PIMS 2025 – Topics in Percolation

Lecturer: Professor Tom Hutchcroft

Notes by: Andrew Lin

June 2025

1 June 2, 2025

This course is about **critical phenomena in percolation** – the basic idea is that we'll be interested in phase transitions (where we have a large system with interacting components, where at some value of parameters there is a stark difference in behavior). Phase transitions occur across various fields of probability, and many of these systems tuned to the exact point of criticality have interesting and mathematically rich things happening. We'll talk about percolation, but many results (random fractal geometry or power-law scaling) will be similar to a range of other applications.

Definition 1

In **Bernoulli bond percolation** on a graph $G = (V, E)$ (which we assume for now to be countable and locally finite, meaning that all vertices have finite degree), we keep or delete each edge independently at random with some **retention probability** p. We write the law as $\mathbb{P}_p$ and associated expectations as $\mathbb{E}_p$. Call the edges that are kept **open** and the edges that are deleted **closed**, and call the connected components **clusters**.

We're generally interested in the geometry of the clusters after our random thinning, and the case $\mathbb{Z}^d$ (that is, $V = \mathbb{Z}^d$ and $x \sim y$ if $||x - y||_1 = 1$) will be one of our main cases of interest. The first basic theorem, which essentially underlies the interest in the rest of the model, says that a phase transition does exist:

Definition 2

Define

$$p_c(G) = \inf\{p : \mathbb{P}_p(\text{infinite clusters exist}) > 0\}.$$

Notice in particular by Kolmogorov's zero-one law, this probability being positive is equivalent to actually being one. In one dimension we have $p_c(\mathbb{Z}) = 1$ (otherwise all components will be finite, since we stop the origin from being in an infinite cluster as long as at least one edge is deleted on the left and on the right), but in any other case we do get something interesting:

Theorem 3

For all $d \geq 2$, we have $p_c(\mathbb{Z}^d) \in (0, 1)$.

This result is essentially due to Rudolf Peierls' work on the Ising model, and there will turn out to be various parallels with other statistical mechanics models as well.

Remark 4. *Notice that the set of p for which $\mathbb{P}_p$(infinite clusters exist) is positive is indeed an interval. This comes from the* **standard monotone coupling** *– for any $p_1 < p_2$, we can couple the two measures so that the p_1-configuration is contained in the p_2-configuration by sampling the p_2-configuration and then further flip a $\frac{p_1}{p_2}$-coin among the remaining open edges. This means that any increasing event has increasing probability in p, so in particular once we've above p_c we will indeed have infinite clusters with probability one.*

To prove this theorem, we'll need to check that $p_c > 0$ and that $p_c < 1$.

Lemma 5

Suppose G has max degree Δ. Then $p_c(G) \geq \frac{1}{\Delta - 1}$.

We will see later that we get equality in this inequality for regular trees.

Proof. We'll do a path-counting argument; we can assume that G is connected. If infinite clusters exist almost surely, then there is some $0 \in V$ which belongs to an infinite cluster with positive probability. On such an event, there is an infinite open simple path from 0 with positive probability (this is König's lemma). Thus if we can show the existence probability of an infinite open simple path is zero at p, then we must be below p_c.

By Markov's inequality, we have

$$\begin{aligned}\mathbb{P}(\text{existence of open simple path of length } n \text{ from } 0) &= \mathbb{E}_p[\text{open simple path of length } n \text{ from } 0] \\ &= p^n \#\{\text{open simple paths of length } n \text{ from } 0\} \\ &\leq p^n \Delta(\Delta - 1)^{n-1},\end{aligned}$$

since we can't go back and revisit a vertex. In particular, if $p < \frac{1}{\Delta - 1}$ this decays exponentially to zero, as desired. □

We can get sharper lower bounds on p_c by finding better combinatorial ways to bound the number of paths, though that's not guaranteed to actually get us the right answer. So the point is that $p_c > 0$ is easy from a counting argument and nothing about the geometry of the graph was necessary.

But proving that $p_c < 1$ will require something more since it's not true in one dimension. And it's not true that it's just about dimension – on a cylinder we have degrees identical to $\mathbb{Z}^2$ but still $p_c = 1$.

Peierls argument for $p_c < 1$. We'll assume $d = 2$, since the square grid is a subgraph of the larger grids and thus having $d > 2$ only makes it easier to get an infinite cluster. The key observation is that if the origin is not in an infinite clusters, then there is a well-defined outer boundary consisting entirely of closed edges. By **planar duality**, we can take the edges in the dual which are orthogonal to the closed edges on our outer boundary; this yields a loop enclosing the origin. So the probability that 0 is in an infinite cluster is

$$\mathbb{P}(0 \not\leftrightarrow \infty) \leq \mathbb{E}_p\left[\text{closed cycles surrounding } 0\right],$$

and summing this over the length of the closed cycles in consideration we find

$$\mathbb{P}(0 \not\leftrightarrow \infty) \leq \sum_{n=1}^{\infty} (1-p)^n \#\{\text{cycles surrounding } 0 \text{ of length } n\}.$$

But we can bound the number of cycles here by n (the place where we intersect the positive real axis, which we can think of as the starting point) times 3^n (the number of choices at subsequent steps). So if p is close enough to 1, this expectation is less than 1, so with positive probability there are no such cycles and thus we do have an infinite cluster. □

There are many interesting things we can study from here (the geometry of the infinite cluster, for example), but we'll mostly care about $p \approx p_c$. We'll mention what the conjectures are:

Conjecture 6

At the critical probability p_c, there are no infinite clusters for all $d \geq 2$.

This is actually known to be true for $d = 2$ (due to Harris and Kesten) and also for large d (by Hara and Slade); in fact, it's true that $p_c(\mathbb{Z}^2) = \frac{1}{2}$ because of planar self-duality. In general, $d = 2$ turns out to be particularly approachable because of special topological properties and also some tools from complex analysis, and high d becomes relatively simple for high enough dimension for reasons that we'll mention later. And "large d" here means that $d \gg 6$, and we'll see later on why 6 is the right threshold (the problem is still open for $d \in \{3, 4, 5, 6\}$).

Remark 7. *Note that it is actually quite atypical to be able to calculate p_c – anything we want to prove about critical percolation will need to be done using the implicit definition as the place where the phase transition occurs, rather than using any sort of formula. And the expectation is that p_c will be some number that's not likely to be algebraic (because there's no reason for that to be the case).*

We want to understand the intricate fractal-like geometry and other properties at p_c quantitatively rather than qualitatively. The idea is that we will be relatively quite easy to get larger finite clusters, and there are two closely related main ways of studying that kind of fractal geometry: **critical exponents** and **scaling limits**. Critical exponents tend to reflect the power-law scaling of quantities of interest, and these are things we should remember the conventional names of (because it helps us identify the same feature exponents across different models):

Definition 8

The **two-point function** is the probability $\tau_{p_c}(x, y) = \mathbb{P}_{p_c}(x \not\leftrightarrow y)$.

In most settings, we can't even prove that these exponents exist without finding their values (and so everything is still wildly open). The conjecture is that

$$\mathbb{P}_{p_c}(x \not\leftrightarrow y) \approx ||x - y||^{-d+2-\eta},$$

where $\eta = 0$ corresponds to Green's function scaling for simple random walk (and in fact in high dimensions η will reduce to 0). We're using $\approx$ to be vague; sometimes $A \approx B$ means $\frac{\log A}{\log B} \to 1$ but sometimes it's just a general statement.

For two other exponents, let K be the cluster containing 0. We then expect the size of the cluster to satisfy (we call this a **volume tail**)

$$\mathbb{P}_{P_c}(|K| \geq n) \approx n^{-1/\delta}.$$

We also get an expression for the **susceptibility** (which is finite if and only if $p < p_c$ – it was proven in the 1980s that finiteness happens at the same point as p_c) near the criticality point:

$$\mathbb{E}_{p_c-\varepsilon}[|K|] \approx \varepsilon^{-\gamma}.$$

We'll just mention one more now, which is the probability of a fixed vertex being part of an infinite cluster

$$\mathbb{P}_{p_c+\varepsilon}(0 \leftrightarrow \infty) \approx \varepsilon^{\beta}.$$

The other way we can understand critical percolation is to take a limit of the whole model (in some sense) as the lattice spacing goes to 0. The most classical case of this is Brownian motion as a scaling limit of simple random walk, but there can be more exotic limits as well. (And often we can extract values of critical exponents from scaling limits if we are able to construct them.)

Example 9

We'll now move to the high-dimensional case; as usual, very similar things are believed to be true in other models, except potentially with different numbers. The point is that once we get to large enough dimensions, we get "mean-field critical behavior," where the lattice is so spacious that it acts like the infinite-dimensional case rather than feeling the geometry.

This is easier to explain in the context of **self-avoiding walk**, in which (forgetting percolation) we want to pick a uniform **simple** path of length n from 0. Self-avoiding walk is basically simple random walk with some self-intersection condition, but if d is large enough (specifically $d > 4$) then almost surely two independent simple random walks only intersect finitely often. In fact Brownian motion in $d \geq 4$ is a simple path, so we're really not getting the effects of interaction at that point. On the other hand, the probability of a randomly-chosen random walk being self-avoiding is still exponentially decaying (because we always have to avoid the previously-traveled edge), so there is some subtlety in actually making the argument work – we're trying to say that self-avoidance is a purely local property, whatever that means, and thus "self-avoiding walk looks like simple random walk." This can actually be proven by the **lace expansion** (by Brydges and Spencer, then Hara and Slade), showing that the self-avoiding interaction only contributes some kind of constant rescaling. And correspondingly, things like the two-point function behave like their analogous objects for $d > 4$. It's conjectured that for $d < 4$, there are different critical exponents and scaling limits ($\text{SLE}_{8/3}$ for $d = 2$, for example), and at $d = 4$ we expect the same scaling limits as $d > 4$ but with some log factor, meaning that instead of $\left(\frac{1}{\sqrt{n}} X_{\lfloor nt \rfloor}\right)$ converging to Brownian motion, we have $\left(\frac{1}{\sqrt{n}(\log n)^{1/8}} X_{\lfloor nt \rfloor}\right) \to \text{BM}$ instead.

Returning back to percolation, then, we should think of breaking up percolation into "branching random walk along with some self-interaction." In words, this can be explained as follows. If we want to do a simulation of percolation on our computer, a natural way to do that is a breadth-first search at the origin (look at the edges coming out of the origin and check if they're open or closed, open up new vertices, and then look at the edges adjacent to those, and keep doing this process). So we always have some front of active vertices, and it's important that we can't reuse vertices or edges if we've already checked that they're open or closed. If we forget to include that constraint, we generate a branching random walk (which is an evolving cloud of particles with rules for creating offspring), and the only modification we need to make to be correct is that "attempting to go back to sites we've already visited destroys the particles." Unfortunately it's rather difficult to prove anything with this framework, and in fact the critical parameters don't agree between percolation and branching random walk – it's really just useful as a heuristic for what is happening.

So for d large, the point is that we expect the same large-scale critical behavior as branching random walk: for example, the expected number of particles visiting y from x in the branching walk is exactly the Green's function, and we can think of this as first picking the genealogy tree and then embedding it uniformly at random. But the expected number of particles alive at time n in the branching random walk can be tuned to be 1 at "criticality," and so the two-point function in critical percolation should also behave like the Green's function

$$\tau_{p_c}(x, y) \asymp ||x - y||^{-d+2}$$

(where $\asymp$ means "up to constants"); in particular $\eta = 0$. And when studying the tail of the cluster size, we can just think about the genealogy tree and not worry about the embedding at all – we expect

$$\mathbb{P}_{p_c}(|K| \geq n) \asymp \frac{1}{\sqrt{n}}$$

(so $\delta = 2$) and then summing via a geometric series,

$$\mathbb{E}_{p_c-\varepsilon}(|K|) \approx \frac{1}{\varepsilon}$$

(so $\gamma = 1$). Similarly we expect $\beta = 1$ and so on, and the point is that we understand things quite well often to the point of having theorems. With some caveats, volume and susceptibility estimates have been proven using the lace expansion by Hara, Slade, and van der Hofstad building on earlier work of Aizenman–Newman and Barsky–Aizenman. So we can prove this kind of thing in general by having a non-interacting model which doesn't do anything interesting, plus some additional interaction.

Fact 10

We've been saying that self-avoiding walk has "high dimension" at $d > 4$, since we try to intersect two simple random walks and $4 = 2 + 2$. In general, the point is that an a-dimensional object and a b-dimensional object will not intersect a lot if $a + b > d$, and it turns out the critical dimension for percolation is 6. This is a bit harder to understand than for self-avoiding walk (and we'll return to it later).

What we're actually able to prove with the lace expansion (which is a perturbative method requiring a certain parameter to be small for inclusion-exclusion to work) is for some certain numerical assumptions to hold; for the original paper it was proven for $d \geq 19$. This has since been optimized with computer assistance to $d \geq 11$ by Fitzner and van der Hofstad, but perhaps the better way to think about this is that in the spread out lattice $\mathbb{Z}_L^d$ where the edge set is the points $||x - y||_1 \leq L$, then for $d > 6$ and L sufficiently large the lace expansion will always work. And it would be a major advancement to find other strategies that don't have this necessary caveat.

Remark 11. *There's also a scaling limit called "super-Brownian motion," which is the same limit as for branching random walk. But understanding proofs for that relies on understanding the lace expansion, which we won't get into much in this course*

Finally, we'll mention what the conjectures look like at the critical $d = 6$, which is the first part where finite-dimensional effects start to become noticeable through logarithmic factors. Physicists (Essam, Gaunt, Guttmann) used heuristic methods in the 1970s to predict

$$\mathbb{P}_{p_c}(|K| \geq n) \asymp \frac{(\log n)^{2/7}}{\sqrt{n}}, \quad \mathbb{E}_{p_c-\varepsilon}[|K|] \asymp \frac{(\log 1/\varepsilon)^{2/7}}{\varepsilon}, \quad \tau_{p_c}(x, y) \asymp ||x - y||^{-d+2}(\log ||x - y||)^{1/21}.$$

Of course, it's very difficult to check these things numerically, but we'll talk next time about "why we can't prove this."

2 June 3, 2025

Last time, we were talking about the conjectured critical behavior of percolation on $\mathbb{Z}^d$, and we'll continue that story today. We mentioned some critical exponents for the two-point function, volume tail, and susceptibility, and the density of the slightly-supercritical cluster, and we mentioned that in high dimensions (for $d > 6$ above the "upper critical dimension"), we should expect $\eta = 0, \delta = 2, \gamma = 1, \beta = 1$. We do also understand the low-dimensional case $d = 2$ quite well; there's a rich theory developed over the years, and it's known specifically only for **site percolation on the triangular lattice** (meaning random vertices are percolated rather than edges)

$$\delta = \frac{91}{5}, \quad \beta = \frac{5}{36}, \quad \gamma = \frac{43}{18}, \quad \eta = \frac{5}{24}$$

These values were first conjectured in non-rigorous contexts in physics and then proven in the 2000s. There aren't actually that many degrees of freedom in these exponents, at least conjecturally – they're expected to satisfy various **scaling relations** (and hyperscaling relations in low dimensions). So the story was first that the scaling relations were proved in the 1980s by Kesten (in other words, that if some of the exponents are well-defined, then others are also well-defined and have the corresponding values). Then Schramm developed the theory of SLE, which is a family of conformally invariant random curves which are the only such curves satisfying certain axioms. So proving that 2D percolation has a scaling limit implies that it must be one of these SLE curves.

Fact 12

We've been talking about studying percolation on $\mathbb{Z}^d$, and we expect that the critical behavior depends quite strongly on d but not actually on the specific lattice. This is the phenomenon of **universality** – it's believed that small-scale effects (like square versus triangular grid) will not change the overall behavior, and that's a guiding principle but not yet a theorem. That being said, many of the strategies in high dimensions are indeed insensitive to the exact lattice and thus in some sense universal.

Note that it is true that the choice of lattice often changes the actual critical point itself – in fact it's determined often by microscopic details. So somehow p_c is a bad quantity to look at because it's not very universal, and it's useful to intsead think about large-scale asymptotic behavior at criticality rather than caring about where that criticality actually is.

Conformal invariance is tied up with universality as well, and so is the idea of **emergent solutions** – if we consider simple random walk on the square or triangular grid, then rotating the lattice by 90 or 120 degrees respectively doesn't change the distribution of the walk, and in fact Brownian motion can have an arbitrary rotations. So the same story happens with percolation – we have invariance under scaling, rotations, and translations, and we also have **conformal invariance** (which locally around any point look infinitesimally like some combination of those three types of transformations, rather than having stretches or shears). And that's very powerful in two dimensions because we can use the Riemann mapping theorem to get conformal maps, and that's where the theory of SLE comes up. Conformal invariance should still be true in higher dimensions, but we have fewer maps (only the **Möbius transformations**, which generally are generated by scaling, rotations, translations, and inversion $x \mapsto \frac{x}{||x||^2}$) and thus the technique is less powerful for current mathematical tools. (The physicists do manage to get things out of this via the **conformal bootstrap** in settings like the 3D Ising model, though.)

Remark 13. *The fact that we get nice rational numbers for $d = 2$ is a very two-dimensional phenomenon; it's tied up with special structure arising from the ability to do complex analysis, and deep behind this is representation theory of the Virasoro algebra. There's no reason to believe that we should also get nice numbers for $d = 3, 4, 5$ in the absence of evidence either way. (And physicists are able to compute critical exponents in some settings to high accuracy, and often those numbers don't look particularly nice.)*

We'll now turn back to the conjecture about the critical case $d = 6$. We mentioned last time that we expect log corrections to the two-point function, volume tail, and susceptibility; tied up with this is the story of intermediate dimensions $d = 3, 4, 5$. We don't really have any idea what the scaling limits should be or how to begin to describe them, and the exponents presumably take very complicated values (we're not even able to express them as infinite series).

The traditional approach in the physicist literature has been to do the **ε-expansion**, introduced by Wilson and Fisher in a paper called "Critical exponents in 3.99 dimensions." The idea is that we take $d = d_c - \varepsilon$ slightly below the critical dimension and try to understand how things change, and the question of "does this d make sense" turns out to be the least significant of the worries here. (We could try doing percolation on the Sierpinski carpet or something, but this isn't quite the right setting because there are various senses of dimension for those fractals.) The conjecture, not precisely stated, is the following:

Conjecture 14

We have in the expansion $d = d_c - \varepsilon$ that

$$\eta = 0 - \frac{1}{21}\varepsilon - \frac{206}{9261}\varepsilon^2 + \cdots, \quad \delta = 2 + \frac{2}{7}\varepsilon + \frac{565}{6174}\varepsilon^2 + \cdots$$

and similar expressions for the other exponents.

Even within this non-rigorous literature, it's not really that we can compute these infinite series – the coefficients get extremely complicated the further we get, and for percolation currently the best results are up to something like ε^4. In fact, the expressions are "asymptotic series" in that they are conjecturally divergent series but have bounded error if we cut it off at any specific power of ε. But the point is that the exponents $\frac{1}{21}, \frac{2}{7}$ also appear in our conjectured powers of log for the critical dimension; this encoding is expected to be a general phenomenon. It seems not realistically on the horizon to use this to prove anything for $d = 3$, for example, but if we plug in $\varepsilon = 1, 2, 3$ and take the first few terms, it does appear to give a pretty reasonable approximation of our numerical simulations.

Remark 15. *To understand why the critical dimension is expected to be an integer, we expect Green's function scaling of two-point functions after a certain point, and we can construct certain "triangle diagrams" for percolation or "bubble diagrams" for Ising for which the criticality occurs at an even integer. But for an arbitrary model, we can probably make the critical dimension whatever we want, and we'll see that with some long-range models later.*

One strategy we do have is the **renormalization group technique** first introduced by Wilson; it's really a vague class of techniques (and it's not actually a group in the mathematical sense) which helps us systematically understand critical phenomena. Theorems very much like our critical exponent conjectures have been proven using rigorous renormalization techniques, and they've been most successful in a framework where we have a "Gaussian field plus some weak self-interaction." (This has been done rigorously for the **ϕ^4 model** and for the **continuous-time weakly self-avoiding walk**, both by Bauerschmidt, Brydges, and Slade.) The point is that this is really something about fields and random functions, and we want to understand how interactions manifest as a dynamical system as we rescale our system.

Fact 16

According to physicists, percolation "should be related" to some ϕ^3 model, but it's not quite clear how to construct such a field theory. And there's also connections to the **Potts model**, which is a generalization of the Ising model with more colors, in that percolation is supposed to be a "one-color version of Potts." So we can do analysis for Potts for general integer $q \geq 2$ and then substitute in $q = 1$, and it turns out this gives the same answer as the ϕ^3 method and thus is very likely to be correct. But of course it doesn't appear to be the basis of a proof for now.

One of the broad goals of this course is to do something akin to renormalization directly from percolation without passing to spin systems, instead staying within the realm of classical probability theory. As it stands, it only really works for long-range models, but we'll see how the techniques play out.

Example 17

We'll bring this back down to earth now; we'll start with the simplest thing, which is to understand the critical behavior of branching processes. For this, we'll understand **percolation on a binary tree**; everything extends to any branching process with finite variance, but we'll stick to the example that keeps things a bit simpler. (Though if the offspring has infinite variance, we can get different behavior.)

Percolation on a tree is much easier than percolation on a grid, because to study the two-point function we only have to look at the unique path connecting the two vertices:

$$\tau_p(x, y) = p^{d(x,y)},$$

so in particular we get exponential decay and won't see the power law scaling. But we'll foreshadow some analytic methods coming up in more complicated models, and then after that we'll show a much easier combinatorial approach.

The first idea is to introduce the concept of a **ghost field**, which is a jargony way of describing something very simple:

Definition 18

A **ghost field** of intensity $h \geq 0$ is a Poisson process on the vertex set independent of the percolation configuration. Specifically, ξ is a random subset of V such that $v \in \xi$ with probability $1 - e^{-h}$, and thus $\mathbb{P}(A \cap \xi \neq \varnothing) = 1 - e^{-h|A|}$ for all subsets A.

Ultimately, it's a way of talking about Laplace transforms if we think about things that way; it'll let us convert questions about "how big is the cluster at the origin" to "probability of connecting to a ghost field of intensity $\frac{1}{n}$." Indeed, the probability of connecting to this field ξ is therefore

$$\mathbb{P}_{p,h}(0 \iff \xi) = \mathbb{E}_p\left[1 - e^{-h|K|}\right],$$

and we will call this constant $\theta(p, h)$. Letting $\theta(p) = \mathbb{P}_p(|K| = \infty)$ be the probability of having an infinite cluster, we then have (because $e^{-x} \approx 1 - x$)

$$\theta(p) = \lim_{h \downarrow 0} \theta(p, h).$$

We'll thus understand percolation on the binary tree by setting up a functional equation for these ghost field connection problems. We have

$$\theta(p, h) = \left(1 - e^{-h}\right) + e^{-h}\left(2p(1-p)\theta(p, h) + p^2(1 - (1 - \theta(p, h))^2)\right),$$

since $1 - e^{-h}$ is the probability that the root is in ξ, and then otherwise $2p(1-p)$ is the probability of the root retaining one edge and p^2 the probability of retaining two edges. Doing the quadratic equation yields

$$\theta = \frac{2e^{-h}p - 1 \pm \sqrt{1 - 4e^{-h}p + 4e^{-h}p^2}}{2e^{-h}p^2};$$

in the special case $h \to 0$ (which indeed corresponds to being connected to the infinite cluster) this yields $\theta(p, 0) \in \{0, \frac{2p-1}{p^2}\}$. Of course, if $p \leq \frac{1}{2}$ this means $\theta(p, 0) = 0$ (since probability can't be negative), and it turns out that if $p > \frac{1}{2}$ and h is positive, actually we need to take $\theta(p, h)$ to be the one with the positive root in the square root because the other one will be negative. Thus we do have $\theta(p, h) = \frac{2e^{-h}p - 1 + \sqrt{1 - 4e^{-h}p + 4e^{-h}p^2}}{2e^{-h}p^2} \implies \theta(p) = \frac{2p-1}{p^2}$, so putting things together, we get that on the binary tree that

$$\boxed{\theta(p) = \frac{2p-1}{p^2} 1\left\{p \geq \frac{1}{2}\right\}}.$$

This tells us the following facts:

Theorem 19

For percolation on the binary tree, we have $p_c = \frac{1}{2}$ and $\theta(p_c + \varepsilon) \sim c\varepsilon$. Thus we have the critical exponent $\beta = 1$.

We can also understand the susceptibility quite well, since we know exactly the formula for the two-point function at each "level:" there are 2^n vertices at distance n from the root, so

$$\mathbb{E}_p[|K|] = \sum_{n=0}^{\infty} p^n 2^n = \frac{1}{1-2p} \text{ if } p < \frac{1}{2},$$

and indeed this is finite up until the critical point of $\frac{1}{2}$ and is in fact asymptotic to $\frac{1}{2\varepsilon}$ for $p = p_c - \varepsilon$ (so $\boxed{\gamma = 1}$).

Next we'll do possibly an unnecessarily complicated strategy for computing δ. One way to understand volume-tail exponents is to understand the behavior of the Laplace transform: we have

$$\theta\left(\frac{1}{2}, h\right) = \frac{2(e^{-h} - 1 + \sqrt{1 - e^{-h}})}{e^{-h}},$$

which is asymptotic to $2\sqrt{h}$ as $h \downarrow 0$. This square root actually already tells us that $\boxed{\delta = 2}$, and we can in fact define the critical exponent that way as well; let's explain the analytic theory for that now.

The general strategy goes under the name **Karamata's Tauberian theorem** – each of the various **Tauberian theorems** gives us conditions under which we can extract asymptotics of a function from its Laplace transform, or a sequence from its generating function. (Going in the other direction is called **Abelian theorems**.) Everything we talk about here can also be made sense of for sequences.

Definition 20

A **regularly varying function** is a function $f : (0, \infty) \to (0, \infty)$ such that $\lim_{x \to \infty} \frac{f(\lambda x)}{f(x)}$ exists for all $\lambda > 0$.

For example, $f(x) = x^\alpha$ always satisfies this condition, since in fact this ratio $\frac{f(\lambda x)}{f(x)}$ is constant and equal to λ^α. That turns out to be a general phenomenon:

Theorem 21

If f is measurable and regularly varying, then

$$\lim_{x \to \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\alpha$$

for some $\alpha \in \mathbb{R}$ called the **index of regular variation**, and convergence is uniform on compact sets $\lambda \in (0, \infty)$.

We say that such a function is **slowly varying** if $\alpha = 0$. Notice that we can always take any $f(x)$ of index of regular variation α and get a slowly varying $x^{-\alpha} f(x)$ – anything that grows subpolynomially on every large scale will work. (For example, $x^\alpha(\log x)^2$ has index of variation α, but something like $x^\alpha(2 + \sin x)$ is not regularly varying because of the oscillation on every scale.)

Theorem 22

Suppose f is a regularly varying function. Then we can write

$$\int_1^x f(y)dy = \int_1^x y^\alpha \ell(y)dy$$

for some slowly varying function ℓ. If $\alpha > -1$, then this integral is asymptotically $\ell(x)\int_1^x y^\alpha dy = \frac{\ell(x)x^{\alpha+1}}{\alpha+1}$, and if $\alpha < -1$, then this is asymptotically $\ell(x)\int_x^\infty y^\alpha dy = -\frac{\ell(x)x^{\alpha+1}}{\alpha+1}$. (But if $\alpha = -1$, then things are a bit more subtle.)

Theorem 23 (Karamata)

Suppose $0 \le \alpha < 1$ and f is (measurable and) regularly varying with index $-\alpha$. Let X be any random variable taking values in $[0, \infty)$. Then the tail $\mathbb{P}(X > x)$ of X is asymptotic to $f(x)$ as $x \to \infty$ if and only if $\mathbb{E}[1 - e^{-hX}] \sim f\left(\frac{1}{h}\right)\Gamma(1-\alpha)$ as $h \to 0$.

Here $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ is the gamma function (and in particular $\Gamma(\frac{1}{2}) = \sqrt{\pi}$). So the Tauberian theorem is saying that regularly varying functions have the same answer of converting between the function and its Laplace transform as the corresponding power law, but it's only a first-order asymptotic property. We'll apply this to $f(x) = 2\sqrt{x}$, since $\theta(\frac{1}{2}, h) = 2\sqrt{h}$, and that yields the following:

Corollary 24

The tail of the cluster for percolation on a binary tree satisfies

$$\mathbb{P}_{p_c}(|K| \ge n) \sim \frac{2}{\sqrt{\pi n}}.$$

And next time, we'll see a way to get this answer which doesn't require all of this Tauberian machinery.

3 June 5, 2025

Last time, we were studying the critical behavior of percolation on the binary tree – we saw that $p_c = \frac{1}{2}$ and found ways of studying behavior around criticality by thinking about the Laplace transform

$$\theta(p, h) = \mathbb{E}[1 - e^{-h|K|}] = \mathbb{P}_{p,h}(0, \xi)$$

as a probability of being connected to a ghost field. In particular, we found that $\theta(\frac{1}{2}, h)$ is asymptotic to $2\sqrt{h}$ as $h \downarrow 0$, and so we can use Karamata's Tauberian theorem to get a bound on the tail of the volume (it's asymptotic to $\frac{2}{\sqrt{\pi n}}$, and so $\delta = 2$). But "this is overkill" and we don't really need such precise results; the following exercise suffices:

Proposition 25

If $\theta(p_c, h) \asymp h^{1/\delta}$, then $\mathbb{P}_{p_c}(|K| \ge n) \asymp n^{-1/\delta}$.

We'll see a combinatorial interpretation of all of this later; it turns out to be easier but only if we assume some standard combinatorial facts which are rather difficult to prove.

Instead, what we'll discuss is the tail of the diameter. Let Q_n be the probability at criticality that the origin is connected to height n of the tree; we want to understand the power-law asymptotics of this new quantity. This is done by establishing a recurrence, since for all $n \geq 1$ we have

$$Q_n = \frac{1}{2}Q_{n-1} + \frac{1}{4}(1 - (1 - Q_{n-1})^2) = Q_{n-1} - \frac{1}{4}Q_{n-1}^2,$$

since $\frac{1}{2}, \frac{1}{4}$ are the probability of having one and two open children. To relate to more complicated situations, we'll analyze this recurrence in a way "tolerant of errors" (since we might only know terms at a $1 + o(1)$ level). It's often useful to take an equation like $Q_n - Q_{n-1} = -\frac{1}{4}Q_{n-1}^2$ and think about what the equivalent differential equation would be (not in the sense of generating functions, just as a first approximation) – morally, this should look like the equation $f' = -\frac{1}{4}f^2$. By separation of variables this yields $f = \frac{4}{x+C}$ for some constant C, and in fact we should think of this as saying that if we had an error $\frac{df}{dx} = \left(\frac{1}{4} + \varepsilon(x)\right) f^2$, we could still integrate $\left(\frac{1}{f}\right)' = -\frac{f'}{f^2} = \frac{1}{4} + \varepsilon(x)$ and understand how errors propagate. But returning to the recurrence, it suggests that studying $\frac{1}{Q_n}$ may be a useful perspective: we get that

$$\frac{1}{Q_n} - \frac{1}{Q_{n-1}} = \frac{Q_{n-1} - Q_n}{Q_n Q_{n-1}} = \frac{1}{4}\frac{Q_{n-1}}{Q_n}.$$

We know that at criticality we don't have infinite clusters, so $Q_n \to 0$ as $n \to \infty$, and our recurrence relation means additionally that $Q_n \sim Q_{n-1}$ (the difference is going to zero faster than the size of the terms). So the right-hand side is asymptotic to $\frac{1}{4}$, and writing as a telescoping series we get

$$\frac{1}{Q_n} - \frac{1}{Q_0} = \sum_{k=1}^{n} \frac{1}{Q_k} - \frac{1}{Q_{k-1}} \sim \frac{n}{4},$$

so this yields $\boxed{Q_n = \frac{4}{n}}$. So this basically ended up being the same strategy as solving the ODE but in the discrete setting, and it gives us another critical exponent. In general this doesn't have a standard notation but is called the **intrinsic one-arm exponent** (intrinsic versus extrinsic is a question of whether we look at distances in the cluster or on the graph, but that's the same on a tree).

Example 26

Returning to the combinatorial interpretation now, the probability $\mathbb{P}_p(|K| = n)$ can be calculated as follows. Any valid tree of n vertices including the origin is specified by having $(n-1)$ specific edges be open and $(n+1)$ specific edges be closed, so each tree occurs with probability $p^{n-1}(1-p)^{n+1}$. (On a more general graph, this isn't as nice, so this exact computation is special to trees.)

Thus we get the overall probability to be

$$\mathbb{P}_p(|K| = n) = C_n p^{n-1}(1-p)^{n+1},$$

where C_n is the number of n-vertex connected subtrees of binary trees. (In particular, this even tells us the exact behavior away from p.) It turns out C_n are the Catalan numbers

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

This can be computed by the recurrence relation $C_n = \sum_{k=1}^{n-1} C_{k-1}C_{n-k}$ (the casework here is the number of vertices in the left and right subtrees, respectively). Solving the recurrence relation using generating functions, we end up with a very similar story as how we compute $\theta(p, h)$ (again we have a quadratic equation and have to pick one of the roots appropriately). But the point is that by Stirling's approximation,

$$C_n = \frac{1}{n+1}\binom{2n}{n} \sim \frac{4^n}{n^{3/2}\sqrt{\pi}},$$

and so plugging in $p = \frac{1}{2}$ and taking the tail by summing over all $m \geq n$ yields the same answer as before.

Remark 27. *The Catalan numbers come up repeatedly across various subjects; they'll also come up later in some random walk calculations and this is not a coincidence. In fact, going further makes use of these bijections between binary walks and excursions.*

Thinking more about this idea of going "away from criticality," we can now for example take $p = \frac{1}{2} + \varepsilon$ and look at the probability

$$\mathbb{P}_p(|K| = n) \approx \frac{1-p}{p} \cdot p^n(1-p)^n 4^n \frac{1}{n^{3/2}} \approx \frac{1-p}{p}\frac{1}{\sqrt{\pi}}\frac{1}{n^{3/2}}e^{-(4\varepsilon^2+o(\varepsilon^2))n}$$

(we're going back to $\approx$ instead of $\sim$ because it's not entirely specified what regime we're taking with multiple variables). And now if $\varepsilon < 0$, we see that the tail of the cluster satisfies $\mathbb{P}_p(|K| = n) \asymp \frac{1}{\sqrt{n}}e^{-\Theta(\varepsilon^2 n)}$, and so we start seeing the decay away from p around $\varepsilon = \frac{1}{\sqrt{n}}$. And notice that this also gives a duality between the subcritical and supercritical cases – we get the exact same scaling on the two sides of the critical point! (Indeed, this is reasonable because above the critical point we still have some exponentially decaying probability of a large finite cluster.) But in a more general graph like $\mathbb{Z}^d$ things are much more complicated.

Definition 28

Let $\zeta(p)$ be the "typical large size cluster" (it's best to actually have different definitions in mind for what this means, depending on the theorem we're trying to prove, and ideally show that all of those are equivalent).

It turns out on our binary tree the right definition to make is $\zeta(p) = |p - p_c|^{-2}$, and then we find that for $p < p_c$,

$$\mathbb{P}_p(|K| \geq n) \approx \frac{1}{\sqrt{n}}e^{-n/\zeta(p)}.$$

When we do heuristic arguments, we usually assume that statements like this are true to prove relations between critical exponents. The question we tend to ask along these lines is "what does a large critical cluster look like?"; one thing we've already seen is that

$$\mathbb{P}_{p_c}(|K| \geq n) \asymp \frac{1}{\sqrt{n}}, \quad \mathbb{P}_{p_c}(\text{diam}(k) \geq r) \asymp \frac{1}{r},$$

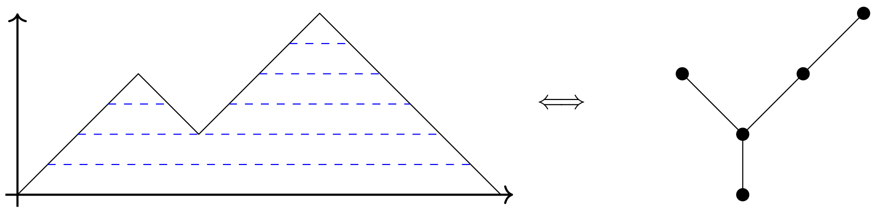
which suggests that the diameter behaves like the square root of the volume (and that somehow "clusters are of dimension 2"). It turns out we can actually understand the large clusters in an extremely detailed way, and we'll give a brief overview of that.

