Five lectures on statistical mechanics methods in combinatorics

Will Perkins *

December 1, 2020

The goal of these lectures is to introduce the basics of statistical physics to people interested in extremal, enumerative, and probabilistic combinatorics. At the most basic level, I hope to provide a guide to translating terms from one field to the other: partition functions, Gibbs measures, ground states, correlation functions etc. At the next level, I want to describe the statistical physics way of looking at things: viewing systems through the lens of correlations, phases, and phase transitions. Finally I want to indicate how all of this can be put to use in combinatorics: what combinatorial methods can be developed based on the statistical physics perspective and what new questions in combinatorics can we ask based on this perspective.

These lectures will necessarily only cover a portion of the applications and connections of statistical physics to combinatorics. In particular I will say very little about several very interesting topics including the Lovász Local Lemma, spin models on random graphs, graphons and dense graphs, and entropy methods.

The five lectures will cover:

1. **Fundamentals of statistical physics**: Gibbs measures, partition functions, phase transitions, correlations. How to approach combinatorics from the perspective of statistical physics.

2. **Extremal combinatorics of sparse graphs**: maximizing and minimizing the number of independent sets in various classes of regular graphs. Linear programming and the occupancy method.

3. **Expansion methods and enumeration**: cluster expansion. Conditions for convergence. Consequences of a convergent cluster expansion.

4. **Combinatorics at low temperatures**: abstract polymer models. Multivariate hard-core model as a universal model. Low-temperature enumeration with polymer models and the cluster expansion.

5. **Sphere packings, kissing numbers, and the hard sphere model**: continuum models and applications in combinatorics.

*Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, math@willperkins.org

1 Fundamentals of statistical physics

Statistical physics is the study of matter via probabilistic and statistical methods. The field was born in the late 1800's with important contributions by Maxwell, Boltzmann, and Gibbs.

The main motivating question in statistical physics is

Question 1.1. *Can the macroscopic properties of matter (gasses, liquids, solids, magnets) be derived solely from their microscopic interactions?*

The beautiful idea behind statistical mechanics is that to understand a system with a huge number of interacting particles or components, instead of tracking the position and velocity of each particle, we can treat them as being distributed randomly, according to a probability distribution that takes into account the microscopic interactions between particles.

1.1 Gibbs measures and partition functions

For now we will focus on *spin models on graphs*.

Fix a finite set of spins Ω. For a graph $G = (V, E)$, the set of configurations is Ω^V, assignments of spins to the vertices of G.

We define an *energy function* (or Hamiltonian) from $\Omega^V \to \mathbb{R} \cup \{+\infty\}$:

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

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* at inverse temperature β is

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

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

$$\mu_G(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z_G(\beta)} .$$

The inverse temperature β controls the strength of the interaction in the model.

- 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 very easy to understand the ground states (e.g. all even/all odd configurations for hard-core; monochromatic configurations for Ising/Potts). In general though, this need not be the case. In particular, finding and understanding the ground states of anti-ferromagnetic models on random graphs is a challenging problem, both mathematically 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 temperature: a weakly interacting system has many of the properties of an independent system, while a strongly interacting system correlates strongly with the extremal object. The switch from one qualitative regime to the other is a phase transition, the main topic of statistical physics.

1.2 Examples

The following are some examples of statistical mechanics models to keep in mind during these lectures. To start thinking like physicist, you can imagine the underlying graph G is a finite box in $\mathbb{Z}^d$ (or even in $\mathbb{Z}^2$).

1. The hard-core model (hard-core lattice gas). Given a graph G, allowed configurations are independent sets. The probability we pick an independent set I is $\frac{\lambda^{|I|}}{Z_G(\lambda)}$ where $\lambda > 0$ is the *fugacity* or *activity*. We can take $\Omega = \{0, 1\}$ with $f(1) = \log \lambda$, $f(0) = 0$, and $g(1, 1) = +\infty$ (a hard constraint).

   The hard-core model is a toy model of gas, and on $\mathbb{Z}^d$ the model exhibits a gas/solid phase transition.

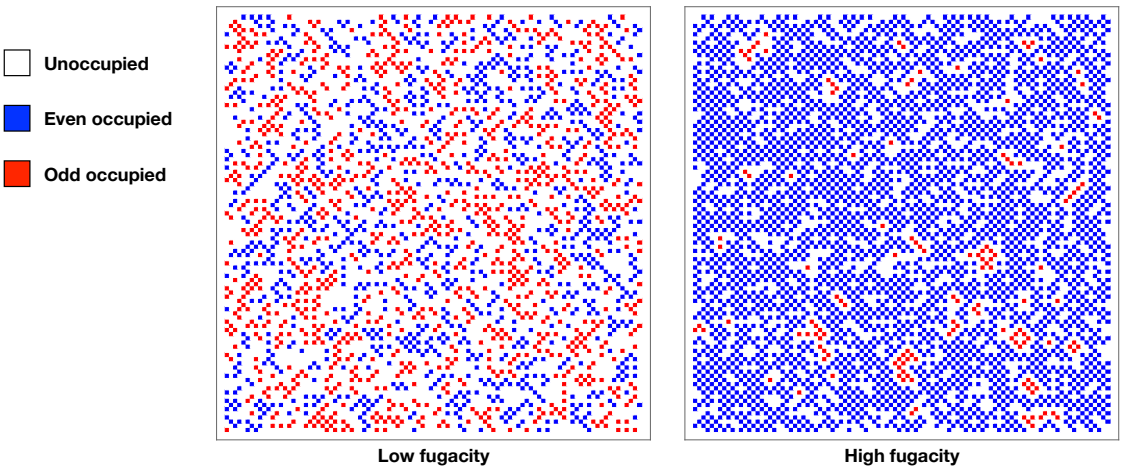


Figure 1: Two instances of the hard-core model on $\mathbb{Z}^2$

2. The Ising model. Configurations are assignments of ± 1 spins to the vertices of a graph. A configuration σ is chosen with probability $\frac{e^{\beta M(G,\sigma)}}{Z_G(\beta)}$ where $M(G, \sigma)$ is the number of edges of G whose vertices receive the same spin. That is, $g(\sigma_u, \sigma_v) = \sigma_u \sigma_v$.

If we think of the spins as being in/out the the Ising model is a probability distribution over cuts of G. The parameter β is the inverse temperature. $\beta \geq 0$ is the ferromagnetic case: same spins are preferred across edges. $\beta \leq 0$ is the antiferromagnetic case. The Ising model is a toy model of a magnetic material (it magnetizes when spins align globally).

3. The Potts model. The Potts model is a generalization of the Ising model to $q \geq 2$ spins (or colors). 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. Again $\beta \geq 0$ is the ferromagnetic and $\beta \leq 0$ the antiferromagnetic case.

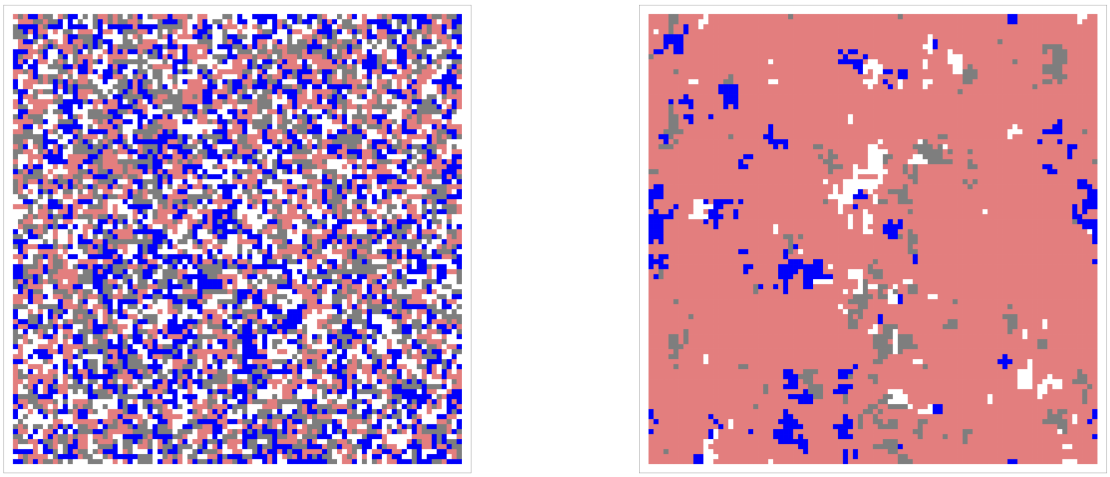


Figure 2: Two instances of the 4-color ferromagnetic Potts model on $\mathbb{Z}^2$

Not all Gibbs measures are spin models on graphs.

4. The monomer-dimer model. Allowed configurations are matchings in G, with $P(M) = \frac{\lambda^{|M|}}{Z_G(\lambda)}$. '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. [74]).

5. The hard sphere model. This is a continuum model of a gas and perhaps the original model in statistical mechanics.

6. The hard-core model on a hypergraph. Configurations are subsets S of vertices that contain no hyperedge, weighted by $\lambda^{|S|}$. The Hamiltonian now has terms corresponding to each hyperedge. Such an interaction is called a *multibody interaction*.

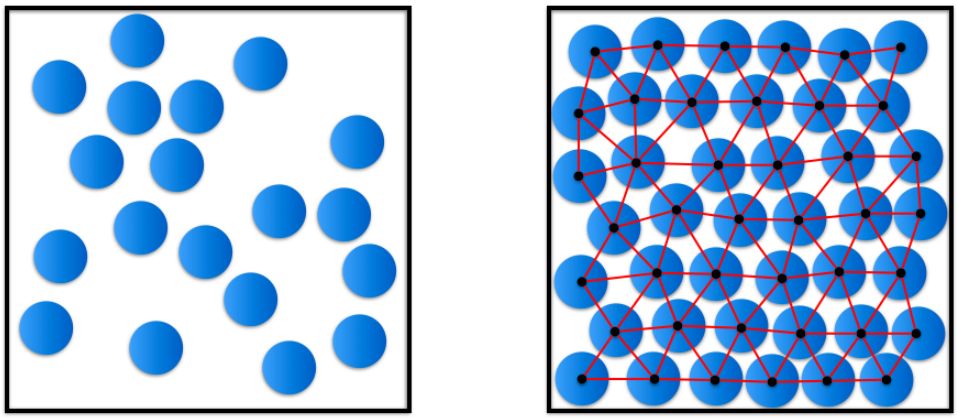


Figure 3: The hard sphere model at low and high density

1.3 Motivation for the form of the distribution

Why does a Gibbs distribution have an exponential (or 'log linear' form)? There are a few ways of answering this.

1. What was the original derivation of this form?

   If we imagine occupied vertices of an independent set are particles in a large box represented by a portion of $\mathbb{Z}^d$ then

2. Why is it useful?

   Gibbs measures have a very important conditional independence property: they are *Markov random fields* and satisfy:

   $$P\left(\sigma_v = \tau_v | \{\sigma_u = \tau_u\}_{u \in V - v}\right) = P\left(\sigma_v = \tau_v | \{\sigma_u = \tau_u\}_{u \in N(v)}\right) .$$

   Equivalently, suppose we partition $V = A \cup B \cup C$ so that there are no edges between A and C. Then if we condition on the spins in B, the spins in A are independent of the spins in C.

   Note that such a property is not true in other natural models of a random independent set, such as choosing a random independent set of size k in G uniformly at random.

3. Is it an 'optimal' distribution in some sense?

   Yes! Say we have a finite set Σ of configurations and a function $H : \Sigma \to \mathbb{R} \cup \{+\infty\}$. Consider the set $\mathcal{P}_B$ of all probability distributions μ on Σ so that $\mathbb{E}_\mu H = B$ where $\min_{\sigma \in \Sigma} H(\sigma) \leq B \leq \max_{\sigma \in \Sigma}$. Then the distribution $\mu_* \in \mathcal{P}_B$ that maximizes the Shannon entropy has the form $\mu_*(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z(\beta)}$ with $Z(\beta) = \sum e^{-\beta H(\sigma)}$. That is, it is a Gibbs distribution.

   For example, the Ising model is the probability distribution on cuts of G that maximizes entropy subject to a given mean number of edges cut. The hard-core model is the distribution on independent sets of G with a given mean size that maximizes entropy.

In combinatorics we are very familiar with the benefits of studying maximum entropy distributions.

Exercise 1. *Let $\mathcal{G}_n$ be the set of all graphs on n vertices. What is the maximum entropy distribution on $\mathcal{G}_n$ with mean number of edges m?*

1.4 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 again for intuition 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.

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 $[q]$).

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 distance between vertices. A natural way to measure the correlation between the spins at vertices u and v is the 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.* One can also define truncated k-point correlation functions.

1.4.1 Decay of correlations

We say μ_G exhibits *exponential decay of correlations* if there exist constants $a, b > 0$ so that for all $u, v \in V$,

$$|\kappa(u, v)| = |\mu_{uv} - \mu_u \mu_v| \leq a e^{-b \cdot \mathrm{dist}(u,v)} \, ,$$

where $\mathrm{dist}(\cdot, \cdot)$ is the graph distance in G. This definition really pertains to an infinite sequence of graphs G_n (or an infinite graph like $\mathbb{Z}^d$) and in this case a and b should be independent of n.

If $\kappa(u, v) \approx e^{-b \cdot \mathrm{dist}(u,v)}$ then we call $1/b$ the *correlation length* of the model: a measure of how far correlations persist. If spins are independent then the correlation length is 0, while if there is long-range order, $|\kappa(u, v)|$ bounded away from 0 independent of the distance, then the correlation length diverges to ∞.

