

Lectures on percolation critical phenomena

Tom Hutchcroft

Princeton Graduate Course, Fall 2026

Contents

1	Percolation Basics	1
1.1	Non-Triviality of p_c	2
1.2	Ergodicity and Tail Triviality	4
1.3	The number of infinite clusters	5
1.4	Sharpness of the phase transition	8
1.5	Consequences for $\mathbb{Z}^2$	11
1.6	Proof of the Harris-FKG inequality	14
2	Introduction to Percolation Critical Phenomena	15
2.1	Conjectures regarding critical exponents and scaling limits	19
2.2	Heuristic scaling theory	19
2.3	Percolation on the binary tree	26
2.4	Branching random walk	31
3	Rigorous relationships between critical exponents	35
3.1	Scaling and hyperscaling in the plane	35
3.2	The BK inequality	35
3.3	Tree-graph inequalities and the mean-field lower bound on the volume tail	35
3.4	The triangle condition	35
3.5	Equivalence of mean-field critical exponents	35
3.6	Critical exponents for intrinsic geometry	35
3.7	Hyperscaling inequalities	35
4	Homework sets	35
4.1	Homework 1	35

1 Percolation Basics

Let $G = (V, E)$ be a countable, locally finite graph. Here “locally finite” means that every vertex has finite degree.

Definition 1 (Bernoulli bond percolation). Fix $p \in [0, 1]$. Bernoulli bond percolation on G with retention probability p is the random subgraph obtained by retaining each edge independently with probability p , and deleting it with probability $1 - p$.

We call retained edges *open* and deleted edges *closed*. The connected components of the open subgraph are called *clusters*. We write $\mathbb{P}_p$ and $\mathbb{E}_p$ for probability and expectation with respect to Bernoulli bond percolation with parameter p .

The standard monotone coupling. Let $(U_e)_{e \in E}$ be independent random variables, each uniform on $[0, 1]$. For $p \in [0, 1]$, define a configuration

$$\omega_p(e) = \mathbf{1}\{U_e \leq p\}, \quad e \in E.$$

Then ω_p has law $\mathbb{P}_p$. Moreover, if $q \geq p$, then

$$\omega_q(e) \geq \omega_p(e) \quad \text{for every } e \in E.$$

Exercise 1. *Is this the only coupling of ω_p and ω_q with this property?*

Thus the open configuration at larger p contains the open configuration at smaller p . As a consequence, if $\mathcal{A} \subseteq \{0, 1\}^E$ is an increasing event (meaning that if $\omega \in \mathcal{A}$ then $\omega' \in \mathcal{A}$ for every $\omega' \geq \omega$), then

$$p \mapsto \mathbb{P}_p(\mathcal{A})$$

is increasing. More generally if $F : \{0, 1\}^E \rightarrow [0, \infty]$ is any increasing function then $p \mapsto \mathbb{E}_p F(\omega)$ is increasing.

Definition 2 (Critical probability). Suppose that G is infinite. Define

$$p_c(G) = \inf\{p \in [0, 1] : \mathbb{P}_p(\text{there exists an infinite cluster}) > 0\}.$$

We will later prove the following proposition as an immediate consequence of Kolmogorov's 0-1 law (which we will prove).

Proposition 3. *If $p > p_c(G)$, then $\mathbb{P}_p(\text{there exists an infinite cluster}) = 1$.*

1.1 Non-Triviality of p_c

Note that if $G = \mathbb{Z}$ is the infinite line graph then $p_c(G) = 1$: If we delete any positive density of edges, then no vertex is connected to infinity. We next discuss why p_c is neither 0 nor 1 in less degenerate examples.

A lower bound from bounded degree. We begin with the following very general lemma yielding *positivity* of $p_c(G)$ when G has bounded degrees.

Lemma 4. *If G has degrees bounded by Δ , then*

$$p_c(G) \geq \frac{1}{\Delta - 1}.$$

Proof. Fix a vertex $x \in V$. Let $N_n(x)$ be the number of open simple paths of length n starting at x . The number of simple paths of length n starting at x is at most

$$\Delta(\Delta - 1)^{n-1},$$

where a **simple path** is a path in the graph that does not visit any vertex more than once. (At the first step we have at most Δ choices, while at each subsequent step we have at most $\Delta - 1$ choices to continue the path.) Each such path is open with probability p^n . Hence

$$\mathbb{E}_p N_n(x) \leq \Delta(\Delta - 1)^{n-1} p^n.$$

If $p < 1/(\Delta - 1)$, this tends to 0 as $n \rightarrow \infty$. If x is in an infinite cluster, then for every n there is an open simple path of length n starting from x . Therefore

$$\mathbb{P}_p(x \leftrightarrow \infty) = 0 \quad \text{when } p < \frac{1}{\Delta - 1}.$$

Taking a countable union over all vertices gives

$$\mathbb{P}_p(\text{there exists an infinite cluster}) = 0$$

for every such p . Hence $p_c(G) \geq 1/(\Delta - 1)$. $\square$

For $\mathbb{Z}^d$, this gives the concrete bound $p_c(\mathbb{Z}^d) \geq 1/(2d - 1)$. This is not optimal. (We will later see that it *is* optimal for trees.)

Exercise 2. Prove that $p_c(\mathbb{Z}^d) > 1/(2d - 1)$ for every $d \geq 2$.

The Peierls argument. We now prove that $p_c(\mathbb{Z}^d) < 1$ when $d \geq 2$.

Theorem 5. $p_c(\mathbb{Z}^d) < 1$ for every $d \geq 2$.

Proof. Since $\mathbb{Z}^2$ is a subgraph of $\mathbb{Z}^d$ for every $d \geq 2$, it suffices to prove the claim with $d = 2$. We use planar duality. If 0 is not connected to infinity in the open configuration, then there exists a closed simple dual loop surrounding 0. Figure 1 illustrates this obstruction.

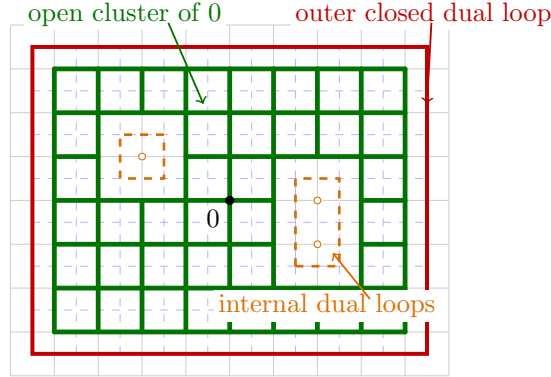


Figure 1: A finite open cluster of the origin in $\mathbb{Z}^2$. The thin grey grid is the primal lattice and the dashed blue grid is the dual lattice. The cluster may have internal holes, but the Peierls argument counts an outer closed dual loop surrounding the whole cluster.

To count the number of dual loops of length n surrounding the origin, note that each such loop must cross one of the first n edges of the positive x axis. Once this edge is chosen there are at most 3 ways to continue the path at each step, so that the total number of paths is at most $n3^n$. Meanwhile, the probability that a fixed such loop is closed is $(1 - p)^n$, so that

$$\mathbb{P}_p(0 \leftrightarrow \infty) \leq \sum_{n \geq 1} n3^n(1 - p)^n.$$

This sum is finite when $p > 2/3$ and small when $1 - p$ is sufficiently small. In particular, if p is sufficiently close to 1 then $\sum_{n \geq 1} n3^n(1 - p)^n$ is strictly smaller than 1 and $\mathbb{P}_p(0 \leftrightarrow \infty) > 0$ as required. $\square$

Exercise 3. Sharpen this argument to show that $p_c(\mathbb{Z}^2) \leq 2/3$.

Exercise 4. Prove that $p_c(\mathbb{Z}^{d+1}) < p_c(\mathbb{Z}^d)$ for each $d \geq 1$.

1.2 Ergodicity and Tail Triviality

We now introduce the basic zero-one laws relevant to percolation, which we will use to prove Proposition 3. Let E be a countably infinite set and let $(X_e)_{e \in E}$ be a family of random variables.

Definition 6 (Tail event). An event $\mathcal{A} \subseteq \{0, 1\}^E$ is a tail event¹ if changing finitely many coordinates cannot change whether $\mathcal{A}$ occurs. That is, for every finite $F \subseteq E$, if ω and ω' agree on $E \setminus F$, then

$$\omega \in \mathcal{A} \iff \omega' \in \mathcal{A}.$$

The collection of all tail events is called the tail σ -algebra and is denoted by $\mathcal{T}$.

Definition 7 (Tail triviality). The family $(X_e)_{e \in E}$ is tail trivial if every tail event has probability 0 or 1.

Theorem 8 (Kolmogorov's 0-1 law). *Independent random variables are tail trivial.*

Proof of theorem 8. Let $\mathcal{A}$ be a tail event. Recall (from introductory measure theory, and more specifically the monotone class theorem) that every event can be approximated by a cylinder event: for every $\varepsilon > 0$, there exists an event $\mathcal{A}_\varepsilon$ depending on only finitely many coordinates such that

$$\mathbb{P}(\mathcal{A} \Delta \mathcal{A}_\varepsilon) \leq \varepsilon.$$

Since $\mathcal{A}$ is a tail event, and $\mathcal{A}_\varepsilon$ depends only on finitely many coordinates, the events $\mathcal{A}$ and $\mathcal{A}_\varepsilon$ are independent. It follows that

$$\mathbb{P}(\mathcal{A} \cap \mathcal{A}_\varepsilon) = \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{A}_\varepsilon) \quad \text{and} \quad \mathbb{P}(\mathcal{A} \cup \mathcal{A}_\varepsilon) = 1 - (1 - \mathbb{P}(\mathcal{A}))(1 - \mathbb{P}(\mathcal{A}_\varepsilon)).$$

and hence that

$$\mathbb{P}(\mathcal{A} \Delta \mathcal{A}_\varepsilon) = \mathbb{P}(\mathcal{A})(1 - \mathbb{P}(\mathcal{A}_\varepsilon)) + (1 - \mathbb{P}(\mathcal{A}))\mathbb{P}(\mathcal{A}_\varepsilon).$$

Letting $\varepsilon \downarrow 0$ we get that $2\mathbb{P}(\mathcal{A})(1 - \mathbb{P}(\mathcal{A})) = 0$ and hence that $\mathbb{P}(\mathcal{A}) \in \{0, 1\}$ as claimed. $\square$

Proof of Proposition 3. The event $\mathcal{A} = \{\text{there exists an infinite cluster}\}$ is a tail event since changing finitely many edges cannot create or destroy the existence of an infinite open cluster. By Kolmogorov's 0-1 law, $\mathbb{P}_p(\mathcal{A}) \in \{0, 1\}$ for every p . If $p > p_c(G)$, then by the definition of $p_c(G)$ and monotonicity there is some $q < p$ such that $\mathbb{P}_q(\mathcal{A}) > 0$, and hence $\mathbb{P}_p(\mathcal{A}) > 0$. Since $\mathbb{P}_p(\mathcal{A})$ is either 0 or 1, it must be 1. $\square$

Ergodicity. The other important 0-1 law for percolation comes from *ergodicity* of product measures. Let E be a countable set, and let Γ be a group acting on E . The main example you should keep in mind is when E is the edge set of a **transitive** graph $G = (V, E)$ such as $\mathbb{Z}^d$ (meaning that for every two vertices $u, v \in V$ there is a graph automorphism mapping u to v). The action induces an action on configurations by

$$(\gamma\omega)(e) = \omega(\gamma^{-1}e).$$

Similarly, given an event $\mathcal{A} \subseteq \{0, 1\}^E$, we define $\gamma\mathcal{A} = \{\gamma\omega : \omega \in \mathcal{A}\}$. A measure μ on $\{0, 1\}^E$ is said to be **invariant** under the action $\Gamma \curvearrowright E$ if whenever ω has law μ , the random variable $\gamma\omega$ also has law μ , or equivalently if $\mu(\gamma\mathcal{A}) = \mu(\mathcal{A})$ for every event $\mathcal{A} \subseteq \{0, 1\}^E$. An event $\mathcal{A} \subseteq \{0, 1\}^E$ is invariant (or, more accurately, invariant under the action $\Gamma \curvearrowright E$) if $\gamma\mathcal{A} = \mathcal{A}$ for every $\gamma \in \Gamma$, or equivalently if

$$\omega \in \mathcal{A} \implies \gamma\omega \in \mathcal{A} \quad \text{for every } \gamma \in \Gamma.$$

We write $\mathcal{I} = \mathcal{I}(\Gamma \curvearrowright E)$ for the sigma-algebra of invariant events.

¹By an "event" we mean a measurable subset of $\{0, 1\}^E$. Usually this will mean Borel measurable.

Definition 9. We say that a measure μ on $\{0,1\}^E$ that is invariant under the action $\Gamma \curvearrowright E$ is **ergodic** (with respect to the action $\Gamma \curvearrowright E$) if every invariant event has probability in $\{0,1\}$.

Proposition 10. *Suppose that for every finite $F \subseteq E$ there exists $\gamma \in \Gamma$ such that $\gamma F \cap F = \emptyset$. Then Γ -invariant product measures $\{0,1\}^E$ are ergodic with respect to the action of Γ .*

Proof of proposition 10. Let $\mathcal{A}$ be an invariant event. Choose a cylinder event $\mathcal{A}_\varepsilon$, depending only on a finite set $F \subseteq E$, such that

$$\mathbb{P}(\mathcal{A} \Delta \mathcal{A}_\varepsilon) \leq \varepsilon.$$

By assumption, choose $\gamma \in \Gamma$ with $\gamma F \cap F = \emptyset$. Then $\mathcal{A}_\varepsilon$ and $\gamma \mathcal{A}_\varepsilon$ depend on disjoint sets of coordinates, and hence are independent. Since $\mathcal{A}$ and $\mathbb{P}$ are invariant,

$$\mathbb{P}(\mathcal{A} \Delta \gamma \mathcal{A}_\varepsilon) = \mathbb{P}(\gamma(\mathcal{A} \Delta \mathcal{A}_\varepsilon)) = \mathbb{P}(\mathcal{A} \Delta \mathcal{A}_\varepsilon) \leq \varepsilon.$$

It follows that

$$\mathbb{P}(\mathcal{A}_\varepsilon \Delta \gamma \mathcal{A}_\varepsilon) \leq \mathbb{P}(\mathcal{A}_\varepsilon \Delta \gamma \mathcal{A}) + \mathbb{P}(\mathcal{A} \Delta \gamma \mathcal{A}_\varepsilon) \leq 2\varepsilon.$$

Since $\mathcal{A}_\varepsilon$ and $\gamma \mathcal{A}_\varepsilon$ are independent, the same calculation as in Kolmogorov's 0-1 law gives

$$2\mathbb{P}(\mathcal{A})(1 - \mathbb{P}(\mathcal{A})) = 0$$

after sending $\varepsilon \downarrow 0$. Thus $\mathbb{P}(\mathcal{A}) \in \{0,1\}$ as claimed. $\square$

Exercise 5. *Prove under the hypotheses of Proposition 10 that every tail-trivial measure that is invariant under Γ is ergodic.*

Exercise 6. *Let $E = \mathbb{Z}$, and let $\Gamma = \mathbb{Z}$ act on $\mathbb{Z}$ by translations. Is every ergodic process tail trivial?*

Exercise 7. *Let $(X_n)_{n \in \mathbb{Z}}$ and $(Y_n)_{n \in \mathbb{Z}}$ be ergodic sequences, and suppose that these two sequences are independent of each other. Must*

$$(\max(X_n, Y_n))_{n \in \mathbb{Z}}$$

be ergodic?

