Non-existence probabilities and lower tails via Gibbs uniqueness on hypertrees

Talk by Matthew Jenssen

Notes by Sanjana Das

April 10, 2025

This is joint work with Will Perkins, Aditya Potukuchi, and Michael Simkin.

§1 Introduction

The title is fairly long; we'll get to each bit of it in sequence.

§1.1 Motivating questions

The kind of question we'll be interested in today is questions of the following flavor.

> **Question 1.1.** What's the probability that the Erdős–Rényi random graph $\mathcal{G}(n, p)$ has no copy of a certain subgraph, such as K_k?

> **Question 1.2.** What's the probability that a p-random subset of $[n]$ avoids some substructure, such as a k-term arithmetic progression?

> **Question 1.3.** What's the probability that a p-random subset of some group, e.g., $\mathbb{Z}/n\mathbb{Z}$, is sum-free?

These are questions about non-existence probabilities. The title also has *lower tails*; this corresponds to asking questions of the following form.

> **Question 1.4.** Instead of completely forbidding these substructures, what's the probability we see fewer copies of these substructures than expected (e.g., 1% of the expected number)?

You could also ask these questions in related models, like $\mathcal{G}(n, m)$ or a random subset of $[n]$ of fixed size.

These are the types of questions we'll be interested in. We'll mostly focus on a specific case for concreteness, but we'll discuss a fairly general approach to these questions, using tools from statistical physics and computer science.

§1.2 Nonexistence probabilities for triangles

For now we'll specialize to Question 1.1 for concreteness, and replace K_k with a triangle.

Question 1.5. What's the probability that $\mathcal{G}(n,p)$ is triangle-free?

First we'll talk about what's known already. As a concatenation of two classical theorems in combinatorics:

Theorem 1.6 (Janson–Łuczak–Rucinski 1988, Łuczak 2000)

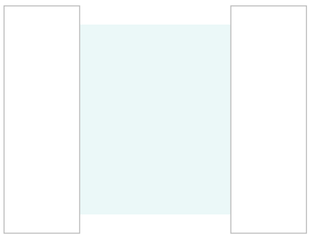
- If $p = o(1/\sqrt{n})$, then $\log \mathbb{P}[\mathcal{G}(n,p) \not\supseteq K_3] \sim -p^3\binom{n}{3}$.
- If $p = \omega(1/\sqrt{n})$, then $\log \mathbb{P}[\mathcal{G}(n,p) \not\supseteq K_3] \sim \frac{n^2}{4}\log(1-p)$.

This is the sort of level of accuracy we're interested in today — determining the first-order asymptotics of $\log \mathbb{P}[\bullet]$. (We write $x \sim y$ to mean $x = (1+o(1))y$.)

The first statement, for the regime $p = o(1/\sqrt{n})$, is the first example of Janson's inequality you'd see in a probabilistic combinatorics course, and is fairly straightforward to prove. The opposite regime $p = \omega(1/\sqrt{n})$ is more tricky, and here the formula is different. It was originally proved using sparse regularity methods; now it can be proved a bit more cleanly using hypergraph containers.

Where do these formulas come from? For the first statement, imagine that the appearance of triangles is independent (this isn't actually true because triangles can overlap, but we'll imagine it is as some Poisson heuristic). For a single triangle, the probability of not getting it is $1 - p^3$. And we want to get none of the $\binom{n}{3}$ possible triangles; so if they're independent this would have probability $(1-p^3)^{\binom{n}{3}}$. That's where this asymptotic comes from — so you should think of it as coming from disordered-looking graphs with no particular structure.

The second statement comes from a very different construction. To get a lower bound, we can just plant a bipartition (let's say a balanced one) and only consider the graphs that respect that bipartition.



What's the probability of seeing such a graph in $\mathcal{G}(n,p)$? We need to have no edges inside each of the two parts; the probability that happens is roughly $(1-p)^{n^2/4}$ (there's roughly $\frac{n^2}{4}$ edges within the parts; the number of ways to choose the bipartition doesn't really contribute). So here the lower bound comes from a very different kind of construction — bipartite graphs, which have structure.