1.5 Phase transitions

There are at least three different but related notions of phase transition in statistical physics. In many situations the three definitions are equivalent.

A phase transition only occurs in the *infinite volume limit*. Let $\Lambda_n \subset \mathbb{Z}^d$ be the box of sidelength n, and let $|\Lambda_n|$ be the number of its vertices. We consider the Gibbs measure and partition function on Λ_n with *boundary conditions*: for vertices on the boundary, we may specify their spins (or leave them 'free'). For instance we may take the all even boundary conditions for the hard-core model: all vertices on the boundary whose sum of coordinates are even are specified to be in the independent set. Under very general conditions the following are true:

1. There is a subsequential weak limit of the Gibbs measures μ_{Λ_n} as $n \to \infty$. Such a limiting measure μ_∞ is an infinite-volume Gibbs measure.

2. The limit
$$f(\beta) = \lim_{n\to\infty} \frac{1}{|\Lambda_n|} \log Z_{\Lambda_n}(\beta)$$
exists and is independent of the sequence of boundary conditions.

   The function $f(\beta)$ is called the infinite volume *pressure* or *free energy*.

1. Disorder vs long-range order.

   A phase transition occurs at β_c if for $\beta < \beta_c$ the model exhibits exponential decay of correlations while for $\beta > \beta_c$ long-range correlations persist (the correlation length diverges).

2. Uniqueness vs non-uniqueness of the infinite volume Gibbs measure.

   A phase transition occurs at β_c if for $\beta < \beta_c$ there is a unique infinite volume Gibbs measure, while for $\beta > \beta_c$ there are multiple infinite volume Gibbs measures. That is, for $\beta < \beta_c$ the effect of the boundary conditions vanishes in the limit while for $\beta > \beta_c$ the effect of boundary conditions persists.

   Often there are extremal boundary conditions: even/odd occupied for hard-core, monochromatic boundary conditions for Ising/Potts. Then we can ask does the choice of extremal boundary conditions affect the marginal of the origin as $n \to \infty$.

3. Analyticity vs non-analyticity of the infinite volume pressure.

   A phase transition occurs at β_c if the function $f(\beta)$ is non-analytic at β_c. A phase transition is first-order if $f'(\beta)$ is discontinuous at β_c and second-order if $f''(\beta)$ is discontinuous at β_c.

   Analyticity of $f(\beta)$ is closely related to the zeroes of $Z_{\Lambda_n}(\beta)$ in the complex plane. As a function $Z_{\Lambda_n}(\beta)$ is a polynomial in $e^{-\beta}$ with positive coefficients and so has no zeros on the positive real axis. If there is a region in the complex plane containing $\beta_0 > 0$ for which $Z_{\Lambda_n} =\neq 0$ for all n, then f must be analytic at β_0 and thus no phase transitions occurs. A phase transition occurs when zeros of Z_{Λ_n} in the complex plane condense as $n \to \infty$ onto a positive β_c. This perspective is called the Lee-Yang theory of phase transitions [80]

1.6 Translation to combinatorics

Here's a basic glossary of objects and concepts in statistical physics with their counterparts in combinatorics.

Statistical physics	Combinatorics
ground state	extremal object
partition function	(weighted) number of objects
Gibbs measure	random object
free energy (pressure)	exponential growth rate of the number of objects
zero-temperature	extremal objects
low-temperature	stability

Take, for example, Mantel's Theorem: the triangle-free graph on n vertices with the most edges is a complete bipartite graph with a balanced bipartition. Classifying the extremal examples is the task of understanding the ground states. Asking 'how many triangle-free graphs are there?' is the counting problem: computing or approximating the partition function. 'What does a typical triangle-free graph look like?' This is the problem of understanding the Gibbs measures and its correlations.

Classical statistical physics focuses on lattices like $\mathbb{Z}^d$ (with special emphasis on the most physically relevant cases $\mathbb{Z}^2$ and $\mathbb{Z}^3$. In particular, these graphs have a few special properties: they are regular, vertex-transitive and of polynomial growth (the number of vertices within distance t of a fixed vertex grows like t^d).

Extremal combinatorics, on the other hand, is the study of extremal, 'worst-case' graphs. Often the graphs studied in combinatorics are very different than lattices: sparse random graphs, for instance, play a leading role in probabilistic combinatorics but their neighborhoods grow exponentially. On the other hand, they are very good expanders and their local structure is particularly simple: typical local neighborhoods are trees.

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

The energy $H(\cdot)$ is a local function: 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. For instance in the hard-core model it counts the size of an independent set while in the Ising and Potts models it counts the number of monochromatic edges (or equivalently the number of crossing edges of a cut).

Understanding the random variable $H(\sigma)$ in the limit $\Lambda_n \to \mathbb{Z}^d$ can tell us a lot about the behavior of the model and any phase transitions that might occur as parameters are varied.

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

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

coefficients in the Taylor series:

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

. 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.

Recall that the partition function looks similar to a moment generating function:

$$Z = \sum_{\sigma} e^{-\beta H(\sigma)} .$$

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

$$\begin{aligned}
\frac{d}{d\beta} \log Z(\beta) &= \frac{\frac{d}{d\beta} Z(\beta)}{Z(\beta)} \\
&= - \frac{\sum_{\sigma \in \Omega^V} H(\sigma) e^{-\beta H(\sigma)}}{Z(\beta)} \\
&= - \sum_{\sigma \in \Omega^V} H(\sigma) \mu(\sigma) \\
&= - \mathbb{E} H(\sigma) \\
&= - \kappa_1(H) .
\end{aligned}$$

The second derivative is

$$\begin{aligned}
\frac{d^2}{d\beta^2} \log Z(\beta) &= \frac{\frac{d^2}{d\beta^2} Z(\beta)}{Z(\beta)} - \left(\frac{\frac{d}{d\beta} Z(\beta)}{Z(\beta)} \right)^2 \\
&= \mathbb{E}[H(\sigma)^2] - (\mathbb{E} H(\sigma)^2 \\
&= \mathrm{var}(H(\sigma)) \\
&= \kappa_2(G) .
\end{aligned}$$

The higher derivatives recover the cumulants of the energy:

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

1.7.1 Multivariate partition functions

To study correlations via the partition function we need to add variables to the partition function to distinguish individual vertices. We add non-uniform *external fields* for every

vertex. Consider the following partition function of a two-spin model with non-uniform external fields given by the vector $\boldsymbol{\alpha}$:

$$Z_G(\boldsymbol{\alpha}) = \sum_{\sigma \in \Omega^V} e^{\sum_{v \in V} \alpha_v \sigma_v} \cdot e^{-\beta H(\sigma)} .$$

Then we can look at the partial derivatives of $\log Z_G$ with respect the variables α_v.

$$\begin{aligned}
\frac{\partial}{\partial \alpha_v} \log Z_G(\boldsymbol{\alpha}) &= \frac{\frac{\partial}{\partial \alpha_v} Z_G}{Z_G} \\
&= \sum_{\sigma \in \Omega^V} \sigma_v \frac{e^{\boldsymbol{\alpha} \cdot \sigma} e^{-\beta H(\sigma)}}{Z_G} \\
&= \sum_{\sigma \in \Omega^V} \sigma_v \mu(\sigma) \\
&= \mathbb{E}[\sigma_v] \\
&= \mu_v ,
\end{aligned}$$

so we have recovered the marginal of v by taking a partial derivative.

We can now take mixed partial derivatives with respect to α_u, α_v:

$$\begin{aligned}
\frac{\partial^2}{\partial \alpha_u \partial \alpha_v} \log Z_G(\boldsymbol{\alpha}) = & \\
&= \mathbb{E}[\sigma_u \sigma_v] - \mathbb{E}[\sigma_u]\mathbb{E}[\sigma_v] \\
&= \mu_{uv} - \mu_u \mu_v \\
&= \kappa(u, v) .
\end{aligned}$$

In fact we can obtain the joint cumulants of any collection of spins by taking partial derivatives. The truncated k-point correlation functions are the joint cumulants of k spin variables.

For more on joint cumulants in the setting of the Ising model at low temperature, see [26].

An important special case of the use of non-uniform external fields is the multivariate hard-core model. The is a probability distribution over independent sets of G in which each vertex has its own fugacity λ_v. The partition function is:

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

This is a multilinear polynomial in n variables. Not only can we use it to study correlations in the hard-core model, but taking the multivariate perspective is also the natural setting of some analytic techniques for understanding complex zeros of the partition function (e.g. [70, 59, 55]).

Note that since we have written the hard-core partition function as a polynomial (univariate or multivariate) we have to adjust the formulas for the cumulants and joint cumulants slightly. For instance, the expected size of an independent set drawn from the hard-core model on G at fugacity λ is

$$\mathbb{E}[|\mathbf{I}|] = \lambda \cdot (\log Z_G(\lambda))' = \frac{\lambda Z_G'(\lambda)}{Z_G(\lambda)} = \frac{\sum_I |I| \lambda^{|I|}}{Z_G(\lambda)} . \tag{1}$$

1.8 Basic tools and tricks

If G_1 and G_2 are disjoint graphs then $Z_{G_1 \cup G_2} = Z_{G_1} Z_{G_2}$. If u and v are in different connected components of G then σ_u and σ_v are independent and $\mu_{uv} = \mu_u \mu_v$.

The following identity is often useful. For any $v \in V$,

$$Z_G(\lambda) = \lambda Z_{G - \overline{N}(v)}(\lambda) + Z_{G-v}(\lambda)\,, \tag{2}$$

where $\overline{N}(v) = \{v\} \cup N(v)$. We can use this to write the marginal

$$\mu_v = \frac{\lambda Z_{G - \overline{N}(v)}(\lambda)}{Z_G(\lambda)}\,. \tag{3}$$

1.9 Summary

- The basic objects in statistical physics are Gibbs measures and partition functions. Statistical physicists are interested in the correlation properties of Gibbs measures in the infinite volume limit on graphs like $\mathbb{Z}^d$.
- The inverse temperature parameter interpolates from independence to optimization
- The form of a Gibbs measure (probability proportional to exponential of an energy, or 'log linear') is physically motivated and provides some very useful properties including conditional independence and the ability to write statistics as derivatives of the log partition function.
- The cumulants of the energy can be obtained by taking derivatives of the log partition function in β. By putting external fields on all vertices, we can obtain the joint cumulants of any set of spins by taking partial derivatives of the log partition function with respect to these external fields.
- Many ideas, themes, questions, and objects in combinatorics have counterparts in statistical physics; knowing a little of the terminology will allow you to move between the two fields.

To read more on the basics of statistical physics, see the recent textbook of Friedli and Vilenik [28]. For many classical and foundational results (including for continuum models), see the classic text of Ruelle [64]. For a computational perspective on statistical physics models and random graphs, see the textbook of Mezard and Montanari [54].

1.10 Exercises

1. Compute $Z_{K_d}(\lambda)$. For $u, v \in K_d$ compute the truncated two-point correlation function.
2. Prove that the following distribution on independent sets of G is the hard-core model at fugacity λ. Pick a subset S by including each vertex independently with probability $\frac{\lambda}{1+\lambda}$ and condition on the event that S is an independent set.

3. Consider the hard-core model on a graph G of maximum degree Δ. Fix a vertex v. Prove that
$$\frac{\lambda}{(1+\lambda)^{\Delta+1}} \leq \mu_v \leq \frac{\lambda}{1+\lambda}\,.$$
Show that the upper bound is tight. Is the lower bound tight? If not, can you prove a tight bound?

4. Let $\Lambda_n \subset \mathbb{Z}^d$ be the box of sidelength n, and let $|\Lambda_n|$ be the number of its vertices. Consider the hard-core model on Λ_n with boundary conditions (vertices on the boundary may be specific 'in' or 'out' of the independent set).
   (a) Prove that the limit $\lim_{n\to\infty} \frac{1}{|\Lambda_n|} \log Z_{\Lambda_n}(\lambda)$ exists. (Hint: look up 'subadditivity').
   (b) Show that the limit does not depend on the boundary conditions.

5. 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 $\log Z_G(\lambda)$.

6. Let P_n be the path on n vertices.
   (a) Write a recursion for the independence polynomial $Z_{P_n}(\lambda)$.
   (b) Solve the recursion to compute the limit $f(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log Z_{P_n}(\lambda)$.
   (c) What can you deduce about phase transitions in the hard-core model on $\mathbb{Z}^1$ from the function $f(\lambda)$?

7. Consider the hard-core model on a bipartite graph G with bipartition (A, B). Prove (by induction?) that if $u, v \in A$ then $\kappa(u, v) = \mu_{uv} - \mu_u\mu_v \geq 0$. (Hint: look up the FKG inequality; it is also possible to prove without using FKG). When does equality hold?

2 Extremal combinatorics of sparse graphs

The field of extremal combinatorics asks for the maximum and minimum of various graph parameters over different classes of graphs. Some examples of classic theorems from extremal combinatorics are Mantel's Theorem mentioned above, or Dirac's Theorem: which graph on n vertices containing no Hamilton cycle has the largest minimum degree?

Here we focus on extremal results for bounded-degree graphs. We first mention three classic results in this area, then we discuss how taking the point of view of statistical physics and correlations allows us to reprove, strengthen, or generalize these results. For a nice overview of results, techniques, and open questions in the area, see the survey of Zhao [82] and [66] for very recent results.