1.3 The number of infinite clusters

Our next goal is to understand the number of infinite clusters in the supercritical regime. We begin with the following easy and general theorem.

Theorem 11 (Newman–Schulman). *If G is transitive, then the number N_∞ of infinite clusters is almost surely equal to a constant in $\{0,1,\infty\}$.*

Proof of Theorem 11. For every $k \in \{0,1,2,\dots\} \cup \{\infty\}$, the event $\{N_\infty = k\}$ is invariant under the automorphism group of G . (Note that this event is *not* a tail event.) Since Bernoulli percolation on a transitive graph is ergodic under this action, there exists some $k \in \{0,1,2,\dots\} \cup \{\infty\}$ such that

$$\mathbb{P}_p(N_\infty = k) = 1.$$

It remains to rule out $1 < k < \infty$.

Suppose for a contradiction that $\mathbb{P}_p(N_\infty = k) = 1$ for some $1 < k < \infty$. Fix a root vertex $o \in V$, and let B_r be the ball of radius r around o . Since there are exactly k infinite clusters almost surely

$$\mathbb{P}_p(B_r \text{ intersects all } k \text{ infinite clusters}) \rightarrow 1 \quad \text{as } r \rightarrow \infty.$$

and hence

$$\mathbb{P}_p(\text{there is an infinite cluster disjoint from } B_r) \rightarrow 0 \quad \text{as } r \rightarrow \infty.$$

Note that the second event depends only on the edges outside B_r , so that the state of the edges in B_r is independent of it. Thus, for all sufficiently large r , with positive probability B_r intersects every infinite cluster and all the edges of B_r are open. On this event, the k infinite clusters are connected together inside B_r , so $N_\infty = 1$. This contradicts the assumption that $N_\infty = k$ almost surely. Hence no finite value $k \geq 2$ is possible and $N_\infty \in \{0, 1, \infty\}$ almost surely. $\square$

Remark 12. The proof of Newman–Schulman works for any percolation process (i.e., an invariant measure on $\{0, 1\}^E$) on a transitive graph that is both ergodic and **insertion tolerant**, meaning that if ω is a sample of this process then the law of $\omega \cup \{e\}$ is absolutely continuous with respect to the law of ω .

The following two exercises show that the hypotheses of Newman–Schulman are necessary.

Exercise 8. Give an example of a connected graph and a percolation process on it such that the number of infinite clusters is random.

Exercise 9 (Harder). Give an example of an automorphism-invariant random subgraph of $\mathbb{Z}^2$ with exactly k infinite clusters almost surely.

Uniqueness of the Infinite Cluster on $\mathbb{Z}^d$. Our next goal is to prove the following theorem.

Theorem 13 (Aizenman–Kesten–Newman; Burton–Keane). *For Bernoulli bond percolation on $\mathbb{Z}^d$, there is at most one infinite cluster almost surely.*

This theorem was originally proven by Aizenman, Kesten, and Newman. We follow the simpler and more general proof of Burton and Keane (although we note that the Aizenman–Kesten–Newman proof is quantitatively stronger and has been highly influential in several recent works on percolation.)

Proof of Theorem 13. By the Newman–Schulman theorem, it remains to rule out the possibility that $N_\infty = \infty$ almost surely. Suppose for a contradiction that

$$\mathbb{P}_p(N_\infty = \infty) = 1.$$

In this case there exists r such that, with positive probability, the box

$$\Lambda_r = [-r, r]^d \cap \mathbb{Z}^d$$

intersects at least three infinite clusters. By changing only the states of the finitely many edges inside Λ_r , we can connect the origin to these three infinite clusters and arrange that the origin is a **trifurcation point**, meaning that it belongs to an infinite open cluster and that deleting the origin and all open edges incident to it from this cluster breaks the cluster into exactly three infinite connected components.

Exercise 10. Prove this in detail.

Since Bernoulli percolation is translation-invariant, this implies that

$$c := \mathbb{P}_p(x \text{ is a trifurcation point})$$

is positive and does not depend on x . Letting Y be the set of trifurcation points, it follows by linearity of expectation that

$$\mathbb{E}_p|Y \cap \Lambda_r| = c|\Lambda_r|.$$

We now use a deterministic counting lemma.

Lemma 14. *Let T be an infinite tree with no leaves, and let $A \subseteq V(T)$ be finite. Then there are at least*

$$2 + \sum_{a \in A} (\deg_T(a) - 2)$$

edge-disjoint paths from A to infinity in T .

Proof of lemma 14. Let Γ be the smallest connected subtree of T containing A , and let I be the set of edges incident to A but not contained in Γ . Each edge of I begins an edge-disjoint path from A to infinity. Let W be the vertex set of Γ . Every vertex in $W \setminus A$ has degree at least 2 inside Γ , since otherwise it could be pruned without disconnecting A . Since Γ is a finite tree,

$$\sum_{w \in W} \deg_\Gamma(w) = 2|W| - 2.$$

It follows that

$$\sum_{a \in A} \deg_\Gamma(a) \leq 2|W| - 2 - 2|W \setminus A| = 2|A| - 2$$

and hence that

$$|I| = \sum_{a \in A} (\deg_T(a) - \deg_\Gamma(a)) \geq 2 + \sum_{a \in A} (\deg_T(a) - 2)$$

as required. $\square$

We now resume the proof of Theorem 13. For each infinite open cluster, choose a spanning tree and then prune leaves until no leaves remain. In the resulting forest, each trifurcation point has degree at least 3. Applying lemma 14 with $A = Y \cap \Lambda_r$, we deduce that there exists a collection of $|Y \cap \Lambda_r|$ edge-disjoint paths from $Y \cap \Lambda_r$ to infinity. Each such path must cross the boundary of Λ_r , so the number of such paths is at most a constant times $|\partial\Lambda_r|$. It follows that

$$|Y \cap \Lambda_r| \leq C|\partial\Lambda_r|.$$

Taking expectations gives

$$c|\Lambda_r| = \mathbb{E}_p|Y \cap \Lambda_r| \leq C|\partial\Lambda_r|.$$

This is impossible as $r \rightarrow \infty$, since $|\Lambda_r| = \Theta(r^d)$ and $|\partial\Lambda_r| = O(r^{d-1})$. $\square$

Remark 15. The Burton–Keane argument extends to any ergodic, insertion-tolerant percolation process on an *amenable* transitive graph, where a graph is said to be amenable if there exists a sequence of finite sets F_n such that $|\partial F_n| = o(|F_n|)$. It is a well-known conjecture of Benjamini and Schramm that the converse is also true, that is, every *nonamenable* transitive graph has a phase in which there are infinitely many infinite clusters.

1.4 Sharpness of the phase transition

The goal of this section is to prove the following theorem.

Theorem 16 (Sharpness of the phase transition, weak form.). *Let $G = (V, E)$ be a transitive graph and let K be the open cluster of a fixed root vertex o . Then the following hold.*

1. $\mathbb{E}_p|K| < \infty$ for every $p < p_c$.
2. There exists $c > 0$ such that $\mathbb{P}_p(|K| = \infty) \geq c(p - p_c)$ for every $p > p_c$.

Later we will improve the conclusion of part 1 of this theorem to state that $|K|$ has an exponential tail when $p < p_c$. This is easy to prove by counting arguments $p < 1/(\Delta - 1)$; the point is to extend this to the entire subcritical regime $[0, p_c)$. This rules out a pathological intermediate regime in which the cluster of the origin is finite but heavy tailed.

There are now many different proofs of this theorem: It was first proven in independent work of Menshikov and Aizenman and Barsky in the mid 1980s. Simpler proofs were discovered by Duminil-Copin and Tassion in 2015 and more recently by Vanneuville, while a proof generalizing to certain dependent percolation models was found by Duminil-Copin, Raoufi, and Tassion. We follow the method of Duminil-Copin and Tassion.

Proof. Although our graphs are unoriented, it will be convenient to think of each edge as corresponding to a pair of oriented edges, one in each direction, and we will write e^+ and e^- for the head and tail of an oriented edge e . Given a finite set $S \subseteq V$ containing o , write

$$\partial_E^{\rightarrow} S = \{e \in E : e^- \in S, e^+ \notin S\}$$

for the oriented edge boundary of S , and write

$$o \xleftrightarrow{S} x$$

for the event that o is connected to x by an open path using only vertices of S . Define

$$\varphi_p(S) = p \sum_{e \in \partial_E^{\rightarrow} S} \mathbb{P}_p(o \xleftrightarrow{S} e^-),$$

and

$$\tilde{p}_c = \sup \left\{ p : \inf_{\substack{S \ni o \\ S \text{ finite}}} \varphi_p(S) < 1 \right\}.$$

We will prove that

$$\mathbb{E}_p|K| < \infty \quad \text{for every } p < \tilde{p}_c$$

and that there exists a positive constant c such that

$$\mathbb{P}_p(|K| = \infty) \geq c(p - \tilde{p}_c) \quad \text{for every } p > \tilde{p}_c.$$

Together these imply that $p_c = \tilde{p}_c$, and hence give the claimed sharpness statement.

Step 1: Finite susceptibility below $\tilde{p}_c$. Fix $p < \tilde{p}_c$, and choose a finite set $S \ni o$ such that $\varphi_p(S) < 1$. We first give a slightly incorrect argument. Let K_S be the cluster of o inside S , and let Δ be the set of edges in $\partial_E^{\rightarrow} S$ such that e is open $e^- \in K_S$. Observe that if we condition on K_S and Δ and let $e^+ \in \Delta$, the probability that e^+ is connected to some vertex x by an open path that does not visit the explored set K_S is at most the unconditional probability $\mathbb{P}_p(e^+ \leftrightarrow x)$: Indeed,

this conditional connection probability is just the connection probability in the subgraph of G in which all edges incident to K_S are deleted. Now observe that if the origin is connected to x but x does not belong to K_S then there must exist an edge e in Δ such that e^+ is connected to x by an open path disjoint from K_S . It follows from the preceding discussion that

$$\mathbb{P}_p(o \leftrightarrow x) \leq \mathbb{P}_p(x \in K_S) + \sum_{e \in \partial_E^+} \mathbb{P}_p(e \in \Delta) \mathbb{P}_p(e^+ \leftrightarrow x) = \sum_{e \in \partial_E^+} p \mathbb{P}_p(o \overset{S}{\leftrightarrow} e^-) \mathbb{P}_p(e^+ \leftrightarrow x),$$

where we used that $\{e \text{ open}\}$ is independent of K_S in the second equality. Summing over $x \in V$, it follows that

$$\mathbb{E}_p|K| \leq \mathbb{E}_p|K_S| + \sum_{e \in \partial_E^+} p \mathbb{P}_p(o \overset{S}{\leftrightarrow} e^-) \sum_{x \in V} \mathbb{P}_p(e^+ \leftrightarrow x) = \mathbb{E}_p|K_S| + \varphi_p(S) \mathbb{E}_p|K|.$$

(Everything we have done so far is correct.) What is tempting to do next, but not fully justified, is to subtract $\varphi_p(S) \mathbb{E}_p|K|$ from both sides and rearrange to conclude that

$$\mathbb{E}_p|K| \leq \frac{\mathbb{E}_p|K_S|}{1 - \varphi_p(S)} \leq \frac{|S|}{1 - \varphi_p(S)} < \infty.$$

Unfortunately these algebraic manipulations are not valid if $\mathbb{E}_p|K|$ is infinite, which is precisely what we want to prove is not the case.

One way we can avoid this issue is using the *ghost field*, i.e., introducing a random subset $\mathcal{G}$ of V independent of our percolation configuration ω in which each vertex is included independently at random with probability $1 - e^{-h} \leq h$. Writing $\mathbb{P}_{p,h}$ for the joint law of ω and $\mathcal{G}$, this has the property that $\mathbb{P}_{p,h}(A \cap \mathcal{G} = \emptyset) = e^{-h|A|}$ for every set $A \subseteq V$ and hence that $\mathbb{P}_{p,h}(o \leftrightarrow \mathcal{G}) = \mathbb{E}_p[1 - e^{-h|K|}]$ is related to the Laplace transform (or moment generating function) of $|K|$. In particular, it follows from standard properties of the Laplace transform (that you should prove yourself if you haven't seen before) that

$$\mathbb{E}_p|K| = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}_{p,h}(o \leftrightarrow \mathcal{G}).$$

Now, essentially the same argument given above yields that

$$\mathbb{P}_{p,h}(o \leftrightarrow \mathcal{G}) \leq \mathbb{P}_{p,h}(o \overset{S}{\leftrightarrow} \mathcal{G}) + \sum_{e \in \partial_E^+} \mathbb{P}_p(e \in \Delta) \mathbb{P}_{p,h}(e^+ \leftrightarrow \mathcal{G}) = \mathbb{P}_{p,h}(o \overset{S}{\leftrightarrow} \mathcal{G}) + \varphi_p(S) \mathbb{P}_{p,h}(o \leftrightarrow \mathcal{G}).$$

We may now safely rearrange to obtain that

$$\mathbb{P}_{p,h}(o \leftrightarrow \mathcal{G}) \leq \frac{\mathbb{P}_{p,h}(o \overset{S}{\leftrightarrow} \mathcal{G})}{1 - \varphi_p(S)} \leq \frac{h|S|}{1 - \varphi_p(S)}$$

where we used Markov's inequality to bound $\mathbb{P}_{p,h}(o \overset{S}{\leftrightarrow} \mathcal{G}) \leq \mathbb{E}|\mathcal{G} \cap S| \leq h|S|$ in the final inequality. Dividing both sides by h and taking $h \downarrow 0$ yields that

$$\mathbb{E}_p|K| \leq \frac{|S|}{1 - \varphi_p(S)} < \infty$$

as claimed.

Step 2: Lower bound on the percolation probability above $\tilde{p}_c$. We next prove that there exists a positive constant c such that

$$\mathbb{P}_p(o \leftrightarrow \infty) \geq c(p - \tilde{p}_c) \quad \text{for } p > \tilde{p}_c.$$

The main input is **Russo's formula**: If $F : \{0, 1\}^E \rightarrow \mathbb{R}$ depends on finitely many edges, then

$$\frac{d}{dp} \mathbb{E}_p F(\omega) = \sum_{e \in E} \mathbb{E}_p (F(\omega \cup \{e\}) - F(\omega \setminus \{e\})).$$

In particular, if $\mathcal{A}$ is an increasing event depending on finitely many edges, then

$$\frac{d}{dp} \mathbb{P}_p(\mathcal{A}) = \sum_{e \in E} \mathbb{P}_p(e \text{ is pivotal for } \mathcal{A}) = \frac{1}{1-p} \mathbb{E}_p \# \{e : e \text{ is closed and pivotal for } \mathcal{A}\},$$

where we say that e is pivotal for $\mathcal{A}$ if $\omega \cup \{e\} \in \mathcal{A}$ and $\omega \setminus \{e\} \notin \mathcal{A}$; the second equality follows from the observation that the events $\{e \text{ is open}\}$ and $\{e \text{ is pivotal for } \mathcal{A}\}$ are independent.