Example 29

We'll switch to a different model to make the combinatorics a little simpler (this part is for motivation). Consider branching processes (Galton-Watson trees) with **geometric** offspring distribution instead of having Bin(2, $\frac{1}{2}$ as in binary trees. (The analysis carries through in all cases but just with different variance constants.)

The nice thing about branching processes is that the critical value always occurs when the mean is 1, and if we condition on such a **geometric** branching process having $|T| = n$, we get a uniform random **plane tree** with n vertices, meaning that we know the order in which the children of a given vertex occur. (This turns out to also be enumerated by the Catalan numbers.)

So what we want to understand is the scaling limit of a uniform random plane tree as $n \to \infty$, and here's the idea (this is theory that was developed by Aldous and Le Gall, going under the name of the **continuum random tree**). Consider the set of **Dyck paths**, which are finite paths in $\mathbb{Z}_{\geq 0}$ which take only ± 1 steps from $(0, 0)$ to $(2n, 0)$. These are also enumerated by Catalan numbers, and the bijection is that if we do a depth-first traversal of a binary tree and record the height of the vertices we visit in order, then we end up with a unique Dyck path. To show this is a bijection, we can construct the inverse map by taking a Dyck path, and for each point on the path we identify it with other points at the same height via lines under the path; this collapses us into a tree where each interval of things seeing each other corresponds to an edge. More intuitively, we squish the path together by gluing underneath, and that yields a tree (we can check that this is indeed a bijection).



This is useful for taking scaling limits, because understanding a uniform random plane tree can be done via uniform Dyck paths, and so we can scale Dyck paths instead of trees. Since Dyck paths are simple random walk conditioned to remain nonnegative and return to the origin at a fixed time, the scaling limit should be Brownian motion conditioned to remain nonnegative and return to the origin at a fixed time; this is called the **Brownian excursion**. (We can't actually condition on a measure-zero event, but we can condition on within an ε-interval of time and space and take $\varepsilon \to 0$, and this turns out to be defined.)

Therefore, we can take this random curve and "squash it together;" there are of course issues that arise when we try to do this, but formally we want to again identify points where they can see each other at the same height under the curve. This is a well-defined object (it's a quotient of a topological space), and we want to show that it's a well-defined metric space, which has some subtleties. But for many things, we don't need to do that – we just need to say that if we pick our favorite finite set of points, then we want to understand the pairwise distances among those points (the tree restricted to those points), and that's easy to do. (We just need to look at local minima among the points, and that's where the tree splits.)

But for the original question, this already gives us the basic reasoning: sampling a tree of n vertices corresponds to running a Brownian excursion of length n, which indeed scales to height roughly $\sqrt{n}$.

Example 30

We'll return now to "heuristic scaling theory" and again understand how we might analyze things when we're not yet trying to prove theorems. There's a really important point here that proving theorems is hard, and it's much easier (even if we solely care about proving theorems) to master non-rigorous methods and understand why things ought to be true.

We've so far talked about various critical exponents, and we're at the point in the course where it's still useful to write them down: we have the volume tail $\mathbb{P}_{p_c}(|K| \geq n) \approx n^{-1/\delta}$, two-point function $\tau_{p_c}(x, y) \approx ||x - y||^{-d+2-\eta}$, the percolation probability $\mathbb{P}_{p_c+\varepsilon}(0 \leftrightarrow \infty) \approx \varepsilon^{\beta}$, and the expected cluster size $\mathbb{E}_{p_c-\varepsilon}[|K|] \approx \varepsilon^{-\gamma}$. The point is that these values are not freely independent of each other and do have to satisfy some relations; to figure out what these should be, we'll assume an ansatz (which we don't know is true) and use that as a basis for our computation. For example, we have $\mathbb{P}_{p_c}(|K| \geq n) \approx n^{-1/\delta}$, and we'll assume that some formula like the one we derived for the binary tree is true:

for $p < p_c$

$$\mathbb{P}_p(|K| \geq n) \approx n^{-1/\delta} f\left(n/\zeta(p)\right)$$

for some rapidly decaying function f (like e^{-x}) and $\zeta(p)$ being the "typical size of the large cluster," where the definition here is whatever makes this formula true. It should be that as $p \to p_c$, we have $\zeta(p) = |p - p_c|^{-\Delta}$, and this is typically called the **gap exponent**.

If we assume a formula like the one above holds, and we want to understand the susceptibility, we use the "integration-by-parts" formula

$$\mathbb{E}_p[|K|] = \sum_{n=0}^{\infty} \mathbb{P}_p(|K| \geq n) \approx \sum n^{-1/\delta} f\left(\frac{n}{\zeta(p)}\right).$$

The heuristic (which is usually easy to verify in practice) is that this kind of sum is of the same behavior as just "summing up until the exponential decay kicks in:" thus

$$\mathbb{E}_p[|K|] \approx \sum_{n=1}^{\zeta(p)} n^{-1/\delta} = \zeta(p)^{(\delta-1)/\delta},$$

and thus we get the relation (assuming that $\delta > 1$ – we'll actually prove a bound of 2 later on)

$$\boxed{\gamma = \frac{\delta - 1}{\delta}\Delta}.$$

Of course in this case we had to introduce a new variable to get a new relation, but there are other cases where we don't need to, and furthermore this perspective of the gap exponent is usually useful to take.

And with higher moments, we can do a very similar computation. We similarly have (again this is an "integration-by-parts") $\mathbb{E}_p[|K|^m] \asymp \sum n^{m-1}\mathbb{P}(|K| \geq n)$, and this sum is roughly $\sum_{n=1}^{\zeta(p)} n^{m-1-1/\delta} \asymp \zeta(p)^{m-1+(\delta-1)/\delta}$. So we have

$$\mathbb{E}_p[|K|^m] \approx \mathbb{E}_p[|K|]\zeta(p)^{m-1},$$

and thus we should think of this as a concentration result: this quantity is basically $\zeta(p)^m \mathbb{P}_p(|K| \geq \zeta(p))$, which is of the same order as $\zeta(p)^m \mathbb{P}_{p_c}(|K| \geq \zeta(p))$ (somehow by the time we are of that size, we can't really tell the difference between p and p_c). So the moments scale as $\boxed{\mathbb{E}_p[|K|^m] \approx |p - p_c|^{-\gamma-(m-1)\Delta}}$, and in fact this is often the definition of the gap exponent (the increase of moments as we go from first to second moments). But of course in many of the settings we care about, we don't have a way of proving these relations or that these asymptotics actually hold.

We'll now do a slightly less convincing argument which relies on some symmetry that we did not see happen in the binary tree. If we now take $p > p_c$ and try to compute $\mathbb{P}_p(|K| = n)$ (which should still decay even above criticality), at criticality we know by taking a derivative of the tail that $\mathbb{P}_p(|K| = n) \approx n^{-1/\delta - 1}$. So by the same reasoning as before we'll assume the ansatz

$$\mathbb{P}_p(|K| = n) \approx n^{-1/\delta-1} g\left(\frac{n}{\zeta(p)}\right).$$

We'll now assume that $\zeta(p_c + \varepsilon)$ and $\zeta(p_c - \varepsilon)$ **have the same scaling**. This is somehow saying something about percolation beyond "quantities behaving in a nice smooth manner," but if we take this as an assumption, then for $p > p_c$ we should again see

$$\mathbb{P}_p(n \leq |K| < \infty) \approx n^{-1/\delta} g\left(\frac{n}{\zeta(p)}\right).$$

In other words, we have that the quantities

$$\mathbb{P}_p(|K| = \infty) \approx \mathbb{P}_p(|K| \geq \zeta(p))$$

should be roughly of the same order because that's the point at which we stop seeing contributions on the right-hand side, and at that point of cluster size we should expect that p and p_c look about the same:

$$\mathbb{P}_p(|K| = \infty) \approx \mathbb{P}_{p_c}(|K| \geq \zeta(p)) \approx \zeta(p)^{-1/\delta}.$$

So matching expressions tells us that $\boxed{\beta = \frac{\Delta}{\delta}}$. This is one of the harder ones to turn into a proof, since the symmetry justification is a bit more interesting than just functional regularity considerations.

Remark 31. *These relations are supposed to always hold regardless of the value of d even though the exponents themselves may change for lower dimensoins, though we'll talk about some additional relations later on.*

The next ansatz we'll make is for the two-point function, namely for $p < p_c$

$$\tau_p(x, y) \approx ||x - y||^{-d+2-\eta} h\left(\frac{||x - y||}{\xi(p)}\right)$$

where instead of the volume scale $\zeta(p)$ we now want a length scale $\xi(p)$ which we call the **correlation length**. We should think of this as the Euclidean distance scale at which we start seeing the effect of not seeing criticality (that is, where decay starts exponentially occurring), and it should have some associated critical exponent as well. We write $\xi(p) \approx |p - p_c|^{-\nu}$, and we'll write down some relations involving it. One thing we can do is write out an expectation over all vertices in our graph

$$\mathbb{E}_p[|K|] = \sum_x \tau_p(0, x),$$

and again by similar logic this is like a sum over only the polynomial term up to a characteristic tail, so

$$\mathbb{E}_p[|K|] \approx \sum_{x:||x|| \leq \xi(p)} ||x||^{-d+2-\eta} = \xi(p)^{2-\eta}$$

by summing over shells of a given radius r. But the point is that the relation we get now is $\boxed{\gamma = \nu(2 - \eta)}$. And next time, we'll see how these hyperscaling relations work out for certain d.

Remark 32. *Often mathematicians hear that these kinds of things are "physics arguments," but it's not really physics: we're just assuming that certain quantities are nice. So we shouldn't be afraid of hearing that things are physics arguments – it typically just means heuristic calculations without all of the assumptions verified.*

4 June 6, 2025

Last time, we did an analysis of percolation on the binary tree and got various asymptotics and critical exponents. We also introduced a variety of scaling relations which really aren't really even about percolation – we were assuming various ansatz formulas, but no real properties of percolation were really being used there. Thus these facts $\gamma = \frac{\delta-1}{\delta}\Delta$, $\beta = \frac{\Delta}{\delta}$, $\gamma = (2 - \eta)\nu$ are fairly general and mostly just using the fact that tail probabilities decay rapidly away from p_c. But things like $\zeta(p)$ and $\xi(p)$ can typically have different definitions, and so for example here is one way we could make our definitions precise:

Definition 33

For K the size of the cluster at the origin, define the **gap exponent** via $\zeta(p) = \frac{\mathbb{E}_p[|K|^2]}{\mathbb{E}_p[|K|]}$.

(It could also make sense to define $\zeta(p) = \frac{\mathbb{E}_p[|K|^3]}{\mathbb{E}_p[|K|^2]}$, though we don't actually know that this has the same behavior as the definition above.) This can be interpreted as a "size-biased estimate"

$$\zeta(p) = \hat{\mathbb{E}}_p[|K|], \quad \hat{\mathbb{E}}_p[F(K)] = \frac{\mathbb{E}_p[F(K)|K|]}{\mathbb{E}_p[|K|]}.$$

Similarly, here's a precise definition for the correlation length:

Definition 34

For $p < p_c$, the **correlation length** is defined as follows. The two-point function $\tau_p(0, (n, 0, \cdots, 0))$ should decay exponentially as $e^{-cn+o(n)}$, and we want $c = \frac{1}{\xi(p)}$. Thus, we should define

$$\xi(p) = \frac{1}{\lim_{n \to \infty} -\log(\tau_p(0, (n, 0, \cdots, 0))^{1/n})}$$

(we'll show later in this course that this limit actually exists).

Of course, it's not clear that this is the definition that makes the scaling limit work, and in fact it's not the definition we want to use for long-range models either.

Example 35

Returning to the binary tree, one idea that came up was that "these large random trees are somehow two-dimensional objects," and we'll come back to that. Let L_r be the vertices at level $r \geq 0$ and $B_r = L_0 \cup \cdots \cup L_r$. We can compute the following probability

$$\mathbb{P}_{p_c}(|K| \geq n, 0 \not\leftrightarrow L_r)$$

at various levels of r.

By Markov's inequality, we can bound this as

$$\mathbb{P}_{p_c}(|K| \geq n, 0 \not\leftrightarrow L_r) \leq \frac{\mathbb{E}_{p_c}[|K| 1\{0 \not\leftrightarrow L_r\}]}{n} \leq \frac{1}{n} \mathbb{E}[K \cap B_{r-1}].$$

But this is exactly equal to $\frac{r}{n}$ because the expected number of vertices in the cluster at each level is 1. We can now notice that this is lower order than the overall probability $\mathbb{P}_{p_c}(|K| \geq n) \asymp \frac{1}{\sqrt{n}}$ for $r \ll \sqrt{n}$. So the way to interpret this is that if we want to get to volume n, we're unlikely to do so within a ball of radius $\sqrt{n}$.

We can also do this the other way around: we have

$$\mathbb{P}_{p_c}(0 \leftrightarrow L_{3r}, |K| \leq n) \leq \mathbb{P}_{p_c}\left(0 \leftrightarrow L_{3r} \text{ and there exists } K \in [r, 2r] \text{ such that } |K \cap L_k| \leq \frac{n}{r}\right),$$

since if we didn't have such a thin level we'd already have n total vertices among the middle r levels. The probability we find such a level is at least the probability that we have any surviving vertices at height r, which we denote Q_r, and then once we find this "thin level" with at most $\frac{n}{r}$ vertices, then at least one of them needs to make it still to level $3r$ which is at least r levels away. Thus by a union bound, this probability is

$$Q_r \cdot \frac{n}{r} Q_r \sim \frac{n}{r^3},$$

which is much smaller than the overall probability $\mathbb{P}_{p_c}(0 \leftrightarrow L_{3r}) = Q_{3r}$ when $n \ll r^2$. So in fact volume being n and radius being $\sqrt{n}$ "really want to happen at the same time."

At criticality, each level contributes exactly 1 expected vertex and thus $\mathbb{E}_{p_c}[|K \cap B_r|] = r + 1$ grows linearly. But now if we condition on getting to a certain distance, we see that the expected number of new vertices we see at the "next layer of r" is

$$\mathbb{E}_{p_c}\left[|K \cap (B_{2r} \setminus B_r)|\Big|0 \leftrightarrow L_r\right] \geq \frac{\mathbb{E}\left[|K \cap (B_{2r} \setminus B_r)|\right]}{Q_r} \asymp r^2,$$

which grows much faster. Similarly, conditioning on getting to some much higher level R and studying the size of the cluster up to some fixed r, we have

$$\mathbb{P}(|K \cap L_r| = k|0 \leftrightarrow L_R) = \frac{\mathbb{P}(|K \cap L_r| = k, 0 \leftrightarrow L_R)}{Q_R} \leq \frac{\mathbb{P}(|K \cap L_r| = k)Q_{R-r}k}{Q_R},$$

where in the last step we use a union bound because we have k chances to get from level r to R independently (and if the events have low probability the union bound is sharp – we could use the Bonferroni inequality to justify this). Now remembering that $\frac{Q_{R-r}}{Q_R} \to 1$ as $R \to \infty$, we see that

$$\mathbb{P}\Big(|K \cap L_r| = k\Big|0 \leftrightarrow L_R\Big) \to k\mathbb{P}(|K \cap L_r| = k),$$

and this is exactly the size-biased measure! (We don't need to normalize again because the mean of the random variable is 1.) So being careful with some exchange of limits we can plug this in to find the conditional **expectation** instead of probability, and we find that

$$\mathbb{E}\left[|K \cap L_r||0 \leftrightarrow L_R\right] \sim \mathbb{E}\left[|K \cap L_r|^2\right] \text{ as } R \to \infty.$$

If we want to understand the second moment of the size of a set, it's the same as summing over pairs of probabilities, of both occurring in the set: we want to know how many pairs of points (possibly the same) at level r are both connected to the origin, so we can sum over intermediate heights a at which the points have their first common neighbor. Since the factor of 2 coming from choices in the binary branching cancels out exactly with the factors of $\frac{1}{2}$ from the percolation probabilities, we get $\mathbb{E}[|K \cap L_r|^2] \asymp r$. Summing up, we thus find that

$$\boxed{\mathbb{E}\left[|K \cap B_r||0 \leftrightarrow L_R\right] \to cr^2 \text{ as } R \to \infty},$$

which again justifies the two-dimensional picture. And it turns out that there is actually a well-defined weak limit of the measures we get as we condition on reaching larger and larger levels – this is called the **incipient infinite cluster**, and we can get IIC measures on $\mathbb{Z}^d$ as well (at least for $d = 2$ or high dimensions). On the lattice it can be a bit complicated to actually make the definition, but on a tree, it's quite simple via this size-biased measure.

Example 36

We can go back to understanding heuristic scaling theory now, thinking about the **hyperscaling relations** (which are supposed to only hold below the critical dimension). We'll present this in a nonstandard way by introducing a new scaling exponent which should take the value 0 in low dimensions.

The idea is that there is a qualitative distinction between low $d < d_c$ and high $d > d_c$ dimensions, and here's one way to interpret that. If we consider a large box at criticality and count the number of crossing clusters that connect one side of the box to the other, this quantity should be $O(1)$ for $d < d_c$ but divergent for $d \geq d_c$, and that'll lead us to a heuristic computation of why the critical dimension is 6. (But a weakness is that it's possible that we have a critical dimension like this which is separate from the critical dimension for mean-field exponents.) What we expect is that we often have a soup of large clusters that are all important in $d \geq d_c$, but we have only a few important large ones for $d < d_c$ and thus there is more important randomness beyond law-of-large-numbers type behavior.

We'll want to work with a slightly-less clear-cut thing than the number of crossing clusters: we'll switch to thinking about volume to make things better-behaved in general. Fix a parameter r and consider $\Lambda_r = [-r, r]^d$. We previously considered $\zeta(p)$ (characteristic large size of clusters at a probability p away from criticality), and we can do the same thing at criticality and define $\zeta(r)$ to be the typical large p_c-size of clusters in Λ_r. We'll also let $N(r)$ denote the number of large clusters (note that in the $d \geq d_c$ regime, we should expect by extreme-value type results that the actual largest cluster is much larger than $\zeta(r)$). The point is that we can use these quantities to do heuristic computations and get the right answer. For example (still at criticality),

$$\mathbb{E}[|K \cap \Lambda_r|^m] \asymp \zeta(r)^m \mathbb{P}(0 \in \text{typical large clusters})$$

since the main contribution should be one of these large ones. Since the origin is like a generic point in the box, this probability is just $\frac{N(r)\zeta(r)}{r^d}$. Thus if we say that $\zeta(r) \approx r^{d_f}$ for the **fractal dimension** d_f, and if we write $N(r) \approx r^{\#}$, then we should expect $d < d_c$ if and only if $\# = 0$ (that is, if the number of large clusters is not growing). In general, this yields the expression

$$\mathbb{E}\left[|K \cap \Lambda_r|^m\right] \approx r^{(m+1)d_f + \# - d}.$$

We already have an easy way to get a relation from this: if we look at the sum of the two-point function, then $\sum_{x \in \Lambda_r} \tau(0, x) \approx \sum_{x \in \Lambda} ||x||^{-d+2-\eta} = \mathbb{E}[|K \cap \Lambda_r|]$, and we can (as usual) sum over the lattice by level to get that this quantity scales as $r^{2-\eta}$. On the other hand, plugging in $m = 1$ above yields $\zeta(r)^2 \frac{N(r)}{r^d} = r^{2d_f + \# - d}$, and therefore we get $2 - \eta = 2d_f + \# - d$ and

$$\boxed{2 - \eta = 2d_f - d \text{ for } d < d_c}.$$

Remark 37. *We're skipping over some details like how 0 can be connected to some vertex but only through a path outside of the box. This turns out to be quite a thorny technical detail in lots of calculations, but it always turns out to not affect the final answer. On the other hand, we could think about doing things on the torus and get wraparound (periodic boundary conditions), and in high dimensions that actually makes a huge difference.*

Next to study the volume tail δ, we're going to make another leap in logic and say heuristically that "the best way to have size n is to be in a typical large cluster of scale r, where $\zeta(r) = n$." Therefore

$$\mathbb{P}(|K| \geq \zeta(r)) \approx \mathbb{P}(0 \text{ is in a typical large cluster}) \approx r^{d_f + \# - d}.$$

But since $r = n^{1/d_f}$, we find that

$$\mathbb{P}(|K| \geq n) \approx n^{(d_f + \# - d)/d_f},$$

and therefore (remember δ is the reciprocal over the exponent in the volume tail) $\delta = \frac{d_f}{d - d_f - \#}$. This yields another hyperscaling relation

$$\boxed{\delta = \frac{d_f}{d - d_f} \text{ for } d < d_c}.$$

We'll have a part of the course later on where we can prove inequalities if not the full asymptotic behavior, and we'll also show where we can compute exponents in long-range percolation. But before that, we'll mention a few more hyperscaling relations. From our relation relating γ, η, ν, some algebra yields

$$\boxed{\gamma = (2d_f - d)\nu, \quad 2 - \eta = d\frac{\delta - 1}{\delta + 1} \quad \text{for } d < d_c}.$$

And interestingly, out of everything we've discussed so far, it turns out we only have two degrees of freedom in $d < d_c$ left if we assume the hyperscaling relations hold. And this already gives the first explanation of why the critical dimension is 6: we should get mean-field exponents for $d > d_c$ and hyperscaling for $d < d_c$, so presumably at $d = d_c$ we should have both with log corrections (for example $N(r)$ should grow as a power of log rather than a power). And since $\delta = 2, \eta = 0$ for mean-field, $2 = d\frac{1}{3} \implies d = 6$.

For one more hyperscaling relation, we can also consider the **one-arm probability** which (for not long-range models) is the probability that the origin hits the boundary of Λ_r. This scales as $r^{-1/\rho}$ and we typically call ρ the **one arm exponent**; another manifestation of hyperscaling is that if we have two points at distance r, the two-point function should be roughly the probability that both x and y connect to boxes of size $\frac{r}{4}$ (at which point they will have constant-order probability to connect via the big clusters), which is of order $r^{-2/\rho}$. Thus we get

$$\boxed{d - 2 + \eta = \frac{2}{\rho} \iff d_F = d - \frac{1}{\rho} \quad \text{for } d < d_c}$$

Indeed, this result can be interpreted as saying that "the size of a cluster is just the number of points connected to the boundary of the box."

Fact 38

These kinds of relations shouldn't hold in high dimensions, because (due to a theorem of Kozma and Nachmias) we have $\rho = \frac{1}{2}$ in $d \gg 6$ via the lace expansion, but $-d + 2$ is significantly more negative than the exponent of $r^{-2/\rho} = r^{-4}$. (So it's significantly harder to connect to another point than just to connect to the correct scale, since we expect to have many large clusters.) Instead it turns out that in high dimensions the fractal dimension is $d_f = 4$, and we have $\# = d - 6$ (so we have roughly a polynomial r^{d-6} number of clusters of size r^4 at criticality).

In terms of heuristic connections with branching random walk, we saw previously that large trees are 2D objects which we embed with random walks with $|X_t| \asymp \sqrt{t}$ in a branching process. Thus the embedding will "double the dimension," and it makes sense that the fractal dimension stabilizes at a constant.

Remark 39. *This fractal dimension is somehow the Hausdorff dimension of a particular scaling limit, but we won't talk about that much here.*

We're now finally ready to move on to proofs. We're transitioning now to working on weighted graphs (in anticipation of working on long-range models), and our first proof next time will be the **sharpness of the phase transition**.

Definition 40

A **weighted graph** is a triple $G = (V, E, J)$ containing vertices V, edges E, and J a function $E : [0, \infty)$. We can do Bernoulli bond percolation on a weighted graph with a parameter which we now call β, where an edge is now open with probability $1 - e^{-\beta J(e)}$. (This is roughly $\beta J(e)$ for $J(e)$ small.)

This particular choice of how we define the model is just algebraically convenient – it's equivalent to a Poisson process on E with intensity βJ, up to ignoring multiple edges. And that's a useful property because the union of two independent Poisson processes with intensities β_1, β_2 is exactly equal in distribution to a single process with intensity $\beta_1 + \beta_2$. And if we set J identically equal to 1, we get percolation as before but with $p = 1 - e^{-\beta}$, and this even still makes some formulas look nicer.

Definition 41

A weighted graph is **(vertex-)transitive** if for all $u, v \in V$, there is an automorphism γ with $\gamma(u) = v$ (meaning that the incidence relation and weights are preserved under the map).

We'll be most interested in the case where J is a kernel of the form $J : \mathbb{Z}^d \times \mathbb{Z}^d \to [0, \infty)$ which is symmetric (meaning $J(x, y) = J(y, x)$) and translation-invariant (meaning $J(x, y) = J(0, y - x)$ – we'll often just write this as $J(y - x)$). In this interpretation, our edge set E is the set of all pairs of vertices in $\mathbb{Z}^d$. Looking toward long-range percolation, we will care about the case where J is **integrable**, meaning that $\sum_{x \in \mathbb{Z}^d} J(0, x) = |J| < \infty$, and thus the expected degree at the origin is bounded by $\beta|J|$ (calculus exercise). Thus we can define the critical threshold

$$\beta_c = \inf\{\beta \geq 0 : \text{an infinite cluster exists}\},$$

and a basic lemma is that $\beta_c \geq \frac{1}{|J|} > 0$ (this is the exact same proof as $p_c \geq \frac{1}{\Delta - 1}$, except we can't really control backtracking in the same way so we don't get the -1).

Of course, we can let $J(x, y) = 1\{||x - y||_1 = 1\}$ and recover nearest-neighbor percolation, but what we'll think about is where we can have infinite-range interactions given by a power law

$$J(x - y) = ||x - y||^{-d-\alpha}$$

for some $\alpha > 0$ (this is the same α as in the α-stable Levy process). What's interesting is that varying the parameter α turns out to be similar to varying d, but we can now do it continuously in a sensible way! And of course, in real-world physical models we do have long-range interactions (like with electrostatics or epidemics).

Theorem 42

For the long-range percolation model, we have $\beta_c < \infty$ if and only if $d \geq 2$ (this actually follows from what we did before) or $d = 1, \alpha \leq 1$. (In particular, we do get phase transitions in one dimension.)

We'll try to prove as much as we can about this long range model; it has many different regimes, not just low- and high-dimension but also nearest-neighbor versus very long range, and there are boundaries of the regimes with various log terms as well.

5 June 9, 2025

Last time, we defined weighted graphs and percolation on weighted graphs – we'll eventually want to study long-range percolation on $\mathbb{Z}^d$, and it makes sense now to study in this generality. Our first goal today is to prove the following result:

Theorem 43 (Sharpness of the phase transition)
Suppose $G = (V, E, J)$ is a transitive weighted graph which is integrable (meaning $|J| < \infty$). Fix some origin o (the choice doesn't matter) and let K be the cluster of o. Then we have the following:

1. For all $\beta < \beta_c$ (the extremal value of β where at least one infinite cluster exists with positive probability), there is some positive constant c_β such that $\mathbb{P}_\beta(|K| \geq n) \leq e^{-c_\beta n}$ is exponentially small; in particular the susceptibility $\mathbb{E}_\beta[|K|]$ is finite.

2. There exists some positive constant C such that

$$\mathbb{P}_\beta(0 \leftrightarrow \infty) \geq c((\beta - \beta_c) \wedge 1) \text{ for all} \beta > \beta_c.$$

   In particular, this latter fact means the critical exponent β is at most 1 (it could be smaller if the growth of the connection probability is faster than linear near β_c – we expect actually in low dimensions that the derivative is infinite at β_c).

For a long time it was open to prove that the p at which infinite clusters exist and where $\mathbb{E}_p[|K|]$ is infinite are equal – old texts will have different names for these two critical values p_t and p_h, but this was proved independently with two completely different methods by Menshikov in 1986 and Aizenman and Barsky in 1987, and then later on using even more different proof strategies which also apply to dependent percolation models (papers from the 2010s and 2020s). All of the proofs manage to show both parts of this theorem, even though those two facts may appear to be completely different, and each strategy tells us something else about our quantities of interest. So it's not just about reproving the same theorem over and over again – we are actually extracting something new.

The fact that the critical exponent β is bounded (in any dimension) by its high-dimensional value is a **mean-field bound** – it says that phase transitions can only get more severe for lower dimensions. The Aizenman-Barsky proof manages to also get another mean-field bound $\delta \geq 2$ (though annoyingly it only shows this for the Laplace transform); we'll show a different proof of that fact later on. And Professor Hutchcroft's proof of this in 2019 also yields $\gamma \leq \delta - 1$.

Proof by Duminil-Copon and Tassion, 2015. For the first point, we'll actually just first prove finite expectation $\mathbb{E}_\beta[|K|] < \infty$; it has to do with redefining β_c and eventually showing that this coincides with the original definition. Somehow this is a useful set of "mental gymnastics" for making the proof easier.

(This argument is often referred to as the "ϕ_p argument.") First define, for any finite subset S of vertices containing the origin,

$$\phi_\beta(S) = \sum_{e \in \partial_E S} \left(1 - e^{-\beta J(e)}\right) \mathbb{P}_\beta(0 \overset{S}{\leftrightarrow} e^-),$$

where $\partial_E S$ is the set of edges where one endpoint e^- is in S and the other e^+ is outside S. Define

$$\tilde{\beta}_c = \inf \{\beta : \phi_\beta(S) \geq 1 \text{ for all } S \subseteq V \text{ finite containing } o\}.$$

It **suffices to prove** now that **(1)** $\mathbb{E}_\beta[|K|] < \infty$ for all $\beta < \tilde{\beta}_c$ and also that **(2)** $\mathbb{P}_\beta(0 \leftrightarrow \infty) \geq \frac{\beta - \tilde{\beta}_c}{\beta}$ for all $\beta > \tilde{\beta}_c$. Indeed, above $\tilde{\beta}_c$ we have an infinite cluster and below we don't, so in fact $\beta_c = \tilde{\beta}_c$ and we will have implied both implications of the theorem.

For **(1)**, we'll present a problematic proof and then explain how to fix it. For $\beta < \tilde{\beta}_c$, by definition there is some subset $S \subseteq V$ containing the origin such that $\phi_\beta(S) < 1$. For any vertex $x \notin S$, we claim that

$$\mathbb{P}_\beta(o \leftrightarrow x) \leq \sum_{e \in \partial_E S} \mathbb{P}\left(o \overset{S}{\leftrightarrow} e^-\right) \mathbb{P}_\beta(e \text{ open}) \mathbb{P}(e^+ \leftrightarrow x).$$

Indeed, what we can do is explore the cluster of the origin in S in whatever way we'd want (for example, breadth-first search), which will reveal some open and closed edges within S. In order for o to be connected to x there must be some open path, which means there must exist some edge in the end boundary connected via this exploration, and then there must be a further connection which **does not reuse** edges from that open path we've traversed so far (which is strictly harder than using an independent configuration, so we're doing percolation inside the smaller subgraph). So then a union bound over all potential boundary edges yields the claim.

But now if we sum over all vertices $x \in V$ and bound the probability for vertices $x \in S$ by 1, we get (we need transitivity in the second line)

$$\begin{aligned}\mathbb{E}_\beta\left[|K|\right] &\le |S| + \sum_{e \in \partial_E S} \mathbb{P}\left(o \overset{S}{\leftrightarrow} e^-\right)\mathbb{P}_\beta(e \text{ open})\sum_{x \in V}\mathbb{P}(e^+ \leftrightarrow x) \\ &= |S| + \sum_{e \in \partial_E S} \mathbb{P}\left(o \overset{S}{\leftrightarrow} e^-\right)\mathbb{P}_\beta(e \text{ open})\mathbb{E}_\beta[|K|] \\ &\le |S| + \phi_\beta(S)\mathbb{E}_\beta[|K|].\end{aligned}$$

Since $\phi_\beta(S) < 1$, we may want to rearrange this as $\mathbb{E}_\beta[|K|] \le \frac{|S|}{(1-\phi_\beta(S))}$ and show that we indeed get a finite value. Unfortunately if $\mathbb{E}_\beta[|K|]$ is infinite we cannot do this last manipulation.

One way to fix this is to do everything inside a finite subgraph, where it's no longer true that the sum $\mathbb{P}_\beta(e^+, x)$ doesn't depend on the base point so things are a bit more annoying (but still work). Alternatively, this is saying that "if we're finite we're bounded by a constant," so we can do a topological argument about approaching from below. We'll show a slightly different way here which is nice and similar to what we did with the binary tree: we introduce a ghost field $\mathcal{G}$, which is a random subset of the vertices V with $\mathbb{P}(v \in \mathcal{G})$ independently of probability $1[e^{-h}$. Then

$$\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}) = \mathbb{E}\left[1 - e^{-h|K|}\right] \implies \mathbb{E}_\beta[|K|] = \lim_{h \downarrow 0}\frac{1}{h}\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}).$$

(Note that for heavy-tailed distributions, the "best way to connect" to the ghost field is no longer to have a cluster of a particularly large size – it's instead just to get lucky and have something near the origin be in the field.) But now we can calculate by exactly the same argument (explore in S, then see how we can connect to the ghost field from there) that

$$\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G} \setminus S) \le \phi_\beta(S)\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}),$$

so that

$$\begin{aligned}\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}) &\le \phi_\beta(S)\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}) + \mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G} \cap S) \\ &\le \phi_\beta(S)\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}) + \mathbb{P}(\mathcal{G} \cap S \ne \varnothing) \\ &\le \phi_\beta(S)\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}) + (1 - e^{-h|S|}),\end{aligned}$$

and now we can rearrange the probabilities (which are definitely finite)

$$\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G}) \le \frac{1 - e^{-h|S|}}{1 - \phi_\beta(S)}.$$

Now dividing by h and taking $h \downarrow 0$ yields the inequality from before but now properly justified, completing the proof. In fact, if we just use the actual value $\mathbb{P}_{\beta,h}(0 \leftrightarrow \mathcal{G} \cap S)$, we get a slightly better numerator

$$\boxed{\mathbb{E}_\beta[|K|] \le \frac{\mathbb{E}\left[\#\{x \in S : 0 \overset{S}{\leftrightarrow} x\}\right]}{1 - \phi_\beta(S)} < \infty}.$$

This turns out to actually be useful in other contexts as well – it's related to the notion of the correlation length. When we talked about heuristic scaling theory, we said that below criticality there should be a characteristic scale at which we start "not feeling critical," and we can define (now going back to the setting of $\mathbb{Z}^d$) the correlation length ξ to be the infimum value of r such that $\phi_\beta([-r, r]^d) \le \frac{1}{2}$. With this definition, we in fact have

$$\mathbb{E}_\beta[|K|] \asymp \mathbb{E}_\beta \left[\# \left\{ x \in [-\xi(r), \xi(r)]^d : 0 \overset{\text{box}}{\leftrightarrow} x \right\} \right].$$

So throwing things away beyond this scale no longer affects the quantities that we care about, and for long-range percolation this does appear to be the most useful correlation length definition (rather than using the rate of exponential decay).

Turning now to the proof of (2), we now want a lower bound on the density of the infinite cluster. This will use **Russo's formula**, which is something that typically comes a lot earlier in a course like this:

Proposition 44 (Russo)
Consider a product measure on $\{0, 1\}^E$ (we can think of this as a percolation configuration), and suppose $F : \{0, 1\}^E \to \mathbb{R}$ is a map depending only on finitely many of the bits. (In practice, we'll want to apply this without this finite restriction, though, so we'll talk about that later.) Assume that $\mathbb{E}_\beta[F(\omega)]$ is a smooth function (in our case it'll be a polynomial, since each event configuration occurs with a certain polynomial probability). Then

$$\frac{d}{d\beta} \mathbb{E}_\beta[F(\omega)] = \sum_e \mathbb{E}_\beta \left[F(\omega \cup \{e\}) - F(\omega) \right]$$

where we can restrict this sum to only the edges e that F depends on.

This is basically calculus, using that we can write the total derivative as a sum of partial derivatives and since each edge "becomes open" at rate $J(\beta)$. This can be interpreted in the following way: for any event A depending on finitely many edges which is increasing (so 1_A is increasing in the coordinates, meaning adding more edges can't make the event stop holding), then letting $F = 1_A$ yields that

$$\frac{d}{d\beta} \mathbb{P}_\beta(A) = \sum_e J_e \mathbb{P}_\beta(e \text{ is a } \textbf{closed pivotal} \text{ edge for } A),$$

where this event in the probability means that the configuration is not in A, but adding e (in particular meaning that e was originally not in A) makes it in A. In general we will say that e is **pivotal** if $\omega \cup \{e\} \in A$ and $\omega \setminus \{e\} \notin A$ and **open pivotal** if it is open and pivotal – notably being pivotal is independent of the edge itself being open or closed.

We do have to be careful when apply Russo's formula in infinite volume. However, it is always true that for an increasing nonnegative function F,

$$\left(\frac{d}{d\beta} \right)_+ \mathbb{E}_\beta \left[F(\omega) \right] \ge \sum_e J_e \mathbb{E}_\beta \left(F(\omega \cup \{e\}) - F(\omega) \right),$$

where these one-sided derivatives

$$\left(\frac{d}{d\beta} \right)_+ f(\beta) = \liminf_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} (f(\beta + \varepsilon) - f(\beta))$$

are always well-defined. Either way, the proof we're about to do is in finite volume anyway.