We will combine the statistical physics and combinatorics perspectives: like statistical physicists we will be interested in correlations, but we will ask *extremal* questions about correlations. For a given class of graphs, when do spins have the strongest positive correlation? The strongest negative correlation? The least correlation?

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 the first approximate answer was given by Noga Alon [2] 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 [46]). *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: 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 [81] 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 [65] for an extension to irregular graphs.

The question of minimizing the number of (weighted) independent sets in a d-regular graph is somewhat simpler: the answer it the clique K_{d+1}, proved by Cutler and Radcliffe [19]; for a short proof see [24].

Independent sets in triangle-free graphs

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 [1], 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 so there is a factor of 2 that could potentially be gained in Shearer's bound. The factor of 2 would immediately give a factor 2 improvement to the bound on $R(3, k)$.

Matchings and perfect matchings

A third 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.

A special case of Bregman's theorem can be stated as an extremal result for d-regular graphs. Let $\mathrm{pm}(G)$ denote the number of perfect matchings of a graph G.

Theorem 2.5 (Bregman). *For all d-regular graphs G,*

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

2.1 The occupancy fraction of the hard-core model

In this section we present a statistical physics based approach to proving extremal theorems for sparse graphs. We will prove extremal results for partition functions and graph polynomials by optimizing the derivative of the log partition function over a given class of graphs. By integrating the resulting bound we obtain a corresponding result for the partition function. As we saw in Lecture 1, the logarithmic derivative has a probabilistic interpretation as the expectation of a locally computable observable of the relevant model.

We start with independent sets and the hard-core model, where the relevant observable is the expected size of an independent set drawn from the model. It will be more convenient for us to divide this by the number of vertices and study the expected independent set density, or the *occupancy fraction*, $\overline{\alpha}_G(\lambda)$:

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

We begin by collecting some basic facts about the occupancy fraction, following our discussion above about the cumulants and logarithmic derivatives of the log partition function.

Lemma 2.6. *The occupancy fraction is λ times the derivative of the free energy:*

$$\overline{\alpha}_G(\lambda) = \lambda \cdot \left(\frac{1}{|V(G)|}\log Z_G(\lambda)\right)' .$$

Lemma 2.7. *The occupancy fraction $\overline{\alpha}_G(\lambda)$ is a strictly increasing function of λ.*

This follows since the second derivative of $\log Z_G$ is, up to scaling, the variance of $|\mathbf{I}|$ which is strictly positive.

The occupancy fraction captures quite a lot of combinatorial information:

- $\overline{\alpha}_G(1)$ is the average size of a (uniformly) random independent set from G.
- $\lim_{\lambda\to\infty}\overline{\alpha}_G(\lambda) = \frac{\alpha(G)}{n}$, the scaled size of the largest independent set in G.
- Since $\overline{\alpha}_G(\lambda)$ is the scaled derivative of $\log Z_G(\lambda)$, we can compute the partition function (or the number of independent sets) of G:

$$\frac{1}{|V(G)|}\log Z_G(\lambda) = \int_0^\lambda \frac{\overline{\alpha}_G(t)}{t}\,dt\,.$$

  In particular if we can prove upper or lower bounds on the occupancy fraction, then by integrating we obtain upper and lower bounds on the partition function (and the number of independent sets).

What is particularly nice about working with the occupancy fraction (or any other observable) is that we can argue about it locally.

In trying to understand correlations between spins in the hard-core model, we can use an idea that has appeared both in combinatorics and computer science (e.g. [46, 27]): instead

of considering correlations between spins (occupancies) we consider correlations between the events that different vertices are allowed to be in the independent set – not blocked by another vertex.

We say v is *uncovered* with respect to an independent set I if $N(v) \cap I = \emptyset$.

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

The follows from the spatial independence property of a Gibbs measure. If $N(v) \cap I = \emptyset$, then v can be either in or out; in the first case it contributes a factor λ in the second case a factor 1.

Fact 2 If G is triangle-free, then $\Pr[v \text{ uncovered} | v \text{ has } j \text{ uncovered neighbors}] = (1+\lambda)^{-j}$.

To prove Fact 2 note that the graph induced by the uncovered neighbors of v consists of isolated vertices since G is triangle free.

Now 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

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

Now consider the following two-part experiment: pick $\mathbf{I}$ from the hard-core model on G and independently choose $\mathbf{v}$ uniformly at random from $V(G)$. Let $\mathbf{Y}$ be 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 $\mathbf{Y}$.

$$\begin{aligned}
\overline{\alpha}_G(\lambda) &= \frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda} (1+\lambda)^{-\mathbf{Y}} \\
\overline{\alpha}_G(\lambda) &= \frac{1}{d} \frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda} \mathbf{Y} \,.
\end{aligned}$$

Thus the identity

$$\mathbb{E}_{G,\lambda} (1+\lambda)^{-\mathbf{Y}} = \frac{1}{d} \mathbb{E}_{G,\lambda} \mathbf{Y} \tag{4}$$

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

We can use this observation to prove a strengthening of Theorem 2.2.

Theorem 2.8 (Davies, Jenssen, Perkins, Roberts [22])**.** *For any d-regular graph G, and any $\lambda > 0$,*

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

Proof of Theorem 2.8. We prove this first for triangle-free G to illustrate the method.

Now the idea is to *relax* the maximization problem; instead of maximizing $\overline{\alpha}_G(\lambda)$ over all d-regular 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 (4).

It is not too hard to 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 (4), 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,

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

Now we give the full proof for graphs that may contain triangles.

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 over the randomness in our two-part experiment.

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

$$\overline{\alpha}_G(\lambda) = \frac{\lambda}{1+\lambda}\Pr_{G,\lambda}[\mathbf{v} \text{ uncovered}] = \frac{\lambda}{1+\lambda}\mathbb{E}_{G,\lambda}\left[\frac{1}{Z_{\mathbf{H}}(\lambda)}\right] \tag{5}$$

$$\overline{\alpha}_G(\lambda) = \frac{1}{d}\mathbb{E}_{G,\lambda}[\mathbf{I} \cap N(\mathbf{v})] = \frac{\lambda}{d}\mathbb{E}_{G,\lambda}\left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)}\right], \tag{6}$$

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

$$\frac{\lambda}{1+\lambda}\mathbb{E}_{G,\lambda}\left[\frac{1}{Z_{\mathbf{H}}(\lambda)}\right] = \frac{\lambda}{d}\mathbb{E}_{G,\lambda}\left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)}\right]. \tag{7}$$

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 (7).

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.

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 insure that the variables $p(H)$ form a probability distribution; the last is constraint (7).

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:

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

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:

$$\frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d-1} + \frac{(1+\lambda)^d-1}{2(1+\lambda)^d-1}\left[\frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)} - \frac{\lambda}{d}\frac{Z_H'(\lambda)}{Z_H(\lambda)}\right] \geq \frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)}\,. \tag{8}$$

Multiplying through by $Z_H(\lambda)\cdot(2(1+\lambda)^d-1)$ and simplifying, (8) reduces to

$$\frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d-1} \geq \frac{\lambda Z_H'(\lambda)}{Z_H(\lambda)-1}\,, \tag{9}$$

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 (9) has a nice probabilistic interpretation: the RHS is simply $\mathbb{E}_{H,\lambda}\left[|\mathbf{I}|\big||\mathbf{I}| \geq 1\right]$, 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. Proving (9) is left for the exercises, and this completes the proof. □

2.2 Minimizing independent sets for triangle-free graphs

Instead of asking for the strongest positive correlations, we can ask for the strongest *negative* correlations. Or, in other words, we can try to minimize the occupancy fraction given our identity (for triangle-free graphs) $\mathbb{E}_{G,\lambda}(1+\lambda)^{-\mathbf{Y}} = \frac{1}{d}\mathbb{E}_{G,\lambda}\mathbf{Y}$.

Theorem 2.9 (Davies, Jenssen, Perkins, Roberts [23])**.** *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.

Proof. We now return to the identity (4) for triangle-free graphs. We remarked that to maximize $\mathbb{E}\mathbf{Y}$ given the constraint $\mathbb{E}(1+\lambda)^{-\mathbf{Y}} = \frac{1}{d}\mathbb{E}\mathbf{Y}$ and $0 \leq \mathbf{Y} \leq d$, we should take $\mathbf{Y}$ to be supported on the two extreme values, 0 and d.

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:

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

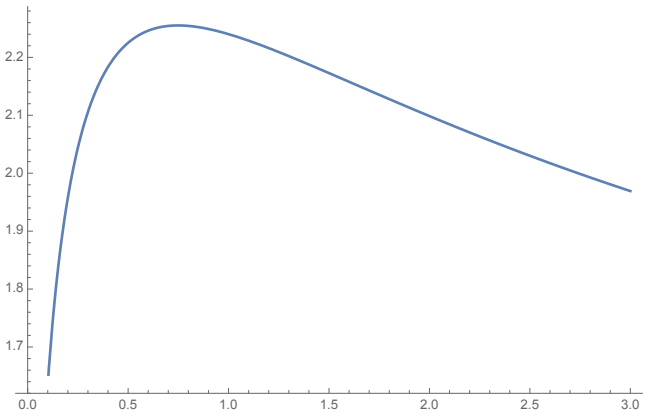


Figure 4: $\frac{\lambda}{1+\lambda} y^*$ as a function of λ with $d = 100$.

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

The solution is

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

where $W(\cdot)$ is the W-Lambert function. This gives

$$\overline{\alpha}_G(\lambda) \geq \frac{1}{d} \frac{\lambda}{1+\lambda} \frac{W(d \log(1+\lambda))}{\log(1+\lambda)} \,. \tag{10}$$

Now although $\overline{\alpha}_G(\lambda)$ is monotone increasing in λ, somewhat surprisingly the bound (10) is not monotone in λ (see Figure 4 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 $\varepsilon > 0$.

We set $\lambda = 1/\log d$ and derive a bound asymptotically in d. We show in the exercises that the Lambert 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 (10) for $\lambda = 0$ to 1 to obtain a lower bound on the partition function.

$$\frac{1}{n} \log i(G) = \frac{1}{n} \log Z_G(1) = \int_0^1 \frac{\overline{\alpha}_G(t)}{t} \, dt$$

$$
\begin{aligned}
&\geq \int_0^1 \frac{1}{d}\frac{1}{1+t}\frac{W(d\log(1+t))}{\log(1+t)}\,dt \quad \text{from (10)} \\
&= \frac{1}{d}\int_0^{W(d\log 2)} 1+u\,du \quad \text{using the substitution } u = W(d\log(1+t)) \\
&= \frac{1}{d}\left[W(d\log 2) + \frac{1}{2}W(d\log 2)^2\right] \\
&= \left(\frac{1}{2}+o_d(1)\right)\frac{\log^2 d}{d}\,.
\end{aligned}
$$

□

Using a similar argument to the proof of the $R(3,k)$ upper bound, we can use Theorem 2.9 to give a lower bound on the number of independent sets in a triangle-free graph without degree restrictions.

Corollary 2.10. *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 2.9. 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

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

□

2.2.1 Max vs. average independent set size?

Theorem 2.9 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 2.11. *Can we use Theorem 2.9 to improve the current asymptotic upper bound on* $R(3,k)$.

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

Conjecture 2.12 ([23])**.** *For any triangle-free graph* G*, we have*

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

Conjecture 2.13 ([23])**.** *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 2.14 ([23])**.** *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 2.12 would imply a factor $4/3$ improvement on the current upper bound for $R(3,k)$, while Conjectures 2.13 and 2.14 would both imply a factor 2 improvement.

2.3 Summary

- Observables – expectations of locally computable statistics of a Gibbs measure – can be calculated by taking derivatives of log partition functions. This means that extremal bounds on observables over a class of graphs imply extremal bounds on partition functions.
- The occupancy fraction of the hard-core model is one such observable and it encodes both the independence number of a graph and the number of its independent sets.
- By using the properties of a Markov random field we wrote an identity for the occupancy fraction of any d-regular triangle-free graph in terms of the number of uncovered neighbors of a randomly chosen vertex when choosing a random independent set from μ_G. We showed that maximizing and minimizing the occupancy fraction subject to the constraint imposed by the identity yields two theorems, one on the number of independent sets and the other on the average size of an independent set.
- This method can be generalized to other partition functions (for matchings, colorings, the Potts model, etc.) and for other classes of graphs (graphs of a given minimum girth for instance).

2.4 Exercises

1. Let $i_k(G)$ denote the number of independent sets of size k in G (these are the coefficients of the hard-core partition function).

(a) Give a probabilistic interpretation of the quantity $F_G(k) = \frac{(k+1)i_{k+1}(G)}{i_k(G)}$ (it is the expectation of...)

(b) Which d-regular graph minimizes the quantity $\frac{1}{n}F_G(k)$ for all k?

(c) What can you conclude about the coefficients $i_k(G)$ and the partition function $Z_G(\lambda)$ for d-regular graphs from the above?

(d) Does some d-regular graph maximize the quantity $\frac{1}{n}F_G(k)$ for all k? (this is an open problem!)

2. Deduce Corollary 2.4 from Theorem 2.3 (or Theorem 2.9).

3. Let G be a d-regular graph (not necessarily triangle-free). Pick a random vertex v from G and pick I according to the hard-core model. For $k = 0, \ldots, d$ let $p_k = P(|\mathbf{I} \cap N(v)| = k)$, the probability that v has exactly k occupied neighbors.

(a) Write an expression for the occupancy fraction $\overline{\alpha}_G(\lambda)$ in terms of p_0 and λ.

