PERCOLATION AND $\theta(p_c) = 0$

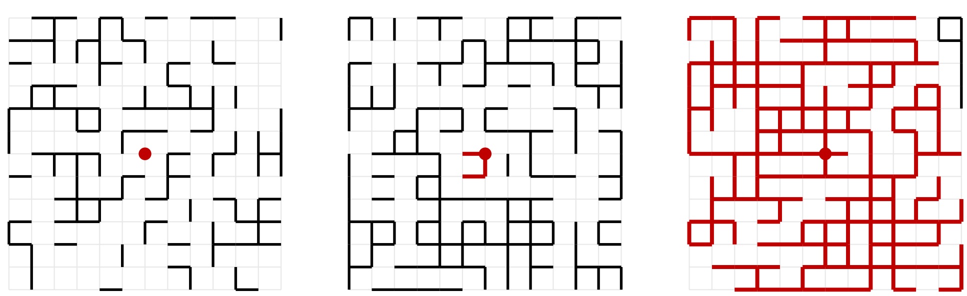
Abstract. These notes prove that the percolation probability vanishes at the critical probability for bond percolation on $\mathbb{Z}^d$ when $d \geq 2$. They also prove the finite labeled-hyperedge inequality needed for site percolation. A companion manuscript supplies the site-specific block and slab argument; together, the two works establish the site result for $d \geq 3$. The planar site result is classical and is discussed separately. It is not part of the site-percolation declaration in the accompanying Lean development. For bond percolation, classical methods cover the plane and, through the lace expansion, dimensions $d \geq 11$; the argument developed here uses neither planar duality nor the lace expansion. Anthropic's earlier Lean development proved and formalized an additive bound, and hence a uniform near-one bound, sufficient for the bond-percolation reduction. The present work extends those conditional covariance inequalities to labeled hyperedges, derives the multiplicative finite inequality in that setting, and provides the finite input for site percolation. The prerequisite is a first graduate course in probability. For the site model, the accompanying Lean declaration has hypotheses $d : \mathbb{N}$ and $3 \leq d$; its printed axioms are `propext`, `Classical.choice`, and `Quot.sound`.

Generated by ChatGPT 5.6, prompted by Ahmed Bou-Rabee, with exposition and formalization support by Claude Opus 5. The earlier bond-percolation proof is due to Anthropic's Claude.

1. Bernoulli Percolation on the Hypercubic Lattice

Broadbent and Hammersley [1957] introduced percolation in 1957 to model a fluid spreading through a porous stone, and posed the existence of the phase transition as the central question. Declare each edge of the lattice $\mathbb{Z}^d$ open with probability p and closed otherwise, independently across edges, and ask whether fluid entering at the origin can travel arbitrarily far. Write $\theta(p)$ for the probability that it can, that is, for the probability that the origin lies in an infinite connected component of open edges. As p increases more edges are open, so θ increases; it vanishes when p is

Date: 25 September 2026.



$p = 0.35$ $|\mathcal{C}_{\mathbf{0}}| = 1$ $p = 0.50$ $|\mathcal{C}_{\mathbf{0}}| = 4$ $p = 0.65$ $|\mathcal{C}_{\mathbf{0}}| = 153$

Figure 1. Samples of Bernoulli bond percolation in a 12×12 window of $\mathbb{Z}^2$ at three parameters, drawn from one seed. The open cluster of the marked origin is thickened, and its size within the window is printed underneath. At $p = 0.35$ the cluster is a small island, and at $p = 0.65$ it already spans the window. The middle picture, at the critical probability $p_c(2) = 1/2$, is the one these notes are about, and the eye cannot tell from it whether the cluster would continue forever.

small and it is positive when p is close to 1. There is therefore a threshold parameter p_c separating the two behaviors, and what happens exactly at that threshold is the subject of these notes.

This section defines the model and establishes two facts. The first is that the threshold is a genuine interior point: for every $d \geq 2$ we have $0 < p_c < 1$, so there really are two phases to separate. The second is that θ is continuous at every parameter other than p_c. The question of whether θ is continuous therefore collapses to the single number $\theta(p_c)$, which is either 0 or the size of a jump. That is why the problem is usually posed as the equation $\theta(p_c) = 0$.

The order is as follows. Section 1.1 introduces the labeled hyperedge framework shared by the finite bond and site arguments and explains its bond-percolation specialization. Section 1.2 constructs the probability space and the class of events used throughout. Section 1.3 introduces open clusters and the percolation probability. Section 1.4 proves that the threshold is an interior point, and Section 1.5 proves continuity away from it. Section 1.6 summarizes the classical dimension-dependent results for $\theta(p_c)$.

Before the results presented here, the bond-percolation cases $3 \leq d \leq 10$ were not covered by either classical method. The planar arguments use duality, which has no analogue above two dimensions, while the high-dimensional arguments use random-walk estimates available through the lace expansion. The approach developed here instead derives a finite connection inequality without either tool; lattice geometry enters only through the final reduction.

We fix three names. The number p is the **parameter**, the function θ is the **percolation probability**, and the threshold p_c is the **critical probability**.

Throughout this section $d \geq 2$ is an integer and p is a number in $[0, 1]$.

Contents

1.1. **Finite hyperedge models for bond and site percolation.** These notes treat both bond and site percolation. Their common finite argument is formulated for labeled hyperedges and then specialized to the two models.

In **site percolation** the vertices, not the edges, carry the randomness: each vertex of $\mathbb{Z}^d$ is open with probability p and closed otherwise, independently, and two open vertices are joined when they are neighbors in the lattice. The cluster of the origin is the set of open vertices reachable from it through open vertices, the percolation probability and the critical probability are defined from it exactly as before, and the same question is asked. Site percolation is not, in general, representable as independent bond percolation. Gladkov and Zimin [2026] gave an obstruction that already appears at a vertex of degree three. Consequently, a bond-only argument does not automatically imply the corresponding site statement.

The shared finite structure consists of independent random objects whose activation creates connections monotonically. Labeled hyperedges represent this structure while allowing one object to meet more than two vertices. Conditioning arguments in this model must retain the labels of hyperedges met by an explored cluster; Sections 2.4 and 2.6 develop the required bookkeeping.

So Sections 2 to 5 are carried out in that model from the start: a finite vertex set, and finitely many elements, each meeting an arbitrary set of vertices and each open or closed independently of the others. Bond percolation is the case in which every element meets exactly two vertices. Site percolation is the case obtained by placing one element at each vertex and one extra vertex on each edge, as Section 6.1 explains. The finite inequality proved in Section 5 therefore applies to both models. Section 6 gives the two specializations, records the bond lattice reduction, and summarizes the separate site reduction.

For the finite bond-percolation specialization, one may usually read "edge" for "hyperedge" and "the two endpoints of the edge" for "incidence set." The exceptions are listed below.

(1) In Definition 2.7 the substitution is exactly the one just described, and the model becomes bond percolation on a finite graph.
(2) The set $\mathcal{K}_S$ of Definition 2.8 becomes the set of open edges meeting the cluster. It cannot be replaced by the set of vertices of the cluster, and that is not an artifact of the general model: the example proving it, Proposition 2.9, is a path on three vertices.
(3) The trace of (5) becomes the set of endpoints outside Z of the open edges meeting Z. The second inclusion of Lemma 2.10 can be strict here too, as soon as Z has two vertices with a common neighbor, so the bookkeeping it forces is not an artifact of the general model either.
(4) Sections 3 to 5 specialize directly to statements about finite graphs.
(5) Of Section 6, only the first sentence of Example 6.1 and the bond half of Section 6.2 are used. The encoding of site percolation and the discussion of it may be skipped.

1.2. **Configurations, events, and the product measure.** We construct the probability space carefully, because every later argument conditions on events and compares measures, and those operations need a fixed underlying space.

