on a labeled graph, introduced in Section 1.4.1. Such reductions are straightforward and are discussed, for instance, in [KF09].

Further, any progress on the more restricted class defined by Eq. (1.1.2) is likely to have an impact beyond it, to a much broader area.

1.4.1 Weighted directed graphs

A weighted directed graph is a triple $G = (V, E, \ell)$ with V a vertex set, D a set of directed edges (i.e. a set of ordered pairs $D \subseteq V \times V$) and ℓ a map $\ell : D \to \mathcal{L}$ that associates to each edge $(i, j) \in D$ a 'weighth' $\ell_{i,j} \in \mathcal{L}$. While in many examples we have $\mathcal{L} \subseteq \mathbb{R}$, more general choices are possible. (For the sake of simplicity, we avoid formal definitions here.)

A factor model can be be defined, by letting the edge interaction $\psi_{\rm e}$ depend on the edge weight as well

$$\psi_{\rm e} : \mathcal{X} \times \mathcal{X} \times \mathcal{L} \to \mathbb{R}_{\geq 0}\,, \tag{1.4.1}$$

where $\psi_{\rm e}$ is no longer required to be symmetric. For the sake of simplicity (and without loss of generality), we let $\psi_{\rm v}(\,\cdot\,) = 1$, and hence have

$$\mu_{G,\psi}(\sigma) = \frac{1}{Z(G;\psi)} \prod_{(i,j)\in D} \psi_{\rm e}(\sigma_i, \sigma_j; \ell_{ij})\,. \tag{1.4.2}$$

Let us emphasize that –as in the previous section– this is a probability distribution on $\sigma \in \mathcal{X}^V$ *given* G and –in particular– *given* the edge labels. In the language of physics, the labels ℓ_{ij} are 'quenched.'

As an example, the Ising spin glass (in zero magnetic field) is defined by $\mathcal{X} = \{+1, -1\}$, $\ell_{ij} \in \mathcal{L} = \mathbb{R}$ and letting

$$\psi_{\rm e}(\sigma_1, \sigma_2; \ell) = e^{\beta \ell \sigma_1 \sigma_2}\,. \tag{1.4.3}$$

For $\ell > 0$, this factor promotes configurations with $\sigma_1 = \sigma_2$ and, for $\ell < 0$ configurations with $\sigma_1 = \sigma_2$. The resulting probability distribution takes the exponential form

$$\mu_{G,\psi}(\sigma) = \frac{1}{Z(G,\psi)} \exp\Big\{\beta \sum_{(i,j)\in D} \ell_{ij}\sigma_i\sigma_j\Big\}\,. \tag{1.4.4}$$

Notice that, by allowing the cardinality of $\mathcal{L}$ to be larger than the number of edges, we can assign a different weigth for each $(i, j) \in D$, and ence a different function $\psi_{ij}(\sigma_i, \sigma_j) = \psi_{\rm e}(\sigma_i, \sigma_j; \ell_{ij})$ for each edge. Graph convergence can be easily generalized to weighted graphs, as it should become apparent from Chapter 2.

1.4.2 Hypergraphs and factor graphs

A second generalization of the factor model introduced above is obtained by replacing the pairwise terms in Eq. (1.1.2) by terms involving arbitrary subsets of the variables. The structure of such a factorization can be captured by an hypergraph, whereby edges are replaced by 'hyperedges' i.e. subsets including an arbitrary number of vertices.

Factor graphs can also be used to describe such generalized models. A factor graph is a triple $G = (V, F, E)$ whereby V is a set of *variable nodes*, F is a set of *factor nodes* and $E \subseteq V \times F$ is a set of edges connecting variable nodes to factor nodes. To each factor node $a \in F$, we associate its neighborhood $\partial a \equiv \{i \in V :\ (i, a) \in E\}$ and a *factor*

$$\psi_a : \mathcal{X}^{\partial a} \to \mathbb{R}_{\geq 0} \tag{1.4.5}$$

$$\sigma_{\partial a} \mapsto \psi_a(\sigma_{\partial a})\,. \tag{1.4.6}$$

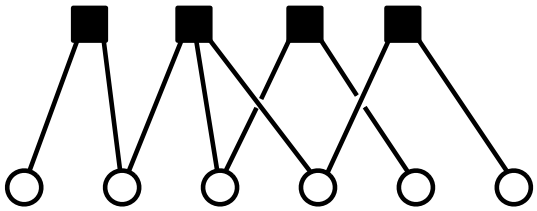

Figure 1.3: A factor graph $G = (V, F, E)$. Factor nodes are represented as filled squares and variable nodes as empty circles.

The probability distribution $\mu_{G,\psi}$ on $\mathcal{X}^V$ is defined by taking the product of these factors

$$\mu_{G,\psi}(\sigma) = \frac{1}{Z(G;\psi)} \prod_{a \in F} \psi_a(\sigma_{\partial a}) \,. \tag{1.4.7}$$

Note that we allowed ψ_a to depend on the specific factor node. Alternatively, we can introduced weigths ℓ_a in analogy to what done in the previous section.

The general pairwise model (1.4.2) is recovered by using factor nodes of degree 2, namely associating one factor node a with $\partial a = (i, j)$ for each $(i, j) \in D$. The inverse reduction (from factor graph to ordinary weightd graph) is also possible by properly enlarging the domain $\mathcal{X}$.

As an example, consider the k-satisfiability problem. An instance of this problem is a Boolean formula over n variables σ_1, σ_2, $\dots$ $\sigma_n \in \mathcal{X} \equiv \{\mathsf{True}, \mathsf{False}\}$. The formula takes the form of the conjunction (logical AND) of m clauses. Each clause is the disjunction (logical OR) of k of the variables or their negation. For instance, the following is a conjunction of $m = 4$ clauses, each being the disjunction of $k = 3$ clauses:

$$(\sigma_1 \vee \overline{\sigma_2} \vee \sigma_5) \wedge (\overline{\sigma_3} \vee \sigma_4 \vee \sigma_5) \wedge (\sigma_2 \vee \overline{\sigma_5} \vee \overline{\sigma_7}) \wedge (\sigma_5 \vee \sigma_6 \vee \sigma_7) \wedge (\overline{\sigma_4} \vee \overline{\sigma_5} \vee \sigma_6) \,. \tag{1.4.8}$$

(Here $\overline{\sigma_i}$ is the negation of σ_i.) The problem is to find an assignment of the variables $\sigma \in \{\mathsf{True}, \mathsf{False}\}^n$ for which the formula evaluates to True, i.e. for which each clause evaluates to True. Once more, this is a famously NP-hard problem [Coo71, GJ79].

To any k-satisfiability formula we can naturally associate a factor graph by having one variable node per each variable, and one factor node per each clause. Factor node a is connected by edges to the subset of variables that appear in clause a.

Equally of interest is the problem of counting the number of satisfying assignments. This can be regarded as the problem of computing the partition function of a factor graph model as per Eq. (1.4.7):

$$Z(G;\psi) = \sum_{\sigma \in \mathcal{X}^V} \prod_{a \in F} \psi_a(\sigma_{\partial a}) \,. \tag{1.4.9}$$

Here $\psi_a(\sigma_{\partial a})$ is a function that is equal to 1 if clause a is satisfied, and to 0 otherwise.

1.4.3 Connections with optimization

In the previous sections we defined several classes of 'graph-structured' probability distributions. In analogy, we can define classes of graph-structured combinatorial optimization

problems, that are directly connected to these probability distributions.

To be definite, given a finite domain $\mathcal{X}$, we introduce a pair of functions $\xi = (\xi_{\rm e}, \xi_{\rm v})$:

$$\xi_{\rm e} : \mathcal{X} \times \mathcal{X} \to [-\infty, \infty)\,, \qquad \xi_{\rm v} : \mathcal{X} \to (-\infty, \infty)\,. \tag{1.4.10}$$

We assume –for the sake of simplicity– $\xi_{\rm e}$ to be symmetric (i.e. $\xi_{\rm e}(\sigma_1, \sigma_2) = \xi_{\rm e}(\sigma_2, \sigma_1)$). Given a graph $G = (V, E)$, we can then define the optimization problem

$$u(G; \xi) = \frac{1}{|V|} \max_{\sigma \in \mathcal{X}^V} U_{G,\xi}(\sigma)\,, \tag{1.4.11}$$

$$U_{G,\xi}(\sigma) \equiv \sum_{(i,j)\in E} \xi_{\rm e}(\sigma_i, \sigma_j) + \sum_{i \in V} \xi_{\rm v}(\sigma_i)\,. \tag{1.4.12}$$

We will drop the subscript G, ξ whenever clear from the context.

A first connection with the probability distribution (1.1.2) is immediate. The problem of finding a mode[4] of the factor model (1.1.2) is equivalent to the prolem of finding a solution of the maximization problem (1.4.11), with the definitions $\xi_{\rm e}(\sigma_1, \sigma_2) = \log \psi_{\rm e}(\sigma_1, \sigma_2)$, $\xi_{\rm v}(\sigma) = \log \psi_{\rm v}(\sigma)$.

As an example, consider again the independent sets model introduced in Section 1.3.2, which has $\mathcal{X} = \{0, 1\}$. The corresponding optimization problem has objective function of the form (1.4.12) with

$$\xi_{\rm e}(\sigma_1, \sigma_2) = \begin{cases} -\infty & \text{if } \sigma_1 = \sigma_2 = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1.4.13}$$

$$\xi_{\rm v}(\sigma) = \sigma\, \log \lambda\,. \tag{1.4.14}$$

For $\log \lambda > 0$, the optimixation problem reduces therefore to finding a maximum size independent set:

$$\text{maximize} \quad \sum_{i \in V} \sigma_i\,, \tag{1.4.15}$$

$$\text{subject to} \quad \sigma \in \mathcal{IS}(G)\,, \tag{1.4.16}$$

where we remember that $\mathcal{IS}(G)$ is the collection of independent sets of G. (For $\log \lambda \le 0$ the optimization problem becomes trivial and is solved by setting $\sigma_i = 0$ for all $i \in V$.)

It is not hard to find relations between the optimal value $u(G; \xi)$ and the free energy density. Define, for $\beta \in \mathbb{R}$

$$\psi^\beta_{\rm e}(\sigma_1, \sigma_2) = e^{\beta \xi_{\rm e}(\sigma_1, \sigma_2)}\,, \qquad \psi^\beta_{\rm v}(\sigma) = e^{\beta \xi_{\rm e}(\sigma)}\,. \tag{1.4.17}$$

Lemma 1.4.1. *Let $\psi^\beta = (\psi^\beta_{\rm e}, \psi^\beta_{\rm v})$ be defined as per Eq. (1.4.17). For $\beta \ge 0$, we have*

$$\beta\, u(G_n; \xi) \le \phi(G_n; \psi^\beta) \le \beta\, u(G_n; \xi) + \log |\mathcal{X}|\,. \tag{1.4.18}$$

Proof. Recalling the definition of partition function, we have

$$\exp\Big\{\beta \max_{\sigma \in \mathcal{X}^V} U(\sigma)\Big\} \le Z(G_n; \psi^\beta) = \sum_{\sigma \in \mathcal{X}^V} \exp\Big\{\beta U(\sigma)\Big\} \le |\mathcal{X}|^n \exp\Big\{\beta \max_{\sigma \in \mathcal{X}^V} U(\sigma)\Big\}\,. \tag{1.4.19}$$

The claim follows by taking logarithms of this inequalities. □

[4] A mode is a configuration $\sigma \in \mathcal{X}^V$ that maximizes the probability $\mu_{G,\psi}(\sigma)$.

Proof of Proposition 1.3.1

First note that the specification of the antiferromagnetic Ising model is of the form (1.4.17) with $\xi_{\mathrm{e}}(\sigma_1, \sigma_2) = -\sigma_1\sigma_2$ and $\xi_{\mathrm{v}}(\sigma) = 0$. The corresponding objective function reads

$$U_{G,\xi}(\sigma) = -\sum_{(i,j)\in E} \sigma_i\sigma_j = -|E| + 2\sum_{(i,j)\in E} \mathbb{I}\big(\sigma_i \neq \sigma_j\big)\,, \tag{1.4.20}$$

This objective is maximized by partitioning the vertex set in two subsets $V_+ = \{i \in V : \sigma_i = +1\}$ and $V_- = \{i \in V : \sigma_i = -1\}$ as to maximize the number of edges across the partition.

We will construct two sequences of random graphs $\{\boldsymbol{G}_n^{(1)}\}_{n\geq 1}$ and $\{\boldsymbol{G}_n^{(2)}\}_{n\geq 1}$ such that, almost surely, $\boldsymbol{G}_n^{(1)} \xrightarrow{\mathrm{loc}} T^{\mathrm{reg},k}$, $\boldsymbol{G}_n^{(2)} \xrightarrow{\mathrm{loc}} T^{\mathrm{reg},k}$, and

$$\liminf_{n\to\infty} u(\boldsymbol{G}_n^{(1)}, \xi) > \limsup_{n\to\infty} u(\boldsymbol{G}_n^{(2)}, \xi)\,, \tag{1.4.21}$$

strictly. Assuming that this claim holds, it follows from Lemma 1.4.1 that there exists $\beta_0 > 0$ such that, for all $\beta \geq \beta_0$,

$$\liminf_{n\to\infty} \phi(\boldsymbol{G}_n^{(1)}, \psi) > \limsup_{n\to\infty} \phi(\boldsymbol{G}_n^{(2)}, \psi)\,. \tag{1.4.22}$$

The claim then follows by constructing aa sequence $\{\boldsymbol{G}_n\}_{n\geq 1}$ that alternates between the two above, e.g. by letting $\boldsymbol{G}_n = \boldsymbol{G}_n^{(1)}$ for n even and $\boldsymbol{G}_n = \boldsymbol{G}_n^{(2)}$ for n odd.

We are therefore left with the task of constructing the two $\{\boldsymbol{G}_n^{(1)}\}_{n\geq 1}$ and $\{\boldsymbol{G}_n^{(2)}\}_{n\geq 1}$ that satisfy the claim (1.4.21). We let $\boldsymbol{G}_n^{(1)}$ be a uniformly random regular bipartite graph of degree k (hence with $n/2$ vertices in each set of the partition). Notice that such a graph exists only for n even, but this is sufficient for our construction. As for the other sequence, we let $\boldsymbol{G}_n^{(2)}$ be a uniformly random regular graph of degree k (not necessarily bipartite). Again, for k odd such a graph exists only if n is even. If both k and n are odd, we can easily overcome this difficulty by constructing a uniformly random k-regular graph over vertices $\{1, \dots, n-1\}$ and leaving vertex n isolated.

It is an easy exercise to show that $\boldsymbol{G}_n^{(1)} \xrightarrow{\mathrm{loc}} T^{\mathrm{reg},k}$ and $\boldsymbol{G}_n^{(2)} \xrightarrow{\mathrm{loc}} T^{\mathrm{reg},k}$ (see Chapter 2 for the case of $\boldsymbol{G}_n^{(2)}$). Next consider the computation of $u(\boldsymbol{G}_n^{(\cdot)}, \xi)$. It is clear fom Eq. (1.4.20) that $U_{G,\xi}(\sigma) \leq |E|$. In the case of bipartite graph, this upper bound is achieved by setting $\sigma_i = +1$ for all vertices i in one set of the partition, and $\sigma_i = -1$ for all vertices i in the other set. We therefore have

$$\lim_{n\to\infty} u(\boldsymbol{G}_n^{(1)}, \xi) = \frac{k}{2}\,. \tag{1.4.23}$$

Consider next the case of a random regular graph $\boldsymbol{G}_n^{(2)}$ and denote by $\boldsymbol{A}_n$ the corresponging adjacency matrix (i.e. the $n \times n$ symmetric matrix whose entry (i, j) is one if $(i, j) \in E(\boldsymbol{G}_n)$ and 0 otherwise). Denoting by $\langle \cdot, \cdot \rangle$ the standard scalar product on $\mathbb{R}^n$, and by $\lambda_{\min}(\boldsymbol{A}_n)$ the minumum eigenvalue of $\boldsymbol{A}_n$, we then have

$$\begin{aligned} u(\boldsymbol{G}_n^{(1)}, \xi) &= -\frac{1}{2n}\,\min\Big\{\langle \sigma, \boldsymbol{A}_n\sigma\rangle :\ \sigma \in \{+1, -1\}^n\Big\} \\ &\leq -\frac{1}{2n}\,\min\Big\{\langle \sigma, \boldsymbol{A}_n\sigma\rangle :\ \sigma \in \mathbb{R}^n\,,\ \|\sigma\|_2^2 = n\Big\} \\ &= -\frac{1}{2}\,\lambda_{\min}(\boldsymbol{A}_n)\,. \end{aligned} \tag{1.4.24}$$

It follows from [FKS89] that there exists $c > 0$ such that $\lambda_{\min}(\boldsymbol{A}_n) > -c\sqrt{k}$ with probability at least $\exp(-n/c)$. We then have, almost surely,

$$\limsup_{n\to\infty} u(\boldsymbol{G}_n^{(1)}, \xi) \le c\sqrt{k}\,, \tag{1.4.25}$$

which completes the proof once we take k_0 so that $k_0/2 > c\sqrt{k_0}$.

1.5 About terminology

The study of 'graph-structured' probability distributions such as the one in Eq. (1.1.2) has a long history originating with statistical physics. Over the last forty years, the same research domain has become increasingly popular, first within probability theory and subsequently within computer science (artificial intelligence, computer vision, machine learning, communications and information theory, and so on). Many names have been introduced in different communities for closely related objects of interest, and it is useful to briefly discuss the differences.

A popular term in statistical physics and probability theory is *Gibbs measure* [Geo11]. This is often written in exponential form (Boltzmann formula), and interactions are not limited to be pairwise, e.g.

$$\mu(\sigma) = \frac{1}{Z}\,\exp\Big\{\sum_{U\in\mathcal{R}} \xi_U(\sigma_U)\Big\}\,, \tag{1.5.1}$$

where $\mathcal{R}$ is a collection of subsets of V. For finite graphs, this form is completely general. The emphasis of the classical theory of Gibbs measures is on the case of infinite graphs were the expression (1.5.1) does not make sense. The theory of Gibbs measures gives a way to generalize this to infinite graphs (see Chapter 3).

Within machine learning, artificial intelligence, computer vision, the name *graphical models* is used broadly to refer to graph-structured probability distributions [KF09]. These include the factor graph models of Section 1.4.2 as well as –for instance– *Bayes networks*. The names *undirected (pairwise) graphical models* and *Markov random field* comprise pairwise models of the form (1.4.2) or, equivalently,

$$\mu_{G,\psi}(\sigma) = \frac{1}{Z(G;\psi)} \prod_{(i,j)\in E} \psi_{ij}(\sigma_i, \sigma_j)\,. \tag{1.5.2}$$

The definition of undirected graphical models or Markov random field is slightly more general and include any probability distribution that is globally Markov with respoect to the graph G. If G has no triangle, and $\mu(\sigma) > 0$ strictly for each $\sigma \in \mathcal{X}^V$, the celebrated Hammersley-Clifford theorem implies that μ takes the form (1.5.2) for some choice of the ψ_{ij}.

Chapter 2

More on graph convergence

The objective of this chapter is threefold. First, we want to introduce some random graph and tree models that have been studied in detail, and provide basic examples of the local weak convergence phenomenon described in Definition 1.2.1. Second, we will reconsider the definition of local weak convergence from a more general point of view, thus filling in some missing mathematical concepts. For instance, in the previous chapter we omitted any discussion of what a 'random tree' is. While this point is trivial for trees over a given finite set of vertices, the question is more subtle for infinite trees. Third, we provide a short introduction to the notion of 'unimodularity'. This answers the natural question: which trees can be obtained as local weak limit of sequances of finite graphs, rooted uniformly at random? It is intuitively clear that the uniform rooting should lead to certain invariance properties: the limit tree should 'look the same' from all its vertices. Unimodularity formalizes this notion.

We regall that all graphs $G = (V, E)$ in these lectures are locally finite, which means that each vertex has a finite number of neighbors. For $i \in V$, we let ∂i denote the set of neighbors of i (i.e. $\partial i = \{j \in V : (i, j) \in E\}$), and $\deg(i) = |\partial i|$ its degree. Whenever useful to make the dependence on the underlying graph G explicit, we will write $\partial_G i$ or $\deg_G(i)$.

The chapter is organized as follows. We begin in Section 2.1 with some preliminary remarks that are useful in clarifying which properties of the graph G are relevant for our purposes. Section 2.2 formally defines the space of (infinite) trees as a measurable space, and introduces a few classical probability distributions over this space. Section 2.3 defines several random graph models, and discusses their local weak convergence properties. Finally, Section 2.4 provides a short introduction to unimodularity focusing on the case of trees.

There is a substantial literature on local weak convergence and unimodularity, and even more substantial on random graphs. Most of the material summarized here can be found in the review/expository papers [AS04, AL07]. Classical references on random graph theory are [Bol01, JLK00].

2.1 Preliminary remarks

It is a simple but important observation that many of the questions investigated in these lectures are insensitive to changes of a small number of edges of the graph G. This statement is somewhat vague but can be made formal in a number of cases.

To be definite consider the problem of computing the free energy density defined by Eq. (1.2.2). It turns out that the free energy density changes at most by an amount $O(1/n)$ if G changes by a bounded number of edges. This conclusion holds under suitable regularity conditions on the specification.

Definition 2.1.1. *A specification ψ is* strictly positive *if $\psi_{\rm e}(\sigma_1,\sigma_2)>0$ for all $\sigma_1,\sigma_2\in\mathcal{X}$.*

It is permissive *if there exist a permitted symbol $\sigma_0\in\mathcal{X}$ such that $\psi(\sigma,\sigma_0)>0$ for all $\sigma\in\mathcal{X}$.*

We then have the announced continuity property.

Lemma 2.1.2. *Let $G=(V,E)$ with $|V|=n$ and $G'=(V,E')$ with $E'=E\cup\{(u,v)\}$, for some $u,v\in V$, $(u,v)\notin E$. If ψ is strictly positive with $\psi_{\max}\equiv\max_{\sigma_1,\sigma_2}\psi_{\rm e}(\sigma_1,\sigma_2)$, $\psi_{\min}\equiv\min_{\sigma_1,\sigma_2}\psi_{\rm e}(\sigma_1,\sigma_2)$, then*

$$\frac{1}{n}\log\psi_{\min}\le\phi(G',\psi)-\phi(G,\psi)\le\frac{1}{n}\log\psi_{\max}\,. \tag{2.1.1}$$

If ψ is permissive with $\psi(\sigma_0)\ge\psi_{\min}$, and $\min_\sigma\psi(\sigma,\sigma_0)\ge\psi_{\min}$, then

$$\frac{1}{n}\log\Big\{\frac{\psi_{\min}}{|\mathcal{X}|}\Big(\frac{\psi_{\min}}{\psi_{\max}}\Big)^{\deg_G(u)+1}\Big\}\le\phi(G',\psi)-\phi(G,\psi)\le\frac{1}{n}\log\psi_{\max}\,. \tag{2.1.2}$$

Proof. By definition

$$Z(G',\psi)=\sum_{\sigma\in\mathcal{X}^V}\prod_{(i,j)\in E}\Big\{\psi_{\rm e}(\sigma_i,\sigma_j)\prod_{i\in V}\psi_{\rm v}(\sigma_i)\Big\}\psi_{\rm e}(\sigma_u\sigma_v)=Z(G,\psi)\;\mu_{G,\psi}\Big(\psi_{\rm e}(\boldsymbol{\sigma}_u,\boldsymbol{\sigma}_v)\Big)\,, \tag{2.1.3}$$

where $\mu_{G,\psi}\big(f(\boldsymbol{\sigma})\big)=\sum_{\sigma\in\mathcal{X}^V}f(\sigma)\,\mu_{G,\psi}(\sigma)$ denotes the expectation of $f(\boldsymbol{\sigma})$ with respect to the probability distribution $\mu_{G,\psi}$. Taking logarithms, we get

$$\phi(G',\psi)=\phi(G,\psi)+\frac{1}{n}\log\mu_{G,\psi}\Big(\psi_{\rm e}(\boldsymbol{\sigma}_u,\boldsymbol{\sigma}_v)\Big)\,, \tag{2.1.4}$$

which immediately implies the bounds (2.1.1) as well as the upper bound in (2.1.2).

For the lower bound in the permissive case, see Eq. (2.1.2), we note that

$$\mu_{G,\psi}\Big(\psi_{\rm e}(\boldsymbol{\sigma}_u,\boldsymbol{\sigma}_v)\Big)\ge\psi_{\min}\,\mu_{G,\psi}(\{\boldsymbol{\sigma}_u=\sigma_0\})\,. \tag{2.1.5}$$

Let $G_{\setminus u}$ be the graph G, with vertex u 'taken out,' i.e. $G_u=(V_u,E_u)$ with $V_u\equiv V\setminus\{u\}$ and $E_u\equiv E\setminus\{(u,w):w\in\partial u\}$. We then have the following identity, for any function $f:\mathcal{X}\to\mathbb{R}$

$$\mu_{G,\psi}\big(f(\boldsymbol{\sigma}_u)\big)=\frac{W(f;u)}{W(1;u)}\,, \tag{2.1.6}$$

$$W(f;u)\equiv\sum_{\sigma_u\in\mathcal{X}}f(\sigma_u)\psi_{\rm v}(\sigma_u)\,\mu_{G_u,\psi}\Big(\prod_{w\in\partial u}\psi_{\rm e}(\sigma_u,\boldsymbol{\sigma}_w)\Big)\,. \tag{2.1.7}$$

We have $W(1;u)\le|\mathcal{X}|\psi_{\max}^{\deg_G(u)+1}$ and, for ψ permissive and $f(\sigma)=\mathbb{I}(\sigma=\sigma_0)$, $W(f;u)\ge\psi_{\min}^{\deg_G(u)+1}$. Therefore

$$\mu_{G,\psi}(\{\boldsymbol{\sigma}_u=\sigma_0\})\ge\frac{1}{|\mathcal{X}|}\left(\frac{\psi_{\min}}{\psi_{\max}}\right)^{\deg_G(u)+1}\,. \tag{2.1.8}$$

whence the claim follows by substituting in Eq. (2.1.4) and (2.1.5). □

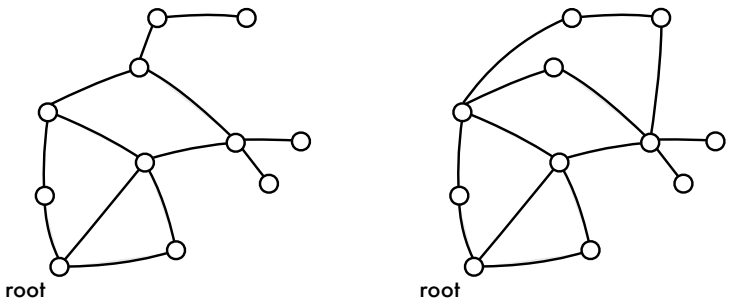

Figure 2.1: Two unlabeled rooted graphs $G_1 = (G_1, o_1)$, $G_2 = (G_2, o_2)$ with distance $d(G_1, G_2) = 1/8$. (Indeed $B_3(o_1; G_1) \not\simeq B_3(o_2; G_2)$, but $B_2(o_1; G_1) \simeq B_2(o_2; G_2)$.)

2.2 Trees

A tree is a graph that does not have loops. Throught this chapter, trees will be rooted, i.e. they have distinguished vertex called the root. We will denote the root by o. Formally, a rooted graph is a pair (G, o) with $G = (V, E)$ a graph and $o \in V$ the root. A rooted tree is such a pair when G is a tree. When clear from the context, we will omit explicit mention of the root and use the same symbol G for the rooted graph.

We are typically interested in properties of the rooted graph (G, o) that do not depend on the labeling of its vertices. We write $(G_1, o_1) \simeq (G_2, o_2)$ if the graphs G_1 and G_2 can be mapped to each other by a root preserving homomorphism (recall that this is a bijection between the vertices of G_1 and G_2 that maps edges onto edtjes bijectively, and o_1 to o_2). The set of equivalence classes of rooted graphs under the equivalne relation $\simeq$ will be denoted by $\mathcal{G}_*$. Its subset corresponding to trees will be denoted by $\mathcal{T}_*$. We refer to $\mathcal{G}_*$ and $\mathcal{T}_*$ as to the spaces of *unlabeled rooted graphs* or *trees*. We will denote elements in $\mathcal{G}_*$, $\mathcal{T}_*$ through their representative elements (G, o) or (T, o). Most of the times we will use the same notation $G = (G, o)$ for a graph and its equivalence class, and it should be clear from the context whether we are referring to one or the other. Whenever useful to be explicit, we will use $[G, o]$ for the equivalence class.

Recall that $B_t(o; G)$ is the ball of radius t around o in G. If $T = (T, o)$ is a rooted tree, we will use the shorthand $T_t = B_t(o; T)$ for the subtree formed by the first t-generations of T. (In particular, T_0 is the tree formed by a unique vertex, the root.) Note that, if $(G_1, o_1) \simeq (G_2, o_2)$, then $B_t(o_1; G_1) \simeq B_t(o_2; G_2)$. In other words, the map from a rooted graph to the ball of tadius t around its root, induces a well defined map on unlabeled trees. For G_1, $G_2 \in \mathcal{G}_*$, let

$$R(G_1, G_2) = \inf \left\{ t \in \mathbb{N} \ : \ B_t(o_1; G_1) \not\simeq B_t(o_2; G_2) \right\}. \tag{2.2.1}$$

(We adopt the usual convention that the infimum of an empty set is equal to $+\infty$.) In words $R(G_1, G_2)$ is the distance from the root of the first vertex where G_1 and G_2 differ. Again, $R(G_1, G_2)$ depends only on the equivalence classes of G_1 and G_2. Further $R(G_1, G_2) \geq 0$ and $R(G_1, G_2) = \infty$ if and only if $G_1 \simeq G_2$ We also define, for $G_1, G_2 \in \mathcal{G}_*$,

$$d(G_1, G_2) = \frac{1}{2^{R(G_1, G_2)}}. \tag{2.2.2}$$

Proposition 2.2.1. *The function d is a distance on $\mathcal{G}_*$, and $(\mathcal{G}_*, d)$ is a complete metric space. The same is true if $\mathcal{G}_*$ is replaced by $\mathcal{T}_* \subseteq \mathcal{G}_*$.*

Proof. It is obvious that d is symmetric and non-negative, with $d(G_1, G_2) = 0$ if and only if $G_1 = [G_1, o_1]$, $G_2 = [G_2, o_2] \in \mathcal{G}_*$ coincide, i.e. $G_1 \simeq G_2$. Further, for $G_1, G_2, G_3 \in UGraph_*$, we have $R(G_1, G_3) = \min(R(G_1, G_2), R(G_2, G_3))$, which implies the triangular inequality.