(b) Write an expression for the occupancy fraction $\overline{\alpha}_G(\lambda)$ in terms of $p_1, \ldots, p_d$, d, and λ.

(c) Write a lower bound for p_{k-1} in terms of k, p_k, d, and λ. Is this bound tight for some d-regular graph G?

(d) Maximize the occupancy fraction subject to the one equality constraint and $d-1$ inequality constraints given above. What can you conclude about $\alpha_G(\lambda)$ for d-regular graphs?

(e) (Open problem) Can you used this proof strategy to prove any new results?

3 Expansion methods and enumeration

An important class of tools in statistical physics are *expansion methods*. The basic idea of expansion methods is to study a system as a perturbation of an easy to understand system. Expansion methods provide a systematic way to understand corrections in approximating a complicated system by a simple system.

In this lecture we will focus on expansion methods for understanding the hard-core model at small fugacities λ. In the following lecture we will show that a wide variety of combinatorial problems can be mapped to multivariate hard-core models and so are amenable to the expansion techniques discussed here.

3.1 The cluster expansion

The *cluster expansion* (or Mayer expansion [52]) is a fundamental tool in statistical physics for understanding systems in the regime of weak interactions. The basic idea is to represent the logarithm of a partition function in terms of an infinite series whose terms measure deviations from a system of non-interacting particles.

Consider the multivariate hard-core model with fugacities $\{\lambda_v\}_{v \in V}$ and partition function

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

The *cluster expansion* is the multivariate Taylor series for $\log Z_G$ in the variables $\{\lambda_v\}_{v \in V}$ around $\vec{0}$. The terms of the cluster expansion admit a convenient combinatorial description as a sum over connected objects called *clusters*.

A *cluster* is an ordered tuple of vertices of G whose induced graph is connected. For instance in the graph C_4, with vertices labeled v_1, v_2, v_2, v_4 in cyclic order, (v_1, v_1, v_2) is a cluster, (v_2) is a cluster, (v_4, v_3, v_1) is a cluster, but (v_1, v_1, v_3) is not a cluster since its induced graph is not connected. The size $|\Gamma|$ of a cluster Γ is the length of the tuple.

The *Ursell function* $\phi(H)$ of a graph H is

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

The Ursell function is an evaluation of the Tutte polynomial (scaled by the factor $1/|V(H)|!$).

For a cluster Γ, let $H(\Gamma)$ be the graph whose vertex set is the set of vertices in Γ (with multiplicities) and with edges between vertices that are neighbors in G and between multiple copies of the same vertex. For instance if $G = C_4$ as above and $\Gamma = (v_1, v_1, v_2, v_3)$ then $H(\Gamma)$ has a triangle formed by (v_1, v_1, v_2) with v_2 attached to the 4th vertex v_3.

The *cluster expansion* is the formal power series

$$\log Z_G(\boldsymbol{\lambda}) = \sum_{\text{clusters } \Gamma} \phi(H(\Gamma)) \prod_{v \in \Gamma} \lambda_v \,. \tag{11}$$

Derivations and applications of the cluster expansion (for both discrete and continuous systems) can be found in [64, 75, 70, 28] among others.

As a simple example take the graph consisting of a single vertex v with fugacity $\lambda_v = \lambda$. Then for $k \geq 1$ there is a single cluster Γ_k of size k consisting of k copies of v. The graph $H(\Gamma_k)$ is the clique on k vertices, and the Ursell function is $\phi(K_k) = \frac{(-1)^{k+1}}{k}$. This gives (as a formal power series)

$$\log Z = \sum_{k \geq 1} \frac{(-1)^{k+1} \lambda^k}{k} \,,$$

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

3.2 Convergence criteria

For the series (11) to be useful, we need to know that it converges, and if so, how fast. The following criteria of Kotecký and Preiss is easy to check and very versatile.

Theorem 3.1 (Kotecký-Preiss [49])**.** *Consider the multivariate hard-core model on a graph G with (possibly complex) fugacities $\lambda_v, v \in V$. Suppose there are functions $a(v) \geq 0$, $b(v) \geq 0$ so that for all $v \in V$,*

$$\sum_{u \in N(v) \cup \{v\}} |\lambda_u| e^{a(u)+b(u)} \leq a(v) \,. \tag{12}$$

Then the following hold:

1. *$Z_G(\boldsymbol{\lambda}) \neq 0$.*
2. *The cluster expansion for $\log Z_G(\boldsymbol{\lambda})$ converges absolutely.*
3. *The following tail bound holds. Let $b(\Gamma) = \sum_{v \in \Gamma} b(v)$. Then for all $t \geq 0$,*

$$\sum_{\substack{\text{clusters } \Gamma \\ b(\Gamma) \geq t}} \left| \phi(H(\Gamma)) \prod_{v \in \Gamma} \lambda_v \right| \leq e^{-t} \sum_{v \in V} a(v) \,. \tag{13}$$

4. *The following 'pinned' bound holds for all $v \in V$. We say $v \in \overline{N}(\Gamma)$ if $\operatorname{dist}(v, \Gamma) \leq 1$.*

$$\sum_{\substack{\text{clusters } \Gamma \\ v \in \overline{N}(\Gamma) \\ b(\Gamma) \geq t}} \left| \phi(H(\Gamma)) \prod_{u \in \Gamma} \lambda_u \right| \leq e^{-t} a(v) \,. \tag{14}$$

Conclusion 3 follows from 4 by summing over v. Conclusion 2 follows from 3 by taking $t = 0$. Conclusion 1 then follows from 2.

Theorem 3.1 is particularly useful for non-uniform activities. In the next lecture we will see a general multivariate hard-core model in which vertices come with a notion of 'size'; the functions $a(\cdot)$ and $b(\cdot)$ will often be taken proportional to size and the theorem will apply when activities decay sufficiently fast in the size.

In the case of bounded-degree graphs with uniform activities, Shearer gave a tight bound for the convergence of the cluster expansion.

Theorem 3.2 (Shearer [72])**.** *Let G have maximum degree Δ and suppose $|\lambda| \leq \frac{(\Delta-1)^{\Delta-1}}{\Delta^{\Delta}}$. Then $Z_G(\lambda) \neq 0$ and the cluster expansion converges absolutely.*

For a wide-ranging and detailed discussion of this result, extensions, and consequences, see the paper of Scott and Sokal [70]. In particular, Theorem 3.2 gives a tight bound on the probabilities of 'bad events' with a Δ-regular dependency graph for which the conclusion of the Lovás Local Lemma holds. Scott and Sokal generalize this and show a remarkable connection between the Local Lemma and cluster expansion: the conclusion of the Local Lemma holds for a dependency graph G and probabilities p_v if and only if $Z_G(\boldsymbol{p}) \neq 0$ for $\boldsymbol{p}$ in a polydisc with radii p_v.

3.2.1 Singularities on the negative real axis

Both Theorem 3.1 and 3.2 give conditions for the non-vanishing of $Z_G(\boldsymbol{\lambda})$ in a polydisc around the origin in $\mathbb{C}$. Scott and Sokal [70, Theorem 2.10] (see also Groeneveld [36]) show that for the (multivariate) hard-core model, the closest zero of Z_G to the origin is on the negative real axis (and this is the zero that controls the applicability of the Local Lemma).

On the other hand, for statistical physics and enumeration we care about positive fugacities (a singularity on the negative real axis does not mark a phase transition). New techniques for proving absence of phase transition and for approximating Z algorithmically that make essential use of positive fugacities have been developed in computer science and applied to statistical physics models (e.g. [79, 62, 59, 55]).

Question 3.3. *Are there any applications of the method of Weitz [79] to enumeration problems in combinatorics?*

3.3 Consequences of convergence

A convergent cluster expansion gives a series expansion for $\log Z$ but by combining the calculations in Section 1.7 with the cluster expansion we can obtain many probabilistic properties of the model as well.

In this section we follow Dobrushin [26] (see also [16] for similar calculations).

We first introduce auxiliary variables. Let

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

so that $Z_G(\lambda) = Z_G(\lambda, \vec{0})$. We can also write

$$\begin{aligned} \mathbb{E}_{G,\lambda}\left[\prod_{v \in \mathbf{I}} e^{t_v}\right] &= \sum_{I \in \mathcal{I}} \frac{\lambda^{|I|}}{Z_G(\lambda)} \prod_{v \in I} e^{t_v} \\ &= \frac{Z_G(\lambda, \vec{t})}{Z_G(\lambda)}, \end{aligned}$$

and so the joint cumulant generating function of the spins σ_v (indicators that $v \in I$) is given by

$$K(\mathbf{t}) = \log Z_G(\lambda, \mathbf{t}) - \log Z_G(\lambda)\,. \tag{15}$$

Similarly the cumulant generating function of $|\mathbf{I}|$ is given by

$$K(t) = \log Z_G(\lambda, t) - \log Z_G(\lambda) \tag{16}$$

where

$$Z_G(\lambda, t) = \sum_{I \in \mathcal{I}(G)} \lambda^{|I|} e^{t|I|}\,.$$

3.3.1 Cumulants

The cluster expansion of $\log Z_G(\lambda, t)$ is

$$\log Z_G(\lambda, t) = \sum_{\text{clusters } \Gamma} \phi(H(\Gamma))\lambda^{|\Gamma|} e^{t|\Gamma|}\,.$$

Differentiating (16) and evaluating at $t = 0$ then gives

$$\kappa_1(|\mathbf{I}|) = \sum_{\text{clusters } \Gamma} \phi(H(\Gamma))|\Gamma|\lambda^{|\Gamma|}\,,$$

and in general

$$\kappa_k(|\mathbf{I}|) = \sum_{\text{clusters } \Gamma} \phi(H(\Gamma))|\Gamma|^k\lambda^{|\Gamma|}\,.$$

3.3.2 Large deviations

Using the cluster expansion we can bound the moment generating function and apply exponential Markov's inequality to bound the probability of a large deviation in $|\mathbf{I}|$.

3.3.3 Joint cumulants and correlation decay

Finally we can obtain the truncated k-point correlation functions by taking partial derivatives of the cluster expansion:

$$\log Z_G(\lambda, \mathbf{t}) = \sum_{\text{clusters } \Gamma} \phi(H(\Gamma))\lambda^{|\Gamma|} \prod_{v \in \Gamma} e^{t_v}\,.$$

Then we have

$$\mu_v = \kappa(v) = \frac{\partial}{\partial t_v} \log Z_G(\lambda, \mathbf{t})\big|_{\mathbf{t}=0} = \sum_{\Gamma \ni v} \phi(H(\Gamma))\lambda^{|\Gamma|} m_v(\Gamma)$$

where $m_v(\Gamma)$ is the multiplicity of v in Γ.

The truncated two-point correlation function is

$$\kappa(u,v) = \frac{\partial^2}{\partial t_u \partial t_v} \log Z_G(\lambda, \mathbf{t})\big|_{\mathbf{t}=0} = \sum_{\Gamma \ni u,v} \phi(H(\Gamma))\lambda^{|\Gamma|} m_v(\Gamma) m_u(\Gamma) \tag{17}$$

This allows us to prove exponential decay of correlations.

Lemma 3.4. *Suppose the Kotecký–Preiss condition (Theorem 3.1) holds with $a(v) = a > 0$ and $b(v) = b > 0$ for all v. Let $\mathrm{dist}(\cdot,\cdot)$ denote the graph distance in G. Then there exists $C = C(a,b) > 0$ so that for all $u, v \in V(G)$,*

$$|\kappa(u,v)| \le C e^{-b \cdot \mathrm{dist}(u,v)}\,. \tag{18}$$

Proof. Since a cluster Γ is a tuple of vertices whose induced graph is connected, if u and v belong to Γ, we must have $|\Gamma| \ge \mathrm{dist}(u,v) + 1$. More generally,

$$|\Gamma| \ge \mathrm{dist}(u,v) + m_u(\Gamma) + m_v(\Gamma) - 1\,.$$

Then by (17) and (23) we have

$$\begin{aligned}
|\kappa(u,v)| &\le \sum_{\Gamma \ni u,v} \left|\phi(H(\Gamma))\lambda^{|\Gamma|} m_v(\Gamma) m_u(\Gamma)\right| \\
&\le a \sum_{s,t \ge 1} e^{-b\cdot(\mathrm{dist}(u,v)+s+t-1)} \\
&= a e^b e^{-b\cdot\mathrm{dist}(u,v)} \sum_{s,t\ge 1} e^{-b(s+t)} \\
&\le C e^{-b\cdot\mathrm{dist}(u,v)}
\end{aligned}$$

for some $C = C(a,b)$. □

We can prove a similar exponentially small upper bound on truncated k-point correlation functions. The prefactor C_k will depend on k (and the dependence is exponential). The right measure of 'distance' of subset of k vertices is the the length of the minimum Steiner tree connecting them (for more see [16]).

3.4 An example

Proposition 3.5. *Let G be a Δ-regular triangle-free graph. Then for $\lambda = o(n^{-1/4})$,*

$$Z_G(\lambda) = (1 + o(1))(1+\lambda)^n \exp\left[-\frac{n\Delta}{2}\lambda^2 + \frac{n\Delta(\Delta+1))}{2}\lambda^3\right] \tag{19}$$

as $n \to \infty$.

