Five lectures on statistical physics methods in combinatorics

Will Perkins*

May 2023

Abstract

How can one use ideas, tools, and intuition from statistical physics to solve problems in extremal, probabilistic, and enumerative combinatorics? These lecture notes give an introduction to this topic, beginning with some basics of statistical physics, showing the connections between statistical physics and combinatorics, and then developing some statistical physics based probabilistic methods in combinatorics.

Contents

*School of Computer Science, Georgia Institute of Technology

1 Fundamental of Statistical Physics

In this lecture we go over the basic objects and questions of statistical physics and introduce much of the terminology we'll use going forward.

Statistical physics is interested in many different types of physical phenomena and the methods of statistical physics have been applied to many problems outside of physics. In these notes we will be primarily concerned with one of the oldest and most fundamental questions in the field: how do the microscopic interactions between molecules or particles influence the macroscopic behavior of a gas, fluid, magnetic material, or other system. Of particular interest is the phase transition phenomenon, in which small changes in the microscopic interactions lead to qualitative changes in the macroscopic behavior.

1.1 Gibbs measures and partition functions

While statistical physics encompasses many different types of models, for now we will focus on *spin models on graphs*. This setting is not only rich enough to exhibit many interesting phase transition phenomena, but it is also most relevant to combinatorial applications. We first define a general model then give some important examples.

Fix a finite set of spins Ω. Typical choices include

- $\Omega = \{0, 1\}$
- $\Omega = \{-1, 1\}$
- $\Omega = \{\text{Red, Blue, Green}\}$
- $\Omega = \{1, \ldots, q\}$.

For a graph $G = (V, E)$, the set of possible *spin configurations* is Ω^V, all assignments of spins from Ω to the vertices of G.

Next define an *energy function* (or Hamiltonian) from $\Omega^V \to \mathbb{R} \cup \{+\infty\}$ that respects the graph structure (as a sum of functions vertex spins and pairs of spins across edges):

$$H(\sigma) = \sum_{v \in V} f(\sigma_v) + \sum_{(u,v) \in E} g(\sigma_u, \sigma_v)$$

where $f : \Omega \to \mathbb{R}$ and $g : \Omega \times \Omega \to \mathbb{R} \cup \{+\infty\}$ is symmetric. If g takes the value $+\infty$ we say that there is a *hard constraint* in the model.

The *partition function* of the spin model at inverse temperature $\beta \in \mathbb{R}$ is

$$Z_G(\beta) = \sum_{\sigma \in \Omega^V} e^{-\beta H(\sigma)} . \tag{1}$$

The *Gibbs measure* is the probability distribution on Ω^V defined by

$$\mu_{G,\beta}(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z_G(\beta)} , \tag{2}$$

where the partition function plays the role of the normalizing constant ensuring $\mu_{G,\beta}$ is a probability distribution.

Example 1. *The ferromagnetic Ising model with no external field. Let* $\Omega = \{+1, -1\}$, $f(\sigma_v) \equiv 0$, $g(\sigma_u, \sigma_v) = -\sigma_u\sigma_v$. *Then* $\mu_{G,\beta}(\sigma)$ *is proportional to* $e^{2\beta M(G,\sigma)}$ *where* $M(G,\sigma)$ *is the number of edges of* G *whose endpoints receive the same spin under* σ *(monochromatic edges). If* $\beta > 0$ *then configurations with more monochromatic edges are preferred; this case is the ferromagnetic case.*

To obtain the antiferromagnetic Ising model we take $g(\sigma_u, \sigma_v) = \sigma_u\sigma_v$*; then we prefer edges with different spins on their endpoints.*

We can add a (uniform) external field by taking $f(\sigma_v) = a\sigma_v$. *If* $a < 0$ *we prefer* $+1$ *spins.*

Consider the ferromagnetic Ising model (with no external field) on $(\mathbb{Z}/n\mathbb{Z})^d$*, the* d*-dimensional discrete torus. If* β *is large, then we more strongly prefer vertices to have the same spin as their neighbors.*

Example 2. *Hard-core model (hard-core lattice gas).* $\Omega = \{0, 1\}$, $f(\sigma_v) = -\sigma_v \cdot \log\lambda$, $g(1,1) = +\infty$, $g(0,0) = g(0,1) = g(1,0) = 0$. $\beta = 1$.

More conveniently, we can associate an independent set I *of* G *(set of vertices that spans no edges) with its indicator vector* σ *so that* $\sigma_v = \mathbf{1}_{v\in I}$. *Then*

$$\mu_G(I) = \frac{\lambda^{|I|}}{Z_G(\lambda)} \tag{3}$$

where

$$Z_G(\lambda) = \sum_{I\in\mathcal{I}(G)} \lambda^{|I|} = \sum_{k\geq 0} \lambda^k i_k(G)\,. \tag{4}$$

Here $\mathcal{I}(G)$ *is the set of all independent sets of* G *and* $i_k(G)$ *is the number of independent sets of size* k *in* G. *The parameter* λ *is the fugacity or activity and governs whether we prefer small or large independent sets.*

The partition function of the hard-core model, $Z_G(\lambda)$*, is also known as the independence polynomial in combinatorics.* $Z_G(\lambda)$ *encodes a lot of combinatorial information.* $Z_G(1) = |\mathcal{I}(G)|$*, the number of independent sets of* G*; the highest order term in* $Z_G(\lambda)$ *is the independence number of* G*,* $\alpha(G)$*, and its coefficient is the number of maximum size independent sets of* G.

Take for example, $G = C_4$*, the cycle on* 4 *vertices. Then* $Z_G(\lambda) = 1 + 4\lambda + 2\lambda^2$. *For any* $v \in V(G)$*, the probability of choosing* $I = \{v\}$ *in the hard-core model on* G *is* $\frac{\lambda}{1+4\lambda+2\lambda^2}$.

Example 3. *The Potts model. The Potts model is a generalization of the Ising model to* $q \geq 2$ *spins (or colors); that is,* $\Omega = \{1, \ldots, q\}$. *Configurations are assignments of* q *colors to the vertices of a graph. A configuration is chosen with probability* $\frac{e^{\beta M(G,\sigma)}}{Z_G(q,\beta)}$ *where* $M(G,\sigma)$ *is the number of monochromatic edges of* G *under the coloring* σ. $\beta \geq 0$ *is the ferromagnetic and* $\beta \leq 0$ *the antiferromagnetic case.*

In general, the inverse temperature β controls the strength of the interaction in the model. At high temperatures (small β, local interactions are weak, while at low temperatures (large β), local interactions are strong. In other words,

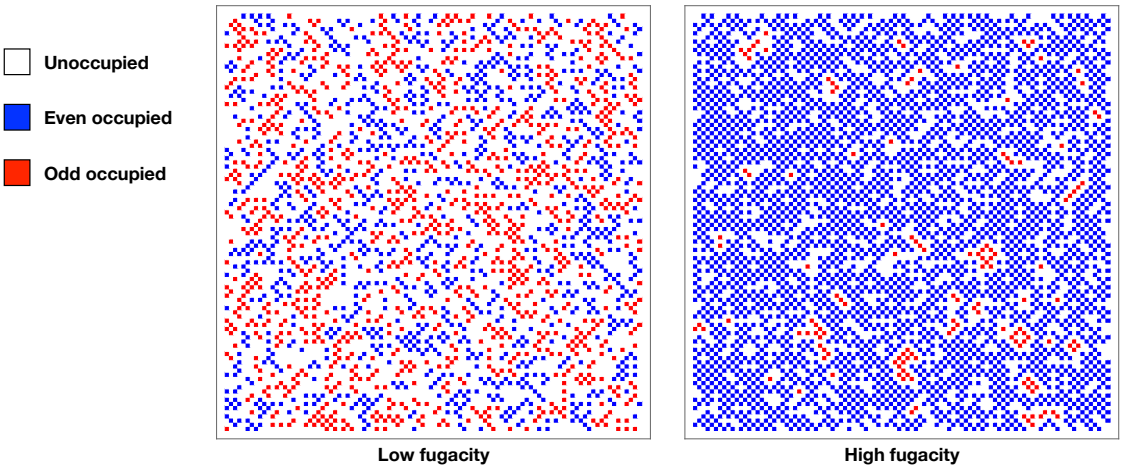

Figure 1: Two instances of the hard-core model on $\mathbb{Z}^2$

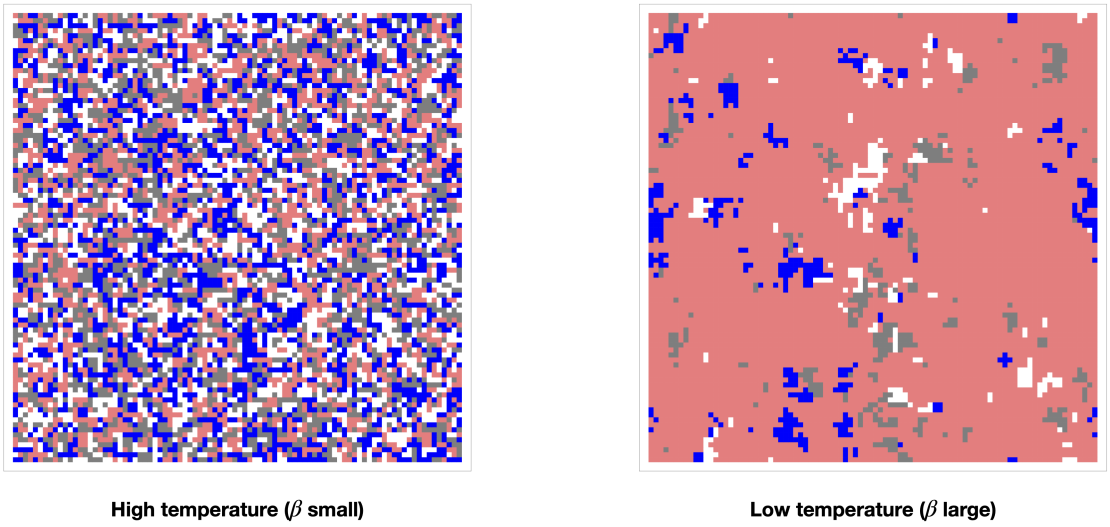

Figure 2: Two instances of the 4-color ferromagnetic Potts model on $\mathbb{Z}^2$

- At $\beta = 0$ (infinite temperature) the Gibbs measure is simply uniform on Ω^V and so each vertex receives a uniform and independent spin from Ω.
- At $\beta = +\infty$ (zero temperature), the Gibbs measure is uniform over the *ground states* of the model: the configurations σ that minimize the energy $H(\cdot)$. For Gibbs measures on lattices like $\mathbb{Z}^d$, it is often[1] easy to understand the ground states (e.g. all even/all odd occupied configurations for hard-core; monochromatic configurations for Ising/Potts). For general graphs though, this need not be the case. In particular, finding a ground state in the hard-core model on a general graph is the max independent set problem, a classic NP-hard problem. Similarly, finding and understanding the ground states of anti-ferromagnetic models on random graphs is a challenging problem, both mathematically

[1]But not always! See e.g. the Edwards-Anderson model.

and algorithmically.

- Taking β positive and finite interpolates between independence (pure entropy) and optimization (pure energy). Understanding the Gibbs measure and partition function at positive temperature requires balancing energy and entropy.

From the combinatorics perspective, the Gibbs measure interpolates between two objects we study a lot: a purely random object (say a uniformly random cut in a graph) and an extremal object (the max cut or min cut in a graph).

An important theme in statistical physics is that the qualitative properties of the two ends of the interpolation persist at positive, finite temperatures: a weakly interacting system has many of the properties of an independent system, while a strongly interacting system correlates strongly with an extremal object. The switch from one qualitative regime to the other is a phase transition, one of the main topics of statistical physics.

Not all Gibbs measures are spin models on graphs. Some other important examples include the following, some of which also have particular importance in combinatorics.

1. The monomer-dimer model. Allowed configurations are matchings in G, with $\mu_{G,\lambda}(M) = \frac{\lambda^{|M|}}{Z_G^{\text{match}}(\lambda)}$. 'Dimers' are edges in the matching while 'monomers' are unmatched vertices. The monomer-dimer model is the hard-core model on the line graph of G. This is an example of an *edge coloring model* (see e.g. [76]).

2. Spin models on hypergraphs. Here the energy function H is a sum of functions on vertices and functions on hyperedges. For example, we can consider the hard-core model on a hypergraph $G = (V, E)$. Configurations are subsets S of vertices that contain no hyperedge, weighted by $\lambda^{|S|}$. An interaction on a hyperedge of size > 2 is called a *multibody interaction*.

3. Gibbs point processes (continuum models). The hard sphere model [54] is a continuum model of a gas and perhaps the original model in statistical mechanics. This is a probability distribution over packings of equal-sized spheres in Euclidean space.

Gibbs measures arise in many other contexts beyond statistical physics, including machine learning, Bayesian statistics, mathematical biology, and many others. They are sometimes called probabilistic graphical models, Markov random fields, log linear models, exponential families, or Boltzmann distributions.

1.2 Markov random fields

A Gibbs measure defined as a spin model on a graph (with interactions across edges) is a *Markov random field* with respect to the graph. This means that it satisfies the following spatial Markov property. Let A, B, S be disjoint subsets of vertices of G so that S separates A and B: any path from a vertex $a \in A$ to a vertex $b \in B$ must pass through S. Then with respect to the Gibbs measure μ_G, if we condition on the spins in S, $\sigma_S = \tau_S$, the spins σ_A in A and σ_B in B are independent.

A special case of the spatial Markov property is that the spin at v, σ_v, is independent of the other spins in the graph conditioned on the spins of its neighbors $\sigma_{N(v)}$. We can write down a formula for the distribution of σ_v given that $\sigma_{N(v)} = \tau_{N(v)}$:

$$\mu_G(\sigma_v = \omega | \sigma_{N(v)} = \tau_{N(v)}) = \frac{\exp\left[-\beta\left(f(\omega) + \sum_{u \in N(v)} g(\omega, \tau_u)\right)\right]}{\sum_{\omega' \in \Omega} \exp\left[-\beta\left(f(\omega') + \sum_{u \in N(v)} g(\omega', \tau_u)\right)\right]}. \tag{5}$$

For the hard-core model, this formula simplifies considerably. Say v is blocked with respect to an independent set I if $N(v) \cap I \neq \emptyset$ and unblocked otherwise. In particular, v can only be in I if it is unblocked. Then we have $\mu_{G,\lambda}(v \in I | v \text{ unblocked}) = \frac{\lambda}{1+\lambda}$.

This spatial Markov property makes it easy to compute the partition function of a spin model on a tree.

1.3 Moments, cumulants, and derivatives of the log partition function

The energy $H(\cdot)$ is a *local function*, with respect to the geometry of the underlying graph: it is a sum of functions on vertices and edges. As a random variable, $H(\sigma)$ is a locally computable statistic or *observable* of the model (it can be computed from σ by summing over vertices and edges). For instance in the hard-core model $H(\sigma)$ counts the size of an independent set while in the Ising and Potts models $H(\sigma)$ counts the number of monochromatic edges (or equivalently the number of crossing edges of a cut).

As we will see, understanding the behavior of the random variable $H(\sigma)$ for large underlying graphs can tell us a lot about the behavior of the model and any phase transitions that might occur as parameters are varied.

To begin to understand the random variable $H(\sigma)$ we'd like to know its expectation, variance, and then perhaps higher moments.

We can write down the expectation:

$$\begin{aligned}
\mathbb{E}[H] &= \sum_{\sigma \in \Omega^V} H(\sigma) \mu_G(\sigma) \\
&= \sum_{\sigma \in \Omega^V} H(\sigma) \frac{e^{-\beta H(\sigma)}}{Z_G(\beta)} \\
&= \frac{\sum_{\sigma \in \Omega^V} H(\sigma) e^{-\beta H(\sigma)}}{Z_G(\beta)} \\
&= \frac{-\frac{\partial}{\partial \beta} Z_G(\beta)}{Z_G(\beta)} \\
&= -\frac{\partial}{\partial \beta} \log Z_G(\beta).
\end{aligned} \tag{6}$$

In the case of the hard-core model (due to the slightly different form of the distribution), we have

$$\mathbb{E}_{G,\lambda}|\mathbf{I}| = \sum_{I \in \mathcal{I}(G)} |I| \mu_{G,\lambda}(I)$$

$$
\begin{aligned}
&= \sum_{I \in \mathcal{I}(G)} |I| \frac{\lambda^{|I|}}{Z_G(\lambda)} \\
&= \frac{1}{Z_G(\lambda)} \sum_{I \in \mathcal{I}(G)} |I| \lambda^{|I|} \\
&= \frac{\lambda \frac{\partial}{\partial \lambda} Z_G(\lambda)}{Z_G(\lambda)} \\
&= \lambda \frac{\partial}{\partial \lambda} \log Z_G(\lambda) \,. \qquad (7)
\end{aligned}
$$

This is an important equation that we will use throughout this lecture series. We now see the same fact in more generality.

For a random variable X, its moment generating function is $M_X(t) = \mathbb{E}e^{tX}$. Its cumulant generating function is its logarithm of the moment generating function $K_X(t) = \log \mathbb{E}e^{tX}$. The *cumulants* of X are the coefficients in the Taylor series around 0:

$$
K_X(t) = \sum_{n=1}^{\infty} \kappa_n(X) \frac{t^n}{n!} \,. \qquad (8)
$$

Or in other words, $\kappa_n(X) = K_X^{(n)}(0)$.

Cumulants are related to moments but are often more convenient to work with in statistical physics. For example, the cumulants of a Gaussian $N(\mu, \sigma^2)$ are $\kappa_1 = \mu, \kappa_2 = \sigma^2$, $\kappa_k = 0$ for $k \geq 3$ (and the vanishing of the higher cumulants characterizes the Gaussian distribution). The cumulants of a Poisson(λ) random variable are all λ.

The partition function (1) looks similar to a moment generating function, and in fact we can write the moment generating function of the random variable $H(\sigma)$ as a ratio of partition functions:

$$
\begin{aligned}
\mathbb{E}e^{tH(\sigma)} &= \frac{1}{Z_G(\beta)} \sum_{\sigma} e^{tH(\sigma)} e^{-\beta H(\sigma)} \\
&= \frac{Z_G(\beta - t)}{Z_G(\beta)} \,, \qquad (9)
\end{aligned}
$$

which gives

$$
K_H(t) = \log Z_G(\beta - t) - \log Z_G(\beta) \,. \qquad (10)
$$

Thus taking derivatives of $\log Z(\beta)$ in β we obtain the cumulants of the random variable $H(\sigma)$. Above we computed

$$
\frac{d}{d\beta} \log Z(\beta) = -\mathbb{E}H(\sigma) = -\kappa_1(H) \,.
$$

We can do a similar calculation with the second derivative:

$$
\frac{d^2}{d\beta^2} \log Z(\beta) = \frac{\frac{d^2}{d\beta^2} Z(\beta)}{Z(\beta)} - \left(\frac{\frac{d}{d\beta} Z(\beta)}{Z(\beta)} \right)^2
$$

$$
\begin{aligned}
&= \mathbb{E}[H(\sigma)^2] - (\mathbb{E}H(\sigma))^2 \\
&= \operatorname{var}(H(\sigma)) \\
&= \kappa_2(G)\,.
\end{aligned}
$$

The higher derivatives recover the higher cumulants of the energy:

$$
\frac{d^k}{d\beta^k} \log Z(\beta) = (-1)^k \kappa_k(H)\,. \tag{11}
$$

1.4 Multivariate partition functions

It is often useful and interesting to generalize partition functions or graph polynomials from a single variable to multiple variables (for a great discussion of this, see [70]).

The multivariate hard-core partition function (or multi-variate independence polynomial) of a graph G is

$$
Z_G(\boldsymbol{\lambda}) = \sum_{I \in \mathcal{I}(G)} \prod_{v \in I} \lambda_v\,, \tag{12}
$$

where $\boldsymbol{\lambda} = (\lambda_v)_{v \in V(G)}$ is a vector of activities indexed by the vertices of G. This is a generalization since we can obtain $Z_G(\lambda)$ by taking $\boldsymbol{\lambda} \equiv \lambda$. $Z_G(\boldsymbol{\lambda})$ is a multi-linear polynomial.

We can do the same for a general Gibbs measure. If our original model has partition function

$$
Z_G(\beta) = \sum_{\sigma \in \{\pm 1\}^V} e^{-\beta H(\sigma)}
$$

we can add *non-uniform external fields* $\mathbf{t} = (t_v)_{v \in V(G)}$ and set

$$
Z_G(\beta, \mathbf{t}) = \sum_{\sigma \in \{\pm 1\}^V} e^{-\beta H(\sigma)} \prod_{v \in V(G)} e^{t_v \sigma_v}\,. \tag{13}
$$

We recover the original partition function by taking $\mathbf{t} \equiv 0$.

More generally if we have a q-spin model; that is, $|\Omega| = q$, we can pick one distinguished spin ω, and put non-uniform external fields in the direction of ω. This gives the partition function

$$
Z_G(\beta, \mathbf{t}) = \sum_{\sigma \in \Omega^V} e^{-\beta H(\sigma)} \prod_{v \in V(G)} e^{t_v \mathbf{1}_{\sigma_v = \omega}} \tag{14}
$$

and Gibbs measure

$$
\mu_{G,\beta,\mathbf{t}}(\sigma) = \frac{e^{-\beta H(\sigma)} \prod_{v \in V(G)} e^{t_v \mathbf{1}_{\sigma_v = \omega}}}{Z_G(\beta, \mathbf{t})}\,. \tag{15}
$$

External fields introduce a bias in favor (or against if the external field is negative) of a certain spin. The (log of the) fugacity λ in the hard-core model is an external field favoring (or penalizing) occupied vertices.

An external field is uniform if it is the same at every vertex. A non-uniform external field is *consistent* if the sign of the external field is the same at every vertex.

1.5 Marginals and correlations

Central to the statistical physics point of view is considering how *correlations* in a given model behave and how this behavior depends on the parameters. All of the discussion below pertains to general graphs, but for intuition it is helpful to keep in mind a graph like $\mathbb{Z}^2$ or $\mathbb{Z}^d$ with very natural geometry.

We will also focus here mostly on two-spin models, like Ising or hard-core where a probability distribution on the spin set Ω can be specified by its expectation.

We will use the following notation: $\sigma_v \in \Omega$ is the spin at vertex v in the configuration σ. For a subset of vertices $S \subseteq V(G)$, $\sigma_S \in \Omega^S$ is the spin assignment to S given by σ (in other words it is the restriction of σ to the coordinates given by S).

The *marginal* or *occupation probability* of a vertex v is $\mu_v = \mathbb{E}[\sigma_v]$; for instance, in the hard-core model $\mu_v = P(v \in \mathbf{I})$. (For a q-spin model like Potts the marginal would be a probability distribution on $\Omega = [q]$, or we could specify a single spin ω and ask for the marginal probability of ω: $P(\sigma_v = \omega)$).

For a pair of vertices u, v, the joint marginal is $\mu_{u,v} = \mathbb{E}[\sigma_u \sigma_v]$. In the hard-core model, this is $\mu_{uv} = \Pr_{G,\lambda}[u, v \in \mathbf{I}]$. (For a q-spin model, the joint marginal would be described by a $q \times q$ matrix).

For a subset $S \subseteq V$, the joint marginal is $\mu_S = \mathbb{E}[\prod_{v \in S} \sigma_v]$. If $|S| = k$, then μ_S is also called the *k-point correlation function.*

We are often interested in how strong correlations between spins are, as a function of the parameters of the model and the graph distance between vertices. A natural way to measure the correlation between the spins at vertices u and v is to compute a covariance:

$$\kappa(u, v) = \mu_{uv} - \mu_u \mu_v \, .$$

If σ_u and σ_v were independent then $\kappa(u, v)$ would be 0; if $\kappa(u, v)$ is small in absolute value then we can say σ_u and σ_v are weakly correlated. The quantity $\kappa(u, v)$ is called the *truncated 2-point correlation function.*

1.5.1 Exponential decay of correlations

In probability and combinatorics we are very happy to work with independent random variables. We can compute variances, prove Chernoff bounds, prove Central Limit Theorems. In many or most interesting situations, however, we are not working with independent random variables. Figuring out how to generalize tools from the independent case is an important topic in probability theory. (For example, Martingales and Azuma's inequality generalize Chernoff bounds).

An important heuristic in statistical physics is that a weakly interacting system is well approximated by a system of independent spins or particles. We will see several precise characterizations of this heuristic, including several notions of 'weakly interacting' and several notions of 'well approximated by'.

One notion of weakly interacting is that of decay of correlations.

Definition 1.1. *Let $\mathcal{G}$ be a family of graphs. We say a family of Gibbs measures μ_G, $G \in \mathcal{G}$, exhibits exponential decay of correlations if there exists constants $A, B > 0$ so that for all $G \in \mathcal{G}$, and all $u, v \in V(G)$,*

$$|\kappa(u,v)| \leq Ae^{-B\text{dist}(u,v)}\,, \tag{16}$$

where $\text{dist}(u,v)$ *is the graph distance from u to v in G.*

Note that exponential decay of correlations only makes sense for an infinite family of graphs (or for an infinite graph G if we define an infinite volume Gibbs measure, see Section 1.8). For any finite graph G (or any finite collection of finite graphs) we can always find A, B satisfying (16); what is important in the definition is that these constants are uniform over the collection of graphs $\mathcal{G}$.

Exercise 1. *Let P_n be the path graph on n vertices. Fix $\lambda > 0$. Show that hard-core model on the family of path graphs $\{P_n\}_{n\geq 1}$ exhibits exponential decay of correlations. Hint: write down a recursive equation for $Z_{P_n}(\lambda)$.*

1.6 Joint cumulants and truncated k-point correlation functions

We can also define truncated k-point correlation functions. To do that it will be helpful to define the joint cumulants of a collection of random variables.