For proving completeness, let $\{G^{(n)}\}_{n\in\mathbb{N}}$ be a Cauchy sequence of unlabeled rootd graphs. For each $t \in \mathbb{N}$, let n_t be such that $d(G^{(n)}, G^{(n')}) \le 2^{-t-1}$ for all $n, n' \ge n_t$. In other words, all the trees $G^{(n)}$ in the sequence share the same first t generations for $n \ge n_t$. Note that this implies that $B_s(o; G^{(n_t)}) \simeq B_s(o; G^{(n_s)})$ for all $t \ge s$. As a consequence, there exist a graph $G^{(\infty)}$ such that $B_t(o; G^{(\infty)}) \simeq B_t(o; G^{(n_t)})$ for all t. It then follows that $\lim_{n\to\infty} G^{(n)} = G^{(\infty)}$ in $(\mathcal{G}_*, d)$, since for all t and all $n > n_t$, $d(G^{(n)}, G^{(\infty)}) \le 2^{-t}$. □

The space $(\mathcal{G}_*, d)$ is not compact but –as shown in the next exercise– compact subsets are obtained by bounding the degrees.

Exercise 2.2.1. Let $\Delta = (\Delta_i)_{i\in\mathbb{N}}$ be a sequence of integers. Denote by $\mathcal{BG}_*(\Delta) \subseteq \mathcal{G}_*$ the set of unlabeled rooted trees such that, for each t, the vertices at distance t from the root have degree at most Δ.

Prove that $\mathcal{BG}_*(\Delta)$ is is sequentially compact and hence compact.

The set $\mathcal{G}_*$ is made into a measurable space by equipping it with the Borel σ-algebra $\mathsf{B}(\mathcal{G}_*)$. A random unlabeled rooted graph (or tree) is simply a random variables with values in $(\mathcal{G}_*, \mathsf{B}(\mathcal{G}_*))$ (respectively, in $(\mathcal{T}_*, \mathsf{B}(\mathcal{T}_*))$). A probability distribution over rooted graphs (rooted trees) is a probability measure over the measurable space $(\mathcal{G}_*, \mathsf{B}(\mathcal{G}_*))$ (respectively, over $(\mathcal{T}_*, \mathsf{B}(\mathcal{T}_*))$).

The following definition is equivalent to the more explicit Definition 1.2.1.

Definition 2.2.2. *Let $\{G_n\}_{n\ge1}$ be a sequence of graphs and, for each n, let $\boldsymbol{I}_n$ a uniformly random vertex in V_n. Let $\boldsymbol{T} = (\boldsymbol{T}, o)$ be a random rooted tree, and let $(G_n, \boldsymbol{I}_n)$ be the unlabeled random rooted graph obtained by rooting G_n at $\boldsymbol{I}_n$.*

We say that the sequence $\{G_n\}_{n\ge0}$ converges locally (weakly) *to $\boldsymbol{T}$ (and write $G_n \xrightarrow{\text{loc}} \boldsymbol{T}$) if $(G_n, \boldsymbol{I}_n)$ converges in distribution to $\boldsymbol{T}$.*

All of the above theory generalizes when the vertices of G carry spin values. Namely, we consider triplets (G, σ, o) with $G = (V, E)$ a graph, $o \in V$ and $\sigma : V \to \mathcal{X}$, $i \mapsto \sigma_i$ an assignment of spin values to the vertices of G. Two such marked graphs are isomorphic (and we write $(G_1, \sigma_1, o_1) \simeq (G_2, \sigma_2, o_2)$) if there exists a bijection $\varphi : V_1 \to V_2$ such that: (i) $\varphi(o_1) = o_2$; (ii) $(i, j) \in E_1$ if and only if $(\varphi(i), \varphi(j)) \in E_2$; (iii) $\sigma_i = \sigma_{\varphi(i)}$ for all $i \in V_1$. The space of marked rooted graphs up to isomorphism will be denoted by $\mathcal{G}_*^{\mathcal{X}}$ and we typically refer to one of its elements by a representative (G, σ, o). When we want ot highlight that we are interested in the equivalence class, we wil write $[G, \sigma, o]$ instead. Then Definition 1.2.2 is equivalent to the following.

Definition 2.2.3. *Let $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ be an unlabeled marked random rooted tree. Let $\{G_n\}_{n\ge1}$ be a sequence of graphs that converges locally to $(\boldsymbol{T}, \boldsymbol{o})$. Let ψ be a specification and for any $n \in \mathbb{N}$, let $\boldsymbol{\sigma}^n \sim \mu_{G_n,\psi}$. Finally, for each n, let $\boldsymbol{I}_n$ be uniformly random in V_n.*

We say that $(G_n, \mu_{G_n,\psi})$ converges locally *to $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ if $(G_n, \boldsymbol{I}_n, \boldsymbol{\sigma}^n)$ converges in distribution to $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$.*

Two remarks are useful for working with unlabeled rooted trees.

First, for each $[G, \sigma, o] \in \mathcal{G}_*^{\mathcal{X}}$ we can pick a unique canonical representative (i.e. a labeled rooted graph (G, σ, o) in the equivalence class $[G, \sigma, o]$) with vertex set in $\mathbb{N}$ and root at 0. In other words, there is a canonical way to put labels on the vertices of an unlabeled graph.

Further, the function f that maps an unlabeled rooted graph to its representative is continuous, with respect to the following topologies. On the domain (the space of unlabeled graphs), we use the topology induced by the distance $d(\,\cdot\,,\,\cdot\,)$, cf. Eqs. (2.2.1) and (2.2.2). On the target space (the space of labeled graphs, we use the obvious generalization of the same distance, nemely the one obtained by replacing $R(G_1, G_2)$ by

$$R'(G_1, G_2) = \inf \Big\{ t \in \mathbb{N} \;:\; B_t(o_1; G_1) \neq B_t(o_2; G_2) \Big\}. \tag{2.2.3}$$

(Here $B_t(o.; G.)$ is understood to be a *labeled* graph and equality includes the labels.) Details of the construcyion of this mapping f can be found in [AL07].

Second, If wet let $\sigma^{\rm can} : \mathbb{N} \to \mathcal{X}$ be the resulting canonical spin configuration, there exists a regular conditional probability $\mathbb{P}_{\sigma^{\rm can}|[G,o]}$. (Note that $\sigma^{\rm can} \in \mathcal{X}^{\mathbb{N}}$, that is a complete separable metric space, and hence Borel-isomorphic to $([0,1], \mathsf{B}([0,1]))$.)

We conclude this section by defining a simple and useful random tree model. When considering a rooted tree T, the *descendants* or *offsprings* of a vertex $i \in V_T$ are all the neighbors of i that are farther from the root than i. The *ancestor* of i is instead the only neighbor of i that s closer to the root (the root has no ancestor).

We already introduced the regular tree $T^{{\rm reg},k}$. An important generalization is the *unimodular Galton-Watson tree* ${\rm GW}(P)$ that is parametrized by a probability distribution P over $\mathbb{N}$, with finite expectation. Let ρ be the size-biased version of P that is, for each $k \in \mathbb{N}$

$$\rho_k = \frac{k\, P_k}{\sum_{\ell=0}^{\infty} \ell P_\ell}\,. \tag{2.2.4}$$

A random tree $\boldsymbol{T} \sim {\rm GW}(P)$ is then generated by letting the number of descendant of the root be $K_o \sim P$, and the number of descendant of any other vertex v be an independent random variable $(K_v - 1)$, with $K_v \sim \rho$. Formally this is described by the following algorithm. (Here $[a, b]$ denotes the concatenation of a and b.)

Algorithm 1 Generate $\boldsymbol{T} = (V, E) \sim {\rm GW}(P)$

Initialize $V = \{o\}$, $E = \emptyset$
Draw $K_o \sim P$ and set $V_1 = \{1, 2, \dots, K_0\}$
Set $V = V \cup V_1$, $E = E \cup \{(o, v) :\; v \in V_1\}$ (connect the new vertices to the root);
for $t \in \{1, 2, \dots\}$ **do**
    for $v \in V_t$ **do**
        Draw $K_v \sim \rho$ independent of the past
        Set $V_{t+1} = V_{t+1} \cup \{[v, 1], \dots, [v, K_v - 1]\}$, $E = E \cup \{(v, [v, 1]), \dots, (v, [v, K_v - 1])\}$
    end for
    Set $V = V \cup V_{t+1}$
end for

Notice that –strictly speaking– the above describes a probability distribution over labeled trees. The random unlabeled tree $\boldsymbol{T} \sim {\rm GW}(P)$ is simply the equivalence class of the resulting labeled tree.

2.3 Random graph models and local convergence

In this section we introduce two standard random graph models and their local weak limits. We will indeed consider graph sequences $\{G_n = (V_n, E_n)\}_{n\geq 1}$ indexed by the number of vertices and will typically take the vertex set to be $V_n = [n]$ (the set of first n integers).

2.3.1 Erdös-Renyi random graphs

The simplest random graph model is arguably the *Erdös-Renyi random graph* with average degree d. This is often referred to as *the random graph* without further specifications. There is more than one definition for this model that are equivalent for our purposes[1]

1. $\boldsymbol{G}_n = (V(\boldsymbol{G}_n) = [n], E(\boldsymbol{G}_n))$ is a simple graph on n vertices, and the events $\{(i,j) \in E(\boldsymbol{G}_n)\}$ are mutually independent for all pairs (i,j), with $\mathbb{P}\{(i,j) \in E(\boldsymbol{G}_n)\} = d/n$.

2. $\boldsymbol{G}_n = (V(\boldsymbol{G}_n) = [n], E(\boldsymbol{G}_n))$ is a simple graph choosen uniformly at random among all graphs with n vertices and $|E(\boldsymbol{G}_n)| = dn/2$ edges.

3. $\boldsymbol{G}_n = (V(\boldsymbol{G}_n) = [n], E(\boldsymbol{G}_n))$ is a multi-graph, i.e. a graph with –potentially– multiple edges and self-loops. The graph has a deterministic number of edges $|E(\boldsymbol{G}_n)| = dn/2$, generated as follows. Start with n vertices and an empty edge set. For each $\ell \in \{1, \dots, |E(\boldsymbol{G}_n)|\}$ choose an edge (i_ℓ, j_ℓ) uniformly at random among the $\binom{n}{2}$ possible choices, and add the new edge to the graph: $E := E \cup \{(i_\ell, j_\ell)\}$.

4. The same as the previous model, except that edges are chosen by drawing i_ℓ, j_ℓ indipendent and uniformly random in $\{1, \dots, n\}$.

Within the first model, it is immediate to see that the degree of a fixed vertex, say of vertex v is a binomial random variable $\deg(v) \sim \mathrm{Binom}(d/n, n-1)$. As n gets large for $d = o(n)$, it is therefore approximately Poisson with mean d. The next result refines this observation.

Proposition 2.3.1. *For any $n \in \mathbb{N}$, let $\boldsymbol{G}_n$ be a random graphs from any of the models 1 to 4 above, with average degree d_n. Further, let $\boldsymbol{T} \sim \mathrm{GW}(\mathrm{Poisson}(d))$ be a unimodular Galton-Watson tree with degree distribution $P = \mathrm{Poisson}(d)$.*

If $\lim_{n\to\infty} d_n = d$, then $\boldsymbol{G}_n \xrightarrow{\mathrm{loc}} \boldsymbol{T}$, almost surely with respect to the distribution of $\{\boldsymbol{G}_n\}_n$.

Proof. This proof is somewhat tedious, and a standard result in the literature. We will present most of its details for the reader who might want develop some familiarity with the various concepts introduced in this chapter.

By the definition, the claim can be restated as follows. For any t-generations unlabeled rooted tree T_t, $\{\boldsymbol{G}_n\}$ almost surely

$$\lim_{n\to\infty} \mathbb{P}\Big(B_t(\boldsymbol{I}_n; \boldsymbol{G}_n) \simeq T_t \;\Big|\; \boldsymbol{G}_n\Big) = \mathbb{P}\big(\boldsymbol{T}_t \simeq T_t\big)\,. \tag{2.3.1}$$

We will limit ourself to proving this convergence in expectation, namely

$$\lim_{n\to\infty} \mathbb{P}\Big(B_t(\boldsymbol{I}_n; \boldsymbol{G}_n) \simeq T_t\Big) = \mathbb{P}\big(\boldsymbol{T}_t \simeq T_t\big)\,. \tag{2.3.2}$$

[1] We adopt here the somewhat unusual practice of parameterizing the model by the average degree, instead of the total number of edges, or the edge probability. The three quantities are easily related to each other.

The step from convergence in expectation to convergence almost sure follows, for instance, from [DM10, Lemma 2.3, 2.4].

Let us begin by computing the right-hand side of Eq. (2.3.2). Note that, in for the Poisson(d) distribution, we have

$$P_k = \rho_{k+1} = \frac{d^k}{k!}\, e^{-d}\,, \tag{2.3.3}$$

i.e. the number of offsprings of the root is distributed as the number of offsprings of any other vertex.

Let us identify T_t with a specific labeled representative. We will do this by labeling a vertex at distance s from the root with a vector $v \in \mathbb{N}^s$, with the convention that the descendants of v take labels $v' = [v, i]$ with $i \in \{1, 2, \dots\}$ following the canonical labeling. We will let $V_{=s}$ (respectively $V_{<s}$, $V_{>s}$) denote the subset of vertices of T_t at distance equal to s from the root (smaller than s, larger than s). We also denote by $\ell(v)$ the number of offsprings of vertex v (i.e. $\ell(v) = \deg_{T_t}(v)$ if v is the root, and $\ell(v) = \deg_G(v) - 1$ otherwise).

It is also useful to choose a labeled representative of the random tree $\boldsymbol{T}$. A specific labeling of the vertices (taking values in $\cup_{k\ge 1}\mathbb{N}^k$) is the one given by the algorithm in the previous section. With these definitions, we have

$$\mathbb{P}\big(\boldsymbol{T}_t = T_t\big) = \prod_{i\in V_{<t}} \Big(\frac{d^{\ell(i)}}{\ell(i)!}\, e^{-d}\Big)\,, \tag{2.3.4}$$

In order to compute $\mathbb{P}\Big(\boldsymbol{T}_t \simeq T_t\Big)$, we have to multiply the last probability by the number of distinct labeled trees T_t' that can be obtained by permuting the labels of siblings in T_t. This factor is given by

$$\frac{1}{|\mathrm{Aut}(T_t)|} \prod_{i\in V_{<t}} \ell(i)! \tag{2.3.5}$$

with $\mathrm{Aut}(T_t)$ the groop of (rooted) automorphisms of T_t. Multiplying these factors and grouping the terms, we thus obtain

$$\mathbb{P}\big(\boldsymbol{T}_t \simeq T_t\big) = \frac{1}{|\mathrm{Aut}(T_t)|}\, d^{|V_{\le t}|-1}\, e^{-d|V_{<t}|}\,. \tag{2.3.6}$$

Next consider the left hand side of Eq. (2.3.2). For the sake of simplicity, we will use model 1 above, but the calculation is similar in the other cases. Notice that, since the law of $\boldsymbol{G}_n$ is invariant under permutation of the vertices, we can replace $B_t(\boldsymbol{I}_n; \boldsymbol{G}_n)$ by $B_t(i; \boldsymbol{G}_n)$ for any fixed $i \in \mathbb{N}$, say $i = 1$. Fix a representative $T_t^{(n)}$ of T_t with vertex labels in $[n]$ and root label 1. Then we have

$$\mathbb{P}\Big(B_t(1;\boldsymbol{G}_n) = T_t^{(n)}\Big) = \Big(\frac{d}{n}\Big)^{|V_{\le t}|-1}\Big(1-\frac{d}{n}\Big)^{(|V_{\le t}|-1)(|V_{\le t}|-2)/2}\Big[\Big(1-\frac{d}{n}\Big)^{n-|V_{\le t}|+1}\Big]^{|V_{<t}|}\,. \tag{2.3.7}$$

This formula is somewhat lengthy but easy to understand. The first factor correspond the probability of all the edges in $T_t^{(n)}$ being present in $\boldsymbol{G}_n$. The second is the probability that all the edges among vertices $u, v \in V(T_t^{(n)})$ but such that $(u, v) \notin E(T_t^{(n)})$ are absent. The

third factor is the probability that each vertex in $T^{(n)}$ that has distance smaller than t from the root, has no neighbors in $V \setminus V(T_t^{(n)})$. The intersection of these three (independent) events is the event $B_t(1; \boldsymbol{G}_n) = T_t^{(n)}$.

Using the fact that $|V_{\le t}|$, $|V_{<t}$ are kept fixed as $n \to \infty$, a simple calculation yields the asymptotic formula

$$\mathbb{P}\Big(B_t(1; \boldsymbol{G}_n) = T_t^{(n)}\Big) = \Big(\frac{d}{n}\Big)^{|V_{\le t}|-1} e^{-d|V_{<t}|}\,. \tag{2.3.8}$$

In order to compute $\mathbb{P}(B_t(1; \boldsymbol{G}_n) \simeq T_t)$ we have to multiply the probability in Eq. (2.3.7) by the number of disctinct ways of labeling vertices of T_t by labels in $[n]$, keeping the root labeled by 1. This factor is

$$\frac{1}{|\mathrm{Aut}(T_t)|}\,(n-1)(n-2)\cdots(n-|V_{\le t}|+1) = \frac{n^{|V_{\le t}|-1}}{|\mathrm{Aut}(T_t)|}\,\big(1+O(n^{-1})\big)\,. \tag{2.3.9}$$

Multiplying this factor by the expresson in Eq. (2.3.8), we get

$$\lim_{n\to\infty} \mathbb{P}\Big(B_t(1; \boldsymbol{G}_n) \simeq T_t^{(n)}\Big) = \frac{1}{|\mathrm{Aut}(T_t)|}\, d^{|V_{\le t}|-1}\, e^{-d|V_{<t}|}\,, \tag{2.3.10}$$

which proves our claim. □

Indeed the graph models 1 to 4 are similar in a stronger sense than t implied by local weak convergence, as you are asked to prove below.

Exercise 2.3.1. Prove that the random graphs $\boldsymbol{G}^{(1)}$, $\dots \boldsymbol{G}^{(4)}$ defined in points 1 to 4 above can be coupled so that so that $\mathbb{E}\{|E(\boldsymbol{G}_n^{(a)}) \triangle E(\boldsymbol{G}_n^{(b)})|\} = O(d^2/n)$.

2.3.2 Graphs with given degree distribution

Given a graph $G = (V = [n], E)$ on n vertices, its *degree sequence* is the vector $d = (d_1, \dots, d_n) \in \mathbb{N}^n$ where, for each $i \in [n]$, $d_i = \deg_G(i)$. A *random graph with given degree sequence* $d \in \mathbb{N}^n$ is a graph $\boldsymbol{G}_n$ that is uniformly random among all simple graphs with degree sequence d.

Note that, for a general $d \in \mathbb{N}^n$, there does not always exists a simple graph with degree sequence d. (Thing for instance of $d = (4, 0, 0, 0)$.) If there exists at least one graph with degree sequence d, then d is called *graphical*. The characterization of graphical sequences is a classical result whose proof can be found, for instance, in [Bol78].

Theorem 1 (Erdös-Gallai)**.** *The degree sequence $d \in \mathbb{N}^n$ is graphical if and only if, assuming without loss of generality $d_1 \ge d_2 \ge \cdots \ge d_n$, the following conditions hold:*

(*a*) *The sum of degrees $\sum_{i=1}^n d_i$ is even.*

(*b*) *For each $k \in \{1, \dots, n\}$ we have $\sum_{i=1}^k d_i \le k(k-1) + \sum_{i=k+1}^n \min(k, d_i)$.*

Note that the situation is much simpler if G is allowed to be a multigraph, i.e. a graph with multiple edges and self loops. In that case, the first condition is necessary and sufficient (the degree is defined in this case as the number of edges incident on a vertex, with self loops counting two) . Namely, if $\sum_{i=1} d_i$ is even, then there exists a multigraph with degree sequence $d = (d_1, \dots, d_n)$.

Since we are interested in sequences of sparse graphs, it is more convenient to consider *random graphs with given degree distribution*, that can be defined as follows. Given a probability distribution P on $\mathbb{N}$, and $n \geq 1$, we define integers $m^{(n)} = (m_\ell^{(n)})_{\ell \geq 1}$, so that for each ℓ, $m_\ell^{(n)} \approx nP_\ell$ will be the number of vertices of degree ℓ.

An explicit construction is as follows. For $\ell \geq 1$, we let $\tilde{m}_\ell^{(n)} = \lfloor nP_\ell \rfloor$, If $\sum_{\ell=0}^{\infty} \ell \tilde{m}_\ell^{(n)}$ is even, then $m^{(n)} \equiv \tilde{m}^{(n)}$. Otherwise, letting $\ell_* \in \arg\max_{\ell \geq 1, \ell \text{ odd}} \tilde{m}_\ell^{(n)}$, we set $m_{\ell_*}^{(n)} = \tilde{m}_{\ell_*}^{(n)} - 1$ and $m_\ell^{(n)} = \tilde{m}_\ell^{(n)}$ for $\ell \neq \ell_*$. Finally, for $\ell = 0$, we let $m_0^{(n)} = n - \sum_{\ell=1}^{\infty} m_\ell^{(n)}$.

We then define a *random graph with n vertices and degree distribution P* as a graph $\boldsymbol{G}_n$ that is uniformly random among all simple graphs that have, for each $\ell \geq 0$, $m_\ell^{(n)}$ vertices of degree ℓ. This definition is well-posed as proved by the following proposition. Its proof can be found in appendix A.

Proposition 2.3.2. *For P a probabilty distribution over $\mathbb{N}$, let $m^{(n)}$ be defined as above. Define the degree sequence $d \in \mathbb{N}^n$ by letting $d_i = \ell$ for all i and ℓ such that $1 + \sum_{k=1}^{\ell} m_k^{(n)} \leq i \leq \sum_{k=1}^{\ell} m_k^{(n)}$.*

If P has finite expectation, then d is graphical for all n large enough.

While the random graph with given degree distribution is simple to define, this definition is rather implicit, and inconvenient for actual calculations. It is much more convenient to work with the so-called *configuration model* [Bol80, Wor81, Wor99]. This defines a probability distribution over multigraphs with n vertices and a given degree sequence $d = (d_1, \dots, d_n)$ such that $\sum_{i=1}^{n} d_i = 2|E|$ is even. A configuration model graph is generated as follows. To each vertex $i \in V = [n]$, associate d_i 'half-edges'. The resulting $2|E|$ half-edges are regarded as distinguishable. Then draw a uniformly random pairing π among $2|E|$ objects, and pair the half-edges using π, to form $|E|$ edges.

This construction defines a random graph *with given degree sequence*. Using the approach described above for uniform random graphs, we can define a configuration model random graph *with given degree distribution P*.

Note that, in general, a configuration-model random graph may contain multiple edges and self loops, i.e. it is in fact a multi-graph. The next result provides a simple connection between the two.

Proposition 2.3.3. *Let $\boldsymbol{G}_n$ be a configuration-model random graph with given degree sequence $d = (d_1, \dots, d_n)$, and denote by $\mathcal{S}_n$ the set of simple graph over n vertices. Then, conditional on $\boldsymbol{G}_n \in \mathcal{S}_n$, $\boldsymbol{G}_n$ is a uniformly random simple graph with degree sequence d.*

Further, consider a sequence of models over n vertices, with degree sequences $d^{(n)} = (d_1^{(n)}, d_2^{(n)}, \dots, d_n^{(n)})$ such that $2|E(\boldsymbol{G}_n)| = \sum_{i=1}^{n} d_i^{(n)} \to \infty$. If $\sum_{i=1}^{n} (d_i^{(n)})^2 = O(\sum_{i=1}^{n} d_i^{(n)})$, then

$$\liminf_{n \to \infty} \mathbb{P}(\boldsymbol{G}_n \in \mathcal{S}_n) > 0\,. \tag{2.3.11}$$

Proof. The first part is straightforward, and analogous to what is done in [Bol80, Wor81, Wor99] for random regular graphs. The second is adapted from [Jan09]. □

In particular, for a sequence of configuration model graphs with given degree distribution P, $\mathbb{P}(\boldsymbol{G}_n \in \mathcal{S}_n)$ remains bounded away from zero as long as P has bounded second moment.

Proposition 2.3.3 is particularly useful for estabilishing typical properties of the uniform random graph. Namely, under the stated assumptions, it implies that, if a certain property

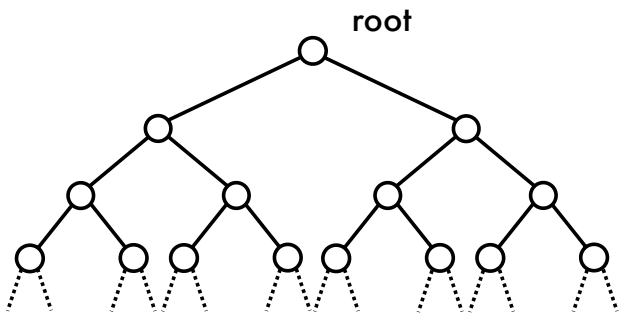

Figure 2.2: The determistic rooted tree $T^{\mathrm{reg},2,3}$, i.e.the infinite tree with degree 2 at the root and degree 3. This tree cannot be obtained as local weak limit of a sequence of randomly rooted graphs.

holds with high probability with respect to the configuration model, it must hold with high probability with respect to the uniform model as well.

Proposition 2.3.4. *Let P be a probability distribution over $\mathbb{N}$ with bounded first moment. For any $n \in \mathbb{N}$, let $\boldsymbol{G}_n$ be a random graphs with degree distribution P, either uniformly random or from the configuration model. Further, let $\boldsymbol{T} \sim \mathrm{GW}(P)$ be a unimodular Galton-Watson tree with degree distribution P.*

Then $\boldsymbol{G}_n \xrightarrow{\mathrm{loc}} \boldsymbol{T}$, almost surely with respect to the distribution of $\{\boldsymbol{G}_n\}$.

Proof. The proof can be found in [DM10] □

2.4 Unimodularity

As we saw, random rooted trees emerge quite naturally as local limits of sequences of graphs rooted at a random vertex. Is is intuitively clear that not any rooted tree can be obtained as such limits.

This point can be illustrated through a simple example. Consider the deterministic rooted tree $T^{\mathrm{reg},k,m}$ that has degree k at the root, and $m \neq k$ at all vertices except the root, cf. Fig. 2.2. It seems unlikely that such tree can be obtained as local limit of a randomly rooted graph sequence. One would expect that such a limit tree should 'look the same' when regarded from any of its vertices. On the contrary, the root of $T^{\mathrm{reg},k,m}$ is 'special.'

This concept is made precise by the the notion of unimodular random rooted graph. In order to define unimodularity, we must first define the space $\mathcal{G}_{**}$ of doubly rooted graphs. A *doubly rooted graph* is a triple (G, u, v) where $G = (V, E)$ is a graph, and $u, v \in V$. Two such graphs (G_1, u_1, v_1) and (G_2, u_2, v_2) are isomorphic if there exists an isomorphism φ between G_1 and G_2 (i.e. a bijection of the vertices that preserves edges), such that $\varphi(u_1) = u_2$ and $\varphi(v_1) = v_2$. We let $\mathcal{G}_{**}$ be the space of equivalence classes of doubly rooted graphs, and will refer to it as to the space of *unlabeled doubly rooted graphs.*

The space $\mathcal{G}_{**}$ can be given a metric by analogously to what we did for $\mathcal{G}_*$. Given $t \in \mathbb{N}$, $G = (V, E)$ and $u, v \in G$, we let $B_t(u, v; G)$ be the subgraph induced by all the vertices $j \in V$ such that $\min(d(u, j), d(v, j)) \leq t$, rooted at the pair u, v. Given two doubly rooted graphs $G_1 = (G_1, u_1, v_1)$, $G_2 = (G_2, u_2, v_2)$, we let $d(G_1, G_2) = 2^{-R(G_1,G_2)}$ with $R(G_1, G_2)$

equal to the smallest t such that $B_t(u_1, v_1; G_1) \not\simeq B_t(u_2, v_2; G_2)$. As for the simply rooted case, $(\mathcal{G}_{**}, d)$ is a complete metric space, we regard it as a measurable space endowed with the Borel σ-algebra.

Exercise 2.4.1. Let $\pi : \mathcal{G}_{**} \to \mathcal{G}_*$ be the mapping that consists in removing the mark from the second root in a doubly rooted graph, i.e. $\pi : (G, u, v) \mapsto (G, u)$. Show that π is continuous.

In the following, we need to consider functions $f : \mathcal{G}_{**} \to \mathbb{R}$ defined on unlabeled doubly rooted graphs. Note that such a function can be identified with a function on labeled rooted graphs such that $f(G_1, u_1, v_1) = f(G_2, u_2, v_2)$ every time $(G_1, u_1, v_1) \simeq (G_2, u_2, v_2)$. This identification will be understood throughout.