So there's a stark transition in structure between the two regimes. This leaves open the regime in between them, which we'll call the *critical* regime (for reasons we'll justify later).

Question 1.7. What if $p = c/\sqrt{n}$?

Remarkably, Janson's inequality still gives you the right order of magnitude.

Theorem 1.8

If $p = c/\sqrt{n}$, then $\log \mathbb{P}[\mathcal{G}(n,p) \not\subseteq K_3] = -\Theta(n^{3/2})$.

So Janson's inequality tells us the asymptotic order of growth of the probability. But what's the right constant — can we get to a similar level of accuracy as the previous results? Let's give this constant a name, for convenience — let $p = c/\sqrt{n}$, and define

$$f(c) = \lim_{n\to\infty} \frac{1}{n^{3/2}} \log \mathbb{P}[\mathcal{G}(n,p) \not\supseteq K_3].$$

So $f(c)$ is just the correct constant in the $\Theta(n^{3/2})$ term; it's also called the *rate function*.

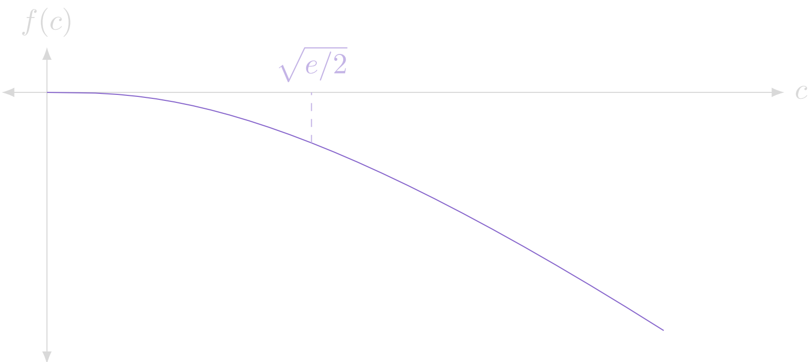
The first result we'll talk about is a determination of $f(c)$ when c is small. This is a strange-looking formula whose exact form is not that important, but we'll draw some pictures and discuss it.

Theorem 1.9 (Jenssen–Perkins–Potukuchi–Simkin)
If $c < \sqrt{e/2}$, then we have

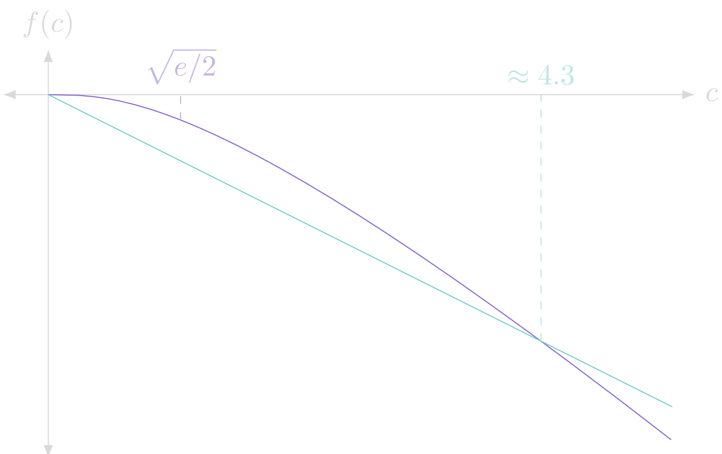
$$f(c) = \frac{1}{2}\left(\frac{W(2c^2)^{3/2} + 3W(2c^2)^{3/2}}{1\sqrt{2}} - c\right),$$

where W is the function such that $W(xe^x) = x$.

This formula looks a bit strange, but we'll motivate it by the end of today. But it's more enlightening to draw a graph of this function; so let's do that. The right-hand side is some downwards-pointing curve (it's negative because we're taking a log of a probability); and this theorem says it's the truth up to $c = \sqrt{e/2}$.