Regard $\mathbb{Z}^d$ as a graph: its vertices are the points of $\mathbb{Z}^d$, and two vertices are joined by an edge exactly when they are at Euclidean distance 1. Write $\mathbb{E}_d$ for the set of edges. This edge set is countably infinite, and each vertex lies in exactly $2d$ edges.

Definition 1.1 (Configuration) A **configuration** is a function $\omega : \mathbb{E}_d \to \{0,1\}$. An edge e is **open** in ω when $\omega(e) = 1$, and **closed** when $\omega(e) = 0$. The set of all configurations is $\Omega_d = \{0,1\}^{\mathbb{E}_d}$.

An event is **determined by** a set S of edges when membership in it depends on ω only through the values $\omega(e)$ for $e \in S$.

Definition 1.2 (Cylinder Event) A **cylinder event** is a subset of Ω_d determined by some finite set of edges. The **product σ-algebra** $\mathcal{F}_d$ is the smallest σ-algebra containing every cylinder event.

Equivalently, a cylinder event is a finite union of sets of the form $\{\omega : \omega(e_i) = a_i \text{ for } 1 \le i \le n\}$, since there are only finitely many assignments to a finite set of edges. The events we care about, such as the event that the origin lies in an infinite open cluster, are not cylinder events.

Definition 1.3 (Bernoulli Bond Percolation) **Bernoulli bond percolation** with parameter p on $\mathbb{Z}^d$ is the probability measure $\mathbb{P}_p$ on $(\Omega_d, \mathcal{F}_d)$ under which the coordinates $\omega(e)$ for $e \in \mathbb{E}_d$ are independent and $\mathbb{P}_p(\omega(e) = 1) = p$ for every edge e.

Existence and uniqueness of $\mathbb{P}_p$ is the Kolmogorov extension theorem applied to a product of countably many copies of the Bernoulli measure on $\{0,1\}$, so we take the measure as given. Write $\mathbb{E}_p$ for the corresponding expectation.

One structural feature of Ω_d drives nearly every argument in these notes. Order configurations pointwise: write $\omega \le \eta$ when $\omega(e) \le \eta(e)$ for every edge e, that is, when every edge open in ω is also open in η.

Definition 1.4 (Increasing Event) An event $A \in \mathcal{F}_d$ is **increasing** when $\omega \in A$ and $\omega \le \eta$ together imply $\eta \in A$. The event A is **decreasing** when its complement is increasing.

Opening more edges can create connections but never destroy them, so every event asserting the existence of an open path is increasing.

The most useful device for comparing two parameters is a coupling that realizes all of them at once on a single probability space.

Definition 1.5 (Standard Coupling) Let $(U_e)_{e \in \mathbb{E}_d}$ be independent random variables, each uniform on $[0,1]$, defined on a probability space with measure $\mathbb{P}$. For $p \in [0,1]$ define a configuration ω_p by setting $\omega_p(e) = 1$ exactly when $U_e \le p$. The family $(\omega_p)_{p \in [0,1]}$ is the **standard coupling**.

Lemma 1.6. (The Standard Coupling Has the Right Laws and Is Monotone) *Let $(\omega_p)_{p\in[0,1]}$ be the standard coupling. Then the following hold.*

- *(i) For every $p \in [0,1]$, the configuration ω_p has law $\mathbb{P}_p$.*
- *(ii) If $p \le q$, then $\omega_p \le \omega_q$ pointwise.*

Proof. Fix p. The variables U_e are independent, so the events $\{U_e \le p\}$ are independent, and each has probability p because U_e is uniform on $[0,1]$. Hence the coordinates of ω_p are independent with $\mathbb{P}(\omega_p(e) = 1) = p$, which is the defining property of $\mathbb{P}_p$ in Definition 1.3.

For the monotonicity, let $p \le q$ and let e be an edge with $\omega_p(e) = 1$. Then $U_e \le p \le q$, so $\omega_q(e) = 1$. Since e was an arbitrary edge, $\omega_p \le \omega_q$. $\square$

A consequence of Lemma 1.6 is that $\mathbb{P}_p(A) \le \mathbb{P}_q(A)$ for every increasing event A and all $p \le q$: realize both parameters in the standard coupling and note that $\omega_p \in A$ forces $\omega_q \in A$. We use this comparison from now on without further comment.

1.3. **Open clusters and the percolation probability.** An **open path** is a finite sequence of vertices $x_0, x_1, \ldots, x_n$ in which consecutive vertices are joined by an edge and every one of those edges is open. Two vertices x and y are **connected**, written $x \leftrightarrow y$, when some open path starts at x and ends at y; every vertex is connected to itself by the path of length 0. Connection is reflexive, symmetric, and transitive, so it partitions the vertex set.

Definition 1.7 (Open Cluster) Let x be a vertex of $\mathbb{Z}^d$. The **open cluster** of x is the set $\mathcal{C}_x = \{\, y \in \mathbb{Z}^d : x \leftrightarrow y \,\}$ of vertices connected to x.

Definition 1.8 (Percolation Probability) The **percolation probability** is

$$\theta(p) := \mathbb{P}_p\big(|\mathcal{C}_{\mathbf{0}}| = \infty\big)\,,$$

the probability that the open cluster of the origin is infinite.

For each $n \geq 1$ let A_n denote the event that the origin is connected to some vertex at Euclidean distance at least n from it. Each A_n is determined by the finitely many edges meeting the ball of radius n, so each A_n is a cylinder event and in particular lies in $\mathcal{F}_d$. The events A_n decrease in n, and their intersection is the event that $\mathcal{C}_{\mathbf{0}}$ is infinite, since a cluster is infinite exactly when it contains vertices at arbitrarily large distance. Therefore

$$\theta(p) = \mathbb{P}_p\Big(\bigcap_{n\geq 1} A_n\Big) = \lim_{n\to\infty} \mathbb{P}_p(A_n) = \inf_{n\geq 1} \mathbb{P}_p(A_n)\,, \tag{1}$$

where the second equality is continuity of the measure along a decreasing sequence of events and the third holds because a nonincreasing sequence of numbers converges to its infimum. Identity (1) is used twice below, once for monotonicity and once for continuity.

Lemma 1.9. (The Percolation Probability Is Nondecreasing)

(i) The function θ is nondecreasing on $[0, 1]$.

(ii) The value $\theta(0)$ equals 0.

Proof. Each event A_n asserts the existence of an open path, so each A_n is increasing, and therefore $\mathbb{P}_p(A_n) \leq \mathbb{P}_q(A_n)$ whenever $p \leq q$ by Lemma 1.6. Taking the infimum over n in (1) preserves the inequality, so $\theta(p) \leq \theta(q)$.

At $p = 0$ no edge is open, so $\mathcal{C}_{\mathbf{0}} = \{\mathbf{0}\}$ is finite, and therefore $\theta(0) = 0$. □

Since θ is nondecreasing and vanishes at 0, the set of parameters at which it vanishes is an interval containing 0. Its right endpoint is the critical probability defined next.

Definition 1.10 (Critical Probability) The **critical probability** is

$$p_{\mathrm{c}} = p_{\mathrm{c}}(d) := \inf\{p \in [0, 1] : \theta(p) > 0\}\,,$$

with the convention that the infimum of the empty set equals 1.

By Lemma 1.9 we have $\theta(p) = 0$ for every $p < p_{\mathrm{c}}$ and $\theta(p) > 0$ for every $p > p_{\mathrm{c}}$. The parameter p_{c} itself is covered by neither statement, and that single value is what these notes are about.

1.4. **The critical probability is strictly between zero and one.** For the critical probability to separate two phases it must avoid both endpoints of the parameter interval. Both bounds come from counting paths, and the two counts run in opposite directions: to rule out an infinite cluster we count open paths leaving the origin, and to produce one we count closed circuits that would have to surround it.

Lemma 1.11. (No Infinite Cluster at Small Parameters) *Let $p < 1/(2d-1)$. Then* $\theta(p) = 0$.