Definition 2.4.1. *Let $(\boldsymbol{G}, \boldsymbol{o})$ be an unlabeled random rooted graph. We say that $(\boldsymbol{G}, \boldsymbol{o})$ is unimodular if, for any measurable function $f : \mathcal{G}_{**} \to \mathbb{R}_{\geq 0}$, we have*

$$\mathbb{E}\Big[\sum_{v \in V(\boldsymbol{G})} f(\boldsymbol{G}, \boldsymbol{o}, v) \Big] = \mathbb{E}\Big[\sum_{v \in V(\boldsymbol{G})} f(\boldsymbol{G}, v, \boldsymbol{o}) \Big] . \tag{2.4.1}$$

Equation (2.4.1) is known as the *mass-transportation principle*. It is a good exercise to parse –say– its right hand side. We sample an unlabeled rooted graph $[\boldsymbol{G}, \boldsymbol{o}]$ according to the estabilished probability distribution. Construct a representative labeled rooted graph $(\boldsymbol{G}, o)$ (for instance the canonical one, with $V(\boldsymbol{G}) = \mathbb{N}$). Evaluate $f(\boldsymbol{G}, \boldsymbol{o}, v)$ for each $v \in V(\boldsymbol{G})$. Notice that for two representatives $(\boldsymbol{G}_1, \boldsymbol{o}_1)$ and $(\boldsymbol{G}_2, \boldsymbol{o}_2)$, the result of this step will be different, but still $f(\boldsymbol{G}_1, \boldsymbol{o}_1, v_1) = f(\boldsymbol{G}_2, \boldsymbol{o}_2, \varphi(v_1))$ if φ is the isomorphism between $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$. We sum this quantity over v: by what we just said, the result depends only on the equivalence class $[\boldsymbol{G}, \boldsymbol{o}]$ and not on the vertex labeling.

The importance of unimodular graphs is clarified by the following result due to Benjamini and Schramm [BS01].

Proposition 2.4.2. *Let $\{G_n\}$ be a sequence of graphs that converges locally to the random rooted graph $(\boldsymbol{G}, \boldsymbol{o})$. Then $\boldsymbol{G}$ is unimodular.*

Proof. The proof proceeds in two steps: (i) Show that, for G_n a finite graph and $\boldsymbol{I}_n$ a uniformly random vertex, the random rooted graph $(G_n, \boldsymbol{I}_n)$ is unimodular; (ii) Prove that the local weak limit of unimodular graphs is also unimodular.

In order to accomplish step (i), we evaluate the left hand side of Eq. (2.4.1):

$$\begin{aligned} \mathbb{E}\{ \sum_{v \in V_n} f(G_n, \boldsymbol{I}_n, v) \Big\} &= \frac{1}{n} \sum_{i \in V_n} \sum_{v \in V_n} f(G_n, i, v) \\ &= \frac{1}{n} \sum_{i \in V_n} \sum_{v \in V_n} f(G_n, v, i) = \mathbb{E}\{ \sum_{v \in V_n} f(G_n, v, \boldsymbol{I}_n) \Big\} . \end{aligned} \tag{2.4.2}$$

Step (ii) is a measure-theory exercise and its details can be found in Appendix A.1.2. □

In fact it turns out that a sligntly simpler condition (introduced by Aldous and Steele [AS04]) is sufficient for unimodularity.

Definition 2.4.3. *Let $(\boldsymbol{G}, \boldsymbol{o})$ be an unlabeled random rooted graph. We say that $(\boldsymbol{G}, \boldsymbol{o})$ is* involution invariant *if, for any measurable function $f : \mathcal{G}_{**} \to \mathbb{R}_{\geq 0}$, we have*

$$\mathbb{E}\Big[\sum_{v \in \partial \boldsymbol{o}} f(\boldsymbol{G}, \boldsymbol{o}, v) \Big] = \mathbb{E}\Big[\sum_{v \in \partial \boldsymbol{o}} f(\boldsymbol{G}, v, \boldsymbol{o}) \Big] . \tag{2.4.3}$$

In other words, $\boldsymbol{G}$ is involution-invariant if it satisfies the mass-transportation principle for $f : \mathcal{G}_{**} \to \mathbb{R}_{\geq 0}$ such that $f(G, u, v) = 0$ unless $(u, v) \in E(G)$.

Proposition 2.4.4. *A random rooted graph $(\boldsymbol{G}, \boldsymbol{o})$ is unimodular if and only if it is involution-invariant.*

Proof. The proof can be found in [AL07, Proposition 2.2]. □

In the following, given a doubly rooted graph (G, u, v), we will rather write $(G, u \to v)$ if $(u, v) \in E(G)$.

Determining whether unimodularity characterizes completely the set of weak limits is an open problem.

Open Problem 2.4.1. Prove that for any $(\boldsymbol{G}, \boldsymbol{o})$ unimodular, there exists a sequence of graphs $\{G_n\}$ that converges locally to $\boldsymbol{G}$.

On the other hand, the problem is settled for the case of unimodular trees, as shown in [EL10, BLS12].

Theorem 2. *Let $(\boldsymbol{T}, \boldsymbol{o})$ be a unimodular random rooted tree. Then there exists a siquence of finite graphs $\{G_n = (V_n, E_n)\}_{n \geq 1}$ that converge locally to $(\boldsymbol{T}, \boldsymbol{o})$.*

Exercise 2.4.2. Prove that the deterministic regular tree $T^{\mathrm{reg},k}$ is unimodular.

Exercise 2.4.3. Let (G, o) be a deterministic rooted unlabeled graph (i.e. a random rooted unlabeled graph that is equal to (G, o) with probability one). Prove that G must be a regular graph.

Prove that the same conclusion holds for a random graph $(\boldsymbol{G}, \boldsymbol{o})$ if $\deg(\boldsymbol{o})$ is deterministic.

The following result provides an equivalent definition of unimodularity.

Proposition 2.4.5. *The unlabeled random rooted graph $(\boldsymbol{G}, \boldsymbol{o})$ is unimodular if and ony if, for any measurable function $f : \mathcal{G}_{**} \to \mathbb{R}_{\geq 0}$, we have*

$$\mathbb{E}\big\{ \deg(\boldsymbol{o}) f(\boldsymbol{G}, \boldsymbol{o} \to \boldsymbol{J}) \big\} = \mathbb{E}\big\{ \deg(\boldsymbol{o}) f(\boldsymbol{G}, \boldsymbol{J} \to \boldsymbol{o}) \big\} , \tag{2.4.4}$$

where $\boldsymbol{J}$ is a uniformly random neighbor of the root.

Proof. This is just a restatement of the condition for involution invariance, cf. Eq. (2.4.3). For instance, the left hand side of that equation reads

$$\mathbb{E}\Big[\sum_{v \in \partial \boldsymbol{o}} f(\boldsymbol{G}, \boldsymbol{o}, v) \Big] = \mathbb{E}\Big[\deg(\boldsymbol{o}) \Big(\frac{1}{\deg(\boldsymbol{o})} \sum_{v \in \partial \boldsymbol{o}} f(\boldsymbol{G}, \boldsymbol{o}, v) \Big) \Big] = \mathbb{E}\Big[\deg(\boldsymbol{o})\, f(\boldsymbol{G}, \boldsymbol{o}, \boldsymbol{J}) \Big] , \tag{2.4.5}$$

with $\boldsymbol{J}$ as in the statement. The right hand side can be written in a similar form. □

There is another definition that will be useful in the following. We let $\mathcal{G}_{\rm e} \subseteq \mathcal{G}_{**}$ be the subset of doubly rooted graphs (G, u, v) such that $(u, v) \in E$. We write $(G, u \to v)$ for an element of $\mathcal{G}_{\rm e}$, and will denote by $\mathcal{T}_{\rm e} \subseteq \mathcal{G}_{\rm e}$ the subset for which $G = T$ is a tree.

If $(\boldsymbol{G}, \boldsymbol{o})$ is a unimodular random rooted graph, with law ν, we let $\tilde{\nu}^{\uparrow}$ denote the law of $(\boldsymbol{G}, \boldsymbol{J} \to \boldsymbol{o})$ for $\boldsymbol{J}$ a uniformly random neighbor of $\boldsymbol{o}$, and $\tilde{\nu}^{\downarrow}$ denote the law of $(\boldsymbol{G}, \boldsymbol{o} \to \boldsymbol{J})$. Notice that both $\tilde{\nu}^{\uparrow}$ and $\tilde{\nu}^{\downarrow}$ are supported on $\mathcal{G}_{\rm e}$. We further define[2] $\nu^{\uparrow}$ and $\nu^{\downarrow}$ by the Radon-Nykodim derivatives

$$\frac{{\rm d}\nu^{\downarrow}}{{\rm d}\tilde{\nu}^{\downarrow}}(G, o \to j) = \frac{\deg(o)}{\mathbb{E}\deg(\boldsymbol{o})}\,, \qquad \frac{{\rm d}\nu^{\uparrow}}{{\rm d}\tilde{\nu}^{\uparrow}}(G, j \to o) = \frac{\deg(o)}{\mathbb{E}\deg(\boldsymbol{o})}\,. \tag{2.4.6}$$

Here expectation is with respect to ν, or $\tilde{\nu}^{\downarrow}$, or $\tilde{\nu}^{\uparrow}$: the three quantities are equal. Also, we assume without substantial loss of generality that $\deg(\boldsymbol{o}) > 0$ with strictly positive probability (in the opposite case $\boldsymbol{G}$ is formed almost surely by a unique vertex.) We can then restate the last Proposition 2.4.5 as follows.

Proposition 2.4.6. *The random rooted graph $(\boldsymbol{G}, \boldsymbol{o})$ with law ν is unimodular if and only if*

$$\nu^{\downarrow} = \nu^{\uparrow}\,. \tag{2.4.7}$$

Note that, for ν unimodular, the five measures ν, $\nu^{\uparrow}$, $\nu^{\downarrow}$, $\tilde{\nu}^{\uparrow}$, $\tilde{\nu}^{\downarrow}$ are mutually absolutely continuous on $\{\deg(\boldsymbol{o}) > 0\}$.

Throughout this section we considered, for economy of notation, unmarked graphs. However, both definitions and results hold verbatimly for marked rooted graps $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ i.e. for probability measures over $\mathcal{G}_*^{\mathcal{X}}$.

The next result is an instructive application that was first estabilished formally in [LPP95] (with a different terminology).

Exercise 2.4.4. Prove that the *unimodular Galton-Watson tree* ${\rm GW}(P)$ defined in Section 2.2 is indeed unimodular.

Exercise 2.4.5. Modify the definition of Galton-Watson tree ${\rm GW}(\,\cdot\,)$ in Section 2.2 by considering generic degree distributions P (at the root) and ρ elsewhere. Prove that the resulting random tree is not unimodular unless P and ρ are related as per Eq. (2.2.4).

[2]The same definitions were made in in [DMS11], but notations here are slightly different.

Chapter 3

Factor models on trees

In this chapter we study factor models, under the additonal condition that the underlying graph is a tree. The motivation for devoting some energy to this special case is self-evident. Our main objective is to understand factor models on large locally tree-like graphs. It is unlikely that we will be able to achieve this objective unless we understand the case of large trees as well.

We will begin in Section 3.1 by considering *finite trees*. This case is particularly simple. The free energy can be show to be the equal to optimal value of a certain optimization problem that admits a convex formulation. The objective function is known as the *Bethe free energy*. The location of the optimum immediately yields a characterization of the local marginals of the probability measure $\mu_{G,\psi}(\sigma)$ introduced in Chapter 1. This characterization is computationally efficient (it allows to compute the free energy and local marginals in polynomial time).

In Section 3.2 we consider the case of infinite trees. In this case, the situation is significantly more subtle. First of all, the definition of the probability measure $\mu_{G,\psi}(\,\cdot\,)$ given in Chapter 1, Eq. (1.1.2) does not make sense any more. We have instead to define μ as a (non-necessarily unique) Gibbs measure corresponding to the specification ψ.

For a subclass of these Gibbs measures and if the underlying tree is unimodular, we can generalize the concepts developed in the case of finite trees. In particular, we can generalize the variational principle based on Bethe free energy. It is natural to conjecture that, for $\{G_n\}$ a sequence of graphs converging locally to a unimodular tree, the free energy density $\phi(G_n,\psi)$ converges to the solution of this variational problem. An analogous conjecture can be formulated for the local limit of the measures $\mu_{G_n,\psi}$. This *ansatz* is known as *Bethe-Peierls approximation*.

While this conjecture has been proved in some cases (proof techniques are discussed in the next two chapters), it is also known not to hold in general. Section 3.3 discusses generalizations based on the idea of 'replica symmetry breaking.'

3.1 Finite graphs

We are interested in the proability distribution on $\mathcal{X}^V$ defined by Eq. (1.1.2), that we copy here for the reader's convenience:

$$\mu_{G,\psi}(\sigma) = \frac{1}{Z(G,\psi)} \prod_{(i,j)\in E} \psi_{\mathrm{e}}(\sigma_i,\sigma_j) \prod_{i\in V} \psi_{\mathrm{v}}(\sigma_i)\,. \tag{3.1.1}$$

This section focuses on the case in which $G = T = (V, E)$ is a finite tree. Some of the results apply to general finite graphs: we will emphaseize when this is the case.

If G is not a tree, it is sometimes necessary to assume special conditions that ensure that the various expressions make sense (e.g. that there is no division by 0, and so on). A simple such condition is permissivity, cf. Definition 2.1.1. While weaker conditions can be used with additional technical work, we will use permissivity for ease of exposition.

3.1.1 Local marginals

Recall that, for $U \subseteq V$, we let $\sigma_U \equiv (\sigma_i)_{i\in U}$ denote that restriction of σ to U. Given $\boldsymbol{\sigma} \sim \mu_{G,\psi}$, we are interested in the in the marginal distribution of $\boldsymbol{\sigma}_U$, whereby U is –typically– a small subset of vertices. Formally we let, for any $\sigma_U \in \mathcal{X}^U$, we let

$$\mu_{G,\psi;U}(\sigma_U) = \mu_{G,\psi}\big(\{\boldsymbol{\sigma}_U = \sigma_U\}\big)\,. \tag{3.1.2}$$

In practice, we will typically drop the subscript U or G, ψ as it will be clear from the context.

It turns out that, when $G = T$ is a tree, all such local marginals can be succinctly encoded in a set of 'messages' and that these quantities can be efficiently computed.

We denote by $(u \to v)$ a ordered pair of vertices $u, v \in V$ such that $(u, v) \in E$ (i.e. $(u \to v)$ is an edge with a choice of direction). For such each ordered pair, we let $T_{u\to v} = (V_{u\to v}, E_{u\to v})$ be the subtree induced by the vertices that can be reached from u without passing through v (note that this excludes v). We let $\mu_{u\to v} = \mu_{T_{u\to v},\psi}$ be the probability measure on $\mathcal{X}^{V_{u\to v}}$ associated to this subtree. Namely, writing $\sigma = \sigma_{V_{u\to v}}$, we define

$$\mu_{u\to v}(\sigma) = \frac{1}{Z(T_{u\to v}, \psi)} \prod_{(i,j)\in E_{u\to v}} \psi_{\mathrm{e}}(\sigma_i, \sigma_j) \prod_{i\in V_{u\to v}} \psi_{\mathrm{v}}(\sigma_i)\,. \tag{3.1.3}$$

Let $\Delta_{\mathcal{X}}$ the $(|\mathcal{X}| - 1)$-dimensional simplex of probability measures over $\mathcal{X}$. We define the *equilibrium message* $h^*_{u\to v} \in \Delta_{\mathcal{X}}$ as the marginal distribution of $\boldsymbol{\sigma}_u$ under $\mu_{u\to v}$. Namely $h^*_{u\to v} = (h^*_{u\to v}(s))_{s\in\mathcal{X}}$ where, for each $s \in \mathcal{X}$,

$$h^*_{u\to v}(s) \equiv \mu_{u\to v}(\{\boldsymbol{\sigma}_u = s\})\,. \tag{3.1.4}$$

The following proposition shows that the equilibrium messages do indeed provide a good description of the local structure of $\mu_{T,\psi}$.

For this statement, and in the following it is useful to introduce the following convention. If p is a probability distribution over a finite set $\mathcal{S}$, and f a non-negative function on the same space, we write $p(s) \cong f(s)$ to mean that p is equal to f 'up to a normalization constant' (explicitly $p(s) = f(s)/\sum_{s'\in\mathcal{S}} f(s')$). Also, remeber that $d_G(i, j)$ is the graph distance between vertices i and $j \in V$. If $A, B \subseteq V$ are sets of vertices, $d_G(A, B) = \min_{i,j} d_G(i, j)$.

Proposition 3.1.1. *Assume G to be a tree, and let $U \subseteq V$ be a subset of vertices such that the induced graph G_U is connected. Let $\partial U \equiv \{(i \to j) : i \in V \setminus U, j \in U, (i, j) \in E\}$ and $U_+ = \{j \in V : d_G(U, j) \le 1\}$. Then*

$$\mu_{G,\psi}(\sigma_{U_+}) \cong \prod_{(i,j)\in E(G_{U_+})} \psi_{\mathrm{e}}(\sigma_i, \sigma_j) \prod_{i\in U} \psi_{\mathrm{v}}(\sigma_i) \prod_{(i\to j)\in\partial U} h^*_{i\to j}(\sigma_i)\,. \tag{3.1.5}$$

Further, for $(i,j) \in E$, there exists a constant z_{ij} (independent of σ_i, σ_j) such that

$$\mu_{G,\psi}(\sigma_i,\sigma_j) = \frac{1}{z_{ij}}\,\psi_{\rm e}(\sigma_i,\sigma_j)\,h^*_{i\to j}(\sigma_i)\,h^*_{j\to i}(\sigma_j)\,. \tag{3.1.6}$$

Proof. To simplify notation, we'll assume $\psi_{\rm v}(\sigma) = 1$ identically (the general case follows from the same calculation). By definition

$$\begin{aligned}\mu_{G,\psi}(\sigma_{U_+}) &\cong \sum_{\sigma_{V\setminus U_+}}\prod_{(i,j)\in E}\psi_{\rm e}(\sigma_i,\sigma_j) \qquad (3.1.7)\\ &= \prod_{(i,j)\in E(G_{U_+})}\psi_{\rm e}(\sigma_i,\sigma_j)\prod_{(i\to j)\in\partial U}Z_{i\to j}(\sigma_i)\\ Z_{i\to j}(\sigma_i) &\equiv \sum_{\sigma_{V_{i\to j}\setminus\{i\}}}\prod_{(l,k)\in E_{i\to j}}\psi_{\rm e}(\sigma_l\sigma_k)\,, \qquad (3.1.8)\end{aligned}$$

where the second identity follows simply by reorganizing the terms thanks to the fact that the trees $T_{i\to j}$, $(i\to j)\in\partial U$ are disjoint. The claim follows by noting that $h^*_{i\to j}(\sigma_i)\cong Z_{i\to j}(\sigma_i)$. □

For a generalization of this characterization to locally tree-like graphs, under additional correlation decay conditions, see [DM10, Theorem 3.14].

3.1.2 Belief-Propagation equations

Given a graph G we let $D(G) = \{(i\to j) : i,j\in V, (i,j)\in E\}$ be the set of edges of G with all choices of direction (in particular $|D(G)| = 2|E|$). For a fixed finite graph G, we define the space of messages as $\mathcal{H}_G \equiv (\Delta_{\mathcal{X}})^{D(G)}$. Hence $h\in\mathcal{H}_G$ is a vector $h = (h_{u\to v} : u,v\in V,(u,v)\in E)$. Whenever useful to indicate the underlying alphabet, we will write $\mathcal{H}_G(\mathcal{X})$.

Definition 3.1.2. *For a graph G and a specification ψ, the* belief propagation *(BP) iteration is the mapping $\mathsf{BP}_{G,\psi} : (\Delta_{\mathcal{X}})^{D(G)}\to(\Delta_{\mathcal{X}})^{D(G)}$ defined by letting for all $(i\to j)\in D(G)$,*

$$(\mathsf{BP}_{G,\psi}h)_{i\to j}(\sigma) \equiv \frac{1}{z_{i\to j}}\psi_{\rm v}(\sigma)\prod_{k\in\partial i\setminus j}\Big(\sum_{\sigma_k\in\mathcal{X}}\psi_{\rm e}(\sigma,\sigma_k)\,h_{k\to i}(\sigma_k)\Big)\,, \tag{3.1.9}$$

with $z_{i\to j}$ a normalization constant given by

$$z_{i\to j} \equiv \sum_{\sigma\in\mathcal{X}}\psi_{\rm v}(\sigma)\prod_{k\in\partial i\setminus j}\Big(\sum_{\sigma_k\in\mathcal{X}}\psi_{\rm e}(\sigma,\sigma_k)\,h_{k\to i}(\sigma_k)\Big)\,. \tag{3.1.10}$$

(If the sum on the right-hand side vanishes, the mapping is defined arbitrarily for that coordinate $(i\to j)$, e.g. setting $(\mathsf{BP}_{G,\psi}h)_{i\to j}(\sigma) = 1/|\mathcal{X}|$ for all $\sigma\in\mathcal{X}$.) Here we use the convention that a product of zero terms is equal to one.

We note that the BP mapping does not have any arbitrary choice if –for instance– ψ is permissive. Indeed, in this case we have

$$\psi_{\rm v}(\sigma_0)\prod_{k\in\partial i\setminus j}\Big(\sum_{\sigma_k\in\mathcal{X}}\psi_{\rm e}(\sigma_0,\sigma_k)\,h_{k\to i}(\sigma_k)\Big) > 0\,, \tag{3.1.11}$$

for $\sigma_0\in\mathcal{X}$ the permitted state, and any $h\in(\Delta_{\mathcal{X}})^{D(G)}$.

Proposition 3.1.3. *Let $G = T$ be a tree and ψ a specification. Then the equilibrium messages $h^* \in \mathcal{H}_T$ are the only solution of the fixed point equation*

$$h = \mathsf{BP}_{G,\psi} h \,. \tag{3.1.12}$$

Further, the sequence $\{h^t\}_{t\ge 0}$ defined by $h^{t+1} = \mathsf{BP}_{G,\psi} h^t$ converges after $t = \mathrm{diam}(T) + 1$ iterations (i.e. $h^t = h^*$ for every $t \ge \mathrm{diam}(T) + 1$).

Proof. For $(i \to j) \in D(G)$, define its depth $\mathrm{depth}_G(i \to j)$ as the length (number of edges) of the longest non-backtracking path from i to a leaf of T, that does not pass through j. In a finite tree $\mathrm{depth}_G(i \to j) \le \mathrm{diam}(G) < \infty$. The proof is by induction over the depth of $(i \to j)$. Namely, we claim that for any $t \ge \mathrm{depth}_G(i \to j) + 1$, $h^t_{i\to j} = h^\infty_{i\to j}$ remains unchanged and satisfies $h^\infty_{i\to j} = (\mathsf{BP}_{G,\psi} h^\infty)_{i\to j}$.

The base case of the induction is trivial since $\mathrm{depth}_G(i \to j) = 0$ implies that i is a leaf and hence the BP update reduces to $h^t_{i\to j}(\sigma) = (\mathsf{BP}_{G,\psi} h^{t-1})_{i\to j}(\sigma) = \psi_{\mathrm{v}}(sigma)/z_{i\to j}$, that does satisfy the claim for all $t \ge 0$.

The induction step is also immediate, by using $h^t_{i\to j} = (\mathsf{BP}_{G,\psi} h^{t-1})_{i\to j}$ and noting that $(\mathsf{BP}_{G,\psi} h^{t-1})_{i\to j}$ only depends on the messages h^{t-1} through its values at edges $(u \to v)$ with $\mathrm{depth}_G(u \to v) \le \mathrm{depth}_G(i \to j) - 1$. □

We will denote by $\mathcal{H}^*_G \subseteq \mathcal{H}_G$ the set of fixed points of the BP iteration, i.e. the set of solutions of Eq. (3.1.12).

The last proposition suggests an algorithm to compute the messages h^*. Simply start from an arbitrary initialization and iterate the BP mapping.The resulting algorithm is known as *belief propagation* or the *sum-product algorithm* and has numerous applications in artificial intelligence, machine learning, communications [RU08, WJ08, MM09, KF09].

3.1.3 Bethe-Peierls free energy

As stated by Proposition 3.1.1, equilibrium messages can be used to compute local marginals of $\mu_{G,\psi}$ when G is a tree. Can they be used also to compute the free energy? The answer is far from clear because the free energy is a 'global' quantity. The *Bethe-Peierls free energy* provides an answer to this question.

Definition 3.1.4. *Let $G = (V, E)$ be a finite graph and ψ a specification. The* Bethe-Peierls free energy *(or –for short–* Bethe free energy*) is a function $\Phi_{G,\psi} : \mathcal{H}_G \to \mathbb{R}$ defined by*

$$\Phi_{G,\psi}(h) \equiv \sum_{i\in V} \Phi^{\mathrm{v}}_{G,\psi}(h; i) - \sum_{(i,j)\in E} \Phi^{\mathrm{e}}_{G,\psi}(h; i, j) \,, \tag{3.1.13}$$

where

$$\Phi^{\mathrm{v}}_{G,\psi}(h; i) \equiv \log\Big\{ \sum_{\sigma\in\mathcal{X}} \psi_{\mathrm{v}}(\sigma) \prod_{j\in\partial i} \Big(\sum_{\sigma_j\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma, \sigma_j)\, h_{j\to i}(\sigma_j) \Big) \Big\} \,, \tag{3.1.14}$$

$$\Phi^{\mathrm{e}}_{G,\psi}(h; i, j) \equiv \log\Big\{ \sum_{\sigma_i,\sigma_j\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma_i, \sigma_j)\, h_{i\to j}(\sigma_i)\, h_{j\to i}(\sigma_j) \Big) \Big\} \,. \tag{3.1.15}$$

If any of the arguments of the logarithms vanishes at h, then $\Phi_{G,\psi}$ is defined arbitrarily.

Note that, if the specification ψ is permissive, then the arguments of logarithms of $\Phi^{\rm v}_{G,\psi}(h;i)$, $\Phi^{\rm e}_{G,\psi}(h;i,j)$ are always strictly positive, and hence there is no arbitrary choice in the definition of $\Phi_{G,\psi}$. The next statement justifies the definition of Bethe free energy.

Proposition 3.1.5. *Let $T = (V, E)$ be a tree and assume ψ to be permissive. Then the* Bethe-Peierls free energy $\Phi_{T,\psi} : \mathcal{H}_T \to \mathbb{R}$ *has a unique stationary point at the equilibrium messages $h^* \in \mathcal{H}_T$. Further, $\Phi_{T,\psi}(h_*) = \log Z(T, \psi)$.*

For proving this statement, it is actually convenient to introduce one more object, that is of independent interest. Given a graph $G = (V, E)$, a vector of *locally consistent marginals* is a vector $b = (b_{ij})_{(i,j)\in E}$ where $b_{ij} \in \Delta_{\mathcal{X}\times\mathcal{X}}$, that satisfies the following constraints

1. b is really indexed by edges. i.e.

$$b_{ij}(\sigma, \sigma') = b_{ji}(\sigma', \sigma) \tag{3.1.16}$$

   for all $(i, j) \in E$ and $\sigma, \sigma' \in \mathcal{X}$. Hence the vector $b = (b_{ij})_{(i,j)\in E}$ is really a point $b \in (\Delta_{\mathcal{X}\times\mathcal{X}})^E$.

2. For all $i \in V$, $(i, j) \in E$, $\sigma \in \mathcal{X}$, the quantity

$$b_i(\sigma) = \sum_{\sigma_j\in\mathcal{X}} b_{ij}(\sigma, \sigma_j)\,, \tag{3.1.17}$$

   only depends on i and not on the edge (i, j).

Note that we can associate bijectively the constraints at the second point with the directed edges $(i \to j) \in D(G)$. The set of locally consistent marginals form a finite-dimensional polytope denoted by $\mathcal{H}^{\rm loc}_G$ and known as the *local polytope* of G. This can be viewed as a polytope $\mathcal{H}^{\rm loc}_G \subseteq (\Delta_{\mathcal{X}\times\mathcal{X}})^E$. Whenever useful to indicate the underlying alphabet, we will use the notation $\mathcal{H}^{\rm loc}_G(\mathcal{X})$.

It is easy to see that the vector of marginals of a probability distribution over $\mathcal{X}^V$, do indeed belong to $\mathcal{H}^{\rm loc}_G$. Explicitly, we define the *marginal polytope* $\mathcal{H}^{\rm marg}_G$ as the set of vectors $b = (b_{ij})_{(i,j)\in E}$ with $b_{ij} \in \Delta_{\mathcal{X}\times\mathcal{X}}$, for wich there exists a random $\boldsymbol{\sigma}$ in $\mathcal{X}^V$, such that

$$b_{ij}(\sigma_1, \sigma_2) = \mathbb{P}(\boldsymbol{\sigma}_i = \sigma_1, \boldsymbol{\sigma}_j = \sigma_2)\,. \tag{3.1.18}$$

Note that $\mathcal{H}^{\rm marg}_G$ is obviously a polytope since it is a projection of the polytope $\Delta_{\mathcal{X}^V}$ of probability distributions over $\mathcal{X}^V$.

Proposition 3.1.6. *For a general graph G, $\mathcal{H}^{\rm marg}_G \subseteq \mathcal{H}^{\rm loc}_G$. If $G = T$ is a tree, then$\mathcal{H}^{\rm marg}_T = \mathcal{H}^{\rm loc}_T$. Finally, if $\mu = \mu_{T,\psi}$ for a specification ψ, then $\mu_{T,\psi}$ is uniquely determined by its vector of marginals $\mu_{ij}(\sigma_1, \sigma_2) = \mu(\{\boldsymbol{\sigma}_i = \sigma_1, \boldsymbol{\sigma}_j = \sigma_2\})$, $\mu_i(\sigma) = \mu(\{\boldsymbol{\sigma}_i = \sigma\}) = \sum_{\sigma'\in\mathcal{X}} \mu_{ij}(\sigma, \sigma')$, namely*

$$\mu_T(\sigma) = \prod_{(i,j)\in E(T)} \frac{\mu_{ij}(\sigma_i, \sigma_j)}{\mu_i(\sigma_i)\mu_j(\sigma_j)} \prod_{i\in V(T)} \mu_i(\sigma_i)\,. \tag{3.1.19}$$

Proof. The inclusion $\mathcal{H}_G^{\rm marg} \subseteq \mathcal{H}_G^{\rm loc}$ obvious. To prove the converse in the tree case, we have to exhibit, for $b \in \mathcal{H}_T^{\rm loc}$ a probability distribution over $\sigma \in \mathcal{X}^V$ whose marginals coincide with the b_{ij}'s. Note that, since both $\mathcal{H}_G^{\rm marg}$ and $\mathcal{H}_G^{\rm loc}$ are closed sets, it is sufficient to choose b in the relative interior of $\mathcal{H}_T^{\rm loc}$. We can therefore assume that $b_{ij}(\sigma_1, \sigma_2) > 0$ for all $(i,j) \in E$, and $\sigma_1, \sigma_2 \in \mathcal{X}$. The probability distribution over $\mathcal{X}^V$ is then given by

$$b_T(\sigma) = \prod_{(i,j)\in E(T)} \frac{b_{ij}(\sigma_i, \sigma_j)}{b_i(\sigma_i) b_j(\sigma_j)} \prod_{i \in V(T)} b_i(\sigma_i)\,. \tag{3.1.20}$$

It is proved by induction over the number of vertices that the marginals are indeed as desired. (Given a tree T over n vertices, let i be a leaf of T. The number of vertices can be reduced by marginalizing –integrating over– the variable σ_i.)