This curve can't be the truth for *all* c, though. Why not? We have a lower bound coming from bipartite graphs. You should think of this bound as coming from 'disordered' graphs; but for large c we get a lower bound of $-\frac{c}{4}$ coming from bipartite graphs; and the two cross at some point, around 4.3.



This means the rate function f can't be analytic for all c — there must be some phase transition in this problem (in the sense that $f(c)$ will be non-analytic in c at some point). This is quite interesting — that one of the earliest questions asked about Erdős–Rényi random graphs, on the probability of triangle-free-ness, has this mysterious question about phase transitions at its heart.

It turns out that we can also determine $f(c)$ for large c, though that's a different story. But there's definitely a region in between where we have no clue. (For example, is the phase transition unique? Is it first order? Is it second order?)

§1.3 Lower tails

Let's talk a little bit about lower tails, to show some interesting different behaviors. (We'll focus on the question of triangles in $\mathcal{G}(n, p)$; but when we talk about the proofs, it'll be clear how the ideas apply to the other problems as well.)

Here we let X denote the number of triangles in $\mathcal{G}(n, p)$.

Question 1.10. Given $\varepsilon \in [0, 1)$, what's $\mathbb{P}[X \leq \varepsilon\mathbb{E}[X]]$?

For example, if $\varepsilon = \frac{1}{2}$, we're asking for the probability $\mathcal{G}(n, p)$ has half its expected number of triangles.

Again, we want to study this in the critical regime $p = c/\sqrt{n}$. So we define a more general rate function

$$f(c, \varepsilon) = \lim_{n\to\infty} \frac{1}{n^{3/2}} \log \mathbb{P}[X \leq \varepsilon\mathbb{E}[X]].$$

Note that when $\varepsilon = 0$, we recover the non-existence problem we were just thinking about (the probability that $\mathcal{G}(n, p)$ is triangle-free) — i.e., $f(c, 0) = f(c)$.

Interestingly, if ε is not too severe, the lower tails problem actually doesn't have a phase transition, and we can determine the rate function for all c! Instead of writing down another big formula, we'll just say:

Theorem 1.11 (Jenssen–Perkins–Potukuchi–Simkin)
If $\varepsilon \geq \frac{1}{2}$, then $f(c, \varepsilon)$ is analytic for all $c > 0$ (i.e., there is no phase transition).

Remark 1.12. The bound $\varepsilon \geq \frac{1}{2}$ is not tight; it's an interesting open question to determine for which ε no phase transition phenomenon appears.

Why do we not have a phase transition anymore? The heuristic is that this curve is interpolating between two different strategies for having fewer triangles than expected. When c is very small, you can take your random graph and delete a few edges from triangles without affecting the global density too much, in order to suppress triangles (this is somehow realizing the Poisson heuristic that triangles are roughly independent).

When c is very large, there's a different strategy for suppressing triangles. In the non-existence problem, we saw one strategy, which was to be bipartite; but it turns out that with these less severe lower tails, there's a different strategy. If you're $\mathcal{G}(n, p)$, what could you do to have fewer triangles than expected? You could pretend that you're $\mathcal{G}(n, q)$, where $q < p$ is chosen so that $\mathcal{G}(n, q)$ has $\varepsilon\mathbb{E}[X]$ expected triangles. So here the strategy is to just pretend you're a sparser random graph. This has a name — it's sometimes called the *replica symmetry* bound.

These two strategies (the Poisson heuristic and the replica symmetry bound) aren't so different — they're similar enough that you can analytically interpolate between the two, which is why we don't get a phase transition. This is very different from the previous situation (on non-existence probabilities), where you had two completely different strategies you couldn't reconcile.

Remark 1.13. You can also study these problems in $\mathcal{G}(n,m)$. That model is interesting because there, this strategy of pretending to be $\mathcal{G}(n,q)$ isn't available — if you have a fixed number of edges, you can't pretend to be a graph of lower density. And that means in this lower tails problem, $\mathcal{G}(n,m)$ actually has a phase transition for *every* ε!