The moment generating function for a collection of random variables $\vec{X} = (X_1, \ldots, X_k)$ defined on the same probability space is

$$M_{X_1,\ldots,X_k}(\mathbf{t}) = \mathbb{E}e^{\sum_{j=1}^k t_j X_j}\,,$$

a function from $\mathbb{R}^k \to \mathbb{R}$. The joint cumulant generating function is

$$K_{X_1,\ldots,X_k}(\mathbf{t}) = \log \mathbb{E}e^{\sum_{j=1}^k t_j X_j}\,.$$

The *joint cumulants* of $\vec{X}$ are the coefficients of the multivariate Taylor series for $K_{X_1,\ldots,X_k}(\mathbf{t})$ around $\mathbf{t} = \vec{0}$. More precisely, for non-negative integers $\ell_1, \ldots, \ell_k$, we define the joint cumulant

$$\kappa(X_1^{(\ell_1)}, \ldots, X_k^{(\ell_k)}) = \frac{\partial^{\sum_{j=1}^k \ell_j}}{\prod_j \partial t_j^{\ell_j}} K_{X_1,\ldots,X_k}(\mathbf{t})\big|_{\mathbf{t}=\vec{0}}\,.$$

Specializing to our multivariate spin model with partition function

$$Z_G(\beta, \mathbf{t}) = \sum_{\sigma \in \Omega^V} e^{-\beta H(\sigma)} \prod_{v \in V(G)} e^{t_v \sigma_v}\,,$$

we can define the *truncated k-point correlation function*

$$\kappa(u_1, \ldots, u_k) = \frac{\partial^k}{\partial t_{u_1} \cdots \partial t_{u_k}} \log Z_G(\beta, \mathbf{t})\big|_{\mathbf{t}=\vec{0}},$$

where $u_1, \ldots, u_k \in V(G)$.

Exercise 2. *Prove that*

$$\mu_v = \frac{\partial}{\partial t_v} \log Z_G(\beta, \mathbf{t})\big|_{\mathbf{t}=\vec{0}} \tag{17}$$

and

$$\mu_{uv} - \mu_u \mu_v = \kappa(u, v) = \frac{\partial^2}{\partial t_u \partial t_v} \log Z_G(\beta, \mathbf{t})\big|_{\mathbf{t}=\vec{0}}\,. \tag{18}$$

1.7 Dynamics

So far we have discussed the *equilibrium measure* μ_G. It is also very natural from several different perspectives (physical, computational, mathematical) to consider dynamics on the space of configurations Ω^V; that is a stochastic process on Ω^V whose behavior at large time scales looks like μ_G.

In this course we will study Markov chains on Ω^V whose stationary distribution is μ_G. A complete study of Markov chains is beyond the scope of this course; for reference see [53]. We recall a few important definitions here.

Definition 1.2. *A discrete-time, discrete space Markov chain on a finite or countable set Σ is a stochastic process $X_0, X_1, X_2, \ldots$ that satisfies the Markov property: the distribution of X_t conditioned on $X_0, \ldots, X_{t-1}$ equals the distribution of X_t conditioned on X_{t-1}.*

This means we can describe a Markov chain by the distribution ν_0 of the initial state X_0 and the transition matrix $P(\cdot, \cdot)$, defined by $P(x, y) = P(X_{n+1} = y | X_n = x)$. We can also define the t-step transition probabilities: $P^t(x, y) = P(X_{n+t} = y | X_n = x)$

Definition 1.3. *A Markov chain is irreducible if for every $x, y \in \Sigma$ there exists t so that $P^t(x, y) > 0$. A Markov chain is aperiodic if for every $x \in \Sigma$ the gcd of the set $\{t \geq 1 : P^t(x, x) > 0\}$ is 1.*

For example, the simple random walk on a connected graph is irreducible, but on a disconnected graph it is not. The simple random walk on a bipartite graph is not aperiodic. We can make any periodic Markov chain aperiodic by making it 'lazy': at each step, with probability $1/2$ (or some other positive constant) stay in the current state and with probability $1/2$ follow the law of the Markov chain.

Definition 1.4. *A probability distribution π on Σ is a stationary distribution for the Markov chain defined by P if for all $x \in \Sigma$,*

$$\pi(X) = \sum_{y \in \Sigma} \pi(y) P(y, x)\,.$$

In particular, if π is a stationary distribution and X_0 is distributed according to $\pi_0 = \pi$ then X_k is distributed according to π for all $k \geq 0$.

How can we find a stationary distribution? There is a useful technique for a special class of Markov chains. A Markov chain P is *reversible* with respect to π if for all $x, y \in \Sigma$,

$$\pi(x) P(x, y) = \pi(y) P(y, x)\,. \tag{19}$$

This is called the *detailed balance* equation.

If P is reversible with respect to π then π is a stationary distribution for P. While most (or perhaps all) the Markov chains we will study in these notes are reversible, non-reversible Markov chains can be very powerful tools for sampling (e.g. [9, 8]) but are often harder to analyze mathematically.

A fundamental result on Markov chains states that if a Markov chain is irreducible and aperiodic then there exists a unique probability distribution μ on Σ so that $\pi_n \to \pi$ as $n \to \infty$, where π_n is the distribution of X_n.

1.7.1 Markov chains for spin models

We now specialize to Markov chains for spin models on graphs (for a short introduction to the topic see [63]).

Here the state space will be the set of all possible configurations, Ω^V, or in the case of models with hard constraints, the set of all allowed configurations ($\mathcal{I}(G)$ in the case of the hard-core model on a graph G).

Our goal will be to find a Markov chain on Ω^V with stationary distribution μ_G, the Gibbs measure. There are many such Markov chains but for now we will be most interested in *local* Markov chains: Markov chains in which at most one (or at most a bounded number of) spins are changed in each step. This is for two reasons: such Markov chains may in some sense capture the way a physical system evolves over time and these Markov chains are often easy to implement computationally.

The Glauber dynamics

Given a configuration $X_t \in \Omega^V$ we obtain X_{t+1} as follows:

1. Pick $v \in V$ uniformly at random.
2. Resample the spin at v from the Gibbs measure μ_G conditioned on the spins at the other vertices.

If μ_G is a Gibbs measure with pairwise interactions across edges, the second step is equivalent to resampling conditioned on the spins of the neighbors of v.

Exercise 3. *1. Show that the Glauber dynamics are reversible with respect to the Gibbs measure μ_G.*

2. Show that the Glauber dynamics are irreducible and aperiodic. (Assume that $H(\sigma) \in [0, \infty)$ for all σ, or prove for the hard-core model).

Exercise 4. *Let $\Sigma_q(G)$ be the set of all proper q-colorings of G and suppose $|\Sigma_q| > 0$. Let $\mu_{G,q}$ denote the uniform distribution on $\Sigma_q(G)$.*

1. Give an example of graph G so that for some $q \geq 2$, $\Sigma_q(G)$ is disconnected under moves that change the color of one vertex at a time.

2. *Give conditions in terms of q and the maximum degree of Δ to ensure that $\Sigma_q(G)$ is connected under single spin updates (changing the color of one vertex at a time).*

3. *Describe precisely a single spin update Markov chain for sampling from $\mu_{G,q}$ (i.e. Glauber dynamics).*

1.7.2 Mixing times

Markov chains are widely used to sample from probability distributions on large sets Σ. If we want to sample from μ_G by running a Markov chain $X_1, X_2, \ldots$ with stationary distribution μ_G, an important question is how long we need to run the chain to get an approximately correct sample from μ_G.

To answer this we need a good measure of 'approximately correct'.

Definition 1.5. *Let μ_1 and μ_2 be two probability measures on the same sample space Σ with the same sigma-field $\mathcal{F}$. Then the total variation distance between μ_1 and μ_2 is*

$$\|\mu_1 - \mu_2\|_{TV} = \sup_{A \in \mathcal{F}} |\mu_1(A) - \mu_2(A)| .$$

For a discrete probability space, we have

$$\begin{aligned} \|\mu_1 - \mu_2\|_{TV} &= \frac{1}{2} \sum_{\sigma \in \Sigma} |\mu_1(\sigma) - \mu_2(\sigma)| \\ &= \sum_{\sigma \in \Sigma : \mu_1(\sigma) > \mu_2(\sigma)} \mu_1(\sigma) - \mu_2(\sigma) \\ &= \mu_1(\{\sigma : \mu_1(\sigma) > \mu_2(\sigma)) - \mu_2(\{\sigma : \mu_1(\sigma) > \mu_2(\sigma)) . \end{aligned}$$

With this definition we can define the Mixing time of a Markov chain.

Definition 1.6. *The ε-mixing time $\tau_{\text{mix}}(\varepsilon)$ of a Markov chain with stationary distribution π on Σ is*

$$\sup_{\pi_0 \in \mathcal{P}(\Sigma)} \min\{t \geq 0 : \|\pi_t - \pi\|_{TV} < \varepsilon\},$$

where π_t is the distribution of X_t given that X_0 has distribution π_0.

Note that for a given π_0, $\|\pi_t - \pi\|_{TV}$ is decreasing in t. We write τ_{mix} for $\tau_{\text{mix}}(1/4)$. This is because for any irreducible, aperiodic Markov chain with stationary distribution π, there are constants $C > 0$ and $\alpha \in (0, 1)$ so that

$$\sup_{\pi_0} \|\pi_t - \pi\|_{TV} \leq C\alpha^t .$$

This means that $\tau_{\text{mix}}(\varepsilon) = O(\log(\varepsilon^{-1})) \cdot \tau_{\text{mix}}$. For these facts and much more on mixing times see [53].

For a Gibbs measure μ_G on a graph G on n vertices we say a Markov chain is fast-mixing or mixes rapidly if τ_{mix} is bounded by a polynomial in n; that is, poly-logarithmic in the size

of the state space Ω^V. We say the Markov chain is slow mixing or torpidly mixing if τ_{mix} is superpolynomial in n (often exponential in some power of n).

Keep in mind that the size of the state space of a spin model on G is exponentially large in n. So in the context of random walks on graphs, a random walk that mixes in time $O(\log n)$ is fast mixing while a random walk with mixing time $\Omega(n^c)$ for some $c > 0$ is slow mixing. Random walks on expander graphs are fast mixing, while a random walk on a cycle, for instance, is slow mixing.

1.8 Phase transitions

A central concept in statistical physics is that of a *phase transition*.

1.8.1 Boundary conditions

It will be very useful to impose *boundary conditions* on a Gibbs measure: that is, fixing the spins on a subsets of vertices (the boundary) and studying the resulting conditional Gibbs measure. We will define and study the notion of boundary conditions for arbitrary graphs (with arbitrary boundaries) but it is useful to keep in mind a box in $\mathbb{Z}^d$ with its natural geometric boundary.

Specify a spin model with spin set Ω and energy function $H(\cdot)$. Let $G = (V, E)$ be a graph and let $S \subseteq V$ be a specified set of boundary vertices. Let $\tau_S \in \Omega^S$ be a fixed assignment of spins to the vertices in S. Then the Gibbs measure with boundary conditions τ_S is the probability distribution $\mu_G^{\tau_S}$ on Ω^V defined by

$$\mu_G^{\tau_S}(\sigma) = \frac{e^{-\beta H(\sigma)} \mathbf{1}_{\sigma_S = \tau_S}}{Z_G^{\tau_S}(\beta)}$$

where

$$Z_G^{\tau_S}(\beta) = \sum_{\sigma \in \Omega^V} e^{-\beta H(\sigma)} \mathbf{1}_{\sigma_S = \tau_S} .$$

In words, $\mu_G^{\tau_S}$ is the distribution of μ_G conditioned on the event $\{\sigma_S = \tau_S\}$.

In the case of the hard-core model, boundary conditions take a particularly simple form. If we specify that a vertex u is not in I, then this is the same as removing u from $V(G)$. If we specify that u is in I, then this is the same as removing $N(u)$ from $V(G)$ and keeping u as an occupied isolated vertex (which will have no effect on the other vertices of G).

1.8.2 Infinite volume limits

How can we define a Gibbs measure on an infinite graph like $\mathbb{Z}^d$ or the infinite Δ-regular tree? We cannot define it via (1) and (2) since the number of vertices is infinite.

One approach to make sense of a Gibbs measure on an infinite graph is that of Dobrushin, Lanford, and Ruelle based on conditional probabilities and boundary conditions.

A *specification* of a spin model is a fixed choice of parameters, say fixing λ in the hard-core model, or fixing the function f, g in the general formulation of a spin model on a graph.

For a given specification and a countably infinite graph G, we say a probability measure μ on $\Omega^{V(G)}$ is a Gibbs measure with the given specification if for every finite set $\Lambda \subset V(G)$ and every configuration $\tau \in \Omega^{V(G)}$, we have the following equality between conditional probabilities and finite volume Gibbs measures with boundary conditions:

$$\mu(\sigma_\Lambda = \cdot \big| \sigma_{\Lambda^c} = \tau_{\Lambda^c}) = \mu_\Lambda^{\tau_{\partial\Lambda}}(\cdot)\,, \tag{20}$$

where $\partial\Lambda$ is the external neighborhood of Λ, vertices in Λ^c that are joined to some vertex in Λ. In words, this condition says that the conditional distributions of μ induced on any finite set by a configuration on its complement is exactly the finite volume Gibbs measure with the same specification and the given boundary condition. For more details on the construction of infinite volume Gibbs measures see [66, 36, 28].

Under very general conditions, there always exists an infinite volume Gibbs measure with a given specification on an infinite graph G. The main question is whether there is only one: is the infinite volume Gibbs measure unique or not?

1.8.3 Notions of phase transition

1. Uniqueness / non-uniqueness. We say a phase transition occurs at β_c if for all ε small enough there exists a unique infinite volume Gibbs measure for $\beta = \beta_c - \varepsilon$ and multiple infinite volume Gibbs measures for $\beta = \beta_c + \varepsilon$.

2. Analyticity. We say a phase transition occurs at β_c if the infinite-volume pressure $p(\beta)$ is non-analytic at β_c.

3. Correlation decay. We say a phase transition occurs at β_c if for all ε small enough correlations decay exponentially fast when $\beta = \beta_c - \varepsilon$ and do not decay exponentially fast when $\beta = \beta_c + \varepsilon$.

1.8.4 Uniqueness methods

Here we list some approaches to proving absence of phase transition or uniqueness of Gibbs measure for a given specification. We will see some of these methods in more details in upcoming lectures.

- Cluster expansion
- Disagreement percolation
- Markov chain mixing
- Computational trees

1.8.5 Phase coexistence methods

Conversely, how does one show that multiple Gibbs measures exist for a given specification? There are a number of different techniques, depending on the model and the type of graph. For

infinite trees, one can set up a recursion for marginal probabilities and show that there exist multiple fixed points. For lattices like $\mathbb{Z}^d$, one usually uses a Peierls argument, bounding the number and 'cost' of contours separating the origin from infinity; a much more sophisticated form of this technique is Pirogov-Sinai theory.

Background reading

The textbook of Friedli and Velenik [28] covers many fundamentals of statistical physics models on lattices from a mathematical point of view. Ruelle's textbook [66] is a classic reference for both lattice and continuum systems. The lecture notes of Adams [1] cover the mathematical fundamentals of statistical physics. It is also well worth reading about statistical physics from a physics perspective; the lecture notes of Tong are a good place to start [77]. Finally, Okounkov's recent overview of the work of Hugo Duminil–Copin [59] gives a nice feeling for some exciting questions and recent results in statistical physics.

1.9 Exercises and examples

1.9.1 Solved examples

1. Let $Z_G(\lambda)$ be the hard-core partition function on a graph G. Let $\mathbf{I}$ be a random independent set from G distributed as $\mu_{G,\lambda}$, the hard-core model at fugacity λ. Let $\alpha_G(\lambda) = \mathbb{E}|\mathbf{I}|$.

   (a) Write an expression for $\alpha_G(\lambda)$ in terms of $\log Z_G(\lambda)$.

   (b) Prove that $\alpha_G(\lambda)$ is strictly increasing as a function of λ (assume the graph G has at least one vertex).

   (c) Write an expression for the number of maximum independent sets of G in terms of a limit involving $\log Z_G(\lambda)$.

   (d) Suppose G has n vertices and maximum degree Δ. Show that

   $$\mathrm{var}(|\mathbf{I}|) \geq cn$$

   where c is a constant that depends on Δ and λ. Hint(s): use the law of total variance; use the fact that G has an independent set of size at least $n/(\Delta+1)$.

   Solution:

   (a) We can write

   $$\begin{aligned}\alpha_G(\lambda) &= \sum_{I \in \mathcal{I}(G)} |I| \mu_G(I) \\ &= \frac{\sum_{I \in \mathcal{I}(G)} |I| \lambda^{|I|}}{Z_G(\lambda)} \\ &= \lambda \frac{\sum_{I \in \mathcal{I}(G)} |I| \lambda^{|I|-1}}{Z_G(\lambda)}\end{aligned}$$

$$= \lambda \frac{Z'_G(\lambda)}{Z_G(\lambda)}$$
$$= \lambda (\log Z_G(\lambda))' .$$

(b) We take the derivative of $\alpha_G(\lambda)$ with respect to λ:

$$\alpha'_G(\lambda) = \frac{\lambda Z_G(\lambda) Z''_G(\lambda) + Z_G(\lambda) Z'_G(\lambda) - \lambda (Z'_G(\lambda))^2}{Z_G(\lambda)^2}$$
$$= \frac{\mathrm{var}(|\mathbf{I}|)}{\lambda} > 0$$

since $\{|\mathbf{I}| = 0\}$ and $\{|\mathbf{I}| = 1\}$ both have positive probability when $\lambda > 0$.

(c) For $\lambda \gg 2^n$, $Z_G(\lambda) \approx i_{\alpha(G)} \lambda^{\alpha(G)}$; in particular,

$$1 \leq \frac{Z_G(\lambda)}{i_{\alpha(G)} \lambda^{\alpha(G)}} \leq 1 + \frac{2^n}{\lambda} .$$

We can recover $\alpha(G)$:

$$\alpha(G) = \lim_{\lambda \to \infty} \frac{\log Z_G(\lambda)}{\log \lambda} .$$

Then once we know $\alpha(G)$ we can recover $i_{\alpha(G)}$:

$$i_{\alpha(G)} = \lim_{\lambda \to \infty} \frac{Z_G(\lambda)}{\lambda^{\alpha(G)}} .$$

(d) Let $J \in \mathcal{I}(G)$ be an independent set of size at least $n/(\Delta + 1)$. Let $X = |\mathbf{I}|$. Let Y be the number of vertices of J 'unblocked' by $\mathbf{I}$; that is,

$$Y = |\{x \in J : N(x) \cap \mathbf{I} = \emptyset\}| .$$

The Law of Total Variance says

$$\mathrm{var}(X) = \mathbb{E}[\mathrm{var}(X | \mathbf{I} \cap J^c)] + \mathrm{var}[\mathbb{E}(X | \mathbf{I} \cap J^c)] .$$

By the spatial Markov property, given $\mathbf{I} \cap J^c$, the unblocked vertices of J appear in $\mathbf{I}$ independently with probability $\lambda/(1+\lambda)$, and so $\mathrm{var}(X | \mathbf{I} \cap J^c) = \frac{\lambda}{(1+\lambda)^2} \cdot Y$. On the other hand, any vertex in a max-degree Δ graph is unblocked with probability at least $\frac{1}{(1+\lambda)^\Delta}$ and so

$$\mathbb{E}Y \geq |J| \frac{1}{(1+\lambda)^\Delta} \geq \frac{n}{(\Delta+1)(1+\lambda)^\Delta} .$$

Putting this together gives

$$\mathrm{var}(X) \geq n \cdot \frac{\lambda}{(\Delta+1)(1+\lambda)^{\Delta+2}} .$$

2. Let G be a graph and $A \subset V(G)$ be a set of d vertices. Suppose you know that $\mu_{G,\lambda}(\mathbf{I} \cap A = \emptyset) = p$. Form the graph G' by adding a single vertex v to $V(G)$ and connecting v to each of the vertices of A.

   (a) Compute $\log Z_{G'}(\lambda) - \log Z_G(\lambda)$.

   (b) Compute $\mu_{G',v,\lambda}$ (the occupation probability of v in the hard-core model on G' at fugacity λ).

   (This is a baby version of the 'cavity method')

   Solution:

   (a) Independent sets of G' can be divided in two sets: those that contain v and those that do not. Independent sets that do not contain v are exactly the independent sets of G. Independent sets that contain v are independent sets of G that contain no neighbor of v, plus v. We can then write

$$Z_{G'}(\lambda) = Z_G(\lambda) + \lambda \sum_{I \in \mathcal{I}: I \cap N(v) = \emptyset} \lambda^{|I|} .$$

   Since

$$p = \mu_{G,\lambda}(\mathbf{I} \cap A = \emptyset) = \frac{\sum_{I \in \mathcal{I}: I \cap N(v) = \emptyset} \lambda^{|I|}}{Z_G(\lambda)}$$

   we have

$$\frac{Z_{G'}(\lambda)}{Z_G(\lambda)} = 1 + \lambda p$$

   and so $\log Z_{G'}(\lambda) - \log Z_G(\lambda) = \log(1 + \lambda p)$.

   (b) By a similar calculation,

$$\begin{aligned} \mu_{G',v,\lambda} &= \frac{\lambda \sum_{I \in \mathcal{I}: I \cap N(v) = \emptyset} \lambda^{|I|}}{Z_{G'}(\lambda)} \\ &= \frac{\lambda p Z_G(\lambda)}{(1 + \lambda p) Z_G(\lambda)} \\ &= \frac{\lambda p}{1 + \lambda p} . \end{aligned}$$

   (As a check, imagine v has no neighbors in G'; then this formula reduces to $\lambda/(1 + \lambda)$, the probability an isolated vertex is occupied).

2 Combinatorics and statistical physics

In this lecture we will look at combinatorics problems from the perspective of statistical physics and see the usefulness of some of statistical physics tools.

2.1 Terminology and perspectives

Some of the central questions in combinatorics can be summarized as follows.

1. **Extremal problems**: what is the largest structure of a given type?
2. **Counting**: how many structures of a given type are there?
3. **Typical structure**: what are the structural properties of a typical (uniformly random) structure of a given type?
4. **Stability**: are all (or most?) near-extremal structures of a given type close in some metric to an extremal structure?

Each of these questions has an analogue in statistical physics.

1. **Ground state**: which configuration σ minimizes $H(\sigma)$?
2. **Partition function**: compute or approximate the partition function Z.
3. **Gibbs measure**: what is the typical structure of a sample from the Gibbs measure μ?
4. **Order/disorder phase transition**: at positive temperature do samples from the Gibbs measure resemble (in some sense) a ground state?

These questions are essentially the same but using different language. On the other hand, the perspectives of combinatorics and statistical physics can be quite different: combinatorics is often concerned with optimizing over all graphs (or hypergraphs), while statistical physics is often focused on specific infinite graphs like $\mathbb{Z}^d$.

This difference in perspectives can often be very useful. Some important results in statistical physics would be very hard to prove without taking the worst-case combinatorics perspective (like the Lee–Yang [52] and Heilmann–Lieb [40] theorems). On the other hand, by focusing on specific graphs, statistical physicists can understand behavior of these models very precisely and provide intuition useful in other settings.

In the next section, we will prove a couple of results about independent sets in graphs, taking a statistical physics approach to combinatorics problems.

2.2 Independent sets in regular graphs

Which d-regular graph has the most independent sets? This question was first raised in the context of number theory by Andrew Granville, and an approximate answer was given by Noga Alon [3] who applied the result to problems in combinatorial group theory.

Jeff Kahn gave a tight answer in the case of d-regular bipartite graphs.

Theorem 2.1 (Kahn [49]). *Let $2d$ divide n Then for any d-regular, bipartite graph G on n vertices,*

$$i(G) \leq i(H_{d,n}) = \left(2^{d+1} - 1\right)^{n/2d},$$

where $H_{d,n}$ is the graph consisting of $n/2d$ copies of $K_{d,d}$.

In terms of the independence polynomial, we can rephrase this as follows. For any d-regular, bipartite G,

$$Z_G(1) \leq Z_{K_{d,d}}(1)^{n/2d},$$

or, more convenient from our perspective,

$$\frac{1}{|V(G)|} \log Z_G(1) \leq \frac{1}{2d} \log Z_{K_{d,d}}(1).$$

Work of Galvin and Tetali [34] and Zhao [80] extended this result to all values of the independence polynomial and all d-regular graphs.

Theorem 2.2 (Kahn; Galvin-Tetali; Zhao). *For all d-regular graphs G and all $\lambda > 0$,*

$$\frac{1}{|V(G)|} \log Z_G(\lambda) \leq \frac{1}{2d} \log Z_{K_{d,d}}(\lambda).$$

See Galvin's lecture notes on the entropy method [31] for an exposition of the proof of Theorem 2.1 and extensions. See also the recent work of Sah, Sawhney, Stoner, and Zhao [67, 68] for an extension to irregular graphs among other results.

The question of minimizing the number of (weighted) independent sets in a d-regular graph is somewhat simpler: the answer is the clique K_{d+1}, proved by Cutler and Radcliffe [21]; for a short proof see the exercises.

Among all d-regular graphs, the graph with the smallest scaled independence number is the clique K_{d+1}. If we impose the condition that G has no triangles, then it is not immediately clear which graph has the smallest independence number $\alpha(G)$.

Following Ajtai, and Komlós, and Szemerédi [2], Shearer proved the following.

Theorem 2.3 (Shearer [71]). *For any triangle-free graph G on n vertices of average degree at most d,*

$$\alpha(G) \geq (1 + o_d(1)) \frac{\log d}{d} n.$$

As a consequence, Shearer obtained the current best upper bound on the Ramsey number $R(3, k)$.

Corollary 2.4 (Shearer [71]). *The Ramsey number $R(3, k)$ satisfies*

$$R(3, k) \leq (1 + o_k(1)) \frac{k^2}{\log k}.$$

The random d-regular graph (conditioned on being triangle-free) satisfies

$$\alpha(G) = (1 + o_d(1))\frac{2\log d}{d} n\,,$$

and this is the smallest independence ratio known for a max degree d graph (in the large d limit), and so there is potentially a factor of 2 that could potentially be gained in Shearer's bound. This factor of 2 would immediately give a factor 2 improvement to the upper bound on $R(3,k)$.

Here we will use statistical physics methods to prove a strengthening of Theorem 2.2 and a result closely resembling Theorem 2.3 but for the average size of an independent set rather than the maximum size.

To start we define the *occupancy fraction* of the hard-core model to be the expected fraction of vertices in the random independent set:

$$\overline{\alpha}_G(\lambda) = \frac{1}{|V(G)|}\mathbb{E}_{G,\lambda}|I|\,.$$

The first theorem states that $K_{d,d}$ maximizes the occupancy fraction over all d-regular graphs and all λ.