Exercise 11. *Prove Russo's formula.*

Exercise 12. *Give an example of an increasing event $\mathcal{A}$ (depending on infinitely many edges) and a value of $p \in [0, 1]$ such that $\sum_{e \in E} \mathbb{P}_p(e \text{ pivotal for } \mathcal{A})$ is finite but is not equal to the derivative of $\mathbb{P}_p(\mathcal{A})$. Bonus: Can the function be taken to be differentiable?*

Let Λ be a finite set containing o , and consider the probability $f_\Lambda(p) = \mathbb{P}_p(o \leftrightarrow \Lambda^c)$. Consider the random set $\mathcal{S}$ of vertices of Λ that are not connected to Λ^c , which contains o if and only if $o \leftrightarrow \Lambda^c$. Observe that, conditional on $\mathcal{S}$, the edges with both endpoints in $\mathcal{S}$ remain distributed as Bernoulli percolation on this graph. (Indeed, we can reveal $\mathcal{S}$ by exploring the cluster of every vertex in Λ^c , and when doing this exploration we do not reveal the status of any edge with both endpoints in $\mathcal{S}$.) Moreover, an edge e is a closed pivotal for the event $\{o \leftrightarrow \Lambda^c\}$ if and only if $o \in \mathcal{S}$, e is in the boundary of $\mathcal{S}$, and the endpoint of e in $\mathcal{S}$ is connected to o inside $\mathcal{S}$. That is, the expected number of closed pivots for $\{o \leftrightarrow \Lambda^c\}$ conditional on $\mathcal{S}$ is precisely $p^{-1} \varphi_p(\mathcal{S})$. It follows that if $p > \tilde{p}_c$ then

$$\frac{d}{dp} f_\Lambda(p) = \frac{1}{1-p} \mathbb{E}_p [\mathbf{1}(o \in \mathcal{S}) \varphi_p(\mathcal{S})] \geq \frac{1 - f_\Lambda(p)}{(1-p)p}.$$

This differential inequality can be analyzed using *Gronwall's lemma*; we will do it “by hand” but the same computation we do can be used to prove the general lemma. (Exercise: read up on Gronwall's lemma if you haven't heard of it before.) We observe that

$$\begin{aligned} \frac{d}{dp} \left(\exp \left[\int_{\tilde{p}_c}^p \frac{1}{q(1-q)} dq \right] f_\Lambda(p) \right) &= \frac{1}{p(1-p)} \exp \left[\int_{\tilde{p}_c}^p \frac{1}{q(1-q)} dq \right] f_\Lambda(p) \\ &\quad + \exp \left[\int_{\tilde{p}_c}^p \frac{1}{q(1-q)} dq \right] f'_\Lambda(p) \geq \frac{1}{p(1-p)} \exp \left[\int_{\tilde{p}_c}^p \frac{1}{q(1-q)} dq \right] \end{aligned}$$

for every $p > \tilde{p}_c$ and hence, integrating, that

$$\exp \left[\int_{\tilde{p}_c}^p \frac{1}{q(1-q)} dq \right] f_\Lambda(p) \geq \int_{\tilde{p}_c}^p \frac{1}{q(1-q)} \exp \left[\int_{\tilde{p}_c}^q \frac{1}{u(1-u)} du \right] dq$$

for every $p > \tilde{p}_c$. Since the integrand is bounded in a neighbourhood of $\tilde{p}_c$, this is easily seen to imply that there exists a positive constant c (not depending on Λ) such that $f_\Lambda(p) \geq c(p - \tilde{p}_c)$ for all $p > \tilde{p}_c$. Since this bound holds for all finite Λ it follows that $\mathbb{P}_p(o \leftrightarrow \infty) \geq c(p - \tilde{p}_c)$ for all $p > \tilde{p}_c$ as claimed. $\square$

Exponential decay of the radius. As an easy consequence of the same argument, if $p < p_c$, then there exists $c_p > 0$ such that

$$\mathbb{P}_p(\text{rad}(K) \geq r) \leq e^{-c_p r} \quad \text{for every } r \geq 1.$$

Indeed, choose a finite set $S \ni o$ with $\varphi_p(S) < 1$. If the cluster of o reaches distance $r + \text{rad}(S) + 1$, then it must first leave S and then continue for distance at least r . Thus

$$\mathbb{P}_p(\text{rad}(K) \geq r + \text{rad}(S) + 1) \leq \varphi_p(S) \mathbb{P}_p(\text{rad}(K) \geq r),$$

and iterating gives the exponential bound. We will return later to prove an analogous bound on the volume tail.

Exercise 13. *Prove that if $G = (V, E)$ is transitive then $\mathbb{E}_{p_c}|K| = \infty$.*

1.5 Consequences for $\mathbb{Z}^2$

We now apply the theorems we have proven so far to prove the following theorem of Harris and Kesten.

Theorem 17. $p_c(\mathbb{Z}^2) = \frac{1}{2}$ and there is no infinite cluster on $\mathbb{Z}^2$ when $p = p_c = 1/2$.

Remark 18. For non-planar graphs like $\mathbb{Z}^3$ there is no apparent reason to expect p_c to have an explicitly computable value.

Historically, Harris proved that there is no infinite cluster on $\mathbb{Z}^2$ when $p = 1/2$, which implies that $p_c \geq 1/2$, while Kesten proved the complementary inequality $p_c \leq 1/2$. The fact that $p_c(\mathbb{Z}^2) = 1/2$ is ultimately a consequence of the fact that $\mathbb{Z}^2$ is (isomorphic to) its own planar dual, with one important consequence of this duality encapsulated in the following proposition:

Proposition 19. *Let R_n be the rectangle $\{0, \dots, n+1\} \times \{0, \dots, n\}$ in $\mathbb{Z}^2$. The probability under $\mathbb{P}_{1/2}$ that the left edge of R_n is connected to the right edge of R_n by an open path in R_n is exactly $1/2$.*

To prove $p_c(\mathbb{Z}^2) = 1/2$, it suffices to prove that the crossing probability must converge to zero when $p < p_c$ and to 1 when $p > p_c$. To also prove there is no infinite cluster at $1/2$, it suffices to prove that the crossing probability converges to 1 under the weaker assumption that an infinite cluster exists.

Proof of Proposition 19. Let $(\mathbb{Z}^2)^\dagger$ be the planar dual of $\mathbb{Z}^2$, whose vertices are the pairs of half-integers $\{(x + 1/2, y + 1/2) : x, y \in \mathbb{Z}\}$ and in which two vertices are connected by an edge if and only if they differ by ± 1 in a single coordinate. There is a bijection between the edge set of $(\mathbb{Z}^2)^\dagger$ and the edge set of $\mathbb{Z}^2$ defined by mapping each primal edge e (i.e., edge of $\mathbb{Z}^2$) to the unique dual edge $e^\dagger$ that crosses it orthogonally. We define a dual percolation configuration $\omega^\dagger$ on $(\mathbb{Z}^2)^\dagger$ by $\omega^\dagger(e^\dagger) = 1 - \omega(e)$, which is distributed as Bernoulli $1 - p$ percolation on $(\mathbb{Z}^2)^\dagger$ when ω is Bernoulli p percolation on $\mathbb{Z}^2$.

If we consider the rectangle $R_n^\dagger = \{1/2, \dots, n + 1/2\} \times \{-1/2, \dots, n + 1/2\}$ in $(\mathbb{Z}^2)^\dagger$, we can map the pair $(R_n, R_n^\dagger)$ to $(R_n^\dagger, R_n)$ by rotating the whole picture by 90 degrees around the middle of the two rectangles, see Figure 2. It is a topological fact that ω contains a path in R_n connecting the left boundary of R_n to the right boundary of R_n if and only if $\omega^\dagger$ does *not* contain a path in $R_n^\dagger$ connecting the top boundary of $R_n^\dagger$ to the bottom boundary of $R_n^\dagger$. When $p = 1/2$, it follows by symmetry that these two events have the same probability, which must be $1/2$ since exactly one of the two events occurs. $\square$



Figure 2: Left: The dual rectangles R_5 (black) and $R_5^\dagger$ (grey). Right: A percolation configuration (blue) and its dual configuration (red). In this case there is a top-bottom crossing but no left-right crossing.

We now use the sharpness of the phase transition to prove the Kesten part of the Harris-Kesten theorem.

Proposition 20. $p_c(\mathbb{Z}^2) \geq 1/2$.

Proof. It suffices to prove that if $p < p_c(\mathbb{Z}^2)$ then the probability of a left right crossing of R_n tends to zero as $n \rightarrow \infty$. If $p < p_c(\mathbb{Z}^2)$, it follows by the sharpness of the phase transition that $\mathbb{P}_p(\text{rad}(K) \geq r) \leq e^{-cr}$ for some constant $c = c_p > 0$. If there is a left right crossing of R_n , then at least one of the vertices in the left hand side of R_n must belong to a cluster of radius at least n , and it follows by a union bound that

$$\mathbb{P}_p(\text{LR crossing of } R_n) \leq (n+1)e^{-cn}.$$

This clearly tends to zero as $n \rightarrow \infty$, completing the proof. $\square$

We now prove the Harris part of Harris-Kesten.

Proposition 21. *Percolation on $\mathbb{Z}^2$ with $p = 1/2$ has no infinite clusters almost surely.*

The proof of this proposition will apply the **Harris-FKG inequality**, which states that if $F, G : \{0, 1\}^E \rightarrow [0, \infty)$ are both increasing then $F(\omega)$ and $G(\omega)$ are non-negatively correlated under $\mathbb{P}_p$ in the sense that

$$\mathbb{E}_p F(\omega) G(\omega) \geq \mathbb{E}_p F(\omega) \mathbb{E}_p G(\omega).$$

In particular, if $\mathcal{A}$ and $\mathcal{B}$ are both increasing events then

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \geq \mathbb{P}_p(\mathcal{A}) \mathbb{P}_p(\mathcal{B}).$$

The same inequality also applies when $\mathcal{A}$ and $\mathcal{B}$ are both decreasing by symmetry. Moreover, it follows by induction that

$$\mathbb{P}_p(\cap_{i=1}^n \mathcal{A}_i) \geq \prod_{i=1}^n \mathbb{P}_p(\mathcal{A}_i)$$

for any family of increasing events $\mathcal{A}_1, \dots, \mathcal{A}_n$. We defer the proof of Harris-FKG to Section 1.6.

Remark 22. It follows from the Harris-FKG inequality that the **two-point function** $\tau_p(x, y) = \mathbb{P}_p(x \leftrightarrow y)$ for Bernoulli bond percolation on any graph satisfies the supermultiplicative inequality $\tau_p(x, z) \geq \tau_p(x, y) \tau_p(y, z)$ for all $x, y, z \in V$. This can be interpreted as stating that $-\log \tau_p(x, y)$ defines a metric on V for each $0 < p < 1$.

For the proof of Proposition 21, we will apply the Harris-FKG inequality via the following corollary known as the *square root trick*.

Lemma 23. *If $\mathcal{A}_1, \dots, \mathcal{A}_n$ are increasing events then*

$$\max_i \mathbb{P}_p(\mathcal{A}_i) \geq 1 - (1 - \mathbb{P}_p(\cup_{i=1}^n \mathcal{A}_i))^{1/n}$$

for every $p \in [0, 1]$.

The main takeaway of this lemma is that if n is fixed and $\cup_{i=1}^n \mathcal{A}_i$ holds with high probability then there must be a specific event $\mathcal{A}_i$ holding with high probability.

Proof of Lemma 23. We apply the Harris-FKG inequality to the complementary events

$$\mathbb{P}_p(\cap_{i=1}^n \mathcal{A}_i^c) \geq \prod_{i=1}^n \mathbb{P}_p(\mathcal{A}_i^c) \geq (\min \mathbb{P}_p(\mathcal{A}_i^c))^n.$$

Rearranging we obtain that $\min \mathbb{P}_p(\mathcal{A}_i) \leq \mathbb{P}_p(\cap_{i=1}^n \mathcal{A}_i^c)^{1/n}$ which is equivalent to the claim. $\square$

Proof of Proposition 21. It suffices to prove that if p is such that an infinite cluster exists almost surely under $\mathbb{P}_p$ then the $\mathbb{P}_p$ -probability of a left-right crossing of R_n converges to 1 as $n \rightarrow \infty$. Fix $p \in (0, 1]$ such that this holds, and note (by continuity of measure) that for each $\varepsilon > 0$ there exists $k < \infty$ such that $\mathbb{P}_p(\{-k, \dots, k\}^2 \leftrightarrow \infty) \geq 1 - \varepsilon$.

Let $n = 2m$ be even and consider the two boxes $\Lambda_n = \{0, \dots, 2m\} \times \{0, \dots, 2m\}$ and $\Lambda'_n = \{1, \dots, 2m+1\} \times \{0, \dots, 2m\}$ in R_n . Given $1 \leq k \leq m$, the boxes $\Lambda_k = \{m-k, \dots, m+k\} \times \{m-k, \dots, m+k\}$ and $\Lambda'_k = \{m-k+1, \dots, m+k+1\} \times \{m-k, \dots, m+k\}$ are centered in Λ_n and Λ'_n respectively, so that the probabilities Λ_k is connected to the left, right, top, or bottom boundaries of Λ_n inside Λ_n are all equal. On the other hand, if Λ_k is connected to infinity then at least one of these connection events must occur, and it follows by the square root trick and by symmetry that for each $\varepsilon > 0$ there exists k such that

$$\begin{aligned} \mathbb{P}_p(\Lambda_k \text{ is connected to the left boundary of } \Lambda_n \text{ inside } \Lambda_n \\ \text{and } \Lambda'_k \text{ is connected to the right boundary of } \Lambda'_n \text{ inside } \Lambda'_n) \geq 1 - \varepsilon \end{aligned}$$

for every $n \geq k$. On the other hand, it follows from the uniqueness of the infinite cluster that the probability that this event occurs but the two relevant clusters in R_n can be taken to be distinct tends to zero as $n \rightarrow \infty$ with k fixed. If the clusters are the same then we have a left-right crossing of R_n , and it follows that the probability of such a left-right crossing converges to 1 as $n \rightarrow \infty$ as claimed. $\square$

Exercise 14. *Consider Bernoulli percolation on $\mathbb{Z}^2$. Prove that if $p > 1/2$ then there exists $c_p > 0$ such that*

$$\mathbb{P}_p(x \leftrightarrow y, x \nleftrightarrow \infty) \leq \exp[-c_p \|x - y\|]$$

for every $x, y \in \mathbb{Z}^2$, and

$$\mathbb{P}_p(n \leq |K| < \infty) \leq e^{-c_p \sqrt{n}}$$

for every $n \geq 1$.

1.6 Proof of the Harris-FKG inequality

We now prove the Harris-FKG inequality.

Theorem 24 (Harris-FKG inequality). *If ω is a random element of $\{0, 1\}^E$ with independent coordinates for some countable E and $F, G : \{0, 1\}^E \rightarrow [0, \infty]$ are both increasing measurable functions then*

$$\mathbb{E}[F(\omega)G(\omega)] \geq \mathbb{E}F(\omega)\mathbb{E}G(\omega).$$

That is, increasing functions are non-negatively correlated under product measure.