Notice the similarity to Janson's Inequality. The quantity $\frac{Z_G(\lambda)}{(1+\lambda)^n}$ is exactly the probability that we obtain an independent set when we take a random subset of V by including each v with probability $\frac{\lambda}{1+\lambda}$ independently. Proposition 3.5 is in fact a direct consequence of a generalization of Janson's Inequality due to Mousset, Noever, Panagiotou, and Samotij [56], that seems to have some similarities to the cluster expansion (see Question 3.6).

Proof of Proposition 3.5. We apply Theorem 3.1 by taking $a(v) = 1$ and $b(v) = w(n)$ chosen so that $n^{1/4} \ll e^{w(n)} \ll \lambda^{-1}$. We compute

$$\sum_{u \in N(v) \cup \{v\}} \lambda e^{a(u)+b(u)} = \sum_{u \in N(v) \cup \{v\}} \lambda e^{1+w(n)} = \lambda(\Delta+1)e^{1+w(n)} = o(1)$$

and so for n large enough the condition (3.1) is satisfied. We can conclude that

$$\sum_{\substack{\text{clusters } \Gamma \\ |\Gamma| \geq 4}} \left| \phi(H(\Gamma))\lambda^{|\Gamma|} \right| \leq n e^{-4w(n)} = o(1)\,,$$

and so to determine the asymptotics of $Z_G(\lambda)$ we only need to consider clusters of size at most 3. We list the cluters 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 copies copies of a vertex and one 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

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

The first three terms give the asymptotics of $\log(1+\lambda)^n$ and so we have

$$Z_G(\lambda) = (1+o(1))(1+\lambda)^n \exp\left[-\frac{n\Delta}{2}\lambda^2 + \frac{n\Delta(\Delta+1))}{2}\lambda^3\right].$$

□

3.5 Summary

- The cluster expansion is an infinite series representation of $\log Z(\boldsymbol{\lambda})$, in fact the multivariate Taylor series around 0.
- Convergence of the cluster expansion has been studied extensively in both statistical physics (for its connection to proving absence of phase transitions) and combinatorics (for its connection to the Lovász Local Lemma)
- It can also be used to (approximately) enumerate in both combinatorics and computer science. Statistical physicists are happy with formulas expressed as infinite series. In combinatorics we might expect
- A convergent cluster expansion implies many nice probabilistic properties: correlation decay, large deviation bounds, Poisson and Gaussian convergence. A spin system with a convergent cluster expansion can be thought of as a generalization of a collection of independent random variables.

Question 3.6. *What is the relation between the cluster expansion and Janson's Inequality? Can the results of [56] be interpreted in the framework of the (hypergraph) cluster expansion?*

3.6 Exercises

1. Let G be a graph consisting of a single edge.
   (a) Compute $Z_G(\lambda)$.
   (b) Describe the set of clusters of G.
   (c) Write down the first few terms of the cluster expansion and then the whole series.
2. Let G be a d-regular graph on n vertices. Let $|\mathbf{I}|$ be the size of the random independent set drawn according to the hard-core model on G at fugacity λ.
   (a) Using the cluster expansion and cumulants prove that if $\lambda = c/n$ then $|\mathbf{I}|$ converges in distribution to a Poisson random variable as $n \to \infty$.
   (b) Using the cluster expansion and cumulants prove that if $n^{-1} \ll \lambda \ll n^{-1/3}$ then after centering and scaling $|\mathbf{I}|$ converges in distribution to a Gaussian random variable as $n \to \infty$. (In fact you can prove it for larger λ as well).
3. We saw in the first set of exercises that the hard-core model on $\mathbb{Z}^1$ has no phase transition. Does the cluster expansion for the hard-core model on the path of n vertices converge for all λ?
4. Let G_n and H_n be two sequences of d-regular graphs on n vertices. Suppose G_n is triangle-free, while H_n has δn triangles for some $\delta > 0$ independent of n.
   (a) Write down the cluster expansion for $\log \frac{Z_{G_n}(\lambda)}{Z_{H_n}(\lambda)}$ out to clusters of size 4.
   (b) Find some $\varepsilon = \varepsilon(d, \delta) > 0$ so that for $\lambda < \varepsilon$ and n large enough, $Z_{G_n}(\lambda) > Z_{H_n}(\lambda)$.

4 Combinatorics at low temperatures

4.1 From low temperature to high temperature

In Figure 5 we see the hard-core model on $\mathbb{Z}^2$ at two very different values of λ. For small λ, we see a *dilute system*: the occupied sites are sparse and disordered. For large λ we see the opposite picture: the occupied sites are essentially frozen, with occupied odd vertices (colored blue) or occupied even vertices (colored red) engulfing the system. We see clear evidence of long-range order.

However, if we focus only on the *defects*: the red (even) vertices in Figure 5(b) or the blue (odd) vertices in Figure 5(c) the the picture (at least superficially) looks like the small λ dilute case. While the spins themselves exhibit long-range order, the defects are sparse and disordered. Putting in another way: while there is long-range order for large λ, perhaps it is all explained by by the 50/50 chance that we see an even or odd dominated configuration. If we condition on the event that there are more odd occupied vertices than even, perhaps we can regain all the nice probabilistic properties of the dilute case (correlation decay etc.).

It turns out that this is indeed the case for a wide variety of lattice models (hard-core, Ising, Potts on $\mathbb{Z}^d$ for example), where absence of phase transition at low enough temperatures (high enough fugacities) can be proved by treating defects from ground states as a new spin model. This is captured by the framework of Pirogov-Sinai theory [60] in which the basic objects are *contours* separating regions of $\mathbb{Z}^d$ dominated by different ground states. For the applications below, however, the simpler *polymers models* will suffice to understand defect distributions.

4.2 Abstract polymer models

At a high level we can describe the multivariate hard-core model in terms of two properties:

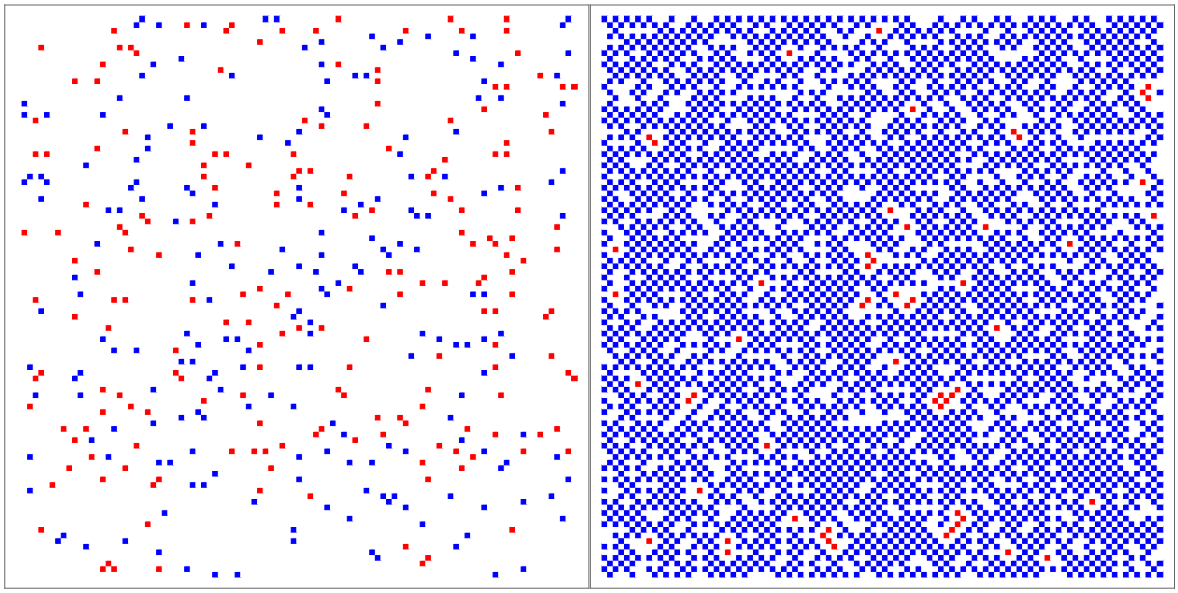
1. Configurations are collections of objects satisfying a pairwise geometric exclusion constraints (vertices cannot be neighbors in G).
2. The weight (probability) of a configuration factorizes over the objects in the collection.

Abstract polymer models provide a way to completely abstract these two properties, and they provide a natural setting in which to apply the cluster expansion. Abstract polymer models were introduced by Kotecký and Preiss [49] following previous definition of lattice polymer models by Gruber and Kunz [37].

Let $\mathcal{C}$ be a finite set of *polymers*. Each polymer comes with a real or complex-valued weight function w_γ. We equip $\mathcal{C}$ with a symmetric compatibility relation and write $\gamma \sim \gamma'$ if polymers γ and γ' are compatible (and $\gamma \nsim \gamma'$ if they are incompatible). We insist that $\gamma \nsim \gamma$ for all $\gamma \in \mathcal{C}$. The polymer model is defined by the triple $(\mathcal{C}, \sim, w)$.

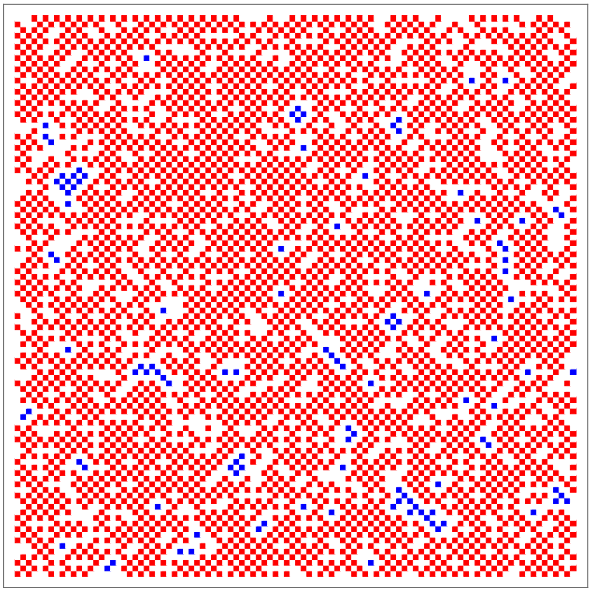
The polymer model partition function is

$$\Xi(\mathcal{C}) = \sum_{\substack{X \subseteq \mathcal{C} \\ compatible}} \prod_{\gamma \in X} w_\gamma$$



(a) Low density, $\lambda = .05$

(b) High density, $\lambda = 4.1$



(c) High density, $\lambda = 4.1$

Figure 5: Hard-core model on a two-dimensional torus

where the sum is over subsets of pairwise compatible polymers. If the weight functions are non-negative we can define the associated Gibbs measure with

$$\mu(X) = \frac{\prod_{\gamma \in X} w_\gamma}{\Xi(\mathcal{C})} .$$

4.2.1 Examples

Both of the following examples are *subgraph polymer models*: polymers are subgraphs of some host graph G (perhaps decorated by a labeling or coloring) and incompatibility is defined by distance or connectivity in the host graph.

Example: The multivariate hard-core model on G can be viewed as a polymer model with $\mathcal{C} = V(G)$, $u \sim v$ if $\mathrm{dist}(u, v) > 1$ and $w_v = \lambda_v$ for all v. Or in other words, the abstract polymer model is a multivariate hard-core model with vertex set $\mathcal{C}$ and graph structure given by the incompatibility graph.

Example: 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

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

where $Z_G(q, \beta)$ is the Potts model partition function and Ξ is the polymer model partition function. Notice that $e^{\beta|E(G)|}$ is the weight of the ground state (all red configuration), and so Ξ captures contributions to Z_G 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 this by restricting polymers to be 'small'.

4.3 Cluster expansion for abstract polymer models

The cluster expansion and Kotecký–Preiss condition for convergence fit very nicely with abstract polymer models.

A cluster Γ is a tuple of polymers whose incompatibility graph $H(\Gamma)$ is connected. The cluster expansion is the multivariate Taylor series of $\log \Xi(\mathcal{C})$ in the variables w_γ around 0:

$$\log \Xi(\mathcal{C}) = \sum_{\text{clusters } \Gamma} \phi(H(\Gamma)) \prod_{\gamma \in \Gamma} w_\gamma . \tag{20}$$

The Kotecký–Preiss result can be stated in this setting.

Theorem 4.1 (Kotecký-Preiss [49]). *Consider an abstract polymer model $(\mathcal{C}, \sim, w)$ with (possibly complex) weights w_γ. Suppose there are functions $a : \mathcal{C} \to [0, \infty)$, $b : \mathcal{C} \to [0, \infty)$ so that for all $\gamma \in \mathcal{C}$,*

$$\sum_{\gamma' \nsim \gamma} |w_{\gamma'}| e^{a(\gamma')+b(\gamma')} \leq a(\gamma)\,. \tag{21}$$

Then the following hold:

1. $\Xi(\mathcal{C}) \neq 0$.
2. *The cluster expansion for* $\log \Xi(\mathcal{C})$ *converges absolutely.*
3. *The following tail bound holds. Let* $b(\Gamma) = \sum_{\gamma \in \Gamma} b(\gamma)$. *Then for all* $t \geq 0$,

$$\sum_{\substack{clusters\ \Gamma \\ b(\Gamma) \geq t}} \left| \phi(H(\Gamma)) \prod_{\gamma \in \Gamma} w_\gamma \right| \leq e^{-t} \sum_{\gamma \in \mathcal{C}} a(\gamma)\,. \tag{22}$$

4. *We say* $\gamma \nsim \Gamma$ *if there exists* $\gamma' \in \Gamma$ *with* $\gamma \nsim \gamma'$. *The following 'pinned' bound holds for all* $\gamma \in \mathcal{C}$.

$$\sum_{\substack{clusters\ \Gamma \nsim \gamma \\ b(\Gamma) \geq t}} \left| \phi(H(\Gamma)) \prod_{\gamma' \in \Gamma} w_{\gamma'} \right| \leq e^{-t} a(\gamma)\,. \tag{23}$$

4.3.1 Cumulants and joint cumulants

By following the same procedure described in Section 3.3, we can compute cumulants and understand correlations in the abstract polymer model.