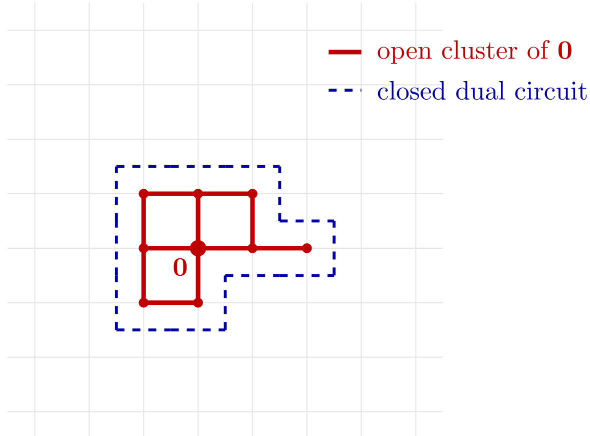


FIGURE 2. Planar duality, as used in Step 2 of the proof of Lemma 1.12. The open cluster of the origin is thickened, and it is finite exactly when a circuit of closed dual edges, drawn dashed, surrounds the origin. Each dual edge crosses exactly one edge of the lattice and carries its state, so a closed dual circuit blocks every open path out of the region it encloses.

Proof. The strategy is to bound the percolation probability by the expected number of long open paths from the origin, and then to show that this expectation tends to 0. We divide the proof into three steps: counting the available paths, bounding the expected number of open ones, and comparing with $\theta(p)$.

Step 1: Count the available paths. Call a path **self-avoiding** when its vertices are distinct. The number of self-avoiding paths of length n starting at the origin is at most $2d(2d-1)^{n-1}$, because the first step has at most $2d$ choices and each later step has at most $2d-1$, the step back along the edge just used being excluded.

Step 2: Bound the expected number of open ones. A fixed self-avoiding path of length n uses n distinct edges, so it is open with probability p^n by independence. Summing over the paths counted in Step 1, the expected number of open self-avoiding paths of length n from the origin is at most $2d(2d-1)^{n-1}p^n$.

Step 3: Compare with the percolation probability. If $\mathcal{C}_{\mathbf{0}}$ is infinite then for every n it contains a vertex at distance at least n, and a shortest open path from the origin to such a vertex is self-avoiding of length at least n. Hence the event $\{|\mathcal{C}_{\mathbf{0}}| = \infty\}$ is contained in the event that at least one open self-avoiding path of length n from the origin exists, and Markov's inequality applied to the count of Step 2 gives

$$\theta(p) \le 2d(2d-1)^{n-1}p^n = \frac{2d}{2d-1}\big((2d-1)p\big)^n .$$

When $(2d-1)p < 1$ the right side tends to 0 as n grows, while the left side does not depend on n, so $\theta(p) = 0$. This proves the lemma. ☐

The other bound uses planar duality, which we describe before using it. Place a dual vertex at the center of each unit square of $\mathbb{Z}^2$, and join two dual vertices by a **dual edge** exactly when their squares share a side. Each edge of $\mathbb{Z}^2$ is crossed by exactly one dual edge, and we declare a dual edge open or closed according to the state of the edge it crosses. A **dual circuit** is a self-avoiding closed path in the dual lattice.

Lemma 1.12. (An Infinite Cluster Exists at Large Parameters) *There is a parameter $p < 1$ with $\theta(p) > 0$.*

Proof. We reduce to the plane and then count dual circuits. The proof has four steps: the reduction, the conversion of finiteness into a dual circuit, the count, and the choice of parameter.

Step 1: Reduce to two dimensions. The subgraph of $\mathbb{Z}^d$ induced by the vertex set $\mathbb{Z}^2 \times \{\mathbf{0}\}$ is a copy of $\mathbb{Z}^2$. An infinite open cluster of this subgraph is contained in an infinite open cluster of $\mathbb{Z}^d$, and the restriction of $\mathbb{P}_p$ to the edges of the subgraph is Bernoulli bond percolation on $\mathbb{Z}^2$ with the same parameter. So it is enough to find $p < 1$ with $\theta(p) > 0$ when $d = 2$.

Step 2: Convert finiteness into a dual circuit. Work in $\mathbb{Z}^2$ and suppose $\mathcal{C}_\mathbf{0}$ is finite. Every edge joining $\mathcal{C}_\mathbf{0}$ to its complement is closed, so the dual edges crossing those edges are all closed. Let W be the union of $\mathcal{C}_\mathbf{0}$ with the bounded components of its complement, a finite set with connected complement, and consider the edges joining W to its complement. The dual edges crossing them form a self-avoiding closed dual circuit surrounding the origin, which is the standard planar duality statement illustrated in Figure 2. Passing to W matters: the dual edges crossing the full edge boundary of $\mathcal{C}_\mathbf{0}$ form in general several circuits, one around each hole, and it is the outermost one that is wanted. Hence the event that $\mathcal{C}_\mathbf{0}$ is finite is contained in the event that some closed dual circuit surrounds the origin.

Step 3: Count the dual circuits. A dual circuit of length n that surrounds the origin must cross the positive horizontal axis, and every vertex of the circuit lies within distance n of the origin, so the crossing occurs at one of at most n dual edges on that axis. Starting from such an edge, the circuit is a self-avoiding dual path with at most three choices at each later step, so the number of dual circuits of length n surrounding the origin is at most $n\,3^{n-1}$. Every such circuit has length at least 4.

Step 4: Choose the parameter. A fixed dual circuit of length n is closed with probability $(1 - p)^n$, so by Step 3 the expected number of closed dual circuits surrounding the origin is at most $\sum_{n \geq 4} n\,3^{n-1}(1-p)^n$. This series converges when $3(1-p) < 1$, and its value tends to 0 as p tends to 1, so we may fix $p < 1$ for which the value is less than 1. For that p, Markov's inequality shows the probability that some closed dual circuit surrounds the origin is less than 1, so by Step 2 the probability that $\mathcal{C}_\mathbf{0}$ is infinite is positive. This proves the lemma. $\square$

Theorem 1.13. (The Critical Probability Is an Interior Point) *We have* $0 < p_c(d) < 1$.

Proof. Lemma 1.11 gives $\theta(p) = 0$ for every $p < 1/(2d-1)$, so no parameter below $1/(2d-1)$ belongs to the set appearing in Definition 1.10, and therefore $p_c \geq 1/(2d-1) > 0$. Lemma 1.12 produces a parameter $p < 1$ with $\theta(p) > 0$, and that parameter does belong to the set, so $p_c \leq p < 1$. $\square$

In other words, both phases are genuinely present, and the parameter interval really does split into a region where every open cluster is finite and a region where an infinite one exists. What Theorem 1.13 does not say is which of the two regions the critical probability itself belongs to.

1.5. **Continuity away from the critical probability.**

By (1) the percolation probability is an infimum of polynomials, which gives one-sided regularity. Together with monotonicity, this reduces continuity to the value $\theta(p_c)$.

Lemma 1.14. (The Percolation Probability Is Right-Continuous) *For every* $p \in [0, 1)$*, the function* θ *is right-continuous at* p.

Proof. Fix n. The event A_n is determined by the finitely many edges meeting the ball of radius n about the origin, so $\mathbb{P}_p(A_n)$ is a finite sum of products of the numbers p and $1 - p$, hence a polynomial in p and in particular a continuous function of p.

By (1) the function θ is an infimum of continuous functions, so it is upper semicontinuous. A nondecreasing upper semicontinuous function is right-continuous, and θ is nondecreasing by Lemma 1.9. $\square$

Left-continuity above the critical probability follows from the almost sure uniqueness of the infinite open cluster, proved by Aizenman, Kesten, and Newman [1987] and given a shorter proof by

Burton and Keane [1989]. That uniqueness implies continuity of the percolation probability except possibly at the critical probability is due to van den Berg and Keane [1984], and Aizenman, Kesten, and Newman [1987] combined the two. Uniqueness is what prevents the percolation probability from being spread across infinitely many competing clusters as the parameter decreases. This uniqueness argument does not determine the value at p_c, because there is no infinite cluster below p_c; left-continuity at p_c is exactly the statement $\theta(p_c) = 0$.