The proof we give applies more generally to any product of totally ordered sets.

Proof of Theorem 24. It suffices to consider the case that F and G are bounded, the unbounded case following by applying this inequality to $\min\{F, M\}$ and $\min\{G, M\}$, taking $M \rightarrow \infty$ and using monotone convergence.

We first prove that the inequality holds for finite sets E , proceeding by induction on $|E|$. The base case $n = 1$ follows from the identity

$$\mathbb{E}[FG] - \mathbb{E}[F]\mathbb{E}[G] = \frac{1}{2}\mathbb{E} \otimes \mathbb{E}[(F(\omega) - F(\omega'))(G(\omega) - G(\omega'))],$$

where $\mathbb{E} \otimes \mathbb{E}$ is the law of two independent copies of the configuration ω , together with the observation that (since $\{0, 1\}$ is totally ordered) the random variables $F(\omega) - F(\omega')$ and $G(\omega) - G(\omega')$ have the same sign when F and G are both increasing. Now suppose that the claim has been proven for finite sets with at most $n - 1 < n = |E|$ coordinates, fix $e_1 \in E$, and for each $y \in \{0, 1\}$ define

$$F_1(y) = \mathbb{E}[F \mid \omega(e_1) = y], \quad G_1(y) = \mathbb{E}[G \mid \omega(e_1) = y], \quad \text{and} \quad (FG)_1(y) = \mathbb{E}[FG \mid \omega(e_1) = y].$$

The induction hypothesis implies that

$$(FG)_1(y) \geq F_1(y)G_1(y)$$

for every $y \in \{0, 1\}$. Taking expectations over $\omega(e_1)$ and applying the one-coordinate case to F_1 and G_1 yields that

$$\begin{aligned} \mathbb{E}[F(\omega)G(\omega)] &= \mathbb{E}[F(\omega)G(\omega) \mid \omega(e_1)] = \mathbb{E}[(FG)_1(\omega(e_1))] \\ &\geq \mathbb{E}[F_1(\omega(e_1))G_1(\omega(e_1))] \geq \mathbb{E}[F_1(\omega(e_1))]\mathbb{E}[G_1(\omega(e_1))] = \mathbb{E}F(\omega)\mathbb{E}G(\omega) \end{aligned}$$

as claimed, where we used the tower property of conditional expectations twice. This completes the induction step and hence the proof of the finite case.

Now suppose that $E = \{e_1, e_2, \dots\}$ is countable, and for each $n \geq 1$ define

$$F_n = \mathbb{E}[F \mid \omega(e_1), \dots, \omega(e_n)] \quad \text{and} \quad G_n = \mathbb{E}[G \mid \omega(e_1), \dots, \omega(e_n)].$$

The functions F_n and G_n are increasing functions of the finitely many coordinates $\omega(e_1), \dots, \omega(e_n)$ and hence satisfy $\mathbb{E}F_n(\omega)G_n(\omega) \geq \mathbb{E}F_n(\omega)\mathbb{E}G_n(\omega)$. Since F and G were assumed to be bounded, it follows from the martingale convergence theorem that $F_n \rightarrow F$ and $G_n \rightarrow G$ almost surely. We also have by bounded convergence that $\mathbb{E}F(\omega) = \lim_{n \rightarrow \infty} \mathbb{E}F_n(\omega)$, $\mathbb{E}G(\omega) = \lim_{n \rightarrow \infty} \mathbb{E}G_n(\omega)$, and $\mathbb{E}F(\omega)G(\omega) = \lim_{n \rightarrow \infty} \mathbb{E}F_n(\omega)G_n(\omega)$, so that the infinite statement follows from the finite statement. $\square$

Remark 25. This inequality was discovered for product measures in the work of Harris, who used it to prove that there are no infinite clusters for percolation on $\mathbb{Z}^2$ at the self-dual point $p = 1/2$. Fortuin, Kastelyn, and Ginibre later developed a more systematic theory of which measures satisfy this property (often known as *non-negative association*), leading the inequality to often be referred to as the FKG inequality.

2 Introduction to Percolation Critical Phenomena

We now begin the part of the course focussed on understanding percolation at, and very close to, the critical point p_c . A longstanding conjecture stated that there is no infinite cluster at criticality on $\mathbb{Z}^d$ for every $d \geq 2$. This was proven for $d = 2$ by Harris and Kesten and in high dimensions by Hara and Slade. The general case was resolved very recently by an autonomous AI agent following using a reduction due to Kozma and Nitzan. Our focus is instead on understanding *quantitative* questions: What do large critical clusters look like?

Broadly speaking, we expect the critical point to be characterised by *interesting* (meaning non-degenerate) *large-scale behaviour*. Indeed, if we fix p and consider a large box $[-r, r]^d$, it can be proven that the largest cluster in this box will have volume $O(\log r)$ with high probability if $p < p_c$, while if $p > p_c$ then there exists a cluster occupying a positive density of the box whose geometry is in some senses very close to that of the Euclidean box itself.² Meanwhile, at the critical point, we expect to have clusters whose *diameter* is of the same order as that of the box, but whose volume is much smaller than that of the box. Some parts of this picture are reasonably easy to verify given what we already know:

Exercise 15. Prove that if $p < p_c(\mathbb{Z}^d)$ then the largest cluster in the box $[-r, r]^d$ has volume $O(\log r)$.

Exercise 16. Consider critical percolation on $\mathbb{Z}^d$. Prove that the probability that there exists a cluster intersecting both $[-r, r]^d$ and $\mathbb{Z}^d \setminus [2r, 2r]^d$ does not converge to zero as $r \rightarrow \infty$.

Remark 26. For critical percolation $\mathbb{Z}^2$ we saw that the probability of crossing an $n \times n$ square does not converge to zero or one as $n \rightarrow \infty$. As we will discuss later, the analogous probability should converge to 1 for *high-dimensional* percolation.

In order to understand critical percolation in detail, we would like to be able to take *scaling limits*, i.e., to zoom out, looking at the lattice on larger and larger scales, and attempt to describe the fractal-like continuum limit of the random geometry that we see.

One way we can understand this fractal geometry is through *critical exponents*: We expect that many quantities of interest are governed by *power laws*, and the exponents in these power laws encode various kinds of “dimensionality” of the model. For example, we expect that there exist exponents η and ρ such that

$$\mathbb{P}_{p_c}(x \leftrightarrow y) \approx \|x - y\|^{-d+2-\eta}$$

and

$$\mathbb{P}_{p_c}(0 \leftrightarrow \partial[-r, r]^d) \approx r^{-1/\rho},$$

where the meaning of the symbol $\approx$ is left intentionally vague for use in heuristic arguments. We will discuss conjectures describing these and other exponents in detail below.

Why power laws? Before proceeding further, let us explain why the existence of scaling limits is so closely tied up with power laws. One way to take scaling limits is by taking limits of re-normalized k -point connectivity functions $\tau_{p_c}(x_1, \dots, x_k) = \mathbb{P}_{p_c}(x_1, \dots, x_k \text{ all connected})$. In particular, it is natural to hope that we have a well-defined continuum two-point function $T : \{(x, y) : x, y \in \mathbb{R}^2, x \neq y\} \rightarrow (0, \infty)$ defined by

$$T(x, y) = \lim_{\lambda \rightarrow \infty} \frac{\mathbb{P}_{p_c}([\lambda x] \leftrightarrow [\lambda y])}{f(\lambda)}$$

²Proving that a precise version of this claim holds for all $p > p_c$ is non-trivial. In the case $d \geq 3$ it is related to the Grimmett-Marstrand theorem.

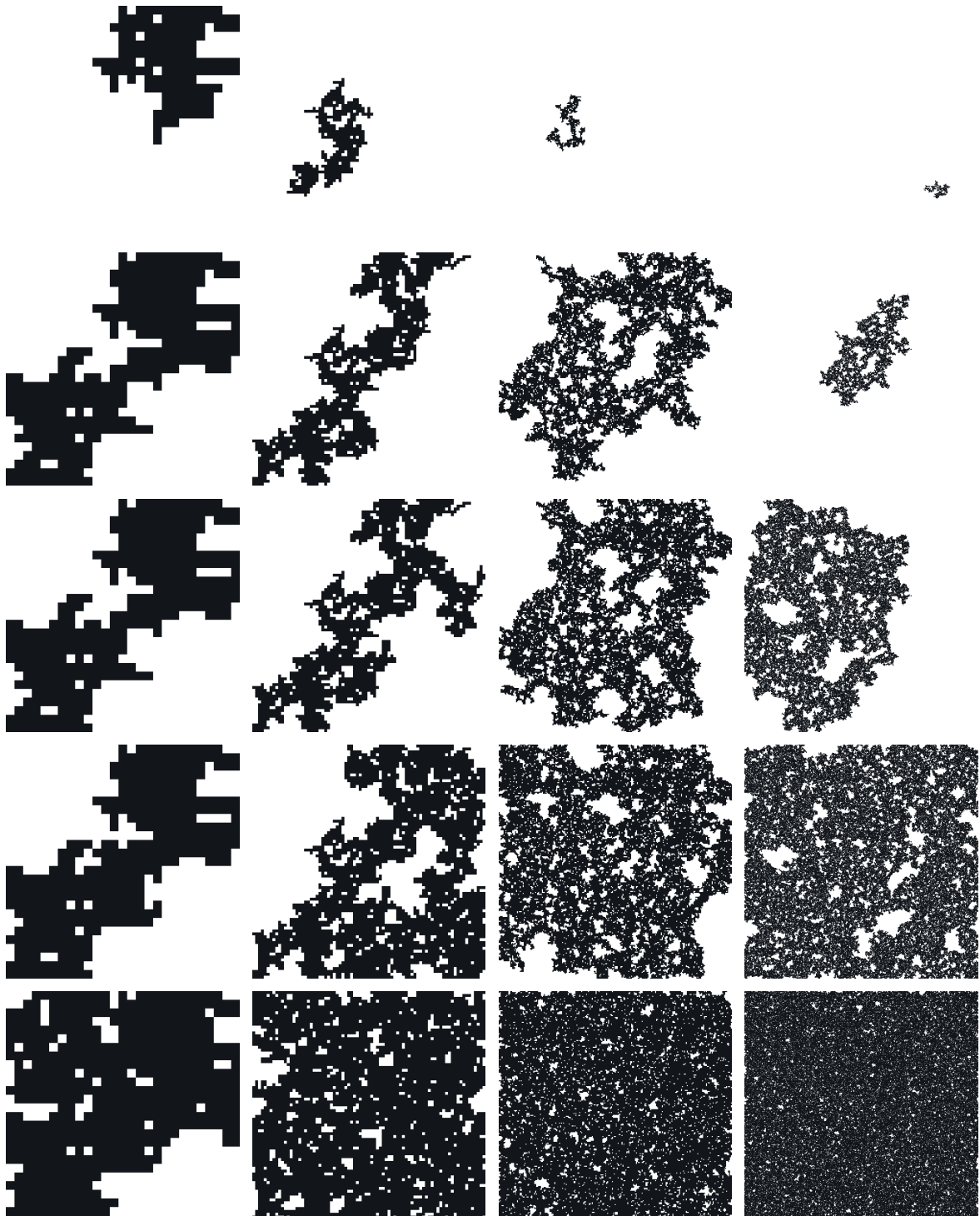


Figure 3: The largest cluster in Bernoulli bond percolation in a box of side length 3^k with $k = 3, 4, 5, 6$ on the square grid with $p = 0.45, 0.49, 0.5, 0.51$, and 0.55 . A single simulation of the monotone coupling of percolation on the box of side length 3^6 was used to generate all images, with the box at each smaller scale taken from the center of the larger scale. In each picture the largest cluster is represented as its set of vertices.

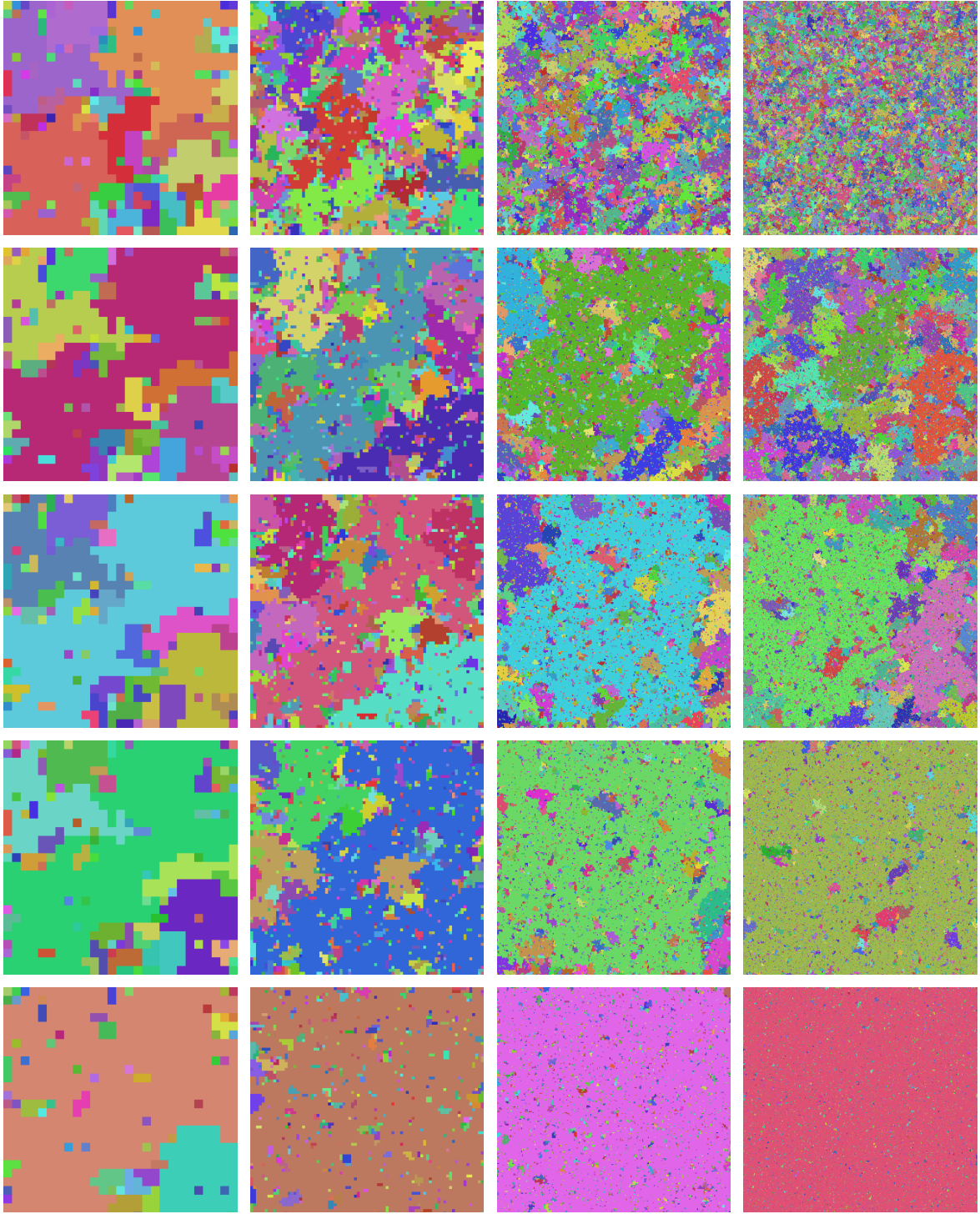


Figure 4: We can represent the entire partition of a box into clusters visually by giving every cluster a random colour. As before, we have taken a single simulation of the standard monotone coupling of Bernoulli bond percolation on the box of side length 3^7 in $\mathbb{Z}^2$ and plotted the associated configurations at the scales 3^k with $k = 4, 5, 6, 7$ and $p = 0.45, 0.49, 0.5, 0.51, 0.55$. The random colouring was done separately for each image.

for some normalizing function f , where $[x]$ denotes the lattice point closest to x . (E.g. we could take $f(\lambda) = \mathbb{P}_{p_c}(0 \leftrightarrow [\lambda x_0])$ for some fixed unit vector x_0 .) If this limit is well-defined as a function to $(0, \infty)$ then for each $\alpha > 0$ we must have that

$$\begin{aligned} T(\alpha x, \alpha y) &= \lim_{\lambda \rightarrow \infty} \frac{\mathbb{P}_{p_c}([\lambda \alpha x] \leftrightarrow [\lambda \alpha y])}{f(\lambda)} = \lim_{\lambda \rightarrow \infty} \frac{f(\alpha \lambda)}{f(\lambda)} \cdot \frac{\mathbb{P}_{p_c}([\lambda \alpha x] \leftrightarrow [\lambda \alpha y])}{f(\alpha \lambda)} \\ &= \left(\lim_{\lambda \rightarrow \infty} \frac{f(\alpha \lambda)}{f(\lambda)} \right) T(x, y). \end{aligned}$$

Thus, the normalizing function f must have a well-defined and positive asymptotic ratio $f(\alpha \lambda)/f(\lambda)$, making it a **function of regular variation**.