The last claim (for the case $\mu = \mu_{T,\psi}$) follows also by induction over the number of vertices. □

Notice that Eq. (3.1.20) defines a mapping $\gamma_T : \mathcal{H}_T^{\rm loc} = \mathcal{H}_T^{\rm marg} \to \Delta_{\mathcal{X}^V}$. The image of this mapping $\mathcal{M}_T \equiv \gamma_T(\mathcal{H}_T^{\rm marg})$ is the set of probability distributions on $\mathcal{X}^V$ that can be written as

$$\mu(\sigma) = \frac{1}{Z} \prod_{(i,j)\in E} \psi_{ij}(\sigma_i, \sigma_j)\,. \tag{3.1.21}$$

Equivalently, this is the set of probability distribution tha satisfy the Markov property[1] with respect to T. This set of probability distributions is a nonlinear algebraic variety.

Given a probability distribution $p = \{p(s)\}_{s\in\mathcal{S}}$ on a finite set $\mathcal{S}$, we denote by $H(p)$ its Shannon entropy

$$H(p) \equiv -\sum_{s\in\mathcal{S}} p(s) \log p(s)\,. \tag{3.1.22}$$

Given probability distribution two probability measures p, q on the space $\mathcal{S}$, the *relative entropy* of p with respect to q is

$$H(p\|q) \equiv \sum_{s\in\mathcal{S}} p(s) \log \frac{p(s)}{q(s)}\,. \tag{3.1.23}$$

As a special case, if $p_{R,S} = \{p_{R,S}(r,s)\}_{r\in\mathcal{R}, s\in\mathcal{S}}$ is the joint distribution of random variables R, S on finite sets $\mathcal{R}$, $\mathcal{S}$, with marginals p_R and p_E, the mutual information of R and S is

$$I(R;S) \equiv H(p_{R,S}\|p_R \times p_S) = \sum_{s\in\mathcal{S}, r\in\mathcal{R}} p_{R,S}(r,s) \log \frac{p_{R,S}(r,s)}{p_R(r) p_S(s)} \tag{3.1.24}$$

Definition 3.1.7. *Let $G = (V, E)$ be a finite graph and ψ a specification. The* dual Bethe free energy *is the function $\Phi^*_{G,\psi} : \mathcal{H}_T^{\rm loc} \to \mathbb{R}$ defined by*

$$\Phi^*_{G,\psi}(b) = \sum_{i\in V} \mathbb{E}_{b_i} \log \psi_{\rm v}(\boldsymbol{\sigma}_i) + \sum_{(i,j)\in E} \mathbb{E}_{b_{ij}} \log \psi_{\rm e}(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j) + \sum_{i\in V} H(b_i) + \sum_{(i,j)\in E} H(b_{ij}\|b_i \times b_j)\,, \tag{3.1.25}$$

[1] A probability distribution μ on $\mathcal{X}^V$ satisfies the global Markov property with respect to the graph G if, for any sets $A, B, C \subseteq V$ such that C separates A from B, we have $\mu(\sigma_A, \sigma_B|\sigma_C) = \mu(\sigma_A|\sigma_C)\mu(\sigma_B|\sigma_C)$.

with the convention $\mathbb{E}_{b_{ij}} \log \psi_{\mathrm{e}}(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j) = -\infty$ *if* $\mathrm{supp}(b_{ij}) \not\subseteq \mathrm{supp}(\psi_{\mathrm{e}})$. *Here* $\mathbb{E}_{b_i}$ *denotes expectation with respect to* b_i *and* $\mathbb{E}_{b_{ij}}$ *is expectation with respect to* b_{ij}. *(Here* b_i *should be interpreted as a shorthand for* $b_i(\sigma) = \sum_{\sigma' \in \mathcal{X}} b_{ij}(\sigma, \sigma')$.*)*

We will drop the 'dual' qualifier whenever clear from the context.

The following statement provides a natural embedding of the set of belief propagation fixed points $\mathcal{H}^*_G$ into the local polytope $\mathcal{H}^{\mathrm{loc}}_G$.

Proposition 3.1.8. *Let* $G = (V, E)$ *be a graph and* ψ *a permissive specification. Let* $h \in \mathcal{H}^*_G$ *and define, for each* $(i, j) \in E$, *and all* $\sigma, \sigma' \in \mathcal{X}$,

$$b_{ij}(\sigma_i, \sigma_j) = \frac{1}{z_{ij}} \psi_{\mathrm{e}}(\sigma_i, \sigma_j)\, h_{i\to j}(\sigma_i)\, h_{j\to i}(\sigma_j)\,, \tag{3.1.26}$$

where $z_{ij} = \sum_{\sigma,\sigma' \in \mathcal{X}} \psi_{\mathrm{e}}(\sigma, \sigma')\, h_{i\to j}(\sigma)\, h_{j\to i}(\sigma')$ *is a normalization constant, which we assume strictly positive* $z_{ij} > 0$ *for all* $(i, j) \in E$. *Then* $b = (b_{i,j})_{(ij)\in E} \in \mathcal{H}^{\mathrm{loc}}_G$, *with* $\Phi^*_{G,\psi}(b) = \Phi_{G,\psi}(h)$.

If $G = T$ *is a tree, then the permissivity assumption is not necessary.*

Proof. The definition (3.1.26) is well-posed with $z_{ij} > 0$ under permissivity, or if G is a tree.

Let us first prove $b \in \mathcal{H}^{\mathrm{loc}}_G$. Note that $b_{ij}(\sigma, \sigma') = b_{ji}(\sigma', \sigma)$ by construction. Next we want to check that $\sum_{\sigma_j \in \mathcal{X}} b_{ij}(\sigma_i, \sigma_j)$ does not depend on $j \in \partial i$. Since b_{ij} is normalized by construction, it is sufficient to check this up to normalization constants.

Recall that the fixed point condition for edge $(i \to j)$ yields (with $\cong$ denoting equality up to normalizations)

$$h_{i\to j}(\sigma_i) = (\mathsf{BP}_{G,\psi} h)_{i\to j}(\sigma_i) \cong \psi_{\mathrm{v}}(\sigma_i) \prod_{k \in \partial i \setminus j} \Big(\sum_{\sigma_k \in \mathcal{X}} \psi_{\mathrm{e}}(\sigma_i, \sigma_k)\, h_{k\to i}(\sigma_k) \Big)\,, \tag{3.1.27}$$

and hence substitutiong in the definition of b_{ij},

$$\sum_{\sigma_j \in \mathcal{X}} b_{ij}(\sigma_i, \sigma_j) \cong \psi_{\mathrm{v}}(\sigma_i) \prod_{k \in \partial i} \Big(\sum_{\sigma_k \in \mathcal{X}} \psi_{\mathrm{e}}(\sigma_i, \sigma_k)\, h_{k\to i}(\sigma_k) \Big)\,. \tag{3.1.28}$$

It is clear that the right-hand side does not depend on the choice of $j \in \partial i$.

The equality $\Phi^*_{G,\psi}(b) = \Phi_{G,\psi}(h)$ follows by substituting the formula (3.1.26) in the definition of $\Phi^*_{G,\psi}(b)$. □

Proposition 3.1.9. *Let* T *be a finite tree and* ψ *a specification. Then*

$$\log Z(T, \psi) = \max_{b \in \mathcal{H}^{\mathrm{loc}}_T} \Phi^*_{T,\psi}(b)\,. \tag{3.1.29}$$

Further the maximum is uniquely achieved at $b^* = (b^*_{ij})$, *where* $b^*_{ij} = \mu_{ij}$ *are the marginals of* $\mu_{T,\psi}$.

Proof. For the sake of simplicity, we will work under the assumption that the specification is permissive, leaving generalizations to the reader.

For this proof it is useful to introduce one more concept of independent interest: the Gibbs free energy $\Phi^{\rm Gibbs}_{T,\psi}$. This is a function over the simplex of probability distributions on $\mathcal{X}^V$: $\Phi^{\rm Gibbs}_{T,\psi} : \Delta_{\mathcal{X}^V} \to \mathbb{R}$. Let us introduce the notation

$$\psi_G(\sigma) \equiv \prod_{i\in V} \psi_{\rm v}(\sigma_i) \prod_{(i,j)\in E} \psi_{\rm e}(\sigma_i,\sigma_j)\,. \tag{3.1.30}$$

Then, for $p \in \Delta_{\mathcal{X}^V}$, the corresponding Gibbs free energy is given by

$$\Phi^{\rm Gibbs}_{T,\psi}(p) \equiv \mathbb{E}_p \log \psi_G(\boldsymbol{\sigma}) + H(p) \tag{3.1.31}$$

$$= \sum_{i\in V} \mathbb{E}_{p_i} \log \psi_{\rm v}(\boldsymbol{\sigma}_i) + \sum_{(i,j)\in E} \mathbb{E}_{p_{ij}} \log \psi_{\rm e}(\boldsymbol{\sigma}_i,\boldsymbol{\sigma}_j) + H(p)\,, \tag{3.1.32}$$

where $\mathbb{E}_q$ denotes expectation with respect to the probability measure p and p_i, p_{ij} are the marginals of p. Two differences can be noted with respect to the Bethe free energy (3.1.25): (i) the domain is the set $\Delta_{\mathcal{X}^V}$ of globally consistent probability distributions instead of the local polytope $\mathcal{H}^{\rm loc}_G$; (ii) the Shannon entropy $H(p)$ replaces the combination of edge and vertex entropies.

Assume, without loss of generality $\psi_T(\sigma) > 0$ for all $\sigma \in \mathcal{X}^V$ (if this is not the case, the domain of σ can be restricted to the support of ψ_T). The following properties of the Gibbs free energy are immediate to verify:

1. $\Phi^{\rm Gibbs}_{T,\psi}$ is continuous bounded and strictly convex on $\Delta_{\mathcal{X}^V}$. Further, it is differentiable in its interior.

2. Differentiating $\Phi^{\rm Gibbs}_{T,\psi}$ with respect to p, we see that the maximum is uniquely achieved at $p(\sigma) = \psi_T(\sigma)/Z(T,\psi) = \mu_{T,\psi}(\sigma)$.

3. The value of the maximum is $\Phi^{\rm Gibbs}_{T,\psi}(\mu_{T,\psi}) = \log Z(T,\psi)$.

4. For $b \in \mathcal{H}^{\rm loc}_T$, we have $\Phi^*_{T,\psi}(b) = \Phi^{\rm Gibbs}_{T,\psi}(\gamma_T(b))$ (recall that γ_T denotes the embedding defined in Eq. (3.1.20).

From the last two properties it follows that, for any $b \in \mathcal{H}^{\rm loc}_T$,

$$\Phi^*_{T,\psi}(b) \le \max_{p\in\Delta_{\mathcal{X}^V}} \Phi^{\rm Gibbs}_{T,\psi}(p) = \log Z(T,\psi)\,. \tag{3.1.33}$$

By taking b the sets of marginals of $\mu_{T,\psi}$, the inequality is satisfied with equality. Finally uniqueness follows from strict convexity of $\Phi^{\rm Gibbs}_{T,\psi}$ with γ_T being one-to-one on its image. □

Theorem 3. *Let ψ be a permissive specification and $G = (V,E)$ a finite graph.*

1. *Let $b \in \mathcal{H}^{\rm loc}_G$ be a maximizer of $\Phi^*_{G,\psi}$ over $\mathcal{H}^{\rm loc}_G$. Then, for all $(i,j) \in E$, $\mathrm{supp}(b_{ij}) = \mathrm{supp}(\psi_{\rm e})$.*

2. *Let $b \in \mathcal{H}^{\rm loc}_G$ be any local maximizer of $\Phi^*_{G,\psi}$ such that, for all $(i,j) \in E$, $\mathrm{supp}(b_{ij}) = \mathrm{supp}(\psi_{\rm e})$. Then b is the image of some $h \in \mathcal{H}^*_G$ under Eq. (3.1.26).*

3. *As a consequence of the last two points*

$$\max_{h\in\mathcal{H}^*_G} \Phi_{G,\psi}(h) = \max_{b\in\mathcal{H}^{\rm loc}} \Phi^*_{G,\psi}(b)\,. \tag{3.1.34}$$

Proof. Let us begin with point 1. First note that $\mathrm{supp}(b_{ij}) \subseteq \mathrm{supp}(\psi_{\mathrm{e}})$ because otherwise $\Phi^*_{G,\psi}(b) = -\infty$. Assume that there exists $(\sigma_1^*, \sigma_2^*) \in \mathrm{supp}(\psi_{\mathrm{e}})$ with $b_{ij}(\sigma_1^*, \sigma_2^*) = 0$. Then define, for $\varepsilon \in (0, 1/2)$, and all $(u, v) \in E$,

$$b^{\varepsilon}_{uv}(\sigma_1, \sigma_2) = \varepsilon \mathbb{I}\big((\sigma_1, \sigma_2) = (\sigma_1^*, \sigma_2^*)\big)\varepsilon \mathbb{I}\big((\sigma_1, \sigma_2) = (\sigma_2^*, \sigma_1^*)\big) + (1 - 2\varepsilon) b_{u,v}(\sigma_1, \sigma_2)\,. \tag{3.1.35}$$

It is then easy to check that $b^{\varepsilon} \in \mathcal{H}^{\mathrm{loc}}_G$ (by convexity of $\mathcal{H}^{\mathrm{loc}}_G$) and that $\Phi^*_{G,\psi}(b^{\varepsilon}) = \Phi^*_{G,\psi}(b) + c\,\varepsilon \log(1/\varepsilon) + O(\varepsilon)$ for some $c > 0$. It then follows that b cannot be a maximixer.

Consider now point 2, and let $\widetilde{\mathcal{H}^{\mathrm{loc}}}_G \subseteq \mathcal{H}^{\mathrm{loc}}_G$ be the subset of the local polytope wich satisfies $b_{ij}(\sigma_1, \sigma_2)$ for all $(i, j) \in E$ and all σ_1, σ_2 such that $\psi_{\mathrm{e}}(\sigma_1, \sigma_2) = 0$. Let $b \in \widetilde{\mathcal{H}^{\mathrm{loc}}}_G$ be local maximizer of $\Phi^*_{G,\psi}$. By the previous point b is in the relative interior of $\widetilde{\mathcal{H}^{\mathrm{loc}}}_G$ and by permissivity this implies $\mathrm{supp}(b_i) = \mathcal{X}$.

Note that $\Phi^*_{G,\psi}$ is differentiable in the relative interior of $\widetilde{\mathcal{H}^{\mathrm{loc}}}_G$. Then there must exist Lagrange multipliers $\lambda_{i\to j}(\sigma)$ indexed by $(i \to j) \in D(G)$ and $\sigma \in \mathcal{X}$ such that $(b_{ij}, b_i)_{(i,j)\in E, i\in V}$ is a stationary point of the Lagrangian

$$\mathcal{L}(b, \lambda) \equiv \Phi^*_{G,\psi}(b) + \sum_{(i\to j)\in D(G)} \sum_{\sigma_i \in \mathcal{X}} \lambda_{i\to j}(\sigma_i) \Big(\sum_{\sigma_j \in \mathcal{X}} b_{ij}(\sigma_i, \sigma_j) - b_i(\sigma_i) \Big)\,, \tag{3.1.36}$$

over the domain $\Delta^E_{\mathcal{X}\times\mathcal{X}} \times \Delta^V_{\mathcal{X}}$. Let us now focus on the dependence of $\mathcal{L}$ on b_{ij}:

$$\mathcal{L}(b, \lambda) = \mathbb{E}_{b_{ij}} \log \psi_{\mathrm{e}}(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j) + H(b_{ij}) + \mathbb{E}_{b_{ij}} \big\{ \lambda_{i\to j}(\boldsymbol{\sigma}_i) + \lambda_{j\to i}(\boldsymbol{\sigma}_j) \big\} + \mathcal{L}_{ij}(b, \lambda)\,, \tag{3.1.37}$$

where $\mathcal{L}_{ij}$ is independent of b_{ij}. From the last expression, stationarity with respect to b_{ij} implies

$$b_{ij}(\sigma_i, \sigma_j) = \frac{1}{z_{ij}}\, \psi_{\mathrm{e}}(\sigma_i, \sigma_j)\, e^{\lambda_{i\to j}(\sigma_i) + \lambda_{j\to i}(\sigma_j)}\,, \tag{3.1.38}$$

which coincides with Eq. (3.1.26), for $h_{i\to j}(\sigma) = e^{\lambda_{i\to j}(\sigma)}$. In particular $h_{i\to j}(\sigma) > 0$ for all $(i \to j) \in D(G)$ and all $\sigma \in \mathcal{X}$.

Considering next the dependence of the Lagrangian on b_i, we get

$$\mathcal{L}(b, \lambda) = \mathbb{E}_{b_i} \log \psi_{\mathrm{v}}(\boldsymbol{\sigma}_i) - (\deg_G(i) - 1) H(b_i) - \mathbb{E}_{b_i} \Big\{ \sum_{j\in\partial i} \lambda_{i\to j}(\boldsymbol{\sigma}_i) \Big\} + \mathcal{L}_i(b, \lambda)\,, \tag{3.1.39}$$

with $\mathcal{L}_i$ independent of b_i. Again, the stationarity condition with respect to b_i implies (using $h_{i\to j}(\sigma) = e^{\lambda_{i\to j}(\sigma)}$)

$$b_i(\sigma_i) \cong \Big\{ \psi_{\mathrm{v}}(\sigma_i)^{-1} \prod_{j\in\partial i} h_{i\to j}(\sigma_i) \Big\}^{1/(\deg_G(i)-1)}\,. \tag{3.1.40}$$

Then from Eq. (3.1.26), we have

$$\prod_{k\in\partial i\setminus j} \Big\{ \sum_{\sigma_k \in \mathcal{X}} b_{ik}(\sigma_i, \sigma_k) \Big\} \cong \prod_{k\in\partial i\setminus j} h_{i\to k}(\sigma_i) \cdot \prod_{k\in\partial i\setminus j} \Big\{ \sum_{\sigma_k \in \mathcal{X}} \psi_{\mathrm{e}}(\sigma_i, \sigma_k) \Big\}\,. \tag{3.1.41}$$

On the other hand, using Eq. (3.1.40) and the fact that $\sum_{\sigma_k\in\mathcal{X}} b_{ik}(\sigma_i, \sigma_k) = b_i(\sigma_i)$ we have

$$\prod_{k\in\partial i\setminus j} \Big\{ \sum_{\sigma_k \in \mathcal{X}} b_{ik}(\sigma_i, \sigma_k) \Big\} = b_i(\sigma_i)^{\deg_G(i)-1} = \psi_{\mathrm{v}}(\sigma_i)^{-1} \prod_{j\in\partial i} h_{i\to j}(\sigma_i)\,.. \tag{3.1.42}$$

Equating Eq. (3.1.41) and (3.1.42), and recalling that $h_{i\to k}(\sigma_i) > 0$ strictly, we get $h_{i\to j} = (\mathsf{BP}_{G,\psi} h)_{i\to j}$, i.e. $h \in \mathcal{H}^*_G$.

Finlly, point 3 follows from from the previous two points, together with Proposition 3.1.8. □

3.2 Infinite trees

3.2.1 Definitions

We consider now a possibly infinite rooted tree (T, o), with $T = (V, E)$. We will assume that the tree contains more than one vertex (or does it with positive probability if the tree is random). Most of the statements elow are trivial if T is finite, so the reader can assume that T is indeed infinite. For $t \in \mathbb{N}$, e denote by $V_{=t}$, $V_{\le t}$, $V_{\ge t}$ the subset of vertices at distance equal to t, at most t, or at least t from o. The subsets $V_{<t}$, $V_{>t}$ are defined analogously, and we use $T_{\dots}$ for the subgraph induced by $V_{\dots}$. Finally we will write σ_t for σ_{V_t}, $\sigma_{\le t}$ for $\sigma_{V_{\le t}}$ and so on.

Of course Eq. (1.1.2) cannot be longer used to define $\mu_{\psi,T}$. For any fixed T, the space $\mathcal{X}^V$ is made into a measure space using the Borel σ-algebra induced by the same distance introduced in Chapter 2, namely $d(\sigma, \sigma') = 2^{-R(\sigma,\sigma')}$ where $R(\sigma, \sigma')$ is the distance from the root of the first vertex i where $\sigma_i \neq \sigma'_i$. This motivates the following definition.

Definition 3.2.1. *Let (T, o) be a rooted tree, and ψ a specification. We say that a probability measure $\mu_{T,\psi}$ on $(\mathcal{X}^V, \mathcal{B}(\mathcal{X}^V))$ is a* Gibbs measure *on (T, ψ) if, for every $t \in \mathbb{N}$, all $\sigma_{<t}$ and $\mu_{T,psi}$-almost all $\sigma_{\ge T}$*

$$\mu_{T,\psi}(\sigma_{<t}|\sigma_{\ge t}) = \mu_{T_{\le t},\psi}(\sigma_{<t}|\sigma_{=t})\,.$$

Explicitly, the right hand side is given by

$$\mu_{T_{\le t},\psi}(\sigma_{<t}|\sigma_{=t}) = \frac{1}{Z(T;\sigma_{\ge t})} \prod_{(i,j)\in E_{\le t}} \psi_{\mathrm{e}}(\sigma_i,\sigma_j) \prod_{i\in V_{<t}} \psi_{\mathrm{v}}(\sigma_i)\,,$$

with $Z(T;\sigma_{\ge t})$ a normalization constant.

Notice that the set of Gibbs measures with respect to (T, ψ) contains –in general– more than one element.

Proposition 3.2.2. *The set of Gibbs measures on a given tree T is convex. Further, it is closed and subsequentially compact with respect to weak topology.*

Proof. Can be found in **???** □

A standard way to construct a Gibbs measure is through a limit procedure.

Definition 3.2.3. *Let $f : \mathcal{X}^{V_{=t}} \to \mathbb{R}_+$ be a measure that is not identically zero on $\sigma_{=t}$. Then, the Gibbs measure on (T, ψ) with* boundary condition *f at level t is the probability measure given by*

$$\mu^{f,t}_{T,\psi}(\sigma_{\le t}) = \frac{1}{Z(f,t)} \prod_{(i,j)\in E_{\le t}} \psi_{\mathrm{e}}(\sigma_i,\sigma_j) \prod_{i\in V_{\le t}} \psi_{\mathrm{v}}(\sigma_i)\, f(\sigma_t)\,, \tag{3.2.1}$$

and defined arbitrarily for $\sigma_{>t}$.

In particular, the Gibbs measure with free boundary conditions *is the one obtained by letting $f(\sigma_t) = 1$ be the constant function. Further, for $\sigma_0 \in \mathcal{X}$, the Gibbs measure with σ_0* boundary conditions *is the one corresponding to $f(\sigma_t) = 1$ if $\sigma_{t,i} = \sigma_0$ for all $i \in V_t$ and $f(\sigma_t) = 0$ otherwise.*

Now, choose $f = \{f^t\}_{t\geq 0}$ a sequence of boundary conditions at level t, and letting $\mu^t_{T,\psi}$ be the Gibbs measure with boundary condition f^t at level t. It follows immediately from Proposition 3.2.2 that the subsequential limits are Gibbs measures. If the subsequential limits coincide (i.e. there exists a limit), then we will speak of the f-boundary Gibbs measure (e.g. free-boundary Gibbs measure, σ_0-boundary Gibbs measure). It is also easy to prove that any Gibbs measure on T can be constructed in this way.

If the underlying rooted tree $(\boldsymbol{T}, \boldsymbol{o})$ is random we can define a Gibbs measure $\mu_{\boldsymbol{T},\psi}$ per realization of the underlying tree. However, in general, the resulting probabilities are not measurable with respect to $\boldsymbol{T}$ i.e. $\mu_{\boldsymbol{T},\psi}(\{\boldsymbol{\sigma}_{\leq t} = \sigma_{\leq t}\})$ is not a random variable for some given $\sigma_{\leq t} \in \mathcal{X}^{V_{\leq t}}$.

Exercise 3.2.1. Construct an example of Gibbs measure that in non-measurable with respect to the underlying tree. (You need to assume that there exists (T, ψ) for which at least two distinct Gibbs measures exist.)

The following definition bypasses this problem.

Definition 3.2.4. *Let $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ be a marked random rooted tree and denote by ν its law. We say that ν is a Gibbs measure with respect to the specification ψ if, for any $t \in \mathbb{N}$, $\sigma_{<t} \in \mathcal{X}^{V_{<t}}$, we have. ν-almost surely,*

$$\nu(\{\boldsymbol{\sigma}_{<t} = \sigma_{<t}\}|\boldsymbol{\sigma}_{\geq t}, \boldsymbol{T}) = \mu_{\boldsymbol{T}_{\leq t}\psi}(\sigma_{<t}\}|\boldsymbol{\sigma}_{=t})\,. \tag{3.2.2}$$

If this is the case, we will also say that the marked tree $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ is

The relevance of Gibbs measures to these lectures is given by the following.

Proposition 3.2.5. *Let $\{G_n\}_{n\geq 1}$ be a sequence of finite graphs and ψ a specification. If $(G_n, \mu_{G_n,\psi})$ converges locally to the marked random rooted tree $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ with law ν, then ν is Gibbs with respect to ψ.*

Proof. Note that, for ν to be Gibbs, i.e. for Definition 3.2.4 to hold, it is sufficient to check that, for any fixed $T \geq t$,

$$\nu(\{\boldsymbol{\sigma}_{<t} = \sigma_{<t}\}|\{\boldsymbol{\sigma}_\ell\}_{t\leq \ell\leq T}, \boldsymbol{T}) = \mu_{\boldsymbol{T}_{\leq t}\psi}(\sigma_{<t}\}|\boldsymbol{\sigma}_{=t})\,. \tag{3.2.3}$$

□

3.2.2 Belief propagation equations

Given an infinite rooted tree (T, o) a *set of messages* is a map $h : D(T) \to \Delta_{\mathcal{X}}$. (This is called a 'boundary law' in the theory of Gibbs measures, but we prefer to keep consistency with the finite-graph terminology.) Note that the BP operator, as defined in Eq. (3.1.9) is well defined even if T is infinite, as a mapping $\mathsf{BP}_{T,\psi} : (\Delta_{\mathcal{X}})^{D(T)} \to (\Delta_{\mathcal{X}})^{D(T)}$.

Consider first the case of 'factorized' boundary conditions, i.e. boundary conditions that, at level t, read

$$f^t(\sigma_t) = \prod_{i\in V_t} f^t_i(\sigma_i)\,. \tag{3.2.4}$$

Proposition 3.2.6. *Let $\mu_{\psi,T}$ be any subsequential limit that can be obtained by a 'factorized' boundary condition. Then there exists a set of messages h such that $h = \mathsf{BP}_{\psi,T} h$ and, for each t, denoting by $p(i)$ the parent of vertex i,*

$$\mu_{T,\psi}(\sigma_{\le t}) = \frac{1}{Z(f,t)} \prod_{(i,j)\in E_{\le t}} \psi_{\mathrm{e}}(\sigma_i,\sigma_j) \prod_{i\in V_{<t}} \psi_{\mathrm{v}}(\sigma_i) \prod_{i\in V_t} h_{i\to p(i)}(\sigma_i). \tag{3.2.5}$$

Viceversa, for any set of messages h such that $h = \mathsf{BP}_{\psi,T} h$, there exists a (unique) Gibbs measure $\mu_{T,\psi}$ satisfying the last equation for all t. Finally, any Gibbs measure with respect to (T,ψ) can be written as convex combination of Gibbs measures of this form.

Proof. Easy except the 'finally' part. □

In the case of random unimodular trees, we look at the tree fro the perspective of the root. Recall that $\mathcal{T}_{\mathrm{e}} \subseteq \mathcal{G}_{\mathrm{e}} \subseteq \mathcal{G}_{**}$ is the space of edge-rooted trees. Also recall the definitions of $\nu^{\uparrow}$ and $\nu^{\downarrow}$ given in section

Definition 3.2.7. *A message function is a measurable function*

$$h : \mathcal{T}_{\mathrm{e}} \to \Delta_{\mathcal{X}}\,, \qquad (T, u\to v) \mapsto (h_{T,u\to v}(\sigma))_{\sigma\in\mathcal{X}}\,. \tag{3.2.6}$$

We will often drop the argument T, when clear from the context. Fiven a measure ν on random rooted tree, and $\nu^{\uparrow}$ on $\mathcal{T}_{\mathrm{e}}$, we consider two message functions equivalent if they coincide $\nu^{\uparrow}$-almost everywhere. We denote the space of message functions as $\mathcal{H} = \mathcal{H}_{\nu}$.

We can extend the above definition by letting h depend on additional randomness, e.g. letting $h : \mathcal{T}_{\mathrm{e}}^{*} \to \Delta_{\mathcal{X}}$ where $\mathcal{T}_{\mathrm{e}}^{*}$ is a space of trees with random marks on the edges. We will possibly discuss this in the last lecture.