Theorem 1.15. (Continuity Away from the Critical Probability)

(i) *The function θ is continuous at every $p \in [0, 1]$ with $p \neq p_c$.*
(ii) *The function θ is continuous on all of $[0, 1]$ if and only if $\theta(p_c) = 0$.*

Proof. We prove continuity below p_c, then continuity above p_c, and finally the stated equivalence.

Step 1: Continuity below the critical probability. If $p < p_c$ then θ vanishes on the whole interval $[0, p_c)$ by Lemma 1.9, and a function that is constant on a neighborhood of p is continuous at p.

Step 2: Continuity above the critical probability. If $p_c < p < 1$ then right-continuity at p is Lemma 1.14, and left-continuity at p holds because the infinite open cluster is almost surely unique at every parameter in the supercritical phase. At $p = 1$ only left-continuity is required, and it holds for the same reason. Steps 1 and 2 together give continuity at every parameter other than p_c.

Step 3: The equivalence. Suppose first that $\theta(p_c) = 0$. Then θ vanishes on $[0, p_c]$, and it is right-continuous at p_c by Lemma 1.14, so it is continuous at p_c; with Steps 1 and 2 this gives continuity on $[0, 1]$. Conversely, suppose θ is continuous on $[0, 1]$. Since θ vanishes on $[0, p_c)$, continuity at p_c forces $\theta(p_c) = \lim_{p\uparrow p_c} \theta(p) = 0$. $\square$

Equivalently, the percolation probability has at most one discontinuity, it can sit only at the critical probability, and it would be a jump from the value 0 up to the value $\theta(p_c)$. Deciding whether that jump occurs is the same as computing one number.

1.6. **Classical results by dimension.** Classical results establish $\theta(p_c) = 0$ in dimension two and in sufficiently high dimensions, using different methods.

In two dimensions the critical probability equals $1/2$ and the percolation probability vanishes there. Harris [1960] proved in 1960 that $\theta(1/2) = 0$, using the correlation inequality that now carries his name together with planar duality, and Kesten [1980] completed the picture in 1980 by proving $p_c(2) = 1/2$, resting on the crossing estimates proved independently by Russo [1978] and Seymour and Welsh [1978]; the book-length account is Kesten [1982]. For site percolation in the plane the corresponding statement, that the percolation probability is continuous on the whole parameter interval, is due to Russo [1981]. Both steps use duality in an essential way, and duality is available only in the plane, so neither argument survives into three dimensions.

For bond percolation in high dimensions the lace expansion applies. Hara and Slade [1990] proved $\theta(p_c) = 0$ in high dimensions, and above six dimensions for a class of models with spread-out steps; together with Hara and Slade [1994] the threshold for the nearest-neighbor model is $d \geq 19$, although the computations behind that number were never published, as Fitzner and van der Hofstad [2017] recorded. They brought the threshold down to $d \geq 11$. The method compares the probability that two vertices are connected with the corresponding quantity for simple random walk, and the comparison needs enough room for two independent walks to avoid each other, which is what forces the dimension to be large.

The dimensions $3 \leq d \leq 10$ were left open by both methods and remained so for decades, as recorded in Grimmett [1999, §8.3] and in Fitzner and van der Hofstad [2017, §1.1]. For site percolation, the cases $d \geq 3$ also remained open. Writing in August 2026, Cerf [2026] described $\theta(p_c) = 0$ for site percolation in dimensions three and above as a central open question, proved that at p_c either the percolation probability vanishes or the tail of the finite cluster size fails to be a stretched exponential, and discussed the inequality of Kozma and Nitzan [2024] as a possible route. Sharpness of the phase transition, which these notes do not use, was proved independently by Aizenman and Barsky [1987] and Menshikov [1986], with short modern proofs by Duminil-Copin and Tassion [2016a] and Duminil-Copin and Tassion [2016b]; Duminil-Copin [2018] surveyed what is known at and near the critical point. The standard references for this section are Grimmett [1999], Bollobás and Riordan [2006] and Friedli and Velenik [2017]. Section 2 develops the correlation inequalities used below, beginning with the one Harris used. Section 3 then explains a route to the middle dimensions that passes through a statement about finite weighted graphs independent of lattice geometry, and Sections 4 and 5 prove that statement.

2. Correlation Inequalities for Percolation

Many arguments in percolation use correlation inequalities expressing that events helped by opening edges tend to occur together. Harris [1960] proved the first such statement in 1960 and used it to show that the percolation probability vanishes at $1/2$ in the plane. Stronger and more flexible versions have been developed since, and this section builds the versions needed below.

This section proves two results. The first is Harris's inequality: two increasing events are positively correlated, so the probability that both occur is at least the product of their probabilities. Section 2.1 proves it by induction on the number of edges. Section 2.2 proves it again, this time from the four functions theorem of Ahlswede and Daykin [1978], which compares four sums at once and which we need in its own right later.

Harris's inequality is not enough for what follows. The later argument compares the cluster of one vertex with a set of other vertices while forbidding that cluster from reaching somewhere else. This avoidance condition is needed to obtain a bound uniform in the size of that set. Conditioning on such an event destroys the independence of the edges, so Harris's inequality does not apply to the conditioned measure. Section 2.3 shows that its conclusion fails as well, by exhibiting two increasing functions that become negatively correlated.

The second is a conditional association statement. Under the avoidance conditioning, positive correlation holds for increasing functions of the cluster of the conditioned vertex. We prove it in the form where the function may read the whole set of open elements inside the cluster and not only the vertices the cluster occupies. That form contains the other, since the vertices are recovered from the open elements, and it is the form the later sections use. Section 2.4 introduces that set and the two cluster-exploration facts used in the proof. Section 2.5 proves the inequality. For graphs this is the theorem of van den Berg, Häggström, and Kahn [2006]; the labeled-hyperedge form used here extends that graph result and supplies the main input to Sections 4 and 5.

2.1. Increasing functions and the Harris inequality.

Throughout this section $n \geq 0$ is an integer and we work on the finite cube $\{0,1\}^n$ rather than on the lattice; the lattice statements follow by a limit that we record at the end of the section. Order the cube pointwise: write $x \leq y$ when $x_i \leq y_i$ for every i with $1 \leq i \leq n$.

Definition 2.1 (Increasing Function) Let $n \geq 0$ be an integer. A function $f : \{0,1\}^n \to \mathbb{R}$ is **increasing** when $f(x) \leq f(y)$ for all $x \leq y$, and **decreasing** when $-f$ is increasing.

An event is increasing exactly when its indicator function is an increasing function, so Definition 2.1 extends Definition 1.4 from events to functions.

Let $p_1, \ldots, p_n$ be numbers in $[0,1]$ and let μ denote the product measure on $\{0,1\}^n$ under which the coordinates are independent and the i-th coordinate equals 1 with probability p_i. The proof applies to unequal coordinate parameters, which are needed later.

The one-dimensional case is a classical rearrangement inequality and forms the base of the induction.