So for **(2)**, let $\Lambda \subseteq V$ be a finite set containing the origin. We define the random set

$$\mathcal{S} = \{ x \in \Lambda : x \not\leftrightarrow \Lambda^c \},$$

and we make the observation that conditional on $\mathcal{S}$, edges with both endpoints in S are distributed as unconditional Bernoulli percolation. Indeed, if we explore the complement of the set $\mathcal{S}$ from outside in, we'll explore everything connected to Λ^c, and similarly to before we know that the boundary of everything else must be all closed, but then anything with both endpoints inside $\mathcal{S}$ was never visited. Also, notice that an edge e is closed pivotal for the event $\{o \leftrightarrow \Lambda^c\}$ if and only if $o \in \mathcal{S}$, $e \in \partial_E \mathcal{S}$, and $o \leftarrow e^-$ in $\mathcal{S}$ (that is, the origin is not connected to the outside now, but there is a path through e). Thus by Russo's formula,

$$\begin{aligned}\frac{d}{d\beta}\mathbb{P}_\beta\left(o \leftrightarrow \Lambda^c\right) &= \sum_e J_e \mathbb{P}(e \text{ closed pivotal}) \\ &= \mathbb{E}\left[\sum_e J_e 1\{e \text{ closed pivotal}\}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_e J_e 1\{e \text{ closed pivotal}\}\middle|\mathcal{S}\right]\right] \\ &= \mathbb{E}\left[\sum_{e\in\partial_E\mathcal{S}} J_e 1\{o \overset{\mathcal{S}}{\leftrightarrow} e^-\}1\{o\in\mathcal{S}\}\right].\end{aligned}$$

Recalling that for any set S, $\phi_\beta(S)$ was the sum over boundary edges of $\mathbb{P}(e \text{ open})\mathbb{P}(o \overset{S}{\leftrightarrow} e^-)$. And (by calculus) J_e is bounded from below by $\frac{1-e^{-\beta J_e}}{\beta}$, so the expression above can be bounded via

$$\mathbb{E}\left[\sum_{e\in\partial_E\mathcal{S}} J_e 1\{o \overset{\mathcal{S}}{\leftrightarrow} e^-\}\right] \geq \frac{1}{\beta}\mathbb{E}\left[\phi_\beta(\mathcal{S})1\{o\in\mathcal{S}\}\right].$$

By definition of $\tilde{\beta}_c$, this is thus bounded by $\frac{1}{\beta}\mathbb{P}(o \not\leftrightarrow \Lambda^c)$. So for any $\beta > \beta_c$, we have

$$\frac{d}{d\beta}\mathbb{P}(o \leftrightarrow \Lambda^c) \geq \frac{1}{\beta}\left(1 - \mathbb{P}(o \leftrightarrow \Lambda^c)\right),$$

and all that's left now is to analyze the differential inequality. Letting $\mathbb{P}(o \leftrightarrow \Lambda^c)$ be $f(\beta)$, we thus have $f'(\beta) \geq \frac{1}{\beta}(1-f)$, and we want to show that this implies $f(\beta) \geq \frac{\beta - \tilde{\beta}_c}{\beta}$. Indeed,

$$(\beta f)' = f + \beta f' \geq f + (1-f) = 1,$$

so integrating (and using that $f(\beta_c) \geq 0$) yields the result. Then by taking Λ larger and larger sets yields the result for $\mathbb{P}(o \leftrightarrow \infty)$, as desired. □

So what we've showed is that below β_c the susceptibility is finite (we'll actually show how this implies exponential decay) and above β_c we have at least linear growth in the probability of an infinite cluster. The next thing we'll discuss are the **correlation inequalities** (Harris and BK specifically); we'll start by stating them and talking a bit and we'll do the proof next time. These are really the main tools along with Russo's formula used for almost everything.

Theorem 45 (Harris)

Let A, B be increasing events. Then $\mathbb{P}_\beta(A \cap B) \geq \mathbb{P}_\beta(A)\mathbb{P}_\beta(B)$ (that is, increasing events are nonnegatively correlated); more generally for any increasing functions $F, G : \{0,1\}^E \to [0,\infty]$, $\mathbb{E}_\beta[FG] \geq \mathbb{E}_\beta[F]\mathbb{E}_\beta[G]$. The same is also true if both functions are decreasing.

The reason this is also called the FKG inequality (Fortuin–Kasteleyn–Ginibre) is that those three authors showed a general condition (the FKG lattice condition) for this to hold for potentially non-product measures, which is useful in dependent models like random cluster or Potts. The BK inequality actually goes in the opposite direction:

Theorem 46 (van den Berg–Kesten)

Consider a configuration $\omega \in A$. We say that a set of edges $W \subseteq E$ is a **witness** for $\omega \in A$ if for any other ω' with $\omega'|_W = \omega|_W$, we also have $\omega' \in A$. (In other words, we can define $[\omega]_W = \{\omega' : \omega'|_W = \omega|_W\}$, and W is a witness if $[\omega]_W \subseteq A$.) For two sets A, B define the **disjoint occurrence** $A \circ B$ to be the event that there are disjoint witnesses for A and B; that is,

$$A \circ B = \bigcup_{W_A \cap W_B = \varnothing} \{[\omega]_{W_A} \subseteq A\} \cap \{[\omega]_{W_B} \subseteq B\}.$$

Then for increasing (or both decreasing) events A, B, we have $\mathbb{P}_\beta(A \circ B) \le \mathbb{P}_\beta(A)\mathbb{P}_\beta(B)$.

The key example to keep in mind is that if $A = \{x \leftrightarrow y\}$ and $B = \{z \leftrightarrow w\}$, then $A \circ B$ is the event that there are disjoint paths from x to y and z to w (because any open path is a witness for connectedness), and we're saying that it is harder to have disjoint paths than just have paths independently because we can't reuse edges. In words, it's less likely for two events to occur "for different reasons."

This is usually stated in terms of only events with finitely many edges, but (as far as Professor Hutchcroft can tell) **that condition is not necessary**. However, there's a subtlety that it's not clear that the event $A \circ B$ is actually measurable (since it involves quantification over uncountably many sets). Luckily, descriptive set theory comes to the rescue, since A and B being increasing means that we can remove some quantification and write

$$A \circ B = \bigcup_{W_A \cap W_B} \{\omega \wedge 1_{W_A} \in A\} \cap \{\omega \wedge 1_{W_B} \in B\}$$

as an **analytic set**. Specifically, a Borel set is some set of existential and universal quantifiers of countably many variables, and an analytic set lets us additionally add existential quantifiers (or just add universal quantifiers) over Polish spaces such as $\mathbb{R}$ instead. Such a result is then **universally measurable**, meaning that it's in the completion of the Borel σ-algebra with respect to any measure (so in particular Lebesgue measurable). **However**, we can't actually have both kinds of quantifiers without the measurability being independent of ZFC – the answer is that we do get measurability under "large cardinal axioms," but whether that makes us happy is unclear.

6 June 10, 2025

We stated the main correlation inequalities last time, which tell us that increasing functions (hence events) of a percolation configuration are positively correlated, so in particular $\mathbb{P}_\beta(A \cap B) = \mathbb{P}_\beta(A)\mathbb{P}_\beta(B)$ and that the "disjoint occurrence" has probability $\mathbb{P}_\beta(A \circ B) \le \mathbb{P}_\beta(A)\mathbb{P}_\beta(B)$. And both of these are really statements about abelian functions rather than percolation itself – we can generalize to strings of variables taking values in any totally ordered set.

Remark 47. *Last time, we described $A \circ B$ as an analytic set in terms of quantifiers, and we'll now explain what that definition means analytically as well. Recall that a* **Polish space** *is a completely metrizable separable space (such as $\mathbb{R}$, the Cantor set, or a countable discrete set). For a Borel subset $A \subseteq X \times Y$, we can ask whether the projection map $\pi : X \times Y \to X$ makes $\pi(A)$ Borel. The answer is* **no** *– this is known as Lebesgue's mistake, and we can instead make the definition that an* **analytic set** *is a projection of a Borel set. Indeed, this projection is exactly $\{x : \exists y \text{ with } (x, y) \in A\}$, so it's basically the same thing as using quantifiers (we can even take a countable product with X and it'll still be a Polish space). And a main theorem is that* **analytic implies universally measurable**, *and when A, B are increasing we can indeed check that $A \circ B$ is analytic under this new definition.*

For example, take the Hausdroff topology on the set of all compact subsets of $[0, 1]$, and the set of uncountable compact sets is analytic (and "analytically complete," which is kind of an analog of NP-complete) but not Borel. But of course this is not the emphasis of this course – it's just nice to know about this because when we encounter non-measurable sets they are often inequality.

It turns out that we can also drop the increasing assumption for the BK inequality if we add back the requirement of finiteness:

Theorem 48 (Reimer)
For all A, B depending on finitely many edges, we have $\mathbb{P}(A \circ B) \le \mathbb{P}(A)\mathbb{P}(B)$.

In particular, this time we can no longer use the same universal quantifier argument to show that $A \circ B$ is analytic (it's in the next level of hierarchy of difficulty to define a set, which as we said last time is independent of ZFC). But if we define a modified version of this operator $A \circ_f B$, which requires us to have **finite** disjoint witnesses, then this is always a Borel set (even when A and B are not Borel), and we can deduce the analogous result. We won't prove this version because it involves a complicated combinatorial argument.

Proof of Theorem 45. We'll want to prove the version with functions $\mathbb{E}_\beta[FG] \ge \mathbb{E}_\beta[F]\mathbb{E}_\beta[G]$. Without loss of generality, we can assume that F and G are bounded (otherwise truncate by some functions $F \wedge M, G \wedge M$, take $M \to \infty$, and use monotone convergence theorem).

The first step is to reduce to the case of finitely many edges $|E|$. If we have an infinite number of edges, we can enumerate them as $\{e_1, e_2, \cdots\}$ and let $\mathcal{F}_n$ be the σ-algebra generated by e_1 through e_n. Then by the martingale convergence theorem (or in the case where F is an indicator, Lévy's zero-one law)

$$\mathbb{E}\left[F(\omega)|\mathcal{F}_n\right] \overset{n\to\infty}{\to} F(\omega)$$

and the same statement for $G(\omega)$ and for $F(\omega)G(\omega)$. So if the theorem is proven for finite $|E|$, we can apply it to the conditional expectations and get

$$\mathbb{E}\Big[\mathbb{E}[F(\omega)|\mathcal{F}_n]\mathbb{E}[G(\omega)|\mathcal{F}_n]\Big] \ge \mathbb{E}\Big[\mathbb{E}[F(\omega)|\mathcal{F}_n]\Big]\mathbb{E}\Big[\mathbb{E}[G(\omega)|\mathcal{F}_n]\Big] = \mathbb{E}[F(\omega)]\mathbb{E}[G(\omega)],$$

but the left-hand side converges to $\mathbb{E}[F(\omega)G(\omega)]$ by the bounded convergence theorem, so the result holds. (We do have to check that the conditional expectations of increasing functions are themselves increasing functions, which is indeed true.)

We're now going to induct on $|E|$ – first, let's do the base case $|E| = 1$, and really the only thing we need is that the state space is totally ordered. We have functions F, G of a single point, and we write

$$\mathbb{E}[F(\omega)G(\omega)] - \mathbb{E}[F(\omega)]\mathbb{E}[G(\omega)] = \frac{1}{2}\mathbb{E} \otimes \mathbb{E}\left[(F(\omega) - F(\tilde{\omega}))(G(\omega) - G(\tilde{\omega}))\right]$$

for $\omega, \tilde{\omega}$ independent copies. (We can just check this by expanding out and using independence.) But this is useful because the two terms on the right-hand side are deterministically always of the same sign (if $\omega > \tilde{\omega}$ then both are positive, and if $\omega < \tilde{\omega}$ then both are negative), so the right-hand side is nonnegative and we're done.

Next assume that the result holds for $|E| = n$; for the inductive step we'll again want to take some conditional expectations. For any $y \in \{0, 1\}$, define the functions $F_1(y) = \mathbb{E}\left[F(\omega)|\omega(e_1) = y\right]$, $G_1(y) = \mathbb{E}\left[G(\omega)|\omega(e_1) = y\right]$, and $(FG)_1(y) = \mathbb{E}\left[F(\omega)G(\omega)|\omega(e_1) = y\right]$. Letting ω_y be the configuration where we set the first edge to y. Then $F(\omega_y), G(\omega_y)$ are increasing in both ω and y by definition, and by our inductive hypothesis $(FG)_1(y) \ge F_1(y)G_1(y)$.

Thus by the tower law of conditional expectation and then applying the inductive hypothesis with $y = \omega(e_1)$,

$$\mathbb{E}[F(\omega)G(\omega)] = \mathbb{E}\Big[(FG)_1(\omega(e_1))\Big] \geq \mathbb{E}\Big[F_1(\omega(e_1))G_1(\omega(e_1))\Big],$$

and then using the case $n = 1$ this is at least $\mathbb{E}[F_1(\omega(e_1))]\mathbb{E}[G_1(\omega(e_1))] = \mathbb{E}[F(\omega)]\mathbb{E}[G(\omega)]$, as desired. □

Proof of Theorem 46. We'll also want to do an inductive proof. First enumerate the edge set $E = \{e_1, e_2, \cdots\}$, and let $\hat{P} = \mathbb{P} \otimes \mathbb{P}$ be two independent copies ω, ω'. Define the sequence of configurations interpolating between them

$$\omega_j(e_i) = \begin{cases} \omega(e_i) & i > j, \\ \omega'(e_i) & i \leq j. \end{cases}$$

Thus $\omega_0 = \omega$ and we have $\lim_{j\to\infty} \omega_j = \omega'$ pointwise. If we now define

$$\hat{A}_j = \{(\omega, \omega') : \omega_j \in A\}, \quad \hat{B} = \{(\omega, \omega') : \omega \in B\}$$

(so we fix the first configuration for $\hat{B}$ but interpolate for $\hat{A}$), then $\hat{A}_0 \circ \hat{B} = (A \circ B) \times \{0, 1\}^E$ (since the event doesn't depend on ω' at all) and thus $\hat{\mathbb{P}}(\hat{A}_0 \circ \hat{B}) = \mathbb{P}(A \circ B)$. It now suffices to prove the following two claims:

1. We have $\limsup_{j\to\infty} \hat{\mathbb{P}}(\hat{A}_j \circ \hat{B}) \leq \mathbb{P}(A)\mathbb{P}(B)$. In fact if $|E| = n$ is finite, then for $j = n$ $\hat{A}_n$ depends only on ω' and thus $\hat{A}_n \circ \hat{B} = B \times A$ we have equality.
2. We have $\hat{\mathbb{P}}(\hat{A}_j \circ \hat{B}) \leq \hat{\mathbb{P}}(\hat{A}_{j+1} \circ \hat{B})$ for all j.

Note that it's not always true that we have equality in the limit in (1) if there are infinitely many edges (even though we're replacing more and more edges of one configuration with another). For example, if A and B are each the event that the density of open edges is at least θ, then this is a tail event and thus we won't find two disjoint witnesses.

For claim (1), we note that $\hat{A}_j \circ \hat{B} \subseteq \hat{A}_j \cap \hat{B}$, and for the intersection we do in fact have the limit

$$\lim_{j\to\infty} \hat{\mathbb{P}}(\hat{A}_j \cap \hat{B}) = \mathbb{P}(A)\mathbb{P}(B).$$

This is a basic exercise in measure theory using cylinder events and bounding with $\frac{\varepsilon}{3}$ factors and so on, and it implies the claim because containment means smaller probability. And for claim (2), for any $j \geq 1$ we can condition on all other edges $\mathcal{F}_{-j} = \{\omega(e_i), \omega'(e_i) : i \neq j\}$ besides e_j. Conditional on these values, we have three possible cases:

(a) $\hat{A}_j \circ \hat{B}$ does not occur even when $\omega(e_j) = \omega'(e_j) = 1$,

(b) $\hat{A}_j \circ \hat{B}$ occurs even when $\omega(e_j) = \omega'(e_j) = 0$,

(c) Neither of the two cases above occur.

It suffices to prove that the inequality we care about even holds conditionally – that is, $\hat{\mathbb{P}}(\hat{A}_j \circ \hat{B} | \mathcal{F}_{-j}) \leq \hat{\mathbb{P}}(\hat{A}_{j+1} \circ \hat{B} | \mathcal{F}_{-j})$. In case (a) the left-hand side is 0 and in case (b) both sides are 1 since the event always occurs, and thus the tricky case is part (c). In this case (where we have dependence on the jth bit) $\hat{A}_j \circ \hat{B}$ does not depend on $\omega'(e_j)$ (since that edge is just not being used for either of our events) and thus (c) holds if and only if $\omega(e_j) = 1$. Thus

$$\hat{\mathbb{P}}(\hat{A}_j \circ \hat{B} | \mathcal{F}_{-j}) = \mathbb{P}(\omega(e_j) = 1 | \mathcal{F}_{-j}) = \mathbb{P}(\omega(e_j) = 1).$$

But now remember that we can only use e_j in one of the two witnesses. If we use the edge for A, then $\hat{A}_{j+1} \circ \hat{B}$ will hold as long as $\omega'(e_j) = 1$. And if we use it for B, then the event will hold as long as $\omega(e_j) = 1$. Thus the right-hand side $\hat{\mathbb{P}}(\hat{A}_{j+1} \circ \hat{B} | \mathcal{F}_{-j})$ is at least $\mathbb{P}(\omega'(e_j) = 1)$ (it's also possible that either of them being 1 is sufficient, which is why we only get an inequality), and so we've proved that the left-hand side is less than the right-hand side, as desired (since ω, ω' are independent copies of the same distribution). $\square$

This is a rather tricky proof to understand, and it's probably recommended for us to look over it ourselves again to internalize it. And the actual proof is not so relevant, most of the time, for actual usage of the result.

We're now ready to enter a long part of the course in which we understand relations between critical exponents and sufficient conditions for taking on mean-field values. We'll begin with the **tree graph inequalities** (due to Aizenman and Newman), which allow us to bound k-point functions by two-point functions. Much like the two-point function tells us the probability two points are connected by an open path, we have

$$\tau_\beta(x_1, \cdots x_k) = \mathbb{P}_\beta\left(x_1, \cdots, x_k \text{ all in the same cluster}\right).$$

First consider the case of three points. If we have a path from x to y and from x to z, then we can look at the last point where the x-z path intersected the x-y path – calling that w, we then have disjoint open paths from w to each of x, y, z. Therefore

$$\{x, y, z \text{ connected}\} \subseteq \bigcup_w \{w \leftrightarrow x\} \circ \{w \leftrightarrow y\} \circ \{w \leftrightarrow z\}.$$

We should briefly mention a point now about taking disjoint occurrences of more than two events. For events $A_1, \cdots, A_n$ we define

$$A_1 \circ \cdots \circ A_n = \{\text{there exist disjoint witnesses for } A_1, \cdots, A_n\}$$

note that this is **not a binary associative operation** (we should think of it as an n-ary operation instead), but $A \circ B \circ C \subseteq (A \circ B) \circ C$ and more generally $A_1 \circ \cdots \circ A_n$ is contained in the iterated disjoint occurrence $(A_1 \circ A_2) \circ A_3 \cdots$. Therefore, the BK inequality still works in that

$$\mathbb{P}(A_1 \circ \cdots \circ A_n) \le \prod_{i=1}^n \mathbb{P}(A_i).$$

So applying that to what we had above, we have

$$\boxed{\tau_\beta(x, y, z) \le \sum_w \tau_\beta(x, w)\tau_\beta(w, y)\tau_\beta(w, z)}.$$

One consequence of this is that the second moment of the cluster at the origin can be written

$$\mathbb{E}_\beta\left[|K|^2\right] = \sum_{x,y} \tau_\beta(0, x, y)$$

by linearity of expectation, and using our inequality this yields the bound (here we do need that the susceptibility is the same at every vertex)

$$\begin{aligned}\mathbb{E}_\beta\left[|K|^2\right] &\le \sum_{w,x,y} \tau_\beta(0, w)\tau_\beta(w, x)\tau_\beta(w, y) \\ &= \sum_w \tau_\beta(0, w) \sum_x \tau_\beta(w, x) \sum_y \tau_\beta(w, y) \\ &= \mathbb{E}_\beta[|K|]^3.\end{aligned}$$

Jensen's gives us $\mathbb{E}_\beta\left[|K|^2\right] \ge \mathbb{E}_\beta[|K|]^2$, but near criticality we have rather heavy-tailed random variables, so it's not surprising if there's a significant difference. And we can continue on: looking at the four-point function, we can do the same thing where we look at the point p where the path from x to w intersects the "tripod" that we have among

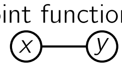
x, y, z. But this time that intersection point can be on the tripod leg for x, for y, or for z, and therefore it's actually a bit of a pain to write out the inequality: calling the interior points a, b,

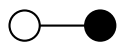
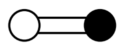
$$\tau_\beta(x, y, z, w) \le \sum_{a,b} \Big(\tau_\beta(x, a)\tau_\beta(a, y)\tau_\beta(a, b)\tau_\beta(b, z)\tau_\beta(b, w) + \tau_\beta(x, a)\tau_\beta(a, b)\tau_\beta(b, y)\tau_\beta(b, z)\tau_\beta(a, w)$$
$$+ \tau_\beta(x, a)\tau_\beta(a, b)\tau_\beta(b, y)\tau_\beta(b, w)\tau_\beta(a, z)\Big).$$

(we don't need to account for the degenerate cases where $a = b$ and so on, since those are already considered). So now summing over all points yields the third moment – every time a point occurs on its own we get a copy of the susceptibility $\mathbb{E}_\beta[|K|]$, and we end up with

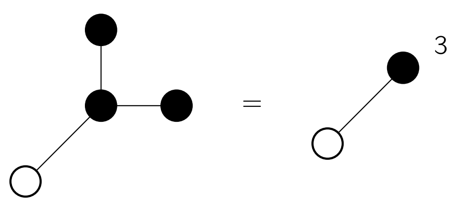
$$\mathbb{E}_\beta\left[|K|^3\right] \le 3\left(\mathbb{E}[|K|]\right)^5.$$

We don't want to keep doing this, and thus we'll introduce some diagrammatic notation, using graphs to tell us how to write out these sums. Each "white vertex" means something that is not summed over (which helps us represent functions), and no label means we have the origin; otherwise the label means that point is something fixed and not summed over. Black vertices are then summed over (and should be thought of as dummy variables), while edges represent two-point functions.

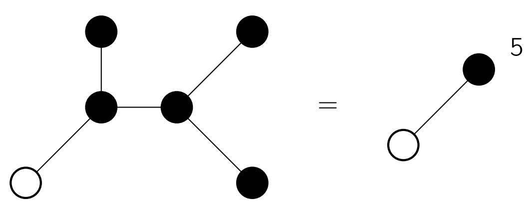
For example, 

is the two-point function,  is the susceptibility because we sum over two-point functions of everything to the origin, and  is the sum of the squares of the two-point functions $\sum \tau_\beta(0, x)^2$ and is called the **bubble diagram** (this comes up a lot more in self-avoiding walk and the Ising model).

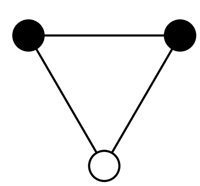
The first tree-graph inequality says that $\mathbb{E}_\beta[|K|^2]$ is bounded from above by the following:



Similarly, the second tree-graph inequality we derived says that $\mathbb{E}_\beta[|K|^3]$ is bounded by 3 times the following diagram (the actual tree is always the same here, it's only up to the choice of labels):



We also have the following **triangle diagram** which will play an important role in percolation: this is the quantity $\sum \tau_\beta(0, x)\tau_\beta(x, y)\tau_\beta(y, 0)$.



In the lace expansion it turns out we can also shade in polygons and that has some other interpretation, but we won't get into that here. And notice that whenever we have one vertex pinned and the others not, we can always switch (via the "mass transfer principle") which one is pinned; we'll talk about this more later.

We can always continue this inductively and get bounds on higher k-point functions, and it turns out the result we get is

$$\tau_\beta(x_1, \cdots, x_k) = \sum_{T \in \mathbb{T}_k} \sum_{\substack{x_i \in \mathbb{Z}^d \\ \text{for } k+1 \le i \le 2k-2}} \prod_{i<j, i \sim j} \tau_\beta(x_i, x_j)$$

where $\mathbb{T}_k$ is the set of all isomorphism class representatives of trees whose leaves are labeled $1, 2, \cdots, k$ and all interior vertices unlabeled and of degree 3. We can always take such trees to have vertex $\{1, \cdots, 2k-2\}$ because such a tree always has $(k-2)$ internal vertices.

The resulting bound is that we always have

$$\mathbb{E}_\beta[|K|^p] \le |\mathbb{T}_{p+1}| \left(\mathbb{E}[|K|]\right)^{2p-1},$$

and we can check that the number of such leaf-labeled binary trees $|\mathbb{T}_{p+1}|$ is exactly $(2p-3)!! = (2p-3)(2p-5)\cdots(3)(1)$. And we'll discuss why these tree-graphs give us a certain correct picture in high dimensions next time!

7 June 12, 2025

Remark 49. *We mentioned last time that $A_1 \circ A_2 \circ \cdots \circ A_n$ is contained in the event $((A_1 \circ A_2) \circ A_3) \circ \cdots$; it turns out that for increasing events we actually have equality, but in general it's possible to have strict containment.*

We used the tree-graph inequalities last time to derive an upper bound on $\mathbb{E}_\beta[|K|^p]$ in terms of the first moment $(2p-3)!!(\mathbb{E}_\beta[|K|])^{2p-1}$ as long as our graph is vertex-transitive. If we want to do this on a general graph, we would have to replace K on the left side by K_x (the cluster containing some specific vertex x), and we'd have to replace K on the right by the supremum over all clusters. Importantly, double factorials arise as the moments of Gaussians, and so what we have here is kind of like a chi-square distribution. We'll come back to that fact later.

We can package the inequality into an appropriate generating function as follows:

Corollary 50

We have (by writing out the Taylor series of the exponential) when G is transitive that

$$\mathbb{E}_\beta\left[|K|e^{r|K|}\right] \le \mathbb{E}\left[|K|\right]\left(1 - 2r(\mathbb{E}[|K|)^2)\right)^{-1/2}$$

for all $r < \frac{1}{2\mathbb{E}[|K|^2}$. In particular, this gives us a tail bound for the cluster size

$$\mathbb{P}_\beta(|K| \ge n) \le \left(\frac{e}{n}\right)^{1/2} \exp\left(-\frac{n}{2\mathbb{E}[|K|]^2}\right).$$

for all $n \ge \mathbb{E}[|K|]^2$.

What we should think is that this is a one-sided version of the postulated ansatz for heuristic scaling theory: we expected to get $n^{-1/\delta} f\left(\frac{n}{\xi(\beta)}\right)$ for some rapidly decaying function f, and that's indeed the form we got above. Of course, it's not true that $\mathbb{E}[|K|]^2$ is the characteristic large length in general, but the inequality itself is always true.

We can also get the following directly from the moment bounds (filling in the missing part of our proof of sharpness of the phase transition from last time):

Corollary 51

If $\mathbb{E}_\beta[|K|] < \infty$, then $\mathbb{P}_\beta(|K| \ge n) \le e^{-c_\beta n}$ for some appropriate constant c_β.

The way to think about such a result is that for a geometric or exponential random variable X of mean λ, the growth of moments looks like $\mathbb{E}[X^p] \asymp p!\lambda^p$. So seeing the double factorials indicates that something similar is happening here, because $(2p)!! \approx p!2^p$.

One thing we'll want to investigate is to what extent the bounds we've written down are of the correct order. We'll see that essentially that they're correct in high dimensions but not low dimensions, and that in fact this is one of the equivalent characterizations of mean-field behavior. To see this, we'll study mean-field lower bounds on the volume tail.

Theorem 52 (Aizenman-Barsky)

Suppose G is transitive and the sum of weights satisfies $|J| < \infty$. Then at criticality, we have

$$\mathbb{E}_{\beta_c}[|K|] \wedge N = \sum_{N=1}^{N} \mathbb{P}_{p_c}(|K| \geq n) \geq cN^{1/2}.$$

In other words, the volume tail morally decays as $\frac{1}{\sqrt{N}}$, though even this fact $\mathbb{P}_{P_c}(|K| \geq n) \geq cn^{-1/2}$ is still open (it's hard to go from Laplace transform to function without some regularity assumptions).

The original proof was a corollary of their proof of sharpness of the phase transition (via considering the magnetization and establishing a partial differential inequality). We'll show a different proof via the tree graph inequalities.

Proof. First we establish a useful lemma:

Lemma 53

For any nonnegative random variable,

$$\mathbb{E}\left[X \cdot 1\left\{X \leq 2\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}\right\}\right] \geq \frac{1}{2}\mathbb{E}[X].$$

Proof of lemma. Recall that $\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ can be thought of as the expectation of X in the size-biased measure, in which $\hat{\mathbb{E}}[F(X)] = \frac{\mathbb{E}[Xf(X)]}{\mathbb{E}[X]}$. So actually this fact is just Markov's inequality with respect to the size-biased measure:

$$\frac{1}{\mathbb{E}[X]}\mathbb{E}\left[X \cdot 1\left\{X \leq 2\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}\right\}\right] = \hat{\mathbb{P}}(X \leq 2\hat{E}[X]) \geq \frac{1}{2},$$

and multiplying both sides by $\mathbb{E}[X]$ yields the result. ☐

We now define $\zeta(\beta) = \frac{\mathbb{E}[|K|^2]}{\mathbb{E}[|K|]}$ (remember heuristically this is supposed to be the size of the typical large cluster). By the tree-graph inequality for the second moment, we have $\zeta(\beta) \leq \mathbb{E}[|K|]^2$; **for simplicity of notation** from now on we'll write $\chi_\beta = \mathbb{E}_\beta[|K|]$ for the susceptibility so that we have $\boxed{\zeta(\beta) \geq \chi_\beta^2}$. Taking some $\beta < \beta_c$, we can truncate and write

$$\mathbb{E}_{\beta_c}\left[|K| \wedge 2\zeta(\beta)\right] \geq \mathbb{E}_\beta[|K| \wedge 2\zeta(\beta)]$$

by monotonicity, and then by our lemma this is at least $\frac{1}{2}\chi_\beta$. Therefore

$$\mathbb{E}_{\beta_c}\left[|K| \wedge 2\zeta(\beta)\right] \geq \frac{1}{2}\chi_\beta \geq \frac{1}{2}\sqrt{\zeta(\beta)}$$

by the boxed inequality. We can now use that $\chi_{\beta_c} = \infty$ (there are various ways of proving this, and it's left as an exercise to us), and that as a corollary to sharpness the moments $\mathbb{E}_\beta[|K|^p]$ are actually real analytic on $[0, \beta_c)$ (also left as an exercise – it's because we can write this down as a series over all clusters, and below β_c we have exponential decay so we can use the Weierstrass M series test), hence continuous. So by the intermediate value theorem, for all N there is some $\beta < \beta_c$ with $2\zeta(\beta) = N$; plugging this in yields that

$$\mathbb{E}_{\beta_c}[|K| \wedge N] \geq \frac{1}{2\sqrt{2}} N^{1/2},$$

which is what we wanted to show. □

We only did this for the case of criticality, but notice that for any $\beta \leq \beta_c$ and any $N \leq 2\zeta(\beta)$, we can still apply this inequality. Thus we also get a lower bound on the tail up until the characteristic size of the large cluster.

Now let's suppose that we know something stronger than the tree-graph inequalities, such as $\zeta(\beta) \leq f(\chi_\beta)$ for some f continuous and strictly increasing. We then find that

$$\mathbb{E}_\beta\left[|K| \wedge N\right] \geq \frac{1}{2} f^{-1}\left(\frac{1}{2}N\right);$$

in particular this gives us an inequality between critical exponents, since if we knew that $\chi_\beta \approx |\beta - \beta_c|^{-\gamma}$ and $\zeta(\beta) \approx |\beta - \beta_c|^{-\Delta}$ as $\beta \uparrow \beta_c$, then we can take $f(x) = Cx^{\Delta/\gamma}$ and get at criticality that

$$\mathbb{E}_{\beta_c}\left[|K| \wedge N\right] \geq C' N^{\gamma/\Delta}.$$

If we also know that the volume tail at criticality decays as $\mathbb{P}_{\beta_c}(|K| \geq n) \approx n^{-1/\delta}$, then $N^{1-\frac{1}{\delta}} \geq N^{\gamma/\Delta}$, meaning that we have derived $\boxed{\frac{\delta - 1}{\delta} \geq \frac{\gamma}{\Delta}}$. And the equality case of this is actually supposed to be a correct scaling relation – we'll prove several inequalities of this form (conditional on exponents being well-defined).

Corollary 54

Suppose the tree-graph inequality is "divergently bad" in the sense that $\mathbb{E}_\beta[|K|^2] = o(\mathbb{E}_\beta[|K|]^3)$ as $\beta \uparrow \beta_c$. Then the tail probability is also divergently large, in that

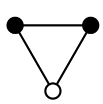
$$\sum_{i=1}^{N} \mathbb{P}_{\beta_c}(|K| \geq n) = \omega(N^{1/2}),$$

where $\omega(g)$ is like $o(g)$ but in the other direction.

So indeed this is a sign that $\mathbb{E}_\beta[|K|^2] \asymp \mathbb{E}_\beta[K]^3$ is a characterization of mean-field behavior, and in the end our goal will be to show that various conditions imply each other to get all of the mean-field exponents.

Example 55

We'll now turn to the triangle condition and how it implies mean-field critical behavior. This will lead us to a couple of other heuristic derivations for $d = 6$ being critical.

Recall the diagram  from last time, encoding the quantity $\nabla_\beta = \sum_{x,y} \tau_\beta(0,x)\tau_\beta(x,y)\tau_\beta(y,0)$.

Definition 56

The **triangle condition** is that $\nabla_{\beta_c} < \infty$.

Theorem 57 (Aizenman-Newman)

If the triangle condition $\nabla_{\beta_c} < \infty$ holds, then the susceptibility satisfies

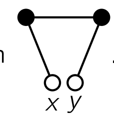
$$\chi_\beta \asymp \frac{1}{|\beta - \beta_c|} \text{ as } \beta \uparrow \beta_c$$

(that is, we're bounded from above and below by constant multiples).

We'll also mention the following theorem (which we will not prove):

Theorem 58 (Hara-Slade)

The triangle condition $\nabla_{\beta_c} < \infty$ holds for nearest-neighbor percolation on $\mathbb{Z}^d$ for d large. (In fact, this further tells us that the two-point function has the correct scaling, though we won't go into the details of this.)

Start of proof of an easier special case of Theorem 57. Consider the open triangle diagram 

. Noticing that the previous triangle condition basically evaluates $\tau_\beta^3(0,0)$ (doing matrix multiplication on the diagonal), our triangle condition is now $\tau_\beta^3(x,y)$ for x and y not necessarily equal. (They don't necessarily have to be close together, either.) On $\mathbb{Z}^d$, notice in particular that $\tau_\beta(x,y) = \tau_\beta(0, y-x)$ and we can rewrite matrix multiplication as convolution.

Remark 59. *It turns out that it can be proved that if $\nabla_{\beta_c} < \infty$, then $\inf_{x,y} \tau_{\beta_c}^3(x,y) = 0$ (that is, the open triangle is small for some values of x, y) via Fourier theory. We'll get to some similar ideas when we do the more general proof.*

The weaker statement will be that we assume that

$$\sup_{x \neq y} \tau_{\beta_c}^3(x,y) < 1$$

for the sake of transparency. To do this, we'll start with another mean-field lower bound: we have by Russo's formula that the derivative of the susceptibility satisfies (this relies on being real analytic)

$$\chi'_\beta = \sum_x \tau'_\beta(0,x) = \sum_x \sum_e J_e \mathbb{P}(e \text{ closed pivotal for } 0 \leftrightarrow x).$$

As an exercise, we can prove this for $\beta < \beta_c$ via sharpness of the phase transition so that we can swap orders of limits and derivatives and so on. We can get the following bound that is always true (and for which we'll try to prove a matching bound for later on):

Lemma 60

We have $\chi'_\beta \leq |J| \chi_\beta^2$.