§2 Gibbs measures on hypergraphs

The problems we've discussed — e.g., $\mathcal{G}(n,p)$ being triangle-free or random subsets of $[n]$ being AP-free — can all be phrased as problems about independent sets in hypergraphs, where the hyperedges encode your structure and an independent set means you avoid that structure. We'll see examples soon, but hopefully this motivates why we'll discuss hypergraphs.

§2.1 From non-existence to independent sets in hypergraphs

Suppose we have a hypergraph $\mathcal{H} = (V, E)$; this means the edge set E is some collection of subsets of the vertex set V. We'll assume it's a k-uniform hypergraph, so every edge has exactly k elements.

Definition 2.1. An independent set in $\mathcal{H}$ is a collection of vertices which doesn't fully contain any edge; we write $\mathcal{I}(\mathcal{H})$ for the set of all independent sets in $\mathcal{H}$.

We'll consider percolation on the *vertex* set of $\mathcal{H}$:

Definition 2.2. We define $\mathcal{H}_p$ as the hypergraph formed from $\mathcal{H}$ by taking a p-random subset of V.

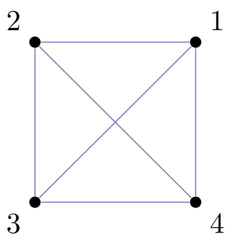
So we look at the vertices of $\mathcal{H}$, and select them independently at random with probability p.

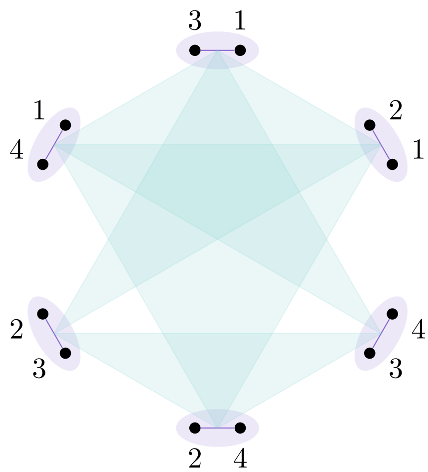
What does this have to do with our question, the probability that $\mathcal{G}(n,p)$ is triangle-free?

Example 2.3

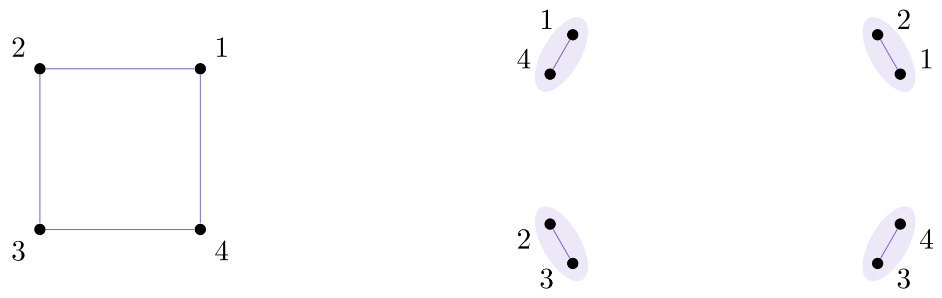
Consider the hypergraph $\mathcal{H} = \mathcal{H}_{K_3}$ with vertex set $V(\mathcal{H}) = E(K_n)$ and edge set

$$E(\mathcal{H}) = \{efg \mid efg \text{ forms a triangle in } K_n\}.$$





So the vertices of $\mathcal{H}$ are *edges* of K_n (the complete graph on n vertices), and the edges of $\mathcal{H}$ are triples of edges in K_n which would form a triangle. What's the point of constructing this auxiliary hypergraph? An independent set in this hypergraph is a subset of vertices containing no edges of $\mathcal{H}$, i.e., a subset of edges in K_n containing no triangles. So independent sets in $\mathcal{H}_{K_3}$ are in one-to-one correspondence with triangle-free subgraphs of K_n.