Theorem 2.5 (Davies, Jenssen, Perkins, Roberts [23])**.** *For all $d \geq 2$, all $\lambda \geq 0$, and all d-regular graphs G,*

$$\overline{\alpha}_G(\lambda) \leq \frac{\lambda(1+\lambda)^d}{2(1+\lambda)^d - 1} = \overline{\alpha}_{K_{d,d}}(\lambda)\,.$$

This implies Theorem 2.2 as follows:

$$\begin{aligned}
\frac{1}{n}\log Z_G(\lambda) &= \frac{1}{n}\log Z_G(0) + \frac{1}{n}\int_0^\lambda (\log Z_G(t))'\,dt \\
&= 0 + \int_0^\lambda \frac{\overline{\alpha}_G(t)}{t}\,dt \\
&\leq \int_0^\lambda \frac{\overline{\alpha}_{K_{d,d}}(t)}{t}\,dt \\
&= \frac{1}{2d}\log Z_{K_{d,d}}(\lambda)\,.
\end{aligned}$$

The next theorem states that the expected density of an independent set in a triangle-free graph G of max degree d is at least $(1 + o_d(1))\frac{\log d}{d}$, the same bound that Shearer achieves for the maximum density (though he proves this with the weaker condition of average degree d). This result also yields a lower bound on the number of independent sets in a triangle-free graph.

Theorem 2.6 (Davies, Jenssen, Perkins, Roberts [24])**.** *For all triangle-free graph G of maximum degree d,*

$$\overline{\alpha}_G(1) \geq (1 + o_d(1))\frac{\log d}{d}\,.$$

Moreover,

$$i(G) \geq e^{\left(\frac{1}{2}+o_d(1)\right)\frac{\log^2 d}{d}n}.$$

The respective constants 1 *and* 1/2 *are best possible and attained by the random d-regular graph.*

We start by proving Theorem 2.5 for triangle-free graphs along with Theorem 2.6. Consider the hard-core model on a d-regular, triangle-free G on n vertices.

We imagine the following two-part experiment: pick a random independent set $\mathbf{I}$ from the hard-core model on G and independently pick a uniformly random vertex $\mathbf{v}$ from $V(G)$. We then will record some local information about $\mathbf{I}$ from the perspective of v. In particular, we say v is *uncovered* with respect to an independent set I if $N(v) \cap I = \emptyset$, and we will record $\mathbf{Y}$, the number of uncovered neighbors of $\mathbf{v}$ with respect to the random independent set $\mathbf{I}$. The random variable $\mathbf{Y}$ takes integer values between 0 and d and its distribution depends on both G and λ.

We will also use the following two facts that follow from the spatial Markov property.

Fact 1 $\Pr[v \in I | v \text{ uncovered}] = \frac{\lambda}{1+\lambda}$.

Fact 2 $\Pr[v \text{ uncovered} | v \text{ has } j \text{ uncovered neighbors}] = (1+\lambda)^{-j}$.

Fact 2 relies on the fact that G is triangle-free: the graph induced by the uncovered neighbors of v consists of isolated vertices.

Now with our two-part experiment in mind, we write $\overline{\alpha}_G(\lambda)$ in two ways:

$$\begin{aligned}
\overline{\alpha}_G(\lambda) &= \frac{1}{n}\sum_{v\in V(G)} \Pr[v \in \mathbf{I}] \\
&= \frac{1}{n}\frac{\lambda}{1+\lambda}\sum_{v\in V(G)} \Pr[v \text{ uncovered}] \quad \text{by Fact 1} \\
&= \frac{1}{n}\frac{\lambda}{1+\lambda}\sum_{v\in V(G)}\sum_{j=0}^{d} \Pr[v \text{ has } j \text{ uncovered neighbors}]\cdot(1+\lambda)^{-j} \quad \text{by Fact 2,}
\end{aligned}$$

and

$$\begin{aligned}
\overline{\alpha}_G(\lambda) &= \frac{1}{n}\frac{1}{d}\sum_{v\in V(G)}\sum_{u\sim v} \Pr[u \in \mathbf{I}] \quad \text{since } G \text{ is d-regular} \\
&= \frac{1}{n}\frac{1}{d}\frac{\lambda}{1+\lambda}\sum_{v\in V(G)}\sum_{u\sim v} \Pr[u \text{ uncovered}] \quad \text{by Fact 1.}
\end{aligned}$$

Recall $\mathbf{Y}$ is the number of uncovered neighbors of $\mathbf{v}$ with respect to $\mathbf{I}$. Now our two expressions for $\overline{\alpha}_G(\lambda)$ can be interpreted as expectations over $\mathbf{Y}$.

$$\overline{\alpha}_G(\lambda) = \frac{\lambda}{1+\lambda}\mathbb{E}_{G,\lambda}(1+\lambda)^{-\mathbf{Y}}$$

$$\overline{\alpha}_G(\lambda) = \frac{1}{d}\frac{\lambda}{1+\lambda}\mathbb{E}_{G,\lambda}\mathbf{Y}\,.$$

Thus the identity

$$\mathbb{E}_{G,\lambda}(1+\lambda)^{-\mathbf{Y}} = \frac{1}{d}\mathbb{E}_{G,\lambda}\mathbf{Y} \tag{21}$$

holds for all d-regular triangle-free graphs G.

Theorems 2.5 and 2.6 are optimization problems (maximization and minimization respectively) over the set of all d-regular graphs. Now we can *relax* these optimization problems: instead of maximizing or minimizating $\overline{\alpha}_G(\lambda)$ over all d-regular triangle-free graphs, we can maximize $\frac{\lambda}{1+\lambda}\mathbb{E}(1+\lambda)^{-\mathbf{Y}}$ over all distributions of random variables $\mathbf{Y}$ that are bounded between 0 and d and satisfy the constraint (21). In particular all d-regular triangle-free graphs induce a distribution $\mathbf{Y}$ satisfying these conditions, but there may be additional distributions that do not arise from graphs.

Consider the maximization problem first. By convexity we see that to maximize $\mathbb{E}\mathbf{Y}$ subject to these constraints, we must put all of the probability mass of $\mathbf{Y}$ on 0 and d. Because of the constraint (21), there is a unique such distribution.

Now fix a vertex v in $K_{d,d}$. If any vertex on v's side of the bipartition is in I, then v has 0 uncovered neighbors. If no vertex on the side is in I, then v has d uncovered neighbors. So the distribution of $\mathbf{Y}$ induced by $K_{d,d}$ (or $H_{d,n}$) is exactly the unique distribution satisfying the constraints that is supported on 0 and d. And therefore,

$$\overline{\alpha}_G(\lambda) \le \overline{\alpha}_{K_{d,d}}(\lambda)\,.$$

This proves Theorem 2.5 in the special case of triangle-free graphs.

What if we want to *minimize* $\mathbb{E}\mathbf{Y}$ subject to these constraints? In this case, by convexity, we should take $\mathbf{Y}$ to be constant: $\mathbf{Y} = y^*$ where $(1+\lambda)^{-y^*} = \frac{y^*}{d}$, or in other words, $y^* \cdot e^{y^*\log(1+\lambda)} = d$.

Formally, we can use Jensen's inequality:

$$\frac{1}{d}\mathbb{E}\mathbf{Y} = \mathbb{E}(1+\lambda)^{-\mathbf{Y}} \ge (1+\lambda)^{-\mathbb{E}\mathbf{Y}}$$

and so $\mathbb{E}\mathbf{Y} \ge y^*$ as above.

The solution is

$$y^* = \frac{W(d\log(1+\lambda))}{\log(1+\lambda)}$$

where $W(\cdot)$ is the W-Lambert function (that is, $x = W(x)e^{W(x)}$). This gives

$$\overline{\alpha}_G(\lambda) \ge \frac{1}{d}\frac{\lambda}{1+\lambda}\frac{W(d\log(1+\lambda))}{\log(1+\lambda)}\,. \tag{22}$$

Now although $\overline{\alpha}_G(\lambda)$ is monotone increasing in λ, somewhat surprisingly the bound (22) is not monotone in λ (see Figure 3 for example).

It turns out that it is best to take $\lambda = \lambda(d) \to 0$ as $d \to \infty$, but not as quickly as any polynomial, that is $\lambda(d) = \omega(d^{-\varepsilon})$ for every $\varepsilon > 0$.

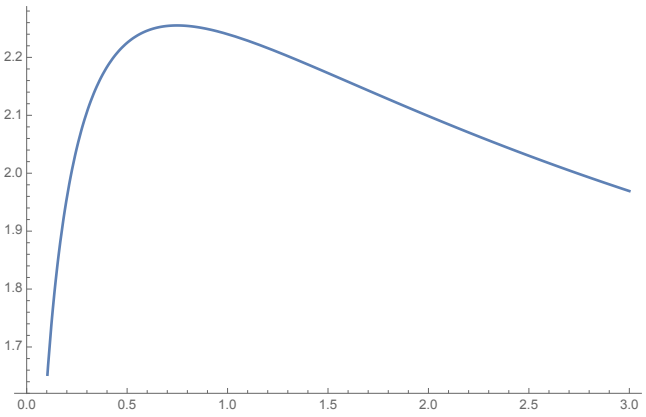

Figure 3: $\frac{\lambda}{1+\lambda} y^*$ as a function of λ with $d = 100$.

We set $\lambda = 1/\log d$ and derive a bound asymptotically in d. We show in the exercises that the Lambert-W function satisfies

$$W(x) = \log(x) - \log\log(x) + o(1)$$

as $x \to \infty$. If $\lambda \to 0$ then $\frac{\lambda}{(1+\lambda)\log(1+\lambda)} \to 1$, and $W(d\log(1+\lambda)) = (1 + o_d(1))\log d$. This gives, for $\lambda = 1/\log d$,

$$\overline{\alpha}_G(\lambda) \geq (1 + o_d(1))\frac{\log d}{d},$$

and by monotonicity this extends to all larger λ.

To obtain the counting result we integrate the bound (22) for $\lambda = 0$ to 1 to obtain a lower bound on the partition function.

$$\begin{aligned}
\frac{1}{n}\log i(G) = \frac{1}{n}\log Z_G(1) &= \int_0^1 \frac{\overline{\alpha}_G(t)}{t}\,dt \\
&\geq \int_0^1 \frac{1}{d}\frac{1}{1+t}\frac{W(d\log(1+t))}{\log(1+t)}\,dt \quad \text{from (22)} \\
&= \frac{1}{d}\int_0^{W(d\log 2)} 1 + u\,du \quad \text{using the substitution } u = W(d\log(1+t)) \\
&= \frac{1}{d}\left[W(d\log 2) + \frac{1}{2}W(d\log 2)^2\right] \\
&= \left(\frac{1}{2} + o_d(1)\right)\frac{\log^2 d}{d}.
\end{aligned}$$

Using a similar argument to the proof of the $R(3,k)$ upper bound, we can use Theorem 2.6 to give a lower bound on the number of independent sets in a triangle-free graph without degree restrictions.

Corollary 2.7 ([24])**.** *For any triangle-free graph G on n vertices,*

$$i(G) \geq e^{\left(\frac{\sqrt{2\log 2}}{4} + o(1)\right)\sqrt{n\log n}}.$$

Proof. Suppose the maximum degree of G is equal to d. Then $i(G) \geq 2^d$ since we can simply take all subsets of the neighborhood of the vertex with largest degree, and $i(G) \geq e^{\left(\frac{1}{2}+o_d(1)\right)\frac{\log^2 d}{d}n}$ from Theorem 2.6. As the first lower bound is increasing in d and the second is decreasing in d, we have

$$i(G) \geq \min_d \max \left\{ 2^d, e^{\left(\frac{1}{2}+o_d(1)\right)\frac{\log^2 d}{d}n} \right\} = 2^{d^*}$$

where d^* is the solution to $2^d = e^{\left(\frac{1}{2}+o_d(1)\right)\frac{\log^2 d}{d}n}$, that is,

$$d^* = (1+o_d(1))\frac{\sqrt{2}\sqrt{n}\log n}{4\sqrt{\log 2}}\,,$$

and so

$$i(G) \geq e^{\left(\frac{\sqrt{2\log 2}}{4}+o(1)\right)\sqrt{n}\log n}\,.$$

□

This improves the bound of $e^{\left(\frac{\sqrt{\log 2}}{4}+o(1)\right)\sqrt{n}\log n}$ from [19].

In the next section we give the full proof of Theorem 2.5, dispensing with the triangle-free assumption.

2.2.1 Linear programming and occupancy fractions

We begin by reviewing the basics of linear programming, duality, and complementary slackness.

Linear programming review

Suppose we have the linear program in standard form with variables $x_1, \ldots x_n$:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} c_i x_i \\ \text{subject to} \quad & x_i \geq 0 \ \forall i \\ & \sum_{i=1}^{n} A_{ij} x_i \leq b_j \quad \text{for } j = 1, \ldots m\,. \end{aligned}$$

This is the *primal LP*. The corresponding *dual LP* has variables $\Lambda_1, \ldots \Lambda_m$ for each constraint of the primal and constraints for each variable of the primal:

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{m} b_j \Lambda_j \\ \text{subject to} \quad & \Lambda_j \geq 0 \ \forall j \\ & \sum_{j=1}^{m} A_{ij} \Lambda_j \geq c_i \quad \text{for } i = 1, \ldots n\,. \end{aligned}$$

Theorem 2.8 (Weak duality theorem)**.** *If $x_1, \ldots, x_n$ and $\Lambda_1, \ldots, \Lambda_m$ are feasible solutions to the primal and dual LP's respectively, then*

$$\sum_{i=1}^{n} c_i x_i \leq \sum_{j=1}^{m} b_j \Lambda_j \, .$$

In particular, the objective value of *any* feasible dual solution gives an upper bound on the optimum of the primal LP. The strong duality theorem says that an optimal upper bound can be found if both primal and dual are feasible.

Theorem 2.9 (Strong duality theorem)**.** *If the primal and dual linear programs are both feasible, then their objective values coincide.*

In particular, if we have a feasible primal solution that we believe is optimal, we can prove this by finding a feasible dual solution with the same objective value.

The strong duality theorem implies that there are four possibilities for a primal/dual pair of LP's:

1. Both primal and dual are feasible and their optima coincide.
2. Both primal and dual are infeasible.
3. The primal is infeasible and the dual is unbounded.
4. The dual is infeasible and the primal is unbounded.

Very often it is useful to use *complementary slackness* to check for optimality. Say $x_1, \ldots, x_n$ and $\Lambda_1, \ldots, \Lambda_m$ are feasible solutions to the primal and dual LP's. We say complementary slackness holds for the pair of solutions if:

- For each $j = 1, \ldots, m$ either $\Lambda_j = 0$ or the jth constraint of the primal holds with equality under the solution $x_1, \ldots, x_n$ (or both are true).
- For each $i = 1, \ldots, n$ either $x_i = 0$ or the ith constraint of the dual holds with equality under the solution $\Lambda_1, \ldots, \Lambda_m$ (or both are true).

Theorem 2.10. *The feasible solutions $x_1, \ldots, x_n$ and $\Lambda_1, \ldots, \Lambda_m$ to the primal and dual LP's respectively are an optimal pair of solutions if and only if complementary slackness holds.*

For much more on linear programming, see for example Boyd and Vandenberghe's book [13].

Proof of Theorem 2.5

Let G be a d-regular n-vertex graph (with or without triangles). Do the following two part experiment: sample $\mathbf{I}$ from the hard-core model on G at fugacity λ, and independently choose $\mathbf{v}$ uniformly from $V(G)$. Previously we considered the random variable $\mathbf{Y}$ counting the number of uncovered neighbors of $\mathbf{v}$. When G was triangle-free we knew there were no

edges between these uncovered vertices, but now we must consider these potential edges. Let $\mathbf{H}$ be the graph induced by the uncovered neighbors of $\mathbf{v}$; $\mathbf{H}$ is a random graph on at most d vertices over the randomness in our two-part experiment. Specifically we mean the neighbors of $\mathbf{v}$ not covered by vertices in $V(G) \setminus N(\mathbf{v})$; so we don't consider covering among the neighbors.

We now can write $\overline{\alpha}_G(\lambda)$ in two ways as expectations involving $\mathbf{H}$.

$$\overline{\alpha}_G(\lambda) = \frac{\lambda}{1+\lambda} \Pr_{G,\lambda}[\mathbf{v} \text{ uncovered}] = \frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda}\left[\frac{1}{Z_{\mathbf{H}}(\lambda)}\right] \tag{23}$$

$$\overline{\alpha}_G(\lambda) = \frac{1}{d}\mathbb{E}_{G,\lambda}[\mathbf{I} \cap N(\mathbf{v})] = \frac{\lambda}{d}\mathbb{E}_{G,\lambda}\left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)}\right], \tag{24}$$

and so for any d-regular graph G, we have the identity

$$\frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda}\left[\frac{1}{Z_{\mathbf{H}}(\lambda)}\right] = \frac{\lambda}{d}\mathbb{E}_{G,\lambda}\left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)}\right]. \tag{25}$$

Now again we can relax our optimization problem from maximizing $\overline{\alpha}_G$ over all d-regular graphs, to maximizing $\frac{\lambda}{1+\lambda}\mathbb{E}\left[\frac{1}{Z_{\mathbf{H}}(\lambda)}\right]$ over all possible distributions $\mathbf{H}$ on $\mathcal{H}_d$, the set of graphs on at most d vertices, satisfying the constraint (25).

We claim that the unique maximizing distribution is the one distribution supported on the empty graph, $\emptyset$, and the graph of d isolated vertices, $\overline{K_d}$. This is the distribution induced by $K_{d,d}$ (or $H_{d,n}$) and is given by

$$\Pr_{K_{d,d}}(\mathbf{H} = \emptyset) = \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1}$$

$$\Pr_{K_{d,d}}(\mathbf{H} = \overline{K_d}) = \frac{(1+\lambda)^d}{2(1+\lambda)^d - 1}.$$

To show that this distribution is the maximizer we will use linear programming duality.

Both our objective function and our constraint are linear functions of the variables $\{p(H)\}_{H \in \mathcal{H}_d}$, so we can pose the relaxation as a linear program.

$$\begin{aligned}
\text{maximize} \quad & \sum_{H \in \mathcal{H}_d} p(H) \cdot \frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)} \\
\text{subject to} \quad & p(H) \geq 0 \ \forall H \in \mathcal{H}_d \\
& \sum_{H \in \mathcal{H}_d} p(H) = 1 \\
& \sum_{H \in \mathcal{H}_d} p(H) \left[\frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)} - \frac{\lambda}{d} \frac{Z'_H(\lambda)}{Z_H(\lambda)}\right] = 0.
\end{aligned}$$

The first two constraints ensure that the variables $p(H)$ form a probability distribution; the last is constraint (25).

Our candidate solution is $p(\emptyset) = \frac{(1+\lambda)^d-1}{2(1+\lambda)^d-1}$, $p(\overline{K_d}) = \frac{(1+\lambda)^d}{2(1+\lambda)^d-1}$, with objective value $\overline{\alpha}_{K_{d,d}}(\lambda) = \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d-1}$. To prove that this solution is optimal (and thus prove the theorem), we need to find some feasible solution to the dual with objective value $\overline{\alpha}_{K_{d,d}}(\lambda)$.

The dual linear program is:

$$\begin{aligned} &\text{minimize} && \Lambda_p \\ &\text{subject to} && \Lambda_p + \Lambda_c \cdot \left[\frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)} - \frac{\lambda}{d}\frac{Z'_H(\lambda)}{Z_H(\lambda)}\right] \geq \frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)} \quad \text{for all } H \in \mathcal{H}_d\,. \end{aligned}$$

For each variable of the primal, indexed by $H \in \mathcal{H}_d$, we have a dual constraint. For each constraint in the primal (not including the non-negativity constraint), we have a dual variable, in this case Λ_p corresponding to the probability constraint (summing to 1) and Λ_c corresponding to the remaining constraint. (Note that we do not have non-negativity constraints $\Lambda_p, \Lambda_c \geq 0$ in the dual because the corresponding primal constraints were equality constraints).

Now our task becomes: find a feasible dual solution with $\Lambda_p = \overline{\alpha}_{K_{d,d}}(\lambda)$. What should we choose for Λ_c? By complementary slackness in linear programming, the dual constraint corresponding to any primal variable that is strictly positive in an optimal solution must hold with equality in an optimal dual solution. In other words, we expect the constraints corresponding to $H = \emptyset, \overline{K_d}$ to hold with equality. This allows us to solve for a candidate value for Λ_c. Using $Z_\emptyset(\lambda) = 1$ and $Z'_\emptyset(\lambda) = 0$, we have the equation

$$\overline{\alpha}_{K_{d,d}}(\lambda) + \Lambda_c \left[\frac{\lambda}{1+\lambda} - 0\right] = \frac{\lambda}{1+\lambda}\,.$$

Solving for Λ_c gives

$$\Lambda_c = \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1}\,.$$

Now with this choice of Λ_c, and $\Lambda_p = \overline{\alpha}_{K_{d,d}}(\lambda) = \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d-1}$, our dual constraint for $H \in \mathcal{H}_d$ becomes:

$$\frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d - 1} + \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1}\left[\frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)} - \frac{\lambda}{d}\frac{Z'_H(\lambda)}{Z_H(\lambda)}\right] \geq \frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)}\,. \tag{26}$$

Multiplying through by $Z_H(\lambda) \cdot (2(1+\lambda)^d - 1)$ and simplifying, (26) reduces to

$$\frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1} \geq \frac{\lambda Z'_H(\lambda)}{Z_H(\lambda) - 1}\,, \tag{27}$$

and we must show this holds for all $H \in \mathcal{H}_d$ (except for $H = \emptyset$ for which we know already the dual constraint holds with equality). Luckily (27) has a nice probabilistic interpretation: the RHS is simply $\mathbb{E}_{H,\lambda}\left[|\mathbf{I}| \middle| |\mathbf{I}| \geq 1\right]$, the expected size of the random independent set given that it is not empty, and the LHS is the same for the graph of d isolated vertices, $\overline{K_d}$. Proving (27) is left for the exercises, and this completes the proof.

2.3 Further directions and open questions

2.3.1 Sphere packings

The proof of Theorem 2.6 works in a very different setting: lower bounding the average density of a random sphere packing of $\mathbb{R}^d$ chosen according to the *hard sphere model*.

2.3.2 Some open questions

Zhao has a nice survey on the area of extremal problems for regular graphs [81]. See also the paper of Csikvári [20].

Max vs average independent set size

Theorem 2.6 implies the upper bound on $R(3,k)$ in exactly the same way as Shearer's bound, as the occupancy fraction is of course a lower bound on the independence ratio. But we might hope that it gives more – that in triangle-free graphs there is a significant gap between the independence number and the size of a uniformly random independent set (i.e. at $\lambda = 1$ in the hard-core model).

Question 1. *Can we use Theorem 2.6 to improve the current asymptotic upper bound on $R(3,k)$.*

We give three specific conjectures whose resolution would improve the bound.

Conjecture 2.11 ([24])**.** *For any triangle-free graph G, we have*

$$\frac{\alpha(G)}{|V(G)| \cdot \overline{\alpha}_G(1)} \geq 4/3\,.$$

Conjecture 2.12 ([24])**.** *For any triangle-free graph G of minimum degree d, we have*

$$\frac{\alpha(G)}{|V(G)| \cdot \overline{\alpha}_G(1)} \geq 2 - o_d(1)\,.$$

Conjecture 2.13 ([24])**.** *For any $\varepsilon > 0$, there is λ small enough so that for any triangle-free graph G we have*

$$\frac{\alpha(G)}{|V(G)| \cdot \overline{\alpha}_G(\lambda)} \geq 2 - \varepsilon\,.$$

Conjecture 2.11 would imply a factor 4/3 improvement on the current upper bound for $R(3,k)$, while Conjectures 2.12 and 2.13 would both imply a factor 2 improvement.

Matchings and perfect matchings

A classic result that can be interpreted as an extremal problem for bounded degree graphs is Bregman's Theorem [14]. This theorem gives an upper bound on the permanent of a 0/1 matrix with prescribed row sums.

Recall that the permanent of an $n \times n$ matrix A, $\mathrm{per}(A)$, is

$$\mathrm{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \,,$$

where the sum is over permutations on n elements.

Theorem 2.14 (Bregman)**.** *Let A be an $n \times n$ matrix with $\{0,1\}$-valued entries and row sums $d_1, \ldots d_n$. Then*

$$\mathrm{perm}(A) \leq \prod_{i=1}^{n} (d_i!)^{1/d_i}.$$

Bregman's theorem can be stated as an upper bound on the number of perfect matchings in a balanced bipartite graph with a given degree sequence on one side. Let $\mathrm{pm}(G)$ denote the number of perfect matchings of a graph G.

Corollary 2.15. *Suppose G is a bipartite graph on two parts of $n/2$ vertices each, with left degrees $d_1, \ldots d_{n/2}$, then*

$$\mathrm{pm}(G) \leq \prod_{i=1}^{n/2} (d_i!)^{1/d_i}.$$

One can also ask about perfect matchings in not-necessarily bipartite graphs.

Theorem 2.16 (Kahn and Lovasz)**.** *Let G be a graph on $2n$ vertices with vertex degrees $d_1, \ldots, d_{2n}$. Then*

$$\mathrm{pm}(G) \leq \prod_{i=1}^{2n} (d_i!)^{1/(2d_i)}.$$

In the case of a d-regular graph on n vertices, this means that

$$\mathrm{pm}(G) \leq \mathrm{pm}(K_{d,d})^{n/2d} \,.$$

One can then ask which graph maximizes the number of total matchings, not just perfect matchings. Let $\mathcal{M}(G)$ be the set of all matchings of G. Let $m(G) = |\mathcal{M}(G)|$ be the number of matchings and $m_k(G)$ be the number of matchings of size k (so $m_{n/2}(G) = \mathrm{pm}(G)$). The matching polynomial, or the partition function of the monomer-dimer model, is

$$Z_G^{\mathrm{match}}(\lambda) = \sum_{M \in \mathcal{M}(G)} \lambda^{|M|} \,.$$

(Sometimes the matching polynomial is defined differently, as $\sum_{k \geq 0} (-1)^k \lambda^{n-2k} m_k(G)$ but these are equivalent up to scaling of the polynomial and the argument). The matching polynomial is the independence polynomial of the *line graph* of G: the graph on the edges of G with adjacency determined by incidence.