Lemma 2.2. (Positive Correlation in One Coordinate) *Let $p \in [0,1]$ and let f and g be increasing functions on $\{0,1\}$. Let X be a random variable with $\mathbb{P}(X=1) = p$ and $\mathbb{P}(X=0) = 1-p$. Then*

$$\mathbb{E}\big[f(X)g(X)\big] \geq \mathbb{E}\big[f(X)\big]\mathbb{E}\big[g(X)\big]\,.$$

Proof. Let X and Y be independent with the same law. Since f and g are both increasing on the two-point set, the product $\big(f(X)-f(Y)\big)\big(g(X)-g(Y)\big)$ is a product of two numbers with the same sign, so it is nonnegative, and therefore its expectation is nonnegative. Expanding,

$$0 \le \mathbb{E}\big[(f(X)-f(Y))(g(X)-g(Y))\big] = 2\mathbb{E}\big[f(X)g(X)\big] - 2\mathbb{E}\big[f(X)\big]\mathbb{E}\big[g(X)\big]\,,$$

where we used that X and Y are independent with the same law to identify the cross terms. Dividing by 2 gives the inequality. □

Theorem 2.3. (Harris Inequality) *Let $n \ge 0$ be an integer, let $p_1, \ldots, p_n$ be numbers in $[0,1]$, and let μ be the product measure on $\{0,1\}^n$ with these parameters. If f and g are increasing functions on $\{0,1\}^n$, then*

$$\int fg \,\mathrm{d}\mu \ge \int f \,\mathrm{d}\mu \int g\,\mathrm{d}\mu\,.$$

In particular, if A and B are increasing events then $\mu(A\cap B) \ge \mu(A)\mu(B)$.

The inequality is due to Harris [1960]. It was later derived in much greater generality by Fortuin, Kasteleyn, and Ginibre [1971], for measures on a distributive lattice satisfying a positivity condition, and on a product space the two statements agree; the lattice form is due to Kleitman [1966].

Proof. We argue by induction on the number of coordinates n, that is, we prove for every $n \ge 0$ that the displayed inequality holds for all increasing f and g on $\{0,1\}^n$ and all parameter vectors.

Step 1: Base case. For $n = 0$ the cube is a single point and both sides are the same number. For $n = 1$ the assertion is Lemma 2.2.

Step 2: Condition on the last coordinate. Suppose $n \ge 2$ and the assertion holds for $n-1$ coordinates. Write a point of $\{0,1\}^n$ as (x,t) with $x \in \{0,1\}^{n-1}$ and $t \in \{0,1\}$, and for $t \in \{0,1\}$ define

$$F(t) \coloneqq \int f(x,t)\,\mathrm{d}\mu_{n-1}(x) \qquad G(t) \coloneqq \int g(x,t)\,\mathrm{d}\mu_{n-1}(x)\,,$$

where μ_{n-1} is the product measure on the first $n-1$ coordinates. For each fixed t the functions $x \mapsto f(x,t)$ and $x \mapsto g(x,t)$ are increasing on $\{0,1\}^{n-1}$, so the induction hypothesis gives

$$\int f(x,t)g(x,t)\,\mathrm{d}\mu_{n-1}(x) \ge F(t)G(t) \qquad \text{for } t \in \{0,1\}\,. \tag{2}$$

Step 3: The outer coordinate. The functions F and G are increasing on $\{0,1\}$: if $s \le t$ then $f(x,s) \le f(x,t)$ for every x because f is increasing, and integrating preserves the inequality. Let T denote a random variable with $\mathbb{P}(T=1) = p_n$. Integrating (2) against the law of T and then applying Lemma 2.2 to F and G,

$$\int fg\,\mathrm{d}\mu = \mathbb{E}\Big[\int f(x,T)g(x,T)\,\mathrm{d}\mu_{n-1}(x)\Big] \ge \mathbb{E}\big[F(T)G(T)\big] \ge \mathbb{E}\big[F(T)\big]\mathbb{E}\big[G(T)\big]\,,$$

and the right side equals $\int f\,\mathrm{d}\mu \int g\,\mathrm{d}\mu$ by Fubini's theorem. This completes the induction.

The statement for events follows by taking f and g to be the indicator functions of A and B, which are increasing because A and B are. □

The lattice version follows from the finite one by conditioning. Let A and B be increasing events in $\mathcal{F}_d$ and let $\mathcal{F}_n$ be the σ-algebra generated by the edges meeting the ball of radius n about the origin. The conditional expectations $\mathbb{P}_p(A \mid \mathcal{F}_n)$ and $\mathbb{P}_p(B \mid \mathcal{F}_n)$ are increasing functions of those finitely many coordinates, since averaging an increasing function over the remaining coordinates preserves monotonicity, so Theorem 2.3 applies to them on the finite cube and gives

$$\mathbb{E}_p\big[\mathbb{P}_p(A \mid \mathcal{F}_n)\mathbb{P}_p(B \mid \mathcal{F}_n)\big] \ge \mathbb{P}_p(A)\mathbb{P}_p(B)\,.$$

By the martingale convergence theorem the two conditional expectations converge almost surely to $\mathbf{1}_A$ and $\mathbf{1}_B$ as n grows, and they are bounded by 1, so the left side converges to $\mathbb{P}_p(A\cap B)$ by dominated convergence. We use the lattice form without further comment.

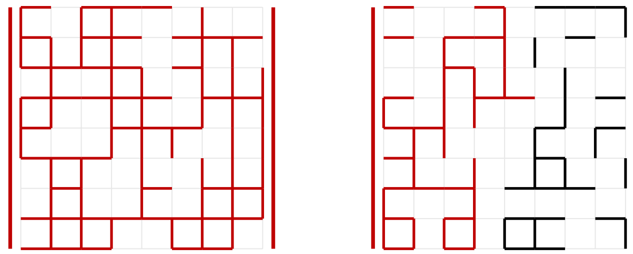


FIGURE 3. Harris's inequality in the case of two crossing events. Both the event that an open path crosses the box from left to right and the event that one crosses from bottom to top become easier when edges are opened, so the two are positively correlated and the probability of seeing both is at least the product of the two probabilities. The left panel shows a configuration where both occur, and the right panel one where only the horizontal crossing occurs.

2.2. **The four functions theorem.** Harris's inequality compares two integrals. The theorem in this section, due to Ahlswede and Daykin [1978], compares four, and it is the flexible tool from which the later arguments are built. For $x, y \in \{0,1\}^n$ write $x \vee y$ for the coordinatewise maximum and $x \wedge y$ for the coordinatewise minimum.

Theorem 2.4. (Four Functions Theorem) *Let $n \geq 0$ be an integer and let f_1, f_2, f_3, f_4 be nonnegative functions on $\{0,1\}^n$ satisfying*

$$f_1(x)f_2(y) \leq f_3(x \vee y)f_4(x \wedge y) \qquad \text{for all } x, y \in \{0,1\}^n \,. \tag{3}$$

Then

$$\Big(\sum_x f_1(x)\Big)\Big(\sum_y f_2(y)\Big) \leq \Big(\sum_x f_3(x)\Big)\Big(\sum_y f_4(y)\Big)\,,$$

where each sum is over all of $\{0,1\}^n$.

Proof. We induct on n. The strategy is that the hypothesis (3) is stable under summing out one coordinate, so the inductive step is a verification that the four one-coordinate sums again satisfy the hypothesis.

Step 1: Base case. For $n = 0$ the cube is a single point and the hypothesis at that point is the conclusion. Let $n = 1$ and write $a_i = f_i(0)$ and $b_i = f_i(1)$. Applying (3) to the four pairs (x, y) gives

$$a_1a_2 \leq a_3a_4 \qquad b_1b_2 \leq b_3b_4 \qquad a_1b_2 \leq b_3a_4 \qquad b_1a_2 \leq b_3a_4\,,$$

and we must show $(a_1+b_1)(a_2+b_2) \leq (a_3+b_3)(a_4+b_4)$. Abbreviate $s = b_3a_4$ and $t = a_3b_4$, and let $u = a_1b_2$ and $v = b_1a_2$, so that the last two displayed inequalities read $u \leq s$ and $v \leq s$. Expanding the left side,

$$(a_1 + b_1)(a_2 + b_2) = a_1a_2 + b_1b_2 + u + v\,, \tag{4}$$

while $(a_3 + b_3)(a_4 + b_4) = a_3a_4 + b_3b_4 + s + t$. Since the first two terms of (4) are already bounded by a_3a_4 and b_3b_4, it suffices to prove $u + v \leq s + t$.