Also, if the specificarion $\psi = \psi^{\theta}$ depends parametrcally on $\theta \in Q \subseteq \mathbb{R}^k$, it is sometimes useful to generalize the definition by letting

$$h : \mathcal{T}_{\mathrm{e}} \times Q \to \Delta_{\mathcal{X}}\,, \qquad ((T, u\to v),\theta) \mapsto (h^{\theta}_{T,u\to v}(\sigma))_{\sigma\in\mathcal{X}}\,. \tag{3.2.7}$$

Definition 3.2.8. *The BP mapping on the space $\mathcal{H}$ is the measurable mapping $\mathsf{BP}_{\psi} : \mathcal{H} \to \mathcal{H}$ defined by letting, for each $(T, i\to j) \in \mathcal{T}_e$ and each $\sigma \in \mathcal{X}$,*

$$(\mathsf{BP}_{\psi} h)_{T,i\to j} = \frac{1}{z_{i\to j}} \psi_{\mathrm{v}}(\sigma) \prod_{k\in\partial i\setminus j} \Big(\sum_{\sigma_k\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma,\sigma_k)\, h_{T,k\to i}(\sigma_k)\,. \tag{3.2.8}$$

If ν is a measure on random rooted trees, we write $\mathcal{H}_{\nu}^{*}$ for the set of message functions such that

$$h = \mathsf{BP}_{\psi} h\,, \qquad \nu^{\uparrow} - \textit{almost surely.} \tag{3.2.9}$$

(In case h is random, as below Definition (3.2.7), this equation holds in distribution, with the h's on right-hand sides independent.)

The definition of message function can appear rather abstract, but can be explicitly evaluated in concrete cases.

The case of regular deterministic trees

Assume that the random rooted tree is indeed deterministic and hence equal to the k-regular tree $T^{\mathrm{reg},k}$. In this case $(T, u \to v)$ is also constant (up to the labelings) and equal to the k-regular rooted at an edge. Hence it is necessary to define h only for this case. Equivalently, we can take the function h to be constant over $\mathcal{T}_{\mathrm{e}}$. The set of BP fixed points $\mathcal{H}^*_\nu$ is therefore identified with the set of solutions in $\Delta_{\mathcal{X}}$ of the equation

$$h(\sigma) = \frac{1}{z}\psi_{\mathrm{v}}(\sigma)\left(\sum_{\sigma'\in\mathcal{X}}\psi_{\mathrm{e}}(\sigma,\sigma')\,h(\sigma')\right)^{k-1}, \tag{3.2.10}$$

with z a normalization constant. As demonstrated by the next exercise, even in this case there can be multiple solutions.

Exercise 3.2.2. Consider the ferromagnetic Ising model in zero field, i.e. the model with $\mathcal{X} = \{+1, -1\}$, $\psi_{\mathrm{v}}(\sigma) = 1$ and $\psi_{\mathrm{e}}(\sigma_1, \sigma_2) = e^{\beta\sigma_1\sigma_2}$, with $\beta \in \mathbb{R}_+$.

Prove that, for $(k-1)\tanh\beta \le 1$, Eq. (3.2.10) admits the unique solution $h(+1) = h(-1) = 1/2$, while, for $(k-1)\tanh\beta > 1$, it admits tree distinct solutions.

A general construction

A general construction of messages in $\mathcal{H}^*_\nu$ is as follows. As in the finite case, for T a tree, and $(u, v) \in E$, we let $T_{u\to v}$ be the subtree induced by the vertices that can be erached from u without passing through v. Let $f_1 : \mathcal{X} \to \mathbb{R}_{\ge 0}$ be a function that is not identically vanishing.

For a random tree $(\boldsymbol{T}, o \to \boldsymbol{J}) \sim \nu^{\downarrow}$ we consider the Gibbs measure $\mu^{f_1,t}_{\boldsymbol{T}_{o\to\boldsymbol{J}},\psi}$ on the subtree $\boldsymbol{T}_{o\to\boldsymbol{J}}$ with factorized boundary conditions $f_1 \times \cdots \times f_1$ at level t. Assume that, as $t \to \infty$, $\mu^{f_1,t}_{\boldsymbol{T}_{o\to\boldsymbol{J}},\psi}$ converges to a limit almost surely with respect to $\boldsymbol{T}$. Set for each $(T, o \to J)$,

$$h_{T,o\to j}(\sigma) \equiv \mu^{f_1}_{T_{o\to J},\psi}(\{\boldsymbol{\sigma}_o = \sigma\})\,. \tag{3.2.11}$$

(Note that the definition is well posed only ν-almost surely.)

The following is easy to prove.

Proposition 3.2.9. *Assume that, for some boundary condition f_1, the function $h : (T, o \to j) \mapsto h_{T,o\to j}$ defined above is measurable. Then $h \in \mathcal{H}^*_\nu$.*

Notice that, if the function h defined here is measurable with respect to $(T, o \to j)$, is indeed measurable with respect to $T_{o\to j}$. Physical intuition suggests that this should be the case in general. We will call such message function *causal*, and denote their space by $\mathcal{H}^{\mathrm{caus}}_\nu$, letting $\mathcal{H}^{\mathrm{caus},*}_\nu \subseteq \mathcal{H}^{\mathrm{caus}}_\nu$ be the subset that solves Eq. (3.2.9).

Open Problem 3.2.1. Prove that, for studying local weak limits of Gibbs measures $(G_n, \mu_{\psi,G_n}))$ on finite graphs, it is sufficient to consider elements of $\mathcal{H}^*$ that are measureable on $T_{o\to j}$.

(To be made more precise.)

The case of Galton-Watson trees

Consider now a unimodular Galton-Watson tree $(\boldsymbol{T}, \boldsymbol{o}) \sim \mathrm{GW}(P)$. Recall that, for $(u, v) \in E(T)$, $T_{u\to v}$ is the subtree of T rooted at u and induced by all the vertices of T that can be reached from u wothout passing through v. For a Galton-Watson tree, it is easy to check that, for $\boldsymbol{J}$ a uniformly random neighbor of the root (conditional on $\partial\boldsymbol{o} \neq \emptyset$) and $\boldsymbol{J}_1, \dots \boldsymbol{J}_K$ the decendants of $\boldsymbol{J}$, we gave the following equalities in distribution

$$\boldsymbol{T}_{\boldsymbol{J}\to\boldsymbol{o}} \stackrel{\mathrm{d}}{=} \boldsymbol{T}_{\boldsymbol{J}_1\to\boldsymbol{J}} \stackrel{\mathrm{d}}{=} \cdots \stackrel{\mathrm{d}}{=} \boldsymbol{T}_{\boldsymbol{J}_k\to\boldsymbol{J}}\,, \tag{3.2.12}$$

and further $\boldsymbol{T}_{\boldsymbol{J}_1\to\boldsymbol{J}}, \boldsymbol{T}_{\boldsymbol{J}_2\to\boldsymbol{J}}, \dots$ are independent identically distributed copies, independent of $K \sim \rho$.

Exercise 3.2.3. Prove the last statements.

Let now $h \in \mathcal{H}_\nu^{\mathrm{caus},*}$ be a causal solution of Eq. (3.2.9). Let $\boldsymbol{h} = h_{(\boldsymbol{T},\boldsymbol{J}\to\boldsymbol{o})}$ be a random variable whose distribution is the same as the one of h when the underlying tree is $(\boldsymbol{T}, \boldsymbol{o}) \sim \mathrm{GW}(P)$. Then Eq. (3.2.9) together with the above distributional properties imply

$$\boldsymbol{h} \stackrel{\mathrm{d}}{=} \mathsf{BP}_{\psi,K}(\boldsymbol{h}_1, \dots, \boldsymbol{h}_K)\,, \tag{3.2.13}$$

where $\boldsymbol{h}_1, \boldsymbol{h}_2, \dots$ are i.i.d. copies of $\boldsymbol{h}$, independent of $K \sim \rho$. We further defined, with a slight abuse of notation, the mapping $\mathsf{BP}_{\psi,k} : (\Delta_{\mathcal{X}})^k \to \Delta_{\mathcal{X}}$ by letting, for all $\sigma \in \mathcal{X}$,

$$\mathsf{BP}_{\psi,k}(h_1, \dots, h_k)(\sigma) \equiv \frac{1}{z}\psi_{\mathrm{v}}(\sigma) \prod_{l=1}^{k} \Big(\sum_{\sigma_l \in \mathcal{X}} \psi_{\mathrm{e}}(\sigma, \sigma_l)\, h_l(\sigma_l) \Big)\,. \tag{3.2.14}$$

Equation (3.2.13) is known in statistical physics as the *replica symmetric cavity equation*. It is related to the *density evolution* method from coding theory. Relations with the latter will be further discussed in Chapter ???.

In general, a solution of the distributional equation (3.2.13) does not yield –by itself- a measurable function $h : \mathcal{T}_{\mathrm{e}} \to \Delta_{\mathcal{X}}$. Such a function can be constructed for instance, using the general method in the previous section.

Note howver that a solution of Eq. (3.2.13) can be used to construct a joint distribution of $\boldsymbol{h}$ and $\boldsymbol{T}_{\boldsymbol{J}\to\boldsymbol{o}}$ as follows. (Discussion of joint distribution to be added.)

Exercise 3.2.4. How does Eq. (3.2.13) generalize to the case of multi-type Galton-Watson trees?

3.2.3 Bethe-Peierls free energy

Definitions

We want now to define the Bethe-Peierls free energy of a (possibly infinite) random rooted tree $(\boldsymbol{T}, \boldsymbol{o})$, with specification *psi*. We hope that this formula will serve as a possible limit value for $\phi(G_n; \psi)$ when $\{G_n\}$ is a graph sequence that converges locally to $(\boldsymbol{T}, \boldsymbol{o})$.

Of course we cannot use Definition 3.1.4 because this will typically yield infinite values. A natural idea would be the following. Truncate the tree $(\boldsymbol{T}, \boldsymbol{o})$ after a finite number of generations t, thus obtaining the finite graph $B_t(o; \underline{\ })$. Evaluate the corresponding Bethe free

energy $\Phi_{B_t(\boldsymbol{o};\boldsymbol{T}),\psi}$, and define the free enrgy density of $\boldsymbol{T}$ as the limit of $\Phi_{B_t(\boldsymbol{o};\boldsymbol{T}),\psi}/|B_t(o;\boldsymbol{T})|$ as $t \to \infty$.

It is important to stress that this recipe is *wrong* except for special cases. The underlying reason is the finite graph $B_t(o;\boldsymbol{T})$ is dominated by its boundary (again, except 'special cases'). and therefore the free energy $\Phi_{B_t(\boldsymbol{o};\boldsymbol{T}),\psi}$ is dominated by the contributions of vertices that are within $O(1)$ distance from the boundary, i.e. with $d(\boldsymbol{o}, v) \geq t - O(1)$.

Exercise 3.2.5. Try to apply the above recipe to some simple model (e.g. the feroomagnetic Ising model) and convince yourself that it is incorrect.

Definition 3.2.10. *Let $(\boldsymbol{T}, \boldsymbol{o})$ be unimodular random rooted tree, with law ν and ψ a specification. The* Bethe-Peierls free energy *of $(\boldsymbol{T}, \boldsymbol{o})$, ψ a function $\Phi_{\nu,\psi} : \mathcal{H}_\nu \to \mathbb{R}$ defined by*

$$\Phi_{\nu,\psi}(h) \equiv \mathbb{E}\Phi_{\boldsymbol{T},\psi}(h)\,, \tag{3.2.15}$$

$$\Phi_{T,\psi}(h) \equiv \Phi^{\mathrm{v}}_{T,\psi}(h) - \sum_{j\in\partial o} \Phi^{\mathrm{e}}_{T,\psi}(h;j)\,, \tag{3.2.16}$$

where

$$\Phi^{\mathrm{v}}_{T,\psi}(h) \equiv \log\Big\{ \sum_{\sigma\in\mathcal{X}} \psi_{\mathrm{v}}(\sigma) \prod_{j\in\partial o} \Big(\sum_{\sigma_j\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma,\sigma_j)\, h_{T,j\to o}(\sigma_j)\Big)\Big\}\,, \tag{3.2.17}$$

$$\Phi^{\mathrm{e}}_{T,\psi}(h;j) \equiv \log\Big\{ \sum_{\sigma_1,\sigma_2\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma_1,\sigma_2)\, h_{T,o\to j}(\sigma_1)\, h_{T,j\to o}(\sigma_2)\Big)\Big\}\,. \tag{3.2.18}$$

In order to justify this definition, consider again Definition 3.1.4. If $h_{i\to j}$ is just a function of $B_t(i\to j; G_n)$ for some finite t, then

$$\frac{1}{n}\Phi_{G_n,\psi} = \frac{1}{n}\sum_{i\in V_n} \Phi_{T=B(i;G_n),\psi}(h) \overset{n\to\infty}{\to} \Phi_{\nu,\psi}(h)\,. \tag{3.2.19}$$

Definition 3.2.10 extends the latter to arbitrary measurable functions h.

The Bethe-Peierls free-energy prediction

The so called *Bethe-Peierls prediction* for the free energy consists in postulating that, for a sequence of graphs $\{G_n\}$ converging localy to a random rooted tree $(\boldsymbol{T}, \boldsymbol{o})$, with law ν,

$$\lim_{n\to\infty} \phi_n(G_n;\psi) = \sup_{h\in\mathcal{H}^*_\nu} \Phi_{\nu,\psi}(h)\,. \tag{3.2.20}$$

This is also caled by physicists, the *(cavity) replica symmetric prediction*. In the next chapters we will review some cases in which this prediction has been proved to hold. It is important to keep in mind that this is not always the case. In particular, the Bethe-Peierls prediction only depends on the limiting random tree. On the other hand, there are simple examples of graph sequences that have the same local limit but different limiting free-energy densities.

Open Problem 3.2.2. Characterize the graph sequences $\{G_n\}$ and specifications ψ for which the Bethe-Peierls prediction holds.

Dual Bethe free energy

It is also possible to develop an limit version of the dual free energy. First of all, we introduce a version of the local polytope.

Definition 3.2.11. *Given a random rooted tree $(\boldsymbol{T}, \boldsymbol{o})$ with law ν, the* local polytope $\mathcal{H}^{\mathrm{loc}}_{\nu}$ *is the set of measurable functions*

$$b : \mathcal{T}_{\mathrm{e}} \to \Delta_{\mathcal{X}\times\mathcal{X}} \tag{3.2.21}$$

$$(T, u \to v) \mapsto b_{(T,u\to v)} = (b_{(T,u\to v)}(\sigma_u, \sigma_v))_{\sigma_u,\sigma_v\in\mathcal{X}}\,, \tag{3.2.22}$$

that satisfy the following properties

1. $b_{(T,u\to v)}(\sigma_u, \sigma_v) = b_{(T,v\to u)}(\sigma', \sigma)$ *for all* $\sigma_u, \sigma_v \in \mathcal{X}$.
2. *Define the marginal*

$$b_{(T,u)}(\sigma) \equiv \sum_{\sigma_v\in\mathcal{X}} b_{(T,u\to v)}(\sigma, \sigma_v)\,. \tag{3.2.23}$$

Then $(T, u) \mapsto b_{(T,u)}$ is a well defined function on $\mathcal{T}_*$, taking values in $\Delta_{\mathcal{X}}$.

We will consider two such functions equivalent if they coincide $\nu^{\uparrow}$-almost everywhere.

Explicitly, the second condition above means the following. If $(T_1, u_1 \to v_1)$ and $(T_2, u_2 \to v_2)$ are equivalent up to a root preserving isomorphism (an isomorphism φ form T_1 to T_2 such that $\varphi(u_1) = u_2$ but possibly $\varphi(v_1) \neq v_2$), then $\sum_{\sigma'\in\mathcal{X}} b_{(T_1,u_1\to v_1)}(\sigma, \sigma') = \sum_{\sigma'\in\mathcal{X}} b_{(T_2,u_2\to v_2)}(\sigma, \sigma')$.

In the following, for $b \in \mathcal{H}^{\mathrm{loc}}_{\nu}$, we will use the shorthand $b_{u,v}(\sigma, \sigma')$ instead of $b_{(T,u\to v)}(\sigma, \sigma')$ and b_u instead of $b_{(T,u)}$, where the latter is defined as per Eq. (3.2.23).

Definition 3.2.12. *Let $(\boldsymbol{T}, \boldsymbol{o})$ be a unimodular random rooted tree with law ν and ψ a specification. The* (dual) Bethe free energy *is a function $\Phi^*_{\nu,\psi} : \mathcal{H}^{\mathrm{loc}}_{\nu} \to \mathbb{R}$ defined by*

$$\Phi^*_{\nu,\psi}(b) = \mathbb{E}\Phi^*_{\boldsymbol{T},\psi}(b)\,, \tag{3.2.24}$$

where, for $(T, o) \in \mathcal{T}_{\mathrm{e}}$,

$$\Phi^*_{T,\psi}(b) = \mathbb{E}_{b_o} \log \psi_{\mathrm{v}}(\boldsymbol{\sigma}_o) + \frac{1}{2}\sum_{j\in\partial o} \mathbb{E}_{b_{o,j}} \log \psi_{\mathrm{e}}(\boldsymbol{\sigma}_o, \boldsymbol{\sigma}_j) + H(b_o) + \frac{1}{2}\sum_{j\in\partial o} H(b_{o,j}\|b_o \times b_j)\,. \tag{3.2.25}$$

As in the finite graph case, there exists a natural embedding of $\mathcal{H}^*_{\nu}$ into $\mathcal{H}^{\mathrm{loc}}_{\nu}$.

Proposition 3.2.13. *Let $(\boldsymbol{T}, \boldsymbol{o})$ be a unimodular random rooted tree with law ν and $h \in \mathcal{H}^*_{\nu}$. Define, for $(T, u \to v) \in \mathcal{T}_{\mathrm{e}}$,*

$$b_{(T,u\to v)}(\sigma_u, \sigma_v) \equiv \frac{1}{z}\, h_{(T,u\to v)}(\sigma_u)\psi_{\mathrm{e}}(\sigma_u, \sigma_v)h_{(T,v\to u)}(\sigma_v)\,. \tag{3.2.26}$$

Then $b \in \mathcal{H}^{\mathrm{loc}}_{\nu}$ with $\Phi^*_{\nu,\psi}(b) = \Phi_{\nu,\psi}(h)$

Proof. To be written. □

Finally, we have the following limit version of Theorem ????.

Theorem 4. *Let ψ be permissive specification and $(\boldsymbol{T}, \boldsymbol{o})$ a random rooted tree with law ν such that $\mathbb{E}[\deg(\boldsymbol{o})] < \infty$. Then*

1. *Any local maximizer b of $\Phi^*_{\nu,\psi}$ is such that, for $\boldsymbol{J}$ a uniformly random neighbor of $\boldsymbol{o}$,*

$$\mathbb{P}_\nu\Big(\mathrm{supp}(b_{\boldsymbol{o},\boldsymbol{J}}) = \mathrm{supp}(\psi_{\mathrm{e}})\Big) = 1\,. \tag{3.2.27}$$

2. *Any stationary point b of $\Phi^*_{\nu,\psi}$ that satisfies the latter condition (3.2.27) can be written as the image of some $h \in \mathcal{H}^*_\nu$ threough the embedding (3.2.26).*

3. *As a consequence of the last two points, if the supremum on the right-hand side is attained in $\mathcal{H}^{\mathrm{loc}}_\nu$, we have*

$$\sup_{h\in\mathcal{H}^*_\nu} \Phi_{\nu,\psi}(h) = \sup_{b\in\mathcal{H}^{\mathrm{loc}}_\nu} \Phi^*_{\nu,\psi}(b)\,. \tag{3.2.28}$$

Proof. To be written. □

Corollary 3.2.14. *Assume that $\psi = \psi^\theta$ with $\theta \mapsto \psi^\theta$ continuously differentiable for $\theta \in Q$ an open domain $\mathbb{R}^k$. Further assume that $\mathrm{supp}(\psi^\theta_{\mathrm{e}})$ is independent of $\theta \in Q$. Define*

$$\Phi_\nu(\theta) \equiv \sup_{h\in\mathcal{H}^*_\nu} \Phi_{\nu,\psi^\theta}(h). \tag{3.2.29}$$

If, for $\psi = \psi^\theta$ $\sup_{b\in\mathcal{H}^{\mathrm{loc}}_\nu} \Phi^*_{\nu,\psi}(b)$ is achieved for some $b \in \mathcal{H}^{\mathrm{loc}}_\nu$, then Φ_ν is continuous at θ.

Proof. To be written. □

The case of regular trees

Assume that $\{G_n\}$ is a graph sequence converging locally to the k-regular rooted $T^{\mathrm{reg},k}$. As we saw, we can take $h \in \mathcal{H}_\nu$ to be constant over $\mathcal{T}_{\mathrm{e}}$, and hence $\mathcal{H}_\nu$ is equivalent to the simplex $\Delta_{\mathcal{X}}$. Hence $\Phi_{\nu,\psi} : \mathcal{H}_\nu\mathbb{R}$ is equivalent to a function $\Phi_{k,\psi} : \Delta_{\mathcal{X}} \to \mathbb{R}$

$$\Phi_{k,\psi}(h) \equiv \log\Big\{ \sum_{\sigma\in\mathcal{X}} \psi_{\mathrm{v}}(\sigma)\Big(\sum_{\sigma'\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma,\sigma')\, h(\sigma')\Big)^k\Big\} - \frac{k}{2}\log\Big\{ \sum_{\sigma_1,\sigma_2\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma_1,\sigma_2)\, h(\sigma_1)\, h(\sigma_2)\Big)\Big\}\,. \tag{3.2.30}$$

Analogously, the local polytope can be identified in this case with

$$\mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k} \equiv \Big\{\, b \in \Delta_{\mathcal{X}\times\mathcal{X}} :\, b(\sigma,\sigma') = b(\sigma',\sigma) \forall \sigma,\sigma' \in \mathcal{X}\,\Big\}\,. \tag{3.2.31}$$

Hence the dual free energy is identified with $\Phi^*_{k,\psi} : \mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k} \to \mathbb{R}$ given by

$$\Phi^*_{k,\psi}(b) = \mathbb{E}_{b_1} \log \psi_{\mathrm{v}}(\boldsymbol{\sigma}) + \mathbb{E}_b \log \psi_{\mathrm{e}}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) + H(b_1) + \frac{k}{2}\, H(b\|b_1\times b_1)\,. \tag{3.2.32}$$

It is interesting to consider the connection with large-deviations theory. Assume $\psi_{\mathrm{v}}(\sigma) = 1$ and $\psi_{\mathrm{e}}(\sigma_1,\sigma_2) = e^{\xi(\sigma_1,\sigma_2)}$. Let

$$\boldsymbol{L}_n = \frac{1}{n}\sum_{(i,j)\in E_n} (\delta_{(\boldsymbol{\sigma}_i,\boldsymbol{\sigma}_j)} + \delta_{(\boldsymbol{\sigma}_j,\boldsymbol{\sigma}_i)})\,, \tag{3.2.33}$$

be the empitical distribution of the edge spins. Clearly, $\boldsymbol{L}_n$ takes values in $\Delta_{\mathcal{X}\times\mathcal{X}}$. Then it is easy to see that

$$\phi(G_n,\psi) = \log|\mathcal{X}| + \log \mathbb{E}\big(e^{\langle \xi, \boldsymbol{L}_n\rangle}\big)\,, \tag{3.2.34}$$

where E denotes expectation with respect to the uniform measure on $\mathcal{X}^{V_n}$. The dual Bethe free energy corresponds to a specific form of the large deviaton functional of $\boldsymbol{L}_n$.

The case of Galton-Watson trees

Assume that $(\boldsymbol{T},\boldsymbol{o}) \sim \mathrm{GW}(P)$. We saw that, in this case, the distribution of the random variable $\boldsymbol{h} \equiv h_{\boldsymbol{T},\boldsymbol{J}\to\boldsymbol{o}}$ must satisfy the distributional equation (3.2.13). Correspondingly, the Bethe free-energy $\Phi_{\nu,\psi}$ only depends on h through the law of $\boldsymbol{h}$. With a slight abuse of notation $\Phi_{\nu,\psi}(h) = \Phi_{P,\psi}(\mathcal{L}_{\boldsymbol{h}})$ where $\mathcal{L}_{\boldsymbol{h}}$ is the law of the random variable $\boldsymbol{h}$ and

$$\Phi_{P,\psi}(\lambda) \equiv \mathbb{E}_K \mathbb{E}_{\boldsymbol{h}_1,\cdots,\boldsymbol{h}_K} \log\Big\{ \sum_{\sigma\in\mathcal{X}} \psi_{\mathrm{v}}(\sigma) \prod_{i=1}^{K} \Big(\sum_{\sigma_i\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma,\sigma_i)\, \boldsymbol{h}_i(\sigma')\Big)\Big\} \tag{3.2.35}$$

$$- \frac{1}{2}\mathbb{E}(K)\mathbb{E}_{\boldsymbol{h}_1,\boldsymbol{h}_2} \log\Big\{ \sum_{\sigma_1,\sigma_2\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma_1,\sigma_2)\, \boldsymbol{h}_1(\sigma_1)\, \boldsymbol{h}_2(\sigma_2)\Big)\Big\}\,. \tag{3.2.36}$$

Here expectations are with respect to $K \sim P$ and $\boldsymbol{h}_1, \boldsymbol{h}_2, \dots,$ i.i.d. with common distribution λ.

3.3 Beyond Bethe approximation

Chapter 4

The second moment method

The moment methods aims at characterizing the asymptotic free energy density $\lim_{n\to\infty}\phi(\boldsymbol{G}_n,\psi)$ for sequences of random graph $\{\boldsymbol{G}_n\}_{n\ge 0}$. Ideally one would try to compute –say– the expectaton $\mathbb{E}\phi(\boldsymbol{G}_n,\psi)$ but this is often as hard as the original task. The basic observation is that it is simpler to compute moments of the partition function $Z(\boldsymbol{G}_n,\psi)$. In particular, if the second moment is roughly the square of the first moment, one can conclude that $\phi(\boldsymbol{G}_n,\psi)\approx(1/n)\log\mathbb{E}Z(\boldsymbol{G}_n,\psi)$.

While the basic idea is very simple, the moment calculations are very insightful. Further, it is instructive to undestand the cases in which the second moment approach breaks down, namely $\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\}$ becomes exponentially larger than $(\mathbb{E}Z(\boldsymbol{G}_n,\psi))^2$. Finally, various refinements have been developed of the basic second moment approach.

This chapter is organized as follows. Section 4.1 lays down the general foundations of the second moment method. Section 4.2 carries out the moment calculation for random regular graphs. Section 4.3 discusses two applications of the second moment method. The first one is successful and allows to compute the asymptotic free energy of ferromagnetic Ising models on random regular graphs. The second is only partially successful, and concerns the independent sets model. We then discuss the fundamental reasons for this failure. We conclude in Section 4.4 by introducing two techniques that refine the second moment method and allow to overcome some of its failures.

4.1 General approach

Our starting point is the following general bound, which is an immediate application of Paley-Ziegmund inequality (which in turn is a simple corollary of Cauchy-Schwarz inequality).

Lemma 4.1.1. *Let $\{\boldsymbol{G}_n\}_{n\ge 0}$ be a sequence of random graphs anf ψ a specification. Define $\phi^{\rm ann}(n,\psi)\equiv(1/n)\log\mathbb{E}Z(\boldsymbol{G}_n,\psi)$. Then, letting $\gamma_0=4\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\}/\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)\}^2$, we have*

$$\mathbb{P}\Big\{|\phi(\boldsymbol{G}_n,\psi)-\phi^{\rm ann}(n,\psi)|\le\frac{1}{n}\log\gamma_0\Big\}\ge\frac{1}{\gamma_0} \tag{4.1.1}$$

Proof. Let $\boldsymbol{Z}=Z(\boldsymbol{G}_n,\psi)$. The probability that $\phi(\boldsymbol{G}_n,\psi)\ge\phi^{\rm ann}(n,\psi)+(\log\gamma)/n$ is upper bounded by Markov inequality. For $\gamma>1$;

$$\mathbb{P}(\boldsymbol{Z}\ge\gamma\mathbb{E}\boldsymbol{Z})\le\frac{1}{\gamma}\,. \tag{4.1.2}$$

The probabilty that $\phi(\boldsymbol{G}_n, \psi) \le \phi^{\rm ann}(n, \psi) - (\log \gamma)/n$ is instead upper bounded by Paley-Ziegmund inequality[1]:

$$\mathbb{P}(\boldsymbol{Z} \le \gamma^{-1}\mathbb{E}\boldsymbol{Z}) \le 1 - (1-\gamma^{-1})^2 \frac{(\mathbb{E}\boldsymbol{Z})^2}{\mathbb{E}(\boldsymbol{Z}^2)}\,. \tag{4.1.3}$$

The claim follows using union bound on Eqs. (4.1.2), (4.1.3), and selecting $\gamma = \gamma_0$ as prescribed. □

This Lemma implies that $\phi(\boldsymbol{G}_n, \psi)$ is close to $\phi^{\rm ann}(n, \psi)$ with probability of order $\mathbb{E}(\boldsymbol{Z})^2/\mathbb{E}(\boldsymbol{Z}^2)$. We use concentration of measure to boost this probability.