Theorem 2.17 (Davies, Jenssen, Perkins, Roberts [23])**.** *For any d-regular graph G, and any $\lambda > 0$,*

$$\frac{1}{|V(G)|} \log Z_G^{\mathrm{match}}(\lambda) \leq \frac{1}{2d} \log Z_{K_{d,d}}^{\mathrm{match}}(\lambda).$$

This is the analogue of Theorem 2.2 for matchings. In particular, taking $\lambda = 1$ we have $m(G)^{1/n} \leq m(K_{d,d})^{1/2d}$ for all d-regular graphs G.

Question 2. *For a given d, λ, what is the minimum and minimizer of*

$$\frac{1}{|V(G)|} \log Z_G^{\text{match}}(\lambda)$$

over all d-regular graphs?

See discussion in [22, 11].

Question 3. *Is there a degree sequence version of Theorem 2.17, that is, a generalization of Bregman's theorem to the matching polynomial?*

Question 4. *Is there an entropy-based proof of Theorem 2.17? (Or simply an easier proof than that in [23]).*

Computational thresholds and extremal problems

In some cases, the extremal problem of maximizing or minimizing the expectation of an observable (like the occupancy fraction) over a class of graphs can be used to identify a computation threshold in an algorithmic problem about approximating coefficients of partition functions, see [25, 17].

3 Cluster expansion

In this lecture we focus on one very important tool from statistical physics: the cluster expansion. At a very high level, the cluster expansion provides a way to show that a collection of random variables is 'close' to independent. When it applies, you can deduce almost any probabilistic information you can deduce for collections of independent random variables. We will see that it is very useful for combinatorial enumeration.

3.1 Ideal gas laws

An *equation of state* for a gas relates the quantities that determine the state of the gas: volume, pressure, and temperature.

The *ideal gas* is a gas of non-interacting particles. In the continuum we can model the ideal gas as a Poisson process of intensity λ on a region Λ of finite volume in $\mathbb{R}^d$. The partition function of the ideal gas is given by

$$Z_\Lambda(\lambda) = \sum_{k\geq 0} \int_{\Lambda^k} \frac{\lambda^k}{k!}\, dx_1 \dots dx_k = e^{\lambda|\Lambda|}\,.$$

The density of the ideal gas is simply λ, while the pressure is the normalized logarithm of the partition function: $p_\Lambda(\lambda) = \frac{1}{|\Lambda|}\log e^{\lambda|\Lambda|} = \lambda$. So for the ideal gas the equation of state is

$$p = \rho$$

where p is the infinite volume pressure and ρ is the infinite volume density. In physics this is stated with various constants (and can be derived from the equation $PV = nRT$ which you may be familiar with).

What happens when we deviate from the ideal gas and particles interact with each other? The equation of state must change to account for the interaction, but we can hope that when the density is small enough or the temperature large enough, the new equation of state will be a small perturbation of the ideal gas law.

The idea of the cluster expansion is to measure deviations from the ideal gas law due to interactions via an infinite series. To begin with the series is just a formal power series, but when the activity λ is small enough it can be shown that the series is convergent. The cluster expansion is also known as the 'Mayer series' or 'Mayer expansion' after the work of Mayer and Mayer and Mayer and Montroll [55, 56]. Rigorous proofs of convergence came later, e.g. [38, 65, 62].

In a discrete setting we can define an ideal gas as well, by taking, for instance, the hard-core model on a graph of n isolated vertices (we could have defined instead a gas with a Poisson number of particles at each site, which would lead to a different discrete ideal gas). Then for a given fugacity λ, the density is $\rho = \frac{\lambda}{1+\lambda}$ while the pressure is $p = \log(1+\lambda)$. Eliminating λ gives an equation of state:

$$p = \log\frac{1}{1-\rho}\,.$$

When we introduce interactions by adding edges to the empty graph, the equation of state will change. But again we can hope that when λ is small enough and interactions are weak enough (in terms of, say, the number of edges of the graph), the new equation of state will be a small perturbation of the ideal law.

3.2 Cluster expansion for low density hard-core model

It will be very useful for what we do later to introduce the cluster expansion in the setting of the multivariate hard-core model on a graph G; that is, each vertex v is assigned its own fugacity $\boldsymbol{\lambda}_v$.

We record some notation before we begin. Let K_k be the complete graph on k vertices and $\mathcal{G}_k$ be the set of all graphs on k vertices; that is, all subgraphs of K_k. Let $\mathcal{C}_k$ be the set of all connected graphs on k vertices. We also define the *Ursell function* (with respect to G): for $v_1, \dots v_k \in V(G)$,

$$\phi(v_1, \dots, v_k) = \frac{1}{k!} \sum_{H \in \mathcal{C}_k} \prod_{(i,j) \in E(H)} \left(-\mathbf{1}_{v_i = v_j \text{ OR } (v_i, v_j) \in E(G)}\right).$$

In particular if $v_1, \dots, v_k$ induce a disconnected subgraph graph of G then $\phi(v_1, \dots, v_k) = 0$.

We begin by writing

$$\begin{aligned} Z_G(\boldsymbol{\lambda}) &= \sum_{I \in \mathcal{I}(G)} \prod_{v \in I} \boldsymbol{\lambda}_v \\ &= \sum_{k \geq 0} \frac{1}{k!} \sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \prod_{1 \leq i < j \leq k} (1 - f(v_i, v_j)) \end{aligned}$$

where $f(v_i, v_j) = 1$ if $v_i = v_j$ or $(v_i, v_j) \in E(G)$. We can expand the last product

$$\prod_{1 \leq i < j \leq k} (1 - f(v_i, v_j)) = \sum_{H \subseteq K_k} \prod_{(i,j) \in E(H)} (-f(v_i, v_j))$$

where H is a subgraph of K_k with edges $E(H)$. We have

$$\begin{aligned} Z_G(\boldsymbol{\lambda}) &= 1 + \sum_{k \geq 1} \frac{1}{k!} \sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \sum_{H \in \mathcal{G}_k} \prod_{(i,j) \in E(H)} (-f(v_i, v_j)) \\ &= 1 + \sum_{k \geq 1} \frac{1}{k!} \sum_{H \in \mathcal{G}_k} \sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \prod_{(i,j) \in E(H)} (-f(v_i, v_j)). \end{aligned}$$

The expression $\sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \prod_{(i,j) \in E(H)} (-f(v_i, v_j))$ factorizes over the connected components $H_1, \dots H_\ell$ of H, and so we can write

$$Z_G(\boldsymbol{\lambda}) = 1 + \sum_{k \geq 1} \frac{1}{k!} \sum_{\ell=1}^{k} \sum_{\substack{H \in \mathcal{G}_k \\ H = (H_1, \dots, H_\ell)}} \prod_{r=1}^{\ell} \sum_{v_1, \dots, v_{m_r} \subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j) \in E(H_r)} (-f(v_i, v_j))$$

and summing over possible sizes of the connected components and noting that the summand corresponding to H_r only depends on its isomorphism class as a graph,

$$= 1 + \sum_{k\geq 1} \sum_{\ell=1}^{k} \frac{1}{\ell!} \sum_{\substack{m_1,\dots,m_\ell \\ \sum m_r = k}} \frac{k!}{m_1! \cdots m_\ell!} \sum_{H_1 \in \mathcal{C}_{m_1},\dots,H_\ell \in \mathcal{C}_{m_\ell}} \prod_{r=1}^{\ell} \sum_{v_1,\dots,v_{m_r} \subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \sum_{k\geq \ell} \frac{1}{\ell!} \sum_{\substack{m_1,\dots,m_\ell \\ \sum m_r = k}} \frac{k!}{m_1! \cdots m_\ell!} \sum_{H_1 \in \mathcal{C}_{m_1},\dots,H_\ell \in \mathcal{C}_{m_\ell}} \prod_{r=1}^{\ell} \sum_{v_1,\dots,v_{m_r} \subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \sum_{m_1,\dots,m_\ell} \sum_{H_1 \in \mathcal{C}_{m_1},\dots,H_\ell \in \mathcal{C}_{m_\ell}} \prod_{r=1}^{\ell} \frac{1}{m_r!} \sum_{v_1,\dots,v_{m_r} \subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \sum_{m_1,\dots,m_\ell} \prod_{r=1}^{\ell} \frac{1}{m_r!} \sum_{H\in\mathcal{C}_{m_r}} \sum_{v_1,\dots,v_{m_r} \subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \prod_{r=1}^{\ell} \sum_{m\geq 1} \frac{1}{m!} \sum_{H\in\mathcal{C}_m} \sum_{v_1,\dots,v_m \subseteq V(G)} \prod_{j=1}^{m} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \left(\sum_{m\geq 1} \frac{1}{m!} \sum_{H\in\mathcal{C}_m} \sum_{v_1,\dots,v_m \subseteq V(G)} \prod_{j=1}^{m} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j)) \right)^{\ell}$$

$$= \exp\left(\sum_{m\geq 1} \sum_{v_1,\dots v_m \in V(G)} \phi(v_1,\dots,v_m) \prod_{j=1}^{\ell} \boldsymbol{\lambda}_{v_j} \right)$$

or in other words,

$$\log Z_G(\boldsymbol{\lambda}) = \sum_{m\geq 1} \sum_{v_1,\dots v_m \in V(G)} \phi(v_1,\dots,v_m) \prod_{j=1}^{\ell} \boldsymbol{\lambda}_{v_j} \,.$$

There was a step (the second line in the last series of equalities) that was not justified: exchanging the order of summation over k and ℓ. Thus the result is a formal power series for $\log Z_G(\boldsymbol{\lambda})$ which may or may not be convergent. We consider the question of convergence in the next section.

We can rewrite this series in terms of *clusters*. Given the graph G, let $\Gamma = (v_1,\dots,v_\ell)$ be an ordered tuple of vertices of G (with possible repetitions). The incompatibility graph $H(\Gamma)$ is the graph with vertex set Γ and edges between $v_i, v_j \in \Gamma$ if $v_i = v_j$ or if $(v_i, v_j) \in E(G)$. A *cluster* is an ordered tuple Γ whose incompatibility graph $H(\Gamma)$ is connected. The size of the cluster is the length of the tuple.

We can interpret the Ursell function in terms of the incompatibility graph H:

$$\phi(H) = \frac{1}{|V(H)|!} \sum_{A\subseteq E(H), \text{spanning, connected}} (-1)^{|A|} \,.$$

Then the cluster expansion becomes

$$\log Z_G(\boldsymbol{\lambda}) = \sum_{\Gamma} \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \,,$$

where the sum is over all clusters Γ of G. It will be very important later, for combinatorial, algorithmic, and probabilistic reasons, that the cluster expansion is the sum over *connected* objects.

3.2.1 Examples

1. Suppose G is a single vertex v with fugacity λ. Then for each $\ell \geq 1$ there is a single cluster consisting of ℓ copies of v. The incompatibility graph is the complete graph on ℓ vertices with Ursell function $\frac{(-1)^{\ell+1}}{\ell}$ (exercise: prove this!). Then the cluster expansion is

$$\log Z_G(\lambda) = \sum_{\ell \geq 1} (-1)^{\ell+1} \frac{\lambda^\ell}{\ell}$$

   which is of course the Taylor series for $\log(1+\lambda)$ around 0.

2. Suppose G is a Δ-regular, triangle-free graph on n vertices. Consider the (univariate) hard-core model at fugacity λ. Then we can use symmetry to compute the first few terms of the cluster expansion.

   We list the clusters by size:

   - There are n clusters of size 1 (each a single vertex)
   - There are n clusters of size 2 consisting of two copies of a single vertex; Δn clusters of size 2 consisting of (ordered) edges.
   - There are n clusters of size 3 consisting of 3 copies of the same vertex; $3\Delta n$ clusters of size three with two copies of one vertex and one copy of a vertex joined to it by an edge; $3n\Delta(\Delta-1)$ clusters consisting of a 'v' of three vertices.

Cluster	Size	Count	Ursell function
1 vertex	1	n	1
2 copies of 1 vertex	2	n	$-1/2$
Ordered edge	2	Δn	$-1/2$
3 copies of 1 vertex	3	n	$1/3$
An edge with a repeated vertex	3	$3\Delta n$	$1/3$
A path of three vertices	3	$3n\Delta(\Delta-1)$	$1/6$

Table 1: A list of clusters up to size 3 in a regular, triangle-free graph

   Putting this together we get

$$\frac{1}{n} \log Z_G(\lambda) = \lambda - \frac{\lambda^2}{2} + \frac{\lambda^3}{3} - \frac{\Delta}{2}\lambda^2 + \frac{2\Delta + \Delta(\Delta-1)}{2}\lambda^3 + \dots$$

We can also measure the deviation from the ideal gas by removing the terms corresponding to clusters consisting of k copies of a single vertex:

$$\frac{1}{n}\log\frac{Z_G(\lambda)}{(1+\lambda)^n} = \frac{1}{n}\log Z_G(\lambda) - \log(1+\lambda) = -\frac{\Delta}{2}\lambda^2 + \frac{2\Delta+\Delta(\Delta-1)}{2}\lambda^3 + \dots$$

The expression $\frac{Z_G(\lambda)}{(1+\lambda)^n}$ has a nice interpretation: it is the probability that when picking a subset $S \subset V(G)$ by including each vertex independently with probability $\lambda/(1+\lambda)$ that the set S is an independent set. Janson's Inequality bounds probabilities such as this, and the first term of the expansion is exactly the term in the exponent in Janson's Inequality. The cluster expansion gives a way to get arbitrarily many terms in the expansion (but of course we need to ask about convergence). A generalization of Janson's Inequality due to Mousset, Noever, Panagiotou, and Samotij [58] deals with the probability of obtaining an independent set in a hypergraph when picking vertices independently at random.

3.3 Convergence criteria

The cluster expansion will only be a useful tool if we can show that it converges (and even more, bound the truncation error). There is a very large body of literature in mathematical physics devoted to finding convergence criteria for the cluster expansion. This criteria seek to balance three qualities: how sharp the bound is, how general its applicability, and how easy it is to check and apply. In this course we will give two such criteria: one that is very general and easy to check; the other that is sharp in a specific situation and has connections to combinatorics.

While we are usually interested in non-negative valued fugacities $\boldsymbol{\lambda}$, the convergence criteria will often be written in terms of complex-valued $\boldsymbol{\lambda}$. This is for several reasons. The cluster expansion is a power series and so will be convergence in a (poly-)disk in the complex plane. Showing convergence in complex domain allows one to deduce analyticity properties of the pressure and prove absence of phase transition in the Lee–Yang sense. Finally, we will see that evaluating the independence polynomial at negative fugacities is closely related to the Lovász Local Lemma. For much more on this last connection see the paper of Scott and Sokal [70].

The following result of Shearer gives the optimal zero-free (poly)disk for graphs of maximum degree Δ

Theorem 3.1 ([72])**.** *Let G be a graph of maximum degree Δ and suppose that for all $v \in V(G)$,*

$$|\boldsymbol{\lambda}_v| \le \lambda_s(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{\Delta^\Delta}\,. \tag{28}$$

The the cluster expansion for $\log Z_G(\boldsymbol{\lambda})$ converges absolutely, and in particular, $Z_G(\boldsymbol{\lambda}) \neq 0$.

Note that $\lambda_s(\Delta) = \frac{1+o(1)}{e\Delta}$ as $\Delta \to \infty$, so it has the same asymptotic performance as the Kotecký–Preiss bound. Shearer's bound, however, is tight: by taking finite truncations of the infinite Δ-regular tree, we can find zeroes of the independence polynomials of max degree Δ

graphs that approach $-\lambda_s(\Delta)$ from below on the negative real axis. In fact, Groeneveld [38] (and Scott–Sokal [70]) show that the coefficients of the cluster expansion for $\log Z_G(\lambda)$ alternate in sign and so the closest complex zero of Z_G must lie on the negative real axis, and this zero determines the radius of convergence of the cluster expansion.

We can derive bounds on the rate of convergence using Theorem 3.1 an idea of Barvinok [5]. Define the kth order truncation of the cluster expansion as

$$T_k(G, \boldsymbol{\lambda}) = \sum_{\Gamma : |\Gamma| \le k} \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \,.$$

Theorem 3.2 ([72, 5])**.** *Suppose G is a graph of maximum degree Δ on n vertices and for $\eta \in (0,1)$ suppose $|\boldsymbol{\lambda}_v| \le \eta \lambda_s(\Delta)$ for all $v \in V(G)$. Then*

$$|\log Z_G(\boldsymbol{\lambda}) - T_k(G, \boldsymbol{\lambda})| \le \frac{n\eta^k}{1-\eta} \,.$$

Proof. Let $\hat{Z}(t) = Z(t\boldsymbol{\lambda})$. $\hat{Z}$ is a polynomial in t of degree at most $N \le n$ and $Z(t) \neq 0$ for any complex t satisfying $|t| \le \eta^{-1}$. We are interested in evaluating $\log \hat{Z}(1)$. By the fundamental theorem of algebra we can write

$$\hat{Z}(t) = \prod_{j=1}^{N} \left(1 - \frac{t}{r_j}\right)$$

where $r_1, \ldots, r_N$ are the complex roots of $\hat{Z}(t)$. We write

$$\log \hat{Z}(t) = \sum_{j=1}^{N} \log \left(1 - \frac{t}{r_j}\right)$$

and then Taylor expand $\log\left(1 - \frac{t}{r_j}\right)$ at $t = 1$ around $t = 0$ using the fact that $|1/r_j| \le \eta$. This gives

$$\begin{aligned} |\log Z_G(\boldsymbol{\lambda}) - T_k(G, \boldsymbol{\lambda})| &\le \sum_{j=1}^{N} \sum_{i=k+1}^{\infty} \frac{\eta^i}{i} \\ &\le \frac{N\eta^k}{1-\eta} \\ &\le \frac{n\eta^k}{1-\eta} \,. \end{aligned}$$

□

The following is a specialization of the Kotecký–Preiss condition. We will see the general condition shortly.

Theorem 3.3 ([51]). *Consider the multivariate hard-core model on a graph G with complex-valued fugacity vector $\boldsymbol{\lambda}$. Suppose that for some $b \geq 0$ and each $v \in V(G)$,*

$$\sum_{u \in N(v) \cup \{v\}} |\boldsymbol{\lambda}_u| e^{1+b} \leq 1 \,. \tag{29}$$

Then the cluster expansion for $\log Z_G(\boldsymbol{\lambda})$ converges absolutely, and moreover, for each $v \in V(G)$ and each $t \geq 0$,

$$\sum_{\substack{\Gamma \nsim v \\ |\Gamma| \geq t}} \left| \phi(H(\Gamma)) \prod_{u \in \Gamma} \boldsymbol{\lambda}_u \right| \leq e^{-bt} \,. \tag{30}$$

In particular, $Z_G(\boldsymbol{\lambda}) \neq 0$.

Here we write $v \nsim \Gamma$ if there is some $u \in \Gamma$ so that $u = v$ or $u \in N(v)$; that is, if $\Gamma \cup \{v\}$ is a cluster.

We can apply this condition to a Δ-regular graph with uniform fugacity λ, taking $b = 0$. In this case, the condition is

$$|\lambda| \leq \frac{1}{e(\Delta + 1)} \,.$$

Taking $b > 0$ allows us to deduce strong bounds on the tail of the cluster expansion and correlation decay properties. We will see this in Section 3.4.

Connections to Lovász Local Lemma

Shearer proved Theorem 3.1 in the context of the Lovász Local Lemma. Consider a set of events $E_1, \ldots, E_n$ on a probability space, and a *dependency graph* G on the vertex set $[n]$ with the property that the event E_i is independent of the events $\{E_j : j \notin \{i \cup N(i)\}\}$. Let $p_j = P(E_j)$. The Lovász Local Lemma gives conditions in terms of G and $\{p_j\}$ to ensure that $\Pr\left(\bigcap_{j=1}^n \overline{E}_j\right) > 0$. Shearer showed a very nice connection between such a condition and the independence polynomial of the dependency graph G.

Theorem 3.4 ([72]). *Given a set of events $E_1, \ldots, E_n$ with probabilities $p_1, \ldots, p_n$ and dependency graph G, suppose that*

$$Z_G(-\boldsymbol{q}) \neq 0$$

for all $\boldsymbol{q} \leq \boldsymbol{p}$ coordinatewise. Then

$$\Pr\left(\bigcap_{j=1}^n \overline{E}_j\right) > 0 \,.$$

Combining the two theorems gives an optimal condition for the Lovász Local Lemma.

Corollary 3.5. *Given a set of events $E_1, \ldots, E_n$ with probabilities $p_1, \ldots, p_n$ and dependency graph G of maximum degree Δ. Then if $p_j \leq \frac{(\Delta-1)^{\Delta-1}}{\Delta^\Delta}$ for all j. Then $\Pr\left(\bigcap_{j=1}^n \overline{E}_j\right) > 0$.*

3.4 Cluster expansion and (joint) cumulants

An important application of a convergent cluster expansion is to proving strong correlation decay properties.

Recall from Lecture 1 that we defined the truncated k-point correlation functions in terms of the partial derivatives of a log partition function with external fields. We restate the definition here in the setting of the multivariate hard-core model.

With $\boldsymbol{\lambda}$ and $\mathbf{t}$ both vectors of non-negative numbers indexed by $V(G)$, let

$$Z_G(\boldsymbol{\lambda}, \mathbf{t}) = \sum_{I \in \mathcal{I}(G)} \prod_{v \in V(G)} \boldsymbol{\lambda}_v e^{t_v} .$$

For $u_1, \ldots, u_k \in V(G)$, the truncated k-point correlation function is

$$\kappa(u_1, \ldots, u_k) = \frac{\partial^k}{\partial t_{u_1} \cdots \partial t_{u_k}} \log Z_G(\boldsymbol{\lambda}, \mathbf{t})\big|_{\mathbf{t}=\vec{0}}.$$

In particular, the marginal of v is $\mu_v = \frac{\partial}{\partial t_v} \log Z_G(\boldsymbol{\lambda}, \mathbf{t})$ and the covariance of spins at u and v is $\kappa(u, v) = \mu_{uv} - \mu_u \mu_v = \frac{\partial^2}{\partial t_u \partial t_v} \log Z_G(\boldsymbol{\lambda}, \mathbf{t})$.

When the cluster expansion for $\log Z_G(\boldsymbol{\lambda})$ converges, we can differentiate it term-by-term and get a convergent expansion for the truncated correlation functions. Moreover, because the cluster expansion is a sum over connected objects, we can deduce correlation decay properties. (See [26] or [16] for more details).

Lemma 3.6. *Consider the multivariate hard-core model on G with fugacity vector $\boldsymbol{\lambda}$. Suppose that for some $b > 0$ the Kotecký–Preiss condition* (29) *holds. Then the model exhibits exponential decay of correlations, i.e. for all $u, v \in V(G)$,*

$$|\kappa(u, v)| \leq C(b) e^{-b \cdot \mathrm{dist}(u,v)} ,$$

where the constant C only depends on the constant b. More generally, for any $k > 0$ and $u_1, \ldots, u_k \in V(G)$, we have

$$|\kappa(u_1, \ldots, u_k)| \leq C(b, k) e^{-b \cdot \mathrm{MST}(u_1, \ldots, u_k)} ,$$

where $\mathrm{MST}(u_1, \ldots, u_k)$ is the minimum number of edges of G in a connected subgraph containing $u_1, \ldots, u_k$.

Recall that joint cumulants of independent random variables are 0, and so upper bounds on the magnitude of joint cumulants show a kind of approximate independence. It is not difficult to convert these bounds into bounds on the total variation distance between the joint distribution of spins at $u_1, \ldots, u_k$ and independent spins with marginals $\mu_{u_1}, \ldots, \mu_{u_k}$.

Proof. To start, we give a formula for the truncated correlation functions.

$$\kappa(u_1, \ldots, u_k) = \sum_{\Gamma \supseteq \{u_1, \ldots, u_k\}} \prod_{i=1}^{k} Y_{u_i}(\Gamma) \cdot \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \tag{31}$$

where $Y_u(\Gamma)$ is the number of occurrences of v in the cluster Γ. In particular,

$$\mu_v = \sum_{\Gamma \ni v} Y_v(\Gamma)\phi(H(\Gamma)) \prod_{u \in \Gamma} \boldsymbol{\lambda}_u \,.$$

Observe that

$$\frac{\partial^k}{\partial t_{u_1} \cdots \partial t_{u_k}} \prod_{v \in \Gamma} \boldsymbol{\lambda}_v e^{\mathbf{t}_v}\big|_{\mathbf{t}=\vec{0}} = \prod_{i=1}^{k} Y_{u_i}(\Gamma) \cdot \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \,,$$

and then apply this term-by-term to the cluster expansion to obtain (31). Differentiating the power series term-by-term is justified since we are inside the radius of convergence.

Now we can apply the pinned tail estimate (30) from Theorem 3.3. Each cluster that contains all vertices $u_1, \ldots, u_k$ has size at least $\mathrm{MST}(u_1, \ldots, u_k)$ since its incompatibility graph must be connected. Moreover, each additional copy of a vertex from $u_1, \ldots, u_k$ adds at least 1 to the size of a cluster. Therefor, by applying (30), for each $r_1, \ldots r_k \geq 0$, we have

$$\begin{aligned}\sum_{\Gamma : Y_{u_j}(\Gamma) = r_j + 1, 1 \leq j \leq k} \left| \prod_{i=1}^{k} Y_{u_i}(\Gamma) \cdot \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \right| &\leq e^{-b \cdot (\mathrm{MST}(u_1, \ldots, u_k) + \sum_j r_j - 1)} \prod_{j=1}^{k} (1 + r_j) \\ &\leq C(b,k) e^{-b \cdot \mathrm{MST}(u_1, \ldots, u_k)} \,.\end{aligned}$$

The result for $\kappa(u, v)$ follows immediately since $\mathrm{MST}(u, v) = \mathrm{dist}(u, v)$.

□

Note that in the setting of the univariate hard-core model on graphs of maximum degree Δ, Weitz's result proves exponential decay of correlations (in fact, strong spatial mixing) for a wider range of parameters than we obtain by cluster expansion convergence. The power of the cluster expansion method will become evident later when we apply it to models with very non-uniform fugacities.

3.5 Limit theorems and large deviations

A general principle is that whenever we have a convergent cluster expansion, we can, without too much work, deduce any probabilistic information we want to know, and in particular, any qualitative result that holds for independent random variables will likely hold in our setting. Convergence of the cluster expansion tells us that the perturbation from independent spins is small enough that we can control it.

