Five lectures on statistical mechanics methods in combinatorics

Will Perkins [*]

November 20, 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 Lovasz 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. *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.

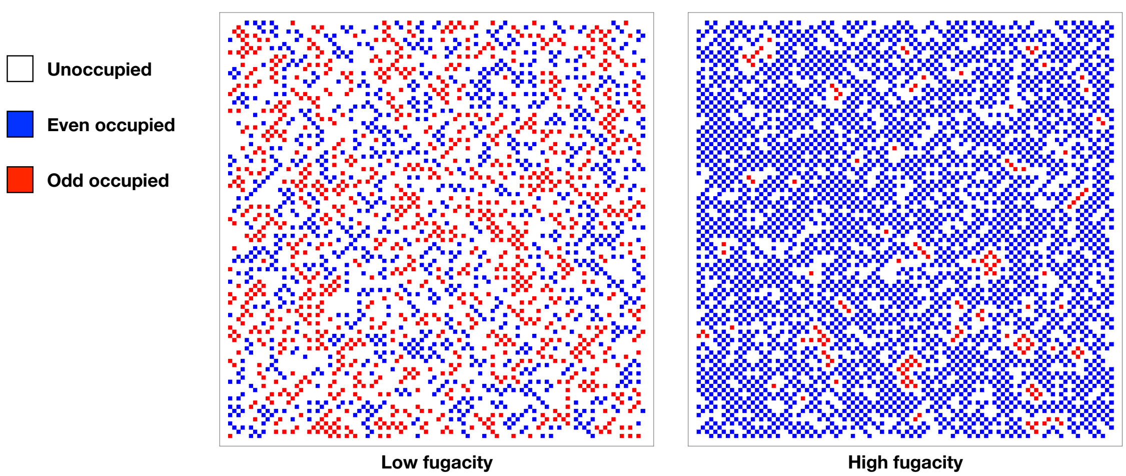
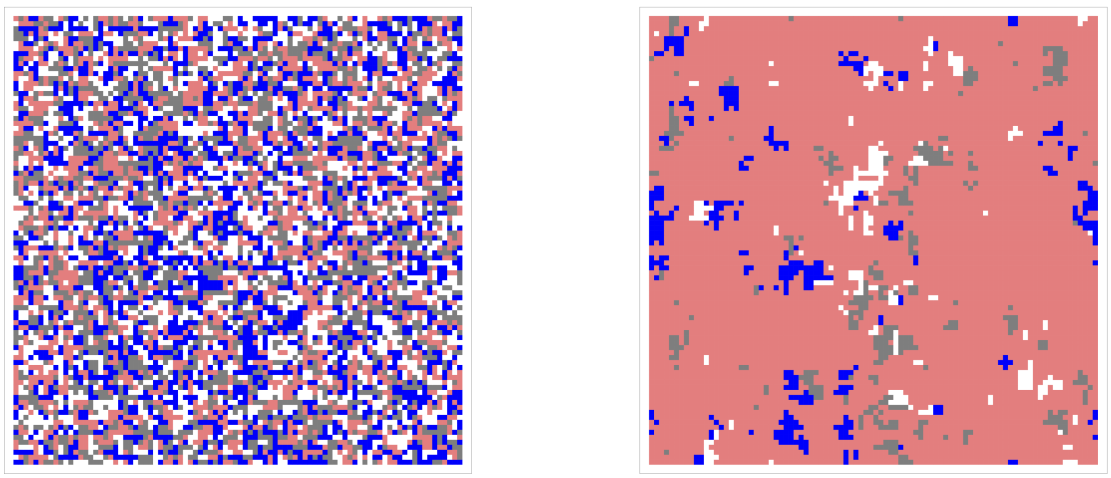


Figure 1: Two instances of the hard-core model on $\mathbb{Z}^2$

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.



High temperature (β small) **Low temperature (β large)**

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. [22]).

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*.