And what is $\mathcal{H}_p$? We're taking a p-random subset of the vertices of $\mathcal{H}$, i.e., the edges of K_n; so this is just $\mathcal{G}(n,p)$. Then Question 1.5 asks, what's the probability that $\mathcal{H}_p$ is an independent set?

Question 2.4. What is $\mathbb{P}[\mathcal{H}_p \in \mathcal{I}(\mathcal{H})]$?

In this particular example, this probability corresponds to $\mathbb{P}[\mathcal{G}(n,p) \not\supseteq K_3]$. Hopefully we can see how this general setup also encodes all the other problems we mentioned (for example, to encode Question 1.2 you can create a hypergraph on $[n]$ with an edge around each k-AP).

§2.2 The hardcore model

We're going to do something slightly confusing and define another measure on this hypergraph $\mathcal{H}$, but then we'll talk about how they're related. We previously defined a very simple measure $\mathcal{H}_p$ by percolation on the vertex set of $\mathcal{H}$. But here's another distribution on vertex subsets of $\mathcal{H}$ (actually just on independent sets).

Definition 2.5. For $\lambda > 0$, the hardcore model is the distribution on $\mathcal{I}(\mathcal{H})$ defined by

$$\mathbb{P}_{\mathcal{H},\lambda}[I] = \frac{\lambda^{|I|}}{Z_{\mathcal{H}}(\lambda)},$$

where $Z_{\mathcal{H}}(\lambda) = \sum_{I \in \mathcal{I}(\mathcal{H})} \lambda^{|I|}$ is called the partition function.

This is more confusing than the previous measure, but if you squint at it, it's just percolation conditioned on the resulting vertex subset being an independent set. (The hardcore model is often studied on graphs, but it makes sense on hypergraphs as well.)

Why is it useful to think about this object? First, as a philosophical point: Why are people in statistical physics interested in partition functions? One reason is that these partition functions contain lots of statistical information about your model which you can extract by taking derivatives.

Example 2.6

What's $\mathbb{E}_{\mathcal{H},\lambda}[|I|]$, i.e., the average size of an independent set in this model?

(If λ is large, we're biasing towards larger independent sets; if λ is small, we're biasing towards smaller ones.) Writing out the definition, we have

$$\mathbb{E}_{\mathcal{H},\lambda}[|I|] = \sum_{I \in \mathcal{I}(\mathcal{H})} |\mathcal{I}| \cdot \frac{\lambda^{|I|}}{Z_{\mathcal{H}}(\lambda)}.$$

This looks a lot like a derivative, and indeed we can write it as

$$\mathbb{E}_{\mathcal{H},\lambda}[|I|] = \frac{\lambda \cdot Z'_{\mathcal{H}}(\lambda)}{Z_{\mathcal{H}}(\lambda)} = \lambda \cdot \frac{d}{d\lambda}(\log Z_{\mathcal{H}}(\lambda)). \tag{2.1}$$

So just by taking a logarithmic derivative, we get this nice first moment.

Now, what on earth does this model have to do with $\mathbb{P}[\mathcal{H}_p \in \mathcal{I}(\mathcal{H})]$, which is the object we were interested in? We're looking at the probability that if we choose a random subset, we get an independent set. If we go over every independent set and sum the probabilities of picking that set, we get

$$\mathbb{P}[\mathcal{H}_p \in \mathcal{I}(\mathcal{H})] = \sum_{I \in \mathcal{I}(\mathcal{H})} p^{|I|}(1-p)^{|V|-|I|} = (1-p)^{|V|} Z_{\mathcal{H}}\left(\frac{p}{1-p}\right).$$

So we've rephrased the problem — to answer Question 2.4, it's enough to understand this partition function $Z_{\mathcal{H}}$; and to understand this partition function, it's enough to understand $\mathbb{E}_{\mathcal{H},\lambda}[|I|]$ (if we understand this, then by (2.1), we can recover the partition function by taking an integral).