Proof of lemma. For any edge $e = \{e^-, e^+\}$,

$$\begin{aligned}\mathbb{P}(e \text{ closed pivotal for } 0 \leftrightarrow x) &= \mathbb{P}(e \text{ closed pivotal for } 0 \leftrightarrow x, e^- \leftrightarrow 0) + \mathbb{P}(e \text{ closed pivotal for } 0 \leftrightarrow x, e^+ \leftrightarrow 0) \\ &= \mathbb{P}\left(\{0 \leftrightarrow e^-\} \circ \{e^+ \leftrightarrow x\}\right) + \mathbb{P}\left(\{0 \leftrightarrow e^+\} \circ \{e^- \leftrightarrow x\}\right),\end{aligned}$$

so if we apply the BK inequality and sum over all x, we get the claim. □

And thus as a corollary, integrating yields the following (note that the direction has flipped):

Corollary 61

For any transitive graph, $\chi_\beta \geq \frac{1}{|J|(\beta_c - \beta)}$ for all $\beta < \beta_c$.

Proof. We have $\left(\frac{1}{\chi_\beta}\right)' = -\frac{\chi'_\beta}{\chi^2_\beta} \geq -|J|$, and $\chi_{\beta_c} = \infty$ implies that we have

$$-\frac{1}{\chi_\beta} = \frac{1}{\chi_{\beta_c}} - \frac{1}{\chi_\beta} \geq \int_\beta^{\beta_c} -|J|ds = -(\beta_c - \beta)|J|;$$

rearranging yields the result. □

Heuristically, something behaving like a power function should satisfy $\chi'_{\beta_c - \varepsilon} \asymp \frac{1}{\varepsilon}\chi_{\beta_c - \varepsilon} \implies \frac{1}{\varepsilon}\chi \leq \chi^2 \implies \chi \geq \frac{1}{\varepsilon}$, and so that tends to tell us a good starting point for what to expect.

Returning to our main proof, the key step will be to prove a complementary differential inequality $\chi'_\beta \geq c\chi^2_\beta$ (in the other direction as what we had above) for $\beta < \beta_c$. What Aizenman and Newman proved is that the triangle condition does actually imply differential inequalities of this form.

Definition 62

A vertex-transitive graph G is **unimodular** if the **mass transport principle**

$$\sum_{x \in V} F(0, x) = \sum_{x \in V} F(x, 0)$$

holds for all $F : V^2 \to [0, \infty]$ such that $F(\gamma u, \gamma v) = F(u, v)$ for any symmetry $\gamma \in \text{Aut}(G)$. Similarly, with functions with more than one variable we have statements like $\sum_{x,y,z} F(0, x, y, z) = \sum_{x,y} F(x, 0, y, z)$ and so on.

Almost all transitive graphs that we're likely to encounter will have this property; in particular for models on $\mathbb{Z}^d$,

$$\sum_x F(0, x) = \sum_x F(-x, 0) = \sum_x F(x, 0)$$

by using that translation by $-x$ and reflection are both automorphisms. And in diagrammatic notation this is exactly the fact we mentioned about "allowing ourselves to make a different node of our diagram white." (For non-unimodular things we can actually say much more and prove everything about them, but we'll leave that aside.) We can use this to transform the derivative: take the Russo's formula calculation above and write

$$\begin{aligned}
\chi'_\beta = \frac{d}{d\beta}\mathbb{E}_\beta[|K|] &= \mathbb{E}_\beta\left[\sum_{a \sim b, x} J_{ab} 1\{(a, b) \text{ closed pivotal for } 0 \leftrightarrow x\right] \\
&= \mathbb{E}_\beta\left[\sum_{a \sim b, x} J_{0b} 1\{(0, b) \text{ closed pivotal for } a \leftrightarrow x\right] \\
&= \mathbb{E}_\beta\left[|K| \sum_b 1\{b \not\leftrightarrow 0\}|K_b| J_{0b}\right],
\end{aligned}$$

where we've used the mass-transport principle to switch the role of 0 and a. So in particular this is looking at the size $|K||K_b|$ but only on the event where the clusters are distinct; previously we said this makes them stochastically smaller than K to get an upper bound, and now we want to know what conditions makes this conditioning not actually affect the distribution very much so that we still get mean-field behavior.

For this, we can do a coupling with independent clusters: if we fix b and let $\tilde{K}_b$ be a copy of the cluster at b, we then get a pair of clusters $(K, \tilde{K}_b)$ independent with $\tilde{K}_b$ distributed as a cluster of 0 and $\tilde{K}$ as one of b. We couple this with the actual pair of clusters in the configuration by first exploring K (and setting this to be the cluster of 0 in the actual percolation configuration). Then if $y \in K$ we forget $\tilde{K}_b$ and sample remaining edges in the usual way; otherwise we explore edges of $\tilde{K}_b$ via unless they were already explored in our first step. In words, this is basically saying that we take the edges from our independent K and $\tilde{K}_b$, except that if they both use the same edge then K overrides $\tilde{K}_b$. Finally, we sample all remaining edges as usual.

We can verify for ourselves that this configuration that we cook up is distributed as Bernoulli bond percolation, and next time we'll see how to use this to finish the argument! □

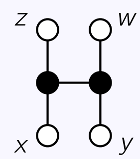
8 June 13, 2025

We've been deriving the mean-field susceptibility from the triangle condition, and we have a geometric expression for the susceptibility in terms of pivotal edges near the origin. In particular, we can always upper bound $\chi'_\beta \le |J|\chi_\beta^2$, and now we want to argue that we have a matching lower bound given the triangle inequality. This allows us to prove the following lemma:

Lemma 63

We have

$$\mathbb{P}_\beta(x \leftrightarrow z, y \leftrightarrow x, x \not\leftrightarrow y) \ge \mathbb{P}_\beta(x \leftrightarrow z)\mathbb{P}_\beta(y \leftrightarrow w) - $$



.

Proof. We'll analyze the coupling when we sample the clusters $(K_x, \tilde{K}_y)$ independently and let K_x get priority over $\tilde{K}_y$. By independence we have $\mathbb{P}_\beta(z \in K_x, w \in \tilde{K}_y) = \mathbb{P}_\beta(x \leftrightarrow z)\mathbb{P}_\beta(y \leftrightarrow w)$. Now if $y \notin K_x$ (in the event we're interested in) but $w \in \hat{K}_y$, then w is in the actual cluster K_y in the real configuration if and only if there is a path from y to w in $\tilde{K}_y$ disjoint from K_x. On the other hand, if $y \in K_x$, then every path from y to w in $\tilde{K}_y$ does indeed intersect K_x. Thus the probability of the event we're interested in can be written

$$\mathbb{P}_\beta(x \leftrightarrow z, y \leftrightarrow w, x \not\leftrightarrow y) = \mathbb{P}_\beta(z \in K_x, w \in \tilde{K}_y) - \mathbb{P}_\beta\left(\left\{z \in K_x, w \in \tilde{K}_y\right\} \setminus \{x \leftrightarrow z, y \leftrightarrow w, x \not\leftrightarrow y\}\right),$$

but this last bad event is contained in the following occurrence: x and z are connected in K, y and w are connected in a different cluster, and then there's some path from x to some point b along a path from y to w. Letting a be the last point at which this path intersects the path from x to z, the points a, b are exactly our black vertices above. So we can then apply the union bound, then BK inequality to the disjoint occurrence

$$\bigcup_{a,b} \left\{x \overset{K_x}{\leftrightarrow} a\right\} \circ \left\{a \overset{K_x}{\leftrightarrow} z\right\} \circ \left\{a \overset{K_x}{\leftrightarrow} b\right\} \circ \left\{y \overset{\tilde{K}_x}{\leftrightarrow} b\right\} \circ \left\{b \overset{\tilde{K}_x}{\leftrightarrow} w\right\},$$

and that proves the claim. In words, the only way we can fail to have the connections in two separate clusters in the real configuration is that the first cluster overwrites part of the path between y and w in the second cluster. So we can sum over the locations along the paths where we intersect and cut the path off, and we also sum over the locations from x to z where we split off. □

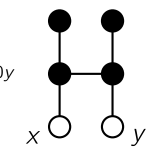
This coupling also gives us another intuitive reason for understanding why the critical dimension is 6. We saw from heuristic theory earlier that "clusters like to be four-dimensional," and we may initially expect that mean-field behavior occurs when two large clusters don't intersect each other (and hence $4+4=8$ is maybe what we naively expect). But the right way to think about this instead is that mean-field occurs when one large cluster intersects the **path between two typical points** in another large cluster, which only stops working at $d \le 4 + 2 = 6$ (since paths are like Brownian motion, which is two-dimensional). Indeed, what's important in the proof above is that the point b is actually along the path from y to w, rather than just being in the same cluster.

Turning back to our bounds on the cluster size, we can now finish the proof:

Proof of special case of Theorem 57, continued. We have (setting x to be the origin now)

$$\frac{d}{d\beta}\mathbb{E}_\beta[|K|] = \sum_{z,y,w} J_{0y}\mathbb{P}_\beta(0 \leftrightarrow z, y \leftrightarrow w, 0 \not\leftrightarrow y)$$

(here we just took the expression $\mathbb{E}_\beta\left[|K|\sum_b 1\{b \not\leftrightarrow 0\}|K_b|J_{0b}\right]$ from last lecture and expanded out by linearity of expectation, since $|K| = \sum_z 1\{z \in K\}$). Applying our inequality and using transitivity, we thus find that

$$\frac{d}{d\beta}\mathbb{E}_\beta[|K|] = |J|\mathbb{E}_\beta[|K|]^2 - \sum_y J_{0y}$$  ,

noting here that z and w are now dummy variables since they are no longer pinned. But we can now sum over those two free nodes and replace them with a factor of $\mathbb{E}_\beta[|K|]^2$, leaving us with an open triangle. Now remembering that we're **assuming the stronger condition** that the supremum $\nabla^*_{\beta_c}$ of these open triangles (over all y) is strictly less than 1 at β_c,

$$\frac{d}{d\beta}\mathbb{E}_\beta[|K|] = |J|\left(1 - \nabla^*_{\beta_c}\right)\mathbb{E}_\beta[|K|]^2.$$

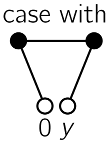
So by the same calculus argument as before, we now get the matching bound (with inequalities flipped)

$$\mathbb{E}_\beta[|K|] \le \frac{C}{|\beta - \beta_c|} \text{ for all } \beta < \beta_c,$$

and we've recovered the mean-field behavior. □

Fact 64

The general proof is quite similar to this, but it just becomes messier. We'll just do a quick sketch without mentioning all of the details.

The idea is to reduce to this case with some fiddling: the first step is to prove that if $\nabla_{\beta_c} < \infty$, then for all $\varepsilon > 0$ there is some vertex y such that  is at most ε. (We can do this with Fourier analysis and Riemann-Lebesgue on $\mathbb{Z}^d$, but in general it's harder.) We can then show that

$$\frac{d}{d\beta}\mathbb{E}_\beta[|K|] \ge c\mathbb{E}\left[|K||K_y|1\{0 \not\leftrightarrow y\}\right]$$

even when y is not connected to the origin. (We do this by saying that if we can make two clusters be distinct without changing their expectation, then via some "surgery argument" where we force edges to be open and closed, and we get two neighboring vertices with the same property by only changing finitely many edges.) Such arguments tend to be pretty annoying to do, but what we presented is the main argument.

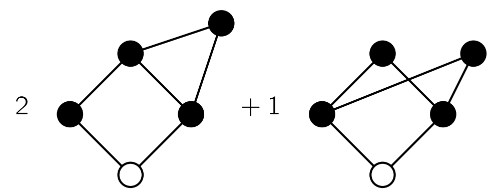
One feature is that quantitatively it's extremely poor when the triangle is large – in fact because we use Riemann-Lebesgue it's actually an ineffective bound (we have no control on the constant). Even if we know it, the constant is exponentially small in the distance we have to go because we're forcing edges along a path. So we'll state a different derivation (but omit the proof):

Proposition 65 (Hutchcroft, 2021)
(Assume modularity, but we'll basically assume that from now on.) We have the bound (remember $\chi_\beta = \mathbb{E}_\beta[|K|]$)

$$\chi'_\beta = \frac{\chi_\beta(\chi_\beta - \nabla_\beta)}{3\beta^2\nabla_\beta^2}$$

(this quantity is roughly, up to some constants, $\frac{\chi_\beta^2}{\nabla_\beta^2}$).

So indeed if the triangle is finite at β_c, we'll get the same asymptotic behavior up to constants for χ_β near criticality. And this alternative way also lets you handle cases where the triangle is actually slowly diverging (for example, $\nabla_{\beta_c-\varepsilon} \le (\log \frac{1}{\varepsilon})^{O(1)}$, then $\chi_{\beta_c-\varepsilon} \le \frac{\left(\log\frac{1}{\varepsilon}\right)^{O(1)}}{\varepsilon}$). The proof of this doesn't really even involve the triangle diagram: what we get in the denominator can be instead represented as



in place of $3\nabla_\beta^2$, and we can prove that the latter diagram is bounded by the former diagram, which is bounded by the square of a triangle.

Remark 66. *If we form the* **Green's function triangle**

$$\sum_{x,y} G(0,x)G(x,y)G(y,0)$$

for the simple random walk, then this quantity turns out to be finite if and only if $d > 6$ (exercise). We've only shown that the triangle condition a sufficient condition, but if we believe that the two-point function behaves like the Green's function in high dimension then this gives us another reason to believe $d_c = 6$.

So what we've proved is that $\nabla_{\beta_c} < \infty$ implies $\chi_{\beta_c-\varepsilon} \asymp \frac{1}{\varepsilon}$, and what we'll do next is deduce other exponent values from this. There's a number of ways of doing this – we'll start by talking about the volume tail δ and the density of the infinite cluster above criticality β. These were proven by Barsky and Aizenman (a different paper than Aizenman and Barsky) originally, but we'll show a different proof:

Theorem 67 (Hutchcroft, 2021)
Again everything is for transitive weighted graphs. For any $\beta_1 \le \beta_2$, we have

$$\mathbb{P}_{\beta_2}(|K| \ge n) \le \left(\frac{2}{n} + |J|\frac{4}{\beta_1}|\beta_2 - \beta_1|^2\right)\chi_{\beta_1}.$$

First, let's notice why result this yields mean-field behavior of the volume tail. Setting $\beta_1 = \beta_c - \varepsilon, \beta_2 = \beta_c$, we know that **if** $\chi_{\beta_c-\varepsilon} \lesssim \varepsilon^{-\gamma}$ (for γ not necessarily 1) then we get

$$\mathbb{P}_{\beta_c}(|K| \ge n) \lesssim \left(\frac{1}{n} + \varepsilon^2\right)\varepsilon^{-\gamma},$$

and now we can optimize over ε. In general note that if f is increasing and d is decreasing, then $\inf f(x) + g(x)$ is the same as $f(x_0)$ when x_0 satisfies $f(x_0) = g(x_0)$, up to a constant of 2, so this can save us some calculus if we don't care about constants. Taking $\varepsilon = n^{-1/2}$ yields

$$\mathbb{P}_{\beta_c}(|K| \geq n) \lesssim n^{-1+\frac{\gamma}{2}},$$

and in particular if $\gamma = 1$ we get the $\frac{1}{\sqrt{n}}$ mean-field behavior that we expect. More generally this gives us the inequality between critical exponents

$$\delta \leq \frac{2}{2-\gamma} \text{ if } \gamma < 2,$$

assuming these exponents are well-defined.

Additionally, we can also plug in $\beta_1 = \beta_c - \varepsilon, \beta_2 = \beta_c + \varepsilon$ and take $n \to \infty$; the result is that

$$\lim_{n\to\infty} \mathbb{P}_{\beta_c+\varepsilon}(|K| \geq n) \lesssim \varepsilon^2 \chi_{\beta_c-\varepsilon},$$

and so under our assumption the right-hand side is bounded as $\varepsilon^{2-\gamma}$. Since the left-hand side is supposed to behave as ε^β we then get

$$\beta \geq 2 - \gamma.$$

When $\gamma \neq 1$ both of these inequalities will not be sharp (and in fact when $\gamma > 2$ they tell us nothing) but they will still be true. But at mean-field what we've proven overall is the following corollary:

Corollary 68

If we have the mean-field scaling of the susceptibility $\chi_{\beta_c-\varepsilon} \asymp \frac{1}{\varepsilon}$, then we also get mean-field scaling $\mathbb{P}_{\beta_c}(|K| \geq n) \asymp \frac{1}{\sqrt{n}}$ and $\mathbb{P}_{\beta_c+\varepsilon}(|K| = \infty) \asymp \varepsilon$.

(The lower bounds turn out to always hold – we proved the latter one last time, and we can deduce the other one from combining the pointwise upper bound and the partial sum lower bound on the volume tail.)

Start of proof of Theorem 67. This proof relies on **relative entropy arguments** of Dewan and Muirhead which come out very cleanly here. (There was an old way of doing things similarly due to Newman, which used more classical percolation but produced some extraneous log factors.) First recall that the relative entropy, also called the Kullback-Leibler divergence, is defined for two measures μ, ν on the same probability space via

$$D_{\text{KL}}(\mu||\nu) = \begin{cases} \infty & \mu \text{ is not absolutely continuous with respect to } \nu, \\ \int \log\left(\frac{d\mu}{d\nu}\right) d\mu & \text{otherwise.} \end{cases}$$

(Having finite KL divergence is a strictly stronger condition than having absolute continuity $\mu \ll \nu$, though.) This is not symmetric in μ and ν, but otherwise it's kind of like a metric (in that it's zero only when $\mu = \nu$). Indeed, one of the basic facts about this quantity is **Pinsker's inequality**, which says that

$$||\mu - \nu||_{\text{TV}} \leq \sqrt{\frac{1}{2} D_{\text{KL}}(\mu||\nu)}.$$

In the original paper of Dewan and Muirhead, they in fact extract a stronger estimate from the proof that for any event A,

$$|\mu(A) - \nu(A)|^2 \leq 2D_{\text{KL}}(\mu||\nu) \max\{\mu(A), \nu(A)\}$$

(so this actually gives us some control over the low-probability events as well). Relative entropy tends to be much easier to compute than other notions of distance between probability distribution, in particular because of the **chain**

rule (plugging in random variables means we're applying to the law of the variable)

$$D_{\mathrm{KL}}((X_1, X_2)||(Y_1, Y_2)) = D_{\mathrm{KL}}(X_1||Y_1) + \mathbb{E}_{x\sim X_1}\left[D_{\mathrm{KL}}\left(X_2|X_1 = x\middle|\middle|Y_2|Y_1 = x\right)\right],$$

and of course we can generalize this to finite sequences of random variables.

Example 69

The relative entropy of Bernoulli random variables satisfies (this is just a calculus exercise)

$$D_{\mathrm{KL}}\left(\mathrm{Ber}(1-e^{-a})\middle|\middle|\mathrm{Ber}(1-e^{-b})\right) \le \frac{|a-b|^2}{2\min(a,b)}$$

We'll end up making use of these entrpoy methods via **decision trees**, which is a formalization of the idea of "exploring a cluster and picking what edge to explore next based on what we've seen."

Definition 70

A **decision tree** is a map $T : \{0,1\}^E \to (E \cup \{\mathrm{Halt}\})^{\mathbb{N}}$ from configurations ω to sequences $(T_1(\omega), T_2(\omega), \cdots)$ of edges which may terminate at finite times, such that the following are true:

1. We never look at the same edge twice.
2. $T_1(\omega) = e_1$ must be deterministic.
3. $T_n(\omega)$ is determined only by $\omega(e_1), \cdots, \omega(e_{n-1})$.
4. Once $T_m(\omega) = \mathrm{Halt}$, then all future terms in the sequence are also Halt.

Given an event $A \subseteq \{0,1\}^E$, we say that a decision tree T **Borel-computes** A if there is some Borel set $\mathcal{A} \subseteq \{0,1,\mathrm{Halt}\}^{\mathbb{N}}$ with $\omega \in A$ if and only if $(\omega(T_n(\omega)))_{n=1}^{\infty} \in \mathcal{A}$ (here $\omega(\mathrm{Halt})$ just outputs Halt).

Note that we do not need to halt – in fact, for long-range models it takes forever to even explore the neighborhood of the origin. And the point is that events depending on only finitely many edges just means "running the decision tree tells us whether the event holds or not," though in infinite-volume we have to be a bit more careful (we can't actually say that this condition of Borel-computation holds up to null sets, or else we get incorrect statements). The following result is what we'll use:

Theorem 71

If T Borel-computes A, then we have

$$|\mathbb{P}_{\beta_1}(A) - \mathbb{P}_{\beta_2}(A)|^2 \le \frac{|\beta_1 - \beta_2|^2}{\min(\beta_1, \beta_2)} \max(\mathbb{P}_{\beta_1}(A), \mathbb{P}_{\beta_2}(A)) \sum_e J_e \mathrm{Rev}_{\beta_1}(e, T),$$

where $\mathrm{Rev}_{\beta_1}(e, T)$ is the "revealment probability" that T ever queries e during its decision process.

Proof. Explore the β_1 and β_2 configurations (not coupled in any particular way – we can just have them on different probability spaces) with the same decision tree T. Say that we end up with configurations ω_1, ω_2. We have the sequence $X_1, X_2, \cdots$ obtained from $\omega_1(T_1(\omega_1)), \omega_1(T_2(\omega_1)), \cdots$ and $Y_1, Y_2, \cdots$ obtained from $\omega_2(T_1(\omega_2)), \omega_2(T_2(\omega_2)), \cdots$. By the chain rule and letting X^k, Y^k denote the first k of the X_is and Y_is,

$$D_{\mathrm{KL}}(X^{k+1}||Y^{k+1}) = D_{\mathrm{KL}}(X^k||Y^k) + \mathbb{E}_{x\sim X^k}\left[D_{\mathrm{KL}}\left(X_{k+1}|X^k = x\middle|\middle|Y_{k+1}|Y^k = x\right)\right].$$

Because we have a decision tree, both of these conditional random variables must pick the same next edge since they've seen the same output. So **either** this initial sequence has already told us to halt, so both variables will output Halt and the relative entropy is zero, **or** we have two Bernoulli random variables because the decision tree can't influence the distribution of the next revealed edge. Thus the contribution is $D_{\text{KL}}(\text{Ber}(1 - e^{-\beta_1 J_{T_{k+1}}(\omega)}), \text{Ber}(1 - e^{-\beta_2 J_{T_{k+1}}(\omega)}))$, which we can bound by Example 69. Thus

$$D_{\text{KL}}(X^{k+1}||Y^{k+1}) \le D_{\text{KL}}(X^k||Y^k) + \mathbb{E}\left[1\{\text{not halted}\}\frac{|\beta_1 - \beta_2|^2}{2\beta_1 \wedge \beta_2} J_{T_{k+1}(\omega)}\right],$$

and summing over k and telescoping yields that

$$D_{\text{KL}}\left(X||Y\right) \le \frac{|\beta_1 - \beta_2|^2}{2\beta_1 \wedge \beta_2}\mathbb{E}\left[\sum_{k=1}^{\infty} J_{T_k(\omega)} 1\{k \le \text{halting time}\}\right];$$

by rearranging, this expectation is exactly $\sum_e J_e \text{Rev}_{\beta_1}(e, T)$, as desired (and we get β_1 rather than β_2 because we're always taking conditional expectation with respect to the first distribution). And we did do some exchange of limits for relative entropy, but that's okay because it's semicontinuous in exactly the way that lets us say $D_{\text{KL}}(X||Y) = \lim_{k\to\infty} D_{\text{KL}}(X^k||Y^k)$ because we have monotonicity. ☐

We'll see the rest of the proof next time! ☐

9 June 16, 2025

Last time, we were discussing the relative entropy method for comparing probabilities (which is good for "stability-type estimates"), which implies that if $\gamma = 1$, then we do have $\delta = 2$ and $\beta = 1$. We had a notion of a decision-tree-computing event (that is, some procedure for telling us which edges to look at, where we always look at what comes next based on what we've seen so far)

Proof of Theorem 67, continued. In Theorem 71, we proved a bound on probabilities in terms of revealment probabilities. We'll define an appropriate decision tree for exploring the cluster size: let T "explore a cluster of the origin," meaning more precisely that we enumerate all edges $E = \{e_1, e_2, \cdots\}$, and at each step we query the smallest-index edge that is adjacent to the explored cluster of the origin so far but has not already been queried. So we must start off by picking the edge of smallest index adjacent to the origin, and then if this yields another vertex in the cluster then we now have a larger set of edges we can choose from for the next step.

If we do this, then the quantity on the right-hand side is

$$\text{Rev}_{\beta_1}(e, T) = \mathbb{P}_{\beta_1}\left(e \text{ adjacent to } K\right),$$

since any such edge will be revealed eventually if and only if it's visible from the cluster. Therefore if we apply Theorem 71 with the event $|K| \ge n$, we find (take $\beta_1 \le \beta_2$ without loss of generality)

$$|\mathbb{P}_{\beta_1}(|K| \ge n) - \mathbb{P}_{\beta_2}(|K| \ge n)| \le \frac{|\beta_1 - \beta_2|^2}{\beta_1}\mathbb{P}_{\beta_2}(|K| \ge n)\mathbb{E}\left[\sum_e J_e 1\{e \text{ adjacent to } K\}\right],$$

and now that expectation is essentially bounded by $|J|\chi_\beta$. So now either $\mathbb{P}_{\beta_2}(|K| \ge n) \le 2\mathbb{P}_{\beta_1}(|K| \ge n)$ or not; if that inequality does hold then using Markov's inequality gives us the $\frac{2}{n}\chi_{\beta_1}$ term in the theorem. And otherwise,

$$\mathbb{P}_{\beta_2}(|K| \ge n) - \mathbb{P}_{\beta_1}(|K| \ge n) \ge \frac{1}{2}\mathbb{P}_{\beta_2}(|K| \ge n),$$

meaning that

$$\mathbb{P}_{\beta_2}(|K| \geq n)^2 \leq \frac{4|\beta_1 - \beta_2|^2}{\beta_1} \mathbb{P}_{\beta_1}(|K| \geq n)|J|\chi_{\beta_1}.$$

But now the term $\mathbb{P}_{\beta_1}(|K| \geq n)$ on the right-hand side is at most $\mathbb{P}_{\beta_2}(|K| \geq n)$, so we can remove the square and that yields the result. □

We can also run our argument "in the other direction," which isn't necessary if our strategy was to prove mean-field behavior via the triangle condition. But it's interesting to show that different characterizations of mean-field behavior are equivalent, especially since we won't always make use of the triangle condition. (And as we mention, we usually get some other inequalities between critical exponents.)

Theorem 72

As usual, assume G is transitive and $|J| < \infty$. Then the following are equivalent:

1. $\chi_\beta \asymp \frac{1}{|\beta - \beta_c|}$ as $\beta \uparrow \beta_c$,
2. $\mathbb{P}_{\beta_c}(|K| \geq n) \asymp n^{-1/2}$ as $n \to \infty$,
3. The derivative of the susceptibility satisfies $\chi'_\beta \asymp \chi_\beta^2$ as $\beta \uparrow \beta_c$,
4. $\mathbb{E}_\beta[|K|^2] \asymp \mathbb{E}_\beta[|K|]^3$ as $\beta \uparrow \beta_c$.

Also see Theorem 83 in the next lecture.

The proof we've just completed says that (1) implies (2). We've also seen that (2) implies (4) when proving $\delta \geq 2$ via size-biasing and Markov's inequality, and also that (3) implies (1) by calculus, but so far that's all that we know.

Theorem 73 (Durrett-Nguyen, 1985)

We have for $\beta < \beta_c$ that

$$\frac{d}{d\beta} \mathbb{E}_\beta[|K|] \leq \sqrt{\frac{|J|}{\beta} \mathbb{E}_\beta[|K|] \mathbb{E}_\beta[|K|^2]}.$$

To interpret this, we saw earlier on that $\chi'_\beta \lesssim \chi_\beta^2$, and we also found by the tree-graph inequality that $\mathbb{E}_\beta[|K|^2] \lesssim \chi_\beta^3$. So if we happen to know that the second moment is much smaller than χ_β^3, plugging that in will get us a better upper bound. But at the least, it yields the implication (3) implies (4) (since if the derivative is large then the second moment must be larger).

Proof idea in the nearest-neighbor case for simplicity. For any function F of the cluster K, we can (it's usually not a good idea but it works well here) expand over every specific cluster that could be the origin

$$\mathbb{E}_p[F(K)] = \sum_C p^{\#\text{ open edges}}(1-p)^{\#\text{ closed edges}} F(C).$$

We can now differentiate both sides of this inequality in p; assuming we're allowed to exchange the order of summation and differentiation (which is fine for F of polynomial growth below criticality, because we have fast-decaying tails) we get

$$\frac{d}{dp} \mathbb{E}_p[F(K)] = \sum_C \left(\frac{\#\text{ open edges}}{p} - \frac{\#\text{ closed edges}}{1-p} \right) p^{\#\text{ open edges}}(1-p)^{\#\text{ closed edges}} F(C).$$

by the product rule applied to the two terms with ps. But now if we explore the cluster at the origin one edge at a time (like we discussed earlier) and keep track of the running value of this quantity

$$Z_n = \left(\frac{\# \text{ open up to time } n}{p} - \frac{\# \text{ closed up to time } n}{1-p} \right),$$

then Z_n is a martingale with bounded increments. In particular if our function $F(K)$ is just the cluster size $|K|$, we find

$$\frac{d}{dp}\mathbb{E}_p[|K|] \leq \mathbb{E}_p[TZ_T]$$

where T is the stopping time at which we finish exploring the cluster (keeping in mind that at each step we can only get at most one more vertex added to the cluster; more precisely we have $\frac{1}{2}|K| \leq T \leq |K|$). We can then check by Doob's inequality that $\mathbb{E}_p[TZ_T] \lesssim \sqrt{\mathbb{E}[T]\mathbb{E}[T^2]}$. □

Theorem 74 (Hutchcroft, 2019)
With the same assumptions as usual, we have

$$\frac{d}{d\beta}\mathbb{P}_\beta(|K| \geq n) \gtrsim \frac{n}{\mathbb{E}_\beta\left[|K| \wedge n\right]}\mathbb{P}_\beta(|K| \geq n).$$

Equivalently, we can write this as $\frac{d}{d\beta} \log \mathbb{P}_\beta(|K| \geq n) \gtrsim \frac{n}{\mathbb{E}_\beta[|K| \wedge n]}$.

We can use this to derive that (2) implies (1) – indeed, for any $\beta < \beta_c$ we can bound the denominator $\mathbb{E}_\beta[|K| \wedge n]$ by the value at criticality, so if we assume the volume tail in condition (2), we have $\mathbb{E}_\beta\left[|K| \wedge n\right] \lesssim \sqrt{n}$ and thus

$$\frac{d}{d\beta}\log \mathbb{P}_\beta(|K| \geq n) \gtrsim \sqrt{n} \implies \mathbb{P}_\beta(|K| \geq n) \leq \mathbb{P}_{\beta_c}\left(|K| \geq n\right) e^{-c|\beta-\beta_c|\sqrt{n}} \lesssim n^{-1/2}e^{-c|\beta-\beta_c|\sqrt{n}},$$

and we can check with another calculus exercise that this actually implies the desired result. But then if know that (1) holds, then plugging that in for the bound instead lets us improve the exponent and get a sharp estimate of the form we were writing down in heuristic scaling theory:

$$\mathbb{P}_\beta(|K| \geq n) \lesssim n^{-1/2}e^{-c|\beta-\beta_c|^2 n}$$

As an exercise, we can also try to use the lower bound with a possibly different constant.

This same inequality can also be used for the final implication we need to tie everything together – that is, (4) implies (3). Indeed, take the inequality Theorem 74, bound $\mathbb{E}_\beta[|K|] \wedge n \leq \mathbb{E}_\beta[|K|]$, and sum over n. Again glossing over analysis details, this yields

$$\frac{d}{d\beta}\mathbb{E}_\beta\left[|K|\right] \gtrsim \frac{\mathbb{E}_\beta[|K|^2]}{\mathbb{E}_\beta[|K|]},$$

since summing over $n\mathbb{P}(|K| \geq n)$ is exactly the second moment. (We should compare this with the Durrett-Nguyen bound.) So even for a specific value of β, $\chi'_\beta \asymp \chi^2_\beta$ indeed holds if and only if $\mathbb{E}_\beta[|K|^2] \asymp \mathbb{E}_\beta[|K|]^3$.

Remark 75. *It's an open question to show that the triangle condition is equivalent to the other conditions (we only saw that it implies mean-field behavior). There's also a similar question for the slightly supercritical cluster scaling* $\mathbb{P}_{\beta+c}(|K| = \infty) \asymp \varepsilon$.

Theorem 74 is a consequence of what is called the **OSSS inequality**, which is named after O'Donnell, Saks, Schramm and Servedio. This is a fact about boolean functions (as many of the considerations with martingales or correlation so far have been).

Theorem 76 (OSSS inequality)
Suppose $A, B \subseteq \{0,1\}^E$ are two events, and suppose we have a decision tree T which computes the event B. Then
$$|\text{Cov}_\beta(A, B)| \le \sum_e \text{Rev}_\beta(e, T)\text{Cov}_\beta(A, \omega(e)).$$

(We often apply this in the case $A = B$, but not here.) We can think of this as T computing 0 or 1 in a way that is well-correlated with A by letting B be the event that the tree outputs 1. And if A is increasing (meaning that all covariances on the right-hand side are nonnegative), we can rearrange this to find

$$\sum_e \text{Cov}_\beta(A, \omega(e)) \ge \frac{|\text{Cov}(A, B)|}{\max_e \text{Rev}_\beta(e, T)}.$$

At the moment this is useless because the revealment probability is just 1 for all edges adjacent to the origin. But the inequality is linear, and thus we can also use some external randomness which allows us to potentially make the maximum revealment probability smaller.

Earlier when we talked about Russo's formula, we saw that derivatives of increasing events are expressed in terms of pivotality. But pivotality probabilities can be written in terms of these covariances, and if we do that we end up with

$$\frac{d}{d\beta}\mathbb{P}_\beta(A) = \sum_e \frac{J_e}{e^{\beta J_e} - 1}\text{Cov}(A, \omega(e)),$$

and we should think of this as being approximately the sum of the covariances if β is bounded away from 0 and ∞ (which is okay because we care about β near criticality). So what this says is that

$$\frac{d}{d\beta}\mathbb{P}_\beta(A) \gtrsim \frac{|\text{Cov}(A, B)|}{\max_e \text{Rev}_\beta(e, T)},$$

and in some sense this goes in the opposite direction as the Dewan and Muirhead result, which says that small total revealment keeps probability about the same (in our case, small maximal revealment means probabilities are large).

Proof sketch of Theorem 74. Introduce an independent ghost field ξ of intensity $\frac{1}{n}$, and consider the events $A = \{|K| \ge n\}$ and $B = \{o \leftrightarrow \xi\}$. We let our decision tree T reveal the ghost field and then explore the cluster of every vertex in ξ. We'd like to apply OSSS to the larger product space with both revealing edges and ghost vertices; even though all ghost vertices are always revealed, the covariance of the event A with each such revealment is zero and thus we do not need to include those terms. So the key fact that we use is that

$$\text{Rev}_\beta(e, T) = \mathbb{P}(e^- \text{ or } e^+ \text{ are connected to } \xi) \le 2\mathbb{P}(0 \leftrightarrow \xi)$$

by translation-invariance. Now we can check that the probability of the origin being connected to a ghost is comparable to $\frac{1}{n}\mathbb{E}[|K| \wedge n]$, so plugging this in yields the result after some further computation. ☐

The OSSS inequality is quite a powerful tool in current work; it works not just for Bernoulli percolation, and in fact can use this to get sharpness of the phrase transition even for something like the random cluster model.

We'll now turn to the **intrinsic geometry of clusters**:

Definition 77

The **intrinsic distance** or **chemical distance** is given by

$$d_{\text{int}} = \text{graph distance on the cluster.}$$

Letting $B_{\text{int}}(x, r)$ be the ball of radius r around x under d_{int} (and taking the ball to be at the origin if not otherwise specified), we'll show that when we have mean-field critical behavior, various geometric things associated to the cluster viewed as a metric space behave similar to percolation on the binary tree.

Lemma 78

Let $\partial B_{\text{int}}(r)$ denote the points exactly at distance r from the origin. We have (for a transitive graph)

$$\mathbb{E}_{\beta_c}\left[|\partial B_{\text{int}}(r)|\right] \geq 1.$$

In particular, this implies that $\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(r)|\right] \geq r + 1$ (by summing over levels).