4.4 Low temperature expansions

We can use abstract polymer models to study the defects from a given ground state in a model at low temperature (strong interactions). The general approach involves three steps:

1. Express small deviations from each ground state of a model as an abstract polymer model; in particular the weight of a configuration of defects must factorize over polymers.
2. Control the polymer models by proving convergence of the cluster expansion.
3. If there is more than one ground state, approximate the partition function of the full model as a sum over polymer partition functions for each ground state. Prove that the weight of configurations that are either missed or double counted is small.

The last step is essentially proving that there is phase coexistence in the model; or in computer science terms, proving that local Markov chains exhibit *slow mixing*. It is similar to proving a stability result in combinatorics. Such a step is not always possible: ground states may not

be sufficiently separated (for example the ground states of the monomer-dimer model on $\mathbb{Z}^d$, perfect matchings, are not well separated).

If we can carry out these steps we achieve several things:

- A counting result: a good approximation of the partition function Z as the sum of polymer model partition functions.
- Probabilistic approximation: the Gibbs measure is well approximated by a weighted mixture of the polymer model Gibbs measures (with the appropriate translation of defect configurations to spin configurations).
- Conditional correlation decay: while there may be long-range correlations driven by the multiple ground states, we achieve *conditional* decay of correlations by conditioning on being close to one of the ground states.

Next we work through these steps in an example from [43].

4.4.1 Potts model on expander graphs

We continue with the Potts model example. Now we will make two assumptions on G:

1. G has maximum degree Δ
2. G is an α-edge-expander for some $\alpha > 0$; that is, for all $S \subset V, |S| \leq n/2$, $|\partial_e S| \geq \alpha |S|$.

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.

Fix one of the q colors and call it red. As above we define polymers to be connected induced subgraphs of G with vertices colored by the remaining $q-1$ colors; now we insist that $|\gamma| \leq n/2$ for all polymers γ. Here $|\gamma|$ denotes the number of vertices of γ. We will show that for β large enough as a functions of q, Δ, α, the Kotecký–Preiss condition is satisfied. To do this we need one lemma from, e.g. [33]:

Lemma 4.2. *Let G be a graph of maximum degree Δ. Then for every $v \in V(G)$, the number of connected induced subgraphs of size k containing v is at most $(e\Delta)^k$.*

This means that for a given polymer γ, the number of polymers γ' of size k incompatible with γ is at most $(e\Delta(q-1))^k \Delta |\gamma|$. The expansion condition gives us an upper bound on the weight of a polymer:

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

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

$$\sum_{\gamma' \nsim \gamma} w_{\gamma'} e^{a(\gamma')+b(\gamma')} \leq \sum_{k \geq 1} (e\Delta(q-1))^k \Delta |\gamma| e^{-\alpha\beta k} e^{2k}$$

$$\leq |\gamma|\Delta \sum_{k\geq 1} \exp\left[k\left(3+\log\Delta+\log(q-1)-\alpha\beta\right)\right]$$

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

Now we need to understand the separation of the q ground states. By using the expansion properties of G we can show that $q \cdot e^{\beta|E(G)|} \cdot \Xi$ approximates $Z(G)$ to within e^{-n} relative error.

Lemma 4.3 ([43])**.** *Let G be a n-vertex, α expander of maximum degree at most Δ. For $\beta \geq$, ..*

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

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 $\beta \geq$,

$$\sum_{r\in[q]} Z_r(\beta) = qZ_r(\beta) \leq Z_G(q,\beta) \leq (1+e^{-n})qZ_r(\beta)\,. \tag{24}$$

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 (24) 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 Ξ_r (after scaling by $e^{\beta|E(G)|}$) approximates $Z_r(\beta)$ up to an exponentially small relative error.

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

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.

4.5 More examples

4.5.1 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 $i(Q_d)$ [48].

Theorem 4.4 (Korshunov and Sapozhenko)**.** *As $d \to \infty$,*

$$i(Q_d) = (2 + o(1))\sqrt{e}2^{2^{d-1}}. \tag{25}$$

Sapozhenko later gave another proof of this result [68] 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 4.4 to the setting of weighted independent sets, the hard-core model on Q_d. He found the asymptotics of $Z_{Q_d}(\lambda)$ for $\lambda > \sqrt{2} - 1$:

$$Z_{Q_d}(\lambda) = (2 + o(1)) \cdot \exp\left[\frac{\lambda}{2}\left(\frac{2}{1+\lambda}\right)^d\right](1+\lambda)^{2^{d-1}}. \tag{26}$$

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 $\lambda > 2^{1/3} - 1$,

$$Z_{Q_d}(\lambda) = (2 + o(1)) \cdot \exp\left[\frac{\lambda}{2}\left(\frac{2}{1+\lambda}\right)^d\left(1 + \frac{(2\lambda^2+\lambda^3)d(d-1)-2}{4(1+\lambda)^d}\right)\right](1+\lambda)^{2^{d-1}}. \tag{27}$$

Compared to (26), the formula (27) 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 $\lambda > 2^{1/k} - 1$.

We sketch a proof of (27) 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. What is remarkably fortuitous is how well this 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.

In the case of the Potts model we took a ground state to be a single configuration, the all-red configuration for instance. Here we will do something different and take for a generalized ground state 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. This notion of a ground state is evident already in the formulas (25), (26), and (27) 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.

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

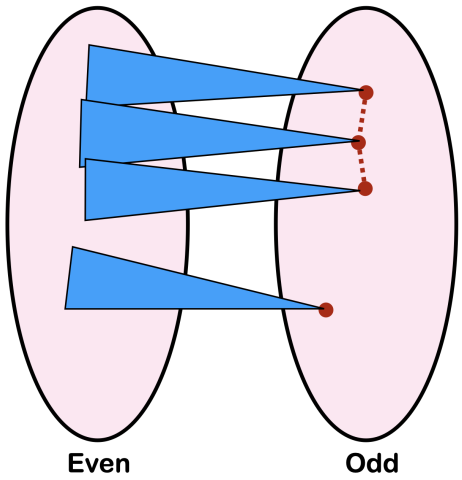


Figure 6: 2-linked components of occupied odd vertices, one of size 3 (top) and one of size 1 (bottom). Odd vertices are 2-linked if their neighborhoods overlap.

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

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

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).

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 4.5. *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 4.5 is a little more complicated than that of Lemma 4.3. It combines ideas from the slow-mixing result of Galvin and Tetali [35] with large deviation estimates from the polymer model itself.

Proving convergence of the cluster expansion is also 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 [68] to weighted independent sets. Let

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

Lemma 4.6 ([30])**.** *There exist constants $C_0, C_1 > 0$, so that for all $\lambda \geq C_0 \log d/d^{1/3}$, all $a \leq 2^{d-2}$,*

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

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

1. For $S \subset \mathcal{O}$, $|S| \leq d/10$, $|N(S)| \geq d|S| - |S|^2$.
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 estimate on the number of polymers containing a given vertex based on Lemma 4.2 is also far too pessimistic for large 2-linked subsets of Q_d. The balance between these two quantities is captured by Lemma 4.6.

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

Lemma 4.7. *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.

The proof of this lemma breaks 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 for the first two sums and the container lemma, Lemma 4.6, for the third sum.

As a consequence, we obtain a convergent series for $\log \Xi$ 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 $t - 1$.

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, 58, 35] for the best current lower bounds on phase coexistence.

4.5.2 Independent sets in unbalanced bipartite graphs

This example comes from [16] and is in some sense simpler than the two previous examples since we will work in a setting with a single ground state and so we will not need the step of showing that the full partition function is well approximated by a sum of polymer model partition functions. The example also shows that 'phase coexistence' – the balancing of phases around multiple ground states – is really a special phenomenon usually induced by symmetry, either in the model (the q-colors of the Potts model) or in the host graph (the symmetry between even and odd vertices in $\mathbb{Z}^d$ or Q_d).

Let G be a bipartite graph with bipartition (L, R) in which every vertex in L has degree Δ_L and every vertex in R has degree Δ_R. Consider the hard-core model on G in which vertices in L have fugacity λ_L and vertices in R have fugacity λ_R. Let $n_L = |L|, n_R = |R|$ (in particular $n_L \Delta_L = n_R \Delta_R$). The larger λ_L is relative to λ_R and the smaller Δ_L is to Δ_R, the more we expect typical independent sets from the model to be dominated by vertices from L. From this perspective, the ground state is the set of all independent sets that include no vertex from R; its weight is $(1 + \lambda_L)^{n_L}$.

To capture this in a polymer model we proceed as above and define polymers to be 2-linked components of R (with no restriction on their size), with two polymers incompatible if their union is 2-linked. The weight of a polymer γ is $w_\gamma = \lambda_R^{|\gamma|}(1 + \lambda_L)^{-|N(\gamma)|}$. Since we have no restriction on the size of polymers, we have the identity

$$Z_G(\lambda_L, \lambda_R) = (1 + \lambda_L)^{n_L} \cdot \Xi_G$$

where Ξ_G is the polymer model partition function.

We can check the Kotecký–Preiss convergence criteria for convergence of the cluster expansion for $\log \Xi_G$. We will bound the weight functions using the bi-regularity of G: $|N(\gamma)| \geq \frac{\Delta_R}{\Delta_L}|\gamma|$, and so

$$w_\gamma \leq \left(\frac{\lambda_R}{(1+\lambda_L)^{\Delta_R/\Delta_L}} \right)^{|\gamma|}.$$

We can also bound the number of polymers of size k incompatible with a given polymer γ by $(e\Delta_L\Delta_R)^k|\gamma|$. Then with $a(\gamma) = b(\gamma) = |\gamma|$ we have

$$\begin{aligned}\sum_{\gamma' \nsim \gamma} w_{\gamma'} e^{a(\gamma')+b(\gamma')} &\leq \sum_{k\geq 1} e^{2k}(e\Delta_L\Delta_R)^k|\gamma| \left(\frac{\lambda_R}{(1+\lambda_L)^{\Delta_R/\Delta_L}} \right)^k \\ &= |\gamma| \sum_{k\geq 1} \exp\left[k\left(3 + \log\Delta_L + \log\Delta_R + \log\lambda_R - \frac{\Delta_R}{\Delta_L}\log(1+\lambda_L) \right) \right]\end{aligned}$$

4.5.3 Cluster expansion in the canonical ensemble

In these lectures so far we have primarily discussed counting or approximating the number of independent sets of a graph or the weighted number of independent sets, $Z_G(\lambda)$. Another very natural quantity to consider is $i_k(G)$, the number of independent sets of size k in G, for various values of k. This parameter is associated to another very natural probabilistic model: choosing an independent set of size k uniformly at random from all independent sets of size k in G (assuming $i_k(G) > 0$). In the terminology of statistical physics this model is the 'canonical ensemble' while the hard-core model is the 'grand canonical ensemble'.

There is a nice formula relating the grand canonical partition function to $i_k(G)$ (which can be thought of as the partition function of the canonical ensemble):

$$i_k(G) = \frac{Z_G(\lambda)}{\lambda^k} \cdot \mu_{G,\lambda}(|\boldsymbol{I}| = k) \tag{28}$$

where λ is chosen so that $\mathbb{E}_{G,\lambda}[|\boldsymbol{I}| = k$.

[61]

[21]

4.6 Further applications and open problems

4.6.1 Combinatorics

Sokal [73] (and later Borgs [11]) used abstract polymer models and the cluster expansion to show that the chromatic polynomial of any graph of max degree Δ is not zero for q outside the disc of radius 8Δ in the complex plane.

Applications to enumeration are more recent. As described in Section 4.5.1, the methods of [44] combine polymer models and the cluster expansion with the graph container method of Sapozhenko [68] (and extensions by Galvin [30]).

Balogh, Garcia, Li [4] used a similar combination of tools to find the asymptotics of the number of independent sets in the middle two layers of the hypercube. In their work the cluster expansion proves useful since the asymptotic formula involves two terms of the cluster expansion. Jenssen and Keevash [42] study the very general setting of asymptotic enumeration of the number of homomorphisms from the hypercube (and other even side-length tori) to fixed graphs. One of their results is finding the asymptotics of the number of q-colorings of Q_d for all q (following the cases $q = 3$ by Galvin [29] and $q = 4$ by Kahn and Park [47]). Again the cluster expansion proves very useful: while the asymptotic formulas for $q = 3, 4$ only involve the first term of the cluster expansion (only the smallest defects matter) the formulas for $q \geq 5$ involve more terms of the cluster expansion and even guessing the right form would be difficult without the cluster expansion framework.

Question 4.8. *Can polymer models and the cluster expansion be used to prove sharper results in other situations in which the method of graph containers is used? (See [67] for an exposition of graph containers). Can polymers models be used in concert with the method of hypergraph containers [5, 69]?*

4.6.2 Algorithms

[6, 8, 7, 57, 9]

[40]

[43, 39, 12]

4.6.3 Contour models and Pirogov-Sinai theory

The original application of polymer models was in understanding phase diagrams of statistical mechanics on lattices. Unlike the examples discussed above, $\mathbb{Z}^d$ is not a good expander, and in many cases polymer models by themselves cannot capture the behavior of deviations from ground states. Instead what is needed is to represent a low temperature model in terms of contours: boundaries between regions dominated by different ground states. Contours differ from polymers in that contours This is captured by Pirogov-Sinai theory [60].