We can use the cluster expansion to prove convergence of random variables. Recall that a random variable X has a $N(\mu, \sigma^2)$ distribution if and only if $\kappa_1(X) = \mu$, $\kappa_2(X) = \sigma^2$ and $\kappa_k(X) = 0$ for $k \geq 3$. Since the Normal distribution is determined by its moments, we can show that a sequence of random variables X_n converges to $N(\mu, \sigma^2)$ in distribution if $\kappa_1(X_n) \to \mu$, $\kappa_2(X_n) \to \sigma^2$, and $\kappa_k(X_n) \to 0$ for each fixed $k \geq 3$.

Consider the multivariate hard-core model, and assume the Kotecký–Preiss condition holds with some $b > 0$. Let $X = |\mathbf{I}|$ be the size of the random independent set drawn from the model (X implicitly depends on G and we can take the size of G to ∞). We can compute

the first two cumulants of X by taking $\mathbf{t} = t$ above and taking derivatives of $\log Z_G(\boldsymbol{\lambda}, t)$ in t:

$$\kappa_1(X) = \sum_{\Gamma} |\Gamma| \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v$$
$$\kappa_2(X) = \sum_{\Gamma} |\Gamma|^2 \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v$$

Now let us assume that $\kappa_2(X) = \operatorname{var}(X) = \Theta(n)$ (which we can prove in many cases, assuming, say, that $\boldsymbol{\lambda}$ is bounded above and below and G has an independent set of linear size). Let $\tilde{X} = (X - \mathbb{E}X)/\sqrt{\operatorname{var}(X)}$ be the centered and scaled version of X. Then to show that $\tilde{X} \Rightarrow N(0,1)$ as $n \to \infty$, it is enough to show that $\kappa_k(\tilde{X}) \to 0$ for each $k \geq 3$. Since centering by the mean does not change higher cumulants, and since $\operatorname{var}(X) = \Theta(n)$ it is enough to show that $\kappa_k(X/\sqrt{n}) \to 0$ for $k \geq 3$. We can compute

$$\begin{aligned}
\kappa_k(X/\sqrt{n}) &= n^{-k/2} \sum_{\Gamma} |\Gamma|^k \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \\
&\leq n^{-k/2} \sum_{t \geq 1} t^k \sum_{|\Gamma| = t} \left| \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \right| \\
&\leq n^{-k/2} n \sum_{t \geq 1} t^k e^{-bt} \\
&= O(n^{1-k/2}) \\
&= o(1)\,,
\end{aligned}$$

where we used (30) and summed over all vertices in the second inequality.

Next we can ask about large deviations for X; that is, prove an upper bound on probabilities like $\Pr(X \geq (1+\varepsilon)\mathbb{E}X)$ for $\varepsilon > 0$ fixed. If we assume $\mathbb{E}X = \Theta(n)$ and $\operatorname{var}(X) = \Theta(n)$ we would like to say that the large deviation probability is exponentially small. We will be able to prove this using the exponential Markov inequality if we can bound the moment generating function $\mathbb{E}e^{tX}$ for some $t > 0$. We can compute, with Z_t denoting the partition function with weights $\boldsymbol{\lambda} e^t$ and Z denoting the partition function with weights $\boldsymbol{\lambda}$,

$$\begin{aligned}
\log \mathbb{E}e^{tX} &= \log Z_t - \log Z \\
&= \sum_{\Gamma} (e^{t|\Gamma|} - 1) \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v
\end{aligned}$$

and so

$$\begin{aligned}
\log \mathbb{E}e^{t(X - \mathbb{E}X)} &= \sum_{\Gamma} (e^{t|\Gamma|} - 1 - t) \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \\
&\leq t^2 C n
\end{aligned}$$

for some absolute constant $C = C(b)$ when the Kotecký–Preiss condition holds. Now applying Markov's inequality we have

$$\Pr(X - \mathbb{E}X \geq \varepsilon n) \leq \frac{\mathbb{E}e^{t(X - \mathbb{E}X)}}{e^{t\varepsilon n}}$$

$$\leq \exp(t^2 Cn - t\varepsilon n)$$
$$\leq \exp\left(-\frac{\varepsilon^2}{4C} n\right),$$

which is our desired exponentially small upper bound.

3.6 Polymer models and cluster expansion

An important framework in which to apply the cluster expansion is that of abstract polymer models. This setting is not so far from that of the multivariate hard-core model on a graph, but with an added notion of 'size' that makes polymer models very useful in geometric settings. Gruber and Kunz defined 'subset polymer models' in which polymers were subgraphs of a lattice like $\mathbb{Z}^d$ [39]; Kotecky and Preiss defined a fully abstract notion of a polymer model and gave a very convenient convergence criteria for the cluster expansion [51].

An abstract polymer model consists of a finite or infinite set of polymers $\mathcal{C}$; a (complex-valued) weight function w_γ for each $\gamma \in \mathcal{C}$; and a symmetric, reflexive incompatibility relation '$\nsim$'. We write $\gamma' \nsim \gamma$ if γ and γ' are incompatible and write $\gamma \sim \gamma'$ if they are compatible.

Let Ω be the collection of all sets of mutually compatible polymers from $\mathcal{C}$ (including the empty set). Then the polymer model partition function is

$$Z = \sum_{X \in \Omega} \prod_{\gamma \in X} w_\gamma \,. \tag{32}$$

If the weights are non-negative then there is a natural probability measure on sets of mutually compatible polymers given by

$$\mu(X) = \frac{\prod_{\gamma \in X} w_\gamma}{Z} \,. \tag{33}$$

Example 4. *The canonical example of a polymer model is the multivariate hard-core model on a graph G with fugacities $\lambda_v, v \in V$. We let $\mathcal{C} = V$ and say two vertices u, v are incompatible if $d_G(u, v) \leq 1$. Sets of pairwise compatible polymers are exactly the independent sets of G, that is $X = \mathcal{I}(G)$. Assigning weights functions λ_v to each polymer v, we see that the polymer model partition function equals the multivariate hard-core partition function.*

Example 5. *Consider the ferromagnetic Ising model at inverse temperature β with external field λ on a graph G. That is,*

$$Z_G(\beta, \lambda) = \sum_{\sigma \in \{\pm 1\}^V} \lambda^{N(\sigma,+)} e^{\beta M(G,\sigma)} \,.$$

Now let polymers be connected induced subgraphs of G, with $w_\gamma = \lambda^{-|\gamma|} e^{-\beta |\partial_E \gamma|}$, where $|\gamma|$ is the number of vertices in the polymer γ and $\partial_E \gamma$ is the edge boundary of the subgraph induced by γ. Then we have

$$Z_G(\beta, \lambda) = \lambda^{|V|} e^{\beta |E|} Z \,. \tag{34}$$

If $\lambda > 1$ we expect to see more $+$ spins in a typical configuration and so the polymers represent deviations from the all $+$ ground state.

When the polymer weights w_γ are small we might expect few polymers in a typical configuration X, with weak correlations between the polymers. The cluster expansion gives us a way to measure this, by expanding $\log Z$ around the empty configuration.

As the example of the polymer model representation of the Ising model shows, a polymer will often come with a natural notion of 'size'; in this case, the size might be the number of vertices in the subgraph. The Kotecký–Preiss condition will provide a way to use the size of polymers in balancing the weight functions and the number of polymers a given polymer is incompatible with (in the hard-core setting, the activities and the degrees of vertices).

The following gives the Kotecký–Preiss condition for convergence of cluster expansion for abstract polymer models.

Theorem 3.7 ([51])**.** *Consider a polymer model defined by $(\mathcal{C}, w, \nsim)$. Suppose that there exist functions $a : \mathcal{C} \to [0, \infty)$, $b : \mathcal{C} : [0, \infty)$ so that for every $\gamma \in \mathcal{C}$,*

$$\sum_{\gamma' \nsim \gamma} |w_{\gamma'}| e^{a(\gamma')+b(\gamma')} \leq a(\gamma)\,. \tag{35}$$

Then for every $\gamma \in \mathcal{C}$ and all $t \geq 0$,

$$\sum_{\substack{\Gamma \nsim \gamma \\ b(\Gamma) \geq t}} \left| \phi(H(\Gamma)) \prod_{\gamma' \in \Gamma} w_{\gamma'} \right| \leq a(\gamma) e^{-bt}\,, \tag{36}$$

where $b(\Gamma) = \sum_{\gamma' \in \Gamma} b(\gamma')$.

Note that if $\mathcal{C}$ is a finite set (and so the partition function Z is finite), then (36) implies convergence of the cluster expansion for $\log Z$ by summing over all $\gamma \in \mathcal{C}$.

In many applications of Theorem 3.7 we will choose the functions a and b to incorporate the 'size' of a polymer (denote it by $|\gamma|$); often we will take $a(\gamma) = a|\gamma|$ and $b(\gamma) = b|\gamma|$ for some constants $a, b > 0$,

Example 6. *Consider the Ising model polymer model as above and suppose the host graph G has maximum degree Δ. For a polymer γ let $a(\gamma) = |\gamma|$, where $|\gamma|$ denotes the number of vertices of the induced subgraph that defines γ. For convenience take $b(\gamma) = 0$, but it is straightforward to extend the analysis below to the case $b(\gamma) = b|\gamma|$ for some $b \geq 0$.*

We need to show that for every γ,

$$\sum_{\gamma' \nsim \gamma} \lambda^{-|\gamma'|} e^{-\beta |\partial_E \gamma'|} e^{|\gamma'|} \leq |\gamma|$$

The number of induced subgraphs of size k containing a given vertex v in a graph of maximum degree Δ is at most $\frac{(e\Delta)^{k-1}}{2}$. In particular, this means that the number of γ' of size k that are incompatible with γ is at most $|\gamma|(\Delta+1)\frac{(e\Delta)^{k-1}}{2}$. Since $\beta \geq 0$ we can bound $w_{\gamma'} \leq \lambda^{-|\gamma'|}$. Then we can bound

$$\sum_{\gamma' \nsim \gamma} \lambda^{-|\gamma'|} e^{-\beta |\partial_E \gamma'|} e^{|\gamma'|} \leq \sum_{k \geq 1} |\gamma|(\Delta+1)\frac{(e\Delta)^{k-1}}{2} e^k \lambda^{-k}$$

$$\leq \frac{|\gamma|}{2} \sum_{k \geq 1} \left(\frac{e^2 \Delta}{\lambda} \right)^k$$
$$\leq |\gamma|$$

if $\lambda \geq 3e^2\Delta/2$, and so the Kotecký–Preiss holds when λ is large enough as a function of Δ. Of course the bound is not tight – there are many points at which we made rough estimates, but we do not expect the condition to be tight itself and so we are usually ok with bounds like this.

3.7 Example: ferromagnetic Potts model on expander graphs

Consider the q-color ferromagnetic Potts model on a graph G and suppose we want to model defects from the all 'red' ground state. Define polymers to be connected induced subgraphs of G with vertices of the subgraph colored by the remaining $q-1$ non-red colors (each different coloring of the same subgraph yields a different polymer). Two polymers γ and γ' are incompatible if their union is connected. The weight of a polymer is

$$w_\gamma = e^{-\beta|\partial_e \gamma| - \beta|E_b(\gamma)|}$$

where $\partial_e \gamma$ is the set of edges from γ to γ^c and $E_b(\gamma)$ are the bichromatic edges of γ. Then we have

$$Z_G(q, \beta) = e^{\beta|E(G)|} \cdot Z$$

where $Z_G(q,\beta)$ is the Potts model partition function and Z is the polymer model partition function. Notice that $e^{\beta|E(G)|}$ is the weight of the ground state (all red configuration), and so Z captures contributions to $Z_G(q,\beta)$ from deviations from the ground state (the empty polymer configuration corresponds to the ground state – no defects). Of course we haven't really gained anything from this representation – the polymer model includes all the configurations that are dominated by blue or by green, etc, while we wanted to capture deviations from the red ground state. We will see below that we can address (in some settings) this by restricting polymers to be 'small'.

We will work in the setting of *expander graphs* (see the excellent survey [42]). The topic of expander graphs is too big to go into here, but expander graphs are extremely useful in mathematics, computer science, and information theory in:

- Derandomizing randomized algorithms
- Building error correcting codes
- Constructing pseudorandom graphs

Random walks mix extremely fast on expander graphs. There are many different constructions of expander graphs including those based on randomness (random regular graphs for example), group theory, number theory, and iterative constructions (the zig-zag product).

There are several different but related definitions of expander graphs: there are spectral expanders, vertex expanders, edge expanders. We will specialize here to Δ-regular graphs to make the connections clearer.

For a vertex subset S, let $\partial_e S$ denote the set of edges with one endpoint in S and one endpoint in S^c.

Definition 3.8. *An n-vertex Δ-regular graph G is an α-edge-expander if for all $S \subset V(G), |S| \leq n/2$, $|\partial_e S| \geq \alpha |S|$.*

Definition 3.9. *An n-vertex Δ-regular graph G is an α-vertex-expander if for all $S \subset V(G), |S| \leq n/2$, $|N(S)| \geq \alpha |S|$.*

Let $A(G)$ be the adjacency matrix of an n-vertex graph G and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of A sorted in decreasing order. Recall that for a Δ-regular graph, $\lambda_1 = \Delta$.

Definition 3.10. *An n-vertex Δ-regular graph G is an ε-spectral-expander if $\lambda_2(G) \leq (1 - \varepsilon)\Delta$*

Later we will use the notion of a bipartite α-expander.

Definition 3.11. *A Δ-regular bipartite graph G with bipartition (L, R) is a bipartite α-expander if for all $S \subset L, |S| \leq |L|/2$, $|N(S)| \geq (1 + \alpha)|S|$ (and likewise for $S \subset R$).*

Returning to the low-temperature Potts model, we will make two assumptions on G:

1. G has maximum degree Δ
2. G is an α-edge-expander for some $\alpha > 0$

For instance, the random Δ-regular graph satisfies these conditions whp.

For large β, we expect configurations to be dominated by one of the q-colors – we expect to see sparse, disordered deviations from one of the q monochromatic ground states. We will control these deviations via polymer models and the cluster expansion.

How do we ensure that our polymer model only captures small deviations from the red ground state and not, for example, small deviations from the green ground state? And what can we do about configurations for which there is no clearly dominant color? We will do this in two steps. In the first we argue that we can divide the state space of the Potts model into q disjoint parts (each associated to a dominant color) so that the complement of these parts has exponentially small probability. We will see later that this is in fact a proof of slow mixing for the Glauber dynamics. In the second step we will show that a polymer model with a convergent cluster expansion can approximate each one of these q parts.

In the polymer model, we will insist that $|\gamma| \leq n/2$ for all polymers γ. Here $|\gamma|$ denotes the number of vertices of γ. This ensures, for example, that the monochromatic green configuration is not represented in the red polymer model. Let Z denote the partition function of this polymer model (note that by symmetry Z does not depend on the ground state color).

We start by understanding the separation of the q ground states. By using the expansion properties of G we can show that $q \cdot e^{\beta |E(G)|} \cdot Z$ approximates $Z_G(q, \beta)$ to within e^{-n} relative error.

Lemma 3.12 ([43])**.** *Let G be a n-vertex, α expander of maximum degree at most Δ. For $\beta \geq$,*

$$(1 - e^{-n}) q \cdot e^{\beta |E(G)|} Z \leq Z_G(q, \beta) \leq (1 + e^{-n}) q \cdot e^{\beta |E(G)|} Z \,.$$

Proof. Proving this takes two steps. The first is to show that when β is large Potts configurations in which no color has a majority have exponentially small relative weight. For $r \in [q]$, let $Z_r(\beta) = \sum_{\sigma:|\sigma^{-1}(r)|>n/2} e^{\beta M(G,\sigma)}$. Then for $\beta \geq 2\log(eq)/\alpha$, we will show that

$$\sum_{r\in[q]} Z_r(\beta) = qZ_r(\beta) \leq Z_G(q,\beta) \leq (1+e^{-n})qZ_r(\beta)\,. \tag{37}$$

The lower bound is immediate since configurations can have at most one majority color. The upper bound is a simple consequence of expansion: when there is no majority there must be many bichromatic edges, and these are penalized heavily for large β. In particular, there must be at least $\frac{n\alpha}{2}$ bichromatic edges, giving a penalty to each configuration of $e^{-n\alpha\beta/2}$ relative to one of the ground state configurations. There are at most q^n configurations with no majority, and taking $\beta \geq$, we have $q^n e^{-n\alpha\beta/2} \leq e^{-n}$. The approximation given by (37) allows us to partition configurations into $q+1$ subsets; one for each color plus an addition error class (no majority) that we can neglect.

The next step is to show that the color r polymer model partition function Z (after scaling by $e^{\beta|E(G)|}$) approximates $Z_r(\beta)$ up to an exponentially small relative error.

$$(1-e^{-n})e^{\beta|E(G)|}Z \leq Z_r(\beta) \leq e^{\beta|E(G)|}Z\,.$$

Every configuration with a majority of vertices colored r is captured by the polymer model since fewer than $n/2$ vertices receive a non-r color; this gives the upper bound. To prove the lower bound we show that configurations in which all non-r connected components are of size at most $n/2$ but which do not have a majority r have small total weight; this also follows from an expansion argument: such configurations must have at least $\alpha n/2$ bichromatic edges and so their total weight is at most $q^n e^{-\beta n/2} \leq e^{-n}$ when $\beta \geq 2\log(eq)/\alpha$, relative to the empty polymer configuration. □

The restriction on polymer sizes will also allow us to show that for β large enough as a functions of q, Δ, α, the Kotecký–Preiss condition is satisfied.

As above we need to bound the number of polymers of a given size and their weight. The number of polymers of size k incompatible with γ is at most $\frac{|\gamma|}{2}(\Delta+1)(q-1)^k(e\Delta)^{k-1}$.

The expansion condition gives us an upper bound on the weight of a polymer:

$$w_\gamma \leq e^{-\alpha\beta|\gamma|}$$

since $|\partial_e\gamma| \geq \alpha|\gamma|$; here we have crucially used the upper bound on the size of γ.

With these two bounds we can verify the Kotecký–Preiss condition. Let $a(\gamma) = b(\gamma) = |\gamma|$. For a given polymer γ,

$$\begin{aligned}\sum_{\gamma'\nsim\gamma} w_{\gamma'} e^{a(\gamma')+b(\gamma')} &\leq \sum_{k\geq 1} (e\Delta(q-1))^k \Delta|\gamma| e^{-\alpha\beta k} e^{2k} \\ &\leq |\gamma|\Delta \sum_{k\geq 1} \exp\left[k\left(3+\log\Delta+\log(q-1)-\alpha\beta\right)\right]\end{aligned}$$

which is at most $a(\gamma) = |\gamma|$ if $\beta \geq \frac{4+2\log(q\Delta)}{\alpha}$.

4 Independent sets in the hypercube

In this lecture we apply abstract polymer models and the cluster expansion to a combinatorial enumeration problem: counting weighted independent sets in the hypercube.

4.1 Independent sets in the hypercube

Let Q_d be the Hamming cube $\{0,1\}^d$ with edges between vectors that differ in exactly one coordinate. The Hamming cube has two maximum independent sets, each of size 2^{d-1}: $\mathcal{O}$, the set of vectors whose coordinates sum to an odd number, and $\mathcal{E}$, the set of vectors whose coordinates sum to an even number. Since any subset of an independent set is an independent set, we have a trivial lower bound on the total number of independent sets of the Hamming cube: $i(Q_d) \geq 2 \cdot 2^{2^{d-1}} - 1$.

In a classic result, Korshunov and Sapozhenko determined the asymptotics of $i(Q_d)$ [50].

Theorem 4.1 (Korshunov and Sapozhenko)**.** *As $d \to \infty$,*

$$i(Q_d) = (2 + o(1))\sqrt{e}2^{2^{d-1}}. \tag{38}$$

Sapozhenko later gave another proof of this result [69] that introduced an influential variant of the method of graph containers. See also Galvin's exposition of this result [32].

Galvin [30] later extended Theorem 4.1 to the setting of weighted independent sets, the hard-core model on Q_d. He found the asymptotics of the partition function $Z_{Q_d}(\lambda)$ for $\lambda > \sqrt{2} - 1$.

Theorem 4.2 (Galvin)**.** *If $\lambda > \sqrt{2}-1$,*

$$Z_{Q_d}(\lambda) = (2 + o(1)) \cdot \exp\left[\frac{\lambda}{2}\left(\frac{2}{1+\lambda}\right)^d\right](1+\lambda)^{2^{d-1}}. \tag{39}$$

He also found the asymptotics of the logarithm of $Z_{Q_d}(\lambda)$ for $\lambda = \Omega(d^{-1/3}\log d)$.

Using the cluster expansion we can obtain asymptotics of $Z_{Q_d}(\lambda)$ for all fixed λ [44]. For instance, if $\lambda > 2^{1/3} - 1$,

$$Z_{Q_d}(\lambda) = (2 + o(1)) \cdot \exp\left[\frac{\lambda}{2}\left(\frac{2}{1+\lambda}\right)^d\left(1 + \frac{(2\lambda^2+\lambda^3)d(d-1)-2}{4(1+\lambda)^d}\right)\right](1+\lambda)^{2^{d-1}}. \tag{40}$$

Compared to (39), the formula (40) has an extra term in the exponent. As we will see this extra term comes from taking more terms in the cluster expansion of a polymer model and reflects a structural change in typical independent sets from the hard-core model on Q_d when $\lambda < \sqrt{2}-1$. More generally, for each $k \geq 1$, there is a structural change in typical independent sets when λ passes $2^{1/k} - 1$ and this is reflected in the asymptotic formula for $Z_{Q_d}(\lambda)$ having $k-1$ terms in the exponent for $\lambda > 2^{1/k} - 1$.

Theorem 4.3 ([44]). *There is a sequence of polynomials $R_j(d,\lambda)$, $j \in \mathbb{N}$, such that for any fixed $t \geq 1$ and $\lambda > 2^{1/t} - 1$,*

$$Z_{Q_d}(\lambda) = (2+o(1))(1+\lambda)^{2^{d-1}} \exp\left[2^{d-1}\sum_{j=1}^{t-1} R_j(d,\lambda)(1+\lambda)^{-dj}\right]$$

as $d \to \infty$. Moreover the coefficients of the polynomial R_j can be computed in time $e^{O(j \log j)}$.

In particular, $R_1 = \lambda$, recovering the formula in Theorem 4.2 and $R_2 = \frac{(2\lambda^3+\lambda^4)d(d-1)-2\lambda^2}{4}$, giving (40).

4.2 Polymer model

We sketch a proof of (40) here. We emphasize that an essential part of the proof is Galvin's weighted generalization of Sapozhenko's graph container lemma [30, Lemma 3.10] which we will take as a black box. The statistical physics approach via the cluster expansion does not replace the container argument; rather uses the container result and builds on it to obtain detailed results. What is remarkably fortuitous is how well this container lemma works with the cluster expansion: along with expansion properties of Q_d it provides exactly what is needed to verify the Kotecký–Preiss condition.

As in the case of unbalanced bipartite graphs, we will consider the generalized ground state consisting of the set of all independent sets not containing an odd (respectively even) occupied vertex. There are two such ground states (even and odd dominated) and each has weight $(1+\lambda)^{2^{d-1}}$. They overlap on the empty independent set of weight 1 (which is neglgible asymptotically). This notion of a ground state is evident already in the formulas (38), (39), and (40) in the factors $2 \cdot 2^{2^{d-1}}$ and $2 \cdot (1+\lambda)^{2^{d-1}}$ respectively. The remaining factors in the formulas capture the contribution from typical deviations from the two ground states.

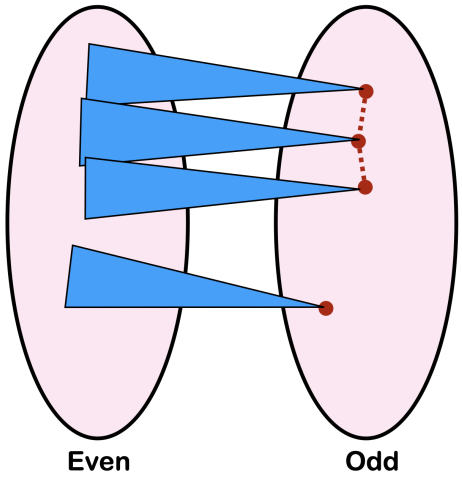

Figure 4: 2-linked components of occupied odd vertices, one of size 3 (top) and one of size 1 (bottom). Odd vertices are 2-linked if their neighborhoods overlap.

We focus now on even-dominated independent sets, and view odd occupied vertices as defects. To capture the behavior of these defects with a polymer model we fix $X \subseteq \mathcal{O}$ and consider all independents sets I so that $I \cap \mathcal{O} = X$. The vertices in X contribute a factor $\lambda^{|X|}$ to the weight of such I. Any even vertex in the neighborhood of X is blocked from being in an independent set, and any even vertex not in the neighborhood of X is free to be in or out of an independent set. This means

$$\begin{aligned}\sum_{I: I \cap \mathcal{O} = X} \lambda^{II|} &= \lambda^{|X|}(1+\lambda)^{|\mathcal{E}|-|N(X)|} \\ &= (1+\lambda)^{2^{d-1}} \frac{\lambda^{|X|}}{(1+\lambda)^{|N(X)|}}\,.\end{aligned}$$

We set $w(X) = \frac{\lambda^{|X|}}{(1+\lambda)^{|N(X)|}}$; this weight measures the penalty relative to the ground state of the set of defect vertices X. The weight $w(X)$ does not factorize over the vertices of X since two vertices may or may not have overlapping neighborhoods. However, the weight does factorize over 2-linked components of X; that is, subsets of X which are connected under the adjacency relation defined by having overlapping neighborhoods. (Or in other words, a 2-linked component of $\mathcal{O}$ is a set $S \subseteq \mathcal{O}$ which is connected in the graph Q_d^2). We can write

$$w(X) = \prod_{\gamma \in X} \frac{\lambda^{|\gamma|}}{(1+\lambda)^{|N(\gamma)|}}$$

where the product is over the 2-linked components of X. This is exactly what we need to define a polymer model: a compatibility relation and a weight that factorizes over pairwise compatible components.