Proposition 4.1.2. *Assume that $e^{-\xi} \le \psi_{\rm e}(\sigma, \sigma') \le e^{\xi}$ for all $\sigma, \sigma' \in \mathcal{X}$. If $\boldsymbol{G}_n$ is a uniformly random k-regular graph then, for any $t > 0$*

$$\mathbb{P}\big\{|\phi(\boldsymbol{G}_n, \psi) - \mathbb{E}\phi(\boldsymbol{G}_n, \psi)| \ge t\big\} \le 2\, e^{-nt^2/(64k\xi)}\,. \tag{4.1.4}$$

Proof. Let $m = nk/2$ denote the number of edges. With a slignt abuse of notation, we let $\boldsymbol{G}_n$ denote the graph in which edges have been labeled, for instance uniformly at random. We also let $\boldsymbol{G}_n(\ell)$ be the subgraph induced by the first ℓ edges and define the random variables $(\boldsymbol{X}_\ell)_{0\le\ell\le m}$ by

$$\boldsymbol{X}_\ell \equiv \mathbb{E}\{\log Z(\boldsymbol{G}_n, \psi) | \boldsymbol{G}_n(\ell)\}\,. \tag{4.1.5}$$

Note that $(\boldsymbol{X}_\ell)_{0\le\ell\le m}$ is a martingale. The claim follows from Azuma-Hoeffding inequality if we can prove that, almost surely

$$|\boldsymbol{X}_\ell - \boldsymbol{X}_{\ell-1}| \le 8\xi\,. \tag{4.1.6}$$

We have

$$\begin{aligned}
|\boldsymbol{X}_\ell - \boldsymbol{X}_{\ell-1}| &= \big|\mathbb{E}\{\log Z(\boldsymbol{G}_n, \psi)|\boldsymbol{G}_n(\ell)\} - \mathbb{E}\{\log Z(\boldsymbol{G}_n, \psi)|\boldsymbol{G}_n(\ell-1)\}\big| && (4.1.7)\\
&\le \max_{(G_n(\ell), G'_n(\ell))\in P} \big|\mathbb{E}\{\log Z(\boldsymbol{G}_n, \psi)|\boldsymbol{G}_n(\ell) = G_n(\ell)\} && (4.1.8)\\
&\qquad - \mathbb{E}\{\log Z(\boldsymbol{G}_n, \psi)|\boldsymbol{G}_n(\ell) = G'_n(\ell)\}\big| && (4.1.9)
\end{aligned}$$

where the maximum is over the graph pairs $(G_n(\ell), G'_n(\ell))$ that differ only in one edge, namely the edge with label ℓ. Let $(\boldsymbol{G}_n^{(1)}, \boldsymbol{G}_n^{(2)})$ be a coupling of the conditional law of $\boldsymbol{G}_n$ given $\boldsymbol{G}_n(\ell) = G_n(\ell)$ and the conditional law of $\boldsymbol{G}_n$ given $\boldsymbol{G}_n(\ell) = G'_n(\ell)$. We then have

$$\begin{aligned}
|\boldsymbol{X}_\ell - \boldsymbol{X}_{\ell-1}| &\le \max_{(G_n(\ell), G'_n(\ell))\in P} \big|\mathbb{E}\big\{\log Z(\boldsymbol{G}_n^{(1)}, \psi) - \log Z(\boldsymbol{G}_n^{(2)}, \psi)\big\}\big| && (4.1.10)\\
&\le \xi \max_{(G_n(\ell), G'_n(\ell))\in P} \mathbb{E}\Delta(\boldsymbol{G}_n^{(1)}, \boldsymbol{G}_n^{(2)})\,, && (4.1.11)
\end{aligned}$$

where $\Delta(\boldsymbol{G}_n^{(1)}, \boldsymbol{G}_n^{(2)})$ is the number of edges in which the two graphs differ (formally, the cardinality of the symmetric difference of the edge sets), and the last bound follows from Lemma 2.1.2.

The proof is completed by showing that there exists a coupling $(\boldsymbol{G}_n^{(1)}, \boldsymbol{G}_n^{(2)})$ such that $\Delta(\boldsymbol{G}_n^{(1)}, \boldsymbol{G}_n^{(2)}) \le 8$. In order to prove this claim, consider the unique edge in which

[1] This follows from Cauchy-Schwarz since $(1-\gamma^{-1})\mathbb{E}\boldsymbol{Z} \le \mathbb{E}\{\boldsymbol{Z}\mathbb{I}(\boldsymbol{Z} \ge \gamma^{-1}\mathbb{E}\boldsymbol{Z})\} \le \mathbb{E}(\boldsymbol{Z}^2)^{1/2}\mathbb{P}(\boldsymbol{Z} \ge \gamma^{-1}\mathbb{E}\boldsymbol{Z})^{1/2}$

$G_n(\ell)$ and $G'_n(\ell)$ differ, and assume first that this edge is (i,j) (for $G_n(\ell)$) and (i,j') (for $G'_n(\ell)$): i.e. that the edge is incident on a common vertex i in the two graph. In order to construct the coupling, let (i',j') the lowest order edge in $\boldsymbol{G}_n^{(1)}$ after the ℓ-th one, that is incident on j' (check that such an edge exists indeed). In other words $E(\boldsymbol{G}_n^{((1)}) = E(G_n(\ell-1))\cup\{(i,j),(i',j')\}\cup E'$. We then construct $\boldsymbol{G}_n^{(2)}$ by letting $E(\boldsymbol{G}_n^{((1)}) = E(G_n(\ell-1))\cup\{(i,j'),(i',j)\}\cup E'$. The reader can check that this is indeed distributed uniformly conditional on $G'_n(\ell)$. Of course we have $\Delta(\boldsymbol{G}_n^{(1)},\boldsymbol{G}_n^{(2)})\le 4$ almost surely.

The general case in $G_n(\ell)$ and $G'_n(\ell)$ differ in an arbitrary vertex can be treated by triangular inequality. □

Theorem 5. *Let $\{\boldsymbol{G}_n\}$ be a sequence of random regular graphs from the configuration model, and ψ a specification with $\psi_{\mathrm{e}}(\sigma,\sigma')>0$ for all $\sigma,\sigma'\in\mathcal{X}$. Further assume that*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}Z(\boldsymbol{G}_n,\psi)=\lim_{n\to\infty}\frac{1}{2n}\log\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\}=\phi\,. \tag{4.1.12}$$

Then, almost surely

$$\lim_{n\to\infty}\phi(\boldsymbol{G}_n,\psi)=\phi\,. \tag{4.1.13}$$

Proof. Follows immediately from the previous Lemma and Proposition. □

Exercise 4.1.1. Generalize Proposition 4.1.2 to permissive specifications.

Exercise 4.1.2. Generalize Proposition 4.1.2 to configuration-model random graphs with given degree distribution.

We state a generalization of the above that follows immediately from the solution of Exercises 4.1.1 and 4.1.2.

Theorem 6. *Let P be a probability distribution over $\mathbb{N}$ with $\mathbb{E}_P\{K\}<\infty$ and, for each n, let $\boldsymbol{G}_n$ be a random graph from the configuration model with degree distribution P. Further assume that ψ is a permissive specification.*

Then the conclusion of Theorem 5, namely Eq. (4.1.12) implies Eq. (4.1.13).

4.2 The case of random regular graphs

While the second moment method can be very general, it works best when the random graph has little variability. Also, moment calculations becomes more difficult for more complicate graph distributions. We shall therefore assume hereafter that $\boldsymbol{G}_n$ is a random regular graph of degree k from the configuration model. Note that such set is non-empty if and only if kn is even, which will be assumed thoughout.

As often, rather than using uniformly random regular graphs, is much more convenient to work with regular graphs of degree k distributed according to the configuration model, as introduced in Section 2.3.2. By proposition 2.3.3, any property that holds with high proability for the configuration model, holds with high probability for the uniform model as well.

4.2.1 First moment

The application of the second moment to random regular method is based on a combinatorial calculation that we will carry out in some generality in Lemma 4.2.1 below.

Before proceeding, let us introduce some combinatorial notation. For $2M$ an even integer, we denote by $\mathfrak{P}(2M)$ the number of pairings of $2M$ objects, namely:

$$\mathfrak{P}(2M) = \frac{(2M)!}{2^M(M!)}\,. \tag{4.2.1}$$

Note that there are exactly $\mathfrak{P}(kn)$ configuration-model graphs over n vertices V_n with uniform degree k. Indeed, these graphs are in one-to-one correspondence with the set of pairings of $2|E_n|$ objects, and $2|E_n| = kn$ in the case of regular graphs.

Given an integer N, and a function $M : \mathcal{S} \to \mathbb{N}$ defined on a finite set $\mathcal{S}$, with $\sum_{s\in\mathcal{S}} M(s) = N$, we define the binomial coefficient

$$\binom{N}{M(\,\cdot\,)} \equiv \frac{N!}{\prod_{s\in\mathcal{S}} M(s)!} \tag{4.2.2}$$

Lemma 4.2.1. *Let $\boldsymbol{G}_n$ be a configuration-model random regular graph of degree k, and ψ a specification. Then*

$$\mathbb{E}Z(\boldsymbol{G}_n,\psi) = \frac{1}{\mathfrak{P}(kn)} \sum_{m\in\mathcal{M}_{n,k}} Z_{n,k}(m) \prod_{\sigma,\sigma'\in\mathcal{X}} \psi_{\mathrm{e}}(\sigma,\sigma')^{m(\sigma,\sigma')/2} \prod_{\sigma\in\mathcal{X}} \psi_{\mathrm{v}}(\sigma)^{m_1(\sigma)}\,, \tag{4.2.3}$$

where the sum is over the set $\mathcal{M}_{n,k}$ of matrices $m = (m(\sigma,\sigma'))_{\sigma,\sigma'\in\mathcal{X}} \in \mathbb{N}^{\mathcal{X}\times\mathcal{X}}$ satisfying:

1. *$m(\sigma,\sigma') = m(\sigma',\sigma)$ for all σ, $\sigma' \in \mathcal{X}$, and $m(\sigma,\sigma)$ is even.*
2. *Defining*

$$m_1(\sigma) = \sum_{\sigma'} m(\sigma,\sigma')\,, \tag{4.2.4}$$

we have $m_1(\sigma) \in k\mathbb{N}$ and $\sum_{\sigma\in\mathcal{X}} m_1(\sigma) = nk$.

Finally

$$Z_{n,k}(m) \equiv \binom{n}{m_1(\,\cdot\,)/k} \prod_{\sigma\in\mathcal{X}} \binom{m_1(\sigma)}{m(\sigma,\,\cdot\,)} \prod_{\sigma\in\mathcal{X}} \mathfrak{P}(m(\sigma,\sigma)) \prod_{(\sigma,\sigma')} m(\sigma,\sigma')!\,, \tag{4.2.5}$$

where the product is over unordered pairs (σ,σ') with $\sigma \neq \sigma'$.

Proof. In order to simplify notation, we will drop the subscript n from $\boldsymbol{G}_n$, and similar.

The factor $1/\mathfrak{P}(kn)$ is the probability that $\boldsymbol{G}$ takes any specific value. We therefore need to prove that the remaining sum is the sum over all the (multi-)graphs G in the configuraton model, of the partition function $Z(G,\psi)$.

Given a (multi-)graph $G = (V,E)$, and a spin configuration $\sigma \in \mathcal{X}^V$, we let, for $\sigma_1,\sigma_2 \in \mathcal{X}$

$$m_{\sigma,E}(\sigma_1,\sigma_2) = \sum_{(i\to j)\in D(G_n)} \mathbb{I}\big((\sigma_i,\sigma_j) = (\sigma_1,\sigma_2)\big)\,. \tag{4.2.6}$$

Here it is understood that, for each self-loop (i,i), the corresponding edge $(i \to i)$ is counted twice. For each normal edge (i,j), counted with its multiplicity, we have two directed edges $(i \to j)$ and $(j \to i)$. Note that $m_{\sigma,E}(\,\cdot\,,\,\cdot\,)$ satisfies both conditions 1 and 2 in the statement, i.e. $m_{\sigma,E} \in \mathcal{M}_{n,k}$. Let $\mathcal{C}_{n,k}$ be the set og graphs in the configuration model, and define

$$\mathcal{C}_{n,k}(m) \equiv \big\{G,\sigma):\ \ G \in \mathcal{C}_{n,k}\,,\ \ \sigma \in \mathcal{X}^V\,,\ \ m_{E,\sigma} = m\big\}\,, \tag{4.2.7}$$

$$Z_{n,k}(m) \equiv \big|\mathcal{C}_{n,k}(m)\big|\,. \tag{4.2.8}$$

For any $(G,\sigma) \in \mathcal{C}_{n,k}(m)$, we have

$$\prod_{(i,j)\in E} \psi_{\rm e}(\sigma_i,\sigma_j) \prod_{i\in V} \psi_{\rm v}(\sigma_i) = \prod_{\sigma,\sigma'\in\mathcal{X}} \psi_{\rm e}(\sigma,\sigma')^{m(\sigma,\sigma')/2} \prod_{\sigma\in\mathcal{X}} \psi_{\rm v}(\sigma)^{m_1(\sigma)}\,. \tag{4.2.9}$$

Therefore

$$\mathbb{E}Z(\boldsymbol{G}_n,\psi) = \frac{1}{\mathfrak{P}(kn)} \sum_{G\in\mathcal{C}_{n,k}} \sum_{\sigma\in\mathcal{X}^V} \prod_{(i,j)\in E} \psi_{\rm e}(\sigma_i,\sigma_j) \prod_{i\in V} \psi_{\rm v}(\sigma_i) \tag{4.2.10}$$

$$= \frac{1}{\mathfrak{P}(kn)} \sum_{m\in\mathcal{M}_{n,k}} \sum_{(G,\sigma)\in\mathcal{C}_{n,k}(m)} \prod_{(i,j)\in E} \psi_{\rm e}(\sigma_i,\sigma_j) \prod_{i\in V} \psi_{\rm v}(\sigma_i) \tag{4.2.11}$$

$$= \frac{1}{\mathfrak{P}(kn)} \sum_{m\in\mathcal{M}_{n,k}} \big|\mathcal{C}_{n,k}(m)\big| \prod_{\sigma,\sigma'\in\mathcal{X}} \psi_{\rm e}(\sigma,\sigma')^{m(\sigma,\sigma')/2} \prod_{\sigma\in\mathcal{X}} \psi_{\rm v}(\sigma)^{m_1(\sigma)}\,. \tag{4.2.12}$$

We are left with the task of showing that $|\mathcal{C}_{n,k}(m)| = Z_{n,k}(m)$ is given by Eq. (4.2.5). In words, an element of $\mathcal{C}_{n,k}(m)$ is a graph/spin configuration pair, whereby exactly $m_1(\sigma)/k$ vertices i have spin $\sigma_i = \sigma$ for each $\sigma \in \mathcal{X}$, and $m(\sigma,\sigma')$ directed edges have endpoints $\sigma_i = \sigma$, $\sigma_j = \sigma'$. The formula (4.2.5) corresponds to the construction of a such a pair:

- We first gave to assign spin values to the vertices. The first multinomial coefficient counts the distinct ways of doing this.
- For each $\sigma \in \mathcal{X}$, and each half-edge leaving a vertex with spin σ (there is $m_1(\sigma)$) we have to assign it a second spin value $\sigma' \in \mathcal{X}$. The corresponding directed edge will be of type (σ,σ'). The number of ways of doing this is given by the second product of multinomial coefficients.
- Finally we have to pair half-edges, consistently with the spin value assignments. If the spin values are distinct (σ,σ'), there is $m(\sigma,\sigma')!$ of doing that. If they coincide (σ,σ), then the number of ways is $\mathfrak{P}(m(\sigma,\sigma))$.

□

The sum in Eq. (4.2.3) can be evaluated by the saddle point method. The result can be stated in terms of the dual Bethe free energy for k-regular trees introduced in the previous chapter. This is the function $\Phi^*_{k,\psi} : \mathcal{H}^{\rm loc}_{{\rm reg},k} \to \mathbb{R}$ given by Eq. (4.2.13), that we copy here for the reader's convenience

$$\Phi^*_{k,\psi}(b) = \mathbb{E}_{b_1} \log \psi_{\rm v}(\boldsymbol{\sigma}) + \frac{k}{2}\mathbb{E}_b \log \psi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) + H(b_1) + \frac{k}{2}\,H(b\|b_1\times b_1)\,. \tag{4.2.13}$$

Here $\mathcal{H}^{\rm loc}_{{\rm reg},k} \equiv \{\, b \in \Delta_{\mathcal{X}\times\mathcal{X}} :\ b(\sigma,\sigma') = b(\sigma',\sigma)\}$ is the local polytope for the regular tree, and $b_1(\sigma) = \sum_{\sigma'\in\mathcal{X}} b(\sigma,\sigma')$.

Theorem 7. *Let $\boldsymbol{G}_n$ be a configuration-model random regular graph of degree k, and ψ a specification. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} Z(\boldsymbol{G}_n, \psi) = \max_{b\in\mathcal{H}^{\rm loc}_{{\rm reg},k}} \Phi^*_{k,\psi}(b)\,. \tag{4.2.14}$$

Proof. First notice that $|\mathcal{M}_{n,k}|$ is polynomial in n. Therefore, to leading exponential order

$$\mathbb{E} Z(\boldsymbol{G}_n, \psi) = \frac{1}{\mathfrak{P}(kn)} \max_{m\in\mathcal{M}_{n,k}} Z_{n,k}(m;\psi)\,, \tag{4.2.15}$$

$$Z_{n,k}(m;\psi) \equiv Z_{n,k}(m) \prod_{\sigma,\sigma'\in\mathcal{X}} \psi_{\rm e}(\sigma,\sigma')^{m(\sigma,\sigma')/2} \prod_{\sigma\in\mathcal{X}} \psi_{\rm v}(\sigma)^{m_1(\sigma)}\,. \tag{4.2.16}$$

Let $b(\sigma,\sigma') \equiv m(\sigma,\sigma')/(nk)$, $b_1(\sigma) \equiv m_1(\sigma)/(nk)$. Notice that, for $m \in \mathcal{M}_{n,k}$, we have $b \in \mathcal{H}^{\rm loc}_{{\rm reg},k}$. Further, we claim that

$$Z_{n,k}(m;\psi) = \exp\left\{ n\Phi^*_{k,\psi}(b) + o(n) \right\}, \tag{4.2.17}$$

where the $o(n)$ is uniform[2] with respect to $b \in \mathcal{H}^{\rm loc}_{{\rm reg},k}$. Together with Eq. (4.2.15) this implies the desired result (noting that the set $\{b \in \mathcal{H}^{\rm loc}_{{\rm reg},k} : (nk\, b) \in \mathcal{M}_{n,k}\}$ becomes dense in $\mathcal{H}^{\rm loc}_{{\rm reg},k}$ as $n \to \infty$).

The last claim (4.2.17) simply follows by recalling standard approximations of the binomial coefficients. In particular, for any probability distribution p on a finite set $\mathcal{S}$, we have [CT12], uniformly in $p \in \Delta_{\mathcal{S}}$,

$$\binom{N}{N\,p(\,\cdot\,)} = \exp\left\{ N\,H(p) + o(N)\,\right\}. \tag{4.2.18}$$

Further

$$\mathfrak{P}(2N) = \exp\left\{ N\,\log(2N) + o(N)\,\right\}. \tag{4.2.19}$$

Substituting these formulae in Eq. (4.2.15) yiends the desired result. □

Surprisingly, the moment calculation yields as a result the Bethe free enenrgy!

By Markov inequality, we also obtain an upper bound on the almost sure limit of the free energy density.

Corollary 4.2.2. *Let $\boldsymbol{G}_n$ be a configuration-model random regular graph of degree k, or a uniformly random regular graph of degree k, and ψ a specification. Then almost surely*

$$\limsup_{n\to\infty} \frac{1}{n} \log Z(\boldsymbol{G}_n, \psi) \le \max_{b\in\mathcal{H}^{\rm loc}_{{\rm reg},k}} \Phi^*_{k,\psi}(b)\,. \tag{4.2.20}$$

Proof. For the configuration model, this follows follows from Theorem 7, together with Markov inequality (4.1.2). For the uniformly random regular graph it follows by the general relation between configuration model and random regular graphs, see Section 2.3.2. □

[2]This means that there exists a function $f(n) = o(n)$ independent of b such that $\exp\{n\Phi^*_{k,\psi}(b) - f(n)\} \le Z_{n,k}(m;\psi) \le \exp\{n\Phi^*_{k,\psi}(b) + f(n)\}$.

Thanks to the correspondence between dual Bethe free energy and Bethe free energy, we can restate the last results in terms of the latter. This is often more convenient for calculations. Recall that in the present case, the Bethe free energy is the function $\Phi_{k,\psi} : \Delta_{\mathcal{X}} \to \mathbb{R}$ given by

$$\Phi_{k,\psi}(h) \equiv \log\Big\{ \sum_{\sigma\in\mathcal{X}} \psi_{\rm v}(\sigma)\Big(\sum_{\sigma'\in\mathcal{X}} \psi_{\rm e}(\sigma,\sigma')\, h(\sigma')\Big)^k\Big\} - \frac{k}{2}\log\Big\{ \sum_{\sigma_1,\sigma_2\in\mathcal{X}} \psi_{\rm e}(\sigma_1,\sigma_2)\, h(\sigma_1)\, h(\sigma_2)\Big\}\,. \tag{4.2.21}$$

Corollary 4.2.3. *Let $\boldsymbol{G}_n$ be a configuration-model random regular graph of degree k, and ψ a specification. Then*

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbb{E} Z(\boldsymbol{G}_n,\psi) = \max_{h\in\mathcal{H}^*_{\psi,k}} \Phi_{k,\psi}(h)\,, \tag{4.2.22}$$

where $\mathcal{H}^*_{\psi,k}\subseteq\Delta_{\mathcal{X}}$ is the set of BP fixed points, i.e. the set of solution of equation (3.2.10), that we copy here for the reader's convenience

$$h(\sigma) = \frac{1}{z}\psi_{\rm v}(\sigma)\left(\sum_{\sigma'\in\mathcal{X}}\psi_{\rm e}(\sigma,\sigma')\,h(\sigma')\right)^{k-1}\,. \tag{4.2.23}$$

4.2.2 Second moment

At first sight, the second moment calculation might appear too involved. Fortunately, it can be reduced to the first moment calculation by a simple remark. Given a specification ψ on $\mathcal{X}$, we define the specification $\psi^{\otimes 2}$ on $\mathcal{X}\times\mathcal{X}$ by letting

$$\psi_{\rm v}^{\otimes 2}((\sigma^{(1)},\sigma^{(2)})) \equiv \psi_{\rm v}(\sigma^{(1)})\psi_{\rm v}(\sigma^{(2)})\,, \tag{4.2.24}$$

$$\psi_{\rm e}^{\otimes 2}((\sigma^{(1)},\sigma^{(2)}),(\tau^{(1)},\tau^{(2)})) \equiv \psi_{\rm e}(\sigma^{(1)},\tau^{(1)})\,\psi_{\rm e}(\sigma^{(2)},\tau^{(2)})\,. \tag{4.2.25}$$

Then

$$Z(G,\psi)^2 = \sum_{\sigma^{(1)},\sigma^{(2)}\in\mathcal{X}^V}\prod_{(i,j)\in E}\psi_{\rm e}(\sigma_i^{(1)},\sigma_j^{(1)})\prod_{i\in V}\psi_{\rm v}(\sigma_i^{(1)}) \tag{4.2.26}$$

$$\cdot\prod_{(i,j)\in E}\psi_{\rm e}(\sigma_i^{(2)},\sigma_j^{(2)})\prod_{i\in V}\psi_{\rm v}(\sigma_i^{(2)}) = Z(G,\psi^{\otimes 2})\,. \tag{4.2.27}$$

In other words, we can write the square of the partition function of a factor model as the partition function a new factor model with the same graph and a larger alphabet.

Using this remark, the results of the previous section can be 'exported' to the computation of the second moment.

Corollary 4.2.4. *Let $\boldsymbol{G}_n$ be a configuration-model random regular graph of degree k, and ψ a specification. Then*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\} = \max_{b\in\mathcal{H}^{\rm loc}_{{\rm reg},k}(\mathcal{X}\times\mathcal{X})}\Phi^*_{k,\psi^{\otimes 2}}(b) \tag{4.2.28}$$

$$= \max_{h\in\mathcal{H}^*_{\psi^{\otimes 2},k}(\mathcal{X}\times\mathcal{X})}\Phi_{k,\psi^{\otimes 2}}(h)\,, \tag{4.2.29}$$

where we recall that $\mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k}(\mathcal{X}\times\mathcal{X})$ in the first line denotes the local polytope for the alphabet $\mathcal{X}\times\mathcal{X}$. Further $\mathcal{H}^{*}_{\psi^{\otimes 2},k}(\mathcal{X}\times\mathcal{X})\subseteq\Delta_{\mathcal{X}\times\mathcal{X}}$ is the set of BP fixed points for the specification $\psi^{\otimes 2}$, i.e. the set of solution of the equation

$$h(\sigma_1,\sigma_2)=\frac{1}{z}\psi_{\mathrm{v}}(\sigma_1)\psi_{\mathrm{v}}(\sigma_2)\left(\sum_{\sigma_1',\sigma_2'\in\mathcal{X}}\psi_{\mathrm{e}}(\sigma_1,\sigma_1')\psi_{\mathrm{e}}(\sigma_2,\sigma_2')\,h(\sigma_1',\sigma_2')\right)^{k-1}. \tag{4.2.30}$$

In words, the second moment of the partition function is computed by solving a statistical mechanics model whose configurations correspond to two copies of the original model. Following the spin-glass literature, we will call these copies *replicas*.

It is an important to note that the set of BP fixed points for the specification ψ, i.e. $\mathcal{H}^{*}_{\psi,k}$, can be embedded in a natural way in the set of fixed points of $\psi^{\otimes 2}$, $\mathcal{H}^{*}_{\psi^{\otimes 2},k}$.

Proposition 4.2.5. *For $b\in\mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k}(\mathcal{X})$, define $b^{\otimes 2}\in\mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k}(\mathcal{X})$ by letting*

$$b^{\otimes 2}((\sigma^{(1)},\sigma^{(2)}),(\tau^{(1)},\tau^{(2)}))=b(\sigma^{(1)},\tau^{(1)})\,b(\sigma^{(1)},\tau^{(1)})\,. \tag{4.2.31}$$

Then

$$\Phi^{*}_{k,\psi^{\otimes 2}}(b^{\otimes 2})=2\Phi^{*}_{k\psi}(b)\,. \tag{4.2.32}$$

Further, for $h\in\Delta_{\mathcal{X}}$, define $h^{\otimes 2}\in\Delta_{\mathcal{X}\times\mathcal{X}}$ by letting $h^{\otimes 2}(\sigma,\sigma')\equiv h(\sigma)\,h(\sigma')$. If $h\in\mathcal{H}^{*}_{\psi,k}$, then $h^{\otimes 2}\in\mathcal{H}^{*}_{\psi^{\otimes 2},k}$ and

$$\Phi_{k,\psi^{\otimes 2}}(h^{\otimes 2})=2\Phi_{k,\psi}(h)\,. \tag{4.2.33}$$

Proof. The proof is a straightforward calculation by applying the definitions. □

Fixed points of the form $h^{\otimes 2}$ with $h\in\mathcal{H}^{*}_{\psi,k}$ describe a situation in which the two replicas are –asymptotically– independent. The common randomness between the two replicas is not sufficient to create dependence, and therefore $\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\}\approx\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)\}^2$.

Recall that the second moment methos succeeds if the limit of $(1/n)\log\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\}$ is twice the limit of $(1/n)\log\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)\}$. Using Theorem 6, and Corollaries 4.2.3 and 4.2.4, we obtain the following general result.

Theorem 8. *Let $\{\boldsymbol{G}_n\}$ be a sequence of random regular graphs from the configuration model, and ψ a permissive specification. Let $\mathcal{H}^{*}_{\psi,k}$ be the set of BP fixed points and $(\mathcal{H}^{*}_{\psi,k})^{\otimes 2}=\{\,h^{\otimes 2}\in\Delta_{\mathcal{X}\times\mathcal{X}}:\ h\in\mathcal{H}^{*}_{\psi,k}\}$, $(\mathcal{H}^{*}_{\psi,k})^{\otimes 2}\subseteq\mathcal{H}^{*}_{\psi^{\otimes 2},k}$.*

If the maximum of $\Phi_{k,\psi^{\otimes 2}}(h)$ over $\mathcal{H}^{*}_{\psi^{\otimes 2},k}$ is achieved on $(\mathcal{H}^{*}_{\psi,k})^{\otimes 2}\subseteq\mathcal{H}^{*}_{\psi^{\otimes 2},k}$, then almost surely

$$\lim_{n\to\infty}\frac{1}{n}\log Z(\boldsymbol{G}_n,\psi)=\max_{h\in\mathcal{H}^{*}_{\psi,k}}\Phi_{k,\psi}(h)=\max_{b\in\mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k}}\Phi^{*}_{k,\psi}(b)\,. \tag{4.2.34}$$

The same conclusion holds if the maximum of $\Phi^{*}_{k,\psi^{\otimes 2}}(b)$ over $b\in\mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k}(\mathcal{X}\times\mathcal{X})$ is achieved for $b=\tilde{b}^{\otimes 2}$ where $\tilde{b}\in\mathcal{H}^{\mathrm{loc}}_{\mathrm{reg},k}(\mathcal{X})$.

4.2.3 Simplifications

The most difficult part in applying Theorem 8 is tocharacterize the maximizer of $\Phi_{k,\psi^{\otimes 2}}(h)$ or the maximizer of $\Phi^*_{k,\psi^{\otimes 2}}(b)$. The lemma below yields a useful constraint.

We say that $b \in \mathcal{H}^{\rm loc}_{{\rm reg},k}(\mathcal{X}\times\mathcal{X})$ has *positive correlation* if, for any function $f : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$, we have

$$\mathbb{E}_b(f\otimes f) - \mathbb{E}_b(f\otimes 1)\mathbb{E}_b(1\otimes f) \geq 0\,. \tag{4.2.35}$$

Here $f\otimes g\big((\sigma^{(1)},\sigma^{(2)}),(\tau^{(1)},\tau^{(2)})\big) \equiv f(\sigma^{(1)},\tau^{(1)})\cdot g(\sigma^{(2)},\tau^{(2)})$, and 1 represents the function that is identically equal to 1.

Lemma 4.2.6. *There exists a convec combination of maximizer of $\Phi^*_{k,\psi^{\otimes 2}}(b)$ over $b \in \mathcal{H}^{\rm loc}_{{\rm reg},k}(\mathcal{X}\times\mathcal{X})$ that has positive correlation.*

Proof. Let $\mathcal{O}_{{\rm reg},k}(\mathcal{X}\times\mathcal{X}) \subseteq \mathcal{H}^{\rm loc}_{{\rm reg},k}(\mathcal{X}\times\mathcal{X})$ be the convex hull of set of maximizers of $\Phi^*_{k,\psi^{\otimes 2}}(b)$. It follows from continuity of $\Phi^*_{k,\psi^{\otimes 2}}(b)$ that this is a closed set.