See [28, Chapter 7] for a nice introduction to these models. See [51] for the classic application to the random cluster and Potts models, [13] for an application to prove slow-mixing bounds, and [40, 12] for algorithmic applications.

Question 4.9. *Are there any combinatorial applications of the more sophisticated contour models of Pirogov-Sinai theory?*

4.7 Summary

- Abstract polymer models provide a method for understanding collections of interacting 'defects' in a probabilistic model. Often these defects are defined geometrically as some connected objects in a host graph.
- Formally an abstract polymer model is simply a multivariate hard-core model with fugacities given by weights and graph structure given by the incompatibility relation.

The cluster expansion and Kotecký–Preiss convergence criteria apply to the model, and so under sufficient decay of the weights, we can obtain asymptotics of the partition function and obtain detailed probabilistic information about the model.

- Using abstract polymer models we can switch from low temperature to high temperature and analyze the distribution of defects from ground states using the cluster expansion.
- The algorithmic perspective on counting problems can be very useful.
- The combinatorial perspective can be useful in statistical physics.

4.8 Exercises

1. Let $G = (V, E)$ be a graph for $q > 0, \beta > 0$ defined the *random cluster model*, a probability distribution on subsets of E with partition function given by:

$$Z_G(q, \beta) = \sum_{A \subseteq E} (e^\beta - 1)^{|A|} q^{c(A)}$$

where $c(A)$ is the number of connected components of the graph $G_A = (V, A)$.

   (a) What are the two possible ground states of the model? (Maximum weight configurations)

   (b) When β is very small what do you expect typical configurations to look like?

   (c) Can you write $Z_G(q, \beta)$ in terms of a polymer model that captures defects from the small β ground state?

   (d) Suppose G has maximum degree Δ. Find $\beta_0 = \beta_0(q, \Delta)$ so that if $\beta < \beta_0$, the Kotecký–Preiss condition is satisfied for this polymer model.

2. Prove (28). Be sure to use the specific choice of λ.

5 Sphere packings and kissing numbers

Here we will discuss continuum statistical physics models: models in which particles live in some region of Euclidean space rather than on vertices of a finite graph. In this setting some of the things we took for granted in spin models are not at all easy: for instance, determining the ground states of a model is closely related to the problem of determining the maximum sphere packing density in Euclidean space. We will focus here on the *hard sphere model*, one of the oldest models in statistical mechanics, and one in which ground states are exactly maximum density sphere packings.

5.1 Gibbs point processes and the hard sphere model

There are several equivalent ways to define the hard sphere model. We will do so via the Poisson process. Let $\Lambda \subset \mathbb{R}^d$ be bounded and let $\lambda \geq 0$. The *Poisson process* of intensity λ on Λ is a random point set $\mathbf{X} \subset \Lambda$ satisfying two properties:

1. For $B \subseteq \Lambda$, the distribution of the number of points in B, $|B \cap \mathbf{X}|$, is $\text{Pois}(\lambda|B|)$ where $|B|$ is volume of B.
2. For $B, B' \subset \Lambda$ with $B \cap B' = \emptyset$, the random variables $|X \cap B|$ and $|X \cap B'|$ are independent.

The Poisson process is the canonical non-interacting spatial point process. There are many extensions and generalizations of the Poisson process. We may replace the intensity λ with a measurable intensity function $\boldsymbol{\lambda}$, and in this case property 1 is replaced with the property that $|B \cap \mathbf{X}|$ has a Poisson distribution with mean $\int_{x \in B} \boldsymbol{\lambda}(x)\, dx$.

A *Gibbs point processes* generalizes the Poisson process by adding interactions between points. Given a function H from finite point sets to $\mathbb{R} \cup \{+\infty\}$, the Gibbs points process on Λ with intensity λ, Hamiltonian H, and inverse temperature β is the point process with density $e^{-\beta H(X)}$ against the Poisson process of intensity λ on Λ. In other words, the likelihood of each set of points is reweighted by the factor $e^{-\beta H(X)}$. The partition function of the Gibbs point process is

$$Z_\Lambda(\lambda) = 1 + \sum_{k \geq 1} \frac{\lambda^k}{k!} \int_{\Lambda^k} e^{-\beta H(x_1, \ldots, x_k)}\, dx_1 \cdots dx_k \,.$$

The most commonly studied models have Hamiltonians defined by a *pair potential*, a function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. Given such a ϕ, the Hamiltonian is

$$H(x_1, \ldots, x_k) = \sum_{i<j} \phi(x_i - x_j) \,.$$

For more about Gibbs point processes see [64, 25].

An important special case of a pair potential in the hard sphere potential. For some $r > 0$,

$$\phi(x) = \begin{cases} +\infty & \text{if} \quad \|x\| < r \\ 0 & \text{if} \quad \|x\| \geq r \,. \end{cases}$$

In particular, this potential forbids configurations in which a pair of points is at distance less than r; or in other words, valid configurations are the centers of a sphere packing of spheres of radius $r/2$. Temperature plays no role in the model: the only interaction between points is the hard-core exclusion. The hard sphere model represents a gas and it expected (in dimension 3) to exhibit a gas/solid phase transition. The model has been studied since the 1800's: Boltzmann wrote about computing coefficients of its virial expansion – an expansion of the pressure in the density. The model plays an important historical role in computer science as well: the Metropolis algorithm (Markov chain Monte Carlo) was first devised to sample from the two-dimensional hard sphere model [53].

Despite its long history, the main mathematical question about the model remains open.

Question 5.1. *Does the hard sphere model in $\mathbb{R}^3$ (or in any other dimension) exhibit a phase transition?*

It is an exercise to show that the one-dimensional hard sphere model does not exhibit a phase transition. In dimension two, it is believed that there is a phase transition but the details are debated even in physics [10]. In dimension three it is widely believed there is a gas/solid phase transition. Dimensions 8 and 24 have very dense maximum sphere packings (see below) and so one would expect a phase transition there too. In very high dimensions not very much is known at all – either about potential phase transitions or about the density and structure of the densest sphere packings.

5.2 Sphere packings in high dimensions

Let $B_r(x)$ be the ball of radius r around $x \in \mathbb{R}^d$.

Definition 5.2. *The maximum sphere packing density of d-dimensional Euclidean space, $\theta(d)$, is*

$$\theta(d) = \sup_{\mathcal{P}} \limsup_{R\to\infty} \frac{\mathrm{vol}(\mathcal{P} \cap B_R(0))}{\mathrm{vol}(B_R(0))},$$

where the supremum is over all sphere packings $\mathcal{P}$ of equal-sized spheres.

Of course $\theta(1) = 1$. $\theta(2) = \pi/\sqrt{12} = .9068\ldots$ with the packing given by the hexagonal lattice and this was proved by Thue in 1894. $\theta(3) = \pi/\sqrt{18} = .7404\ldots$ with the packing given by stacking hexagonal packings (exactly how you might try to stack oranges); this was Kepler's Conjecture and it was only proved in 2005 by Thomas Hales [38].

In a 2016 paper Maryna Viazovska proved that $\theta(8)$ is achieved by the E_8 lattice [78], and along with collaborators quickly proved that $\theta(24)$ is given by the Leech lattice [17]. All other dimensions are currently unknown.

What about optimal sphere packings in very high dimensions? Almost nothing is known! We do not know if the optimal packings are lattice packings or disordered. And our upper and lower bounds on $\theta(d)$ as $d \to \infty$ are very far apart.

A lower bound of $\theta(d) \geq 2^{-d}$ is trivial.

Proposition 5.3. *In all dimensions $\theta(d) \geq 2^{-d}$.*

Proof. Take any saturated (maximal) sphere packing, and double all the radii; because the original packing was saturated, the doubled balls must cover space (or else there would have been space for another center). Since the fraction of space covered increases by at most a factor 2^d, the original packing must have covered at least a 2^{-d} fraction of space. □

Compare Proposition 5.3 (and its proof) to the following bound on the independence number of a graph. (We use D here to distinguish vertex degree from dimension of Euclidean space).

Proposition 5.4. *For all graphs G of maximum degree D, $\alpha(G) \geq |V(G)|/(D+1)$.*

Proof. Take any maximal independent set I of G. Let $B(I)$ be the set of vertices in I and their neighbors. Because I was maximal, we must have $B(I) = V(G)$. And because G has maximum degree D, $|B(I)| \leq (D+1)|I|$, and so $|I| \geq |V(G)|/(D+1)$. □

This suggests an analogy between independent sets and sphere packings with the maximum degree D of a graph equivalent in some sense to the size of the excluded neighborhood of a center of a sphere packing (2^d times the volume of a sphere). Of course the centers of a sphere packing are in fact an independent set in the infinite graph with vertex set $\mathbb{R}^d$ in which two vertices are joined if their distance is at most $2r_d$.

The 2^{-d} bound has been improved by a factor of d by Rogers [63], with subsequent constant factor improvements by Rogers and Davenport [20], Ball [3], Vance [76], and finally Venkatesh [77] who proved that $\theta(d) \geq (65963 + o_d(1))d \cdot 2^{-d}$. Venkatesh also gains an additional $\log\log d$ factor in a sparse sequence of dimensions.

An upper bound of $2^{-(.599\cdots+o_d(1))\cdot d}$ was proved by Kabatiansky and Levenshtein [45]; there is a recent constant factor improvement by Cohn and Zhao [18].

The lower bounds mentioned above in fact show the existence of a lattice packing of the given density. This is clearly a stronger result that the existence of some packing, but it is not clear that by considering lattice packings only we will be able to close the gap in the bounds. There is a proof of $\theta(d) \geq .01 \cdot d2^{-d}$ by Krivelevich, Litsyn, and Vardy [50] using graph theory and the above mentioned result of Ajtai, Komlós, and Szemerédi [1], strengthening this analogy between Ramsey theory and sphere packing. It remains a major open problem to improve the exponential order of the upper or lower bound on $\theta(d)$.

5.2.1 A lower bound on the occupation density of the hard sphere model

Recall the hard sphere model on a bounded, measurable set $S \subset \mathbb{R}^d$. It is a random set $\mathbf{X}$ of centers in S at pairwise distance at least $2r_d$. Its partition function is

$$Z_S(\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \int_{S^k} \mathbf{1}_{\mathcal{D}(x_1,\ldots,x_k)}\, dx_1 \cdots dx_k\,.$$

The analogue of the occupancy fraction is the expected number of centers per unit of volume.

Definition 5.5. *The occupation density of the hard sphere model on S at fugacity λ is*

$$\alpha_S(\lambda) = \frac{1}{\mathrm{vol}(S)} \mathbb{E}_{S,\lambda} |\mathbf{X}| \,.$$

Theorem 5.6 (Joos, Jenssen, P. [41])**.** *Let B_n be the ball of volume n in $\mathbb{R}^d$. Then for $\lambda \geq 3^{-d/2}$,*

$$\alpha_{B_n}(\lambda) \geq (1 + o_d(1)) \frac{\log(2/\sqrt{3}) \cdot d}{2^d} \,.$$

The corresponding bound $\theta(d) \geq (1 + o_d(1)) \frac{\log(2/\sqrt{3}) \cdot d}{2^d}$ follows:

Lemma 5.7. *Let B_n be the ball of volume n in $\mathbb{R}^d$. Then for any $\lambda > 0$,*

$$\theta(d) \geq \limsup_{n \to \infty} \alpha_{B_n}(\lambda) \,.$$

The proof of this lemma uses the fact that volume grows subexponentially fast in $\mathbb{R}^d$.

5.2.2 Proof of Theorem 5.6

Given a set of centers X in S, we can partition S into three sets: points that are covered, blocked, and free. A point $x \in S$ is covered if $d(x, X) \leq r_d$; it is blocked if $d(x, X) \in (r_d, 2r_d]$ and free if $d(x, X) > 2r_d$.

Definition 5.8. *The expected free volume of the hard sphere model on S is*

$$\mathrm{FV}_S(\lambda) = \frac{1}{\mathrm{vol}(S)} \int_S \Pr_{S,\lambda}[d(x, \mathbf{X}) > 2r_d] \, dx \,.$$

That is, $\mathrm{FV}_S(\lambda)$ is the expected fraction of points in S that could be added to $\mathbf{X}$ and still result in a packing.

Lemma 5.9.

$$\alpha_S(\lambda) = \lambda \cdot \mathrm{FV}_S(\lambda) \,.$$

Proof. □

Now consider the following two-part experiment. Pick $\mathbf{X}$ from the hard sphere model on S and choose $\mathbf{v} \in S$ uniformly at random. Let $\mathbf{T}_S$ be the externally uncovered volume in the $2r_d$ neighborhood of $\mathbf{v}$; that is,

$$\mathbf{T}_S = \{x \in B_{2r_d}(\mathbf{v}) : d(x, y) > 2r_d \, \forall \, y \in \mathbf{X} \cap B_{2r_d}(\mathbf{v})^c\} \,.$$

Note that only centers outside of the $2r_d$ ball around $\mathbf{v}$ affect the set $\mathbf{T}_S$.

Lemma 5.10. *Let S be bounded and measurable, and consider the above two-part experiment. Then*

1. $\alpha_S(\lambda) = \lambda \mathbb{E}_{S,\lambda}\left[\frac{1}{Z_{\mathbf{T}_S}(\lambda)}\right]$.