Proof. We claim that it suffices to prove for all n, m that

$$\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n+m)|\right] \leq \mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n)|\right] \mathbb{E}_{\beta_c}\left[|\partial B_{\text{int}}(m)|\right] + \mathbb{E}_{\beta_c}\left[|B_{\text{int}}(m)|\right].$$

Indeed, similarly to how we proved sharpness, suppose for the sake of contradiction that there is some r with $\mathbb{E}_{\beta_c}[|\partial B_{\text{int}}(r)|] \leq 1$. Then rearranging the inequality yields (plugging in $m = r$ and using that $\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n)|\right] \leq \mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n+r)|\right]$

$$\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n+r)|\right] \leq \frac{\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(r)|\right]}{1 - \mathbb{E}_{\beta_c}\left[|\partial B_{\text{int}}(r)|\right]}.$$

Taking $n \to \infty$, we see that the right-hand side is finite, while the left-hand side approaches the susceptibility. Since the susceptibility should be infinite at criticality, we get a contradiction.

One way to prove this claim is to use Reimer's inequality by splitting up the path before and after exiting, but we can also calculate directly that

$$\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n+m) \setminus B_{\text{int}}(m)|\right] \leq \mathbb{E}_{\beta_c}\left[|\partial B_{\text{int}}(m)|\right] \mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n)|\right],$$

using that because we've revealed some negative information about our edges when finding the boundary at level m, that can only make it harder to connect outward. □

Lemma 79

We have

$$\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n)|\right] \lesssim \chi_{\beta_c - 1/n}.$$

Proof. If $x \in B_{\text{int}}(n)$, then there's a decent chance that $x \in K$ is still in the cluster at $\beta_c - \frac{1}{n}$ under the standard monotone coupling, since for all of the n edges from the origin to x there's a constant-size probability that all of those edges are still there. □

Corollary 80
If $\chi_\beta \asymp \frac{1}{|\beta-\beta_c|}$, then the intrinsic ball has size scaling as expected:

$$\mathbb{E}_{\beta_c}\left[|B_{\text{int}}(n)|\right] \asymp n+1.$$

We'll next discuss the same thing as what we did with the binary tree – in that case the expected cluster size is trivial since actually we expect exactly one vertex at each level, and so the inequality is tight. What we'll consider next is the intrinsic one-arm exponent: define

$$Q_r = \mathbb{P}_{\beta_c}\left(\partial B_{\text{int}}(r) \text{ is nonempty}\right).$$

It turns out to be annoying that this is not a monotone event (it's possible that points become closer to the origin and thus we might reduce distances by adding edges). So for the arguments it's useful to introduce a modified version of this

$$Q_r^* = \sup_{H \text{ subgraph}} \mathbb{P}^H_{\beta_c(G)}\left(\partial B_{\text{int}}(r) \text{ is nonempty}\right),$$

where $\mathbb{P}^H_{\beta_c(G)}$ means we only percolate on H and thus are forbidden from using some other edges.

Remark 81. *Q_r and Q_r^* are morally the same at criticality, but at supercriticality it's possible to have much more "winding paths" in the cluster and thus Q_r^* will be much larger than Q_r.*

Theorem 82 (Kozma-Nachmias)
If $\mathbb{P}_{\beta_c}(|K| \geq n) \asymp n^{-1/2}$, then $Q_r^* \lesssim \frac{c}{r}$.

(As an exercise, we can check that we always have $Q_r \gtrsim \frac{c}{r}$ with no assumptions besides transitivity.) We'll prove this next time, and then after that we'll get rigorous versions of hyperscaling relations (in inequality form).

10 June 17, 2025

At the end of last lecture, we started discussing "intrinsic geometry exponents:" looking at the graph distance structure, the expected size of $B_{\text{int}}(r)$ always grows at least linearly and is bounded from above by the susceptibility at $\beta_c - \frac{1}{r}$ (so we get matching bounds in mean-field).

Proof of Theorem 82. Recall that we're working with a supremum over subgraphs to help with monotonicity issues, and this means we also prove a stronger statement than just working with Q_r. We'll do a general version **assuming any power-law bound** of the form

$$\mathbb{P}_{\beta_c}(|K| \geq n) \lesssim n^{-1/\delta}$$

(in particular caring about the case $\delta = 2$). We'll work out the bound (without saying ahead of time what it is) using the **bootstrap method**. There are lots of precise ways to do this and the exact one to use depends on context; the one we'll use is to define

$$Q_r^*(\beta) = \sup_{H \text{ subgraph}} \mathbb{P}^H_\beta(|\partial B_{\text{int}}(r)| > 0),$$

and we claim it suffices to bound this quantity for every subcritical $\beta < \beta_c$ with constants that don't blow up as $\beta \uparrow \beta_c$. (Then we can take the limit by monotone convergence.) To do that, we want to prove a bound of the form $Q_r^*(\beta_c) \lesssim r^{-\theta}$ for some θ we don't know yet.

For this, we fix some $\theta > 0$ and we'll later choose it to be the largest constant that will make the proof work. Define

$$A(\beta) = \inf\left\{A : Q_r^*(\beta) \le Ar^{-\theta} \text{ for all } r \ge 1\right\}.$$

For all $\beta < \beta_c$ we have an exponential tail on the volume, hence certainly on the radius, so $A(\beta)$ will always be finite for each β. The hope is then to show something of the form $A(\beta) = CA(\beta)^{1-\varepsilon}$, which rearranges to $A(\beta) \le C^{1/\varepsilon}$ and thus implies that $A(\beta_c) \le C^{1/\varepsilon}$ as well. (Again, there are slight variations on these methods – in other contexts we might want to do an induction on different radius scales instead of approaching criticality.)

The first observation to make (which was always used in our analysis on the binary tree) is that

$$\begin{aligned}\mathbb{P}_\beta^H(|\partial B_{\text{int}}(3r)| > 0) &\le \mathbb{P}_\beta^H(|K| \ge n) + \mathbb{P}_\beta^H(|\partial B_{\text{int}}(3r)| > 0, |K| \le n)\\ &\le \mathbb{P}_\beta^H(|K| \ge n) + \mathbb{P}_\beta^H(|\partial B_{\text{int}}(3r)| > 0, |K| \le n)\\ &\le \mathbb{P}_\beta(|K| \ge n) + \mathbb{P}_\beta^H(|\partial B_{\text{int}}(3r)| > 0, |K| \le n)\end{aligned}$$

because being in a large cluster is a monotone event. Now reaching $3r$ means we need to get to r, then $2r$, then $3r$, and if our cluster is small we need something in the middle r levels of size at most $\frac{n}{r}$. So we can explore our cluster by breadth-first search until we find such a layer, and we can bound the resulting probability by Q_r^* (reaching level r) times 1 (the chance of finding a small layer) times $\frac{n}{r}Q_r^*$ (since each of our $\frac{n}{r}$ things has probability at most Q_r^* of further reaching layer $3r$). So plugging this in and taking the supremum over H yields

$$\begin{aligned}Q_r^* &\le \mathbb{P}_\beta(|K| \ge n) + \frac{n}{r}(Q_r^*)^2\\ &\lesssim n^{-1/\delta} + \frac{n}{r}(Q_r^*)^2\end{aligned}$$

for all $\beta \le \beta_c$. Now using the definition of $A(\beta)$ and taking $\beta < \beta_c$ (so that $A(\beta)$ is finite), we have

$$Q_{3r}^* \le Cn^{-1/\delta} + C\frac{n}{r}A(\beta)^2 r^{-2\theta}.$$

We can now optimize over the free parameter n, and setting the two terms equal yields $n^{-1/\delta} = \frac{n}{r}A^2 r^{-2\theta} \implies n^{1+1/\delta} = r^{1+2\theta}/A^2$. This means we plug in $n = r^{(1+2\theta)\frac{\delta}{\delta+1}}A^{-2\frac{\delta}{\delta+1}}$, and the result is that

$$Q_{3r}^* \le \tilde{C}r^{-\frac{1+2\theta}{\delta+1}}A^{\frac{2}{\delta+1}}.$$

Now if $\delta > 1$, the power of A is less than 1, which is good because it means we have a constant getting smaller. And we want the bound to come out in the same way in terms of the r-exponent scaling, so in fact we should be setting

$$\frac{1+2\theta}{\delta+1} = \theta \implies \boxed{\theta = \frac{1}{\delta-1}}.$$

What we've thus deduced is that

$$Q_{3r}^* \le \tilde{C}A^{\frac{2}{\delta+1}}r^{\theta} \implies Q_r^* \le \tilde{\tilde{C}}A^{\frac{2}{\delta+1}}r^{-\theta}.$$

Because $A(\beta)$ is defined to be the best constant we could put in this equation, we find that

$$A(\beta) \le \tilde{\tilde{C}}A(\beta)^{2/(\delta+1)},$$

so $A(\beta)$ is bounded by a constant and taking $\beta \uparrow \beta_c$ yields the result

$$Q_r^*(\beta_c) \lesssim r^{-1/(\delta-1)}.$$

In particular, starting with $\delta = 2$ yields the Kozma-Nachmias result. ☐

There are often "self-improving constants" of this form in similar arguments. Another way we may see this formulated (as we saw in short talks on the lace expansion) is to prove a "forbidden zone argument" where a constant being bounded by 3 means it must actually be bounded by 2, hence by intermediate value theorem it can never go above 2. (We do of course usually need some continuity.)

We'll prove one more result, which adds to Theorem 72:

Theorem 83

The conditions of Theorem 72 are further equivalent to the following condition:

5. $\mathbb{E}\left[|B_{\text{int}}(r)|\right] \asymp r$ and $Q(r) \asymp \frac{1}{r}$.

Proof. We've just seen a proof that the other conditions imply (5), so now we just need to show the other direction. We'll show that (5) implies (2); this is again similar to our argument on the binary tree in which we said that size being at least n implies radius of at least $\sqrt{n}$. If we want to have large volume, we can split it into two cases

$$\begin{aligned}\mathbb{P}_{\beta_c}(|K| \geq n) &\leq \mathbb{P}_{\beta_c}(|\partial B_{\text{int}}(r)| > 0) + \mathbb{P}_{\beta_c}\left(|B_{\text{int}}(r)| \geq n\right) \\ &\lesssim \frac{1}{r} + \frac{r}{n}\end{aligned}$$

(last step by Markov's inequality and our assumption on the boundary size), and taking $r = \sqrt{n}$ optimizes this inequality and gives the correct $n^{-1/2}$ scaling. ☐

So again, we see the phenomenon of "two-dimensional clusters in internal graph distance, but four-dimensional clusters in the ambient space." Notice that we still only have one example, the binary tree, where we've proved that these critical exponents even exist, but soon we'll show some more models where that proof is indeed possible.

Before looking at such examples, we'll turn to the promised hyperscaling inequalities (which only hold in low dimension and also be seen as scaling relations depending on an additional ♯), which will always hold but only be sharp in low dimensions.

Theorem 84 (Universal tightness theorem (Hutchcroft, 2020))

For any weighted graph $G = (V, E, J)$ (not assuming transitivity) and any finite set $\Lambda \subseteq V$, define the random variable

$$|K_{\max}(\Lambda)| = \max\left\{|K_x \cap \Lambda| : x \in V\right\}.$$

(Note that in this definition the clusters themselves can go outside Λ.) Define $M_\beta(\Lambda)$ to be the "median value but with $\frac{1}{e}$" (this could be called the "edian")

$$M_\beta(\Lambda) = \min\left\{n \in \mathbb{N} : \mathbb{P}_\beta(|K_{\max}(\Lambda)| \geq n) \leq \frac{1}{e}\right\}.$$

Then we have the following:

1. $\mathbb{P}_\beta(|K_{\max}| \geq \alpha M_\beta) \leq e^{-\alpha/9}$.
2. $\mathbb{P}_\beta(|K_{\max}| \geq \varepsilon M_\beta) \leq 27\varepsilon$.
3. For any specific vertex $x \in V$, $\mathbb{P}_\beta(|K_x \cap \Lambda| \geq \alpha M_\beta) \leq e^{1-\alpha/9}\mathbb{P}_\beta(|K_x \cap \Lambda| \geq M_\beta)$.

The idea of this statement is that "the size of the largest cluster" is very well-behaved, and the constants don't depend on the graph itself.

Corollary 85

Assume that we have a global volume tail bound $\sup_x \mathbb{P}_{\beta_c}(|K| \geq n) \lesssim n^{-1/\delta}$ (by taking this supremum we don't need to assume transitivity). Then

$$\mathbb{P}_{\beta_c}(|K_x \cap \Lambda| \geq n) \lesssim n^{-1/\delta} e^{-n/(18M_{\beta_c}(\Lambda))}.$$

Indeed, we just apply part (3) of theorem to get $(n \wedge M_\beta)$ in place of n and an exponent of 9 instead of 18, and we just need to do a little bit of calculation to turn it into the result we want. So this looks a lot like our two-sided scaling limit assumptions involving some typical large size of the cluster – it's not clear that M_β is the right quantity to look at, but it does give us an inequality. And conjecturally it is of the correct order in low dimensions when Λ is a box, a high-dimensional torus, or an Erdos-Renyi graph.

Fact 86

On the other hand, in a high-dimensional box, we expect a typical large cluster to satisfy $\zeta(r) \asymp r^4$, while $M_{\beta_c}([-r, r]^4) \asymp r^4 \log r$. Aizenman has proved the upper bound here, but the lower bound is still open. And in high dimensions we should be using the squared susceptibility instead of M_β.

We can then take this to get hyperscaling inequalities:

Corollary 87

Again assume that $\sup_x \mathbb{P}_{\beta_c}(|K_x| \geq n) \lesssim n^{-1/\delta}$. Then we have

$$M_{\beta_c}(\Lambda) \lesssim |\Lambda|^{\delta/(\delta+1)}, \quad \frac{1}{|\Lambda|} \sum_{v \in \Lambda} \mathbb{P}_{\beta_c}(x \leftrightarrow v) \lesssim |\Lambda|^{-2/(\delta+1)}.$$

One thing we can see is that these bounds are indeed optimal for Erdos-Renyi (in which we have a $\frac{1}{\sqrt{n}}$ tail on the volume, and if we take Λ to be the entire vertex set, we have cluster size $|\Lambda|^{2/3}$ and susceptibility of order $|\Lambda|^{1/3}$ at criticality.

Proof. Fix a vertex u. We have

$$\begin{aligned}\sum_{v \in \Lambda} \mathbb{P}(u \leftrightarrow v) &= \mathbb{E}\left[|K_u \cap \Lambda|\right] \\ &= \sum_n \mathbb{P}(|K_u \cap \Lambda| \geq n) \\ &\lesssim \sum_n n^{-1/\delta} e^{-n/(18M)}\end{aligned}$$

by the previous corollary. And now this "rapidly decaying factor" has essentially the same effect as summing up until the characteristic scale, so this behaves as $\sum_{n=1}^{M} n^{-1/\delta} \lesssim M^{(\delta-1)/\delta}$. So if we sum over u as well, we end up with

$$\sum_{u,v \in \Lambda} \mathbb{P}(u \leftrightarrow v) \lesssim |\Lambda| M^{(\delta-1)/\delta}.$$

On the other hand, we can also use linearity of expectation and sum over clusters:

$$\begin{aligned}\sum_{u,v\in\Lambda}\mathbb{P}(u\leftrightarrow v) &= \mathbb{E}\left[\sum_{u,v}1\{u\leftrightarrow v\}\right] \\ &= \mathbb{E}\left[\sum_{C \text{ cluster}}|C|^2\right] \\ &\geq \mathbb{E}\left[|K_{\max}(\Lambda)|\right] \\ &\gtrsim M^2,\end{aligned}$$

and so if we compare our bounds we get

$$M^2 \lesssim |\Lambda| M^{(\delta-1)/\delta} \implies M \lesssim |\Lambda|^{\delta/(\delta+1)},$$

which was the first thing we wanted to prove, and so now feeding this back into our earlier upper bound yields

$$\sum_{u,v\in\Lambda}\mathbb{P}(u\leftrightarrow v) \lesssim |\Lambda| M^{(\delta-1)/\delta} \lesssim |\Lambda|^{(\delta-1)/(\delta+1)},$$

which implies the result. ☐

In the case where Λ is a box in $\mathbb{Z}^d$, we find that a typical large cluster has (recall that we're saying the typical size is $\zeta(r) \asymp r^{d_f}$)

$$\frac{d_f}{d} \leq \frac{\delta}{\delta+1}, \quad (2-\eta) \leq d\frac{\delta-1}{\delta+1}$$

where we assume the two-point function scales as $r^{-d+2-\eta}$ and $|\Lambda|^{-2/(\delta+1)}$ scales as $r^{-2d/(\delta+1)}$. Both are conjectured to be equalities for $d < d_c$.

Corollary 88

For $d < 6$, we cannot have $\delta = 2, \eta = 0$, because this would contradict our latter relation above. So this is a rigorous proof of the inequality $d_c \geq 6$.

Remark 89. *In $d = 6$, we expect to have sharpness in the inequalities up to a* $\log\log$ *factor because there are* $\log$ *many clusters. Also, it turns out tori and boxes behave differently – when we impose periodic boundary conditions we end up with $O(1)$ big clusters even in high dimensions.*

We'll now return to the proof of the universal tightness theorem:

Proof of Theorem 84. We'll deduce this from the BK inequality, together with the following combinatorial lemma.

Lemma 90

Let $H = (V, E)$ be a connected graph with at least one edge, and let A be a set of vertices (for example, a cluster). If $|A| \geq 3^k$ for some k, then there exists $m \geq 3^{k-1}+1$ and some collection $\{E_i : 1 \leq i \leq m\}$ of disjoint subsets of the edges E such that (1) every vertex $v \in V$ is adjacent to at least one edge in $\bigcup_i E_i$, (2) the graph spanned by E_i is connected for all i, and (3) if we define V_i to be the set of endpoints of E_i, then

$$3^{-k} \leq \frac{|A\cap V_i|}{|A|} \leq 3^{-k+1}.$$

In other words, we can take our graph and split it up into connected subgraphs, so that the special set A we care about is equally distributed among the parts.

Proof sketch of lemma. There are essentially two steps, which are to split into two pieces and then iterate. (If we can take a graph and find two edge-disjoint subgraphs that cover the vertex set and partition the set A into roughly equal pieces, then we apply the same algorithm to each piece until it would be too small.) More precisely, for this first step, we wish to find disjoint E_1, E_2, each a connected edge set, such that $V_1 \cup V_2 = V$ and such that $\frac{1}{3} \leq \frac{|A \cap V_i|}{|A|} \leq \frac{2}{3}$.

Without loss of generality we can always assume H is a tree (we can just restrict our graph to a spanning tree). We will grow one edge set and have it eat the other – we get a sequence starting with $E_1^0 = \varnothing, V_1^0 = \{V_0\}$ for some arbitrary starting vertex, and then we preserve the property that it and its complement are both connected while eating up as little of the set as possible. This means that at each step (from (E_1^n, V_1^n) to (E_1^{n+1}, V_1^{n+1})), we need to add an entire branch of edges (along with the adjacent vertices), and we can always pick the branch with the least number of vertices added.

But we can check that at the first time $|A \cap V_1^i| \geq \frac{|A|}{3}$, we must actually have $|A \cap V_i| \leq \frac{2|A|}{3}$ (if we jumped from less than $\frac{1}{3}$ to more than $\frac{2}{3}$, we could have eaten a smaller branch instead).

From here, we dynamically iterate each piece and split into smaller and smaller pieces; we won't go from a density 3^{-k+1} all the way down to below a density 3^{-k} in one splitting by the same logic. $\square$

Returning to the main proof now, take an integer $K \geq 1$ and a real number $\lambda > 0$. The event $\left\{|K_{\max}(\Lambda)| \geq 3^k \lambda\right\}$ is contained in the disjoint occurrence

$$\{|K_{\max}(\Lambda)| \geq \lambda\} \circ \cdots \circ \{|K_{\max}(\Lambda)| \geq \lambda\}$$

where we have $m \geq 3^{k-1} + 1$ of these things being put together. (Indeed, we apply the lemma with H being the cluster that attains the bound $|K_{\max}(\Lambda)| \geq 3^k \lambda$ and A the cluster; each disjoint edge set is a separate witness for the event $|K_{\max}(\Lambda)| \geq \lambda$.) Similarly, we also have

$$\left\{|K_x \cap \Lambda| \geq 3^k \lambda\right\} \subseteq \{|K_{\max}(\Lambda)| \geq \lambda\} \circ \cdots \circ \{|K_{\max}(\Lambda)| \geq \lambda\} \circ \{K_x \cap \Lambda| \geq \lambda\},$$

where we now have $m-1$ copies instead of m, since the set of edge endpoints covers the vertex set, one of them covers x, and the others are just a witness for the largest cluster being large. So using the BK inequality and optimizing over K and λ yields the result (since if λ is larger than the edian, each event $\{|K_{\max}(\Lambda)| \geq \lambda\}$ has probability at most $\frac{1}{e}$.).

Finally, the lower-tail statement actually uses a similar bound – we get something like $\mathbb{P}(|K_{\max}(\Lambda)| \geq M) \leq \mathbb{P}(|K_{\max}(\Lambda)| \geq \varepsilon M)^{1/\varepsilon}$, and the probability on the right-hand side is close to 1. $\square$

We'll see how to prove the hyperscaling relations in two dimensions next time!

11 June 19, 2025

Last time, we discussed the universal tightness theorem and its implications for hyperscaling inequalities. We expect that these are equalities in low dimensions but not high dimensions, and we also saw that this implies that we do not have $\delta = 2, \eta = 0$ when $d < 6$. However, do note that it is an open problem to ask whether $\delta = 2$ on its own – it shouldn't happen in low dimensions, but we can't prove that.

Example 91

We'll now turn to understanding how this all works in the two-dimensional case. In many courses, we'd spend the entire time on $d = 2$, but that isn't the focus of this course and we'll just be discussing it a bit here.

For concreteness, we'll consider bond percolation on $\mathbb{Z}^2$ – we don't actually know how to prove the critical exponents here (we only know it for site percolation on the triangular lattice), but it's analogous to the work we've been doing so far.

Fact 92

For bond percolation on $\mathbb{Z}^2$, we have $p_c = \frac{1}{2}$. We'll only consider criticality at percolation here, and when we say "box of radius n" we mean $[-n, n]^2$ and "annulus (r, R)" we mean $[-R, R]^2 \setminus [-r, r]^2$. Then we have the following facts:

- The probability that there is an open path surrounding a box of radius n contained within an annulus $(n, 2n)$ is bounded from below by some $c_1 > 0$.
- Similarly, the probability that there is a path connecting the box of radius n to the box of radius $4n$ is bounded from below by some $c_2 > 0$.

These are called the **RSW estimates** – the key fact is that crossing an n-by-n box has probability $\frac{1}{2}$ by a duality argument. What we'll try to do is show how these facts imply inequalities on critical exponents. First, define the one-arm probability

$$A(r) = \mathbb{P}\left(0 \leftrightarrow \partial[-r, r]^2\right).$$

We first want to show that this function is well-behaved (it doesn't drop drastically if r changes by a constant factor), via the "doubling" or 'halving" relation

$$A(2r) \geq cA(r).$$

This is like asking for polynomial decay but much weaker, and the way we prove it is as follows: if we have a connection to level r, and we also have a circuit around $\frac{r}{2}$ within the annulus $(\frac{r}{2}, r)$, and we also have a connection from the $\frac{r}{2}$-box to the $2r$ box, then (this is just a topological fact about two dimensions) then we can put those three things together and get a connection (first follow the path until we hit the circuit, then follow the circuit until we hit the outward connection). These three events are increasing, and therefore by FKG and the fact above we find that $A(2r) \geq c_1 c_2 A(r)$ as desired.

On the other hand, we can show that the probability of a connection from n to $4n$ is strictly smaller than $1 - c$ for some c; iterating these inequalities tells us that $r^{-c_2} \leq A(r) \leq r^{-c_1}$ for some constants c_1, c_2.

So what we'll do now is assume that we have a well-defined exponent $A(r) \asymp r^{-1/\rho}$. We claim that

$$\mathbb{P}(x \leftrightarrow y) \asymp A(||x - y||)^2.$$

Indeed, for the lower bound we can draw two boxes of radius $\frac{|x-y||}{4}$ around x, y, and both x and y need to have paths out to those boxes. So $\mathbb{P}(x \leftrightarrow y) \leq A\left(\frac{||x-y||}{4}\right)^2 \asymp A(||x - y||)^2$ by our doubling inequality. On the other hand, for the upper bound we can draw a big annulus $(R, 2R)$ around both x and y (so that both x and y are inside the box of radius R), and we'll look at the event that both of those points are connected to the large box **and** that there's a circuit within the annulus $(R, 2R)$. This implies connection (by following the path, then circuit, then path), so by

FKG again we thus get $\mathbb{P}(x \leftrightarrow y) \gtrsim A(||x-y||)^2$, as desired. We have thus shown that

$$A(r) \asymp r^{-1/\rho} \iff \mathbb{P}(x \leftrightarrow y) \asymp ||x-y||^{-\eta}, \quad \eta = \frac{2}{\rho}.$$

(The $-d+2$ factor no longer shows up because $d=2$.) So for this particular example, it's easy to prove assuming these RSW estimates.

Next, we want to understand the volume tail. We've seen the following kind of argument before: just like in the mean-field case, either we connect to some large distance or get the volume within a ball, so

$$\begin{aligned}\mathbb{P}(|K| \ge n) &\le A(r) + \mathbb{P}(|K \cap [-r,r]^2| \ge n) \\ &\le A(r) + \frac{1}{n} r^{2-\eta} \\ &\lesssim r^{-\eta/2} + \frac{1}{n} r^{2-\eta}\end{aligned}$$

by Markov's inequality, where we're assuming that we can get the expectation of the volume size by summing the specified scaling of the two-point function. So now if we optimize over r (by setting the two terms equal), we want to take $r^{-\eta/2} = \frac{r^{2-\eta}}{n} \implies r = n^{2/(4-\eta)}$, and we get that

$$\mathbb{P}(|K| \ge n) \lesssim n^{-\eta/(4-\eta)}.$$

The left-hand side is supposed to scale as $n^{-1/\delta}$, so what this implies is the inequality

$$\frac{1}{\delta} \ge \frac{\eta}{4-\eta} \implies \delta \le \frac{4-\eta}{\eta}.$$

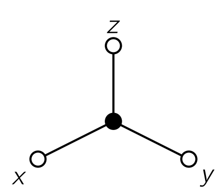
Recall that we also had the bound $2-\eta \le d\frac{\delta-1}{\delta+1}$ from universal tightness (with $d=2$), but it turns out these are actually matching lower and upper bounds: rearranging yields $\delta \ge \frac{d+(2-\eta)}{d-(2-\eta)} = \frac{4-\eta}{\eta}$. So actually the very simple bound we found is tight: assuming all of these exponents exist we get that in $d=2$ this universal tightness bound actually holds.

Remark 93. *Recall that the conjecture for 2D percolation in general is to have $\eta = \frac{5}{24}$; taking this relation yields $\delta = \frac{4-\frac{5}{24}}{\frac{5}{24}} = \frac{91}{5}$. And Kesten more generally showed relations among γ and ν and the other off-criticality exponents (using other more direct techniques than universal tightness), which is more difficult than what we saw here.*

Example 94

We'll next discuss hyperscaling for the k-point function. Recall that the k-point function $\tau_\beta(x_1, \cdots, x_k)$ is the probability that all k of those points are connected (in the same cluster).

Beginning with the three-point function, we've already seen the tree-graph inequality

$\tau_\beta(x, y, z) \le$ 

This is always true, but it's further conjectured to be of the correct order if and only if $d > d_c$. Recalling how we proved the tree-graph inequality, we said that "if x, y are connected and so are x, z, there is some last point where they intersect and we apply BK." It's possible that we're overcounting from the different ways of intersecting, and it's also possible that BK is quite bad (because once we form this first path it becomes very hard to form the next one because it takes up a lot of space). Once we have the lace expansion, it's not so bad to prove that for $d > d_c$ we're good; what's much harder is to show that $\tau_{\beta_c}(x, y, z)$ is actually asymptotic to some constant times that diagram (as the minimum distance between points goes to infinity). Estimates of the three-point function's Fourier transform are due to Hara and Slade, and the pointwise version will be in an upcoming paper of Professor Hutchcroft's with Blanc and Bernadie.

What we'll discuss today is instead the **low-dimensional picture**. We'll start by explaining this in a kind of complicated "physics way" and then see how it's actually not so mysterious for $d = 2$. We've mentioned that we can talk about criticality at the level of critical exponents but also at the level of scaling limits (where we try to take a limit of the whole model). As mathematicians we like to talk about the actual limiting object, but in physics they tend to think about the limit in terms of k-point functions (observables of the system). What we do is consider points in $\mathbb{R}^d$ and define a rescaling (here the brackets indicate "nearest integer part")

$$T(x_1, x_2, \cdots, x_k) = \lim_{\lambda\to\infty} f(\lambda)^k \tau_{\beta_c}([\lambda x_1], \cdots, [\lambda x_k])$$

for some function $f(\lambda)$ which governs scaling for all k-point functions. Here we should have $f(\lambda)$ be regularly varying of index $\sigma = \frac{d-2+\eta}{2}$ (because for two points we should cancel out the power law decay of the two-point function); often we will actually just take $f(\lambda) = \lambda^\sigma$, but sometimes there are logs to account for and so we allow more generality. **Let's suppose we can do this** to get a nondegenerate limit (which is a strong assumption). Then if this continuum k-point function T is well-defined, it must satisfy the scale-invariance condition

$$\begin{aligned} T(\alpha x_1, \cdots, \alpha x_k) &= \lim_{\lambda\to\infty} f(\lambda)^k \tau_{\beta_c}([\lambda\alpha x_1], \cdots, [\lambda\alpha x_k]) \\ &= \lim_{\tilde{\lambda}\to\infty} f\left(\frac{\tilde{\lambda}}{\alpha}\right)^k \tau_{\beta_c}([\tilde{\lambda} x_1], \cdots, [\tilde{\lambda} x_k]) \\ &= \alpha^{-\sigma k} T(x_1, \cdots, x_k) \end{aligned}$$

by making a change of variables $\hat{\lambda} = \lambda\alpha$. (Notice that factors like log in the regularly varying function actually disappear in the limit – they're not reflected in the actual limiting object, only in how we take the limit.) If we now take another leap of faith and also assume rotation invariance in addition to scale invariance, that "usually" implies conformal invariance. Remember that in $\mathbb{R}^d$ for $d \neq 2$, this is equivalent to Möbius invariance, which is invariance under Euclidean isometries, scaling, and inversion $x \mapsto \frac{x}{||x||_2^2}$. The law that we get is therefore that for some $\psi \in \text{Mö}(\mathbb{R}^d)$, we get

$$T(\psi(x_1), \cdots, \psi(x_k)) = \left(\prod_{i=1}^n \text{local scaling factor}(x_i)\right)^{-\sigma} T(x_1, \cdots, x_k)$$

(the idea is that locally this looks like a rotation times some scaling factor around each point, so instead of an $\alpha^{-\sigma}$ for each one we get a different term). More precisely, this local scaling factor is given by the Jacobian via $|\det D\psi(x_i)|^{1/d}$ (if the function is just scaling this gives us back the scaling factor). These formal algebraic properties are typically maybe not how we want to think about these things – probabilistically we'd want to think about preserving a distribution. And in $d = 2$ we do often think about things that way, especially since we have many more conformal maps. We often think about things in terms of domain crossing (the chance of having a path from one marked boundary arc to another agrees with that of a disk after the Riemann mapping theorem), and such ideas should also be true for three-dimensional percolation whenever we can conformally map one domain to another (and also for higher-dimensions, though the probability becomes trivially 1).

We actually already have strong constraints on the two-point function in the setting that we're in: we can transform any two distinct points in $\mathbb{R}^d$ to any other two, so scaling and rotation-invariance give us everything:

$$T(x_1, x_2) = C_2(||X_1 - x_2||_2^{-2\sigma},$$

where we can evaluate the constant by just plugging in x_1, x_2 that are a distance 1 away. And if we also have Möbius invariance, we end up getting that we can map any **three** points to any three points; this yields something similar of the form

$$T(x_1, x_2, x_3) = C_3||x_1 - x_2||_2^{-\sigma}||x_2 - x_3||^{-\sigma}||x_3 - x_1||_2^{-\sigma}.$$

The form of the continuum function is only fixed for two- and three-point functions though, and that's where the theory of the conformal bootstrap starts (not being able to deduce the functional form but having constraints and extracting information from that). For the four-point function we get something called the "cross-ratio," and the dependence can be arbitrary as a function of that; things only get more complicated from there.

But what we do know is that in particular we have

$$T(x_1, x_2, x_3) \sim C\sqrt{T(x_1, x_2)T(x_2, x_3)T(x_3, x_1)};$$

in fact this constant is universal in two dimensions (and probably always); this was recently proven for site percolation of the triangular lattice by Ang, Cai, Sun, and Wu.

Remark 95. *Interestingly, physics papers never actually talk about the distinction between low and high dimensions – it's just that in high dimensions we have $C_3 = 0$ and we have a "trivial model" where only the two-point function is nonzero (this is called "Gaussian" because that same fact is also true for Gaussian fields).*

We can now talk about the k-point function hyperscaling relation via the following inequality:

Theorem 96 (Gladkov)
For percolation on any graph, we always have

$$\tau_\beta(x, y, z) \le \sqrt{8\tau_\beta(x, y)\tau_\beta(y, z)\tau_\beta(z, x)}.$$

We should compare this with the tree-graph inequality from above (also true for any graph and for any β) $\tau_\beta(x, y, z) \le \sum_w \tau_\beta(x, w)\tau_\beta(w, y)\tau_\beta(w, z)$; interestingly this theorem should be sharp for $d < d_c$ (as long as C_3 is nonzero) while the tree-graph inequality should be sharp above $d > d_c$, and both should be off at $d = d_c$ by some powers of log.

Fact 97
This relation between the two-point and three-point functions is easy to prove in the planar case: we do have on $\mathbb{Z}^2$ that

$$\tau_{\beta_c}(x, y, z) \asymp \sqrt{\tau(x, y)\tau(y, z)\tau(z, x)}$$

We should think of $\tau(x, y)$ as the square of the one-arm probability $A(||x - y||)$, so this equation can be restated as saying that $\tau_{\beta_c}(x, y, z) \asymp A(||x - y||)A(||y - z||A(||z - x||)$.

It's instructive to think about the interesting case where two points are close and another is not: we're saying that the two close points x, y need to get to an appropriate scale to connect, and then that and z both need to get to an appropriately large scale too – that's clearly a necessary condition. So the upper bound is pretty clear, and if we want to prove the lower bound we do the same kind of argument as before with the circuits around annuli.

Proof of Theorem 96. Say we have points x, y, z, and assume we're in a finite graph to avoid some set theory considerations (we can always take the limit). First explore the cluster K_x by a **depth-first** search (so we'll always explore from the furthest-away point), and let T be the time where the exploration reaches $\{y, z\}$. We'll always reveal a tree (because there's no reason to look at edges between already-connected edges), and because we did a depth first search, all edges off of that tree are closed (or else we would have already explored them) except potentially the edges along the path from x to y. Define $\mathcal{R}_y$ to be the event that x is connected to at least one point in $\{y, z\}$, but the depth-first search finds y first; similarly define $\mathcal{R}_z$.

Now consider taking two conditionally independent continuations of the depth-first search that explore the rest of the cluster after T. Let $\mathcal{A}$ be the event that both explorations find both points $\{y, z\}$ (the clusters we end up with will now be different other than the initial bit revealed). We then have (here $x \leftrightarrow y, z$ means that we're connected to both y and z rather than just the set $\{y, z\}$)

$$\begin{aligned} \mathbb{P}(\mathcal{A}, \mathcal{R}_y) &= \mathbb{E}\left[\mathbb{P}(x \leftrightarrow y, z | \mathcal{F}_T)^2 \cdot 1_{\mathcal{R}_y}\right] \\ \implies \mathbb{P}(\mathcal{A} | \mathcal{R}_y) &= \mathbb{E}\left[\mathbb{P}(x \leftrightarrow y, z | \mathcal{F}_T)^2 | \mathcal{R}_y\right] \\ &\geq \mathbb{P}(x \leftrightarrow y, z | \mathcal{R}_y)^2 \end{aligned}$$

by Jensen's inequality, and therefore writing back in terms of unconditioned probabilities we have

$$\mathbb{P}(\mathcal{A}, \mathcal{R}_y) \geq \frac{\mathbb{P}(x \leftrightarrow y, z, \mathcal{R}_y)^2}{\mathbb{P}(\mathcal{R}_y)},$$

and the same holds for $\mathcal{R}_z$. But any unqueried edges lie on the line along x to y, so there must be a path from this line to z in both of the two configurations. There's actually a disjoint occurrence at play here – those paths only intersect in the continuations, and so (up to justifying this claim with a lot of notation) the BK inequality gives us that

$$\mathbb{P}(\mathcal{A}, \mathcal{R}_y) \leq 2\mathbb{P}(x \leftrightarrow z)\mathbb{P}(y \leftrightarrow z)$$

and also the same for y and z swapped. From here, we just do a bit of rearrangement:

$$\begin{aligned} \mathbb{P}(x \leftrightarrow y, z, \mathcal{R}_y)^2 &\leq \mathbb{P}(\mathcal{R}_y)\mathbb{P}(\mathcal{A}, \mathcal{R}_y) \\ &\leq \mathbb{P}(x \leftrightarrow y) \cdot 2\mathbb{P}(x \leftrightarrow z)\mathbb{P}(y \leftrightarrow z), \end{aligned}$$

using that in order to have seen y it must have been in the cluster of x to begin with. The three-point function is thus

$$\mathbb{P}(x \leftrightarrow y, z)^2 = (\mathbb{P}(x \leftrightarrow y, z, \mathcal{R}_y) + \mathbb{P}(x \leftrightarrow y, z, \mathcal{R}_z))^2 \leq 2\left(\mathbb{P}(x \leftrightarrow y, z, \mathcal{R}_y)^2 + \mathbb{P}(x \leftrightarrow y, z, \mathcal{R}_z)^2\right),$$

and plugging in the bound above yields the result. □

This same method also lets us bound the higher k-point functions as well; the inequality that we get is rather curious and we'll discuss it more next time. (This is the part of the course which starts to get into unpublished work not yet on the arxiv.)