For the even-dominated polymer model, the set of polymers $C_{\mathcal{E}}$ is the set of all 2-linked components γ from $\mathcal{O}$ with $|[\gamma]| \leq 2^{d-2}$ where $[\gamma] = \{y \in \mathcal{O} : N(y) \subseteq N(\gamma)\}$ (this notion of the *closure* of γ appears in [35, 30]). Two polymers γ, γ' are incompatible if $\gamma \cup \gamma'$ is 2-linked. The weight function is $w_\gamma = \frac{\lambda^{|\gamma|}}{(1+\lambda)^{|N(\gamma)|}}$. The analogous odd-dominated polymer model features 2-linked components of even vertices. Let Ξ be the polymer model partition function (its value is the same for the even and odd models).

4.3 Cluster expansion convergence

Proving convergence of the cluster expansion is more complicated than in the case of expander graphs, and relies crucially on Sapozhenko's graph containers. In particular, we will use the following result of Galvin [30], generalizing a lemma from [69] to weighted independent sets. Let

$$\mathcal{G}(a,b) = \{\gamma \subseteq \mathcal{O} : \gamma \text{ 2-linked}, |[\gamma]| = a, |N(\gamma)| = b\}.$$

Lemma 4.4 ([30])**.** *There exist constants $C_0, C_1 > 0$, so that for all $\lambda \geq C_0 \log d / d^{1/3}$, all $a \leq 2^{d-2}$,*

$$\sum_{\gamma \in \mathcal{G}(a,b)} \frac{\lambda^{|\gamma|}}{(1+\lambda)^b} \leq 2^d \exp\left(-\frac{C_1(b-a)\log d}{d^{2/3}}\right).$$

Proof. □

We also use three different expansion properties of Q_d (collected in [30]):

Lemma 4.5. *The following bipartite expansion estimates hold for Q_d:*

1. *For $S \subset \mathcal{O}$, $|S| \leq d/10$, $|N(S)| \geq d|S| - 2|S|^2$.*
2. *For $S \subset \mathcal{O}$, $|S| \leq d^4$, $|N(S)| \geq d|S|/10$.*
3. *For $S \subset \mathcal{O}$, $|S| \leq 2^{d-2}$, $|N(S)| \geq \left(1 + \frac{1}{2\sqrt{d}}\right)|S|$.*

For very small sets (polylogarithmic in the size of the graph) the hypercube is a very good expander, and the expansion-based arguments we used for the Potts model will work here too. For larger sets, however, the expansion guarantees are much too weak. However the exponential estimate on the number of polymers containing a given vertex based from the previous lecture is also far too pessimistic for large 2-linked subsets of Q_d. The balance between these two quantities is captured by Lemma 4.4.

4.3.1 Approximation by polymer models

Using the above ingredients, we will show that for λ sufficiently large ($\Omega(\log d/d^{1/3})$) the cluster expansion for the even (or odd) dominated polymer model converges and that, after scaling, the polymer model partition function is a very good approximation to $Z_{Q_d}(\lambda)$.

The following [44, Lemma 15] proves convergence of the cluster expansion for the defect polymer model.

Lemma 4.6. *Suppose $\lambda \geq C_0 \log d/d^{1/3}$. Then with $a(\gamma) = \frac{|\gamma|}{d^{3/2}}$ and*

$$b(\gamma) = \begin{cases} \log(1+\lambda)(d|\gamma| - 3|\gamma|^2) - 7|\gamma|\log d & \text{if} \quad |\gamma| \leq \frac{d}{10} \\ \frac{|\gamma| d \log(1+\lambda)}{20} & \text{if} \quad \frac{d}{10} < |\gamma| \leq d^4 \\ \frac{|\gamma|}{d^{3/2}} & \text{if} \quad |\gamma| > d^4 \end{cases}$$

the Kotecký–Preiss condition is satisfied.

Proof. We break up the sum $\sum_{\gamma' \nsim \gamma} w_{\gamma'} e^{a(\gamma')+b(\gamma')}$ into three parts, based on the size of γ' and then uses expansion properties of Q_d for the first two sums and the graph container lemma, Lemma 4.4, for the third sum.

To show that

$$\sum_{\gamma' \nsim \gamma} w_{\gamma'} e^{a(\gamma')+b(\gamma')} \leq a(\gamma) = \frac{|\gamma|}{d^{3/2}}$$

it is enough to show that for all $v \in \mathcal{E}$,

$$\sum_{\gamma \ni v} w_\gamma e^{a(\gamma)+b(\gamma)} \leq d^{-7/2}\,.$$

We break up the sum into three parts according to the size of γ; we will show that each sum is at most $\frac{1}{3}d^{-7/2}$.

First consider $|\gamma| \le d/10$. By counting rooted, connected induced subgraphs in a bounded degree graph, the number of polymers of size k containing v is at most $\exp(3k \log d)$; by Lemma 4.5, for $|\gamma| \le d/10$, $|N(\gamma)| \ge d|S| - 2|S|^2$. This gives

$$\begin{aligned}\sum_{\substack{\gamma \ni v \\ |\gamma| \le d/10}} w_\gamma e^{a(\gamma)+b(\gamma)} &\le \sum_{k=1}^{d/10} e^{3k\log d} e^{kd^{-3/2}} e^{\log(1+\lambda)(dk-3k^2)-7k\log d} \frac{\lambda^k}{(1+\lambda)^{dk-2k^2}} \\ &\le \sum_{k\ge 1} \exp\left(3k\log d + k\log\lambda + kd^{-3/2} - k^2\log(1+\lambda) - 7k\log d\right) \\ &\le \sum_{k\ge 1} \exp\left(-4k\log d + kd^{-3/2}\right)\end{aligned}$$

which is at most $\frac{1}{3}d^{-7/2}$ for d large enough.

Next consider $d/10 < |\gamma| \le d^4$. In this case $|N(\gamma)| \ge d|\gamma|/10$, and so

$$\begin{aligned}\sum_{\substack{\gamma \ni v \\ d/10 < |\gamma| \le d^4}} w_\gamma e^{a(\gamma)+b(\gamma)} &\le \sum_{k=d/10}^{d^4} e^{3k\log d} e^{kd^{-3/2}} e^{d\log(1+\lambda)k/20} \frac{\lambda^k}{(1+\lambda)^{dk/10}} \\ &= \sum_{k=d/10}^{d^4} \exp\left(k\left(d^{-3/2} + \log\lambda + 3\log d - d\log(1+\lambda)/20\right)\right)\end{aligned}$$

which is at most $\frac{1}{3}d^{-7/2}$ for d large enough and $\lambda \ge C\log d/d$.

Finally consider $|\gamma| > d^4$. Here we have $|N(\gamma)| \ge |\gamma|(1 + 1/(2\sqrt{d}))$. We bound

$$\begin{aligned}\sum_{\substack{\gamma \ni v \\ |\gamma| > d^4}} w_\gamma e^{a(\gamma)+b(\gamma)} &= \sum_{\substack{d^4 < a \le 2^{d-2} \\ (1+1/(2\sqrt{d}))a \le b \le 2^{d-1}}} \sum_{\substack{\gamma \ni v \\ |[\gamma]|=a, |N(\gamma)|=b}} \frac{\lambda^{|\gamma|}}{(1+\lambda)^b} e^{2|\gamma|d^{-3/2}} \\ &\le \sum_{\substack{d^4 < a \le 2^{d-2} \\ (1+1/(2\sqrt{d}))a \le b \le 2^{d-1}}} e^{2ad^{-3/2}} \sum_{\substack{\gamma \ni v \\ |[\gamma]|=a, |N(\gamma)|=b}} \frac{\lambda^{|\gamma|}}{(1+\lambda)^b} \\ &\le \sum_{\substack{a > d^4 \\ b \ge (1+1/(2\sqrt{d}))a}} e^{2ad^{-3/2}} d \exp\left(-\frac{C_1(b-a)\log d}{d^{2/3}}\right),\end{aligned}$$

where the last inequality comes from applying Lemma 4.4. In the sum, we have $(b-a) \ge a/(2\sqrt{d})$ and $a > d^4$, and so

$$\frac{2a}{d^{3/2}} + \log d - \frac{C_1(b-a)\log d}{d^{2/3}} \le -ad^{-7/6}$$

for large enough d, and so

$$\begin{aligned}\sum_{\substack{\gamma \ni v \\ d^4 < |\gamma| \le 2^{d-2}}} w_\gamma e^{a(\gamma)+b(\gamma)} &\le \sum_{\substack{d^4 < a \le 2^{d-2} \\ (1+1/(2\sqrt{d}))a \le b \le 2^{d-1}}} \exp(-ad^{-7/6}) \\ &\le 2^d \sum_{a > d^4} \exp(-ad^{-7/6}) \\ &\le \frac{1}{3d^{7/2}}\end{aligned}$$

for d large enough.

□

The following lemma [44, Lemma 14] shows that $2(1+\lambda)^{2^{d-1}}\Xi$ is a very good approximation to the full partition function.

Lemma 4.7. *For $\lambda = \Omega(\log d \cdot d^{-1/3})$,*

$$\left|\log Z_{Q_d}(\lambda) - \log\left(2(1+\lambda)^{2^{d-1}}\Xi\right)\right| = O\left(e^{-2^d/d^4}\right) .$$

The proof of Lemma 4.7 combines ideas from the slow-mixing result of Galvin and Tetali [35] with large deviation estimates from the polymer model itself.

We state the large deviation result here. The proof of this lemma uses the techniques outlined in the previous lecture on the cluster expansion.

Lemma 4.8. *Let $\mathbf{X}$ be a random configuration drawn from the defect polymer distribution. Let $\|\mathbf{X}\| = \sum_{\gamma \in \mathbf{X}} |\gamma|$. Then*

$$\Pr[\|\mathbf{X}\| > 2^d/d^2] \le \exp\left(-2^d/d^4\right) . \tag{41}$$

Now we turn to the proof of Lemma 4.7.

Proof of Lemma 4.7. We say an independent set I is captured by the odd-dominated polymer model if every 2-linked component γ of $\mathcal{E} \cap I$ is small; that is, $|[\gamma]| \le 2^{d-2}$ (and likewise for the even-dominated polymer model). We can show that $2(1+\lambda)^{2^{d-1}}\Xi$ is a good approximation to Z_{Q_d} by showing that the weight of independent sets a) not captured by either polymer model and b) captured by both polymer models is small.

To address a), note that every independent set is captured by either the odd or the even polymer model. Suppose for the sake of contradiction that there exists $I \in \mathcal{I}(Q_d)$ that contains a 2-linked set $\gamma \subseteq \mathcal{O}$ with $|[\gamma]| > 2^{d-2}$ and a 2-linked set $\gamma' \subseteq \mathcal{E}$ with $|[\gamma']| > 2^{d-2}$. We have $|N(\gamma)| = |N([\gamma])| > 2^{d-2}$ (since Q_d contains a perfect matching). Then $N(\gamma) \cap [\gamma'] \neq \emptyset$ and so $\gamma \cap N(\gamma') = \gamma \cap N([\gamma']) \neq \emptyset$, contradicting the fact that I is an independent set.

It remains to bound the contribution from independent sets that are captured twice. To do that it is enough to show that with probability $1 - O\left(\exp(-2^d/d^4)\right)$ the number of odd occupied vertices in an independent set formed from the odd-dominated polymer model is

greater than the number of even occupied vertices. This follows from Lemma 4.8: with probability $1 - O\left(\exp(-2^d/d^4)\right)$ the number of even occupied vertices is at most $2^d/d^2$; conditioned on this event, the number of odd occupied vertices stochastically dominates a $\mathrm{Bin}(2^{d-1} - 2^d/d, \lambda/(1+\lambda))$ random variable. This random variable is greater than $2^d/d^2$ with probability $1 - O\left(\exp(-2^d/d^4)\right)$ by a Chernoff bound.

All together this gives

$$\left(1 - O\left(\exp(-2^d/d^4)\right)\right) 2(1+\lambda)^{2^{d-1}}\Xi \leq Z(\lambda) \leq 2(1+\lambda)^{2^{d-1}}\Xi\,,$$

and so

$$\log[2(1+\lambda)^{2^{d-1}}\Xi] - O(\exp(-2^d/d^4)) \leq \log Z(\lambda) \leq \log[2(1+\lambda)^{2^{d-1}}\Xi]\,,$$

which proves Lemma 4.7.

□

4.4 Consequences

As a consequence of Lemmas 4.6 and 4.7, we obtain a convergent series approximation for $\log Z_{Q_d}$ with bounds on the truncation error. The larger λ is the fewer terms of the series needed to get the asymptotics of Ξ (and thus Z_{Q_d}). In fact, the number of terms of the cluster expansion needed corresponds exactly to the size of the largest defect polymers typically seen in a random independent set.

In particular, if $\lambda > 2^{1/t} - 1$, then we have the asymptotic formula

$$Z_{Q_d}(\lambda) = (2 + o(1)) \exp\left[\sum_{k=1}^{t-1} L_k\right] (1+\lambda)^{2^{d-1}}\,,$$

where L_k is the sum of cluster expansion terms over clusters of size k. Moreover, when $\lambda > 2^{1/t} - 1$ then whp the largest 2-linked occupied component on the side of the bipartition with fewer occupied vertices is of size at most $t - 1$.

In the proof of Lemma 4.7 we implicitly used the following. Consider the distribution ν_λ on $\mathcal{I}(Q_d)$ defined as follows:

1. With probability 1/2 each choose even or odd as the dominant side of the bipartition.
2. Choose a polymer configuration $\mathbf{X}$ from the odd (even) dominated polymer model. Place each of the vertices of this configuration in the independent set.
3. From the vertices of $\mathcal{O}$ ($\mathcal{E}$) unblocked by $\mathbf{X}$ add each to the independent set independently with probability $\lambda/(1+\lambda)$.

Lemma 4.7 immediately implies that ν_λ is very close in total variation distance to μ_λ.

Corollary 4.9. *For $\lambda \geq C_0 \log d/d^{1/3}$,*

$$\|\mu_\lambda - \nu_\lambda\|_{TV} = O\left(\exp(-2^d/d^4)\right)\,.$$

(It is not hard to see that the two distributions are not exactly equal - the empty independent set gets twice the weight in ν_λ as in μ_λ).

4.5 Further directions for Q_d and beyond

One area in which combinatorial ideas can be applied to statistical physics questions is in identifying or bounding the values of parameters at which phase transitions occur. In particular, the location of the phase transition for the hard-core model on $\mathbb{Z}^d$ or Q_d is an open problem (and a subtle problem: Brightwell, Häggström, and Winkler [15] show that on some infinite graphs there may be more than one phase transition in the hard-core model).

See [33, 61, 35] for the best current lower bounds on phase coexistence.

See [4] for additional applications of the cluster expansion to combinatorial enumeration.

5 The algorithmic perspective

In this lecture we look at the algorithmic problems of approximating partition functions and sampling from Gibbs measures.

5.1 Basics of approximate counting and sampling

An ε-relative approximation to a non-negative real number Z is a number $\hat{Z}$ so that

$$e^{-\varepsilon}\hat{Z} \leq Z \leq e^{\varepsilon}\hat{Z}\,.$$

We could have also used $(1-\varepsilon)$ and $(1+\varepsilon)$ but e^{ε} is slightly more convenient.

There are two standard notions of efficient approximate counting algorithms: one deterministic and one randomized.

A *fully polynomial-time approximation scheme* (FPTAS) for computing a partition function Z_G is an algorithm that given $\varepsilon > 0$ and G outputs an ε-relative approximation to Z_G and runs in time polynomial in $n = |V(G)|$ and $1/\varepsilon$.

A *fully polynomial-time randomized approximation scheme* (FPRAS) for computing a partition function Z_G is a randomized algorithm that given $\varepsilon > 0$ and G outputs, with probability at least $2/3$, an ε-relative approximation to Z_G and runs in time polynomial in $n = |V(G)|$ and $1/\varepsilon$.

The choice of $2/3$ is arbitrary – any probability in $(1/2, 1)$ would suffice. By repeating the algorithm $O(\log(1/\delta))$ times and taking a median one can boost the probability of success to $1-\delta$.

The notion of an efficient approximate sampling algorithm is not quite standard. We give one definition here.

A *polynomial-time sampling scheme* for a Gibbs measure μ_G on spin configurations $\Sigma \subseteq \Omega^V$ is a randomized algorithm that given G and $\varepsilon > 0$ outputs a configuration $\sigma \in \Sigma$ with distribution $\hat{\mu}$ so that $\|\hat{\mu} - \mu_G\|_{TV}$. The algorithm should run in time polynomial in $n = |V(G)|$ and $1/\varepsilon$.

The other commonly used notion requires the sampling algorithm to run in time polynomial in n and $\log(1/\varepsilon)$.

In many situations randomized efficient approximate counting (FPRAS) and efficient approximate sampling are equivalent. The precise setting is that of *self-reducible problems* [48, 73]; in the exercises we will show implications in the case of the hard-core model on a family of graphs closed under taking subgraphs (e.g. all graphs, all graphs of maximum degree Δ, all triangle-free graphs etc.).

5.2 Major results and open questions in approximate counting and sampling

We next give a brief survey of some of the major results and major open problems in the field of approximate counting. Very broadly, statistical physics spin models (with exceptions!) fit into one of three categories:

- Problems for which there is an efficient algorithm for *all graphs* (ferromagnetic Ising and monomer-dimer models). Here the main open questions are whether there are deterministic algorithms for all graphs. Intriguingly, the two models in this category are exactly the two models for which there are very strong zero-freeness results (the Lee-Yang and Heilmann-Lieb theorems). Is there a rigorous connection between these phenomena?

- Problems which are hard NP-hard on general graphs but for which there are efficient algorithms in the weak interaction regime on, say, bounded-degree graphs (hard-core model is one example). Here a main question is to find *computational thresholds* separating parameter regimes of tractability and intractability on some classes of graphs. For the hard-core model this has been done by Weitz and Sly, while for other models the location of the threshold is still open.

- Problems for which neither efficient algorithms nor hardness of approximation is known either. The class of problems defined by #BIS (below) lies in this category, and the major open problem is to determine its complexity.

5.2.1 Positive results

The following are some of the landmark results in the field.

Theorem 5.1 (Jerrum-Sinclair [46])**.** *There is an FPRAS for the partition function of the ferromagnetic Ising model for all inverse temperatures $\beta \geq 0$ and consistent, non-uniform external fields.*

Consistent external fields mean that all external fields point in the same direction (either favoring + or favoring −).

Theorem 5.2 (Jerrum-Sinclair-Vigoda [47])**.** *There is an FPRAS for approximating the permanent of an $n \times n$ matrix with non-negative entries. Thus there is an FPRAS for approximating the number of perfect matchings in a bipartite graph.*

Theorem 5.3 (Jerrum-Sinclair [45])**.** *For every $\lambda > 0$ there is an FPRAS for approximating the monomer-dimer partition function on all graphs.*

While we can sample weighted matchings in a general graph or perfect matchings in a bipartite graph, it is not known how to sample perfect matchings from a general graph.

Question 5. *Is there an FPRAS for approximating the number of perfect matchings in a general graph?*

An important theme in the study of computational complexity is the power of randomness. In the setting of approximate counting and sampling there are several important cases in which randomized approximation algorithms are known but not deterministic approximation algorithms

Question 6. *Is there an FPTAS (deterministic approximation algorithm) for the number of perfect matchings of a bipartite graph or the permanent of a non-negative matrix?*

5.2.2 Computational thresholds

For some problems we know an FPRAS for some range of parameters, know NP-hardness for other ranges of parameters, but do not know the precise computational threshold. The canonical example is the hard-core model on the class of graphs of maximum degree Δ.

Theorem 5.4 (Weitz, Sly)**.**

1. *If $\lambda < \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, then there is an FPTAS and a polynomial-time sampling scheme for the hard-core model on graphs of maximum degree Δ [79].*
2. *If $\lambda > \lambda_c(\Delta)$, then there is no FPRAS for the hard-core model on graphs of maximum degree Δ unless NP=RP [74, 75, 29].*

For other problems there is a range of tractable parameters and a range of intractable parameters but we do not know the precise computational threshold. One example is the problem of counting and sampling proper q-colorings of a graph.

Theorem 5.5 (Vigoda [78])**.** *For $q > \frac{11}{6}\Delta$ there is an FPRAS for counting the number of q-colorings of a graph of maximum degree Δ. In particular the Glauber dynamics mix rapidly for these parameters.*

See also the small recent improvement in [18].

Question 7. *Is there an FPTAS for counting the number of q-colorings in graphs of maximum degree Δ when $q \geq \Delta + 1$?*

5.2.3 Intermediate complexity: $\#BIS$

Sampling independent sets (uniformly or from the hard-core model) is hard in general: in fact, finding large independent sets in graph is already NP-hard. What if we restrict to a class of graphs for which the search problem is tractable? In particular, there is a polynomial-time algorithm to find a maximum independent set in a bipartite graph. Can we in fact sample (or approximately count) independent sets in bipartite graphs?

The problem of counting the number of independent sets in a bipartite graph is known as #BIS (counting bipartite independent sets). Exact counting is #P-hard but one can ask about approximate counting.

Question 8. *Is there an FPRAS for approximating the number of independent sets in a bipartite graph? Or is the problem NP-hard?*

Dyer, Goldberg, Greenhill, and Jerrum [27] showed that many other important approximate counting and sampling problems with unknown complexity are equivalent to or as hard as #BIS. These include the ferromagnetic Potts model ($q \geq 3$), counting q-colorings in bipartite graphs, counting stable matchings.

5.3 MCMC approach

Perhaps the most widely used approach to approximate counting and sampling is Markov Chain Monte Carlo (MCMC). The basic idea is to find a Markov chain with the following properties:

- The unique stationary distribution of the chain is the target distribution μ.
- A single step of the chain is easy to implement.
- The chain converges rapidly to the stationary distribution.

Then the sampling algorithm is to start at an arbitrary state, run the chain for long enough, and output the final configuration. Using random sampling we can then approximate a partition function. The MCMC method was first developed in [57], to sample from the 2-dimensional hard disk model (for some history on its developments see [64]).

For background on Markov chains see [53]. We recall a few important definitions here.

Definition 5.6. *A discrete-time, discrete space Markov chain on a finite or countable set Σ is a stochastic process $X_0, X_1, X_2, \ldots$ that satisfies the Markov property: the distribution of X_t conditioned on $X_0, \ldots, X_{t-1}$ equals the distribution of X_t conditioned on X_{t-1}.*

This means we can describe a Markov chain by the distribution ν_0 of the initial state X_0 and the transition matrix $P(\cdot, \cdot)$, defined by $P(x, y) = P(X_{n+1} = y | X_n = x)$. We can also define the t-step transition probabilities: $P^t(x, y) = P(X_{n+t} = y | X_n = x)$

Definition 5.7. *A Markov chain is irreducible if for every $x, y \in \Sigma$ there exists t so that $P^t(x, y) > 0$. A Markov chain is aperiodic if for every $x \in \Sigma$ the gcd of the set $\{t \geq 1 : P^t(x, x) > 0\}$ is 1.*

For example, the simple random walk on a connected graph is irreducible, but on a disconnected graph it is not. The simple random walk on a bipartite graph is not aperiodic. We can make any periodic Markov chain aperiodic by making it 'lazy': at each step, with probability $1/2$ (or some other positive constant) stay in the current state and with probability $1/2$ follow the law of the Markov chain.

Definition 5.8. *A probability distribution π on Σ is a stationary distribution for the Markov chain defined by P if for all $x \in \Sigma$,*

$$\pi(X) = \sum_{y \in \Sigma} \pi(y) P(y, x) \,.$$

In particular, if π is a stationary distribution and X_0 is distributed according to $\pi_0 = \pi$ then X_k is distributed according to π for all $k \geq 0$.

How can we find a stationary distribution? There is a useful technique for a special class of Markov chains. A Markov chain P is *reversible* with respect to π if for all $x, y \in \Sigma$,

$$\pi(x) P(x, y) = \pi(y) P(y, x) \,. \tag{42}$$

This is called the *detailed balance* equation.

If P is reversible with respect to π then π is a stationary distribution for P. While most (or perhaps all) the Markov chains we will study in this class are reversible,

A fundamental result on Markov chains states that if a Markov chain is irreducible and aperiodic then there exists a unique probability distribution μ on Σ so that $\pi_n \to \pi$ as $n \to \infty$, where π_n is the distribution of X_n.

5.3.1 Markov chains for spin models

We now specialize to Markov chains for spin models on graphs (for a short introduction to the topic see [63]).

Here the state space will be the set of all possible configurations, Ω^V, or in the case of models with hard constraints, the set of all allowed configurations ($\mathcal{I}(G)$ in the case of the hard-core model on a graph G).

Our goal will be to find a Markov chain on Ω^V with stationary distribution μ_G, the Gibbs measure. There are many such Markov chains but for now we will be most interested in *local* Markov chains: Markov chains in which at most one (or at most a bounded number of) spins are changed in each step. This is for two reasons: such Markov chains may in some sense capture the way a physical system evolves over time and these Markov chains are often easy to implement computationally.

The Glauber dynamics

Given a configuration $X_t \in \Omega^V$ we obtain X_{t+1} as follows:

1. Pick $v \in V$ uniformly at random.
2. Resample the spin at v from the Gibbs measure μ_G conditioned on the spins at the other vertices.

If μ_G is a Gibbs measure with pairwise interactions across edges, the second step is equivalent to resampling conditioned on the spins of the neighbors of v.

Exercise 5. *1. Show that the Glauber dynamics are reversible with respect to the Gibbs measure μ_G.*

2. Show that the Glauber dynamics are irreducible and aperiodic. (Assume that $H(\sigma) \in [0, \infty)$ for all σ, or prove for the hard-core model).

Exercise 6. *Let $\Sigma_q(G)$ be the set of all proper q-colorings of G and suppose $|\Sigma_q| > 0$. Let $\mu_{G,q}$ denote the uniform distribution on $\Sigma_q(G)$.*

1. *Give an example of graph G so that for some $q \geq 2$, $\Sigma_q(G)$ is disconnected under moves that change the color of one vertex at a time.*
2. *Give conditions in terms of q and the maximum degree of Δ to ensure that $\Sigma_q(G)$ is connected under single spin updates (changing the color of one vertex at a time).*

3. *Describe precisely a single spin update Markov chain for sampling from $\mu_{G,q}$ (i.e. Glauber dynamics).*

5.3.2 Mixing times

Markov chains are widely used to sample from probability distributions on large sets Σ. If we want to sample from μ_G by running a Markov chain $X_1, X_2, \ldots$ with stationary distribution μ_G, an important question is how long we need to run the chain to get an approximately correct sample from μ_G.