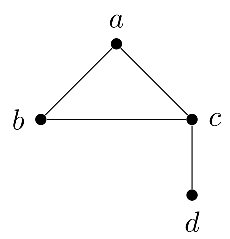
So we started with a confusing large deviation probability (the probability of a graph having no triangles), and we've turned it into the simplest first moment you could ask for in this new measure. The tradeoff is that this first moment is with respect to a confusing measure — we started with a simple measure but confusing probability, and transformed it into a complicated measure but simple moment. So it's unclear whether we've gained anything. But it turns out that we have, because statistical physics has lots of tools for dealing with confusing measures.

§3 Trees

We've been talking about hypergraphs, but to illustrate the nice idea from statistical physics that we use, we'll go back to graphs (and then we'll wave our hands and say this works for hypergraphs as well).

§3.1 A self-avoiding walk tree

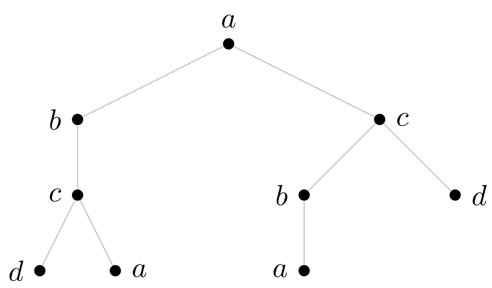
We'd like to understand the hardcore model on graphs. So let's draw a simple graph $\mathcal{H}$ as an example.



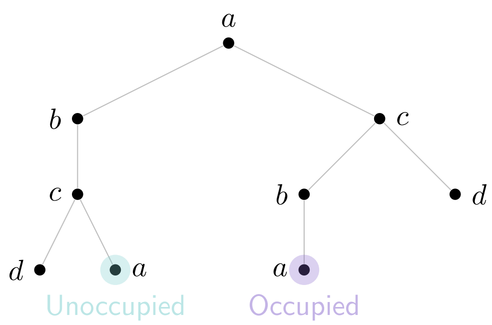
Suppose we want to understand the marginal at a:

Question 3.1. If we draw an independent set according to the hardcore model, what's the probability that a is in it (i.e., $\mathbb{P}_{\mathcal{H},\lambda}[a \in I]$)?

There's a theorem that gives a really neat way of computing this marginal, due to Weitz. We first construct a self-avoiding walk tree. This means we want to consider self-avoiding walks starting at a. So $a \to b \to c \to d$ is in the tree (we stop at d because we're not allowed to backtrack along an edge). We could also walk $a \to b \to c$, and then we'd get stuck because we're forced to close a cycle; so we add the walk $a \to b \to c \to a$ to the tree as well. Here's the full tree for this example:



This defines some tree, which we denote by $\mathcal{T}_{\mathcal{H}}(a)$. There's one twist: Whenever we close a cycle, we put a boundary condition on the last vertex, setting it to Occupied or Unoccupied. This means if we're considering an independent set in this graph $\mathcal{T}_{\mathcal{H}}(a)$, we're forcing that vertex to be either in it (if it's occupied) or out of it (if it's unoccupied). An alternative way of thinking about this is that if we're forcing v to be in some independent set, then none of its neighbors can be; so we can imagine that if v is labelled Occupied then we delete v and all its neighbors, while if it's labelled Unoccupied then we just delete v. (Which boundary condition we put depends on the direction we went along the cycle; it's not super important.)



The point of this is that we can compute $\mathbb{P}_{\mathcal{H},\lambda}[a \in I]$ by moving over from $\mathcal{H}$ to this tree:

Theorem 3.2 (Weitz)
We have $\mathbb{P}_{\mathcal{H},\lambda}[a \in I] = \mathbb{P}_{\mathcal{T}_{\mathcal{H}}(a),\lambda}[\mathsf{root} \in I]$.

This is really neat because cycles are what make statistical physics models confusing; hardcore models on trees are much simpler to analyze.