Theorem 98 (Higher-order Gladkov)
Let A be a set of points. The probability all points in A are in the same cluster is bounded as

$$\tau_\beta(A) \leq C_{|A|} \max_{\substack{A_1, A_2, A_3 \\ \text{partition of } A}} \sqrt{\tau(A_1 \cup A_2)\tau(A_2 \cup A_3)\tau(A_3 \cup A_1)};$$

in fact you can take A_3 to be a singleton.

The proof is the same as the theorem above – we run the exploration until we find all but one of the points and then run the same argument with independent continuations. And we'll see how this actually is very closely related to Möbius invariance!

12 June 20, 2025

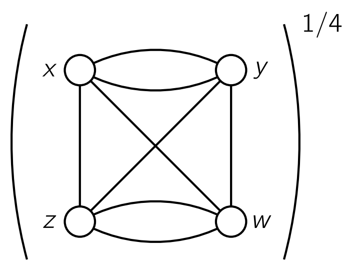
Last time, we were discussing the conjectures of Möbius invariance for critical percolation – indeed, we're saying that if a limit obtained via rescaling is well-defined, then we automatically get scale-invariance. We're then generalizing that to allowing a different local scale factor at each vertex x_i in the k-point function, and in fact we get power-law scaling for the two-point and three-point function assuming scale and Möbius inversion, respectively.

We then stated the inequality Theorem 98, and we'll start today by explaining what this has to do with Möbius invariance. This inequality is "quite weird" in full generality – when we had only three points, there was only one way of partitioning and the square root of the connection probability could be interpreted as certain "arm events" happening, but it's harder to interpret the general version.

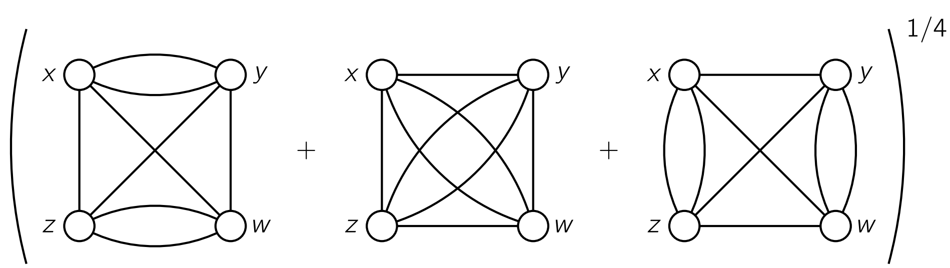
One thing we can do is look at the four-point function and try to apply the result recursively: if we partition $\{x, y, z, w\}$ into three parts, they must be of size 2, 1, 1 and we just need to choose which two points are put together. For example if $A_1 = \{x, y\}, A_2 = \{z\}$ and $A_3 = \{w\}$, the right-hand side (as a candidate for the maximum) would have a term which looks like

$$\sqrt{\tau(x,y,z)\tau(z,w)\tau(w,x,y)} \leq \sqrt{C\sqrt{\tau(x,y)\tau(x,z)\tau(y,z)}\tau(z,w)\sqrt{\tau(x,y)\tau(y,w)\tau(x,w)}}$$

where we've applied the three-point inequality. Now the weird thing is that some of the two-point functions are raised to the power $\frac{1}{4}$, while others are raised to the power $\frac{1}{2}$. Diagrammatically, we have the following picture for this particular term:



If we only care about things up to constants, we can thus just bound by the sum of the various diagrams, and we get a strange bound of the following form:

$\tau(x, y, z, w) \lesssim$  .

And as we increase the complexity we'll get a huge number of edges, making it harder to picture what is going on. So what we'll instead explain now is an alternative interpretation of this recursion which makes the connection to hyperscaling much clearer. Suppose we're now on $\mathbb{Z}^d$ and we have a bound on the two-point function $\tau_{\beta_c}(x, y) \lesssim ||x - y||_2^{-d+2-\eta}$; we saw that we get a bound on the three-point function from this.

Definition 99

Define the quantity $S(A)$ for vertex sets A by saying that for two points we have $S(x,y) = ||x-y||_2^2$, and otherwise recursively we have

$$S(A) = \min_{\substack{A_1, A_2, A_3 \\ \text{partition of } A}} \sqrt{S(A_1 \cup A_2) S(A_2 \cup A_3) S(A_3 \cup A_1)}$$

(we use min instead of max here because τ is like a negative power of the distance).

As a corollary of higher-order Gladkov, we find that if $\tau)\beta(x,y) \lesssim ||x-y||^{-d+2-\eta}$, then the connection probability has a bound

$$\tau_\beta(A) \le c_{|A|} s(A)^{\frac{-d+2-\eta}{2}}.$$

Of course, all we've done is introduce notation for the quantity in the recursion, but we have the following nice fact:

Lemma 100

The function S is Möbius-invariant.

This can be checked by induction – any power of the Euclidean distance is always Möbius-invariant with "dimension -1" (that's the quantity σ in our scaling law), and so in fact for any Möbius transformation ψ we have

$$S(\psi(A)) = \left(\prod_{x \in A} |\det D\psi(x)|^{1/d} \right)^1 S(A).$$

So we're bounding by a quantity which is Möbius-invariant, and we claim that while this may not be the right function, it is of the right order (so we can change our recursion rule in slightly different ways – for example, taking the harmonic mean instead of the minimum – and it's not so clear which is the most natural). What we claim is that for $d < d_c$ we should actually have

$$\tau_\beta(A) \asymp S(A)^{-\frac{d+2-\eta}{2}}.$$

Indeed, for this kind of thing one bound is always immediate (we showed the upper bound already), and our goal will be to show that this is correct in two-dimensions for all k-point functions. Instead of analyzing the recursion directly, we'll define another quantity with a more geometric interpretation:

Definition 101

An **arborescence** on a set A is an abstract directed tree on A with all edges pointing toward the root. For each point $x \in A$, let A_x be the set $\{x, \text{parent of } x, \text{descendants of } x\}$. With this notation, define the quantity

$$\text{Sweep}(A) = \min_{\text{arborescences on } A} \prod_{x \in A} \text{diam}(A_x),$$

where $\text{diam}(A_x)$ is now the $\mathbb{Z}^d$-diameter.

Lemma 102

We have $S(A) \asymp_{|A|} \text{Sweep}(A)$.

Indeed, if we have two points, there's only two ways to form an arborescence on $\{x, y\}$ and we always get $\text{Sweep}(x,y) = ||x-y||_2^2$. And to fill in the "fiddly details" needed in general, we require a bit of combinatorial geometry that shows we also satisfy the same recurrence up to constants as $S(A)$.

So we can express the bound for $\tau_\beta(A)$ in terms of Sweep instead of S, and let's now understand where the the matching lower bound comes from in two dimensions – that is, why we have

$$\tau_{p_c}(A) \overset{?}{\gtrsim} c_{|A|}\text{Sweep}(A)^{\eta/2}.$$

Indeed, what we're saying is that (now $\mathcal{A}$ will denote the one-arm probability instead)

$$\tau_{p_c}(A) \overset{?}{\asymp}_{|A|} \max \prod_x \mathcal{A}(\text{diam}(A_x)),$$

but this now makes sense in our previous geometric context: we're saying each point in A wants to reach some A which is appropriate to the "correct scale," and if this event occurs we can start drawing all of the circuits that connect our paths. That line of reasoning does lead us to the following result:

Theorem 103

If $\mathbb{Z}^2$, if $\tau_{p_c}(x, y) \asymp |x - y|^{-\eta}$, then we have $\tau_{p_c}(A) \asymp_{|A|} S(A)^{-\eta/2}$ (or equivalently $\tau_{p_c}(A) \asymp_{|A|} \text{Sweep}(A)^{-\eta/2}$).

So there's a weak sense in which we get Möbius invariance from scale invariance: this is weaker in that we only get equivalence up to constants with a Möbius-invariant function. But if we look at the critical two-point function, it will satisfy up to constants the transformation law

$$\tau_{p_c}(\psi(A)) \asymp_{|A|} \left(\prod_{x \in A} |\det D(\psi(x))|^{1/d} \right)^{-\eta/2} \tau_{p_c}(A)$$

(after doing some rounding and assuming that we don't end up mapping two points to the same point). It's not so clear how useful these kinds of relations are just yet, though – it's not the end of the story, but it's where we're at now.

Example 104

We're now going to enter the last part of the course in which we talk about long-range models. Everything so far (other than sharpness of the phase transition) has been proving conditional statements of the form 'if we can prove one property of percolation we can prove another one," and now we want to move to some more unconditional results. The analysis we see will be quite different and able to study things in quite a lot more generality.

We'll save the proofs for next week and just give a bit of a historical overview for now. We'll be primarily interested in the model on $\mathbb{Z}^d$ with power-law scaling of the form

$$J(x, y) = ||x - y||^{-d-\alpha}$$

(meaning any two points can be connected by an edge, but with polynomially decaying probability). Recall that this means the probability of having an open edge is then $1 - e^{-\beta J(x,y)}$, which is approximately $\beta J(x, y)$ when x, y are far apart.

In the case $\alpha > 0$, the model is integrable (that is, $|J| = \sum J(0, x)$ is finite – we'll always take α positive from now on. Our first question is to figure out the location of criticality:

Theorem 105

We have $0 < \beta_c < \infty$ if and only if $d \geq 2$ and $\alpha > 0$, or $d = 1$ and $0 < \alpha \leq 1$.

The fact that it's finite for $d \geq 2$ follows from the corresponding statement for nearest-neighbor percolation (since when x, y are adjacent $J(x, y)$ is a constant). But it's interesting that we also have a phase transition in one dimension and no longer have a trivial model. We won't prove this right now but will see the $\alpha < 1$ case later on.

Remark 106. *These models have been worked on for a long time in physics, primarily in the Ising model (though the statement for having a phase transition turns out to be equivalent).*

The first thought might be that we've made the model more complicated, but these long-range models are natural because they do have applications in various models (for example in disease). And remarkably, there are various ways in which it is actually easier to study.

Theorem 107 (Berger, 2002)
If $\alpha < d$, then there are no infinite clusters at β_c.

We'd very much like to prove this result for something like three-dimensional nearest-neighbor random walk; we should think of small α as being "more long-range," and the conjecture is that for $d \geq 2$ this fact should actually be true for all α. (The value of β_c is known if we parameterize the model in a different way and tune all of the parameters nicely, fixing the intensity of long edges and then tuning the short edge intensity. But in our setting the value is not exactly known.)

Theorem 108 (Aizenman–Newman)
The result above is not true – that is, there is an infinite cluster at $\alpha = d = \alpha$ (so we have a discontinuous phase transition).

With many of the techniques used with long-range percolation, it's hard to use knowledge of what the dimension is – it's usually down to something like the ratio between α to d. So going beyond Berger's theorem requires something new so that d itself actually shows up.

Turning to specific problems we want to solve in this new model, we have everything that we've already been considering for nearest-neighbor percolation (critical exponents, scaling limits, mean-field behavior). But there's an additional interesting question here of "when is the model really long-range" – that is, is there a large enough value of α that the decay makes things more localized. And we can study off-critical behavior as well, though that's not something we'll discuss in the remainder of the course.

One phenomenon with these long-range models is there is an "analogy between varying α and varying d." Indeed, one motivation (which hasn't been carried out for percolation yet) is that often in physical contexts we like to understand "$6 - \varepsilon$ dimensional percolation," and we can vary α continuously in a way that ends up being similar to continuously varying dimension. Indeed, we define the **effective dimension**

$$d_{\text{eff}} = \max\left(d, \frac{2d}{\alpha}\right).$$

We can think of this as being the **spectral dimension** of the random walk with jump kernel $P(x, y) \propto J(x, y)$ (normalized to be a probability measure), meaning that the return probabilities satisfy

$$P^{2n}(0, 0) \approx n^{-d_s/2}.$$

This is a useful notion but unfortunately imperfect – we'll see that $d_{\text{eff}} > 6$ is in fact equivalent to mean-field behavior (with one subtlety we'll come back to), but it's not so useful for thinking about low dimensions and using it to compute non-mean-field exponents.

Example 109

We can understand the question of "when the model is long range" in the more classical **random walk** setting.

If we have a walk with the jump kernel $P(x,y) \propto ||x-y||^{-d-\alpha}$, the idea is that α governs the **number of moments** that the kernel has. Indeed, after one step of the random walk started at the origin, up to constants we have (summing over spheres)

$$\mathbb{E}\left[||X_1||_p\right] \asymp \sum_r r^{d-1} r^p r^{-d-\alpha} = \sum_r r^{-1+p-\alpha},$$

which is infinite for $p \geq \alpha$ and finite otherwise. In particular, we get a central limit theorem and Donsker's invariance principle, which say that if the random walk has finite variance ($\alpha > 2$) then $(\lambda^{-1/2} X_{\lambda t})_{t \geq 0}$ converges to a Brownian motion $(B_t)_{t \geq 0}$, which is a **random continuous function**. On the other hand, for $\alpha < 2$ we have instead a different scaling $(\lambda^{-1/\alpha} X_{\lambda t})$, and we end up converging to a **discontinuous** process $(L_t)_{t \geq 0}$ which is the **α-stable Lévy process**. So in the case of random walk, we get a transition from "being really long range" to "being the same as nearest-neighbor" at $\alpha = 2$. And at the exact boundary point, we get one or the other with some logarithmic corrections; indeed for $\alpha = 2$ we get convergence to Brownian motion but with the correction $\left(\frac{1}{\sqrt{\lambda \log \lambda}} X_{\lambda t}\right)$. So we'll call the former regime **SR** (for "short range") and the latter regime **LR** (for "long range"), and the boundary case will be called **mSR** (for "marginally short range").

We can also look at the random walk Green's function; we know that

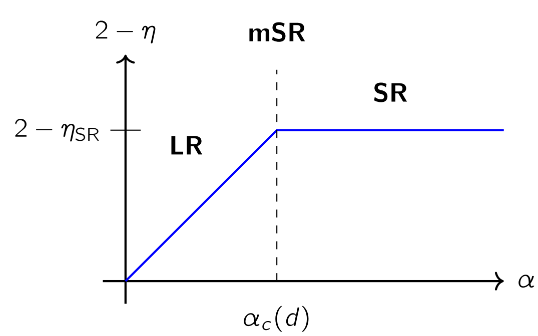
$$G(x,y) = \sum_{n \geq 0} P^n(x,y) \asymp \begin{cases} ||x-y||^{-d+\alpha} & \alpha < 2 \text{ and } d > \alpha \text{ (LR)}, \\ \frac{||x-y||^{-d+2}}{\log ||x-y||} & \alpha = 2 \text{ and } d \geq 2 \text{ (mSR)}, \\ ||x-y||^{-d+2} & \alpha > 2 \text{ and } d > 2 \text{ (SR)}. \end{cases}$$

And indeed it makes sense now why $\alpha = 2$ gives that the two terms d, $\frac{2d}{\alpha}$ are equal in the definition of d_{eff}.

However, percolation is a much more complicated object and what goes on there is a long story. Our first guess might be that we also have this transition between being effectively long or short range at $\alpha = 2$, but that turns out to be incorrect in low dimensions:

Conjecture 110 (Sak)

Fix $d > 1$, and plot $2 - \eta$ (the exponent for the two-point function) versus α. For large α the graph should be constant at $2 - \eta_{\text{SR}}$ (whose exact value is a difficult thing to understand in general), and for small α we should follow the line $y = x$ from the Lévy scaling. The conjecture is then that the switch from long-range to short-range just comes from connecting those two lines, meaning that the value of transitioning from LR to SR should be $\alpha_c(d) = 2 - \eta_{\text{SR}}$. (And the conjecture for $d = 1$ is that the graph just looks linear up until $\alpha = 1$.)

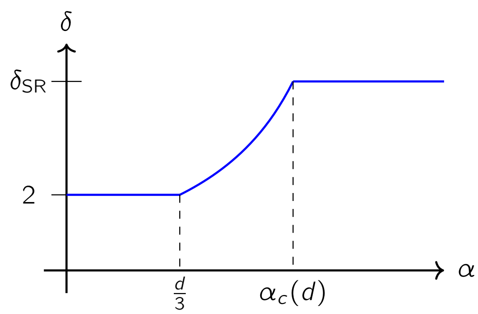


That is, in two dimensions we should have $\alpha_c(2) = 2 - \frac{5}{24}$, in high dimensions we should have $\alpha_c(d) = 2$, and in intermediate dimension it's quite hard to understand. Numerically (there's a Wikipedia page with tables of values that have been computed numerically or via non-rigorous methods, but some of them may not be so trustworthy) the current simulations say that

$$\alpha_c(3) \in [2.03, 2.08], \quad \alpha_c(4) \in [2.08, 2.13], \quad \alpha_c(5) \in [2.05, 2.08].$$

And we can do the same ε-expansion as we introduced at the start of the course, so that we would expect $\alpha_c(6-\varepsilon) = 2 + \frac{\varepsilon}{21} + \cdots$ and so on. And we already see that it's not just the effective dimension that matters when d is small.

Say now that $1 < d < 6$; we now have an additional important number $\frac{d}{3}$, which is where the effective dimension is at 6 and thus we should be critical in some sense. This doesn't have an important role in the picture above, but that's certainly not the case for other exponents. One way people interpret the prediction is that "the two-point function sticks to its mean-field above the critical regime." If we try plotting δ instead, what we should expect is that above the critical dimension we have $\delta = 2$. On the other hand, below the critical dimension we're determined by the hyperscaling relations $(2 - \eta) = d\frac{\delta-1}{\delta+1} \implies \delta = \frac{d+2-\eta}{d-2+\eta}$, and putting the predictions together gives us a curve of the following shape:



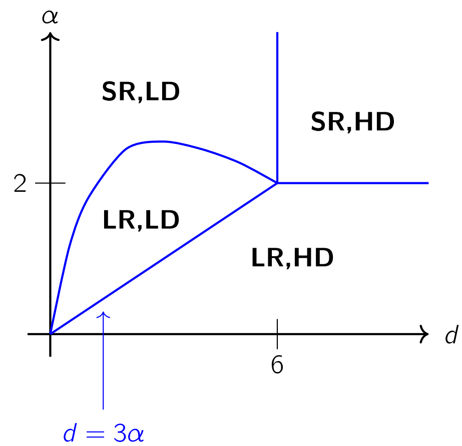
Here the middle part follow the curve $\delta = \frac{d+\alpha}{d-\alpha}$, and so now this other point does play an important role!

Fact 111

In branching random walks with heavy-tail jumps, the genealogical tree is what determines the volume tail and that's why we still have $\delta = 2$. But we now longer have Brownian scaling for the processes, which is why we wouldn't expect $\eta = 0$.

Remark 112. *One way we can check all of this is to think about when the triangle diagram $\sum G(0,x)G(x,y)G(y,0)$ is finite for the Green's function. It then ends up being finite if and only if* **either** *$d_{eff} = 6$,* **or** *if $d = 6$ and $\alpha = 2$, and so this is the slight caveat about mean-field behavior that we mentioned earlier today.*

In summary, we have the following picture for the phase diagram as we vary the dimension and α with four different regimes. **Importantly, this is a cartoon** (because d is actually only integer-valued); see the beginning of Lecture 14 for a more precise picture.



As we change α and d, there are thus two things that can change. We can be effectively high-dimensional versus effectively low-dimensional, and we can also see whether we're effectively long-range or effectively short-range. On all of the boundaries, we can have various behaviors of criticality as well (critical dimension, marginally short range, or both).

Next week, we'll essentially be able to analyze the model in quite a lot of the diagram, namely all of the high-dimensional and long-range regimes. (Of course, we should remember that d is actually an integer and that it doesn't really make sense to continue this down to 0.) But of course, the short-range low-dimension regime and its boundaries will remain difficult, much like in the nearest-neighbor case.

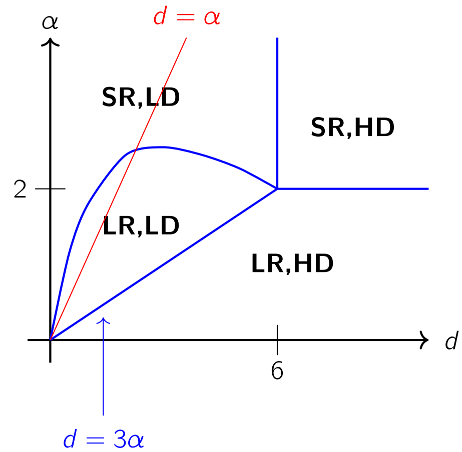
13 June 23, 2025

We started talking about long-range percolation on $\mathbb{Z}^d$ last time; recall that this means each edge $\{x, y\}$ is open with probability $1 - e^{-\beta J(x,y)}$ and we're interested in the case where $J(x, y) \asymp ||x - y||^{-d-\alpha}$ and J is symmetric and translation-invariant.

We drew a diagram of the two-point function exponent $2 - \eta$ as a function of α last time (recall $\mathbb{P}_{\beta_c}(x \leftrightarrow y) \approx ||x - y||^{-d+2-\eta}$); the conjecture is that we have a crossover value $\alpha_c(d)$ such that we have $2 - \eta = \alpha$ below α_c, until we hit the short-range value $2 - \eta_{\text{SR}}$. We also drew the volume tail exponent δ, which is a curve with three pieces (constant on the outside two, and curve in the middle depending on the η conjecture and hyperscaling relations). And as we vary d and α, we described four regimes (and their boundaries) for the behavior of percolation.

Remark 113. *We might guess that something particularly interesting happens where the points meet, but that point basically behaves like the "marginally short-range, high-dimensional" boundary to its right. This has to do with how the Green's function $G(x, y)$ for the random walk $P(x, y) \propto ||x - y||^{-d-\alpha}$ behaves as a function of α; in particular the Green's function triangle is indeed finite (heuristically this is mean-field behavior) at this middle point $d = 6, \alpha = 2$. (The extra log in the Green's function is enough to make this triangle converge.)*

Our main goal today is to state all of the theorems that can actually be proved, and we'll try to make it all precise. We've already mentioned Berger's result, Theorem 107 (saying that there are no infinite clusters at criticality for $\alpha < d$), but keep in mind that $\alpha = d$ doesn't correspond to any of these changes in behavior except in one dimension. Interestingly, it does tell us results for models that are conjectured to be in the same universality class (think something like $d = 3, \alpha = 2.5$, which is a short-range, low-dimension part of the diagram), though actually showing that we're in the same universality class is likely a much harder question.



Fact 114

The lace expansion has also been carried out for long-range models, yielding mean-field criticality for **spread-out models** – we'll need some additional numerical assumptions beyond d, α, and typically we can think about that as saying that the triangle for the random walk is not just finite but actually small (think $1+\varepsilon$). Here spread-out is not the same thing as having a large range anymore – it means that we have a function of the form

$$J(x,y) = \max\left(||x-y||, L\right)^{-d-\alpha}.$$

One secondary goal is to remove perturbative conditions needed in the lace expansion – in some sense this is related to being able to understand the critical dimension.

Remark 115. *Note that in long-range percolation, the exponents like η and δ tend to take nicer values in dimensions like $d=3$ than in the nearest-neighbor case. That gives us some hope that they might be easier to analyze, but it's not believed that some other critical exponents like γ will also have such nice forms.*

To prove that we don't have percolation at criticality, there are two main strategies (typically called the "sub-critical strategy" and "supercritical strategy.") The sub-critical strategy involves proving estimates that we take to the limit, such as the lace expansion, and it tends to give quite strnog quantitative information in addition to proving that we don't have infinite clusters. But the alternative is to suppose that we have infinite clusters at some β and show that we must actually have them still at a slightly smaller value, and that's what Berger does. That type of argument doesn't tend to give good quantitative results about the critical point, though (since we never really look at it in the argument).

Professor Hutchcroft's first involvment in this story was a few years ago:

Theorem 116 (Hutchcroft '20)

Berger's result can be improved to a quantitative bound

$$\mathbb{P}_{\beta_c}(|K| \geq n) \lesssim n^{-\frac{d-\alpha}{2d+\alpha}}$$

for all $\alpha < d$.

This is never optimal, but it is a power-law bound and tells us a lot more than just that we don't have infinite clusters. Many of the methods used have been made obsolete now, but the key idea is the universal tightness theorem we mentioned last week (and the related rigorous hyperscaling inequalities).

Remark 117. *To quickly explain the significance of this condition $\alpha < d$, suppose we take two boxes of side length n separated by a distance n. To count the number of edges in the configuration between these two boxes, we know that in expectation we multiply the possible number of edges $(n^d)^2$ by the probability $\approx n^{-d-\alpha}$ of each edge appearing; this quantity is of order $n^{d-\alpha}$, and so the basic geometric feature is that we get a difference in behavior at $\alpha = d$. As mentioned before, it's only supposed to correspond to an actual change in behavior at $d = 1$, but it is a quantitative change that we can see.*

One way to understand the shift from long-range to short-range is to put two boxes of side length n next to each other and see how the clusters merge. It's possible that they mostly only merge in a "bulk-to-bulk" way (only connected by some long edges), or that they merge mostly in a "boundary-to-boundary" way (via short edges). And as we change α, we change which of these strategies is better for growing large clusters; the exact value for when this occurs is the very complicated $\alpha_c(d)$ threshold, but for most values of α the point is that one or the other of these strategies will be much better than the other one. Bulk-to-bulk then corresponds to being effectively long-range, and boundary-to-boundary corresponds to being effectively short-range.

The idea in the latter case (short-range) is that we might have the same behavior as nearest-neighbor percolation. And when bulk-to-bulk is dominant, we should behave the same as a model which **only** has bulk-to-bulk interactions – that's the motivation for the hierarchical model, which we'll introduce soon. (And perhaps on the boundary we get some complicated thing where all scales are contributing, and that can lead to logarithmic corrections, but we won't go there.)

The relevance to what we're saying is that if $\alpha > d$, then there's no hope of bulk-to-bulk being relevant, since there's basically no edges of that kind. (The change is actually occurring much before that anyway, but it does tell us the inequality $\alpha_c(d) \le d$ even if that inequality is really bad.) Let's introduce what such a model would look like:

Definition 118

Define the **hierarchical lattice** $\mathbb{H}_L^d$ (with two parameters $d \ge 1, L \ge 2$) by starting with the integer lattice and grouping together vertices in a nested way at levels $L, L^2, L^3, \cdots$. More precisely, we define (this is not really a norm, but we'll write it as one)

$$\langle\langle x - y \rangle\rangle = \text{side length of the smallest block containing both } x \text{ and } y.$$

For example, to construct $\mathbb{H}_2^2$, we first group together the lattice into 2×2 chunks (in one of the 4 ways possible); then we group together those chunks into 4×4 chunks, and so on. And this can be interpreted by forming a tree whose leaves are the individual lattice points, their parents are the blocks containing those points, the parents of those the "next-level blocks," and so on; the value of $\langle\langle x - y \rangle\rangle$ is then exponential in how far up in the tree we need to go to connect two vertices. (We do have to be a bit careful with avoiding having points at infinite distance from each other, and there are various ways of making that precise.)

We can then do long-range percolation exactly as before but using the hierarchical distance

$$J(x, y) = \langle\langle x - y \rangle\rangle^{-d-\alpha}.$$

Note that this is now **not** translation-invariant, but it is invariant under the isometries of the hierarchical lattice $\mathbb{H}_L^d$. We can think about this as being "typically of the same order" as $||x - y||_\infty$, though two points on the boundary of large blocks will be close in Euclidean distance but very large in hierarchical distance. The isometries of this lattice look quite different from the Euclidean isometries; we can't do the usual translations, but we can transpose any two blocks with the same child and so on. In particular this makes the hierarchical lattice **distance-transitive**, meaning that for any points $(x, y), (z, w)$, there is some isometry γ such that

$$(\gamma x, \gamma y) = (z, w) \iff \langle\langle x - y \rangle\rangle = \langle\langle z - w \rangle\rangle.$$

In particular, the hierarchical model doesn't run into any issues like the Euclidean one wanting "rotation at non-90-degree angles" as long as we specify the correct isometries.

The following is the first result of determining a critical exponent outside the mean-field regime:

Theorem 119 (Hutchcroft '21)
For the hierarchical model $\mathbb{H}^d$, if the kernel J satisfies $J(x, y) \asymp \langle\langle x - y \rangle\rangle^{-d-\alpha}$ for $\alpha < d$, then the two-point function is given up to constants by

$$\mathbb{P}_{\beta_c}(x \leftrightarrow y) \asymp \langle\langle x - y \rangle\rangle^{-d+\alpha}.$$

(Note that for hierarchical models, β_c is only finite when $\alpha < d$ anyway.)

That is, we have no transition to an effectively short-range regime for η; we just have linear behavior. And that should make sense if we think about the "bulk-to-bulk versus boundary-to-boundary" discussion from above – there are no points that are closer than others to the next box, so there is no boundary in the first place. This basically means that we are actually getting mean-field behavior for the two-point function regardless of the value of α.

Corollary 120
In $\mathbb{H}^d$, the triangle condition holds if and only if $\alpha < \frac{d}{3}$. With a bit more work, this actually shows that the critical dimension is $d_c = 3\alpha$ for hierarchical models.

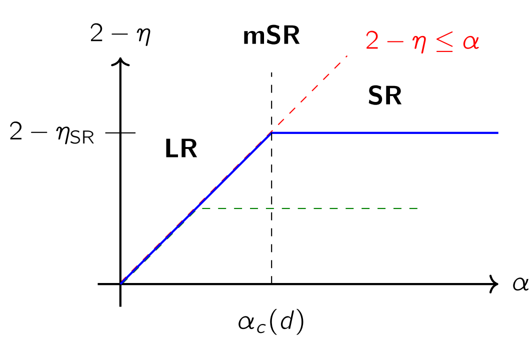
In fact, this also shows, using the rigorous hyperscaling inequalities, that $\delta \neq 2$ for $d < 3\alpha$. But we'll show a much stronger statement later anyway.

Theorem 121 (Hutchcroft '22)
For the integer lattice $\mathbb{Z}^d$, suppose the kernel is translation-invariant and satisfies $J(x, y) \gtrsim ||x - y||^{-d-\alpha}$. Then we get a corresponding bound

$$\frac{1}{r^d} \sum_{x \in [-r,r]^d} \mathbb{P}_{\beta_c}(0 \leftrightarrow x) \lesssim r^{-d+\alpha}.$$

This phenomenon of "monotonicity going in the opposite direction" as what we might expect is that the critical point changes as we vary J (in fact the larger J is, the smaller β_c tends to be), and that's the overwhelming effect that affects these things. Notice that the direction we're able to reason about for $\mathbb{Z}^d$ takes advantage of the fact that the hierarchical distance is bounded from below by the usual distance.

This result basically tells us that the red dashed line is indeed an upper bound for Sak's conjecture:



When Professor Hutchcroft posted this paper, he posed the question of posing conditions under which the converse is true.

Theorem 122 (Bäumler–Berger '22)
For $\mathbb{Z}^d$, assume we have the lower bound $J(x,y) \lesssim ||x-y||^{-d-\alpha}$. Then

$$\frac{1}{r^d} \sum_{x \in [-r,r]^d} \mathbb{P}_{\beta_c}(0 \leftrightarrow x) = \begin{cases} r^{-d+\alpha} & \alpha < 1, \\ \frac{r^{-d+1}}{\log r} & \alpha = 1, \\ r^{-d+1} & \alpha > 1. \end{cases}$$

This is basically specifying the green line above as a lower bound. And the point is that for $\alpha = 1$, all three curves agree and thus we do have Sak's prediction $\boxed{2 - \eta = \alpha}$ when $\alpha < 1$ and also always in one dimension. Meanwhile in $d = 2$, we have a regime $\alpha \in \left(\frac{2}{3}, 1\right)$ where Sak's prediction is true and we're not in mean-field behavior.

It's a bit annoying to not have pointwise versions of these estimates (only averages over boxes); it would be interesting to improve them. One thing that's been done is the following:

Theorem 123 (Hutchcroft '24)
On $\mathbb{Z}^d$, if we have the two-sided estimates $J(x,y) \asymp ||x-y||^{-d-\alpha}$ and furthermore $\alpha < 1$, then we in fact get pointwise upper and lower bounds of the same order

$$\mathbb{P}_{\beta_c}(x \leftrightarrow y) \asymp ||x-y||^{-d+\alpha}.$$

In particular (subject to $\alpha < 1$), the triangle ∇_{β_c} converges at criticality if and only if $\alpha < \frac{d}{3}$.

Interestingly, many of the proofs of these results actually use the structure and results of the hierarchical model, and we'll see that when we get more into the proofs.

Remark 124. *What we're seeing now is that we have the bounds $1 \le \alpha_c(d) \le d$; however when d is bigger than 1 they should both be strict inequalities. The reason why $\alpha = 1$ comes up in proofs even if it doesn't govern the actual behavior is that it also has a geometric interpretation. In the isoperimetric problem, we consider supercritical long-range percolation and ask whether most edges leaving a large box are long or short. It turns out that for $\alpha < 1$ most edges will be of the order of the sidelength of the box, but for $\alpha > 1$ most will be of smaller order.*

Everything we've stated so far is about the two-point function; one thing we don't see in these theorems is the distinction between the high and low-dimensional regimes. The proof methods also don't really see this distinction. We'll now turn to volume tails where this changes:

Theorem 125 (Hutchcroft '22)
For the hierarchical lattice $\mathbb{H}^d_L$, suppose we have the exact kernel $J(x,y) = \langle\langle x-y \rangle\rangle^{-d+\alpha}$. Then the volume tail satisfies

$$\mathbb{P}_{\beta_c}(|K| \ge n) \asymp \begin{cases} n^{-1/2} & d > 3\alpha \\ n^{-1/2}(\log n)^{1/4} & d = 3\alpha \\ n^{-(d-\alpha)/(d+\alpha)} & d < 3\alpha \end{cases}$$

Note that this is not the same as the conjectured $\frac{2}{7}$ log factor for nearest-neighbor percolation in $d = 6$; we're indeed expecting physically different corrections along different "lines" in our phase diagram. So what this is saying is that even though in $\mathbb{Z}^d$ the effective dimension is $\max(d, \frac{2d}{\alpha})$, in $\mathbb{H}^d$ we should just be defining the effective dimension to be $\frac{2d}{\alpha}$. In particular $d = 3\alpha$ still corresponds to "effective dimension 6."

The proof this result actually shows, in the hierarchical context, the validity of the various scaling and hyperscaling relations we've previously mentioned. And understanding the shift in the geometry between the clusters can be done much more explicitly: letting $N(r)$ be the number of large clusters and $\zeta(r)$ their size, we have the following table **for the hierarchical case**:

	$N(r)$	$\zeta(r)$
$d > 3\alpha$	$r^{d-3\alpha}$	$r^{2\alpha}$
$d = 3\alpha$	$\log r$	$\frac{r^{2\alpha}}{\sqrt{\log r}}$
$d < 3\alpha$	$O(1)$	$r^{(d+\alpha)/2}$

So to make a precise statement out of this, we're saying that we get the correct answer by using these predictions: as part of the proof of the result above, we can check that

$$\mathbb{E}_{\beta_c}\left[\left|K \cap [-r, r]^d\right|^p\right] \asymp_p \zeta(r)^p \frac{N(r)\zeta(r)}{r^d}$$

And we can think about these volume-tail things in a uniform way: if $\eta(r) = \frac{N(r)\zeta(r)}{r^d}$ is the probability of being part of a large cluster, we're saying that

$$\mathbb{P}_{\beta_c}(|K| \geq \zeta(r)) \asymp \eta(r)$$

in all regimes, meaning that the best way of having K be large is to actually be in one of these large clusters.