Definition 27 (Regular variation). A function $f : (0, \infty) \rightarrow (0, \infty)$ is **regularly varying** if

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)}$$

exists in $(0, \infty)$ for every $\lambda > 0$.

Fact 28. *If f is regularly varying and measurable, then there exists $\alpha \in \mathbb{R}$ such that*

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\alpha \quad \text{for every } \lambda > 0. \quad (1)$$

The number α is called the **index** of regular variation of f . Moreover, the convergence (1) holds uniformly for compact sets of λ in $(0, \infty)$.

Regularly varying functions of index 0 are said to be **slowly varying**. Examples include

$$\log x, \quad \log \log x, \quad \exp(\sqrt{\log x}).$$

Note that x^α is regularly varying of index α and that the product of two regularly varying functions of index α and β is regularly varying of index $\alpha + \beta$, so that every regularly varying function of index α can be written in the form

$$f(x) = x^\alpha L(x),$$

where L is slowly varying. The following exercise relates regular variation to asymptotic growth estimates.

Exercise 17. *Prove that every regularly varying function of index α satisfies $f(x) = x^{\alpha+o(1)}$ as $x \rightarrow \infty$. Does the converse hold?*

Thus, if the continuum two-point function $T(x, y) = \lim_{\lambda \rightarrow \infty} f(\lambda)^{-1} \mathbb{P}_{p_c}([\lambda x] \leftrightarrow [\lambda y])$ is well-defined as a function from pairs of distinct points in $\mathbb{R}^d$ to $(0, \infty)$ for some measurable function f then f must be regularly varying of some index $d - 2 + \eta$, so that

$$\mathbb{P}_{p_c}(0 \leftrightarrow x) \sim f(\|x\|) T(0, x/\|x\|) = \|x\|^{-d+2-\eta+o(1)}$$

as $x \rightarrow \infty$.

Similarly, we could try to take a scaling limit by conditioning on the cluster of the origin having radius at least some large number r , rescaling the cluster by a factor of r , and taking its weak limit as a random closed subset of $\mathbb{R}^d$. If this weak limit exists and is supported on bounded sets of radius strictly larger than 1, the following exercise shows that the radius of the cluster of the origin must have a power law in the sense that $\mathbb{P}_{p_c}(\text{rad}(K) \geq r)$ is regularly varying.

Exercise 18. *Let X be a non-negative random variable. Prove that the measures $\mathbb{P}(X/x \in \cdot \mid X \geq x)$ have a weak limit as $x \rightarrow \infty$ that is not a single atom at 1 if and only if $\mathbb{P}(X \geq x)$ is regularly varying of strictly negative index.*

2.1 Conjectures regarding critical exponents and scaling limits

We now describe the main conjectures regarding critical exponents and scaling limits for percolation on $\mathbb{Z}^d$ and briefly survey the progress that has been made on them.

The expected picture.

In high dimensions the critical behaviour is expected to be the same as that of branching random walk. More precisely, when $d > 6$ one expects

$$\begin{aligned}\mathbb{E}_p|K| &\approx \frac{1}{p_c - p} && \text{as } p \uparrow p_c, \\ \mathbb{P}_{p_c}(0 \leftrightarrow x) &\approx \|x\|^{-d+2} && \text{as } \|x\| \rightarrow \infty, \\ \mathbb{P}_{p_c+\varepsilon}(0 \leftrightarrow \infty) &\approx \varepsilon && \text{as } \varepsilon \downarrow 0,\end{aligned}$$

and

$$\mathbb{P}_{p_c}(0 \leftrightarrow \mathbb{Z}^d \setminus [-r, r]^d) \approx r^{-2} \quad \text{as } r \rightarrow \infty.$$

Thus the high-dimensional values are

$$\gamma = 1, \quad \eta = 0, \quad \beta = 1, \quad \rho = \frac{1}{2}.$$

Below six dimensions the exponents are expected to be different. In two dimensions their values are

$$\gamma = \frac{43}{18}, \quad \eta = \frac{5}{24}, \quad \beta = \frac{5}{36}, \quad \rho = \frac{48}{5}.$$

The physics prediction near the upper-critical dimension is expressed by the ε -expansion, which gives for example

$$\eta(6 - \varepsilon) = -\frac{\varepsilon}{21} - \frac{206}{9261}\varepsilon^2 + \dots$$

and

$$\delta(6 - \varepsilon) = 2 + \frac{2}{7}\varepsilon + \frac{565}{6174}\varepsilon^2 + \dots$$

[This section to be fleshed out.]

2.2 Heuristic scaling theory

*A little nonsense now and then
is relished by the wisest men.*

—Anon.

In this section we discuss *heuristic scaling theory*: non-rigorous arguments giving conjectural relationships between different critical exponents. As we will see in the exercises in this section, several of the arguments can be made into rigorous one-sided inequalities between critical exponents assuming these exponents exist in appropriate senses. Note that most of the heuristic arguments in this section have nothing to do with Bernoulli percolation *per se* and should apply to a very wide class of critical models.

Scaling theory for the two-point function. The central assumption underlying our heuristic derivation of the scaling relations will be that we can interpolate between critical (power-law) and off-critical (exponential or stretched-exponential) rates of decay in simple ways in the near-critical

	Quantity		Behaviour	Exponent	HD	2d
Percolation probability	$\theta(p) = \mathbb{P}_p(K = \infty)$	$\approx (p - p_c)^\beta$	as $p \downarrow p_c$	β	1	5/36
Susceptibility	$\chi(p) = \mathbb{E}_p K $	$\approx p - p_c ^{-\gamma}$	as $p \uparrow p_c$	γ	1	43/18
Cluster volume tail	$\mathbb{P}_{p_c}(K \geq n)$	$\approx n^{-1/\delta}$	as $n \uparrow \infty$	δ	2	91/5
Two-point function	$\mathbb{P}_{p_c}(0 \leftrightarrow x)$	$\approx \ x\ ^{-d+2-\eta}$	as $\ x\ \uparrow \infty$	η	0	5/24
Correlation length	$\xi(p)$	$\approx p - p_c ^{-\nu}$	as $p \uparrow p_c$	ν	1/2	4/3
One-arm probability	$\mathbb{P}_{p_c}(0 \leftrightarrow \partial[-r, r]^d)$	$\approx r^{-1/\rho}$	as $r \uparrow \infty$	ρ	1/2	48/5
Characteristic volume	$\zeta(p)$	$\approx p - p_c ^{-\Delta}$	as $p \uparrow p_c$	Δ	2	91/36
	$\zeta(p_c, r)$	$\approx r^{d_f}$	as $r \uparrow \infty$	d_f	4	91/48

Table 1: Some of the most important critical exponents for nearest-neighbour percolation and their values in high dimensions and in two dimensions. We write K for the cluster of the origin.

regime. We begin by explaining this in the case of the two-point function. Recall that the critical two-point function is conjectured to be described by

$$\mathbb{P}_{p_c}(x \leftrightarrow y) \approx \|x - y\|^{-d+2-\eta} \quad \text{as } \|x - y\| \rightarrow \infty.$$

for some $\eta \in \mathbb{R}$, where the notation $\approx$ remains intentionally vague and may suppress logarithmic terms. To derive the scaling relations, for each $p < p_c$ we posit the existence of a **correlation length** $\xi(p)$ and a rapidly decaying function f such that

$$\mathbb{P}_p(x \leftrightarrow y) \approx \mathbb{P}_{p_c}(x \leftrightarrow y) f\left(\frac{\|x - y\|}{\xi(p)}\right) \approx \|x - y\|^{-d+2-\eta} f\left(\frac{\|x - y\|}{\xi(p)}\right). \quad (2)$$

Here “rapidly decaying” means decaying faster than any power; as we will see in the exercises, the most important content of this assumption is that the scale on which the subcritical two-point function begins to become smaller than the critical two-point function by a constant factor is (up to constants) the same scale on which the subcritical two-point function begins to decay much more rapidly than the critical two-point function. The correlation length $\xi(p)$ is itself expected to have a power-law divergence as $p \uparrow p_c$ (with possible logarithmic corrections), and the traditional name for this exponent is ν :

$$\xi(p) \approx (p_c - p)^{-\nu} \quad \text{as } p \uparrow p_c.$$

Note that we have not yet given any precise definition of the correlation length, but we think of it intuitively as the scale at which the model “begins to look subcritical”. As we will soon see, there are different ways of making this precise that are not obviously equivalent, and proving the equivalence of these different definitions is closely entwined with proving the scaling relations rigorously.

Let us now see how the ansatz (2) leads to our first scaling relation: Summing the two-point function gives

$$\mathbb{E}_p|K| = \sum_{x \in \mathbb{Z}^d} \mathbb{P}_p(0 \leftrightarrow x) \approx \sum_{x \in \mathbb{Z}^d} \|x\|^{-d+2-\eta} f\left(\frac{\|x\|}{\xi(p)}\right) \approx \sum_{\|x\| \leq \xi(p)} \|x\|^{-d+2-\eta} \approx \xi(p)^{2-\eta}, \quad (3)$$

where we applied the following exercise in the third equality:

Exercise 19. Let $f : [0, \infty) \rightarrow (0, \infty)$ be decreasing and satisfy $f(t) = O(t^{-A})$ as $t \rightarrow \infty$ for every $A < \infty$. Let g be regularly varying of index $\alpha > -d$. Prove that

$$\sum_{x \in \mathbb{Z}^d} g(\|x\|) f\left(\frac{\|x\|}{\lambda}\right) \asymp \sum_{\|x\| \leq \lambda} g(\|x\|)$$

as $\lambda \rightarrow \infty$.

Writing the relation (3) in terms of critical exponents yields the (non-rigorous) scaling relation

$$\gamma = (2 - \eta)\nu. \quad (4)$$

Two precise definitions of the correlation length for $p < p_c$ that are compatible with this calculation are the **radius of gyration** (a.k.a. correlation length of order 2)

$$\xi_{\text{gyr}}(p) = \sqrt{\frac{\sum_{x \in \mathbb{Z}^d} \|x\|_2^2 \mathbb{P}_p(0 \leftrightarrow x)}{\mathbb{E}_p|K|}}$$

and the **stability radius**

$$\xi_{\text{stab}}(p) = \min \left\{ \|x\| : x \in \mathbb{Z}^d, \frac{\mathbb{P}_p(0 \leftrightarrow x)}{\mathbb{P}_{p_c}(0 \leftrightarrow x)} \leq \frac{1}{2} \right\}.$$

The following exercise shows that the scaling relation $\gamma = (2 - \eta)\nu$ is rigorously justified under the assumption that these two notions of correlation length are the same order (which would be a consequence of the ansatz formula (2).)

Exercise 20. In this question we investigate the scaling relation $\gamma = (2 - \eta)\nu$ under different precise definitions of the correlation length. We will assume that the exponents γ and η are well-defined in the sense that $\mathbb{P}_{p_c}(x \leftrightarrow y) = \|x - y\|^{-d+2-\eta+o(1)}$ as $\|x - y\| \rightarrow \infty$ and $\mathbb{E}_{p_c-\varepsilon}|K| = \varepsilon^{-\gamma+o(1)}$ as $\varepsilon \downarrow 0$.

1. Prove that if the exponent ν is well-defined in the sense that $\xi_{\text{gyr}}(p_c - \varepsilon) = \varepsilon^{-\nu+o(1)}$ as $\varepsilon \downarrow 0$ then

$$\gamma \leq (2 - \eta)\nu.$$

2. Prove that if the exponent ν is well-defined in the alternative sense that $\xi_{\text{stab}}(p_c - \varepsilon) = \varepsilon^{-\nu+o(1)}$ as $\varepsilon \downarrow 0$ then

$$\gamma \geq (2 - \eta)\nu.$$

Deduce that if these two measures of the correlation length ξ_{gyr} and ξ_{stab} are equal to within constant factors then the scaling relation $\gamma = (2 - \eta)\nu$ holds whenever all three exponents are well-defined.

Scaling theory for cluster volumes. We can also formulate a scaling hypothesis for the volume of the critical cluster. Recall that the critical exponent for the volume tail is traditionally written as

$$\mathbb{P}_{p_c}(|K| \geq n) \approx n^{-1/\delta} \quad \text{as } n \rightarrow \infty,$$

where this notation may hide logarithmic corrections to scaling. For $p < p_c$, our scaling hypothesis for the cluster volume tail posits the existence of a **characteristic volume** $\zeta(p)$ and a rapidly decaying function g such that

$$\mathbb{P}_p(|K| \geq n) \approx n^{-1/\delta} g\left(\frac{n}{\zeta(p)}\right). \quad (5)$$

We further posit that $\zeta(p)$ has power-law divergence as $p \uparrow p_c$; the exponent for this divergence is known as the **gap exponent** and traditionally denoted by

$$\zeta(p) \approx (p_c - p)^{-\Delta} \quad \text{as } p \uparrow p_c.$$

Using these assumptions, a similar calculation to before yields a relationship between the exponents γ, δ , and Δ : Indeed, we can write

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

which yields the (heuristic) scaling relation

$$\gamma = \Delta \frac{\delta - 1}{\delta}. \quad (6)$$

Note moreover that the same computation for higher moments gives that

$$\mathbb{E}_p|K|^k \approx \sum_{n \geq 1} n^{k-1} \mathbb{P}_p(|K| \geq n) \approx \sum_{n=1}^{\zeta(p)} n^{k-1-1/\delta} \approx \zeta(p)^{k-1/\delta} \approx (p_c - p)^{-\gamma - (k-1)\Delta}$$

or equivalently that

$$\mathbb{E}_p|K|^k \approx \mathbb{E}_p|K| \cdot \zeta(p)^{k-1}. \quad (7)$$

Again, we have not yet given a precise definition of the characteristic volume $\zeta(p)$ not relying on an appeal to the guessed formula (5). One reasonable definition is to take $\zeta(p) = \mathbb{E}_p|K|^2 / \mathbb{E}_p|K|$ to make (7) vacuously true in the case $k = 2$.