Define the shorthand, for $\sigma \in \mathcal{X}^V$,

$$\psi_G(\sigma) \equiv \prod_{(i,j)\in E}\psi_{\rm e}(\sigma_i,\sigma_j)\prod_{i\in V}\psi_{\rm v}(\sigma_i)\,. \tag{4.2.36}$$

Further, for $\sigma^{(1)},\sigma^{(2)} \in \mathcal{X}^V$, we define the empirical edge distribution in analogy to Eq. (4.2.6)

$$m_{\sigma^{(1)},\sigma^{(2)},E}(s,t) = \sum_{(i\to j)\in D(G_n)} \mathbb{I}\big((\sigma^{(1)}_i,\sigma^{(2)}_i)=s\big)\cdot\mathbb{I}\big((\sigma^{(1)}_j,\sigma^{(2)}_j)=t\big)\,, \tag{4.2.37}$$

and introduce the average $b^{(n)} = \big(b(s,t)\big)_{s,t\in\mathcal{X}\times\mathcal{X}}$ by letting

$$b^{(n)} = \frac{1}{(kn)\mathbb{E}\{Z(\boldsymbol{G},\psi)^2\}}\sum_{\sigma^{(1)},\sigma^{(2)}}\mathbb{E}\Big\{\psi_{\boldsymbol{G}}(\sigma^{(1)})\psi_{\boldsymbol{G}}(\sigma^{(2)})\,m_{\sigma^{(1)},\sigma^{(2)},E(\boldsymbol{G})}\Big\}\,. \tag{4.2.38}$$

By construction $b^{(n)} \in \mathcal{H}^{\rm loc}_{{\rm reg},k}(\mathcal{X}\times\mathcal{X})$. We claim that $b^{(n)}$ has positive correlation. Indeed for any two functions $f,\, g : \mathcal{X}\times\mathcal{X}\to\mathbb{R}$, and letting $(i(1),j(1))$ denote the first directed edge in $D(\boldsymbol{G})$,

$$\begin{aligned}\mathbb{E}_{b^{(n)}}\{f\otimes g\} &= \frac{1}{\mathbb{E}\{Z(\boldsymbol{G},\psi)^2\}}\sum_{\sigma^{(1)},\sigma^{(2)}\in\mathcal{X}^n}\mathbb{E}\{\psi_{\boldsymbol{G}}(\sigma^{(1)})\psi_{\boldsymbol{G}}(\sigma^{(2)})f(\sigma^{(1)}_{i(1)},\sigma^{(1)}_{j(1)})g(\sigma^{(2)}_{i(1)},\sigma^{(1)}_{j(2)})\}\\ &= \mathbb{E}'\big\{\mu_{\boldsymbol{G},\psi}\big(f(\boldsymbol{\sigma}^{(1)}_{i(1)},\boldsymbol{\sigma}^{(1)}_{j(1)})\big)\mu_{\boldsymbol{G},\psi}\big(g(\boldsymbol{\sigma}^{(1)}_{i(1)},\boldsymbol{\sigma}^{(1)}_{j(1)})\big)\big\}\end{aligned} \tag{4.2.39}$$

where $\mathbb{E}'$ denotes expectation with respect to the probability measure over graphs $\boldsymbol{G}$ defined by the Radon-Nykodim derivative

$$\frac{{\rm d}\mathbb{P}'}{{\rm d}\mathbb{P}}(G_n) = \frac{Z(G_n,\psi)^2}{\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\}}\,. \tag{4.2.40}$$

We therefore have

$$\mathbb{E}_{b^{(n)}}\{f\otimes f\} - \mathbb{E}_{b^{(n)}}\{f\otimes 1\}\mathbb{E}_{b^{(n)}}\{1\otimes f\} = \mathbb{E}'\big\{\mu_{\boldsymbol{G},\psi}\big(f(\boldsymbol{\sigma}^{(1)}_{i(1)},\boldsymbol{\sigma}^{(1)}_{j(1)})\big)^2\big\} - \mathbb{E}'\big\{\mu_{\boldsymbol{G},\psi}\big(f(\boldsymbol{\sigma}^{(1)}_{i(1)},\boldsymbol{\sigma}^{(1)}_{j(1)})\big)\big\}^2 \geq 0\,. \tag{4.2.41}$$

Since $b^{(n)}$ belongs to a convex set, we have $b^{(n_\ell)} \to b^{(\infty)} \mathcal{H}^{\rm loc}_{{\rm reg},k}(\mathcal{X})$ along some subsequence $\{n_\ell\}$. Further, $b^{(\infty)}$ must be positively correlated. The same proof of Theorem 7 and Corollary 4.2.4 implies that $b^{(\infty)} \in \mathcal{O}_{{\rm reg},k}(\mathcal{X}\times\mathcal{X})$ which imples the claim. □

Next we state a useful condition for the case of multiple maximizers.

Lemma 4.2.7. *Let $(\psi^\beta)_\beta$ be a family of specifications parameterized by $\beta \in \mathbb{R}$ with $\psi^\beta_{\rm e}(\sigma_1,\sigma_2) = \psi^0_{\rm e}(\sigma_1,\sigma_2)\, e^{\beta\xi_{\rm e}(\sigma_1,\sigma_2)}$, $\psi^\beta_{\rm v}(\sigma) = \psi^0_{\rm v}(\sigma)$, and assume $\min_{\sigma_1,\sigma_2\in\mathcal{X}} \psi^0_{\rm e}(\sigma_1,\sigma_2) > 0$, $\min_{\sigma\in\mathcal{X}} \psi^0_{\rm v}(\sigma) > 0$.*

We call $\beta\in\mathbb{R}$ 'exceptional' if the following happens. There is two global maximizers b, $b' \in \mathcal{H}^{\rm loc}_{k,{\rm reg}}(\mathcal{X})$ of $\Phi^*_{k,\psi}$ such that

$$\mathbb{E}_b \xi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) \neq \mathbb{E}_{b'} \xi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2)\,. \tag{4.2.42}$$

Then the set of exceptional points is at most countable.

Proof. Write, with a slight abuse of notation, $\Phi^*_{k,\beta}(b) = \Phi^*_{k,\psi^\beta}(b)$ for the Bethe free energy with specification ψ^β. This takes the form

$$\Phi^*_{k,\beta}(b) = \Phi^*_{k,0}(b) + \beta \mathbb{E}_b \xi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_1)\,. \tag{4.2.43}$$

Let $v_1 \equiv \min_{b\in Hloc_{{\rm reg},k}} \mathbb{E}_b \xi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_1)$, $v_2 \equiv \max_{b\in Hloc_{{\rm reg},k}} \mathbb{E}_b \xi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_1)$, and, for any $\beta\in\mathbb{R}$, define

$$v_+(\beta) \equiv \sup\left\{ \mathbb{E}_b \xi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_1) :\ \ b \in \arg\max_{b\in\mathcal{H}^{\rm loc}_{{\rm reg},k}} \Phi^*_{k,\beta}(b) \right\}, \tag{4.2.44}$$

$$v_-(\beta) \equiv \inf\left\{ \mathbb{E}_b \xi_{\rm e}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_1) :\ \ b \in \arg\max_{b\in\mathcal{H}^{\rm loc}_{{\rm reg},k}} \Phi^*_{k,\beta}(b) \right\}. \tag{4.2.45}$$

Our claim is then equivalent to the claim that $v_+(\beta) = v_-(\beta)$ for all β apart from a countable set. The claim follows from checking the following properties (that are straightforward):

- For any $\beta\in\mathbb{R}$, $v_1 \le v_+(\beta) \le v_-(\beta) \le v_2$.
- $v_+(\beta)$, $v_-(\beta)$ are non-decreasing. Further $v_+(\beta)$ is right-continuous, and $v_-(\beta)$ is left-continuous.
- For any $\beta_0\in\mathbb{R}$, we have $v_-(\beta)\downarrow v_+(\beta_0)$ as $\beta\downarrow\beta_0$, and $v_+(\beta)\uparrow v_-(\beta_0)$ as $\beta\uparrow\beta_0$.

Hence the set of exceptional points coincides with the set of discontinuity points of the non-decreasing bounded function v_+, and this is necessarily countable. □

4.3 Applications

4.3.1 The ferromagnetic Ising model

As a simple successful applications of the method developed in the previous section, we can consider the ferromagnetic Ising model in zero magnetic field. In this case $\mathcal{X} = \{+1,-1\}$, $\psi_{\rm v}(\sigma) = 1$ and $\psi_{\rm e}(\sigma,\sigma') = \exp(\beta\sigma\sigma')$.

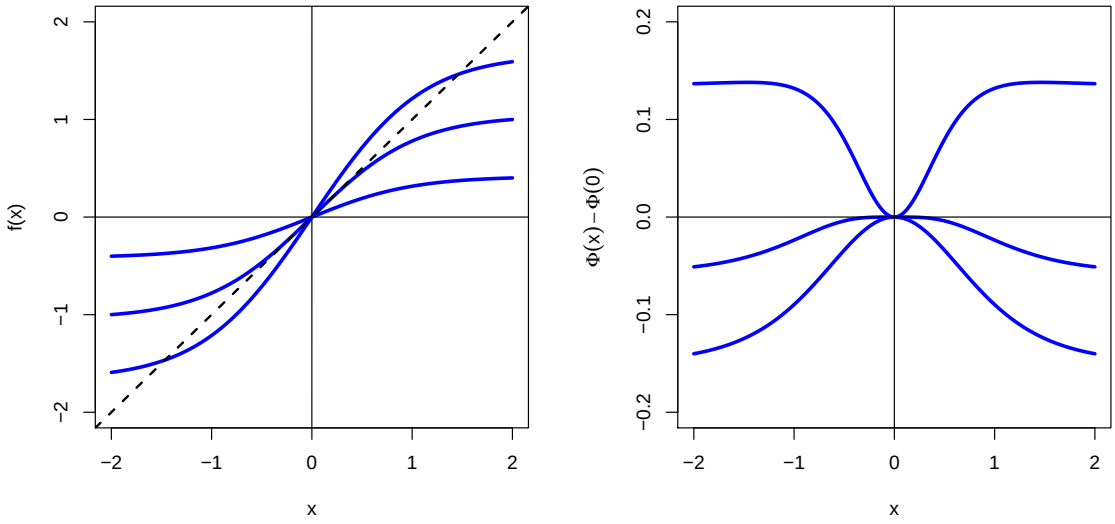

Figure 4.1: Left frame: graphical solution of the stationarity condition for the free energy of ferromagnetic Ising model, Eq. (4.3.2). Here $k = 4$ and the three curves refer to different values of β. From bottom to top in the right part of the plot: $(\beta/\beta_c) = 0.4$, 1, 1.6. Here $\beta_c = \mathrm{atanh}(1/(k-1)) \approx 0.34657$. Right frame: free energy functional $\Phi_{k,\beta}(x) - \Phi_{k,\beta}(0)$, for the same values of β. From bottom to top $(\beta/\beta_c) = 0.4$, 1, 1.6. We subtract the value $\Phi_{k,\beta}(0)$ for clarity.

It is convenient and customary to parameterize $h \in \Delta_{\mathcal{X}}$ as

$$h(\sigma) = \frac{e^{x\sigma}}{2\cosh x} = \frac{1}{2}\big(1 + \sigma \tanh x\big)\,, \tag{4.3.1}$$

with $x \in \overline{\mathbb{R}}$ (recall that $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ denotes the completed real line). In terms of this variable, and letting $\theta = \tanh\beta$, Eq. (4.2.23) reads

$$x = (k-1)\mathrm{atanh}(\theta \tanh x)\,. \tag{4.3.2}$$

With a slight abuse of notation, we will write $\mathcal{H}^*_{\beta,k}$ for the set of solutions of this equation, see Fig. ???.

The function $x \mapsto f(x) \equiv (k-1)\mathrm{atanh}(\theta \tanh x)$ is antisymmetric ($f(-x) = -f(x)$) and concave on $\mathbb{R}_{\geq 0}$. Further $f'(0) = (k-1)\theta$ and $\lim_{x\to+\infty} f(x) = (k-1)\beta < \infty$. Ths implies immediately the following picture:

1. For $(k-1)\tanh\beta \leq 1$, Eq. (4.3.2) admits a unique solution $b = 0$, and hence $\mathcal{H}^*_{\beta,k} = \{0\}$.

2. For $(k-1)\tanh\beta > 1$, $b = 0$ still solves Eq. (4.3.2) but in addition we have two symmetric solutions, i.e. $\mathcal{H}^*_{\beta,k} = \{-b_*, 0, b_*\}$ for some $b_* = b_*(\beta) > 0$.

The free-energy (4.2.21) can be written explicitely in terms of b as

$$\Phi_{k,\beta}(x) \equiv \frac{k}{2}\log\cosh\beta + \log\Big\{(1+\theta\tanh x)^k + (1-\theta\tanh x)^k\Big\} - \frac{k}{2}\log\Big\{1+\theta(\tanh x)^2\Big\}\,. \tag{4.3.3}$$

A simple study of this function shows that:

1. For $(k-1)\tanh\beta \le 1$, the function $b \mapsto \Phi_{k,\beta}(b)$ has a unique global maximum at $b = 0$, with value $\Phi_{k,\beta}(0) = \log 2 + (k/2)\log\cosh\beta$.

2. For $(k-1)\tanh\beta > 1$, $b = 0$ is a local minimum of $b \mapsto \Phi_{k,\beta}(b)$, while $\{+b_*, -b_*\}$ are degenerate global maxima.

It turns out that in this case thesecond moment method succeeds at all temperatures, and allows to prove the following result.

Theorem 9. *Let $\{\boldsymbol{G}_n\}_{n\ge 1}$ be a sequence of random k-regular graphs and ψ be the specification of a ferromagnetic Ising model at inverse temperature β (i.e. $\psi_{\rm e}(\sigma,\sigma') = \exp(\beta\sigma\sigma')$, $\psi_{\rm v}(\sigma) = 1$).*

If $(k-1)\tanh\beta \le 1$, then almost surely

$$\lim_{n\to\infty} \phi(\boldsymbol{G}_n, \psi) = \log 2 + \frac{k}{2}\log\cosh\beta\,. \tag{4.3.4}$$

If $(k-1)\tanh\beta > 1$, then almost surely

$$\lim_{n\to\infty} \phi(\boldsymbol{G}_n, \psi) = \Phi_{k,\beta}(b_*)\,, \tag{4.3.5}$$

with $\Phi_{k,\beta}$ given by Eq. (4.3.3) and b_* the only positive solution of Eq. (4.3.2).

Proof. The proof cosists in showing that the maximum of $\Phi_{\psi^{\otimes 2},k}$ over $\mathcal{H}^*_{\psi^{\otimes 2},k}$ is indeed achieved on $(\mathcal{H}^*_{\psi,k})^{\otimes 2}$, and hence using Theorem 8.

By Lemma **??**, we can write any element of $\mathcal{H}^*_{\psi^{\otimes 2},k}$ as

$$h(\sigma,\sigma') = \frac{1}{z}\exp\left(b\sigma + b\sigma' + Q\sigma\sigma'\right), \tag{4.3.6}$$

with $b \in \mathcal{H}^*_{\beta,k}$. Correspondingly, we have $\mathcal{H}^*_{\psi^{\otimes 2},k} = \mathcal{H}^{0,*}_{\psi^{\otimes 2},k} \cup \mathcal{H}^{+,*}_{\psi^{\otimes 2},k}$ where the partition corresponds to $b = 0$ and $b = +b_*$ (the case $-b_*$ can be eliminated by symmetry).

We first consider the solutions in $\mathcal{H}^{+,*}_{\psi^{\otimes 2},k}$. We claim that, for $b = b_*$, the only solution of Eq,. (4.2.30) of the form Eq. (4.3.6) has $Q = 0$. This can be proved by Griffiths inequality (to be written).

Next consider the solutions in $\mathcal{H}^{0,*}_{\psi^{\otimes 2},k}$, i.e. the case $b = 0$. Equation (4.2.30) yields in this case

$$Q = (k-1)\,\mathrm{atanh}(\theta^2\tanh(Q))\,, \tag{4.3.7}$$

i.e. the same equation as for $b_*(\beta)$, excepth that θ is replaced by θ^2. Hence for $(k-1)(\tanh\beta)^2 \le 1$,$Q = 0$ is the only soution, i.e. $\mathcal{H}^{0,*}_{\psi^{\otimes 2},k} = (\mathcal{H}^*_{\psi,k})^{\otimes 2}$. For $(k-1)(\tanh\beta)^2 \le 1$, two more solutions appear $\{+Q_*(\beta), -Q_*(\beta)\}$ where, letting $\beta_2(\beta) \equiv \mathrm{atanh}(\tanh(\beta)^2)$, we have $Q_*(\beta) = b_*(\beta_2(\beta)) > 0$. By symmetry we can focus on the positive solution $+Q_*(\beta)$.

Let $h_+ \in \mathcal{H}^*_{\psi^{\otimes 2},k}$ be given by

$$h_{\beta,+}(\sigma,\sigma') \equiv \frac{e^{Q_*(\beta)\sigma\sigma'}}{4\cosh Q_*(\beta)}\,. \tag{4.3.8}$$

This is the only solution that is not in $(\mathcal{H}^*_{\psi,k})^{\otimes 2}$. We are left with the task of proving that $\Phi_{\psi^{\otimes 2},k}(h_{\beta,+}) \le 2\Phi_{\beta,k}(b_*(\beta))$ for all $(k-1)\tanh(\beta)^2 > 1$, with $\Phi_{\beta,k}(b)$ given by Eq. (4.3.3). A straightforward calculation yields

$$\Phi_{\psi^{\otimes 2},k}(h_{\beta,+}) = \Phi_{\beta,k}(b=0) + \Phi_{\beta_2,k}(b = b_*(\beta_2))\,, \tag{4.3.9}$$

with $\beta_2 = \beta_2(\beta) < \beta$. Note that $\Phi_{\beta,k}(0) \le \Phi_{\beta,k}(b_*(\beta))$. Further we claim that $\beta \mapsto \Phi_{\beta,k}(b_*(\beta))$ is monotone increasing and therefore $\Phi_{\beta_2,k}(b = b_*(\beta_2)) \le \Phi_{\beta,k}(b_*(\beta))$ because $\beta_2(\beta) < \beta$.

To be written: Prove of last claim. □

Another example/exercise: Independent sets on pipartite regular graphs.

4.3.2 The anti-ferromagnetic Ising model

Consider now the antiferromagnetic Ising case in zero field, namely $\psi_{\rm v}(\sigma) = 1$, $\psi_{\rm e}(\sigma, \sigma') = \exp(-\beta\sigma\sigma')$, with $\beta > 0$. (Note that we changed parameterization so that β remains positive.)

We will use the same parameterization introduced in Eq. (4.3.1) for $h \in \Delta_{\mathcal{X}}$ in terms of $b \in \mathbb{R}$, and let $\theta = \tanh(\beta) > 0$. We thus obtain the following equation that replaces Eq. (4.3.2):

$$b = -(k-1)\text{atanh}(\theta \tanh b)\,. \tag{4.3.10}$$

The function $f(x) = -(k-1)\text{atanh}(\theta \tanh x)$ is antysimmetric ($f(-x) = -f(x)$) and decreasing. Hence the only solution is $b = 0$, i.e. $\mathcal{H}^*_{\beta,k} = \{0\}$. The Bethe free enrgy functional has the same form (4.3.3). Evaluating it at $b = 0$, we obtain the following almost sure bound.

$$\limsup_{n\to\infty} \frac{1}{n} \log Z(\boldsymbol{G}_n, \psi) \le \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} Z(\boldsymbol{G}_n, \psi) = \log 2 + \frac{k}{2} \log\cosh\beta\,. \tag{4.3.11}$$

The second moment method does not allow to prove a matching lower bound in this case.

Theorem 10. *Let $\{\boldsymbol{G}_n\}_{n\ge 1}$ be a sequence of random k-regular graphs and ψ be the specification of a ferromagnetic Ising model at inverse temperature β (i.e. $\psi_{\rm e}(\sigma,\sigma') = \exp(\beta\sigma\sigma')$, $\psi_{\rm v}(\sigma) = 1$).*

If $(k-1)\tanh(\beta)^2 \le 1$, then almost surely

$$\lim_{n\to\infty} \phi(\boldsymbol{G}_n, \psi) = \log 2 + \frac{k}{2} \log\cosh\beta\,. \tag{4.3.12}$$

Proof. To be written □

Another example: Independent sets on random regular graphs.

4.4 Refining the second moment method

4.4.1 Truncation

While the second moment method is elementary, its implementation can become fairly sophisticated using the trick in the next exercise.

Exercise 4.4.1. Prove that the conclusions of Theorem 5 and 6 continue to hold if Eq. (4.1.12) is replaced by

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{Z(\boldsymbol{G}_n, \psi)\, \mathbb{I}(\boldsymbol{G}_n \in \mathcal{G}_n)\} = \lim_{n\to\infty} \frac{1}{2n} \log \mathbb{E}\{Z(\boldsymbol{G}_n, \psi)^2\, \mathbb{I}(\boldsymbol{G}_n \in \mathcal{G}_n)\} = \phi\,, \tag{4.4.1}$$

where $\boldsymbol{G}_n$ is a set of graphs such that $\mathbb{P}(\boldsymbol{G}_n \in \mathcal{G}_n) \geq e^{-o(n)}$.

Further, prove that, if $(1/n)\mathbb{E}\log Z(\boldsymbol{G}_n, \psi)$ has a limit, then there always exists sets $\mathcal{G}_n$ such Eq. (4.4.1) holds, with $\mathbb{P}(\boldsymbol{G}_n \in \mathcal{G}_n) \geq e^{-o(n)}$. Unfortunately, te sets $\mathcal{G}_n$ are not explicit.

4.4.2 Clustering

Chapter 5

The weak convergence method

The second moment method has a few limitations:

1. It only applies to sequences of random graphs $\{\boldsymbol{G}_n\}_{n\geq 1}$.

2. It aims at computing the asymptotic free energy density and does not estabilish a local limit for the measures $\mu_{\boldsymbol{G}_n,\psi}$. While it is sometimes possible to use the limit free energy to identify a specific local limit, there is no general technique to achieve this goal.

3. The key condition under which the second moment method succeeds is $\mathbb{E}\{Z(\boldsymbol{G}_n,\psi)^2\} \doteq \mathbb{E}\{Z(\boldsymbol{G}_n,\psi)\}^2$. This condition often often holds for random regular graphs in a weak-dependencies regime and sometimes beyond that regime. However, it generically does not holds for other graph models, such as Erdös-Renyi random graphs. The intuition is that other random graph models have more 'variability.' For instance, $B_t(i;\boldsymbol{G}_n)$ is asymptotically deterministic (as $n\to\infty$) for random regular graphs. It is instead random, and asymptotically distributed as a Galton-Watson tree (see Proposition **????**) for Erdös-Renyi random graph. This results in a much larger variance for $Z(\boldsymbol{G}_n,\psi)$.

   The domain of applicability of the second moment method can be expanded by using truncation. However, the resulting calculations are significantly more complicated, and there is no general guarantee that they yield matching first and second moment.

This chapter discusses a very different approach that is instead well-suited to deterministic graph sequences $\{G_n\}$ and directly addresses the problem of determining the local limit of $(G_n,\mu_{G_n,\psi})$. As is often the case with weak convergence, one proceeds in two steps. The first step estabilish, by a general soft argument, that the local limit of $(G_n,\mu_{G_n,\psi})$ exists along subsequences. The second step is the most challenging and is model-specific. One has to show that, among all the possible limt distributions, only one can be realized. Hence all the subsequential limits coincide with that unique distribution and $(G_n,\mu_{G_n,\psi})$ converges to it.

The first step of this procedure is explained, in some generality, in Section 5.1. The second step is illustrated on two cases. The independent sets model is discussed in Section **??**: in this case the approach allows to estabilish the local weak limit only in a 'weakly interacting' regime. This is to be expected because in the 'strongly interacting' regime, the behavior of this model depends on the global structure of the graph G_n, e.g. on whether or not G_n is bipartite, cf. Sections **??** and 5.3 for further discussion. The ferromagnetic

Ising model is considered in Section 5.3. In this case a nearly complete characterization has been developed, although some interesting questions remain open.

5.1 General approach

The weak convergence method is based on the following general result. Note that this statement holds for non-positive, non-permissive specifications, under the condition that $\mu_{G_n,\psi}$ is well-defined for all n, i.e. there is at least one configuration $\sigma \in \mathcal{X}^{V_n}$ with non-zero weight.

Theorem 11. *Let $\{G_n\}_{n\geq 1}$ be a graph sequence (not necessarily indexed by the number of vertices) that converges locally to a random rooted tree $(\boldsymbol{T}, \boldsymbol{o}) \sim \nu$. Let ψ a specification such that $Z(G_n, \psi) > 0$ for all n. Then there exist a subsequence $\{n(k)\}_{k\geq 1}$, $\lim_{k\to\infty} n(k) = \infty$, and a random marked rooted tree $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ such that $(\boldsymbol{T}, \boldsymbol{o}) \sim \nu$ and $(G_{n(k)}, \mu_{G_{n(k)},\psi})$ converges locally to $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$.*

Further, the limit $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$ is unimodular and Gibbs with respect to ψ.

Proof. Let $\boldsymbol{I}_n$ a uniformly random vertex in V_n, and fix $t \in \mathbb{N}$. Further, for each n, let $\boldsymbol{\sigma}^n \sim \mu_{G_n,\psi}$. We claim that, for each sequence $\{m(k)\}_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} m(k) = \infty$, there exists a subsequence $\{n(k)\}_{k\in\mathbb{N}} \subseteq \{m(k)\}_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} n(k) = \infty$, and a t-generations marked tree $(\boldsymbol{T}_t, \boldsymbol{o}_t, \boldsymbol{\sigma}_t)$, such that

$$(B_t(\boldsymbol{I}_{n(k)}; G_{n(k)}), \boldsymbol{I}_{n(k)}, \boldsymbol{\sigma}^{n(k)}_{B_t(\boldsymbol{I}_{n(k)};G_{n(k)})}) \stackrel{\mathrm{d}}{\Longrightarrow} (\boldsymbol{T}_t, \boldsymbol{o}_t, \boldsymbol{\sigma}_t)\,, \tag{5.1.1}$$

along the subsequence $\{n(k)\}_{k\in\mathbb{N}}$.

In order to prove this claim, note the following. Let $\mathcal{K}_*(t, \Delta) \subseteq \mathcal{G}_*$ be the set of of unlabeled rooted graph with radius (maximum distance of any vertex from the root) at most t and maximum degree at most Δ. Clearly $\mathcal{K}_*(t, \Delta)$ is finite. Also, let $\mathrm{maxdeg}_t : \mathcal{G}_* \to \mathbb{N}$ be the function such that $\mathrm{maxdeg}_t(G, o)$ is the maximum degree of the vertices of (G, o) whose distance from the root is at most t. Clearly, maxdeg_t is continuous in the topology of $\mathcal{G}_*$.

Since $B_t(\boldsymbol{I}_n; G_n)$ converges in distribution, $\mathrm{maxdeg}_t(B_t(\boldsymbol{I}_n; G_n))$ also converges in distribution. Hence, for each $\varepsilon > 0$, there exists $\Delta(\varepsilon) < \infty$ such that $\mathrm{maxdeg}_t(B_t(\boldsymbol{I}_n; G_n)) \leq \Delta(\varepsilon)$ with probability at least $1-\varepsilon$ for all t. In other words $B_t(\boldsymbol{I}_n; G_n) \in \mathcal{K}_*(t, \Delta(\varepsilon))$ with probability at least $1-\varepsilon$. Define the shorthand $\mathcal{K}(\varepsilon) = \mathcal{K}_*(t, \Delta(\varepsilon))$ and let $\mathcal{K}^{\mathcal{X}}(\varepsilon)$ denote the set of marked rooted graphs whose underlying unmarked graph is in $\mathcal{K}(\varepsilon)$. Note that for each $(G, o) \in \mathcal{K}(\varepsilon)$ marks can be assigned in at most $|\mathcal{X}|^{|V(G)|}$ ways. Since $\mathcal{K}(\varepsilon)$ is finite, it follows that $\mathcal{K}^{\mathcal{X}}(\varepsilon)$ is finite as well and therefore compact in $\mathcal{G}_*^{\mathcal{X}}$. Therefore

$$\{(B_t(\boldsymbol{I}_{m(k)}; G_{m(k)}), \boldsymbol{\sigma}^{m(k)}_{B_t(\boldsymbol{I}_{m(k)};G_{m(k)})})\}_{k\in\mathbb{N}} \tag{5.1.2}$$

is a tight family, whence the claim follows by Prohorov theorem.

The convergence of $(G_{n(k)}, \mu_{G_{n(k)},\psi})$ is then obtained by a diagonal argument. Namely, by the above, we can construct subsequences $\{n_1(k)\} \supseteq \{n_2(k)\} \supseteq \{n_3(k)\} \supseteq \dots$ such that, along $n_t(\,\cdot\,)$,

$$(B_t(\boldsymbol{I}_{n_t(k)}; G_{n_t(k)}), \boldsymbol{\sigma}^{n_t(k)}_{B_t(\boldsymbol{I}_{n_t(k)};G_{n_t(k)})}) \stackrel{\mathrm{d}}{\Longrightarrow} (\boldsymbol{T}_t, \boldsymbol{\sigma}_{\boldsymbol{T}_t}). \tag{5.1.3}$$

By chosing $n(k) = n_k(k)$, we have, for any $t < \infty$

$$(B_t(\boldsymbol{I}_{n(k)}; G_{n(k)}), \boldsymbol{\sigma}^{n(k)}_{B_t(\boldsymbol{I}_{n(k)}; G_{n(k)})}) \stackrel{\mathrm{d}}{\Longrightarrow} (\boldsymbol{T}_t, \boldsymbol{\sigma}_{\boldsymbol{T}_t}), \tag{5.1.4}$$

whence the desired convergence follows by Lemma [**Write a Lemma to clarify that it is sufficient to check convergence over balls.**].