To answer this we need a good measure of 'approximately correct'.

Definition 5.9. *Let μ_1 and μ_2 be two probability measures on the same sample space Σ with the same sigma-field $\mathcal{F}$. Then the total variation distance between μ_1 and μ_2 is*

$$\|\mu_1 - \mu_2\|_{TV} = \sup_{A \in \mathcal{F}} |\mu_1(A) - \mu_2(A)| .$$

For a discrete probability space, we have

$$\begin{aligned} \|\mu_1 - \mu_2\|_{TV} &= \frac{1}{2} \sum_{\sigma \in \Sigma} |\mu_1(\sigma) - \mu_2(\sigma)| \\ &= \sum_{A \subseteq \Sigma : \mu_1(A) > \mu_2(A)} \mu_1(A) - \mu_2(A) . \end{aligned}$$

With this definition we can define the Mixing time of a Markov chain.

Definition 5.10. *The ε-mixing time $\tau_{\text{mix}}(\varepsilon)$ of a Markov chain with stationary distribution π on Σ is*

$$\sup_{\pi_0 \in \mathcal{P}(\Sigma)} \min\{t \geq 0 : \|\pi_t - \pi\|_{TV} < \varepsilon\},$$

where π_t is the distribution of X_t given that X_0 has distribution π_0.

We write τ_{mix} for $\tau_{\text{mix}}(1/4)$. This is because for any irreducible, aperiodic Markov chain with stationary distribution π, there are constants $C > 0$ and $\alpha \in (0, 1)$ so that

$$\sup_{\pi_0} \|\pi_t - \pi\|_{TV} \leq C\alpha^t .$$

This means that $\tau_{\text{mix}}(\varepsilon) = O(\log(\varepsilon^{-1})) \cdot \tau_{\text{mix}}$.

For a Gibbs measure μ_G on a graph G on n vertices we say a Markov chain is fast-mixing or mixes rapidly if τ_{mix} is bounded by a polynomial in n; that is, poly-logarithmic in the size of the state space Ω^V. We say the Markov chain is slow mixing or torpidly mixing if τ_{mix} is superpolynomial in n (often exponential in some power of n).

Keep in mind that the size of the state space of a spin model on G is exponentially large in n. So in the context of random walks on graphs, a random walk that mixes in time $O(\log n)$ is fast mixing while a random walk with mixing time $\Omega(n^c)$ for some $c > 0$ is slow mixing. Random walks on expander graphs are fast mixing, while a random walk on a cycle, for instance, is slow mixing.

5.4 The method of correlation decay

In this section we describe an approach to approximate counting due to Weitz [79]. It is known as 'the method of correlation decay' and it involves constructing a computational tree to compute the marginal of a single vertex. If the computational tree exhibits strong spatial mixing then approximating the marginal can be done efficiently. Then by self-reducibility efficient computation of marginals leads to an efficient approximate counting algorithm for the partition function.

We present the method in its original context (the hard-core model on bounded degree graphs) but it has since been extended to other graphs and other spin systems.

Weitz's main result is both an algorithmic result and a probabilistic result about correlation decay and uniqueness.

Theorem 5.11 (Weitz [79])**.** *There is an FPTAS and polynomial-time sampling scheme for the hard-core partition function $Z_G(\lambda)$ on graphs of maximum degree Δ when*

$$\lambda < \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}\,. \tag{43}$$

Moreover, the hard-core model on any infinite graph of maximum degree Δ exhibits strong spatial mixing and uniqueness when $\lambda < \lambda_c(\Delta) = (1+o_\Delta(1))\frac{e}{\Delta}$.

The value $\lambda_c(\Delta)$ marks a phase transition on an infinite graph: it is the uniqueness threshold for the hard-core model on the infinite Δ-regular tree. Theorem 5.11 and the complementary Theorem **??** below connect the statistical physics phase transition on the tree to a *computational threshold*.

Recall that we can write $Z_G(\lambda)$ as the inverse of a product of marginals of subgraphs of G. This implies that if we have an FPTAS for μ_v for a given λ and class of graphs closed under taking subgraphs (such as max degree Δ graphs) then we have an FPTAS for $Z_G(\lambda)$. Then by the reduction of approximate sampling to approximate counting we have a polynomial-time sampling scheme as well.

5.4.1 The self-avoiding walk tree

Given a graph G and a vertex $v \in V(G)$, Weitz (following Godsil's construction for matchings [37]) constructs a tree $\mathbf{T}_{SAW}(G,v)$ with root r so that for a given fugacity λ,

$$\mu_{G,v} = \mu_{\mathbf{T}_{SAW}(G,v),r}\,.$$

That is, the marginal of v in G equals the marginal of r in $\mathbf{T}_{SAW}(G,v)$. This is somewhat promising since marginals in trees can be computed exactly in polynomial time, but in general $\mathbf{T}_{SAW}(G,v)$ will be of exponential size in the size of G.

To construct $\mathbf{T}_{SAW}(G,v)$ first order the vertices of G with $v_1,\dots v_n$ (and assume $v = v_1$). This puts an order on edges incident to a given vertex w, with the 'smaller' edge being the edge with the vertex appearing first in the order.

We will first construct $\mathbf{T}_{SAW}(G, v)$ as a tree with a boundary condition imposed on leafs; in the hard-core model, however, we can implement boundary conditions by pruning the tree (if a vertex is specified out of the independent set, remove it from the graph; if a vertex is specified in the independent set, remove it and its neighbors).

The tree is the tree of self-avoiding walks of G starting at v, with vertices added for steps that close a cycle. These vertices will be specified in or out of the independent set based on the ordering of edges. In particular, if a path closes a cycle, the vertex in the tree closing the cycle is set to be in if the edge closing the cycle is larger than the edge starting the cycle and set to be out otherwise.

Theorem 5.12. *For any G, any $v \in V(G)$, and any $\lambda \geq 0$,*

$$\mu_{G,v} = \mu_{\mathbf{T}_{SAW}(G,v),r}\,.$$

Moreover if G has maximum degree Δ then $\mathbf{T}_{SAW}(G, v)$ has maximum degree Δ. The depth of $\mathbf{T}_{SAW}(G, v)$ is at most $|V(G)|$.

Proof. It will be convenient to work with ratios of marginal instead of marginals directly. Let

$$R_{G,v} = \frac{\mu_{G,v}}{1 - \mu_{G,v}}\,.$$

On a tree T with root r with degree d, we have a nice recursion for $R_{T,r}$:

$$R_{T,r} = \lambda \prod_{i=1}^{d} \frac{1}{1 + R_{T_i,u_i}}$$

where T_i is the subtree rooted at the ith neighbor u_i of r. Applying this identity recursively gives a procedure to compute the marginal of the root of any finite tree (with the base case that $R_{v,v} = \lambda$).

Now consider a general graph G with specified vertex v whose neighbors are $u_1, \ldots, u_d$. Let G' be the graph G where v is replaced by d copies $v_1, \ldots, v_d$, with each v_i joined only to u_i. We also set the activity of v_i to be $\lambda^{1/d}$. Then we have that $R_{G,v}$ is exactly the ratio of the probability that all of the v_i's are occupied to the probability that none of the v_i's are occupied in G'.

Let τ_i be the boundary condition that u_j is occupied for $j < i$ and unoccupied for $j > i$. Then by the above we have that

$$R_{G,v} = \prod_{i=1}^{d} R^{\tau_i}_{G',v_i}$$

where we have included the boundary conditions in the notation.

Since v_i is only connected to u_i, we have

$$R^{\tau_i}_{G',v_i} = \frac{\lambda^{1/d}}{1 + R^{\tau_i}_{G' \setminus v_i,u_i}}$$

and so

$$R_{G,v} = \lambda \prod_{i=1}^{d} \frac{1}{1+R}\,.$$

As in the tree, this defines a recursive procedure for computing $R_{G,v}$; the recursion terminates since the number of unfixed vertices (by boundary conditions) decreases by 1 at each step.

In fact, the recursion for $R_{G,v}$ is exactly the same recursion as for $R_{\mathbf{T}_{SAW}(G,v),r}$. □

Corollary 5.13. *If the infinite Δ-regular tree $\mathbb{T}_\Delta$ exhibits strong spatial mixing then any infinite graph (or infinite family of finite graphs) of maximum degree Δ exhibits strong spatial mixing.*

5.4.2 Uniqueness threshold of the hard-core model on the Δ-regular tree

Lemma 5.14. *The uniqueness threshold (and the threshold for weak spatial mixing) of the hard-core model on $\mathbb{T}_\Delta$ is $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$.*

Proof. The hard-core model on the tree exhibits a kind of monotonicity. The two extreme boundary conditions are the all occupied and the all unoccupied boundary conditions at depth k; if k is even the the all occupied boundary conditions favor the root being occupied, and if k is odd the all occupied boundary conditions favor the root being unoccupied. Recall that we can implement hard-core boundary conditions by pruning the tree. Therefore to check for weak spatial mixing we can check whether the occupation probability of the root in the depth-k pruning of $\mathbb{T}_\Delta$ converges as $k \to \infty$ or not.

Consider iterating the recursion $F(R) = \lambda \left(\frac{1}{1+R}\right)^{\Delta-1}$ twice to get

$$G(R) = \lambda \left(\frac{1}{1+\lambda \left(\frac{1}{1+R}\right)^{\Delta-1}} \right)^{\Delta-1}\,.$$

We want to understand if $G(R)$ has a unique fixed point or not. Some calculus will tell us that if $\lambda < \lambda_c(\Delta)$ then $|G'(R)| < 1$ for all $R > 0$ and so G has a unique fixed point. On the other hand, when $\lambda > \lambda_c(\Delta)$ there are multiple fixed points, as we can see in Figure 5.

□

The three fixed points can be interpreted as follows. For any Δ, λ there is a unique fixed point $F(R) = R$. This corresponds to the unique *translation-invariant* infinite volume Gibbs measure on $\mathbb{T}_\Delta$ (translation-invariant in the sense that the occupation probability of every vertex is equal).

On the other hand, when $\lambda > \lambda_c(\Delta)$, the two additional fixed points correspond to the two extremal semi-translation-invariant infinite volume Gibbs measures (occupation probabilities depend on the parity of the depth of the vertex). One can be obtained by the limit of all even occupied boundary conditions and the other as the limit of the all odd occupied boundary conditions.

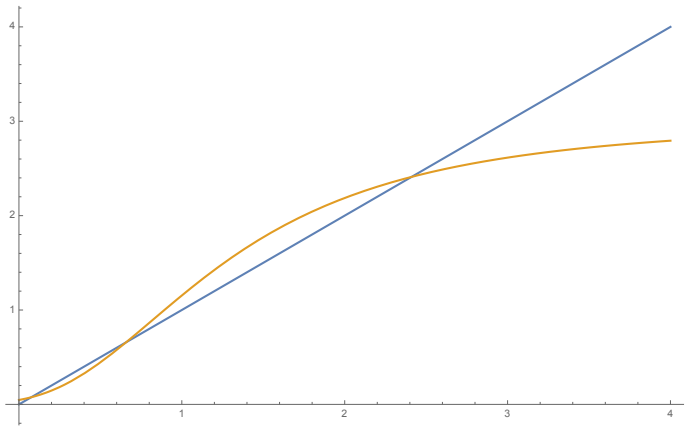

Figure 5: The function $G(R)$ for $\Delta = 4, \lambda = 3.0$

The heart of Weitz's argument is that weak spatial mixing implies strong spatial mixing on the tree. In particular, subtrees of $\mathbb{T}_\Delta$ can be obtained by imposing boundary conditions, strong spatial mixing holds for the self-avoiding walk tree when $\lambda < \lambda_c(\Delta)$.

5.5 Series truncation algorithms

Another approach to approximate counting is the *polynomial interpolation* method of Barvinok [5, 6, 7] which relies on a complex zero-free region of a univariate partition function. This is closely related to an algorithmic version of the cluster expansion which relies on some convergence criterion being satisfied. The two approaches are of course closely linked. The cluster expansion can be used to prove the existence of a zero-free region. Moreover, the cluster expansion is the multivariate Taylor series of the log partition function and so specializing to equal activities the two series are identical. Nonetheless, it is worthwhile considering both approaches separately since they have different conditions and apply in different situations.

A simple but powerful observation of Barvinok is that if a polynomial $p(x)$ of degree n has no zeros in the disk of radius R around 0 in the complex plane, then truncating the Taylor series for $\log p(x)$ around 0 after $O(\log(n/\varepsilon))$ terms gives a good approximation to $\log p(x)$ if x is in the interior of the disk.

Lemma 5.15. *Let $p(x)$ be a polynomial of degree n and suppose that $p(x) \neq 0$ when $|x| < R$ for some $R > 0$. Define*

$$T_m(x) = \log p(0) + \sum_{j=1}^{m} \frac{(\log p)^{(j)}(0)}{j!} x^j$$

be the mth order Taylor series for $\log p(x)$ around 0. Then if $|x| \leq (1-\eta)R$ for some $\eta > 0$,

$$|\log p(x) - T_m(x)| \leq \frac{n(1-\eta)^m}{(m+1)\eta}.$$

Proof. Let $z_1, \ldots, z_n \in \mathbb{C}$ denote the roots of the polynomial $p(x)$ (with multiplicity). We

can write

$$p(x) = p(0) \prod_{j=1}^{n} \left(1 - \frac{x}{z_j}\right)$$

or

$$\log p(x) = \log p(0) + \sum_{j=1}^{n} \log \left(1 - \frac{x}{z_j}\right).$$

Now since $\left|\frac{x}{z_j}\right| \leq 1 - \eta$, the Taylor series for $\log\left(1 - \frac{x}{z_j}\right)$ is absolutely convergent and

$$\begin{aligned}
\left|\log\left(1 - \frac{x}{z_j}\right) - \sum_{i=1}^{m} \frac{(-1)^i (x/z_j)^i}{i}\right| &\leq \sum_{i=m+1}^{\infty} \frac{(1-\eta)^i}{i} \\
&\leq \frac{1}{m+1} \sum_{i=m+1}^{\infty} (1-\eta)^i \\
&= \frac{(1-\eta)^{m+1}}{(m+1)\eta}.
\end{aligned}$$

Summing the error over all the n roots proves the lemma. □

In particular, Lemma 5.15 tells us that $\exp(T_m(x))$ is an ε-relative approximation to $p(x)$ when $m \geq C(\eta)\log(n/\varepsilon)$ and p and x satisfy the conditions of the theorem.

To implement this as an algorithm, we need to compute the coefficients b_k of the power series $T_m(x) = \sum_{k=0}^{m} b_k x^k$. In specific cases, such as the hard-core partition function, the cluster expansion gives us a formula for these coefficients: $b_k = \sum_{|\Gamma|=k} \phi(H(\Gamma))$, where the sum is over all clusters of G of size k.

For a generic polynomial p we do not have such a formula, but we can still compute the coefficients inductively. With $p(x) = \sum_{k=0}^{n} a_k x^k$, we can compute b_k by solving a linear system involving $b_0, \ldots, b_k$ and $a_0, \ldots, a_k$ (exercise!). This reduces the problem to compute the first $O(\log(n/\varepsilon))$ coefficients of p. We will see some examples of this below.

5.5.1 Application: the ferromagnetic Ising model

Consider the ferromagnetic Ising model partition function with external field λ:

$$Z_G(\beta, \lambda) = \sum_{\sigma \in \{\pm 1\}^{V(G)}} e^{\beta M(G,\sigma)} \lambda^{|\{v:\sigma_v = +1\}|}.$$

Viewed as a function of λ, $Z_G(\beta, \lambda)$ is a polynomial of degree n (with coefficients that depend on β and the graph structure). The Lee–Yang Theorem tells us that all of the zeros of this polynomial lie on the unit circle $|\lambda| = 1$ (in the ferromagnetic case $\beta \geq 0$). Now fix λ with $|\lambda| = 1 - \eta < 1$. Then when $m \geq C(\eta)\log(n/\varepsilon)$, $\exp(T_m)$ is an ε-relative approximation to

$Z_G(\beta, \lambda)$. We just need to compute the coefficients a_k of $Z_G(\beta, \lambda)$ to compute the coefficients of the Taylor series T_m. We have

$$a_k = \sum_{\sigma \in \binom{V}{k}} e^{\beta M(G,\sigma)},$$

and so we can compute this in time $O(n^k)$ by summing over all sets of k vertices of G. This does not give a polynomial-time algorithm, since we need to consider $k = \Omega(\log(n/\varepsilon))$; rather it gives a quasi-polyniomial-time algorithm running in time $n^{O(\log n)}$ for fixed ε. Several applications of Barvinok's polynomial interpolation method have running times of this form. In the case of the Ising model (and other spin models on graphs), when G is of maximum degree Δ the running time for computing a_k can be improved to $n^{O(\log \Delta)}$ which is polynomial time when Δ is fixed.

5.6 Algorithms at low temperatures

All of the algorithmic approaches to approximate counting and sampling we have seen so far work at high temperatures: when interactions (or densities) are weak enough. Each of the algorithmic techniques has a corresponding technique for proving uniqueness of infinite Gibbs measures, and so phase transitions present an inherent barrier to these methods. What can we do at low temperatures? Or in the presence of phase coexistence?

Polymer models will give us one way to design low temperature algorithms, by allowing us to switch from a low-temperature (or high-density) spin model to a high-temperature (weakly interacting) polymer model to which we can apply several different algorithmic techniques. This type of low temperature algorithm first appeared in [41] in the form of efficient approximate counting and sampling algorithms for the large β Potts model and large λ hard-core models on $\mathbb{Z}^d$. These algorithms used Pirogov-Sinai theory, a class of contour models which are more complex than polymer models. We will instead start by describing algorithms from [43, 16] which use polymer models.

5.6.1 Making the cluster expansion algorithmic

Using the Kotecký–Preiss condition we can truncate the cluster expansion to get a good approximation of $\log Z$. In particular, if the condition holds for some $b > 0$, then with

$$T_m = \sum_{|\Gamma| \le m} \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v$$

we have

$$|\log Z_G(\boldsymbol{\lambda}) - T_m| \le n e^{-bm}.$$

This tells us that $\exp(T_m)$ is an ε-relative approximation to $Z_G(\boldsymbol{\lambda})$ when $m \ge \log(n/\varepsilon)/b$.

To turn the polynomial interpolation method or the cluster expansion into a polynomial-time algorithm, we must be able to compute terms of the expansion efficiently.

The key to doing this is to be able to enumerate connected, induced subgraphs of a graph G efficiently. To obtain polynomial time algorithms we would like to enumerate all connected, induced subgraphs of G of size at most k in time polynomial in n and exponential in k. Patel and Regts give an algorithm to do this [60]. We give a sketch of their algorithm and proof.

We use the following simple but useful bound from [12].

Lemma 5.16. *Let G be a graph of max degree Δ and fix $v \in V(G)$. Then the number of connected, induced subgraphs of G of size k that contain v is at most $\frac{(e\Delta)^{k-1}}{2}$.*

The proof of this lemma is by comparison with a Δ-regular tree. Question: what is the best upper bound you can prove for connected subgraphs of size k containing v (not necessarily induced)?

The algorithm of Patel and Regts proceeds as follows. Let $\mathcal{T}_k$ be the set of all connected induced subgraphs of G of size k. Constructing $\mathcal{T}_1$ can be done in linear time since it is simply the set $V(G)$. Now given $\mathcal{T}_{k-1}$ we can construct $\mathcal{T}_k$ by appending to each $S \in \mathcal{T}_{k-1}$ one of its neighbors. There are at most $(k-1)\Delta$ such neighbors and at most $n\frac{(e\Delta)^{k-2}}{2}$ such subgraphs S so this can be done in time polynomial in n and exponential in k. We then iterate through the (multi)-sets created and remove duplicates, leaving us with $\mathcal{T}_k$.

Finally, to compute terms of the cluster expansion efficiently we need to compute the Ursell function of a graph H in time exponential in the number of vertices. To see that we can do this, we first recall the *Tutte polynomial*: for an n vertex graph G with c connected components, define

$$T_G(x,y) = \sum_{A \subset E(G)} (x-1)^{c(A)-c}(y-1)^{c(A)+|A|-n}$$

where $c(A)$ is the number of connected components of the graph $G_A = (V(G), A)$. The Tutte polynomial is related to many other important graph polynomials. Up to a change of variables it is the random cluster model partition function. Specializations of the Tutte polynomial include the chromatic polynomial and the Ising and Potts model partition functions.

We use the fact that evaluating $T_H(x,y)$ at $(1,0)$ and scaling gives the Ursell function $\phi(H)$. Naively from the definition of the Tutte polynomial (or the Ursell function) it looks like exact computation would take time $2^{|E(G)|}$, by summing over all edge subsets. Fortunately, exact computation (in fact, computing the coefficients of $T_H(x,y)$) can be done more efficiently, in vertex exponential time using an algorithm of Björklund, Husfeldt, Kaski, and Koivisto [10]. Thus computing the Ursell function of a graph H of size $O(\log(n/\varepsilon))$ can be done in time polynomial in n and $1/\varepsilon$.

5.6.2 Hard-core model on unbalanced bipartite graphs

Consider a Δ regular bipartite graph G with bipartition (L,R) each of size n (so G has $2n$ vertices). Suppose every vertex in L has fugacity λ_L and every vertex in R has fugacity λ_R. Let $Z_G(\lambda_L, \lambda_R)$ denote the partition function. If λ_L and λ_R are both small as a function of Δ (say $\leq 1/(e\Delta)$) then our previous algorithmic approaches will give efficient approximate counting and sampling algorithms. If $\lambda_L = \lambda_R = \lambda$ and λ is large, the approximating $Z_G(\lambda)$

is #BIS-hard: as hard as approximating the number of independent sets in a general bipartite graph.

What if λ_L and λ_R are large but there is an imbalance: say $\lambda_L > \lambda_R$? We will show that if the unbalance is large enough as a function of Δ then there is an FPTAS for $Z_G(\lambda_L, \lambda_R)$.

What might we expect a typical independent set from the hard-core model to look like if $\lambda_L \gg \lambda_R$? We might expect such an independent set to be skewed, with most of its vertices coming from L and few vertices coming from R.

To make this intuition precise, we can consider a generalized notion of a ground state. Previously a ground state was a spin configuration of maximum weight; in this case, a maximum size independent set (either the all L occupied or the all R occupied independent sets) or when λ is small, the empty set. Now we can think of a ground state as being a collection of spin configurations of collectively large weight that we understand well. In this example, following our intuition we will take this generalized ground state to be the set of all independent sets that contain no vertex from R. The weight of this set of configurations is simply $(1 + \lambda_L)^n$, since each vertex in L can be in or out without constraint.

Now suppose a set S of vertices in R is occupied. What is the total weight of the corresponding independent sets? We get a factor $\lambda_R^{|S|}$ from these occupied vertices. The vertices in L that belong to the neighborhood of S cannot be in the independent set, but all others can. So this contributes a factor $(1 + \lambda_L)^{n-|N(S)|}$. All together, and pulling out the ground state weight, we have a contribution of $(1 + \lambda_L)^n \cdot \frac{\lambda_R^{|S|}}{(1+\lambda_L)^{|N(S)|}}$. We call $w(S) = \frac{\lambda_R^{|S|}}{(1+\lambda_L)^{|N(S)|}}$ the (relative) weight of $S \subseteq R$.

We say that two vertices $u, v \in R$ are 2-linked if they are common neighbors of a vertex $w \in L$. The 2-linked components of a subset $S \subseteq R$ are the maximal connected components of S under the 2-linked relation. In particular, if the 2-linked components of S are $\gamma_1, \ldots, \gamma_t$ then $N(\gamma_i) \cap N(\gamma_j) = \emptyset$ for all $i \neq j$. Therefore, we see that the weight of S factorizes over its 2-linked components:

$$w(S) = \prod_{j=1}^{t} \frac{\lambda_R^{|\gamma_j|}}{(1 + \lambda_L)^{|N(\gamma_j)|}} = \prod_{j=1}^{t} w(\gamma_j)\,.$$

With this factorization, we can define a polymer model. Polymers are 2-linked subsets $\gamma \subseteq R$; the weight function is $w(\gamma) = \frac{\lambda_R^{|\gamma|}}{(1+\lambda_L)^{|N(\gamma)|}}$; and two polymers γ, γ' are compatible if their union is not 2-linked. Then from the discussion above we have the identity

$$Z_G(\lambda_L, \lambda_R) = (1 + \lambda_L)^n \cdot Z \tag{44}$$

where Z is the polymer model partition function.

Now we can verify the Kotecký–Preiss condition.

Lemma 5.17. *Let G be a Δ-regular bipartite graph on bipartition (L, R) and consider the hard-core model with fugacities λ_L, λ_R. When*

$$\lambda_L \geq \frac{3}{2} e^3 \Delta^2 \lambda_R - 1 \tag{45}$$

the Kotecký–Preiss condition holds for the polymer model defined above, by taking $a(\gamma) = |\gamma|$ *and* $b(\gamma) = |\gamma|$.

Proof. We first bound the number of polymers of size k incompatible with a given polymer γ by $|\gamma|\Delta^2\frac{(e\Delta^2)^{k-1}}{2}$ using a similar argument to that for the Ising polymer model above. We next bound the weight of a polymer of size k: since G is a regular bipartite graph, we have $|N(\gamma)| \geq |\gamma|$, and so if $|\gamma| = k$,

$$w(\gamma) \leq \left(\frac{\lambda_R}{1+\lambda_L}\right)^k .$$

Then for any polymer γ,

$$\begin{aligned}
\sum_{\gamma' \nsim \gamma} w(\gamma') e^{a(\gamma')+b(\gamma')} &= \sum_{\gamma' \nsim \gamma} w(\gamma') e^{2|\gamma'|} \\
&\leq \sum_{k \geq 1} |\gamma| \Delta^2 \frac{(e\Delta^2)^{k-1}}{2} \left(\frac{\lambda_R}{1+\lambda_L}\right)^k e^{2k} \\
&\leq \frac{|\gamma|}{2} \sum_{k \geq 1} \left(\frac{e^3\Delta^2\lambda_R}{1+\lambda_L}\right)^k \\
&\leq |\gamma|
\end{aligned}$$

if $\lambda_L \geq \frac{3}{2}e^3\Delta^2\lambda_R - 1$. □

Corollary 5.18. *Fix* $\Delta \geq 3$ *and* λ_L, λ_R *so that* (45) *holds. Then there is an FTPAS for the hard-core model on the class of* Δ*-regular bipartite graphs with fugacities* λ_L, λ_R.