The following exercise explores how the scaling relation $\delta\gamma = \Delta(\delta - 1)$ can be proven rigorously under the assumption that this characteristic volume coincides with the first volume scale at which the subcritical volume tail becomes significantly smaller than the critical volume tail.

Exercise 21. *In this question we investigate the scaling relation $\Delta = \delta\gamma/(\delta - 1)$ under different precise definitions of the characteristic cluster volume $\zeta(p)$ for $p < p_c$. We will assume that the exponents γ and δ are well-defined in the sense that $\mathbb{E}_{p_c - \varepsilon}|K| = \varepsilon^{-\gamma+o(1)}$ as $\varepsilon \downarrow 0$ and $\mathbb{P}_{p_c}(|K| \geq n) = n^{-1/\delta+o(1)}$ as $n \rightarrow \infty$. We will also assume that $\delta > 1$.*

1. *Prove that if we define*

$$\zeta(p) = \frac{\mathbb{E}_p|K|^2}{\mathbb{E}_p|K|}$$

for $p < p_c$ and the exponent Δ is well-defined in the sense that $\zeta(p_c - \varepsilon) = \varepsilon^{-\Delta+o(1)}$ as $\varepsilon \downarrow 0$ then

$$\Delta \geq \frac{\delta\gamma}{\delta - 1}.$$

(Hint: First prove that if X is a positive random variable then the function $a \mapsto \log \mathbb{E}X^a$ is convex on the set where it is finite.)

2. *Prove that if we instead define $\zeta(p)$ for $p < p_c$ to be the minimal volume for which $\mathbb{P}_p(|K| \geq n)$ becomes significantly smaller than it is at criticality:*

$$\zeta(p) = \min \left\{ n \geq 1 : \mathbb{P}_p(|K| \geq n) \leq \frac{1}{2} \mathbb{P}_{p_c}(|K| \geq n) \right\}$$

Prove that if the exponent Δ is well-defined in the sense that $\zeta(p_c - \varepsilon) = \varepsilon^{-\Delta+o(1)}$ as $\varepsilon \downarrow 0$ then

$$\Delta \leq \frac{\delta\gamma}{\delta-1}.$$

Deduce that if these two measures of the characteristic volume are equal to within constant factors then the scaling relation $\Delta = \delta\gamma/(\delta-1)$ holds.

Open Problem 29. Is it true that

$$\mathbb{E}_p |K|^k \asymp \mathbb{E}_p |K| \left(\frac{\mathbb{E}_p |K|^2}{\mathbb{E}_p |K|} \right)^{k-1}$$

for Bernoulli percolation on $\mathbb{Z}^d$ for every integer $k \geq 1$ (where the implicit constants may depend on k)? What about on an arbitrary transitive graph?

The percolation probability. We now include the percolation probability exponent $\mathbb{P}_{p_c+\varepsilon}(0 \leftrightarrow \infty) \approx \varepsilon^\beta$ into our scaling relations, which will require a somewhat greater leap of faith than our other heuristic derivations. Finally, one can use the same characteristic-volume picture on both sides of the critical point. First, we suppose that our scaling hypothesis (5) can be extended to both sides of the critical point via

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

where the sign depends on whether p is below or above p_c and the two functions g_- and g_+ are rapidly decaying but possibly distinct. We now suppose (without any strong argument in favour) that the characteristic volume is roughly symmetric between the subcritical and supercritical regimes:

$$\zeta(p_c + \varepsilon) \approx \zeta(p_c - \varepsilon) \quad \text{as } \varepsilon \downarrow 0.$$

For $p > p_c$, the infinite cluster should arise from the finite clusters at criticality whose volume is at least the characteristic volume, giving

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

and hence that

$$\beta = \frac{\Delta}{\delta}. \tag{8}$$

Again, the heuristic argument behind this scaling relation may seem weaker than the other arguments presented in this section, but it is still expected to be broadly correct.

The fractal dimension and arm exponent. Further scaling relations can be derived from scaling hypotheses in which, rather than taking p off-critical, we instead work at criticality but restrict to finite volume. In particular, we may posit the existence of a rapidly decaying function h and a characteristic volume $\zeta(p_c, r)$ such that

$$\mathbb{P}_{p_c}(|K \cap [-r, r]^d| \geq n) \approx n^{-1/\delta} h \left(\frac{n}{\zeta(p_c, r)} \right). \tag{9}$$

Again, the main content of this assumption is that the volume scale at which the tail $\mathbb{P}_{p_c}(|K \cap [-r, r]^d| \geq n)$ begins to become significantly smaller than $\mathbb{P}_{p_c}(|K| \geq n)$ is the same volume scale at

which the same tail begins to decay rapidly. Again, we posit further that the characteristic volume $\zeta(p_c, r)$ has power-law divergence in r described by the **fractal dimension** exponent d_f :

$$\zeta(p_c, r) \approx r^{d_f}.$$

We think of the exponent $\zeta(p_c, r) \approx r^{d_f}$ as describing the “characteristic volume of a large cluster in $[-r, r]^d$.” Using the same argument as for the relations $\gamma = (2 - \eta)\nu$ and $\delta\gamma = \Delta(\delta - 1)$, we can immediately write down a relation between d_f , η , and δ . Indeed, we can express the expected size of the cluster inside a box both as

$$\mathbb{E}_{p_c} |K \cap [-r, r]^d| = \sum_{x \in [-r, r]^d} \mathbb{P}_{p_c}(0 \leftrightarrow x) \approx \sum_{x \in [-r, r]^d} \|x\|^{-d+2-\eta} \approx r^{2-\eta}$$

and as

$$\mathbb{E}_{p_c} |K \cap [-r, r]^d| = \sum_{n=1}^{\infty} \mathbb{P}_{p_c}(|K \cap [-r, r]^d| \geq n) \approx \sum_{n=1}^{\zeta(p_c, r)} n^{-1/\delta} \approx \zeta(p_c, r)^{(\delta-1)/\delta},$$

and writing this in terms of exponents yields that

$$2 - \eta = d_f \frac{\delta - 1}{\delta}. \quad (10)$$

Another scaling relation for d_f concerns the arm exponent $\mathbb{P}_{p_c}(0 \leftrightarrow \partial[-r, r]^d) \approx r^{-1/\rho}$ and arises from somewhat different reasoning. In this case, we suppose that the best strategy for a cluster to have volume at least n is to have diameter at least r where $\zeta(p_c, r) = n$ and therefore $r \approx n^{1/d_f}$. This yields that

$$n^{-1/\delta} \approx \mathbb{P}_{p_c}(|K| \geq n) \approx \mathbb{P}_{p_c}(0 \leftrightarrow \partial[-\zeta^{-1}(n), \zeta^{-1}(n)]^d) \approx n^{-1/(\rho d_f)}$$

and hence that

$$\delta = \rho d_f, \quad (11)$$

where we write ζ^{-1} for the inverse of the function $r \mapsto \zeta(p_c, r)$.

Note that at this stage we have introduced eight exponents $\gamma, \delta, \beta, \eta, \Delta, \nu, d_f$, and ρ and (non-rigorously) derived five relations between them (4), (6), (8), (10), and (11) leaving us with three degrees of freedom to specify these exponents.

Hyperscaling and the upper-critical dimension. Recall that we think of $\zeta(p_c, r)$ as being the typical volume of a large cluster in the box $[-r, r]^d$. We now want to understand *how many* such clusters exist. We write $N(p_c, r)$ for the (as yet undefined) *number of large clusters* in $[-r, r]^d$, and posit the existence of an exponent $\#$ such that

$$N(p_c, r) \approx r^\#.$$

Whatever the precise definition of $N(p_c, r)$ is, we will want that the moments of the volume of the cluster of the origin can be described accurately (up-to-constants) by $N(p_c, r)$ and $\zeta(p_c, r)$ in the sense that

$$\mathbb{E}_{p_c} |K \cap [-r, r]^d|^k \approx \frac{N(p_c, r) \zeta(p_c, r)}{r^d} \zeta(p_c, r)^k. \quad (12)$$

Here $r^{-d} N(p_c, r) \zeta(p_c, r)$ should be thought of as the probability that the cluster belongs to a characteristic large cluster of scale r (this quantity being the ratio of the number of points in

characteristic large clusters to the volume of the box), while the term $\zeta(p_c, r)^k$ is the contribution to the k th moment from this event. Note that the cases $k = 1, 2$ of this relation can be solved to give two reasonable definitions of $\zeta(p_c, r)$ and $N(p_c, r)$, namely

$$\zeta(p_c, r) = \frac{\mathbb{E}_{p_c}|K \cap [-r, r]^d|^2}{\mathbb{E}_{p_c}|K \cap [-r, r]^d|} \quad \text{and} \quad N(p_c, r) = \frac{r^d \mathbb{E}_{p_c}|K \cap [-r, r]^d|}{\zeta(p_c, r)^2} = \frac{r^d (\mathbb{E}_{p_c}|K \cap [-r, r]^d|)^3}{(\mathbb{E}_{p_c}|K \cap [-r, r]^d|^2)^2}.$$

As before this is certainly not the only reasonable definition of these quantities, and proving the validity of the scaling relations is closely related to proving that different definitions of these quantities are equivalent within constant factors. Writing the case $k = 1$ of (12) in terms of exponents yields the scaling relation

$$2 - \eta = \# + 2d_f - d. \quad (13)$$

Thus, we have introduced one additional exponent and one additional relation, leaving three degrees of freedom in total as before.

The relation (13) is unlike the others we have (non-rigorously) derived so far in that it involves the dimension d and can be used to give our first non-rigorous derivation of the fact that the critical dimension is 6. Taking on trust for now that the critical exponents η and d_f should take the values $\eta = 0$ and $d_f = 4$ when $d > d_c$ (these are the values of these exponents for branching random walk as we will show later), we get that if $d > d_c$ then $\# = d - 6$. Since $\#$ cannot be negative, this already shows (non-rigorously) that $d_c \geq 6$. On the other hand, for $d = 2$ it is not too hard to show using proposition 19 that the number of clusters crossing an $n \times n$ square is tight for all $n \geq 1$, which suggests that $\# = 0$ when $d = 2$. As such, it is reasonable to guess that there is some dimension at which $\#$ becomes positive and stays positive, and it is also reasonable to guess that this critical dimension should coincide with the dimension at which the exponents begin to take their mean-field values. Indeed, the scaling relation (13) is usually stated as the **hyperscaling relation**

$$2 - \eta = 2d_f - d \quad \text{for } d < d_c \quad (14)$$

with the term “hyperscaling relation” referring to an additional scaling relation true only below the upper critical dimension. If this is the case (and one assumes that exponents are continuous, so that at $d = d_c$ the exponents both take their mean-field values and satisfy the hyperscaling relations) then the critical dimension must be $d_c = 6$. That is:

$d_c = 6$ is the unique dimension for which the mean-field exponents satisfy the hyperscaling relation $2 - \eta = 2d_f - d$

Later we will explain why we expect the mean-field values $\eta = 0$, $d_f = 4$ in high dimensions and give two further explanations of why $d_c = 6$ for percolation.

Note that when $d \leq d_c$, the scaling and hyperscaling relations we have stated give six relations for eight exponents, leaving two degrees of freedom. A further hyperscaling relation states that for $d \leq d_c$, the probability that two points at distance r are connected should be roughly the same as both points being connected to distance $r/2$ (say), so that

$$\|x - y\|^{-d+2-\eta} \approx \mathbb{P}_{p_c}(x \leftrightarrow y) \approx \mathbb{P}_{p_c}(0 \leftrightarrow \partial[-\|x - y\|, \|x - y\|]^d)^2 \approx \|x - y\|^{-2/\rho}$$

and hence

$$d - 2 + \eta = \frac{2}{\rho}. \quad (15)$$

Unfortunately this does not really give a new constraint on the exponents since it follows from the scaling relations (10) and (11) and the hyperscaling relation (14).

2.3 Percolation on the binary tree

In this section we analyze critical branching random walk (and the related topic of critical percolation on trees) and explain the heuristic relating this to Bernoulli percolation in high dimensions. We will focus on the binary tree (and random walks with binary branching) to allow for quick explicit computations; extensions to k -regular trees and other branching distributions will be outlined in the exercises.

We begin by noting the following simple exact formula for the susceptibility, yielding in particular that $\gamma = 1$ once we prove that $p_c = 1/2$.

Lemma 30. *Let K be the cluster of the root in Bernoulli- p percolation on the binary tree. Then*

$$\mathbb{E}_p|K| = \begin{cases} \frac{1}{1-2p} & p < 1/2 \\ \infty & p \geq 1/2. \end{cases}$$

Proof of lemma 30. Since there is only one possible open path between any two vertices, the probability that two vertices x and y are connected is equal to $p^{d(x,y)}$, where $d(x,y)$ is the graph distance. Since there are 2^n vertices at distance n from the root, it follows that

$$\mathbb{E}_p|K| = \sum_{n=0}^{\infty} p^n 2^n$$

which is equivalent to the claim. □

Corollary 31. *The critical probability of the binary tree is $1/2$.*

We will give a quick proof from the sharpness of the phase transition. Using this theorem to compute the critical probability of the binary tree is completely unnecessary, and a hands-on proof will be given as an exercise below.

Proof. For the 3-regular tree the same computation yields that $\mathbb{E}_p|K| = \frac{1+p}{1-2p}$ when $p < 1/2$ and $\mathbb{E}_p|K| = \infty$ when $p \geq 1/2$. Since this graph is transitive, it follows from the sharpness of the phase transition that the critical probability of the three-regular tree is $1/2$. Since the probability that the origin is connected to infinity in the 3-regular tree is the same as in the binary tree up to a factor of 3, it follows that $1/2$ is also the critical probability in the binary tree. □

Our next goal is to study the volume tail. In the case of the binary tree the volume distribution admits a fairly explicit formula:

Proposition 32. *Let K be the cluster of the root in Bernoulli- p percolation on the binary tree. Then*

$$\mathbb{P}_p(|K| = n) = p^{n-1}(1-p)^{n+1} \cdot \frac{1}{n+1} \binom{2n}{n}$$

for every $p \in [0, 1]$ and $n \geq 1$.