Of course, we're interested in trying to extend this to $\mathbb{Z}^d$ instead of the hierarchical model, and that's the ongoing new work of Professor Hutchcroft's. We'll be using a new approach which is arguably easier and doesn't require numerical assumptions like the lace expansion in high dimensions:

Theorem 126 (Hutchcroft '25+)

On $\mathbb{Z}^d$, take $J(x, y) = ||x - y||^{-d-\alpha}$ (though it doesn't have to be the Euclidean norm here). If $d > 3\alpha$ (which includes both the **LR,HD** regime and also part of the **SR,HD** one in our diagram), then we have the following:

- We have mean-field critical behavior for the volume tail in a very precise sense, namely that

$$\mathbb{P}_{\beta_c}(|K| \geq n) \sim \frac{\text{const}}{\sqrt{n}}.$$

- We can understand the scaling limit: there are functions η, ζ (see the table below for the scaling behavior) such that we have the weak convergence

$$\frac{1}{\eta(r)}\mathbb{P}_{\beta_c}\left(\frac{1}{\zeta(r)}\sum_{x\in K}\delta_{x/r} \in \cdot\right) \to \mathbf{N},$$

  where $\mathbf{N}$ is the canonical measure of the **integrated superprocess excursion**.

One concrete way of saying describing $\mathbf{N}$ is that it's the same limit as a critical branching random walk with the same (appropriately normalized) jump kernel. We may find it upsetting that we're taking a probability measure and divided it by something going to zero, but that turns out to be the best way of thinking about scaling limits in this context. Here $\mathbf{N}$ is a locally finite measure on measures in $\mathbb{R}^d$, and if we want to extract a more standard scaling limit statement about actual probability measures then we can condition on events and then pass that to $\mathbf{N}$.

The way we'll prove this theorem is quite concrete – we just compute asymptotics of moments, and we just need to know how to compute the moments of **N**. And since we get the same limit as branching random walk, we have **super-Brownian behavior** when $\alpha \geq 2$ and **super-Lévy behavior** when $\alpha < 2$. In aprticular, this does mean we get a difference between continuous versus jump paths as we cross over the line from short-range to long-range.

The table for how η and ζ should behave is shown below **for the $\mathbb{Z}^d$ case** – the last line is explained in the theorem at the end of this lecture:

	$N(r)$	$\zeta(r)$	$\eta(r)$
LR-HD	$r^{d-3\alpha}$	$r^{2\alpha}$	$r^{-\alpha}$
mSR-HD	$r^{d-6}(\log r)^6$	$\frac{r^4}{(\log r)^4}$	$\frac{\log r}{r^2}$
SR-HD	r^{d-6}	r^4	r^{-2}
LR-CD	$\log r$	$\frac{r^{2\alpha}}{\sqrt{\log r}}$	$r^{-\alpha}\sqrt{\log r}$

This isn't the entire high-dimensional regime in our diagram – the proof technique breaks down above the line $d = 3\alpha$, but we can extend further as well with some additional help:

Theorem 127

The same result of Theorem 126 also holds if the lace expansion converges.

So up in that upper region where we have a good bound on the two-point function, there's a way to do the argument starting from that point. (In particular, even for the spread-out models at the point $d = 6, \alpha = 2$, these results still apply.)

Finally, consider the line separating the high- and low-dimensional regimes for long-range:

Theorem 128 (Hutchcroft '25+)

For $\mathbb{Z}^d$ and $d = 3\alpha < 6$, we have (just like in the hierarchical case)

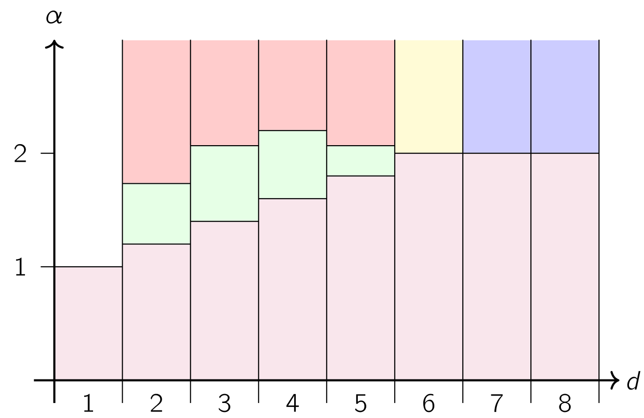
$$\mathbb{P}_{\beta_c}(|K| \geq n) \asymp \text{const}\frac{(\log n)^{1/4}}{\sqrt{n}},$$

and we have the same scaling limit theorem as before but with some extra log factors.

This last result is by far the hardest to prove of all of the new work, but in the remainder of this course we should at least be able to explain how we get the calculations to get the $(\log n)^{1/4}$ factor if we're okay with skipping some details about why errors go to zero.

14 June 24, 2025

First of all, here is a more "realistic" version of the phase diagram we've been working with so far. (In particular in $d = 1$, there is no phase transition, even though that's not reflected in the cartoon.)



Today's goal is to analyze the case $d > 3\alpha$ of the diagram. We mentioned the $\frac{1}{\sqrt{n}}$ scaling of the volume tail (implying varying other facts about critical exponents), as well as a result about superprocess scaling limits. We'll focus on the former result today, and the key tool will be a **form of the renormalization group method** (though it's rather different from other renormalization group techniques we may have seen before). There's a usual strategy of expressing our model as a perturbation of the Gaussian free field and then taking the covariance function and truncating or smoothly cutting it off, understanding how various quantities of interest vary as we send this cutoff to infinity. (So there are many ways we can try to implement this, and it's more a family of strategies than a particular technique.) For long range models the point is to use another startegy: instead of applying a cutoff to the Green's function (the inverse), we can apply the cutoff directly to J. This yields a somewhat "dual" approach to the classical renormalization, and whenever we can do this it's actually much easier.

Example 129

We'll work with a different form of the kernel J – up to some constant scalings we can rewrite instead

$$J(x, y) = \int_{||x-y||}^{\infty} s^{-d-\alpha-1} ds.$$

We can actually let $|| \cdot ||$ be any norm here, though we will assume to make calculations easier that the unit ball $B = \{x \in \mathbb{R}^d : ||X|| \le 1\}$ has Lebesgue measure 1 (so we don't need to worry about the various gamma functions and powers of π).

In particular, we can prove that the ball $B_r = \{x \in \mathbb{Z}^d : ||x|| \le r\}$ of radius r actually satisfies $|B_r| = r^d \pm O(r^{d-1})$, and for simplicity we will pretend it is r^d (this is really easily enough for anything we want to do). Our scheme will be to take, in addition to the untruncated kernel, the truncated versions

$$J_r(x, y) = 1\{||x - y|| \le r\} \cdot \int_{||x-y||}^{r} s^{-d-\alpha-1} ds.$$

(We could also use a smooth cutoff instead; different choices will have different advantages.) This approach will be central to all of the strongest results about the long-range model, not just in the high-dimensional regime, and we end up with a model with these two parameters $\mathbb{P}_{\beta,r}$.

We'll always take $\beta_c = \beta_c(J)$ to be the critical point **of the untruncated kernel**, and what we'll analyze is asymptotics of various functionals $\mathbb{E}_{\beta_c,r}[F(K)]$ as $r \to \infty$. In particular, we'll consider the expected size $F(K) = |K|$ or more generally the pth moments $F(K) = |K|^p$. And to understand full scaling limits, we'll need not just the volume but also things about how we embed the cluster into space. Thus we might also analyze things like $F(K) = \sum_{x \in K} ||x||_2^2$; if we can do this for arbitrary polynomials $\sum_{x_1,\cdots,x_k} P(x_1) \cdots P(x_k)$, then we actually determine the full scaling limit as a measure. So everything is actually quite concrete and we just need to compute moments. (It's actually not obvious that this is enough – in the end we want to prove things for the original model without cutoff – and that does require a nontrivial argument. But we'll come back to this later.)

Fact 130

The choice to take β_c to always be the critical point of the untruncated kernel actually implies that for each $r < \infty$, we have $J_r < (1-\varepsilon)J$ for some $\varepsilon > 0$. Thus actually $\mathbb{P}_{\beta_c,r}$ is stochastically dominated by some subcritical $\mathbb{P}_{(1-\varepsilon)\beta_c}$, so by the sharpness of the phase transition all of the various expectations we might care about are actually finite.

The reason for working with this $d > 3\alpha$ case is that we can compute explicitly from scratch. We'll always understand everything "by taking derivatives:" for $\beta \le \beta_c$ and $r < \infty$, we can compute, via Russo's formula,

$$\begin{aligned} \frac{\partial}{\partial r}\mathbb{E}_{\beta,r}[|K|] &= \sum_{x,y,z} \text{rate of } (x,y) \text{ becoming connected} \cdot 1\{(0,x) \text{ connected}, (y,z) \text{ connected}\} \\ &= \beta r^{-d-\alpha-1}\sum_{x,y,z} 1\{(x,y) \text{ not connected}\} \cdot 1\{(0,x) \text{ connected}, (y,z) \text{ connected}\} \end{aligned}$$

The picture to imagine is that we have a cluster at the origin, and as r increases we have some edges that open up which add another cluster. Furthermore we can only introduce new edges for points $||x-y|| \le r$ (by definition of the truncated kernel), and the rate is the same for all such edges by the fundamental theorem of calculus. Formally, we write this out as (exchanging the roles of 0 and x in one of the clusters)

$$\boxed{\frac{\partial}{\partial r}\mathbb{E}_{\beta,r}[|K|] = \beta r^{-d-\alpha-1}\mathbb{E}_{\beta,r}\left[\sum_{y\in B_r} 1\{0 \not\leftrightarrow y\}|K||K_y|\right]}.$$

(There are some details for applying Russo's formula in infinite volume, but it's fine because we're at subcriticality and so everything has exponential tails.) The idea now is that the expectation will factor asymptotically in high dimensions – $|K|$ and $|K_y|$ will end up being basically independent, and so our **goal is to prove** that this ends up approaching $\beta r^{-\alpha-1}\mathbb{E}[|K|]^2$ as $r \to \infty$. (The r^d points in the sum cancels out with the $-d$ in the leading exponent.)

Remark 131. *This quantity is only once differentiable, not twice differentiable, because of how we've defined things with integrals. But we'll never look at higher derivatives – if we did, we should use some smoother version of cutoff.*

In contrast, the lace expansion for nearest-neighbor percolation would do something similar but take a β-derivative instead and not have an r-cutoff at all. What we'd expect is that in high dimensions, we'd have $\frac{d}{d\beta}\mathbb{E}_\beta[|K|] \sim A\mathbb{E}_\beta[|K|]^2$ for some limiting constant A. But this A is really complicated – from that perspective we'd be adding in short edges and thus the clusters are now touching each other, and we now wouldn't expect independence. In the long-range model we tend to add in long edges as $r \to \infty$, so the two clusters will typically be well-separated and can thus actually have a chance of being asymptotically independent.

If we had such an asymptotic relation, we can then try to solve for f. (There's also a question of whether the asymptotic ODE has solution asymptotic to the actual solution – sometimes the answer is yes and sometimes no, and the difference will play an important role in our analysis. This one has some nice stability conditions and thus we don't run into issues.) We have

$$f' \sim \beta r^{-\alpha-1}f^2 \implies \left(\frac{1}{f}\right)' = -\frac{f'}{f^2} \sim -\beta r^{-\alpha-1}.$$

If we find integrating asymptotic expressions confusing, we can always just give a name to the error function and carrying that out along the way – thus we can write this last relation as $\left(\frac{1}{f}\right)' = (1-\delta_r)\beta r^{-\alpha-1}$. Integrating both sides then yields

$$\frac{1}{f(r)} - \frac{1}{f(\infty)} = \beta \int_r^\infty (1-\delta_s) s^{-\alpha-1} ds \sim \frac{\beta}{\alpha} r^{-\alpha}$$

(and we can get bounds on the final solution if we really wanted). So at $\beta = \beta_c$, if we know that $f(\infty) = \infty$ (as we will for the functions we care about, for example that the expected cluster size is infinite once we remove the cutoff),

$$f(r) \sim \frac{\alpha}{\beta_c} r^\alpha.$$

In this case, if the "asymptotic" function blows up to first order but not before infinity, then the true function will have the same behavior too. But that's not true for every ODE, and we'll see examples where it's not true later. Commenting a bit more on the stability part of this, if $|\delta_r| \le \varepsilon$ for r large and $f(\infty) = \infty$, then $f(r) = (1 + o(\varepsilon))\frac{\alpha}{\beta_c} r^\alpha$.

Example 132

Thus what we need to do is return to proving our claim that the expectation factors asymptotically. This is in fact the first place in the course where we're going to prove an unconditional result about exponents, and it turns out there's a nice easy argument here.

Take our exact boxed expression from above and rewrite it as

$$\begin{aligned}\frac{\partial}{\partial r}\mathbb{E}_{\beta,r}[|K|] &= \beta r^{-d-\alpha-1}\mathbb{E}_{\beta,r}\left[\sum_{y \in B_r} 1\{0 \not\leftrightarrow y\}|K||K_y|\right] \\ &= (1-\varepsilon_{\beta,r})\beta r^{-\alpha-1}\left(\mathbb{E}_{\beta,r}[|K|]\right)^2,\end{aligned}$$

where we define $\varepsilon_{\beta,r}$ to be whatever error function quantity is needed to make this equality true.

Lemma 133

The error function satisfies $0 \le \varepsilon_{\beta,r} \le \frac{\mathbb{E}_{\beta,r}[|K|^2|K \cap B_r|]}{|B_r|\mathbb{E}_\beta[|K|]^2}$. (We should think of this as saying that the numerator is the additive error.)

Proof. The lower bound is a BK inequality type argument – if we explore the first cluster, then the second one "has less space" and thus has expected size smaller than if they were independent. For the upper bound, all we're going to do is take our original quantity and rewrite

$$\mathbb{E}_{\beta,r}\left[\sum_{y \in B_r} 1\{0 \not\leftrightarrow y\}|K||K_y|\right] = \mathbb{E}_{\beta,r}\left[\sum_{y \in B_r} |K||K_y|\right] - \mathbb{E}_{\beta,r}\left[\sum_{y \in B_r} 1\{0 \leftrightarrow y\}|K||K_y|\right].$$

Now using Harris's inequality on the first term bounds it from below by $|B_r|\mathbb{E}[|K|]^2$, and the second term is actually exactly $\mathbb{E}[|K|^2|K \cap B_r|$ (since if y and 0 are connected, they're in the same cluster). Rearranging this yields the result. □

Since we're doing the case $d < 3\alpha$ it's actually okay to be less careful about the $|K \cap B_r|$ part and just bound it by $|K|$; we have by the tree-graph inequality that

$$0 \le \varepsilon_{\beta,r} \le \frac{\mathbb{E}_{\beta,r}[|K|^3]}{|B_r|\mathbb{E}_\beta[|K|]^2} \le \frac{3\mathbb{E}_{\beta,r}[|K|]^5}{|B_r|\mathbb{E}_\beta[|K|]^2} = \frac{3\mathbb{E}_{\beta,r}[|K|]^3}{|B_r|}.$$

Arguing "in a circular fashion" for a moment, we saw previously that if the error goes to zero at criticality, then the first moment $\mathbb{E}_{\beta,r}[|K|]$ is asymptotic to $\frac{\alpha}{\beta_c} r^\alpha$. So the error will actually tend to zero as $r \to \infty$, since the numerator is $O(r^{3\alpha})$ and the denominator is $O(r^d)$. But what this means is that if this goes to zero, it must go to zero at least at some power of r, and that tends to be a good sign. To turn this into a non-circular argument, we use a form of the bootstrap analysis ("choosing some optimal constant") in a slight variation of what we've seen before. We define

$$r_0(\beta) = \inf \left\{ r : \mathbb{E}_{\beta,r}[|K|] \le 2\frac{\alpha}{\beta_c} R^\alpha \text{ for all } R \ge r \right\}.$$

For $\beta < \beta_c$ this is finite (since the susceptibility actually doesn't go to infinity as $r \to \infty$), and now we want to argue that r_0 being finite actually implies that it's bounded by something not dependent on β at all. Indeed, using our bound we have, for any $\frac{\beta_c}{2} < \beta < \beta_c$,

$$r \ge r_0(\beta) \implies \varepsilon_{\beta,r} \le C r^{3\alpha - d} \implies \text{there is some } r_{00} < \infty \text{ independent of } \beta\text{, so that } r \ge r_0(\beta) \vee r_{00} \text{ implies } \varepsilon_{\beta,r} \le \frac{1}{100}.$$

So the case above the scale r_0 is not so interesting. But below r_0, we have

$$r \le r_0(\beta) \implies \varepsilon_{\beta,r} \le \frac{3\mathbb{E}_{\beta,r}[|K|]^3}{|B_r|} \le \frac{3\mathbb{E}_{\beta,r_0(\beta)}[|K|]^3}{|B_r|} \le C r_0^{3\alpha} r^{-d},$$

and notably this right-hand side is small even when r is smaller than r_0 by some power: there is some $\varepsilon > 0$ such that for all $r \ge r_0^{1-\varepsilon} \vee r_{00}$, we have $\varepsilon_{\beta,r} \le \frac{1}{100}$. We can thus rewrite all of this as saying that $\varepsilon_{\beta,r} \le \frac{1}{100}$ for all $r \ge C r_0^{1-\varepsilon}$ for some constant C, which implies that (plugging in what our error bound says, and noting that the $\frac{1}{f(\infty)}$ term actually only makes our inequality better)

$$\frac{\partial}{\partial r} \mathbb{E}_{\beta,r}[|K|] \ge 0.99 \beta r^{-\alpha-1} \mathbb{E}_{\beta,r}[|K|]^2 \text{ for all } r \ge C r_0^{1-\varepsilon}.$$

Integrating this ODE inequality, we thus find that we're quite close to the constant (the 1.5 can be more like 1.01, but the point is just that it's less than 2)

$$\mathbb{E}_{\beta,r}[|K|] \le 1.5 \frac{\alpha}{\beta_c} r^\alpha \text{ for all } r \ge C r_0^{1-\varepsilon}.$$

But by definition of r_0 we must have $r_0 \le C r_0^{1-\varepsilon}$, which rearranges to (knowing that r_0 is finite because $\beta < \beta_c$) $r_0 \le C^{1/\varepsilon}$, which doesn't depend on β. Then taking the limit $\beta \uparrow \beta_c$, we know that $r_0(\beta)$ is bounded and thus there is some $r_0 < \infty$ such that

$$\mathbb{E}_{\beta_c,r}[|K|] \le 2\frac{\alpha}{\beta_c} r^\alpha \text{ for all } r \ge r_0,$$

and this implies that the error $\varepsilon_{\beta_c,r}$ goes to zero at criticality, and so we get the exact first-order asymptotic claimed:

Theorem 134

If $d > 3\alpha$, then the first moment of the cluster size satisfies $\mathbb{E}_{\beta_c,r}[|K|] \sim \frac{\alpha}{\beta_c} r^\alpha$, and in fact the error is up to a factor of $(1 \pm O(r^{-\varepsilon}))$.

So what's nice is that in long-range models, approaching the critical point via this cutoff gives us simple explicit asymptotic formulas where all of the constants are known – there's no infinite inclusion-exclusion sums needed! And now when we look at more complicated things we will be able to use this to bound appropriate error terms instead of having to do more epsilon-delta analysis.

Fact 135

We won't have to work with $\beta < \beta_c$ from now on – everything will be at criticality and we'll write $\mathbb{E}_r = \mathbb{E}_{\beta_c, r}$.

To understand higher moments now, we'll first write an exact expression using Russo's formula: by very similar logic as before, we get

$$\frac{d}{dr}\mathbb{E}_r[|K|^p] = \beta_c r^{-d-\alpha-1}\mathbb{E}\left[\sum_{x \in K}\sum_{y \in B_r(x)} 1\{y \not\leftrightarrow x\}\left(|K| + K_y|)^p - |K|^p\right)\right].$$

Again switching the role of the origin and x, this simplifies to

$$\beta_c r^{-d-\alpha-1}\mathbb{E}\left[|K| \sum_{y \in B_r} 1\{y \not\leftrightarrow 0\}\left(|K| + K_y|)^p - |K|^p\right)\right],$$

and now expanding out the binomial term yields

$$\beta_c r^{-d-\alpha-1}\sum_{k=1}^{p}\binom{p}{k}\mathbb{E}\left[\sum_{y \in B_r} |K|^{p-k+1}|K_y|^k 1\{0 \not\leftrightarrow y\}\right].$$

We now want to do the same thing as before, saying that in the high-dimensional regime each term in the sum factors asymptotically. Each term in this sum can be written

$$\sum_{y \in B_r} |K|^{p-k+1}|K_y|^k 1\{0 \not\leftrightarrow y\} = |B_r|\mathbb{E}[|K|^{p-k+1}]\mathbb{E}[|K|^k] - O(\mathbb{E}[|K|]^{p+2})$$

again using BK and Harris like in the lemma above.

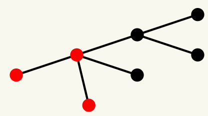
First we do the case $p = 2$. Simplifying the equation above (as before the r^d cancels with the size of the ball B_r, and both terms in the binomial coefficient give us a second moment on one cluster and a first moment on the other so they collapse to a single term), we get that

$$\frac{d}{dr}\mathbb{E}_r[|K|^2] = 3\beta_c r^{-\alpha-1}\mathbb{E}_r[|K|]\mathbb{E}_r[|K|^2] - O(r^{-d-\alpha-1}\mathbb{E}[|K|^4]).$$

We'd like to say that the last term is negligible, but if we try to use the tree-graph inequality it doesn't obviously work (since we don't know that the upper bound for the second moment term is actually of the right order). So first we need to prove a **generalization of the tree-graph inequality**.

Fact 136

Instead of bounding the tree graph in the usual way by pulling off first moments at each step, we can actually "pull off second moments" and just bound via the three-point function. For example, we get that for one particular possible configuration contributing to the fourth moment that



$$\leq \mathbb{E}[|K|^2] \left(\mathbb{E}[|K|]\right)^4 .$$

We can actually do the same for all of the different possible combinatorial structures, and thus we end up with the relation

$$\boxed{\mathbb{E}[|K|^4] \lesssim \mathbb{E}[|K|^2]\mathbb{E}[|K|]^4}.$$

There are extensions of this for arbitrary moments; the point is that when we choose which vertices to replace, we have to always pick an independent set so we don't "merge too much stuff."

Thus plugging this in shows that the error term is negligible, since we have three additional copies of the first moment but also an r^{-d} factor and thus we're smaller even by a power of r:

$$\frac{d}{dr}\mathbb{E}_r[|K|^2] = \left(1 - O(r^{-\varepsilon})\right) 3\beta_c r^{-\alpha-1}\mathbb{E}_r[|K|]\mathbb{E}_r[|K|^2].$$

(Here we do need to make use of the asymptotic estimate of the first moment in the theorem above.) We now find that

$$\frac{d}{dr}\mathbb{E}_r[|K|^2] \sim \frac{3\alpha}{r}\mathbb{E}_r[|K|^2],$$

and this time knowing the ODE to first order does not determine the solution to first order anymore. Indeed, if we have $f' = \frac{3\alpha+\delta_R}{r} f$, then $f = f(1)\exp\left(\int \frac{3\alpha}{r} dr\right)\exp\left(\int \frac{\delta_r}{r} dr\right)$. So the former integral gives us the $r^{3\alpha}$ factor, and if the latter integral converges, we're asymptotic to a constant. **In this case knowing that $\delta_r \to 0$ isn't enough** – it tells us that f is regularly varying of index 3α but not that it behaves exactly as we want. So it's important that our bound on the error is polynomial in r so that we do get convergence.

Note however that this error term $\exp\left(\int \frac{\delta_r}{r} dr\right)$ has effects mostly coming from small r. Thus we actually don't expect this to be universal – the constant will actually depend on the microscopic behavior of our system. But the point is that we have the following result:

Theorem 137

For $d > 3\alpha$, there is some constant $A < \infty$ such that $\mathbb{E}_r[|K|^2] \sim Ar^{3\alpha}$.

To give a preview of what happens in the critical dimension $d = 3\alpha$, the error term now might go to zero as a power of $\log(r)$ instead of a power of r. We can end up getting the same result as above, but instead of $Ar^{3\alpha}$ we get $A(r)r^{3\alpha}$ for some slowly-varying function $A(r)$. So most of the additional work in the critical-dimension case goes into showing that the asymptotic ODEs are true, and in fact the logarithmic corrections enter exactly through this function $A(r)$.

Notice that what we've found is that

$$\mathbb{E}_r[|K|^2] \asymp r^{3\alpha} \asymp \mathbb{E}_r[|K|]^3,$$

and actually **with no further work** this already gives us the bound from tree-graph (for finite r)

$$\mathbb{E}_r[|K|^p] \asymp_p \mathbb{E}_r[|K|]^{2p-1}.$$

Indeed, the upper bound is exactly what tree-graph tells us, and the lower bound is Hölder's inequality: for any random variable we have $\frac{\mathbb{E}[X^{p+1}]}{\mathbb{E}[X^p]}$ increasing as a function of p, so the moments must increase by a factor of $\mathbb{E}[|K|]^2$ each time. And thus we know that all errors are negligible for higher p as well in our binomial expansion; that is, the derivatives of the pth moments satisfy

$$\frac{d}{dr}\mathbb{E}_r[|K|^p] \sim \beta_c r^{-\alpha-1} \sum_{k=1}^{p} \binom{p}{k} \mathbb{E}_r[|K|^{p-k+1}]\mathbb{E}_r[|K|^k].$$

There's an interesting sense in which the asymptotic ODE that we get actually behaves very differently for $p = 1$, $p = 2$, and $p \geq 3$. We had a nonlinear ODE for $p = 1$ (with $f' \sim r^{-\alpha-1}f^2$), yielding stability. We then had a linear ODE for $p = 2$ (with $f' \sim \frac{1}{r}f$), which gives regular variation. But when we go to $p \geq 3$, we actually get many different terms in the binomial sum above. Two of the terms will actually involve the pth moment (when we take $k = 1$ or $k = p$), but all others will involve lower moments that we've already determined in our inductive analysis. The result is that

$$f' \sim \frac{(p+1)\alpha}{r} f + \beta_c r^{-\alpha-1} \sum_{k=2}^{p-1} \binom{p}{k} \mathbb{E}_r[|K|^{p-k+1}]\mathbb{E}_r[|K|^k].$$

Schematically, we can think of this situation using the following lemma:

Lemma 138

Suppose $f' \sim \frac{a}{r} f + \frac{h}{r}$, and h is regularly varying of index b. Then f is regularly varying of index $a \vee b$. Furthermore if $b > a$ (the driving term is dominating), then we actually have $f \sim \frac{h}{b-a}$.

Indeed, this additional condition is the situation we're in here (by looking at our asymptotics of moments). To sketch the proof of this, we introduce an integrating factor to solve the ODE: we have

$$f' = Pf + Q \implies \left(e^{-\int P} f\right)' = e^{-\int P} f' - Pe^{-\int P} f = e^{-\int Q},$$

so

$$e^{-\int_1^r P} f = \int_1^r e^{-\int_1^r P} Q + f(1),$$

and therefore

$$f = e^{\int_1^r P} \int_1^r e^{-\int_1^r P} Q + e^{\int_1^r P} f(1).$$

So if $P = \frac{1+\delta_r}{r} a$, we know that $e^{-\int P}$ is regularly varying of index $-a$, and thus the quantity inside the integral $\int_1^r e^{-\int_1^r P} \frac{h}{r}$ is regularly varying of index $b - 1 - a$, and we know that regularly varying functions can be integrated like powers to first order – thus we just get $\frac{1}{b-a} e^{-\int_1^r P} h$, and that's basically the claim. (And the other term $e^{\int_1^r P} f(1)$ is regularly varying with a smaller index, so it doesn't contribute overall.) We'll finish up the rest of the proof next time!

15 June 26, 2025

Last time, we were discussing a "renormalization group" type analysis in the $d > 3\alpha$ regime. We introduced the truncated kernel J_r (by writing J in integral form and then truncating the integral), and we proved that in this regime we get first-order asymptotics for the cluster size in the cut-off model

$$\mathbb{E}_{\beta_c,r}[|K|] \sim \frac{\alpha}{\beta_c} r^\alpha, \quad \mathbb{E}_{\beta_c,r}[|K|^2] \sim A r^{3\alpha}.$$

We then looked at the general pth moment, finding that the derivative satisfies a complicated first-order asymptotic expression (and in fact we get a different asymptotic ODE than for $p = 1$ or $p = 2$ because of stability considerations). In the higher-moment case, we now group our terms and get something like

$$f' \sim \frac{p+1}{r}\alpha f + \frac{h}{r}$$

(where the first term on the right-hand side comes from $k = 1, p$ in the binomial sum, and the second term is everything else). In this case, it depends on the relation between the coefficient and the function h, and what's important for using Lemma 138 is that our function h is regularly varying of index greater than $(p+1)\alpha$.

Remark 139. *We may notice that so far, there's been no distinction between the short-range $\alpha > 2$ and long-range $\alpha < 2$ case; that'll happen later. It's worth mentioning also that (exercise) we could have bounded the error in our ODEs in a slightly different way by assuming instead that $\nabla_{\beta_c} < \infty$; that will actually get us the bound on the first moment in the full short-range high-dimensional regime, and if we want to bound the second moment instead we need some rate at which the triangles converges as well.*

The way we will prove this is by induction: we want to show that

$$\mathbb{E}_r[|K|^p] \sim A_p \left(\frac{\mathbb{E}_r[|K^2]}{\mathbb{E}_r[|K|]}\right)^{p-1} \mathbb{E}_r[|K|]$$

for some constant A_p which we will figure out later. (We can think of this quantity in parentheses as the size-biased first moment.) Indeed, in particular it follows that the pth moment is regularly varying of order $(2p-1)\alpha$, which is greater than $(p+1)\alpha$ as long as $p \geq 3$. Indeed, by the inductive hypothesis, we have

$$\frac{d}{dr}\mathbb{E}_r[|K|^p] \sim \frac{(p+1)\alpha}{r}\mathbb{E}_r[|K|^p] + \beta_c r^{-\alpha-1}\sum_{k=2}^{p-1}\binom{p}{k}A_k A_{p-k+1}\left(\frac{\mathbb{E}_r[|K^2]}{\mathbb{E}_r[|K|]}\right)^{p-1}\mathbb{E}_r[|K|]^2$$

(here we've used the asymptotics of the first moment to simplify the terms in the binomial sum). But now take our asymptotic first moment on the latter term just once to cancel out the $r^{-\alpha}$, so

$$\frac{d}{dr}\mathbb{E}_r[|K|^p] \sim \frac{(p+1)\alpha}{r}\mathbb{E}_r[|K|^p] + \frac{\beta_c}{r}\sum_{k=2}^{p-1}\binom{p}{k}A_k A_{p-k+1}\left(\frac{\mathbb{E}_r[|K^2]}{\mathbb{E}_r[|K|]}\right)^{p-1}\mathbb{E}_r[|K|],$$

but now we can actually use Lemma 138 (because we know the rate of growth of the latter term and we do have $2p-1 > p+1$) to find that

$$\mathbb{E}_r[|K|^p] \sim \frac{\alpha}{\alpha(p-2)}\sum_{k=2}^{p-1}\binom{p}{k}A_k A_{p-k+1}\left(\frac{\mathbb{E}_r[|K^2]}{\mathbb{E}_r[|K|]}\right)^{p-1}\mathbb{E}_r[|K|].$$

So this proves the induction and further gives us a formula recursively for the A_ps, which is

$$A_p = \frac{1}{p-2}\sum_{k=2}^{p-1}\binom{p}{k}A_k A_{p-k+1}, \quad A_2 = 1.$$

It turns out the first few terms here are $1, 3, 15, 105, \cdots$, so $A_p = (2p-3)!!$. And this shouldn't be so surprising, since in high dimension the tree-graph inequality should be the right order and the coefficients are exactly these counts of leaf-labeled binary trees. (We can prove this algebraically using exponential generating functions, since these double factorials appear in negative binomials.) But the point is that we get the following:

Theorem 140

If $d > 3\alpha$, then

$$\mathbb{E}_r[|K|^p] \sim (2p-3)!! \left(\frac{\mathbb{E}_r[|K^2]}{\mathbb{E}_r[|K|]}\right)^{p-1} \mathbb{E}_r[|K|].$$

In particular we know that $\frac{\mathbb{E}_r[|K^2]}{\mathbb{E}_r[|K|]} \sim \tilde{A}r^{2\alpha}$ and $\mathbb{E}_r[|K|] \sim \frac{\alpha}{\beta_c}r^{\alpha}$, and thus we know a lot about the model now. The question is then what to extract from this – we want to prove something that doesn't require this truncation, and the first thing to notice is the following:

Corollary 141

We have, under the size-biased measure, convergence to a chi-square

$$\frac{|K|}{\hat{\mathbb{E}}_r[|K|]} \stackrel{r\to\infty}{\to} N^2,$$

where N is a standard normal.

Indeed, the pth moments of $\frac{|K|}{\hat{\mathbb{E}}_r[|K|]}$ are $\frac{\mathbb{E}_r[|K|^{p+1}]}{\mathbb{E}_r[|K|]\hat{\mathbb{E}}_r[|K|]^p}$, and the constants cancel out exactly to $(2p-1)!!$. So convergence of moments and moments not growing too quickly yields weak convergence; in particular we have a well-defined moment generating function so this will indeed work in this case. (Notice that this is the same limit we get when we take a slightly subcritical branching process and size-bias it, or do the same with Erdos-Renyi.) But this still isn't quite telling us what we care about, and for that we'll need the following:

Example 142

Let's return to the untruncated model, and suppose we want to connect to a ghost field ξ_h. We have of course that

$$\mathbb{P}_{\beta_c}(0 \leftrightarrow \xi_h) \geq \mathbb{P}_{\beta_c,r}(0 \leftrightarrow \xi_h),$$

and now we want to show that these are actually close on the right scaling of h with r.

To bound the left-hand side from below, we do the obvious bound under the standard monotone coupling: we have

$$\mathbb{P}_{\beta_c}(0 \leftrightarrow \xi_h) \leq \mathbb{P}_{\beta_c,r}(0 \leftrightarrow \xi_h) + \mathbb{P}(0 \leftrightarrow \xi \text{ but not in the } r\text{-truncated version}).$$

Now if the latter event happens, look at our model in the truncation where we know that there are no ghost vertices in the cluster – in order to become connected to the ghost, there must be some edge that opens up that lets us escape the truncated cluster, and then we have to connect from there to the ghost. So by the BK inequality, we actually have

$$\begin{aligned}\mathbb{P}(0 \leftrightarrow \xi \text{ but not in the } r\text{-truncated version}) &\leq \mathbb{E}_{\beta_c,r}[|K|e^{-h|K|}] \cdot \text{edges that might open up} \cdot \mathbb{P}_{\beta_c}(0 \leftrightarrow \xi_h) \\ &\sim \mathbb{E}_{\beta_c,r}[|K|e^{-h|K|}]r^{\alpha}\mathbb{P}_{\beta_c}(0 \leftrightarrow \xi_h)\end{aligned}$$

(just by a calculation of the kernel and summing over all vertices). Now $\mathbb{E}_{\beta_c,r}[|K|e^{-h|K|}]$ can be written as $\mathbb{E}_{\beta_c,r}[|K|]\hat{\mathbb{P}}_{\beta_c,r}(0 \leftrightarrow \xi_h)$, and what we know is that this term is small when h is much larger than $r^{-2\alpha}$ (coming from the limit rule – the only property of the chi-square limit law is that it doesn't have an atom at zero, and so we need the cluster to be of size on the order of $r^{2\alpha}$). So for all $\varepsilon > 0$, there is some finite λ such that

$$r \geq \lambda h^{-1/(2\alpha)} \implies \mathbb{P}_{\beta_c}(0 \leftrightarrow \xi_h) \leq \mathbb{P}_{\beta_c,r}(0 \leftrightarrow \xi) + \varepsilon\mathbb{P}_{\beta_c}(0 \leftrightarrow \xi_h)$$

(because the $EE_{\beta_c,r}[|K|]$ now cancels with the $r^{-\alpha}$ factor, and in fact the coefficients are actually reciprocals as well). So by choosing r appropriately scaled, we get near-equality of the ghost-field connection probability with a constant close to 1.