§3.2 Correlation decay

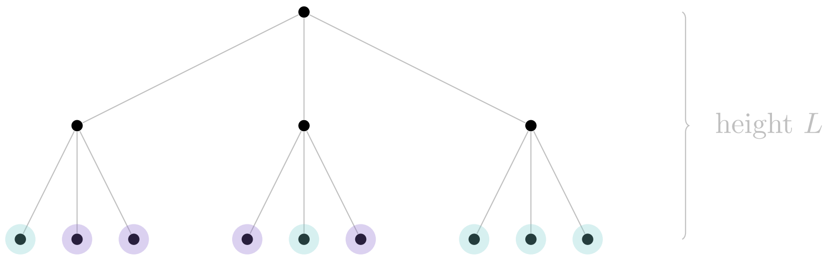
Trees are in some sense the first graph you go to when trying to understand some statistical physics model, and the simplest tree is a regular tree.

Definition 3.3. We use $\mathbb{T}_\Delta$ to denote the infinite Δ-regular tree.

Physicists are normally curious about phase transition phenomena, and this tree is a first testing ground for those. Here what we mean by this is:

Question 3.4. When is the hardcore model on this tree ordered or disordered?

What does that mean? Suppose we take this tree up to depth L and put some boundary conditions on the leaves. Can we do this in some tricky way to influence the probability the root is occupied, as L gets larger?



If the answer is no, we say there's some *decay of correlations*; if the answer is yes, we say there's no correlation decay. Whether the answer is yes or no may depend on λ. If λ is small, then you might expect correlation decay because the independent sets we're drawing are typically small. But if λ is large, you might not expect correlation decay because our independent sets are large, so if we put some structure at the leaves, the rest of the tree might have to align with it.

Theorem 3.5

There is $\lambda_c(\Delta) \sim e/\Delta$ such that if $\lambda < \lambda_c(\Delta)$, the hardcore model on $\mathbb{T}_\Delta$ exhibits decay of correlations.

This means if we truncate the tree at depth L, then for any assignment to the leaves, the root's occupation probability will be within a $(1 + o(1))$-factor of the original probability (as $L \to \infty$).

Remark 3.6. This result, combined with Theorem 3.2 of Weitz, was very influential: One of the earliest problems in CS was to come up with algorithms for sampling from the hardcore model (e.g., this is what the Metropolis algorithm was invented for). And this gives you a very neat algorithm for doing so. To sample from the hardcore model on a graph, it's enough to be able to approximate the marginal very closely at an arbitrary vertex a. So you can build the tree $\mathcal{T}_\mathcal{H}(a)$ (which might be very large), and then truncate this tree at some depth and do brute force on the truncated tree. And if we have correlation decay, this should give a good estimate on the true marginal. So this gives an efficient sampling algorithm when $\lambda < \lambda_c(\Delta)$.

§3.3 Tree-like hypergraphs

Why is this useful for us? We're interested in calculating $\mathbb{E}_{\mathcal{H},\lambda}[|I|]$ (the expected size of an independent set). By linearity of expectation, this is a sum of marginals, i.e.,

$$\mathbb{E}_{\mathcal{H},\lambda}[|I|] = \sum_a \mathbb{P}_{\mathcal{H},\lambda}[a \in I].$$

So it would be great if we could understand the marginals of a hardcore model on a graph (really a hypergraph). If the graph has lots of cycles, then its associated tree $\mathcal{T}_\mathcal{H}(a)$ is going to be confusing and irregular, so it won't necessarily be possible to extract a nice formula. But under some conditions, maybe this tree looks regular enough that we can understand its marginals very well analytically. (There's a hypergraph analog of the Weitz method.)

So we'll now write down some simple conditions on a hypergraph which will mean that its Weitz hypertree looks nice and regular; we'll call such a hypergraph a *tree-like hypergraph.* (You might recognize some of these conditions — they pop up in lots of places in combinatorics.)

As usual, $\mathcal{H} = (V, E)$ is some k-uniform hypergraph.