Proof. Every n -vertex subtree of the binary tree containing the root has $2n$ edges of the binary tree incident to it and $n-1$ edges contained in it, so that the number of edges incident but not contained in it is $n+1$. Thus, every n -vertex subtree containing the root has the same probability to be the cluster of the root, namely $p^{n-1}(1-p)^{n+1}$. To compute the probability that the cluster of the root has n vertices, it suffices therefore to compute the number of n -vertex subtrees of the

binary tree containing the root. Denoting this number by C_n , we observe that C_n satisfies the recurrence

$$C_0 = 1 \quad C_{n+1} = \sum_{k=0}^n C_k C_{n-k}$$

since we can first choose the number of vertices in the subtree descended from each child of the origin before choosing what this subtree is. It is a well-known fact that this recurrence is solved by the **Catalan numbers**

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

which also count many other combinatorial sequences (see https://en.wikipedia.org/wiki/Catalan_number). (The fact that the Catalan numbers solve the recurrence above is perhaps easiest to verify using generating functions: If $G(\lambda) = \sum_{n=0}^{\infty} C_n \lambda^n$ is the generating function of our sequence then the recurrence relation above becomes the quadratic equation $G(\lambda) = 1 + \lambda G(\lambda)^2$. Solving this equation using the quadratic formula then expanding the radical with the generalized Binomial theorem yields the above formula for C_n .) $\square$

Using Stirling's formula, the Catalan number can be expressed asymptotically as

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

so that

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

for some δ_n converging to zero as $n \rightarrow \infty$. This yields in particular that

$$\mathbb{P}_{1/2}(|K| = n) \sim \frac{1}{\sqrt{\pi}} \cdot n^{-3/2} \quad \text{and hence} \quad \mathbb{P}_{1/2}(|K| \geq n) \sim \frac{2}{\sqrt{\pi}} \cdot n^{-1/2}$$

as $n \rightarrow \infty$, and also yields the explicit off-critical estimate

$$\begin{aligned} \mathbb{P}_{1/2+\varepsilon}(|K| = n) &= (1 + \delta_n + O(\varepsilon)) \frac{1}{\sqrt{\pi}} \cdot n^{-3/2} \exp [\log(1 - 4\varepsilon^2)n] \\ &= (1 + \delta_n + O(\varepsilon)) \frac{1}{\sqrt{\pi}} \cdot n^{-3/2} \exp [-(4\varepsilon^2 + O(\varepsilon^4))n]. \end{aligned}$$

Thus, for percolation on the binary tree the exponents δ and Δ are well-defined and satisfy $\delta = \Delta = 2$. It follows in particular that there are no infinite clusters almost surely at the critical point.

Exercise 22. *Prove that percolation on the binary tree satisfies $\mathbb{P}_{1/2+\varepsilon}(|K| = \infty) \asymp \varepsilon$ as $\varepsilon \downarrow 0$, so that the exponent β is well-defined and equal to 1.*

Exercise 23. *We now sketch an alternative analysis of the critical volume tail, which is better suited to more general branching processes. Consider a ghost field of intensity h on the binary tree independent of the percolation configuration, write $\mathbb{P}_{p,h}$ for the joint law of percolation and the ghost field, and let $M(p, h)$ be the probability that the root is connected to the ghost field.*

1. *By conditioning on the number of open edges incident to the root, find a quadratic equation satisfied by $M(p, h)$.*
2. *Use this equation to give an alternative proof that $p_c = 1/2$, being careful to justify how you know which root of the quadratic equation is equal to $M(p, h)$.*

3. Prove that $M(1/2, h)$ is asymptotic to a constant multiple of $\sqrt{h}$ and that $\lim_{h \rightarrow 0} M(1/2 + \varepsilon, h)$ is asymptotic to a constant multiple of ε .
4. Look up Karamata's Tauberian theorem, then use it together with part 3 to compute the asymptotics of $\mathbb{P}_{1/2}(|K| \geq n)$ and $\mathbb{P}_{1/2+\varepsilon}(|K| = \infty)$.

The tail of the radius. We now move on from the cluster volume to instead study the *radius* of critical clusters in the binary tree.

Proposition 33. *For each $r \geq 0$ let $Q(r)$ be the probability that the root is connected to distance r in critical percolation on the binary tree. Then $Q(r) \sim 4r^{-1}$ as $r \rightarrow \infty$.*

Proof of Proposition 33. Conditioning on the number of edges incident to the root yields that

$$Q(r+1) = \frac{1}{2}Q(r) + \frac{1}{4}(1 - (1 - Q(r))^2),$$

where the first term accounts for the probability that the root has exactly one child in its cluster and this child is connected to height r , while the second accounts for the probability that the root is connected to both its children and at least one of these children is connected to height r . This equation can be rearranged to yield that

$$Q(r+1) = Q(r) - \frac{1}{4}Q(r)^2$$

for every $r \geq 1$. We first give some heuristic analyses of this recurrence relation before giving a rigorous argument.

Heuristic 1: Power-law ansatz. Since we are at criticality, it is reasonable to suppose that $Q(r)$ has power-law decay. Let us suppose therefore that $Q(r) \sim Ar^{-a}$ for some $A, a > 0$. Assuming that this asymptotic holds with a small enough error that $Q(r+1) - Q(r) \sim A(r+1)^{-a} - Ar^{-a} \sim aAr^{-a-1}$, we must have that $aAr^{-a-1} \sim \frac{1}{4}A^2r^{-2a}$ and hence that $a = 1$ and $A = 4$.

Heuristic 2: ODE. The difference equation $Q(r+1) - Q(r) = -\frac{1}{4}Q(r)^2$ reminds us of the ODE $f' = -f^2/4$. This ODE can be solved by noting that

$$\left(\frac{1}{f}\right)' = -\frac{f'}{f^2} = \frac{1}{4}.$$

so that

$$\frac{1}{f(x)} = \frac{1}{f(1)} + \frac{x-1}{4}$$

and hence $f(x) \sim 4/x$ as $x \rightarrow \infty$.

Rigorous analysis. We use a similar algebraic manipulation as in our analysis of the analogous ODE, observing that

$$\frac{1}{Q(r+1)} - \frac{1}{Q(r)} = \frac{Q(r) - Q(r+1)}{Q(r)Q(r+1)} = \frac{Q(r)}{4Q(r+1)}$$

where we used the recurrence relation in the second equality. Now, it follows from the recurrence relation and the fact that $Q(r) \rightarrow 0$ as $r \rightarrow \infty$ that $Q(r+1) - Q(r) = o(Q(r))$ as $r \rightarrow \infty$ and hence that $Q(r+1) \sim Q(r)$ as $r \rightarrow \infty$. As such we have that

$$\frac{1}{Q(r+1)} - \frac{1}{Q(r)} \sim \frac{1}{4}$$

as $r \rightarrow \infty$ and hence that

$$\frac{1}{Q(R)} - 1 = \sum_{r=0}^{R-1} \left[\frac{1}{Q(r+1)} - \frac{1}{Q(r)} \right] \sim \frac{R}{4}$$

as $R \rightarrow \infty$. This is easily seen to imply the claim. $\square$

Large critical clusters are two-dimensional. We now explore various senses in which large critical clusters on the binary tree are “two-dimensional”. A hint of this two-dimensionality is already provided by the ratio of the volume exponent $1/2$ to the radius exponent 1 ; we start by sharpening this to show that a critical cluster of volume n typically has radius of order $n^{1/2}$ and vice versa.

Proposition 34. *Consider critical percolation on the binary tree. For each $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\mathbb{P}_{1/2}(|K| \geq n, \text{rad}(K) \leq \delta n^{1/2}) \leq \varepsilon \mathbb{P}_{1/2}(|K| \geq n)$$

for every $n \geq 1$ and

$$\mathbb{P}_{1/2}(\text{rad}(K) \geq r, |K| \leq \delta r^2) \leq \varepsilon \mathbb{P}_{1/2}(\text{rad}(K) \geq r)$$

for every $r \geq 1$.

We write L_r for the set of vertices of the binary tree at distance r from the root and write $L_{r,R} = \cup_{i=r}^R L_i$.

Proof of Proposition 34. For the first claim, we note that for each $n, r \geq 1$ we have that

$$\mathbb{P}_{1/2}(|K| \geq n, \text{rad}(K) \leq r) \leq \mathbb{P}_{1/2}(|K \cap L_{0,r}| \geq n) \leq \frac{1}{n} \mathbb{E}_{1/2} |K \cap L_{0,r}| = \frac{r+1}{n},$$

where in the last line we used that the expected number of vertices in $K \cap L_i$ is 1 for every $i \geq 0$ in critical percolation on the binary tree. This bound easily implies the first claim since $\mathbb{P}_{1/2}(|K| \geq n) \asymp n^{-1/2}$. For the second claim, note that if $\text{rad}(K) \geq 3r$ but $|K| \leq n$ then there must exist a level i_0 between r and $2r$ such that $|K \cap L_{i_0}| \leq n/r$. If we explore the cluster of the origin level-by-level until first finding such a level i_0 , the conditional probability that we reach height $3r$ is at most $\frac{n}{r} \mathbb{P}_{1/2}(\text{rad}(K) \geq r)$ by Markov’s inequality. Since the probability we find the scale i_0 is itself at most $\mathbb{P}_{1/2}(\text{rad}(K) \geq r)$, it follows that

$$\mathbb{P}_{1/2}(\text{rad}(K) \geq 3r, |K| \leq n) \leq \frac{n}{r} \mathbb{P}_{1/2}(\text{rad}(K) \geq r)^2 \preceq \frac{n}{r^3},$$

which again is easily seen to imply the claim since $\mathbb{P}_{1/2}(\text{rad}(K) \geq 3r) \asymp r^{-1}$. $\square$

We now explore another sense in which large critical clusters on the tree are two-dimensional, focussing on the small-scale geometry around the root when the cluster of the root has large radius.

Lemma 35. *Consider critical percolation on the binary tree. We have that*

$$\mathbb{P}_{1/2}(|K \cap L_r| = n \mid \text{rad}(K) \geq R) \sim n \mathbb{P}_{1/2}(|K \cap L_r| = n)$$

as $R \rightarrow \infty$ with $n, r \geq 1$ fixed.

In general, if we take a random variable X with law $\mathbb{P}$ then define a random variable $\hat{X}$ with law $\hat{\mathbb{P}}$ by

$$\hat{\mathbb{E}}F(\hat{X}) = \frac{\mathbb{E}XF(X)}{\mathbb{E}X}$$

then $\hat{X}$ is known as the **size-biasing** of X . Thus, this lemma shows that conditioning the cluster of the root to reach a very high level has the effect of size-biasing the distribution of the cluster of the origin in each constant level.

Proof of Lemma 35. Conditional on what happens in the first r levels, each vertex in $|K \cap L_r|$ is connected to L_R independently at random with probability $Q(R - r)$. Markov's inequality yields that

$$\mathbb{P}_{1/2}(|K \cap L_r| = n, \text{rad}(K) \geq R) \leq nQ(R - r)\mathbb{P}_{1/2}(|K \cap L_r| = n)$$

and hence that

$$\mathbb{P}_{1/2}(|K \cap L_r| = n \mid \text{rad}(K) \geq R) \leq \frac{Q(R - r)}{Q(R)} n \mathbb{P}_{1/2}(|K \cap L_r| = n) \sim n \mathbb{P}_{1/2}(|K \cap L_r| = n)$$

as $R \rightarrow \infty$ by Proposition 33. For the reverse inequality, we have by the Bonferroni inequality (i.e., inclusion-exclusion applied to second order as an inequality) that

$$\mathbb{P}_{1/2}(|K \cap L_r| = n, \text{rad}(K) \geq R) \geq \left[nQ(R - r) - \binom{n}{2} Q(R - r)^2 \right] \mathbb{P}_{1/2}(|K \cap L_r| = n),$$

and taking the limit as $R \rightarrow \infty$ as before yields that $\mathbb{P}_{1/2}(|K \cap L_r| = n \mid \text{rad}(K) \geq R) \sim n \mathbb{P}_{1/2}(|K \cap L_r| = n)$ as claimed. $\square$

Exercise 24. Consider critical percolation on the binary tree. Prove that the probability measures $\mathbb{P}_{1/2}(K \in \cdot \mid \text{rad}(K) \geq r)$ have a well-defined weak limit as $r \rightarrow \infty$, where we think of the cluster as a set of vertices and take limits with respect to the product topology on $\{0, 1\}^V$. You should also prove that this limit measure $\mathbb{P}_{\text{IIC}}$ is determined by the property that

$$\mathbb{P}_{\text{IIC}}(K \cap L_{0,r} \in \cdot) = \mathbb{E}_{1/2} \left[\mathbf{1}(K \cap L_{0,r} \in \cdot) \mid |K \cap L_r| \right]$$

for every $r \geq 1$. This limit is known as the **incipient infinite cluster** (IIC) on the binary tree and was first constructed by Kesten.

The next lemma gives a sense in which the IIC on the binary tree has two-dimensional (i.e., quadratic) volume growth.

Lemma 36. Consider the incipient infinite cluster measure on the binary tree. We have that

$$\mathbb{E}_{\text{IIC}} |K \cap L_r| = \frac{r + 2}{2}$$

for every $r \geq 0$ and hence that

$$\mathbb{E}_{\text{IIC}} |K \cap L_{0,r}| \sim \frac{1}{4} r^2$$

as $r \rightarrow \infty$.

Proof of Lemma 36. It is a general fact about size-biased random variables, immediate from the definition, that the first moment of the size-biased random variable is equal to the ratio of the second and first moments of the original random variable. In our context the first moment is 1, so that

$$\mathbb{E}_{\text{IC}}|K \cap L_r| = \mathbb{E}_{1/2}|K \cap L_r|^2.$$

To compute the second moment, first note that $\mathbb{E}_{1/2}|K \cap L_r|^2$ is equal to the number of ordered pairs $x, y \in L_r$ that both belong to K . To evaluate this expectation, we can consider the sum over $0 \leq i \leq r$, $z \in L_i$, and $x, y \in L_r$ such that x and y are both connected to the root with most recent common ancestor z . This yields that

$$\mathbb{E}_{1/2}|K \cap L_r|^2 = 2^{-r}2^r + \sum_{i=0}^{r-1} 2^{-i-2(r-i)}2^i \cdot 2 \cdot 2^{2(r-i-1)}.$$

Here we have split off the term $z = x = y$ since it has slightly different combinatorics. In the other term, the $2^{-i-2(r-i)}$ accounts for the fact there are $i + 2(r - i)$ edges that must be open for the specified connection to occur, the 2^i accounts for the number of choices of z , the 2 accounts for the fact we must choose whether x is descended from the left or right child of z (the corresponding choice for y being determined by this), and the $2^{2(r-i-1)}$ accounts for the number of ways of choosing x and y once these choices have been made. Simplifying, it follows that

$$\mathbb{E}_{1/2}|K \cap L_r|^2 = 1 + \sum_{i=0}^{r-1} 2^{-1} = \frac{r+2}{2}$$

as claimed. □

Exercise 25. Show that for each $k \geq 1$ there exists a constant C_k such that

$$\mathbb{E}_{\text{IC}}|K \cap L_r|^k \sim C_k r^k$$

as $r \rightarrow \infty$. *Bonus:* Compute the constants C_k , and use this to compute the weak limit of the random variables $r^{-1}|K \cap L_r|$ as $r \rightarrow \infty$.