2. $\alpha_S(\lambda) \geq 2^{-d}\lambda \mathbb{E}_{S,\lambda}\left[\frac{\lambda Z'_{\mathbf{T}_S}(\lambda)}{Z_{\mathbf{T}_S}(\lambda)}\right]$.

Proof. □

Lemma 5.11. *Let S be bounded and measurable.*

1. $\log Z_S(\lambda) \leq \lambda \cdot \mathrm{vol}(S)$.
2. $\alpha_S(\lambda) \geq \lambda \cdot e^{-\lambda \mathbb{E}_{S,\lambda}\mathrm{vol}(\mathbf{T}_S)}$.

Finally we need a simple geometric inequality about spheres in $\mathbb{R}^d$.

Lemma 5.12. *Let $S \subseteq B_{2r_d}(0)$ be measurable. Then*

$$\mathbb{E}[\mathrm{vol}(B_{2r_d}(\mathbf{u}) \cap S)] \leq 2 \cdot 3^{d/2},$$

where $\mathbf{u}$ is chosen uniformly from S.

Proof. □

Proof of Theorem 5.6. Let $B_n = B_{n^{1/d}r_d}(0)$. Let $\alpha_n = \alpha_{B_n}(\lambda)$.

We have

$$\alpha_n = \lambda \cdot \mathbb{E}_{B_n,\lambda}\left[\frac{1}{Z_{\mathbf{T}_{B_n}}(\lambda)}\right] \geq \lambda \cdot e^{-\lambda \mathbb{E}_{B_n,\lambda} \log Z_{\mathbf{T}_{B_n}}(\lambda)}. \tag{29}$$

On the other hand,

$$\begin{aligned}
\alpha_n &\geq 2^{-d}\mathbb{E}_{B_n,\lambda}\left[\frac{\lambda Z'_{\mathbf{T}_{B_n}}(\lambda)}{Z_{\mathbf{T}_{B_n}}(\lambda)}\right] && (30)\\
&= 2^{-d}\mathbb{E}_{B_n,\lambda}[\mathrm{vol}(\mathbf{T}_{B_n}) \cdot \alpha_{\mathbf{T}_{B_n}}(\lambda)] && (31)\\
&\geq 2^{-d}\mathbb{E}_{B_n,\lambda}\left[\lambda \cdot \mathrm{vol}(\mathbf{T}_{B_n}) \cdot e^{-\lambda \mathbb{E}_{\mathbf{T}_{B_n}}\mathrm{vol}(\mathbf{U})}\right] \text{ by Lemma 5.11 part 2} && (32)\\
&\geq 2^{-d}\mathbb{E}_{B_n,\lambda}\left[\log Z_{\mathbf{T}_{B_n}}(\lambda) \cdot e^{-\lambda \mathbb{E}_{\mathbf{T}_{B_n}}\mathrm{vol}(\mathbf{U})}\right] \text{ by Lemma 5.11 part 1} && (33)\\
&\geq 2^{-d}\mathbb{E}_{B_n,\lambda}\left[\log Z_{\mathbf{T}_{B_n}}(\lambda) \cdot e^{-\lambda 2 \cdot 3^{d/2}}\right] \text{ by Lemma 5.12} && (34)\\
&= 2^{-d} \cdot e^{-\lambda 2 \cdot 3^{d/2}}\mathbb{E}_{B_n,\lambda} \log Z_{\mathbf{T}_{B_n}}(\lambda). && (35)
\end{aligned}$$

Now with $z = \mathbb{E}_{B_n,\lambda} \log Z_{\mathbf{T}_{B_n}}(\lambda)$, we have

$$\alpha_n \geq \inf_{z \geq 0} \max\{\lambda e^{-z}, z \cdot 2^{-d} e^{-\lambda 2 \cdot 3^{d/2}}\}.$$

As before, one expression is decreasing in z and the other increasing and so the infimum is achieved at

$$z^* = W\left(\lambda \cdot \left(2e^{-\lambda 2 \cdot 3^{d/2}}\right)^d\right) .$$

Now take $\lambda = d^{-2}3^{d/2}$. Then

$$\begin{aligned} z^* &= W(\lambda 2^d e^{2/d}) \\ &= \log \lambda + d \log 2 - \log d - \log \log(2/\sqrt{3}) + o_d(1) . \end{aligned}$$

This gives

$$\alpha_n \geq (1 + o_d(1)) \frac{\log(2/\sqrt{3}) \cdot d}{2^d} .$$

□

5.3 Summary

-

5.4 Excercises

1.

Acknowledgements

I thank Andrey Kupavskii, Alexander Polyanskii, Yulia Rylova and the Laboratory of Combinatorial and Geometric Structures for inviting me and hosting these lectures. Thanks to all those who attended the lectures, asked great questions, and gave helpful feedback.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. A note on Ramsey numbers. *Journal of Combinatorial Theory, Series A*, 29(3):354–360, 1980.

[2] N. Alon. Independent sets in regular graphs and sum-free subsets of finite groups. *Israel Journal of Mathematics*, 73(2):247–256, 1991.

[3] K. Ball. A lower bound for the optimal density of lattice packings. *Int. Math. Res. Not.*, 1992(10):217–221, 1992.

[4] J. Balogh, R. Garcia, and L. Li. Independent sets in middle two layers of Boolean lattice. *arXiv preprint arXiv:2004.03060*, 2020.

[5] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *Journal of the American Mathematical Society*, 28(3):669–709, 2015.

[6] A. Barvinok. Computing the partition function for cliques in a graph. *Theory of Computing*, 11(13):339–355, 2015.

[7] A. Barvinok. *Combinatorics and complexity of partition functions*, volume 9. Springer, 2016.

[8] A. Barvinok. Computing the permanent of (some) complex matrices. *Foundations of Computational Mathematics*, 16(2):329–342, 2016.

[9] A. Barvinok and G. Regts. Weighted counting of solutions to sparse systems of equations. *Combinatorics, Probability and Computing*, 28(5):696–719, 2019.

[10] E. P. Bernard and W. Krauth. Two-step melting in two dimensions: first-order liquid-hexatic transition. *Physical Review Letters*, 107(15):155704, 2011.

[11] C. Borgs. Absence of zeros for the chromatic polynomial on bounded degree graphs. *Combinatorics, Probability and Computing*, 15(1-2):63–74, 2006.

[12] C. Borgs, J. Chayes, T. Helmuth, W. Perkins, and P. Tetali. Efficient sampling and counting algorithms for the Potts model on F^d at all temperatures. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 738–751, 2020.

[13] C. Borgs, J. T. Chayes, and P. Tetali. Tight bounds for mixing of the Swendsen–Wang algorithm at the Potts transition point. *Probability Theory and Related Fields*, 152(3-4):509–557, 2012.

[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] H. Cohn, A. Kumar, S. D. Miller, D. Radchenko, and M. Viazovska. The sphere packing problem in dimension 24. *Ann. of Math. (2)*, 185:1017–1033, 2017.

[18] H. Cohn and Y. Zhao. Sphere packing bounds via spherical codes. *Duke Math. J.*, 163:1965–2002, 2014.

[19] 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.

[20] H. Davenport, C. Rogers, et al. Hlawka's theorem in the geometry of numbers. *Duke Math. J*, 14(2):367–375, 1947.

[21] E. Davies, M. Jenssen, and W. Perkins. A proof of the Upper Matching Conjecture for large graphs. *arXiv preprint arXiv:2004.06695*, 2020.

[22] 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.

[23] 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.

[24] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. Tight bounds on the coefficients of partition functions via stability. *Journal of Combinatorial Theory, Series A*, 160:1 – 30, 2018.

[25] D. Dereudre. Introduction to the theory of Gibbs point processes. In *Stochastic Geometry*, pages 181–229. Springer, 2019.

[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] C. Efthymiou, T. P. Hayes, D. Stefankovic, E. Vigoda, and Y. Yin. Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. *SIAM Journal on Computing*, 48(2):581–643, 2019.

[28] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017.

[29] D. Galvin. On homomorphisms from the hamming cube to Z. *Israel Journal of Mathematics*, 138(1):189–213, 2003.

[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] J. Groeneveld. Two theorems on classical many-particle systems. *Phys. Letters*, 3, 1962.

[37] C. Gruber and H. Kunz. General properties of polymer systems. *Communications in Mathematical Physics*, 22(2):133–161, 1971.

[38] T. C. Hales. A proof of the Kepler conjecture. *Ann. of Math. (2)*, 162(3):1065–1185, 2005.

[39] T. Helmuth, M. Jenssen, and W. Perkins. Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs. *arXiv preprint arXiv:2006.11580*, 2020.

[40] T. Helmuth, W. Perkins, and G. Regts. Algorithmic Pirogov-Sinai theory. *Probability Theory and Related Fields*, 176:851–895, 2020.

[41] M. Jenssen, F. Joos, and W. Perkins. On the hard sphere model and sphere packings in high dimensions. *Forum of Mathematics, Sigma*, 7:E1, 2019.

[42] M. Jenssen and P. Keevash. Homomorphisms from the torus. *arXiv preprint arXiv:2009.08315*, 2020.

[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] G. A. Kabatjanskiĭ and V. I. Levenšteĭn. Bounds for packings on the sphere and in space. *Problemy Peredači Informacii*, 14:3–25, 1978.

[46] J. Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10(03):219–237, 2001.

[47] J. Kahn and J. Park. The number of 4-colorings of the hamming cube. *Israel Journal of Mathematics*, pages 1–21, 2020.

[48] A. Korshunov and A. Sapozhenko. The number of binary codes with distance 2. *Problemy Kibernet*, 40:111–130, 1983.

[49] R. Kotecký and D. Preiss. Cluster expansion for abstract polymer models. *Communications in Mathematical Physics*, 103(3):491–498, 1986.

[50] M. Krivelevich, S. Litsyn, and A. Vardy. A lower bound on the density of sphere packings via graph theory. *International Mathematics Research Notices*, 2004(43):2271–2279, 2004.

[51] L. Laanait, A. Messager, S. Miracle-Solé, J. Ruiz, and S. Shlosman. Interfaces in the potts model i: Pirogov-sinai theory of the fortuin-kasteleyn representation. *Communications in Mathematical Physics*, 140(1):81–91, 1991.

[52] J. Mayer and M. Mayer. *Statistical Mechanics*. John Wiley, 1940.

[53] 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.

[54] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[55] M. Michelen and W. Perkins. Analyticity for classical gasses via recursion. *arXiv preprint arXiv:2008.00972*, 2020.

[56] 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.

[57] 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.

[58] 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.

[59] H. Peters and G. Regts. On a conjecture of Sokal concerning roots of the independence polynomial. *The Michigan Mathematical Journal*, 68(1):33–55, 2019.

[60] S. A. Pirogov and Y. G. Sinai. Phase diagrams of classical lattice systems. *Teoreticheskaya i Matematicheskaya Fizika*, 25(3):358–369, 1975.

[61] E. Pulvirenti and D. Tsagkarogiannis. Cluster expansion in the canonical ensemble. *Comm. Math. Phys.*, 316(2):289–306, 2012.

[62] R. Restrepo, J. Shin, P. Tetali, E. Vigoda, and L. Yang. Improved mixing condition on the grid for counting and sampling independent sets. *Probability Theory and Related Fields*, 156(1-2):75–99, 2013.

[63] C. A. Rogers. Existence theorems in the geometry of numbers. *Annals of Mathematics*, pages 994–1002, 1947.

[64] D. Ruelle. *Statistical mechanics: Rigorous results*. World Scientific, 1999.

[65] 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.

[66] A. Sah, M. Sawhney, D. Stoner, and Y. Zhao. A reverse Sidorenko inequality. *Inventiones mathematicae*, pages 1–47, 2020.

[67] W. Samotij. Counting independent sets in graphs. *European Journal of Combinatorics*, 48:5–18, 2015.

[68] A. Sapozhenko. On the number of connected subsets with given cardinality of the boundary in bipartite graphs. *Metody Diskret Analiz*, 45:42–70, 1987.

[69] D. Saxton and A. Thomason. Hypergraph containers. *Inventiones mathematicae*, 201(3):925–992, 2015.

[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. D. Sokal. Bounds on the complex zeros of (di) chromatic polynomials and Potts-model partition functions. *Combinatorics, Probability and Computing*, 10(1):41–77, 2001.

[74] B. Szegedy. Edge coloring models and reflection positivity. *Journal of the American mathematical Society*, 20(4):969–988, 2007.

[75] D. Ueltschi. Cluster expansions and correlation functions. *Moscow Mathematical Journal*, 4(2):511–522, 2004.

[76] S. Vance. Improved sphere packing lower bounds from Hurwitz lattices. *Adv. Math.*, 227:2144–2156, 2011.

[77] A. Venkatesh. A note on sphere packings in high dimension. *Int. Math. Res. Not.*, 2013:1628–1642, 2013.

[78] M. S. Viazovska. The sphere packing problem in dimension 8. *Ann. of Math. (2)*, 185:991–1015, 2017.

[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] C.-N. Yang and T.-D. Lee. Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Physical Review*, 87(3):404, 1952.

[81] Y. Zhao. The number of independent sets in a regular graph. *Combinatorics, Probability and Computing*, 19(02):315–320, 2010.

[82] Y. Zhao. Extremal regular graphs: independent sets and graph homomorphisms. *The American Mathematical Monthly*, 124(9):827–843, 2017.