If $s = 0$ then $u = v = 0$ and there is nothing to prove, so assume $s > 0$. If $u = 0$ then $u + v = v \leq s \leq s + t$, so assume $u > 0$ as well. Multiplying the first two displayed inequalities gives

$$uv = (a_1a_2)(b_1b_2) \leq (a_3a_4)(b_3b_4) = st\,.$$

Suppose first that $u \geq t$. Then $u \leq s$ and, from $uv \leq st$, also $v \leq {}^{st}/_{u}$, so

$$u + v \leq u + \frac{st}{u} \leq s + t\,,$$

the last step because $(u-s)(u-t) \le 0$ for $t \le u \le s$, which rearranges to $u^2 + st \le (s+t)u$. Suppose instead that $u < t$. Then $v \le s$ gives $u + v < t + s$. In both cases $u + v \le s + t$, which proves the base case.

Step 2: Inductive step. Suppose $n \ge 2$ and the theorem holds for $n-1$ coordinates. Write a point of $\{0,1\}^n$ as (x,t) with $x \in \{0,1\}^{n-1}$ and $t \in \{0,1\}$, and define functions on $\{0,1\}^{n-1}$ by

$$\widetilde{f}_i(x) \coloneqq f_i(x,0) + f_i(x,1) \qquad \text{for } 1 \le i \le 4\,.$$

We show that the four functions $\widetilde{f}_1, \widetilde{f}_2, \widetilde{f}_3, \widetilde{f}_4$ satisfy the hypothesis (3) on $\{0,1\}^{n-1}$. Fix x and y in $\{0,1\}^{n-1}$ and consider the four functions on the single coordinate t given by

$$g_1(t) = f_1(x,t) \quad g_2(t) = f_2(y,t) \quad g_3(t) = f_3(x \vee y, t) \quad g_4(t) = f_4(x \wedge y, t)\,.$$

For $s, t \in \{0,1\}$ the hypothesis on $\{0,1\}^n$ applied to the points (x,s) and (y,t) reads

$$f_1(x,s) f_2(y,t) \le f_3(x \vee y, s \vee t) f_4(x \wedge y, s \wedge t)\,,$$

which says exactly that g_1, g_2, g_3, g_4 satisfy (3) in one coordinate. The base case therefore gives $\widetilde{f}_1(x)\,\widetilde{f}_2(y) \le \widetilde{f}_3(x \vee y)\,\widetilde{f}_4(x \wedge y)$, which is the claim.

Applying the induction hypothesis to $\widetilde{f}_1, \dots, \widetilde{f}_4$ and noting that $\sum_x \widetilde{f}_i(x) = \sum_{(x,t)} f_i(x,t)$ completes the induction. □

Harris's inequality is a special case of Theorem 2.4, and the choice of four functions that produces it is the same one used later under conditioning, in Lemma 2.13.

Corollary 2.5. (Harris Inequality from Four Functions) *Let $n \ge 1$ be an integer, let μ be a product measure on $\{0,1\}^n$, and let f and g be increasing functions. Then $\int fg \, \mathrm{d}\mu \ge \int f \, \mathrm{d}\mu \int g \, \mathrm{d}\mu$.*

Proof. Write $\mu(x)$ for the mass of the point x and set

$$f_1 = f\mu \qquad f_2 = g\mu \qquad f_3 = fg\mu \qquad f_4 = \mu\,.$$

We check the hypothesis (3). A product measure satisfies $\mu(x)\mu(y) = \mu(x \vee y)\mu(x \wedge y)$, since for each coordinate the two sides record the same pair of values. Since f and g are increasing and nonnegative, $f(x) \le f(x \vee y)$ and $g(y) \le g(x \vee y)$. Multiplying,

$$f_1(x) f_2(y) = f(x) g(y) \mu(x) \mu(y) \le f(x \vee y) g(x \vee y) \mu(x \vee y) \mu(x \wedge y) = f_3(x \vee y) f_4(x \wedge y)\,.$$

Theorem 2.4 now gives $\left(\int f \, \mathrm{d}\mu\right)\left(\int g \, \mathrm{d}\mu\right) \le \left(\int fg \, \mathrm{d}\mu\right) \cdot 1$, which is the claim for nonnegative f and g.

For increasing f and g that are not nonnegative, note that the cube is finite, so f and g are bounded; choose a constant M with $f + M \ge 0$ and $g + M \ge 0$. Both $f + M$ and $g + M$ are increasing and nonnegative, so the case already proved applies to them, and adding a constant to either argument leaves a covariance unchanged. □

2.3. **Avoidance conditioning and the limits of Harris's inequality.** We first locate where the unconditional inequality fails, since the shape of the failure dictates the shape of the repair.

Fix a vertex x, a set Y of vertices, and consider the event $\{x \nleftrightarrow Y\}$ that x is connected to no vertex of Y. This event is decreasing: opening more edges can only create connections. Conditioning on it therefore biases the configuration downward, and the conditioned measure is no longer a product measure.

Example 2.6 Take the graph with three vertices x, y and z and the two edges xz and zy, each open with probability $1/2$, and let $Y = \{y\}$. Under the unconditioned measure the two edges are independent. Conditioned on $\{x \nleftrightarrow Y\}$, which here is the event that the two edges are not both open, the edges are no longer independent: the conditional probability that xz is open is $1/3$, and given that xz is open the conditional probability that zy is open is 0. So conditioning on an avoidance event destroys independence, and Theorem 2.3 does not apply to the conditioned measure.

The example does more than show that the proof of Theorem 2.3 cannot be reused. Take f to be the indicator that the edge xz is open and g the indicator that zy is open, both increasing functions of the configuration. Conditioned on $\{x \nleftrightarrow Y\}$ their expectations are each $1/3$ while the expectation of their product is 0, so they are negatively correlated under the conditioned measure. Positive correlation therefore fails outright for arbitrary increasing functions, and what survives the conditioning is the statement restricted to functions of the cluster of x, which is the subject of the next section.

2.4. **Labeled hypergraphs and cluster exploration.** The subsequent finite-model arguments use labeled hyperedges, in which an open element may join more than two vertices. Section 1.1 explains the choice. The labeled formulation requires explicit trace and exploration bookkeeping and permits the later specialization to site percolation.

Definition 2.7 (Independent Hyperedges) Let U be a finite vertex set and E a finite set of **hyperedges**. Each $e \in E$ carries an **incidence set** $\iota(e) \subseteq U$ with at least two elements, and is **open** with probability p_e, independently of the others. Two vertices are **connected** when a chain of open hyperedges joins them, consecutive hyperedges sharing a vertex. For $S \subseteq U$ write $\mathcal{C}_S$ for the union of the connected components meeting S.

Bond percolation is the case in which every incidence set has two elements. Section 6 shows that site percolation is another case, and that it is not a case of bond percolation.

Conditioning on what a cluster does requires remembering slightly more than the set of vertices it occupies.

Definition 2.8 (The Open Hyperedges of a Cluster) For $S \subseteq U$ write

$$\mathcal{K}_S = \{e \in E : e \text{ is open and } \iota(e) \cap \mathcal{C}_S \neq \varnothing\}$$

for the set of open hyperedges that meet the cluster of S.

The cluster is recovered from $\mathcal{K}_S$ together with S, by

$$\mathcal{C}_S = S \cup \bigcup_{e \in \mathcal{K}_S} \iota(e)\,,$$

since a vertex of $\mathcal{C}_S$ outside S lies on an open hyperedge meeting $\mathcal{C}_S$, and conversely the incidence set of such a hyperedge lies in $\mathcal{C}_S$. If $x \in S$ then the cluster of x is recovered as well, by taking the component of x in the hyperedges of $\mathcal{K}_S$. The set $\mathcal{C}_S$ by itself does not determine that cluster, and this already fails when every incidence set has two elements.