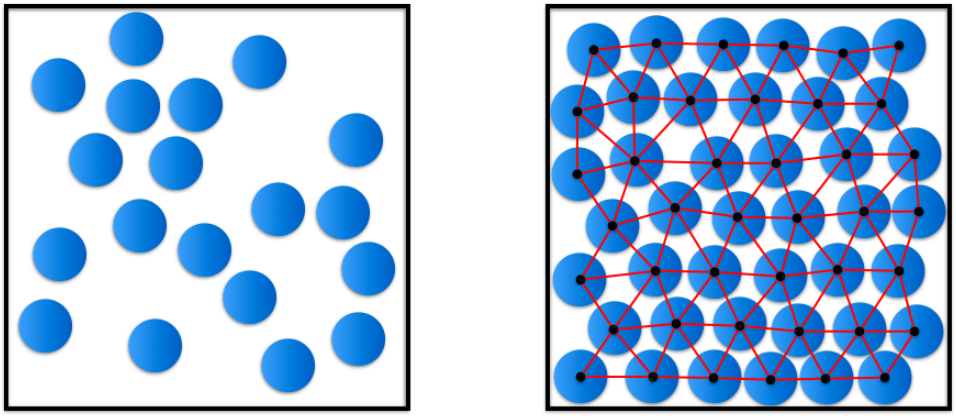


Figure 3: The hard sphere model at low and high density

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 [23]

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 [8].

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. [19, 16, 15]).

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 [10]. For many classical and foundational results (including for continuum models), see the classic text of Ruelle [17]. For a computational perspective on statistical physics models and random graphs, see the textbook of Mezard and Montanari [14].

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
$$\frac{\lambda}{(1+\lambda)^{\Delta+1}} \le \mu_v \le \frac{\lambda}{1+\lambda} .$$
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.

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 $\log Z_G(\lambda)$.

   (a) Prove that the limit $\lim_{n\to\infty} \frac{1}{|\Lambda_n|} \log Z_{\Lambda_n}(\lambda)$ exists. (

   (b) Show that the limit does not depend on the boundary conditions.

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 $\kappa(u, v) = \mu_{uv} - \mu_u \mu_v \ge 0$. 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 [25].

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 [13])**.** *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 [12] and Zhao [24] 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 [11] for an exposition of the proof of Theorem 2.1 and extensions. See also the recent work of Sah, Sawhney, Stoner, and Zhao [18] 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 [4]; for a short proof see [7].

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 [20]). *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 [20]). *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 [3]. 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. [13, 9]): 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 [5]). *For any d-regular graph G, and any $\lambda > 0$,*

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

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,

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

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. $\square$

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 [6])**.** *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:

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

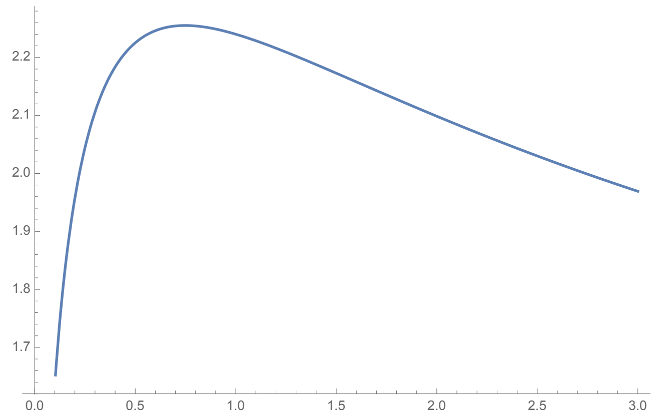


Figure 4: $\frac{\lambda}{1+\lambda} y^*$ as a function of λ with $d = 100$.

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. *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 2.11 ([6]). *For any triangle-free graph G, we have*

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

Conjecture 2.12 ([6]). *For any triangle-free graph G of minimum degree d, we have*

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

Conjecture 2.13 ([6]). *For any $\varepsilon > 0$, there is λ small enough so that for any triangle-free graph G we have*

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

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

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] L. Bregman. Some properties of nonnegative matrices and their permanents. *Soviet Math. Dokl*, 14(4):945–949, 1973.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] D. Galvin. Three tutorial lectures on entropy and counting. *arXiv preprint arXiv:1406.7872*, 2014.

[12] D. Galvin and P. Tetali. On weighted graph homomorphisms. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 63:97–104, 2004.

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

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

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

[16] 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.

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

[18] 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.

[19] 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.

[20] J. B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983.

[21] J. B. Shearer. On the independence number of sparse graphs. *Random Structures & Algorithms*, 7(3):269–271, 1995.

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

[23] 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.

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

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