2.4 Branching random walk

As explained earlier, we expect that when d is sufficiently large, critical percolation on $\mathbb{Z}^d$ begins to behave similarly to critical percolation on the binary tree in some senses. This includes the $n^{-1/2}$ scaling of the volume tail and the ε scaling of the probability the origin belongs to an infinite cluster at $p_c + \varepsilon$. On the other hand, we do not expect the probability the cluster of the origin is connected to distance r in $\mathbb{Z}^d$ to decay like r^{-1} as it does in the tree. Meanwhile, for the two-point function we expect to have polynomial decay on $\mathbb{Z}^d$ whereas the critical two-point function decays exponentially on the tree.

A better toy model for critical percolation in high dimensions is **branching random walk**. Here, branching random walk is a Markov process taking values in the set of finite configurations of particles on $\mathbb{Z}^d$ in which at each step every particle splits independently into a random number of new particles with some fixed offspring distribution, and each new particle simultaneously performs a step of simple random walk. (That is, offspring particles are located at random neighbours of their parents rather than in the same location as their parent.) Branching random walk can also be defined equivalently by first sampling the genealogical tree of the particles as a branching process, then picking a uniform random graph homomorphism from this tree into $\mathbb{Z}^d$; in the case where the offspring distribution is Binomial(2, p) this genealogical tree has the same distribution as the cluster of the root in percolation on the binary tree.

Proposition 37. *Consider critical branching random walk on $\mathbb{Z}^d$. The expected number of particles that visit x when we begin with a single particle at the origin is equal to the simple random walk Greens function $G(0, x) = \sum_{n=0} p_n(0, x)$.*

Proof of Proposition 37. The expected number of particles in generation n is exactly 1. Conditional on the genealogical tree, each particle in generation n is distributed as a simple random walk run for n steps, and the claim follows by linearity of expectation. $\square$

Proposition 38. *Consider critical branching random walk on $\mathbb{Z}^d$ with Binomial(2, 1/2) offspring distribution, where we start with a single particle at the origin. If $d > 4$ then there exists a positive constant c such that the probability x is ever visited by a particle is between $cG(0, x)$ and $G(0, x)$.*

We will use the fact that the simple random walk Greens function on $\mathbb{Z}^d$ satisfies $G(x, y) \asymp \langle x - y \rangle^{-d+2}$ when $d \geq 3$ and where we write $\langle x \rangle = \max\{\|x\|, 2\}$ to avoid division by zero.

Proof of Proposition 38. The upper bound follows from proposition 37 and Markov's inequality. For the lower bound we will estimate the second moment of the number of particles that visit x . We need to count the number of pairs of vertices u, v in the binary tree that are both connected to the root of the tree and are both mapped to x by the random embedding of the binary tree into $\mathbb{Z}^d$. Given two vertices u and v of the binary tree, write $u \wedge v$ for their most recent common ancestor, i.e., the last place where the paths from the root to these vertices coincide. If we write $|u|$ for the height of u in the tree, the probability that u and v are both connected to the root is

$$2^{-|u \wedge v| - (|u| - |u \wedge v|) - (|v| - |u \wedge v|)} = 2^{-|u| - |v| + |u \wedge v|}.$$

Given u and v are both in the cluster of the origin, the probability they are both mapped to x by the embedding is

$$\sum_{y \in \mathbb{Z}^d} p_{|u \wedge v|}(0, y) p_{|u| - |u \wedge v|}(y, x) p_{|v| - |u \wedge v|}(y, x),$$

where $p_{|u \wedge v|}(0, y)$ is the probability $u \wedge v$ is mapped to y and $p_{|u| - |u \wedge v|}(y, x) p_{|v| - |u \wedge v|}(y, x)$ is the probability u and v are both mapped to x conditional on $u \wedge v$ being mapped to y . Thus, the second moment of the number of particles that visit x is

$$\sum_w \sum_{\substack{u, v \\ u \wedge v = w}} \sum_{y \in \mathbb{Z}^d} 2^{-|w|} p_{|u \wedge v|}(0, y) 2^{-(|u| - |w|)} p_{|u| - |w|}(y, x) 2^{-(|v| - |w|)} p_{|v| - |w|}(y, x).$$

As before, we can compute this sum by first choosing whether u or v lies to the left of w , the summing over which height above w they lie in. This yields that

$$\begin{aligned} \sum_{\substack{u, v \\ u \wedge v = w}} \sum_{y \in \mathbb{Z}^d} 2^{-|w|} p_{|u \wedge v|}(0, y) 2^{-(|u| - |w|)} p_{|u| - |w|}(y, x) 2^{-(|v| - |w|)} p_{|v| - |w|}(y, x) \\ = \frac{1}{2} \sum_{r=1}^{\infty} 2^r 2^{-r} p_r(y, x) \sum_{\ell=1}^{\infty} 2^\ell 2^{-\ell} p_\ell(y, x) = \frac{1}{2} G(y, x) \end{aligned}$$

and hence by a similar computation that our second moment is

$$\sum_w \sum_{y \in \mathbb{Z}^d} 2^{-|w|} p_{|u \wedge v|}(0, y) \frac{1}{2} G(y, x)^2 = \frac{1}{2} \sum_{y \in \mathbb{Z}^d} G(0, y) G(y, x)^2.$$

Using that

$$G(0, x) \asymp \langle x - y \rangle^{-d+2}$$

it follows that the second moment is of order

$$\sum_{y \in \mathbb{Z}^d} \langle y \rangle^{-d+2} \langle x - y \rangle^{-2d+4}.$$

To estimate this sum, we can split into cases according to whether $\|y\| \leq \|x\|/2$, $\|x - y\| \leq \|x\|/2$, or otherwise. In the first case $\langle x - y \rangle$ is of the same order as $\langle x \rangle$, in the second case $\langle y \rangle$ is of the same order as $\langle x \rangle$, and in the third case $\langle x - y \rangle$ is of the same order as $\langle y \rangle$. This, the second moment is bounded by a constant multiple of

$$\sum_{\|y\| \leq \|x\|/2} \langle y \rangle^{-d+2} \langle x \rangle^{-2d+4} + \sum_{\|y-x\| \leq \|x\|/2} \langle x \rangle^{-d+2} \langle x - y \rangle^{-2d+4} + \sum_{\|y\|, \|y-x\| > \|x\|/2} \langle y \rangle^{-3d+6}.$$

When $d > 4$ the sum $\sum_{y \in \mathbb{Z}^d} \langle y \rangle^{-2d+4}$ is finite and we obtain that the second moment is bounded by a constant multiple of

$$\langle x \rangle^{-2d+6} + \langle x \rangle^{-d+2} + \langle x \rangle^{-2d+6} \preceq \langle x \rangle^{-d+2}.$$

That is, when $d > 4$ the second moment of the number of particles that visit x is of the same order as the first moment, namely $\langle x \rangle^{-d+2}$. Note that if X is any non-negative random variable then we have by Cauchy-Schwarz that

$$\mathbb{E}X = \mathbb{E}X \mathbf{1}(X > 0) \leq \sqrt{\mathbb{P}(X > 0) \mathbb{E}X^2} \quad \text{so that} \quad \mathbb{P}(X > 0) \geq \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2},$$

and applying this inequality in our case yields the desired lower bound. $\square$

Remark 39. When $d = 3$, this argument bounds the second moment of the total number of particles that visit x by $\langle x \rangle^{-2d+6}$, yielding that the probability x is ever visited is at least $\langle x \rangle^{-2}$. This bound is of the correct order, with the matching upper bound following from the one-arm estimate proven below. In dimension $d = 4$ the bound we get on the second moment and hitting probability are $\langle x \rangle^{-2} \log \langle x \rangle$ and $\langle x \rangle^{-2} (\log \langle x \rangle)^{-1}$ respectively, which are also of the correct order (but this is harder to justify).

We next compute the one-arm probability for critical branching random walk.

Proposition 40. *Consider critical branching random walk on $\mathbb{Z}^d$ with Binomial(2, 1/2) offspring distribution. The probability any particle ever leaves the box $[-r, r]^d$ is of order r^{-2} .*

Proof of Proposition 40. Let $A(r)$ denote the relevant probability. We begin with the lower bound. Let K be the cluster of the origin in percolation on the binary tree, and let $X : K \rightarrow \mathbb{Z}^d$ be the random embedding of K into $\mathbb{Z}^d$. In order for the cluster of the origin to be large we must either have that it has a large number of points mapped into $[-r, r]^d$ or that the arm event occurs, so that

$$\mathbb{P}(|K| \geq n) \leq \mathbb{P}(|X^{-1}([-r, r]^d)| \geq n) + A(r)$$

for every $n, r \geq 1$. Using Markov's inequality, the probability $\mathbb{P}(|X^{-1}([-r, r]^d)| \geq n)$ is bounded by $\frac{1}{n} \sum_{x \in [-r, r]^d} G(0, x) \asymp n^{-1} r^2$, and since we also have that $\mathbb{P}(|K| \geq n) \asymp n^{-1/2}$ it follows that there exist positive constants c and C such that

$$\frac{c}{\sqrt{n}} \leq \frac{Cr^2}{n} + A(r)$$

for every $n, r \geq 1$. The claimed lower bound follows by taking r to be a small multiple of $n^{1/4}$ and rearranging.

We now prove the upper bound. We consider the union bound

$$A(2r) \leq \mathbb{P}(X^{-1}(\partial[-2r, 2r]^d) \neq \emptyset, \text{rad}(K) \leq \ell) + \mathbb{P}(\text{rad}(K) \geq \ell)$$

for $r, \ell \geq 1$, where $\text{rad}(K)$ continues to denote the radius of K as a subset of the binary tree. Observe that if K has radius at most ℓ but the BRW reaches Euclidean distance at least $2r$, there must exist a particle of the BRW within the first ℓ generations that belongs to the boundary of the box $[-r, r]^d$, never had an ancestor that left this box, and has a descendant that reaches the boundary of the box $[-2r, 2r]^d$. Considering the expected number of such particles and applying Markov yields that if we define $\mathcal{P}(r, \ell)$ to be the set of vertices of that belong to $\partial[-r, r]^d$, did not have an ancestor that left the box, and belong to the first ℓ generations then

$$\mathbb{P}(X^{-1}(\partial[-2r, 2r]^d) \neq \emptyset, \text{rad}(K) \leq \ell) \leq A(r) \mathbb{E}|\mathcal{P}(r, \ell)|.$$

Now, if we let $\mathbf{P}_0$ denote the law of a simple random walk $(Y_i)_{i \geq 0}$ started at the origin, we have by linearity of expectation as before that

$$\begin{aligned} \mathbb{E}|\mathcal{P}(r, \ell)| &= \sum_{i=0}^{\ell} \mathbf{P}_0(Y_i \in \partial[-r, r]^d, Y_j \in [-r, r]^d \text{ for all } j \leq i) \\ &\preceq \mathbf{P}_0(\text{hit } \partial[-r, r]^d \text{ before time } \ell), \end{aligned}$$

where in the inequality we used that every time the walk is in the boundary of the box it has a constant probability to leave the box. Now, it is a standard fact about simple random walk on $\mathbb{Z}^d$ that

$$\mathbf{P}_0(\max_{i \leq \ell} \|Y_i\| \geq r) \preceq \exp \left[-\frac{c' r^2}{\ell} \right]$$

for some positive constant c' (see e.g. the Azuma-Hoeffding inequality). As such, there exists a constant δ such that if $\ell \leq \delta r^2$ then $\mathbb{E}|\mathcal{P}(r, \ell)| \leq \frac{1}{8}$. Taking $\ell = \lfloor \delta r^2 \rfloor$ and using that $\mathbb{P}(\text{rad}(K) \geq \ell) \asymp \ell^{-1}$, we deduce that there exists a constant C such that

$$A(2r) \leq \frac{1}{8} A(r) + C r^{-2}.$$

This inequality can be applied recursively to yield that

$$A(2^{k+1}) \leq C 4^{-k} + C 8^{-1} 4^{-(k-1)} + C 8^{-2} 4^{-(k-2)} + \dots \preceq 4^{-k}$$

as required. (The only important feature of the constant $1/8$ is that it is strictly smaller than $1/4$.) $\square$

Exercise 26. State and prove an analogue of Proposition 34 stating that critical branching random walk on $\mathbb{Z}^d$ is unlikely to reach Euclidean distance r while having total particle count much smaller than r^4 , and conversely that it is unlikely to have total particle count at least n without reaching Euclidean distance $n^{1/4}$.

Proposition 41. Consider the IIC version of BRW with Binomial(2, 1/2) branching, in which we embed the IIC on the binary tree into $\mathbb{Z}^d$ via a uniform random homomorphism. Prove that the expected number of particles that visit the site x is of order $\langle x \rangle^{-d+4}$ when $d > 4$. Bonus: Prove that the probability x is visited by at least one particle is also of this order when $d > 4$.

Exercise 27. Using that the simple random walk on $\mathbb{Z}^d$ satisfies

$$cn^{-d/2}e^{-Cd(x,y)^2/n} \leq p_n(x,y) \leq Cn^{-d/2}e^{-cd(x,y)^2/n}$$

for some positive constants c and C , prove that if we consider a branching random walk on $\mathbb{Z}^d$ with mean offspring $1 - m < 1$ started with a single particle at the origin and define $G_m(0, x)$ to be the expected number of particles that ever visit a site x then there exist positive constants c' and C' (not depending on m) such that

$$c'\|x\|^{-d+2}e^{-C'm^{1/2}\|x\|} \leq G_m(0, x) \leq C'\|x\|^{-d+2}e^{-c'm^{1/2}\|x\|}$$

for every $x \neq 0$.

3 Rigorous relationships between critical exponents

3.1 Scaling and hyperscaling in the plane

3.2 The BK inequality

3.3 Tree-graph inequalities and the mean-field lower bound on the volume tail

3.4 The triangle condition

3.5 Equivalence of mean-field critical exponents

3.6 Critical exponents for intrinsic geometry

3.7 Hyperscaling inequalities

4 Long-range and hierarchical models

5 Homework sets

5.1 Homework 1

Exercise 28. The standard monotone coupling of the product measures $\mathbb{P}_p$ on $\{0, 1\}^E$ defines a measure on $[0, 1]$ labels $(U_e)_{e \in E}$ such that $\omega_p(e) := \mathbb{1}(U_e \leq p)$ has independent coordinates for each $p \in [0, 1]$ and $\mathbb{P}(\omega_p(e) = 1) = p$ for each $p \in [0, 1]$. Do there exist other collections $(U_e)_{e \in E}$ with this property for which the different labels U_e are not independent?

Exercise 29. Prove that $p_c(\mathbb{Z}^{d+1}) < p_c(\mathbb{Z}^d)$ for every $d \geq 1$.

Exercise 30. The 7-regular triangulation is a planar graph in which every face is a triangle (i.e., has three sides) and every vertex has degree 7. Prove that the 7-regular triangulation is nonamenable. Hint: Use Euler's formula.

Exercise 31. Consider critical percolation on $\mathbb{Z}^d$ with $d \geq 2$. Prove that the probability that there exists a cluster intersecting both $[-r, r]^d$ and $\mathbb{Z}^d \setminus [2r, 2r]^d$ does not converge to zero as $r \rightarrow \infty$.

Exercise 32. Give an example of a connected graph and a percolation process on it such that the number of infinite clusters is random. Bonus: Give an example of an automorphism-invariant random subgraph of $\mathbb{Z}^2$ with exactly k infinite clusters almost surely.