Proposition 2.9. (The Cluster of a Set Does Not Determine the Cluster of a Member) *There are a graph, a set S of vertices, a vertex $x \in S$, and two configurations giving S the same $\mathcal{C}_S$ but giving x different clusters.*

Proof. Take the path on x, y and z and let $S = \{x, z\}$. If both edges are open then $\mathcal{C}_S = \{x, y, z\}$ and the cluster of x is $\{x, y, z\}$. If xy is open and yz is closed then $\mathcal{C}_S = \{x, y, z\}$ again, since z belongs to S and y is reached from x, while the cluster of x is now $\{x, y\}$. □

This is why the inequalities below are stated for $\mathcal{K}_S$ rather than for $\mathcal{C}_S$. These sets are ordered by inclusion, and $\mathcal{C}_S$ grows with them.

We use one bookkeeping device throughout. For $Z \subseteq U$ write E_Z for the set of hyperedges meeting Z. If $I \subseteq E_Z$ is the set of open ones, then its **trace** outside Z is

$$T_Z(I) = \bigcup_{e \in I} \big(\iota(e) \setminus Z\big)\,. \tag{5}$$

Lemma 2.10. (Union Preservation and the Intersection Inclusion for Traces) *For all $I, J \subseteq E_Z$,*

$$T_Z(I \cup J) = T_Z(I) \cup T_Z(J) \qquad T_Z(I \cap J) \subseteq T_Z(I) \cap T_Z(J)\,.$$

Proof. The first identity is immediate from (5), since a union of unions is the union. For the second, if $e \in I \cap J$ then $\iota(e) \setminus Z$ is contained in both $T_Z(I)$ and $T_Z(J)$. The inclusion may be strict, because two different hyperedges, one in I only and one in J only, can contribute the same vertex. □

Lemma 2.11. (Exploration Identity) *Let $Z \subseteq N \subseteq U$. Condition on the states of the hyperedges in E_Z and let $I \subseteq E_Z$ be the open ones. Then the following hold.*

(i) Whenever a cluster contained in $U \setminus Z$ avoids N, deleting Z replaces the avoided set N by $(N \setminus Z) \cup T_Z(I)$.
(ii) The hyperedges disjoint from Z retain their original independent laws.

Proof. A path that enters Z must use an open hyperedge meeting Z, that is, a member of I, and the last vertex it occupies before leaving Z lies in the trace (5). Conversely every vertex of the trace is joined to Z by the open hyperedge that put it there. So avoiding N from outside Z is the same as avoiding $(N \setminus Z) \cup T_Z(I)$ in the model with Z deleted. The remaining coordinates are disjoint from E_Z, so exposing E_Z leaves their product law unchanged. □

After the open hyperedges meeting a cluster have been recorded, hyperedges disjoint from the resulting vertex set retain their original product law. This is the second bookkeeping fact used below and permits iterated conditioning.

Lemma 2.12. (Conditioning on a Cluster Leaves the Rest Unbiased) *Let $S \subseteq U$ and let $A \subseteq E$ be a set of hyperedges such that $\mathbb{P}(\mathcal{K}_S = A) > 0$, and let W be the union of S with the incidence sets of the members of A. Conditionally on $\{\mathcal{K}_S = A\}$ the hyperedges whose incidence sets are disjoint from W are independent, each open with its original probability p_e.*

Proof. On $\{\mathcal{K}_S = A\}$ we have $\mathcal{C}_S = W$: every member of A is open and meets $\mathcal{C}_S$, so its incidence set lies in $\mathcal{C}_S$, and conversely a vertex of $\mathcal{C}_S$ outside S is reached by a chain of open hyperedges, each of which meets $\mathcal{C}_S$ and is therefore recorded in A. Consequently a hyperedge is open and meets W exactly when it belongs to A, so

$$\{\mathcal{K}_S = A\} = \bigcap_{e \in A} \{e \text{ open}\} \cap \bigcap_{e \notin A,\ \iota(e) \cap W \neq \varnothing} \{e \text{ closed}\}\,.$$

For the reverse inclusion, use $\mathbb{P}(\mathcal{K}_S = A) > 0$: some configuration has $\mathcal{K}_S = A$, and in it every vertex met by a hyperedge of A is joined to S through hyperedges of A alone, since every open hyperedge meeting $\mathcal{C}_S$ belongs to A. So if all of A is open then $W \subseteq \mathcal{C}_S$, and if in addition every hyperedge outside A meeting W is closed then no path leaves W, so $\mathcal{C}_S = W$ and the open hyperedges meeting $\mathcal{C}_S$ are exactly A.

The displayed event is therefore determined by the states of the hyperedges meeting W. The remaining hyperedges are independent of those, so conditioning does not change their joint law. □

2.5. **Conditional association.** We can now prove that conditioning on an avoidance event preserves positive correlation. Write

$$K_U(f; X) = \mathbb{E}\big[f(\mathcal{K}_s); \mathcal{C}_s \cap X = \varnothing\big]$$

for a nonnegative increasing function f of $\mathcal{K}_s$, where s is a vertex.

Lemma 2.13. (Conditional Association for One Explored Cluster) *Let $s \in U$, let $X, Y \subseteq U$, and let f and g be nonnegative increasing functions of $\mathcal{K}_s$. Then*

$$K_U(f; X) K_U(g; Y) \leq K_U(fg; X \cap Y) K_U(1; X \cup Y)\,. \tag{6}$$

Proof. We induct on the number of vertices of U. If $s \in X \cup Y$ then one factor on the right is zero and one on the left is zero, so assume $s \notin X \cup Y$. Let $Z = X \cap Y$.

Step 1: The case $Z = \varnothing$. Write A_X for the event $\mathcal{C}_s \cap X = \varnothing$ and A_Y for $\mathcal{C}_s \cap Y = \varnothing$. The functions $f(\mathcal{K}_s)$ and $g(\mathcal{K}_s)$ are increasing in the hyperedge coordinates and the indicators of A_X and A_Y are decreasing, so Theorem 2.3 gives four inequalities:

$$\mathbb{E}[f; A_X] \le \mathbb{E}[f]\mathbb{P}(A_X) \qquad \mathbb{E}[g; A_Y] \le \mathbb{E}[g]\mathbb{P}(A_Y)\,,$$
$$\mathbb{E}[f]\mathbb{E}[g] \le \mathbb{E}[fg] \qquad \mathbb{P}(A_X)\mathbb{P}(A_Y) \le \mathbb{P}(A_X \cap A_Y)\,.$$

The first two apply the theorem to an increasing function and a decreasing one, which is the theorem applied to the increasing pair f and the negative of the indicator. The third applies it to two increasing functions and the fourth to two decreasing ones. Multiplying the first two and then using the third and the fourth,

$$\mathbb{E}[f; A_X]\mathbb{E}[g; A_Y] \le \mathbb{E}[fg]\mathbb{P}(A_X \cap A_Y)\,.$$

Since $X \cap Y = \varnothing$ the left side is $K_U(f; X)K_U(g; Y)$, while $\mathbb{E}[fg] = K_U(fg; \varnothing)$ and $\mathbb{P}(A_X \cap A_Y) = K_U(1; X \cup Y)$. This is (6).

Step 2: Inductive step. Suppose $Z \neq \varnothing$. Expose the states of the hyperedges in E_Z and delete Z; by Lemma 2.11 each of the four factors in (6) becomes a sum over the exposed open set $I \subseteq E_Z$, with the avoided set replaced by its trace. For two exposed states I and J apply the induction hypothesis in $U \setminus Z$, which has fewer vertices, with avoided sets

$$(X \setminus Z) \cup T_Z(I) \qquad \text{and} \qquad (Y \setminus Z) \cup T_Z(J)\,.$$

By Lemma 2.10 the union of these two sets is $((X \cup Y) \setminus Z) \cup T_Z(I \cup J)$, and their intersection contains $((X \cap Y) \setminus Z) \cup T_Z(I \cap J)$. Enlarging an avoided set can only decrease an avoidance probability, so the induction hypothesis gives, for each pair I and J, the pointwise inequality indexed by I, J, $I \cap J$ and $I \cup J$ that the four functions theorem requires.