What this means is that we can get asymptotics as a double limit: taking $r = \lambda h^{-1/2}$ with λ as above, the truncated probability can be turned into a size-biased quantity

$$\mathbb{P}_{\beta_c,r}(0 \leftrightarrow \xi_h) = \mathbb{E}_{\beta_c,r}[|K|]\hat{\mathbb{E}}_r\left[\frac{1-e^{-h|K|}}{|K|}\right]$$

(since if we have an indicator, we need to divide by the size of the cluster to make it make sense under size-bias). Given this relationship between r and $h = \left(\frac{r}{\lambda}\right)^{-2\alpha}$, we can now plug in that value and get that

$$\begin{aligned}\mathbb{P}_{\beta_c,r}(0 \leftrightarrow \xi_h) &= \mathbb{E}_{\beta_c,r}[|K|]\hat{\mathbb{E}}_r\left[\frac{1-\exp\left(-\frac{\lambda^{2\alpha}|K|}{r^{2\alpha}}\right)}{|K|}\right] \\ &\sim \mathbb{E}_{\beta_c,r}[|K|]\frac{1}{\hat{\mathbb{E}}_r[|K|]}\hat{\mathbb{E}}_r\left[\frac{1-\exp\left(-\frac{A\lambda^{2\alpha}|K|}{\hat{\mathbb{E}}_r[|K|]}\right)}{|K|/\hat{\mathbb{E}}_r[|K|]}\right]\end{aligned}$$

where we've used $\hat{\mathbb{E}}_r[|K|] \sim Ar^{2\alpha} \implies r^{2\alpha} \sim \frac{\hat{\mathbb{E}}_r[|K|]}{A}$. The point is that we've now written everything in terms of the normalized size-biased cluster, so actually that whole last expectation is now $\mathbb{Q}\left[\frac{1-e^{-A\lambda^{2\alpha}X}}{X}\right]$, where $\mathbb{Q}$ is the chi-square. Therefore using our asymptotics and then writing everything back in terms of h, we have

$$\begin{aligned}\mathbb{P}_{\beta_c,r}(0 \leftrightarrow \xi_h) &\sim \frac{\alpha}{\beta_c}r^\alpha \cdot \frac{1}{Ar^{2\alpha}}\mathbb{Q}\left[\frac{1-e^{-A\lambda^{2\alpha}X}}{X}\right] \\ &\sim c \cdot \lambda^{-\alpha}\mathbb{Q}\left[\frac{1-e^{-A\lambda^{2\alpha}X}}{X}\right]\sqrt{h}\end{aligned}$$

for some constant c. Since the untruncated quantity is very close to the truncated one for λ large, we thus find that without truncation,

$$\mathbb{P}_{\beta_c}(0 \leftrightarrow \xi_h) \sim c\left(\lim_{\lambda\to\infty} \lambda^{-\alpha}\mathbb{Q}\left[\frac{1-e^{-A\lambda^{2\alpha}X}}{X}\right]\right)\sqrt{h},$$

and in fact this implies that the limit must exist. We can actually compute the exact constant of the law, but it's not so important – the upshot is that we end up with $\tilde{c}\sqrt{h}$. And remembering the connection to Karamata's Tauberian theorem and the Laplace transform, this now implies that

$$\mathbb{P}_{\beta_c}(|K| \geq n) \sim \frac{\text{const}}{\sqrt{n}}.$$

All of the constants here can indeed be computed in terms of the non-universal A. The gist is that we can actually go from truncated things (if we understand them well enough) to non-truncated things as well, which is what we cared about initially. And in all of this (except actually computing the constant), the only property of this limit law that we used was that we didn't have an atom at zero.

So we have proven, fully, mean-field behavior in some model in quite a strong sense:

Corollary 143

For $d > 3\alpha$, we have

$$\mathbb{P}_{\beta_c}(|K| \geq n) \sim \frac{\text{const}}{\sqrt{n}},$$

which implies mean-field values of many other exponents.

Example 144

The second part of the stated theorem is convergence to the superprocess limit, and given how much time we have left we won't go into too much detail. But we will explain the content of the theorem generally.

We'll start with the simplest moment that we haven't done yet: so far everything has been volumes of clusters, and if we want the full scaling limit we need moments involving spatial information. For a vector $u \in \mathbb{R}^d$, we of course have by symmetry that

$$\mathbb{E}_r\left[\sum_{x \in K} \langle x, u \rangle^p\right] = 0$$

for any p odd, so the first interesting quantity is something like (we can think of this one-dimensional projection as just being the x-coordinate if we'd like) $\mathbb{E}_r\left[\sum_{x \in K} \langle x, u \rangle^2\right]$. As before, we'll try to write an expression for the derivative and show that it simplifies in the high-dimensional case via some ODE. We have by Russo's formula that

$$\frac{d}{dr}\mathbb{E}_r\left[\sum_{x \in K} \langle x, u \rangle^2\right] = \beta_c r^{-d-\alpha-1}\mathbb{E}_r\left[\sum_{x,y,z,||x-y|| \leq r} 1\{(x,y) \text{ not connected}\} \cdot 1\{(0,x) \text{ connected}, (y,z) \text{ connected}\} \cdot \langle z, u \rangle^2\right].$$

Again switching the roles of 0 and x, we thus care about

$$\beta_c r^{-d-\alpha-1}\mathbb{E}_r\left[\sum_{x,y,z,||y|| \leq r} 1\{(0,y) \text{ not connected}\} \cdot 1\{(0,x) \text{ connected}, (y,z) \text{ connected}\} \cdot \langle z, u \rangle^2\right],$$

and we can write $z - x = z - y + y + (-x)$ so that

$$\langle z, u \rangle^2 = \langle z - y, u \rangle^2 + \langle y, u \rangle^2 + \langle -x, u \rangle^2 + \text{cross terms}.$$

We'd like to say that in high dimensions, the expectations approximately factor as before. Writing our expression from above slightly more formally, the terms $\langle y, u \rangle^2$ and $\langle -x, u \rangle^2$ the same amount.

$$\frac{d}{dr}\mathbb{E}_r \sum \langle x, u \rangle^2 = \beta_c r^{-d-\alpha-1}\mathbb{E}_r\left[\sum_{y \in B_r} 1\{0 \not\leftrightarrow y\}\left(2\sum_{x \in K}\langle x, u \rangle^2 |K_y| + |K||K_y|\langle y, u \rangle^2\right)\right] + \text{cross terms}.$$

The details are now very similar to what we did before: if $d > 3\alpha$, then what we have is asymptotic to all these various terms being independent, so

$$\frac{d}{dr}\mathbb{E}_r\left[\sum \langle x, u \rangle^2\right] \sim 2\beta_c r^{-\alpha-1}\mathbb{E}\left[\sum \langle x, u \rangle^2\right]\mathbb{E}[|K|] + \beta_c r^{-d-\alpha-1}\sum_{y \in B_r}\langle y, u \rangle^2 \mathbb{E}[|K|]^2 + \text{cross terms}.$$

If the two clusters actually were independent, then when we took the expectation of cross-terms we'd just get zero because of symmetry of odd moments. And what we can show is that those terms are indeed negligible – we can subtract the quantity we get if we were independent for every triple of points x, y, z, and we use the inequalities from earlier comparing the probability of various connections and non-connections to the independent case (for consequences of the triangle condition). Thus using our asymptotics,

$$\frac{d}{dr}\mathbb{E}_r\left[\sum \langle x, u\rangle^2\right] \sim \frac{2\alpha}{r}\mathbb{E}_r\left[\sum \langle x, u\rangle^2\right] + \frac{1}{r}\frac{\alpha^2}{\beta_c} r^{\alpha+2}\int_{B_1}\langle y, u\rangle^2 dy,$$

where B_1 is now the unit ball. And so we're back in this situation of Lemma 138 – the latter term is regularly varying of index $\alpha + 2$, which is bigger than the coefficient 2α if we're in the long-range regime $\alpha < 2$. (So now we see the importance of this transition in the phase diagram!) So we converge to a super-Lévy process in this regime, but in $\alpha > 2$ we instead get the Brownian limit. And we see now that in fact the combinatorial details of actually computing coefficients does make a big difference.

We can now define the **radius of gyration**

$$\xi_2(r) = \sqrt{\frac{\mathbb{E}\left[\sum ||X||_2^2\right]}{\mathbb{E}[|K|]}}$$

(that is, how far away a typical point is under the size-biased cluster) – this is $\sqrt{\hat{\mathbb{E}}_r[||X||_2^2]}$ if X is uniform from the size-bias. We find from our calculations (and analyzing the ODE) that for $d > 3\alpha$,

$$\xi_2(r) \sim \text{const} \cdot \begin{cases} r & \alpha < 2, \\ r\sqrt{\log r} & \alpha = 2 \\ r^{\alpha/2} & \alpha > 2. \end{cases}$$

(For $\alpha > 2$ we will need to use a similar thing as with the second moment where the errors get small by powers of r.) The critical case where the coefficient and index of variation are equal is actually straightforward via actually doing the computations of the integrating factor – we get a log factor in the integration, which becomes a square root of log here.)

What this radius of gyration is telling us is a notion of the correlation length – this model had some truncation which made it subcritical, and there should be some length above which we feel the subcriticality. And this is one reasonable notion for this – when $\alpha < 2$ a cutoff of scale r is felt at the same scale as where it was introduced, which is characteristic of being really long range since the long edges are important. But when $\alpha > 2$, truncating long edges isn't felt until a much larger scale, which corresponds to saying that the limit of the model doesn't see big jumps – we see something with continuous paths.

Turning now to larger moments, differentiating will again get us in a situation where some terms are negligible and we can use the lemma. Recall that for the volume we could determine the asymptotics from the ODE but with a non-universal constant for the second moment specifically; something similar happens here where

$$\frac{d}{dr}\mathbb{E}_r\left[\sum \langle x, u\rangle^{2p}\right] \sim A_{2p}(u)\left(\frac{\mathbb{E}\left[\sum \langle x, u\rangle^2\right]}{\mathbb{E}[|K|]}\right)^{p-1}\mathbb{E}_r[|K|]$$

with the coefficients satisfying some complicated recurrence relation. We'll skip doing this, but the point is that getting here is very similar to what we've already done before for the volume. But these coefficients (which can now depend on the direction) actually encode the scaling limit. And especially in the super-Lévy case, these coefficients are too complicated to give a closed-form expression and we have to instead "find a reason why the appropriate limit object also satisfies that same relation," so we do get the right thing.

We'll just say what actually happens: for $\alpha > 2$ (the Brownian limit), we can actually solve exactly that

$$A_{2p} = \frac{(2p)!}{2^p}.$$

Interpreting what this means, remembering that X is a uniform size-biased point, the absolute value of the rescaled point converges under rescaling to an Exp(1/2) distribution, which is actually the same as a centered Gaussian with random variance Exp(1). (And the limiting thing is indeed a centered multi-dimensional Gaussian with random variance.) And this is exactly the right thing for super-Brownian behavior – we don't have to understand what that super-process is if we can just keep computing the correct moments. To get the full scaling limit, we do this same thing but with polynomials of k points: we need to be able to compute

$$\mathbb{E}\left[\sum_{x_1,\cdots,x_k\in K} P_1(x_1)\cdots P_k(x_k)\right],$$

and it's exactly the same argument but just with recurrence relations on polynomials which are much messier to write down. So we have the recipe for how to continue the argument, and there's an additional step at the end where we go from understanding everything with the cutoff r to removing the cutoff altogether, which we've already seen a simple example of earlier.

Example 145

To finish this lecture, we'll see what happens in the critical dimension $d = 3\alpha < 6$ at a very general level.

In this case, we really are at long-range, and the main theorem in some sense is the following:

Theorem 146

All first-order asymptotic ODEs that were used in the arguments above are still valid.

We'll talk about low dimensions next time and see the "hydrodynamic condition" (sufficient for all errors to be negligible). And this hydrodynamic condition still holds at criticality, but the main difference is that when we look at the second moment, the sensitivity to the higher-order corrections now yields

$$\mathbb{E}_r[|K|^2] \sim A(r)r^{3\alpha}, \quad A(r) \text{ slowly varying.}$$

In fact, this second moment is the only place where this happens – at all higher moments it's replicated in a simple way, just like with our second-moment calculation before. And it happens with the spatial moments as well, so whatever is happening at criticality that's different from the high-dimensional case is in this one single $A(r)$ function. So we get the scaling limits exactly as before (having first-order asymptotic expressions for all moments), except including this function $A(r)$ in all of the appropriate scale factors. For example, we had that the pth moment was asymptotically $(2p-3)!!\left(\frac{\mathbb{E}_r[|K|^2]}{\mathbb{E}_r[|K|]}\right)^{p-1}\mathbb{E}_r[|K|]$ for high dimensions; this is still true but that ratio in parentheses can now involve log factors. And yet, this doesn't matter because what really matters is the combinatorial prefactor $(2p-3)!!$, so the argument works in that we "compute the scaling limit before knowing what we scale by."

These scaling limits always come from that ODE lemma we proved – we get the same one whether we're in critical dimension or above. And to compute what $A(r)$ actually is, we have something like $f' = \frac{3\alpha}{r}f$ to first order, and what we need is the second-order correction. We do this by writing

$$f' = \frac{3\alpha}{r}f - h$$

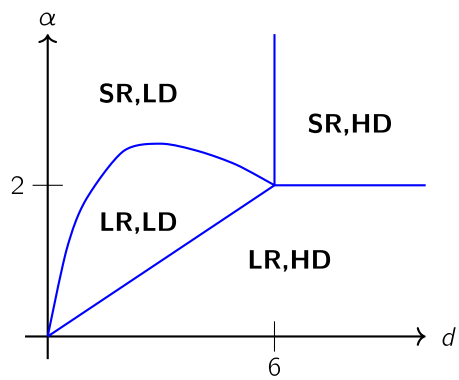
for some function h in terms of the model because we have an exact formula for the derivative, and then we do exactly the same thing but applied to h instead of f, finding an asymptotic ODE for the second-order term! That turns out to have the right form where the driving term has a larger index than the coefficient, so we can get a first-order asymptotic for the second-order term. Doing that yields

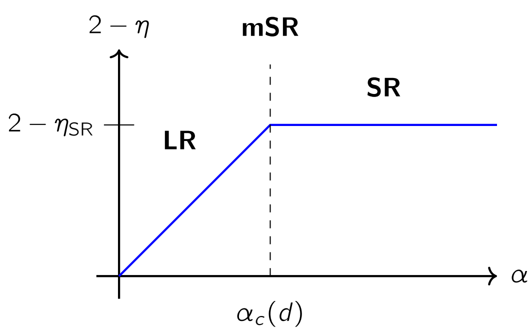
$$f' = \frac{3\alpha}{r}\left(1 - (1 \pm o(1))\, c\frac{f^2}{r^{6\alpha}}\right) f,$$

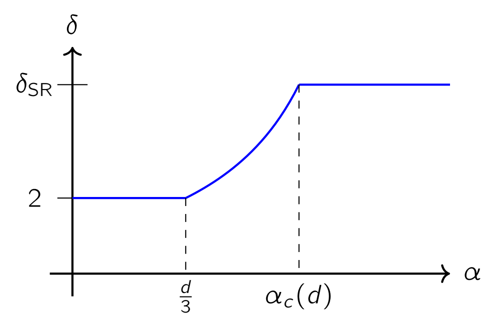
where the constant c is determined by the scaling limit (so we do have to understand the full process and understanding probabilities of returning to a ball after some time), and this error in multiplicative form has a self-referential thing. And from here it's just a calculus exercise to find that $f \asymp \tilde{c}\frac{r^{3\alpha}}{\sqrt{\log r}}$, so we can find $A(r)$ and use that to understand everything about the model and get all of the other log corrections through the Tauberian analysis, such as $\mathbb{P}_{\beta_c}(|K| \geq n) \sim \text{const}\frac{(\log n)^{1/4}}{\sqrt{n}}$.

16 June 27, 2025

The goal of this last lecture is to discuss low-dimensional long-range percolation, but we'll first talk about the two-point function (which doesn't actually distinguish between low and high dimensions in the long-range case). Here are all of the relevant plots that we've shown so far:







Note in particular that in dimensions $d = 3, 4, 5$, it's expected that η is actually negative (and then returns to zero for $d \geq 6$). Our first goal will be to prove estimates on the two-point function **in the hierarchical case**, Theorem 119. (It turns out that even in $\mathbb{Z}^d$, the hierarchical decomposition still ends up being useful by taking some appropriately large L.)

Recall that in the hierarchical model the "dimension" is somehow fake and the main quantity that matters is $\frac{\alpha}{d}$, since we have $\mathbb{H}_L^d \cong \mathbb{H}_{L^d}^1$ if we distort distances by an appropriate power. We'll fix some d, L, and let Λ_n be the n-block containing 0 (so in particular, it has side length L^n and volume L^{dn}). We define the "edian value of the largest cluster"

$$M_n(\beta) = \inf\left\{k \geq 1 : \mathbb{P}(|K_{\max}(\Lambda_n)| \geq k) \leq \frac{1}{e}\right\}$$

where $K_{\max}(\Lambda_n)$ is the largest possible cluster we can form using only the edges in Λ_n (so we're not allowed to go outside the box to extend the cluster). When we last looked at quantities of this sort, we were getting exponential decay in the universal tightness theorem.

Proposition 147

Assume $\alpha < d$ (otherwise the estimates we obtain will generally be trivial anyway). We have for all $\beta \le \beta_c$ that

$$M_n(\beta) \lesssim (L^n)^{\frac{d+\alpha}{2}}.$$

We'll use a "runaway observable" or "scaling catastrophe" type argument for this: the idea is that if this is ever exceeded at some scale, it will be exceeded more and more at larger scales, which fails by the sharpness of the phase transition.

Lemma 148

There is some constant $\ell_0 < \infty$ such that for all $\ell \ge \ell_0$, there is some finite $A = A(\ell)$ such that for all n, we have

$$M_n(\beta)^2 \ge \frac{A}{\beta} L^{(d+\alpha)n} \implies M_{n+\ell}(\beta)^2 \ge \frac{A}{\beta} L^{(d+\alpha)(n+\ell)}.$$

This indeed implies Proposition 147, since if we fix some $\ell \ge \ell_0$ and let A be the corresponding constant (we could just take $\ell = \ell_0$ itself), then it suffices to prove that $M_n(\beta)^2 < \frac{A}{\beta} L^{(d+\alpha)n}$ for all $n \ge 1$, and then we can take the limit $\beta \uparrow \beta_c$ for the equality case. (There's some small technicalities with continuity of the edian, but those don't cause any issues.) Suppose for contradiction that $M_n(\beta) \ge \frac{A}{\beta} L^{(d+\alpha)n}$ for some n; by the lemma we then have

$$M_{n+k\ell}(\beta)^2 \ge \frac{A}{\beta} L^{(d+\alpha)(n+k\ell)},$$

but we also proved earlier on that when $\beta < \beta_c$ we have

$$M_n = O(\log L^{dn}) = O(n).$$

And these two estimates are incompatible at some appropriately large scale, since our upper bound is linear but our lower bound is exponential in n. (So this is somewhat similar to the bootstrap ideas we've seen as well.)

Proof of Lemma 148. The block $\Lambda_{n+\ell}$ contains a total of $L^{\ell d}$ copies of the block Λ_n, and the largest clusters in each of the boxes are independent (or we could use Harris even if they weren't). So by the weak law of large numbers, there must exist ℓ' such that for all $\ell \ge \ell'$, at least $\frac{1}{4} L^{\ell d}$ of these copies have $|K_{\max}| \ge \frac{1}{2} M_n$ with probability at least 0.99 (just using that $\frac{1}{4} < \frac{1}{e}$ here because M_n is the edian).

Now fix some $\ell \ge \ell'$. If at scale n we have $M_n(\beta)^2 \ge \frac{A}{\beta} L^{(d+\alpha)n}$, and then we go to the next scale $n + \ell$, the probability that two of our clusters directly merge via a long-range edge is, up to constants, at least

$$1 \wedge M_n(\beta)^2 \mathbb{P}(\text{particular edge connecting}) = 1 \wedge M_n(\beta)^2 (L^{n+\ell})^{-d-\alpha},$$

which is close to 1 if $A = A(\ell)$ is large enough (here is where we make the choice of constant A). So we've reduced to a statement about a trivial regime of Erdös-Renyi – each pair of these clusters merges with quite high probability, and so all of the large clusters will merge together with probability at least 0.99.

This means that with probability at least 0.98, our block $\Lambda_{n+\ell}$ will have a cluster of size at least (merging all of them together)

$$\left(\frac{1}{4} L^{d\ell}\right) \left(\frac{1}{2} M_n(\beta)\right) \ge \frac{1}{8} L^{d\ell} \sqrt{\frac{A}{\beta} L^{(d+\alpha)n}},$$

and in particular this must mean the new edian satisfies

$$M_{n+\ell}(\beta)^2 \geq \frac{1}{64}\frac{A}{\beta}L^{(d+\alpha)n+2d\ell}.$$

Since we assume $\alpha < d$, rewriting this as

$$M_{n+\ell}(\beta)^2 \geq \frac{L^{(d-\alpha)\ell}}{64}\frac{A}{\beta}L^{(d+\alpha)(n+\ell)},$$

and choosing ℓ_0 so that constant in front is always at least 1 yields the result. $\square$

This bound turns out to be the correct order in the low-dimensional regime, but not in the high-dimensional one.

We'll now work at β_c now that we have one estimate in place. Recall that in the universal tightness theorem, we had the other estimate (letting K_n be the cluster of 0 in Λ_n)

$$\mathbb{P}_{\beta_c}(|K_n| \geq \lambda M_n) \leq \mathbb{P}_{\beta_c}(|K_n| \geq M_n)e^{1-\lambda/9}.$$

As an exercise, we can show that this implies that for all $\varepsilon > 0$, we can find some λ such that

$$\mathbb{E}_{\beta_c}\left[|K_n| 1\{|K_n| \leq \lambda M_n\}\right] \geq (1-\varepsilon)\mathbb{E}_{\beta_c}[|K_n|]$$

(so throwing away large clusters doesn't affect the mean since we're well into the tail).

Proposition 149

We have $\mathbb{E}_{\beta_c}[|K_n|] \lesssim (L^n)^{\alpha}$.

This is very close to what we want to prove with the two-point function estimate – this is connections just inside the box, while two-point function yields connections globally. Again, we'll prove this with some kind of inductive argument – this next result should look somewhat similar to the ODE bounds we've been previously deriving:

Lemma 150

There exists $c > 0$ such that going up one scale yields the bound

$$\mathbb{E}_{\beta_c}\left[|K_{n+1}|\right] \geq c\frac{\mathbb{E}_{\beta_c}[|K_n|]^2}{L^{\alpha n}}.$$

This implies Proposition 149 because (exercise) there's some constant C where if $a_n \geq CL^{\alpha n}$, then in fact we have $a_m \geq e^{e^{\tau m}}$ as $m \to \infty$ (since "once the squaring starts winning we're basically just squaring every step), which cannot happen because the most points we can have in a box is L^{dn}. And having an inequality $a_{n+1} \geq c\frac{a_n^2}{L^{\alpha n}}$ indeed implies that such a C exists.

Proof of Lemma 150. For this proof we just need to take two "sibling" blocks Λ_n, Λ_n' (that is, both contained in the same parent block). Let $\mathcal{C}'$ be the set of clusters in Λ_n'; define the σ-algebra

$$\mathcal{F} = \sigma(\text{edges entirely within } \Lambda_n \text{ or } \Lambda_n').$$

We'll look at the cluster at the next scale conditional on $\mathcal{F}$: we have that we can expand our current cluster by connecting directly to one of them, so

$$\mathbb{E}\left[|K_{n+1}|\middle|\mathcal{F}\right] \geq \sum_{C \in \mathcal{C}'} |C|\left(1 - \exp\left(-c(L^n)^{-d-\alpha}|K_n||C|\right)\right).$$

We can (using our previous proposition to bound the edian) pick some constant λ such that $\mathbb{E}[|K_n| 1\{|K_n| \leq \lambda(L^n)^{(d+\alpha)/2}\}] \geq \frac{1}{2}\mathbb{E}[|K_n|]$, and then restrict to a smaller set of clusters

$$\mathcal{C}'' = \left\{ C \in \mathcal{C}' : |C| \leq \lambda(L^n)^{(d+\alpha)/2} \right\}.$$

We can restrict to a smaller set of clusters in the bound above, so we instead have

$$\mathbb{E}\left[|K_{n+1}| \middle| \mathcal{F} \right] \geq \sum_{C \in \mathcal{C}''} |C| \left(1 - \exp\left(-c(L^n)^{-d-\alpha}|K_n||C|\right)\right).$$

Now if $|K_n| \leq \lambda(L^n)^{(d+\alpha)/2}$, then the term inside the exponential (in blue) is small and thus we can linearize using $1 - e^{-x} \geq cx$ for sufficiently small x:

$$\mathbb{E}\left[|K_{n+1}| \middle| \mathcal{F} \right] \geq c(L^n)^{-d-\alpha} \sum_{C \in \mathcal{C}''} |C|^2 |K_n|,$$

and taking expectations over both sides yields (here because $|C|$ and $|K_n|$ are actually independent)

$$\mathbb{E}[|K_{n+1}|] \geq c\mathbb{E}\left[|K_n| 1\{|K_n| \leq \lambda(L^n)^{(d+\alpha)/2}\} \right] (L^n)^{-d-\alpha} \mathbb{E}\left[\sum_{C \in \mathcal{C}''} |C|^2 \right].$$

Indeed, it's a deterministic fact that the sum of the squared cluster sizes is just summing the cluster sizes over individual vertices:

$$\mathbb{E}\left[\sum_{C \in \mathcal{C}''} |C|^2 \right] = \mathbb{E}\left[\sum_{x \in \Lambda_n'} |K_x| 1\{|K_x| \leq \lambda(L^n)^{(d+\alpha)/2}\} \right],$$

and by transitivity this is just L^{dn} (the number of vertices) times the same expectation as above! So in fact we have by our choice of λ that

$$\mathbb{E}[|K_{n+1}|] \gtrsim L^{-\alpha n} \mathbb{E}[|K_n|]^2,$$

which is what we set out to prove. □

Everything we've said actually also works for $\mathbb{Z}^d$ instead of the hierarchical lattice – it's no longer true that we have transitivity, so we just have to average over the box instead and get instead that

$$\frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} \mathbb{E}\left[|\text{cluster in } \Lambda_n \text{ of } x|\right] \lesssim (L^n)^\alpha.$$

Indeed, the argument is "if clusters are too big, then they grow too quickly afterward," and the Euclidean kernel is just larger than the hierarchical one – we have a sufficient condition for rapid growth.

But now **this next step completely breaks in the Euclidean case**: the hierarchical lattice is distance-transitive in that all points in $\Lambda_n \setminus \Lambda_{n-1}$ have the same connection probability to the origin by symmetry, and so we actually get pointwise estimates of the form

$$\mathbb{P}_{\beta_c}\left(0 \overset{\Lambda_n}{\leftrightarrow} x\right) \lesssim (L^n)^{-d+\alpha} \text{ for all } x \in \Lambda_n \setminus \Lambda_{n-1}.$$

Now we want the same thing but without requiring our paths to lie within the box Λ_n. We do this by decomposing according to the first scale $n+k$ such that the origin is connected to x within Λ_{n+k}; we now use the useful "ultrametric space" fact that we can only leave a large box by taking a long edge. This yields

$$\mathbb{P}_{\beta_c}(0 \leftrightarrow x) = \mathbb{P}_{\beta_c}\left(0 \overset{\Lambda_n}{\leftrightarrow} x\right) + \sum_{k=1}^{\infty} \mathbb{P}_{\beta_c}\left(0 \overset{\Lambda_{n+k}}{\leftrightarrow} x \text{ but } 0 \overset{\Lambda_{n+k-1}}{\not\leftrightarrow} x\right).$$

We have a bound on the first term already, and for the second term (for some fixed k), we can decompose according to the first edge used to enter the top scale $n+k$ (after which we must descend back down into x). In the hierarchical case, we know that edge must have a specific weight $(L^{(n+k)})^{-d-\alpha}$, so then by the BK inequality (since we have disjoint events of the individual parts of the path) we thus get

$$\mathbb{P}_{\beta_c}\left(0 \overset{\Lambda_{n+k}}{\leftrightarrow} x \text{ but } 0 \overset{\Lambda_{n+k-1}}{\not\leftrightarrow} x\right) \lesssim \mathbb{E}[|K_{n+k}|](L^{(n+k)})^{-\alpha}(L^{n+k})^{-d+\alpha}.$$

Since that first expectation (which is encoding the number of possible vertices we can enter at the largest scale) is bounded by $L^{(n+k)\alpha}$ by our previous proposition, we can in fact bound the sum over k by a geometric series, which of the same order as the first term $(L^n)^{-d+\alpha}$. Thus plugging this back in yields the appropriate bound

$$\boxed{\mathbb{P}_{\beta_c}(0 \leftrightarrow x) \lesssim (L^n)^{-d+\alpha}}.$$

Remark 151. *There is actually a way to fix this to get $\mathbb{Z}^d$, but we have to do everything from the beginning in a slightly different way so that we're only sprinkling in long edges at the very end. That's something we'll leave out for another time, though.*

To get the lower bound, recall that in the proof of sharpness of the phase transition we had the quantity $\phi_{\beta_c}(\Lambda_n) \geq 1$. In the hierarchical case this actually directly implies that $\mathbb{E}_{\beta_c}[|K_n|] \gtrsim (L^n)^\alpha$, since every point inside the box has the same relation to everything outside the box. (But of course that isn't going to work in the Euclidean case anymore – Bäumler and Berger showed that in $\mathbb{Z}^d$, we have to apply this inequality to a box of uniform random radius between n and $2n$, and then when $\alpha < 1$ it turns out the argument still works.)

Example 152

As we've said already, the behavior of the two-point function in the long-range regime does not reflect any difference between low and high dimensions. So we'll now say something interesting about the "long-range, low-dimensional regime" – it's going to be tricky because the curve $\alpha = \alpha_c(d) = 2 - \eta_{\text{SR}}$ isn't exactly known at all.

What we'll instead do is define alternatively what it means to be in this regime of interest. In the high-dimensional regime we had a truncated kernel J_r and defined the radius of gyration $\xi_2(r)$, which we saw was of order r for $\alpha < 2$, $r\sqrt{\log r}$ for $\alpha = 2$, and $r^{\alpha/2}$ for $\alpha > 2$ in the high-dimensional regime $d > 3\alpha$. This turns out to also occur for $d = 3\alpha < 6$, and so the long-range regime should actually be characterized by **having linear scaling of the correlation length as a function of the cutoff** (that is, the model feels the cutoff at the same scale as where it was applied). Let's use that as a definition.

When working with the correlation length, the exact definition is rather application-dependent, and here we'll use something a bit ugly which works well (though it would be nice in future work to show that it's equivalent to something nicer):

Definition 153

We'll now say that our model is **effectively long-range** if for all ε, there exists λ such that

$$\mathbb{E}_{\beta_c,r}\left[|K|^p\right] \leq (1+\varepsilon)\mathbb{E}_{\beta_c,r}\left[|K \cap B_{\lambda r}|^p\right]$$

for the moments $p = 1, 2, 3, 4$.

Theorem 154

The condition above holds when $\alpha < 1$. (This is a sufficient condition in low dimensions, covering some non-mean-field regimes, but it should not be sharp).

Theorem 155

If we are in the low effective-dimensional case $d < 3\alpha$, and we are effectively long-range, then the largest cluster in a box of radius r satisfies $M_r \asymp r^{(d+\alpha)/2}$, and the cluster scales as $\mathbb{E}[|K \cap B_r|] \asymp r^\alpha$ (this is morally related to the two-point function scaling $\mathbb{P}_{\beta_c}(x \leftrightarrow y) \asymp ||x-y||^{-d+\alpha}$, but it's just a spatial average). Finally, the tail of the cluster satisfies $\mathbb{P}_{\beta_c}(|K| \geq n) \asymp n^{-(d-\alpha)/(d+\alpha)}$.

In words, this is a verification of Sak's prediction under our particular definition.

Theorem 156

If we instead have the exact two-point function scaling $\mathbb{P}_{\beta_c}(x \leftrightarrow y) \asymp ||x-y||^{-d+\alpha}$, then we have for any set A that $\tau_{\beta_c}(A) \asymp S(A)^{-(d-\alpha)/2}$, with S defined the same way as it was for nearest-neighbor percolation.

So all of the hyperscaling story is actually valid for the model in this regime, letting us get between exponents, and the k-point function is also determined in the appropriate way.

To understand a little bit of what goes into these results, suppose we have as input that the average of the two-point function in a box is bounded above as

$$\frac{1}{r^d} \sum_{x \in [-r,r]^d} \mathbb{P}_{\beta_c}(0 \leftrightarrow x) \lesssim r^{-d+\alpha}.$$

Relatedly we always have (we now let M_r mean the edian of largest cluster in infinite volume, but intersected with the box, under $\mathbb{E}_{\beta_c,r}$) $M_r \lesssim r^{(d+\alpha)/2}$; universal tightness actually lets us deduce this latter bound from the former one. Both of these are always true as upper bounds, and the significance of the power $\frac{d+\alpha}{2}$ is that if we have two boxes next to each other and we ask for the probability of two boxes to merge via a bulk edge, each such edge is present with probability $r^{-d-\alpha}$ and so we get a chance of roughly $r^{-d-\alpha}|A||B|$. So it can never be that these clusters merge with high probability – if the largest cluster M_r is of this maximum order then we have good probability, but never high probability. We can thus define the **hydrodynamic condition** to hold if (we're at criticality here)

$$M_r = o(r^{(d+\alpha)/2}) \text{ as } r \to \infty.$$

If this is the case, then in fact the largest clusters don't really see each other directly – we're instead in a mean-field situation with a soup of large-ish clusters.

Theorem 157

Under the effectively long-range assumption, the hydrodynamic condition holds if and only if $d \geq 3\alpha$.

The case $d > 3\alpha$ is actually quite easy to prove using the tree-graph inequality, since we have $\mathbb{E}_{\beta_c,r}[|K|] \asymp r^\alpha$ and it's very hard for the cluster at the origin to be larger than the square of its mean; we get an exponential tail and using the union bound yields $M_r \lesssim r^{2\alpha} \log r$, and when $d > 3\alpha$ this is actually a better bound than the hydrodynamic condition. At the critical dimension it's actually very difficult – it's like a 50 page proof by contradiction in the paper.

What we'll do is show why this hydrodynamic condition does not actually hold in low dimensions. We want to compute a low-dimensional exponent, and in general that seems to be a difficult thing to do, but here we have that exponents take simple values and that we already have a one-sided bound which is of the correct order in low dimensions. The effectively-long-range condition, together with our hydrodynamic condition, turn out to imply that **all of the ODE analysis for $\mathbb{E}[|K|]$ and $\mathbb{E}[|K|^2]$ works**, meaning that the errors go to zero (for higher moments we need to strengthen our definition since we always need two more moments of control). Indeed, recall that in our ODE we had a bound of the form

$$0 \le \varepsilon_{1,r} \le \frac{\mathbb{E}[|K|^2|K \cap B_r|]}{|B_r|\mathbb{E}[|K|]^2},$$

which we just used tree-graph to bound in the past. But now our LR assumption allows us to bound this by

$$\begin{aligned} \varepsilon_{1,r} &\lesssim \frac{\mathbb{E}[|K \cap B_{\lambda r}|^3}{|B_r|\mathbb{E}[|K|]^2} \\ &\lesssim \frac{M_{\lambda r}^2 \mathbb{E}[|K|]}{|B_r|\mathbb{E}[|K|]^2}, \end{aligned}$$

last step by the universal tightness theorem (we're saying that we can write down $\mathbb{E}[|K \cap \Lambda|^{p+q}] \lesssim M(\Lambda)^q \mathbb{E}[|K \cap \Lambda|^p$ because the tail starts to decay very quickly past this edian size). So if the susceptibility behaves like r^α and the M term is appropriately small, we can use a bootstrapping condition to show that the error actually does go to zero. A similar thing also works for the second moment, and it shows that we must in fact have $\mathbb{E}_r[|K|^2]$ regularly varying of index 3α.

But on the other hand, by the LR assumption, we also have that $\mathbb{E}_{\beta_c,r}[|K|^2] \lesssim \mathbb{E}_{\beta_c,r}[|K \cap B_{\lambda r}|^2]$; again pulling out a factor of the maximum cluster size bounds this as $\mathbb{E}_{\beta_c,r}[|K|^2] \lesssim M_{|lambdar} \mathbb{E}_{\beta_c,r}[|K|]$, which is $o(r^{(d+\alpha)/2} r^\alpha)$ by the hydrodynamic condition. And this is strictly smaller than 3α when $d < 3\alpha$, a contradiction.

Thus in summary, the way to analyze the low-dimensional case and get the exponents is to start with bounds on the two-point function which are always true. We then say that we want them to actually be of the correct order, and we assume for contradiction that they're not the case. That yields exactly the right conditions to let us carry out our high-dimensional analysis, which leads us to prove something false.