Definition 3.7. We define the ℓ-degree of $\mathcal{H}$ as $\Delta_\ell(\mathcal{H}) = \max_{|S|=\ell} \#\{e \in E \mid S \subseteq e\}$.

So we take some subset of size ℓ and ask how many hyperedges it's contained in, and that's the ℓ-degree. (The 1-degree is just the vertex degree; we write $\Delta_1 = \Delta$.)

Definition 3.8. We say a Δ-regular hypergraph $\mathcal{H}$ is tree-like if:

- We have $\Delta_\ell(\mathcal{H}) = o(\Delta^{(k-\ell)/(k-1)})$ for all $\ell \in \{2, \dots, k-1\}$.
- For all u and v, we have $\#\{|S| = k-2 \mid S \cup \{u, v\} \in E\} = o(\Delta)$.

The first condition says that the ℓ-degree of $\mathcal{H}$ is well-controlled in terms of the 1-degree. And the second condition is some codegree condition; it says that for any u and v, the number of $(k-2)$-tuples that complete them into an edge is $o(\Delta)$.

Remark 3.9. Hypergraphs satisfying these conditions are also sometimes called *unclustered.* These are precisely the conditions you need to get the hypergraph container method off the ground, and they're also the conditions you need to study the random greedy independent set algorithm on hypergraphs (this is work of Bohman and Bennet). What's nice is that when you *don't* have uniqueness on the hypertree, that's precisely when the hypergraph container method becomes useful; so these two methods fit together nicely.

The point is that if $\mathcal{H}$ is Δ-regular and tree-like, then the Weitz hypertree $\mathcal{T}_{\mathcal{H}}(a)$ 'looks like' a k-uniform linear Δ-regular hypertree.

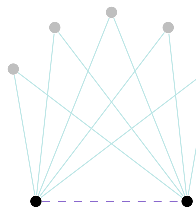
§3.4 Back to triangle-free graphs

To finish, we'll return to Question 1.5 — we'll go back to the example $\mathcal{H} = \mathcal{H}_{K_3}$, and suppose we want to understand $\mathbb{E}_{\mathcal{H},\lambda}[|I|]$ (the expected size of an independent set on this hypergraph). This is $\binom{n}{2}$ times the probability some *given* vertex is in the independent set; so it's enough to understand the marginal

$$q = \mathbb{P}_{\mathcal{H},\lambda}[v \in I].$$

And we know that if we build the Weitz hypertree and our hypergraph is tree-like (which it's easy to check it is), then this should look a lot like the root occupation probability in the hardcore model on this hypertree. And you can just write down a formula for what that root occupation probability is.

We won't write down the formula, but we'll describe the heuristic that this argument (passing to hypertrees) makes rigorous. We have some measure on triangle-free graphs, and we want to understand the probability that a given edge is occupied (i.e., included in the graph). If this edge is occupied, then it had better not have any path of length 2 sitting on top of it, because then it's blocked (adding it would create a triangle).



And if the density of our graph is q, the heuristic is that the probability this edge is unblocked should be $(1-q^2)^{n-2}$ (since there's $n-2$ choices for the top vertex of the 2-path, and we'd expect each 2-path to be present with probability q^2). And if this edge is unblocked, we put it in with probability $\frac{\lambda}{1+\lambda}$. And the probability this edge gets put in is a heuristic for the density itself, so we get that

$$\frac{\lambda}{1+\lambda} \cdot (1-q^2)^{n-2} = q.$$

This gives a heuristic for $\mathbb{E}_{\mathcal{H},\lambda}[|I|]$, and integrating it (according to (2.1)) gives the complicated formula in Theorem 1.9.

> **Remark 3.10.** The authors don't expect $\sqrt{e/2}$ to be the actual cutoff for the phase transition: They also determine $f(c)$ for large c, and it looks kind of like a straight line. And if you analytically continue these two curves and see where they meet, they meet at around 4.3, so maybe that's the truth. The idea is that this tree isn't actually the triangle-free hypergraph — the triangle-free hypergraph has a lot more structure — but our methods are limited by the phase transition on the tree.