We already proved unimodularity in **???** and the Gibbs property in Proposition 3.2.5. □

The weak convergence method amounts to identify, among all possible limit probability measures on $\mathcal{G}_*^{\mathcal{X}}$ (i.e. all possible laws of $(\boldsymbol{T}, \boldsymbol{\sigma}, \boldsymbol{o})$) a unique one, call it ν that can be obtained as subsequential local limit of (G_n, μ_{ψ, G_n}). If this happen, then (G_n, μ_{ψ, G_n}) converges locally to ν. Viceversa, Theorem 11 can be used to prove the existence of unimodular Gibbs measures on $(\boldsymbol{T}, \boldsymbol{o})$. We have the following immediate corollary.

Corollary 5.1.1. *Let $(\boldsymbol{T}, \boldsymbol{o}) \sim \nu_G$ be a unimodular random rooted tree and ψ a specification. Assume that there exists a sequence of finite graphs $\{G_n\}_{n \geq 1}$ that converges locally to $(\boldsymbol{T}, \boldsymbol{o})$, and that, for all n, $Z(G_n, \psi) > 0$. Then there exists at least one unimodular probability measure on $\mathcal{G}_*^{\mathcal{X}}$ that is Gibbs with respect to psi and such that the induced measure on $\mathcal{G}_*$ coincide with ν_G.*

A particularly simple case is the one in which there exists a unique probability measure on $\mathcal{G}_*^{\mathcal{X}}$ that is unimodular and Gibbs with respect to ψ. In that case (G_n, μ_{ψ, G_n}) converges locally to this unique measure. This is stated formally below.

Corollary 5.1.2. *Let $\{G_n\}_{n \geq 1}$ be a graph sequence that converges to a random rooted tree $(\boldsymbol{T}, \boldsymbol{o})$, and ψ a specification. Assume that there exists a unique unimodular probability measure ν_G on $\mathcal{G}_*^{\mathcal{X}}$ that is Gibbs with respect to ψ. Then $(G_n, \mu_{G_n, \psi})$ converges locally to ν.*

In particular, this is the case if, $(\boldsymbol{T}, \boldsymbol{o})$-almost surely, there exists a unique Gibbs measure with specification ψ on $(\boldsymbol{T}, \boldsymbol{o})$.

Proof. The first part of the statement (convergence if there exists a unique probability measure on $\mathcal{G}_*^{\mathcal{X}}$ that is unimodular and Gibbs) follows from Theorem 11 as per the discussion above. The second part follows from the first part after proving the following claim: if, $(\boldsymbol{T}, \boldsymbol{o})$-almost surely, there exists a unique Gibbs measure with specification ψ on $(\boldsymbol{T}, \boldsymbol{o})$, then there exists unique probability measure ν on $\mathcal{G}_*^{\mathcal{X}}$ that is Gibbs with respect to ψ, and such that the induced distribution on $\mathcal{G}_*$ coincides with ν_G.

Indeed by Corollary 5.1.1, there exists at least one probability measure satisfying these conditions. Assume by contradiction that there exists at least two distinct such probability measures ν_1 and ν_2. By Lemma [**Lemma on canonical representation of $\mathcal{G}_*$**] these can be viewed as probability measures on $\mathcal{C}_* \times \mathcal{X}^{\mathbb{N}}$ and hence admit the representation $\nu_1 = \nu_{\boldsymbol{T}} \otimes \nu^{(1)}_{\boldsymbol{\sigma}|\boldsymbol{T}}$, $\nu_2 = \nu_{\boldsymbol{T}} \otimes \nu^{(2)}_{\boldsymbol{\sigma}|\boldsymbol{T}}$. Since th marginal $\nu_{\boldsymbol{T}}$ coincides, we must have $\nu^{(1)}_{\boldsymbol{\sigma}|\boldsymbol{T}}(\,\cdot\,|(T, o)) \neq \nu^{(2)}_{\boldsymbol{\sigma}|\boldsymbol{T}}(\,\cdot\,|(T, o))$ for all $(T, o) \in \mathcal{S}$, where $\mathcal{S} \subseteq \mathcal{C}_*$ is such that $\nu_{\boldsymbol{T}}(\mathcal{S}) > 0$. Since by Theorem 11 both $\nu^{(1)}_{\boldsymbol{\sigma}|\boldsymbol{T}}(\,\cdot\,|(T, o))$ and $\nu^{(2)}_{\boldsymbol{\sigma}|\boldsymbol{T}}(\,\cdot\,|(T, o))$ are Gibbs, this contradicts the hypotheses. □

The *Dobrushin criterion* provides a simple sufficient condition for uniqueness of Gibbs measure, and hence allows to check the conditions of Corollary 5.1.2. We specialize it here

to the present setting. For each ℓ, define the conditional distribution of a spin given its ℓ neighbors:

$$\mu_\ell(\sigma_0|\sigma_1,\dots,\sigma_\ell) = \frac{\psi_{\rm v}(\sigma_0)\prod_{i=1}^{\ell}\psi_{\rm e}(\sigma_0,\sigma_\ell)}{\sum_{\tau\in\mathcal{X}}\psi_{\rm v}(\tau)\prod_{i=1}^{\ell}\psi_{\rm e}(\tau,\sigma_\ell)}\,. \tag{5.1.5}$$

In words, this is the conditional distribution of the spin $\boldsymbol{\sigma}_0$ given $\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_\ell$, where the joint distribution of $\boldsymbol{\sigma}_0,\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_\ell$, is given by the factor model $\mu_\ell = \mu_{S_\ell,\psi}$ on the star graph S_ℓ shown in Fig. **????**. Explicitly $S_\ell = (V(S_\ell), E(S_\ell))$ where $V(S_\ell) = (0,1,2,\dots,\ell)$ and $E(S_\ell) = \{(0,1),(0,2),\dots,(0,\ell)\}$. The *influence coefficients* quantify the effect on a spin distribution, produced by conditioning on the neighbors of that spin. They are defined defined, for each $\ell \ge 1$, by

$$c_\ell(\psi) = \max_{\sigma_1,\sigma_2,\dots,\sigma_\ell,\sigma'_\ell} \big\|\mu_\ell(\,\cdot\,|\sigma_1,\dots,\sigma_\ell) - \mu_\ell(\,\cdot\,|\sigma_1,\dots,\sigma'_\ell)\big\|_{\rm TV}\,. \tag{5.1.6}$$

Corollary 5.1.3. *Let $\{G_n\}_{n\ge1}$ be a graph sequence that converges to a random rooted tree $(\boldsymbol{T},\boldsymbol{o})$, such that $\deg(o)\le\Delta$ almost surely. Let ψ be a specification such that*

$$\max_{1\le\ell\le\Delta}\ \ell\, c_\ell(\psi) < 1\,. \tag{5.1.7}$$

Then there exists almost surely a unique Gibbs measure ν on $\mathcal{G}_*^{\mathcal{X}}$ with the prescribed marginal of $(\boldsymbol{T},\boldsymbol{o})$, and $(G_n,\mu_{G_n,\psi})$ converges locally to ν.

Proof. It is a standard result in the theory of Gibbs measures [**Add citation**] that, for any graph $G=(V,E)$ with degree bounded by Δ, if ψ satisfies condition (5.1.7), then there is a unique probability measure on $\mathcal{X}^V$ that is Gibbs with respect to ψ. The claim therefore follows from Corollary 5.1.2. □

Example 5.1.1 (Independent set)**.** Recall that the independent sets model is defined by$\mathcal{X}=\{0,1\}$, $\psi_{\rm v}(\sigma)=\lambda^\sigma$, $\psi_{\rm e}(\sigma,\sigma') = {\rm I}((\sigma,\sigma')\neq(1,1))$.

In order to evaluate the influence coefficient (5.1.6), note that

$$\big\|\mu_\ell(\,\cdot\,|\sigma_1,\dots,\sigma_\ell) - \mu_\ell(\,\cdot\,|\sigma_1,\dots,\sigma'_\ell)\big\|_{\rm TV} = \big|\mu_\ell(1|\sigma_1,\dots,\sigma_\ell) - \mu_\ell(1|\sigma_1,\dots,\sigma'_\ell)\big| \tag{5.1.8}$$

and that both terms on the right hand side vanish if at least one out od $\sigma_2,\dots\sigma_\ell$ is equal to 1. We therefore have

$$\begin{aligned} c_\ell(\psi) &= \big|\,\mu_\ell(1|0,\dots,0,0) - \mu_\ell(1|0,\dots,0,1)\,\big| && (5.1.9)\\ &= \mu_\ell(1|0,\dots,0,0) && (5.1.10)\\ &= \frac{\lambda}{1+\lambda}\,. && (5.1.11)\end{aligned}$$

Applying Corollary 5.1.3, we immediately have the following.

Corollary 5.1.4. *Let $\{G_n\}_{n\ge1}$ be a graph sequence that converges to a random rooted tree $(\boldsymbol{T},\boldsymbol{o})\sim\nu_{\boldsymbol{T}}$, such that $\deg(o)\le\Delta$, $\nu_{\boldsymbol{T}}$-almost surely. For each n, let $\mu_{G_n,\psi}$ be the indepedent sets measure for the fraph G_n with fugacity λ*

Let $\lambda\in[0,(\Delta-1)^{-1})$. Then there exists a unique probability ν on $\mathcal{G}_*^{\mathcal{X}}$ that is Gibbs for the independent sets model, and whose marginal on $\mathcal{G}_*$ coincides with $\nu_{\boldsymbol{T}}$. Further, $(G_n,\mu_{G_n,\psi})$ converges locally to ν.

In Section 5.2 we discuss how this analysis can be sharpened to obtain the optimal range of λ for which the same limit result applies.

Example 5.1.2 (Ising model)**.** In this case $\mathcal{X} = \{+1, -1\}$, $\psi_{\rm v}(\sigma) = e^{B\sigma}$, $\psi_{\rm e}(\sigma\sigma') = e^{\beta\sigma\sigma'}$. For the sake of simplicity, we stick to the zero magnetic field case, $B = 0$. A simple calculation yields

$$\big\|\mu_\ell(\,\cdot\,|\sigma_1,\dots,\sigma_\ell) - \mu_\ell(\,\cdot\,|\sigma_1,\dots,\sigma'_\ell)\big\|_{\rm TV} = \big|\mu_\ell(1|\sigma_1,\dots,\sigma_\ell) - \mu_\ell(1|\sigma_1,\dots,\sigma'_\ell)\big| \quad (5.1.12)$$

$$= \frac{1}{2}\big|\tanh\big(y+\beta\sigma_\ell\big) - \tanh\big(y+\beta\sigma'_\ell\big)\big|\,, \quad (5.1.13)$$

where $x = \beta(\sigma_1 + \cdots + \sigma_{\ell-1})$. Note that the function $x \mapsto \tanh(x)$ is monotone increasing, its derivative is maximized at $x = 0$, it is symmetric in x, is monotone increasing for $x < 0$ and decreasing for $x > 0$. It follows that the above difference is mazimized for $y = 0$ if ℓ is odd, and $y = \beta$ if ℓ is even. We get

$$c_\ell(\psi) = \begin{cases} \tanh\beta & \text{if } \ell \text{ is odd,} \\ (1/2)\tanh(2\beta) & \text{if } \ell \text{ is even.} \end{cases} \quad (5.1.14)$$

In particular $c_\ell(\psi) \le \tanh\beta$ for all ℓ and we have therefore the following. (Note that the argument above applies to $\beta < 0$ as well and hence we state the next result generally.)

Corollary 5.1.5. *Let $\{G_n\}_{n\ge1}$ be a graph sequence that converges to a random rooted tree $(\boldsymbol{T}, \boldsymbol{o}) \sim \nu_{\boldsymbol{T}}$, such that $\deg(o) \le \Delta$, $\nu_{\boldsymbol{T}}$-almost surely. For each n, let $\mu_{G_n,\psi}$ be the Ising model on the graph G_n with inverse temperature β and magnetic field $B = 0$.*

Assume $\tanh\beta \in (-\Delta^{-1}, \Delta^{-1})$. Then there exists a unique probability ν on $\mathcal{G}_*^{\mathcal{X}}$ that is Gibbs for the Ising model, and whose marginal on $\mathcal{G}_*$ coincides with $\nu_{\boldsymbol{T}}$. Further, $(G_n, \mu_{G_n,\psi})$ converges locally to ν.

The argument based on Dobrushin condition only admits a limited improvement in this case, as demonstrated by the following remark.

Proposition 5.1.6. *Let $T = T^{{\rm reg},\Delta} = (V, E)$ be the infinite regular tree with degree $\Delta \ge 3$, and ψ be the specification for the Ising model with zero magnetic field $B = 0$ and inverse temperature β. For $\tanh\beta \in [-(\Delta-1)^{-1}, (\Delta-1)^{-1}]$ there exists a unique probability measure on $\{+1,-1\}^V$ thatis Gibbs with respect to ψ.*

Viceversa, it $(\Delta-1)\tanh|\beta| > 1$, then there exists at least two distinct Gibbs measures for ψ.

Proof. The proof of this statement is standard and hence we omit it referring, for instance, to **???**. In a nutshell, it amounts to considering the Gibbs measures constructed by taking 'plus' and 'minus' boundary conditions and proving that they are ditinct. The latter is done by computing the probabilities that $\boldsymbol{\sigma}_o = +1$ under the two measures (which can be done by recursion) and proving that they are distinct. □

Example 5.1.3 (Proper colorings)**.** In this case $\mathcal{X} = \{1, 2, \dots, q\}$, $\psi_{\rm v}(\sigma) = 1$, and $\psi_{\rm e}(\sigma, \sigma') = \mathbb{I}(\sigma \ne \sigma')$. Therefore $\mu_\ell(\,\cdot\,|\sigma_1,\dots,\sigma_\ell)$ is the uniform probability distribution on $\mathcal{X} \setminus \{\sigma_1,\dots,\sigma_\ell\}$. (In order for Dobrushin criterion to apply, it is necessary to take $q > \ell$.)

5.2 The case of independent sets

In this section we outline a sharper analysis of the independent sets example in the previous section. Let $\{G_n\}_{n\ge 1}$ be a graph sequence that converges locally to a random rooted tree $(\boldsymbol{T},\boldsymbol{o})$, with $\deg(\boldsymbol{o})\le k$ almost surely (and –as a consequence– $(\boldsymbol{T},\boldsymbol{o})$ has maximum degree k amost surely).

Define

$$\lambda_{\mathrm{u}}(k)\equiv\frac{(k-1)^{k-1}}{(k-2)^k}\,. \tag{5.2.1}$$

It was proved in [Wei06, Theorem 2.3] that, for $\lambda<\lambda_{\mathrm{u}}(k)$ there exists a unique Gibbs measure on any tree with maximum degree k. Applying the last corollary, it follows that, for $\lambda<\lambda_{\mathrm{u}}(k)$, $(G_n,\mu_{G_n,\psi})$ converges to that unique Gibbs measure.

While proving the result of [Wei06] goes beyond the scope of these lectures, it is easy (and instructive) to prove this result for the k-regular tree $T=T^{\mathrm{reg},k}$, $T=(V,E)$. In the rest of this section, we omit the arguments T and ψ since they are fixed throughout.

The key observation is that –on bipartite graphs– the independent sets model enjoys a special monotonicity property.

Given $\sigma^{(1)},\sigma^{(2)}\in\mathcal{X}^V=\{0,1\}^V$, we write $\sigma^{(1)}\preceq\sigma^{(2)}$ if $\sigma^{(1)}_i\le\sigma^{(2)}_i$ when $d(o,i)$ is even *and* $\sigma^{(1)}_i\ge\sigma^{(2)}_i$ when $d(o,i)$ is odd. Given two probability measures $\mu^{(1)}$, $\mu^{(2)}$ on $\mathcal{X}^V$, we write $\mu^{(1)}\preceq\mu^{(2)}$ if there exists a coupling $\mu^{(1,2)}$ of these two measures, such that, letting $(\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)})\sim\mu^{(1,2)}$, $(\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)})\in\mathcal{X}^V\times\mathcal{X}^V$, we have $\boldsymbol{\sigma}^{(1)}\preceq\boldsymbol{\sigma}^{(2)}$ almost surely.

The next lemma states the mentioned monotonicity property.

Lemma 5.2.1. *Let $t\in\mathbb{N}$ and $\sigma^{(1)}_{V_{=t}},\sigma^{(2)}_{V_{=t}}\in\mathcal{X}^{V_{=t}}$ two configurations at level ℓ.*

If t is even and $\sigma^{(1)}_{V_{=t}}\preceq\sigma^{(2)}_{V_{=t}}$, then

$$\mu(\,\cdot\,|\{\boldsymbol{\sigma}_{V_{=t}}=\sigma^{(1)}_{V_{=t}})\preceq\mu(\,\cdot\,|\{\boldsymbol{\sigma}_{V_{=tx}}=\sigma^{(1)}_{V_{=t}})\,. \tag{5.2.2}$$

Viceversa, if t is odd and $\sigma^{(1)}_{V_{=t}}\preceq\sigma^{(2)}_{V_{=t}}$, then

$$\mu(\,\cdot\,|\{\boldsymbol{\sigma}_{V_{=t}}=\sigma^{(1)}_{V_{=t}})\succeq\mu(\,\cdot\,|\{\boldsymbol{\sigma}_{V_{=t}}=\sigma^{(1)}_{V_{=t}})\,. \tag{5.2.3}$$

It is useful to recall the definition of Gibbs measure on T with boundary condition f, cf. Definition 3.2.3.

5.3 The case of the Ising model

As we saw in the last section, the weak convergence method easily estabilishes the limit tree distribution when the functions $\psi_{\mathrm{e}}(\,\cdot\,,\,\cdot\,)$, $\psi_{\mathrm{v}}(\,\cdot\,)$ are close to constants.

In some cases however, one can use additional properties of the model to select one of the Gibbs measures with respect to the specification ψ, even if multiple exist. In this section we consider the special case of the ferromagnetic Ising model, i.e. $\psi_{\mathrm{e}}(\sigma,\sigma')=\exp(\beta\sigma\sigma')$, $\beta>0$, and $\psi_{\mathrm{v}}(\sigma)=\exp(B\sigma)$. Without loss of generality we can assume

Chapter 6

Algorithmic aspects

Appendix A

Proofs omitted from the main text

A.1 Proofs omitted from Chapter 2

A.1.1 Proof of Proposition 2.3.2

The proof amounts to checking the conditions of Erdös-Gallai theorem 1. Condition (a) holds by construction.

As for condition (b), ww will assume that $P_\ell > 0$ for some $\ell \geq 2$, because otherwise the proof is trivial. We also will relabel vertices in such a way that the degree sequence is decreasing $d_1 \geq d_2 \geq d_3 \geq \dots$. Let $\ell_{\max}(n) = \max\{\ell : \quad m_\ell^{(n)} > 0\}$. Assume that $\sum_{\ell=\ell_0+1}^{\ell_{\max}} m_\ell^{(n)} < k \leq \sum_{\ell=\ell_0}^{\ell_{\max}} m_\ell^{(n)}$, and let $q \equiv \sum_{\ell=\ell_0}^{\ell_{\max}} m_\ell^{(n)} - k + 1 \geq 1$.

Let $\overline{P} = \sum_{\ell\geq 0} \ell P_\ell$. First consider $k \geq 1 + \sqrt{n\overline{P}}$. Then

$$\sum_{i=1}^{n} d_i = \sum_{\ell=1}^{\ell_{\max}} m_\ell^{(n)} \ell \leq n\overline{P} \leq k(k-1)\,, \tag{A.1.1}$$

which verify Erdös-Gallai condition in this case.

Next consider $k \leq 1 + \sqrt{n\overline{P}} \leq \sqrt{2n\overline{P}}$. Then, by definition of k

$$\sqrt{2n\overline{P}} \geq \sum_{\ell=\ell_0+1}^{\ell_{\max}} \lfloor nP_\ell \rfloor\,, \tag{A.1.2}$$

which can only happen if $\ell_0 \geq \ell_0(n)$ with for some $\ell_0(n)$ dependent only on P and such that $\ell_0(n) \to \infty$ as $n \to \infty$. On the other hand

$$\sum_{i=1}^{k} d_i = \sum_{\ell=\ell_0+1}^{\ell_{\max}} m_\ell^{(n)} \ell + q\ell_0 \tag{A.1.3}$$

$$\leq n\Big(\sum_{\ell=\ell_0}^{\infty} P_\ell \ell\Big)\,. \tag{A.1.4}$$

Since $\sum_{\ell=0}^{\infty} P_\ell \ell < \infty$, we have $\lim_{n\to\infty} \sum_{\ell=\ell_0(n)}^{\infty} P_\ell \ell = 0$, and hence $\sum_{i=1}^{k} d_i = o(n)$. On the other hand, consider the right hand side of the Erdös-Gallai condition, and fix ℓ independent of n and such that $P_\ell > 0$. Using the fact that $\ell_0(n) > \ell$ for all n large enough, we have

$$\sum_{i=k+1}^{n} \min(d_i, k) \geq m_\ell^{(n)} = nP_\ell + o(n)\,. \tag{A.1.5}$$

and therefore, for ll n large enough

$$\sum_{i=1}^{k} d_i < \sum_{i=k+1}^{n} \min(d_i, k)\,. \tag{A.1.6}$$

This finishes the proof.

A.1.2 Finishing the proof of Proposition 2.4.2

In this appendix we provide the details of step (ii) in the proof of Proposition 2.4.2. This requires to prove the mass-transportation principle (2.4.1) for the limiting random rooted graph, given that it holds for the finite graphs $(G_n, \boldsymbol{I}_n)$.. First, notice that it is suggicient to prove it for functions f such that $f(G,u,v) = 0$ unless $d_G(u,v) \le M$ for some fixed M. Indeed, any f can be obtained as monotone increasing limit of such functions, by setting $f_M(G,u,v) = f(G,u,v)\,\mathbb{I}(d_G(u,v) \le M)$.

Next define the set of simple functions

$$\mathsf{SF}_+(M) = \Big\{ \sum_{\ell=1}^{L} \alpha_\ell\, \mathbb{I}\big([G,u,v] \in A_\ell\big)\mathbb{I}(d_G(u,v) \le M) \;:\quad \alpha_\ell \in \mathbb{R}_{>0}, A_\ell \in \mathsf{B}(\mathcal{G}_{**}) \Big\}\,. \tag{A.1.7}$$

Any function f supported on $d_G(u,v) \le M$ can be obtained as monotone limit of gunctions in $\mathsf{SF}_+(M)$, and hence it is sufficient to prove the claim for such functions. By linearity, it is sufficient to prove it for any f of the type $f(G,u,v) = \mathbb{I}\big([G,u,v] \in A\big)\,\mathbb{I}(d_G(u,v) \le M)$.

It is easy to check that the set of A's for which it holds is closed by countable unions and complements. Hence, it is sufficient to consider a collection of generators of the σ-algebra $\mathsf{B}(\mathcal{G}_{**})$. We take this to be the collection of sets $A_{t,(H,i,j)} = \{(G,u,v) : B_t(u,v;G) \simeq (H,i,j)\}$ where $t \in \mathbb{N}$ is arbitrary, and (H,i,j) is an arbitrary depth t doubly rooted graphs with distance beteen the roots at most M.

For any set $A_{t,(H,i,j)}$, the mass-transportation principle reads

$$\mathbb{E}\Big[\sum_{v \in B_M(\boldsymbol{o},\boldsymbol{G},)} \mathbb{I}\Big(B_t(\boldsymbol{G},\boldsymbol{o},v) \simeq (H,i,j)\Big)\Big] = \mathbb{E}\Big[\sum_{v \in B_M(\boldsymbol{o},\boldsymbol{G},)} \mathbb{I}\Big(B_t(\boldsymbol{G},v,\boldsymbol{o}) \simeq (H,i,j)\Big)\Big]\,. \tag{A.1.8}$$

This only depends on the law of $B_{M+t}(\boldsymbol{G},\boldsymbol{o})$, and therefore holds that G_n converges locally to $(\boldsymbol{G},\boldsymbol{o})$.

Appendix B

Summary of notations

$|S|$: the size of set S.
$\doteq$: Equal to leading exponential order
Δ_S : The simplex of probability distributions over S
$d_G(i,j)$: graph distance on G
$\Delta_{\mathcal{S}}$: The polytope of probability distributions over the finite set $\mathcal{S}$.
$\overline{\mathbb{R}}$: The completed real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$

Bibliography

[AL07] David Aldous and Russell Lyons, *Processes on unimodular random networks*, Electron. J. Probab **12** (2007), no. 54, 1454–1508.

[AN04] Dimitris Achlioptas and Assaf Naor, *The two possible values of the chromatic number of a random graph*, Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, ACM, 2004, pp. 587–593.

[AS04] David Aldous and J Michael Steele, *The objective method: Probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, Springer, 2004, pp. 1–72.

[BD12] Anirban Basak and Amir Dembo, *Ferromagnetic ising measures on large locally tree-like graphs*, arXiv preprint arXiv:1205.4749 (2012).

[BGT10] Mohsen Bayati, David Gamarnik, and Prasad Tetali, *Combinatorial approach to the interpolation method and scaling limits in sparse random graphs*, Proceedings of the 42nd ACM symposium on Theory of computing, ACM, 2010, pp. 105–114.

[Big93] Norman Biggs, *Algebraic graph theory*, Cambridge University Press, 1993.

[Big98] ———, *Constructions for cubic graphs with large girth*, Journal of Combinatorics **5** (1998), 1–26.

[BLS12] Itai Benjamini, Russell Lyons, and Oded Schramm, *Unimodular random trees*, arXiv preprint arXiv:1207.1752 (2012).

[Bol78] Béla Bollobás, *Extremal graph theory*, DoverPublications. com, 1978.

[Bol80] Bela Bollobas, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European Journal of Combinatorics **1** (1980), 311316.

[Bol01] Béla Bollobás, *Random graphs*, vol. 73, Cambridge university press, 2001.

[BS01] I Benjamini and O Schramm, *Recurrence of distributional limits of finit planar graphs*, Electronic Journal of Probability **6** (2001), 1–13.

[Coo71] Stephen A Cook, *The complexity of theorem-proving procedures*, Proceedings of the third annual ACM symposium on Theory of computing, ACM, 1971, pp. 151–158.

[CT12] Thomas M Cover and Joy A Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[DM10] Amir Dembo and Andrea Montanari, *Gibbs measures and phase transitions on sparse random graphs*, Brazilian Journal of Probability and Statistics **24** (2010), no. 2, 137–211.

[DMS11] Amir Dembo, Andrea Montanari, and Nike Sun, *Factor models on locally tree-like graphs*, arXiv preprint arXiv:1110.4821 (2011).

[DS05] Irit Dinur and Samuel Safra, *On the hardness of approximating minimum vertex cover*, Annals of Mathematics (2005), 439–485.

[EL10] Gábor Elek and Gábor Lippner, *Sofic equivalence relations*, Journal of Functional Analysis **258** (2010), no. 5, 1692–1708.

[FKS89] Joel Friedman, Jeff Kahn, and Endre Szemeredi, *On the second eigenvalue of random regular graphs*, Proceedings of the twenty-first annual ACM symposium on Theory of computing, ACM, 1989, pp. 587–598.

[Geo11] Hans-Otto Georgii, *Gibbs measures and phase transitions*, vol. 9, Walter de Gruyter, 2011.

[GJ79] Michael R Garey and David S Johnson, *Computers and intractability*, vol. 174, Freeman New York, 1979.

[Has96] Johan Hastad, *Clique is hard to approximate within* $n^{1-\epsilon}$, Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on, IEEE, 1996, pp. 627–636.

[HTF09] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, Springer, 2009.

[Isi25] Ernst Ising, *Beitrag zur theorie des ferromagnetismus*, Zeitschrift für Physik A Hadrons and Nuclei **31** (1925), no. 1, 253–258.

[Jan09] Svante Janson, *The probability that a random multigraph is simple*, Combinatorics, probability & computing **18** (2009), no. 1-2, 205–225.

[JLK00] Svante Janson, Tomasz Luczak, and VF Kolchin, *Random graphs*, Cambridge Univ Press, 2000.

[Kar72] Richard Karp, *Reducibility among combinatorial problems*, Complexity of Computer Computations (R. Miller and J. Thatcher, eds.), Plenum Press, 1972, pp. 85–103.

[KF09] Daphne Koller and Nir Friedman, *Probabilistic graphical models: principles and techniques*, MIT press, 2009.

[Lau96] Steffen L Lauritzen, *Graphical models*, Oxford University Press, 1996.

[LPP95] Russell Lyons, Robin Pemantle, and Yuval Peres, *Ergodic theory on galton-watson trees: speed of random walk and dimension of harmonic measure*, Ergodic Theory and Dynamical Systems **15** (1995), no. 3, 593.

[MM09] M. Mézard and A. Montanari, *Information, Physics and Computation*, Oxford, 2009.

[MMS12] Andrea Montanari, Elchanan Mossel, and Allan Sly, *The weak limit of ising models on locally tree-like graphs*, Probability Theory and Related Fields **152** (2012), no. 1-2, 31–51.

[RU08] T.J. Richardson and R. Urbanke, *Modern Coding Theory*, Cambridge University Press, Cambridge, 2008.

[Wei06] Dror Weitz, *Counting independent sets up to the tree threshold*, Proceedings of the thirty-eighth annual ACM symposium on Theory of computing, ACM, 2006, pp. 140–149.

[WJ08] Martin J Wainwright and Michael I Jordan, *Graphical models, exponential families, and variational inference*, Foundations and Trends in Machine Learning **1** (2008), no. 1-2, 1–305.

[Wor81] Nicholas C Wormald, *The asymptotic distribution of short cycles in random regular graphs*, Journal of Combinatorial Theory, Series B **31** (1981), no. 2, 168–182.

[Wor99] ———, *Models of random regular graphs*, London Mathematical Society Lecture Note Series (1999), 239–298.