Proof. The FTPAS follows directly from the identity (44). In particular, Lemma 5.17 tells us that

$$\left| \log Z_G(\lambda_L, \lambda_R) - n \log(1+\lambda_L) - \sum_{\Gamma : b(\Gamma) \leq t} \phi(H(\Gamma)) \prod_{\gamma \in \Gamma} w(\gamma) \right| \leq n e^{-t}$$

and so truncating the cluster expansion up to terms of 'size' $\log(n/\varepsilon)$, where the size of a cluster is $b(\Gamma) = \sum_{\gamma \in \Gamma} b(\gamma)$, gives an ε-relative approximation to $Z_G(\lambda_L, \lambda_R)$. For Δ fixed, this can be done in time polynomial in n and $1/\varepsilon$. □

5.6.3 Potts model on expander graphs

Recall the discussion of the q-color ferromagnetic Potts model on expander graphs from Lecture 3. We defined a polymer model where polymers were connected induced subgraphs of G of size less than $n/2$, colored with the $q-1$ non-ground-state colors. Let $\hat{Z}$ denote the polymer model partition function (by symmetry it is identical for each of the q possible ground states). We showed that if G is an α-edge-expander and $\beta \geq \frac{4+2\log(q\Delta)}{\alpha}$ then

$$e^{-e^{-n}} q \cdot e^{\beta|E(G)|} \hat{Z} \leq Z_G(\beta) \leq e^{e^{-n}} q \cdot e^{\beta|E(G)|} \hat{Z} ,$$

or in other words, $qe^{\beta|E(G)|}\hat{Z}$ is an e^{-n}-relative approximation to the partition function $Z_G(\beta)$.

In terms of an FPTAS this suggest two cases: if, say, $\varepsilon \leq 2e^{-n}$ we can just compute the partition function by brute force, enumerating over the q^n configurations; this is acceptable since time $e^{O(n)}$ is polynomial in $1/\varepsilon$ in this case. On the other hand, if $\varepsilon > 2e^{-n}$, then we can find an $\varepsilon/2$-relative approximation to $\hat{Z}$, multiply it by $qe^{\beta|E(G)|}$ and obtain an ε-relative approximation to $Z_G(\beta)$.

6 Exercises

1. Prove that the occupancy fraction of the hard-core model on any non-empty graph G is a strictly increasing function of λ.

2. Prove that for the hard-core model on $\mathbb{Z}$, the truncated two-point correlation function decays exponentially fast in the distance, for any $\lambda \geq 0$.

3. Consider a tree T and a vertex $v \in V(T)$ with neighbors $u_1, \ldots, u_d$. Write a formula for the marginal $\mu_{v,\lambda}$ of the hard-core model at fugacity λ in terms of λ and the marginals $\mu^{-v}_{u_1,\lambda}, \ldots, \mu^{-v}_{u_d,\lambda}$ of the neighbors of v in the graph $T \setminus v$.

4. Derive a formula for the three-point truncated correlation function for vertices u, v, w in a 2-spin model on a graph in terms of marginals and joint marginals.

5. Let $\mathcal{G}_n$ be the set of all (labeled) graphs on n vertices. For $m \in (0, \binom{n}{2})$, determine the maximum entropy probability distribution on $\mathcal{G}_n$ with mean number of edges m. (Recall that the entropy of a probability distribution μ on a finite set Ω is $H(\mu) = -\sum_{x\in\Omega} \mu(x) \log \mu(x)$ with the convention that $0 \log 0 = 0$).

6. Let K_d be the complete graph (clique) on d vertices.

   (a) Compute the hard-core partition function $Z_{K_d}(\lambda)$.

   (b) For $u, v \in K_d$ compute the truncated two-point correlation function.

7. Let $G = G_1 \cup G_2$, the disjoint union of two graphs G_1, G_2. Prove that

$$Z_G(\lambda) = Z_{G_1}(\lambda) Z_{G_2}(\lambda)\,.$$

8. Let $K_{d,d}$ be the complete d-regular bipartite graph (two sets of d vertices L, R) with all d^2 edges between L and R present and no others).

   (a) Compute $Z_{K_{d,d}}(\lambda)$.

   (b) Compute $\mathbb{E}_{K_{d,d},\lambda}|I|$, the expected size of an independent set I drawn from the hard-core model on $K_{d,d}$ at fugacity λ.

9. Prove that the following probability distribution on independent sets of G is the hard-core model on G at fugacity λ. Pick a subset $S \subseteq V(G)$ by including each vertex independently with probability $\frac{\lambda}{1+\lambda}$ and condition on the event that S is an independent set.

10. Consider the hard-core model on a graph G and let F be the set of vertices that are not in the independent set and have no neighbor in the independent set (they are free to be added to the independent set). Calculate $\mathbb{E}[|F|]$ in terms of derivatives of $\log Z_G(\lambda)$.

11. Let $i_k(G)$ be the number of independent sets of size k in a graph G.

    (a) Give a probabilistic interpretation (as, say, and expectation) for the quantity $\frac{i_{k+1}(G)}{i_k(G)}$ in terms of the uniform distribution over independent sets of size k in G.

(b) Prove that for all G of maximum degree Δ on n vertices,

$$\frac{i_{k+1}(G)}{i_k(G)} \geq \frac{n-(\Delta+1)k}{k+1}\,,$$

and find a family of graphs for which the inequality is tight.

(c) Use the above to prove that for all G of maximum degree Δ on n vertices,

$$\frac{1}{n}\log Z_G(\lambda) \geq \frac{1}{\Delta+1}\log(1+(\Delta+1)\lambda)\,,$$

and show that the inequality is tight. (Hint: recall that partition functions are multiplicative over disjoint graphs and that $Z_G(\lambda)$ is a polynomial).

12. Prove that for any $\lambda > 0$,

$$i_k(G) = \frac{Z_G(\lambda)}{\lambda^k}\Pr_{G,\lambda}(|I| = k)$$

where $i_k(G)$ is the number of independent sets of size k in G and the probability (and the partition function) is with respect to the hard-core model on G at fugacity λ.

13. Pick an independent set I from the hard-core model on a d-regular, triangle-free graph G at fugacity $\lambda > 0$, and pick v uniformly at random. Let Y be the number of uncovered neighbors of v with respect to I (an integer-valued random variable bounded between 0 and d).

(a) Show that if Y is supported on 0 and d, then G is the complete bipartite graph $K_{d,d}$ or a union of $K_{d,d}$'s.

(b) Suppose G has no component isomorphic to $K_{d,d}$. Prove a positive lower bound on $\Pr(Y \notin \{0,d\})$ in terms of d and λ. (The lower bound should not depend on the size of the graph G).

(c) Using the previous two results and the theorem we proved in class, prove that for every d and $\lambda > 0$ there exists $\varepsilon > 0$ so that for every d-regular, triangle-free graph G on n vertices without a $K_{d,d}$ component,

$$\frac{1}{n}\log Z_G(\lambda) \leq \frac{1}{2d}\log Z_{K_{d,d}}(\lambda) - \varepsilon\,.$$

(The result extends to graphs that may contain triangles).

14. Gdenko's Local Central Limit Theorem states the following: Let X be an integer valued random variable with mean μ and variance σ^2 whose support has gcd 1. Let $X_1, X_2, \ldots$ be iid copies of X and let $S_n = \sum_{j=1}^n X_j$. Then for every integer k,

$$\Pr(X = k) = \frac{1}{\sqrt{2\pi\sigma^2 n}} e^{-\frac{(k-\mu)^2}{2\sigma^2 n}} + o(n^{-1/2})\,.$$

(Why is the gcd condition necessary?)

Let $H_{d,n}$ be the graph that is a union of $n/2d$ copies of $K_{d,d}$.

(a) Show that for every $k \in \{0, 1, \ldots, n/2\}$ there exists $\lambda \geq 0$ so that the expected size of an independent set drawn from the hard-core model on $H_{d,n}$ at fugacity λ is exactly k.

(b) Fix $\varepsilon > 0$ and suppose $\varepsilon n < k < (1-\varepsilon)n/2$. Choose λ so that the expected size size of an independent set drawn from the hard-core model on $H_{d,n}$ at fugacity λ is exactly k. Show that

$$\Pr(|I| = k) = \Theta(n^{-1/2}) .$$

(c) Prove the following. For every $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon, d)$ large enough so that the following holds: for all $n \geq n_0$ divisible by $2d$, every d-regular G on n vertices that does not contain a component isomorphic to $K_{d,d}$, and every $\varepsilon n \leq k \leq n/2$,

$$i_k(G) < i_k(H_{d,n}) .$$

(Hint: use the result from Question 2 as an input)

(d) (Harder) Prove the same result without the assumption that G contains no component isomorphic to $K_{d,d}$. Hint: consider two cases, depending on the fraction of vertices in $K_{d,d}$ components. In the case that there are many $K_{d,d}$ components, analyze what the distribution of a uniformly random independent set of size k looks like restricted to the part of G that is not in a $K_{d,d}$ component.

15. Compute the moment generating function of a Poisson(λ) random variable and use this to prove a formula for the cumulants of a Poisson random variable.

16. Prove that the clique K_{d+1} has the highest triangle density (number of triangles divided by number of vertices) of any d-regular graph.

17. Show that there is a sequence of d-regular graphs G_d so that the smallest complex root (in complex absolute value) of $Z_{G_d}(\lambda)$ is $\Theta(1/d)$.

18. Fix $\Delta > 0$ and $0 < \lambda < \frac{1}{e(\Delta+1)}$. Let G_n be a sequence of n vertex graphs of max degree Δ. Let X_n be the size of a random independent set drawn from the hard-core model on G_n at activity λ.

(a) Prove that $\mathrm{var}(X_n) = \Omega(n)$. Hint: use the law of total variance and the fact that G has a linear sized set of vertices at pairwise distance at least 3.

(b) Prove that $\mathrm{var}(X_n) = O(n)$. Hint: use cluster expansion convergence.

(c) For $k \geq 3$ fixed, prove an asymptotic upper bound on the kth cumulant of X, $\kappa_k(X)$.

(d) Deduce that X is asymptotically normal; that is, $(X - \mathbb{E}X)/\sqrt{\mathrm{var}(X)} \Rightarrow N(0,1)$.

(e) Write a formula using the cluster expansion for the cumulant generating function of X, $\log \mathbb{E}e^{tX}$. For what t does this converge?

(f) Using the previous result prove a large deviation result for X, i.e. the best upper bound you can on the probability

$$\Pr(X \geq (1+\delta)\mathbb{E}X) .$$

19. Prove that the clique K_{d+1} has the highest triangle density (number of triangles divided by number of vertices) of any d-regular graph, and that there is gap to any graph that does not contain a K_{d+1} component.

20. Use the previous result and the cluster expansion for the generating function of matchings (monomer-dimer partition function) to prove that for some $\lambda^*(d) > 0$, all $0 < \lambda < \lambda^*$ and all d-regular graphs G not containing a K_{d+1} component we have

$$\frac{1}{|V(G)|} \log Z_G^{\text{match}}(\lambda) > \frac{1}{d+1} \log Z_{K_{d+1}}^{\text{match}}(\lambda) .$$

The monomer-dimer partition function is

$$Z_G^{\text{match}}(\lambda) = \sum_{M \in \mathcal{M}(G)} \lambda^{|M|}$$

where the sum is over all matchings of G. A matching is an independent set in $L(G)$, the line graph of G.

(Harder) Can you extend this to prove that for all d-regular G and all $0 < \lambda < \lambda^*$

$$\frac{1}{|V(G)|} \log Z_G^{\text{match}}(\lambda) \geq \frac{1}{d+1} \log Z_{K_{d+1}}^{\text{match}}(\lambda) .$$

21. Let G be a Δ-regular bipartite graph on bipartition (L, R) each of size n, with $\Delta = cn^{1/3}$. Suppose each vertex in L has fugacity $\lambda_L = \ell n^{-2/3}$ and each vertex in R has fugacity $\lambda_R = rn^{-2/3}$. Let $\mathbf{I}_L, \mathbf{I}_R$ be the number of occupied vertices in L and R respectively.

    (a) Use the cluster expansion to write an asymptotic formula for $\mathbb{E}|\mathbf{I}_L|$ and $\mathbb{E}|\mathbf{I}_R|$.

    (b) Use the cluster expansion to write an asymptotic formula for $\text{cov}(\mathbf{I}_L, \mathbf{I}_R)$.

    (c) Prove that after suitable centering and scaling the random vector $(\mathbf{I}_L, \mathbf{I}_R)$ converges to a bivariate Gaussian.

22. Let G be a biregular, bipartite graph with bipartition (L, R) and suppose every vertex in L has degree Δ_L and every vertex in R has degree Δ_R. Assume that $\Delta_R > \Delta_L$.

    (a) When Δ_R is much bigger than Δ_L, what do you expect typical uniformly random independent sets from G to look like?

    (b) Write the hard-core partition function of G as the partition function of a polymer model measuring deviations from the generalized ground state of the independents sets with no vertex from R.

    (c) How large must Δ_R be as a function of Δ_L to guarantee convergence of the cluster expansion for the polymer model when $\lambda = 1$?

23. Prove that the following are equivalent (by providing polynomial-time reductions):

    - There is an FPRAS for $Z_G(\lambda)$ for the class of max degree Δ graphs.

- There is a polynomial-time approximate sampling algorithm for $\mu_{G,\lambda}$ for the class of max degree Δ graphs.

24. Prove that if there is a polynomial-time approximate sampling algorithm for $\mu_{G,\lambda}$ for some class of graphs $\mathcal{G}$ and all $0 < \lambda < \lambda_0$ then there is an FPRAS for $Z_G(\lambda)$ for $\mathcal{G}$ and and all $0 < \lambda < \lambda_0$.

25. The lazy random walk on a graph is the following Markov chain: starting at a vertex v, with probability $1/2$ stay at v and with probability $1/2$ move to a uniformly random neighbor of v.

    (a) Suppose G is regular and connected. Prove that the lazy random walk is irreducible and aperiodic (look up these terms on google if needed!). What is the stationary distribution?

    (b) If G is connected and regular on n vertices, what is the order of the best and worst possible mixing times of the lazy random walk? (I.e. find good and bad graphs for this Markov chain). What if G is Δ-regular for Δ constant? What's the best mixing time you can expect?

26. (Coupling). Consider the lazy random walk on the hypercube $Q_d = \{0,1\}^d$. Start two copies of this Markov chain from different states $X_0, Y_0 \in Q^d$. Find a coupling of the two processes X_n, Y_n so the expected number of steps until they coincide ($X_n = Y_n$) is as small as possible.

27. (Total variation distance) For $m \geq n$ consider the following distributions of configurations of m balls in n labeled bins: 1) place each of the m balls independently in uniformly chosen random bins; 2) start with one ball in each bin and place the remaining $m - n$ balls independently in uniformly chosen random bins. Call the two distribution μ_1, μ_2 respectively (both depend on n and m).

    (a) Find good strategy for the following game: I pick μ_1 or μ_2 with probability $1/2$ each and show you one sample from the given distribution; from the sample you have to guess which distribution it came from. For what $m = m(n)$ can you win this game with probability $1 - o(1)$?

    (b) What does the strategy and probability of winning have to do with $\|\mu_1 - \mu_2\|_{TV}$?

    (c) Can you find the optimal threshold in $m = m(n)$ for $\|\mu_1 - \mu_2\|_{TV} \to 0$?

Acknowledgements

These notes are based on mini-courses given at the Prague Summer School in Discrete Mathematics, July 2022, the IBS ECOPRO Winter School in February 2023, and the Lake Michigan Workshop in Combinatorics and Graph Theory in May 2023. Thanks to Zdeněk Dvořák, Hong Liu, and David Galvin for inviting me to give this lectures.

Thanks to many colleagues and coauthors who have taught me a lot about both combinatorics and statistical physics. Thanks to Xavier Povill for a careful reading of a draft of these notes and for correcting many typos and mistakes.

References

[1] S. Adams. *Lectures on mathematical statistical mechanics*. Number 30. Dublin Institute for Advanced Studies, 2006.

[2] M. Ajtai, J. Komlós, and E. Szemerédi. A note on Ramsey numbers. *Journal of Combinatorial Theory, Series A*, 29(3):354–360, 1980.

[3] N. Alon. Independent sets in regular graphs and sum-free subsets of finite groups. *Israel Journal of Mathematics*, 73(2):247–256, 1991.

[4] J. Balogh, R. I. Garcia, and L. Li. Independent sets in the middle two layers of Boolean lattice. *Journal of Combinatorial Theory, Series A*, 178:105341, 2021.

[5] A. Barvinok. *Combinatorics and complexity of partition functions*, volume 9. Springer, 2016.

[6] A. Barvinok. Computing the permanent of (some) complex matrices. *Foundations of Computational Mathematics*, 16(2):329–342, 2016.

[7] A. Barvinok and G. Regts. Weighted counting of solutions to sparse systems of equations. *Combinatorics, Probability and Computing*, 28(5):696–719, 2019.

[8] E. P. Bernard and W. Krauth. Two-step melting in two dimensions: first-order liquid-hexatic transition. *Physical Review Letters*, 107(15):155704, 2011.

[9] E. P. Bernard, W. Krauth, and D. B. Wilson. Event-chain Monte Carlo algorithms for hard-sphere systems. *Physical Review E*, 80(5):056704, 2009.

[10] A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto. Computing the Tutte polynomial in vertex-exponential time. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 677–686. IEEE, 2008.

[11] M. Borbenyi and P. Csíkvári. Matchings in regular graphs: minimizing the partition functionn. *Transactions on Combinatorics*, 10(2):1–23, 2021.

[12] C. Borgs, J. Chayes, J. Kahn, and L. Lovász. Left and right convergence of graphs with bounded degree. *Random Structures & Algorithms*, 42(1):1–28, 2013.

[13] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[14] L. Bregman. Some properties of nonnegative matrices and their permanents. *Soviet Math. Dokl*, 14(4):945–949, 1973.

[15] G. R. Brightwell, O. Häggström, and P. Winkler. Nonmonotonic behavior in hard-core and widom–rowlinson models. *Journal of statistical physics*, 94(3-4):415–435, 1999.

[16] S. Cannon and W. Perkins. Counting independent sets in unbalanced bipartite graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1456–1466. SIAM, 2020.

[17] C. Carlson, E. Davies, A. Kolla, and W. Perkins. Computational thresholds for the fixed-magnetization Ising model. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1459–1472, 2022.

[18] S. Chen, M. Delcourt, A. Moitra, G. Perarnau, and L. Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2216–2234. SIAM, 2019.

[19] J. Cooper, K. Dutta, and D. Mubayi. Counting independent sets in hypergraphs. *Combinatorics, Probability and Computing*, 23(4):539–550, 2014.

[20] P. Csikvári. Extremal regular graphs: the case of the infinite regular tree. *arXiv preprint arXiv:1612.01295*, 2016.

[21] J. Cutler and A. Radcliffe. The maximum number of complete subgraphs in a graph with given maximum degree. *Journal of Combinatorial Theory, Series B*, 104:60–71, Jan. 2014.

[22] E. Davies, M. Jenssen, and W. Perkins. A proof of the upper matching conjecture for large graphs. *Journal of Combinatorial Theory, Series B*, 151:393–416, 2021.

[23] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. Independent sets, matchings, and occupancy fractions. *Journal of the London Mathematical Society*, 96(1):47–66, 2017.

[24] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. On the average size of independent sets in triangle-free graphs. *Proceedings of the American Mathematical Society*, 146:111–124, 2018.

[25] E. Davies and W. Perkins. Approximately counting independent sets of a given size in bounded-degree graphs. In *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[26] R. Dobrushin. Estimates of semi-invariants for the Ising model at low temperatures. *Translations of the American Mathematical Society-Series 2*, 177:59–82, 1996.

[27] M. Dyer, L. A. Goldberg, C. Greenhill, and M. Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2004.

[28] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017.

[29] A. Galanis, Q. Ge, D. Štefankovič, E. Vigoda, and L. Yang. Improved inapproximability results for counting independent sets in the hard-core model. *Random Structures & Algorithms*, 45(1):78–110, 2014.

[30] D. Galvin. A threshold phenomenon for random independent sets in the discrete hypercube. *Combinatorics, Probability and Computing*, 20(1):27–51, 2011.

[31] D. Galvin. Three tutorial lectures on entropy and counting. *arXiv preprint arXiv:1406.7872*, 2014.

[32] D. Galvin. Independent sets in the discrete hypercube. *arXiv preprint arXiv:1901.01991*, 2019.

[33] D. Galvin and J. Kahn. On phase transition in the hard-core model on $\mathbb{Z}^d$. *Combinatorics, Probability and Computing*, 13(02):137–164, 2004.

[34] D. Galvin and P. Tetali. On weighted graph homomorphisms. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 63:97–104, 2004.

[35] D. Galvin and P. Tetali. Slow mixing of Glauber dynamics for the hard-core model on regular bipartite graphs. *Random Structures & Algorithms*, 28(4):427–443, 2006.

[36] H.-O. Georgii. Gibbs measures and phase transitions. In *Gibbs Measures and Phase Transitions*. de Gruyter, 2011.

[37] C. D. Godsil. Matchings and walks in graphs. *Journal of Graph Theory*, 5(3):285–297, 1981.

[38] J. Groeneveld. Two theorems on classical many-particle systems. *Phys. Letters*, 3, 1962.

[39] C. Gruber and H. Kunz. General properties of polymer systems. *Communications in Mathematical Physics*, 22(2):133–161, 1971.

[40] O. J. Heilmann and E. H. Lieb. Theory of monomer-dimer systems. *Communications in Mathematical Physics*, 25(3):190–232, 1972.

[41] T. Helmuth, W. Perkins, and G. Regts. Algorithmic Pirogov-Sinai theory. *Probability Theory and Related Fields*, 176:851–895, 2020.

[42] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[43] M. Jenssen, P. Keevash, and W. Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM Journal on Computing*, 49(4):681–710, 2020.

[44] M. Jenssen and W. Perkins. Independent sets in the hypercube revisited. *Journal of the London Mathematical Society*, 102(2):645–669, 2020.

[45] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM journal on computing*, 18(6):1149–1178, 1989.

[46] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on computing*, 22(5):1087–1116, 1993.

[47] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004.

[48] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical computer science*, 43:169–188, 1986.

[49] J. Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10(03):219–237, 2001.

[50] A. Korshunov and A. Sapozhenko. The number of binary codes with distance 2. *Problemy Kibernet*, 40:111–130, 1983.

[51] R. Kotecký and D. Preiss. Cluster expansion for abstract polymer models. *Communications in Mathematical Physics*, 103(3):491–498, 1986.

[52] T.-D. Lee and C.-N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Physical Review*, 87(3):410, 1952.

[53] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[54] H. Löwen. Fun with hard spheres. In *Statistical physics and spatial statistics*, volume 554, pages 295–331. Springer, 2000.

[55] J. Mayer and M. Mayer. *Statistical Mechanics*. John Wiley, 1940.

[56] J. E. Mayer and E. Montroll. Molecular distribution. *The Journal of Chemical Physics*, 9(1):2–16, 1941.

[57] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[58] F. Mousset, A. Noever, K. Panagiotou, W. Samotij, et al. On the probability of nonexistence in binomial subsets. *The Annals of Probability*, 48(1):493–525, 2020.

[59] A. Okounkov. The Ising model in our dimension and our times. *arXiv preprint arXiv:2207.03874*, 2022.

[60] V. Patel and G. Regts. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM Journal on Computing*, 46(6):1893–1919, 2017.

[61] R. Peled and W. Samotij. Odd cutsets and the hard-core model on $\mathbb{Z}^d$. *Annales de l'Institut Henri Poincare, Probabilites et Statistiques*, 50:975–998, 2014.

[62] O. Penrose. Convergence of fugacity expansions for fluids and lattice gases. *Journal of Mathematical Physics*, 4(10):1312–1320, 1963.

[63] D. Randall. Rapidly mixing markov chains with applications in computer science and physics. *Computing in Science & Engineering*, 8(2):30–41, 2006.

[64] M. N. Rosenbluth. Genesis of the Monte Carlo algorithm for statistical mechanics. In *AIP Conference Proceedings*, volume 690, pages 22–30. American Institute of Physics, 2003.

[65] D. Ruelle. Correlation functions of classical gases. *Annals of Physics*, 25:109–120, 1963.

[66] D. Ruelle. *Statistical mechanics: Rigorous results*. World Scientific, 1999.

[67] A. Sah, M. Sawhney, D. Stoner, and Y. Zhao. The number of independent sets in an irregular graph. *Journal of Combinatorial Theory, Series B*, 138:172–195, 2019.

[68] A. Sah, M. Sawhney, D. Stoner, and Y. Zhao. A reverse Sidorenko inequality. *Inventiones mathematicae*, pages 1–47, 2020.

[69] A. Sapozhenko. On the number of connected subsets with given cardinality of the boundary in bipartite graphs. *Metody Diskret Analiz*, 45:42–70, 1987.

[70] A. D. Scott and A. D. Sokal. The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *Journal of Statistical Physics*, 118(5-6):1151–1261, 2005.

[71] J. B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983.

[72] J. B. Shearer. On a problem of Spencer. *Combinatorica*, 5(3):241–245, 1985.

[73] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

[74] A. Sly. Computational transition at the uniqueness threshold. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 287–296, Las Vegas, NV, USA, Oct. 2010. IEEE.

[75] A. Sly and N. Sun. Counting in two-spin models on d-regular graphs. *Ann. Probab.*, 42(6):2383–2416, Nov. 2014.

[76] B. Szegedy. Edge coloring models and reflection positivity. *Journal of the American mathematical Society*, 20(4):969–988, 2007.

[77] D. Tong. Lectures on statistical physics. `https://www.damtp.cam.ac.uk/user/tong/statphys.html`.

[78] E. Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.

[79] D. Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 140–149. ACM, 2006.

[80] Y. Zhao. The number of independent sets in a regular graph. *Combinatorics, Probability and Computing*, 19(02):315–320, 2010.

[81] Y. Zhao. Extremal regular graphs: independent sets and graph homomorphisms. *The American Mathematical Monthly*, 124(9):827–843, 2017.