Let $\pi(I)$ be the probability that the exposed open set is exactly I. Since the hyperedges are independent, $\pi(I)\pi(J) = \pi(I \cap J)\,\pi(I \cup J)$. Multiplying the pointwise inequality by these weights puts it in the form of the hypothesis (3), so Theorem 2.4, applied on the Boolean lattice of subsets of E_Z, gives (6) after summing over I and J. □

Corollary 2.14. (Conditional Association Given an Avoidance Event) *Let $s \in U$ and $X \subseteq U$ with $\mathbb{P}(\mathcal{C}_s \cap X = \varnothing) > 0$. Then increasing functions of $\mathcal{K}_s$ are positively correlated under the conditional law given $\mathcal{C}_s \cap X = \varnothing$.*

Proof. Take $X = Y$ in Lemma 2.13 and divide by $\mathbb{P}(\mathcal{C}_s \cap X = \varnothing)^2$, which is positive. This gives the assertion for nonnegative f and g. For real-valued increasing f and g, add a constant to each to make them nonnegative, which is possible because the state space is finite; adding a constant to either argument leaves a covariance unchanged. □

The restriction to functions of $\mathcal{K}_s$ cannot be dropped. Example 2.6 exhibits two increasing functions of the configuration that are negatively correlated under such a conditioning.

2.6. **Cluster-exploration lemmas.** Six consequences of the preceding results are collected here for use in the proof of the conditional covariance inequalities, Theorem 4.9. The first two are known for graphs; all six are formulated here for labeled hyperedges.

One piece of notation runs through the six. For $W \subseteq U$ write $\mathbb{P}_W$ and $\mathbb{E}_W$ for the model induced on W, in which a hyperedge is retained only when its incidence set is contained in W, and write $\mathcal{C}_S^W$ and $\mathcal{K}_S^W$ for the cluster of S there and for the open hyperedges of that cluster. If a named vertex lies outside W, then its cluster is empty and every connection to it fails, and $\mathcal{C}_\varnothing^W = \varnothing$.

The first of the six compares two clusters rather than one. The conditioning forces them apart, so growing one shrinks the other, and the inequality records that the two effects still point the same way. For graphs this is a conditional correlation inequality of van den Berg, Häggström, and Kahn [2006], and the resampling argument below is theirs.

Lemma 2.15. (Two Clusters Given Disconnection) *Let S and T be disjoint nonempty subsets of U with $\mathbb{P}(\mathcal{C}_S \cap T = \varnothing) > 0$. Conditional on $\mathcal{C}_S \cap T = \varnothing$, two nonnegative functions of $(\mathcal{K}_S, \mathcal{K}_T)$ that are increasing in the first argument and decreasing in the second are positively correlated.*

Proof. Let F and G be the two functions. Condition first on $\mathcal{K}_S = J$, and let $K = \mathcal{C}_S$ be the support of J. Lemma 2.12 says that the hyperedges contained in $U \setminus K$ retain their product law. With J fixed, both F and G are decreasing functions of $\mathcal{K}_T$, so the Harris inequality, applied to their negatives, gives

$$\mathbb{E}[FG \mid \mathcal{K}_S = J] \geq \mathbb{E}[F \mid \mathcal{K}_S = J]\mathbb{E}[G \mid \mathcal{K}_S = J]\,.$$

Both conditional means on the right are increasing functions of J on the set of values $\mathcal{K}_S$ can take. Indeed, suppose $J \subseteq L$. The support of L contains the support of J. Couple the two induced models outside these supports by using the same hyperedge states. Removing the larger support can only decrease $\mathcal{K}_T$. At the same time the first argument of F and G has increased. Thus both functions, and hence both conditional means, can only increase.

A conditional mean m is so far defined only at the values $\mathcal{K}_S$ can take. Extend it to every set A of hyperedges by

$$\overline{m}(A) = \max\bigl(\{m(I) : I \subseteq A \text{ and } m \text{ is defined at } I\} \cup \{0\}\bigr)\,.$$

This extension is increasing and agrees with m wherever m was defined. Corollary 2.14 also holds for a finite source set: adjoin a new vertex s, add a deterministic open hyperedge with incidence set $S \cup \{s\}$, and apply that corollary to the cluster of s. Consequently the two conditional means are positively correlated under the law of $\mathcal{K}_S$ given $\mathcal{C}_S \cap T = \varnothing$. Averaging the displayed conditional Harris inequality proves the lemma. □

The next lemma lets a cluster be explored one hyperedge at a time before the Harris inequality is applied. Without it, exposing a cluster would destroy the product structure that inequality needs. Gladkov [2024] proved a form of the Harris inequality for decision trees; this is the version the argument below uses.

Lemma 2.16. (Harris Inequality at a Stopping Time) *Let μ be a product Bernoulli measure on finitely many coordinates. Let $\mathcal{T}$ be an adaptive procedure that reveals previously unrevealed coordinates, one at a time, and eventually stops. Let $\mathcal{F}_\mathcal{T}$ be the information it has revealed. If f and g are increasing real functions, then*

$$\operatorname{Cov}_\mu\bigl(\mathbb{E}_\mu[f \mid \mathcal{F}_\mathcal{T}], g\bigr) \geq 0\,. \tag{7}$$

Proof. Let ω and η be independent samples from μ. Run $\mathcal{T}$ on ω, and form a third configuration ξ by taking each revealed coordinate from ω and each unrevealed coordinate from η. We first prove

$$\mathbb{E}[f(\omega)g(\xi)] \geq \mathbb{E}[f]\mathbb{E}[g]\,. \tag{8}$$

We induct on the largest number of further queries that the procedure can make. If it stops without a query, then $\xi = \eta$ and equality holds. Otherwise let i be its first query, let p_i be the opening probability, and let $\mathcal{T}_0$ and $\mathcal{T}_1$ be the procedures followed after the answers 0 and 1. Write f_b and g_b for the restrictions to the branch with coordinate i equal to b. The induction hypothesis in the two branches gives

$$\mathbb{E}[f(\omega)g(\xi)] \geq (1-p_i)\mathbb{E}[f_0]\mathbb{E}[g_0] + p_i\mathbb{E}[f_1]\mathbb{E}[g_1]\,.$$

Since f and g are increasing, $\mathbb{E}[f_0] \leq \mathbb{E}[f_1]$ and $\mathbb{E}[g_0] \leq \mathbb{E}[g_1]$. Lemma 2.2 applied to these two pairs of numbers shows that the last display is at least $\mathbb{E}[f]\mathbb{E}[g]$. This proves (8).

Conditional on $\mathcal{F}_\mathcal{T}$, the unrevealed coordinates of ω and η are independent and have their original product laws. Therefore

$$\mathbb{E}[f(\omega)g(\xi)] = \mathbb{E}\left[\mathbb{E}[f \mid \mathcal{F}_\mathcal{T}]\mathbb{E}[g \mid \mathcal{F}_\mathcal{T}]\right] = \mathbb{E}\left[\mathbb{E}[f \mid \mathcal{F}_\mathcal{T}]g\right]\,.$$

Together with (8), this is (7). □

The two lemmas just proved combine into the estimate the induction actually uses: what a covariance loses when a whole cluster is deleted from the model.

Lemma 2.17. (Covariance after Exploring a Cluster) *Let $x, v \in U$, let $S \subseteq U$ satisfy $x \in S$ and $v \notin S$, and let f be a nonnegative increasing function of $\mathcal{K}_x$. For $W \subseteq U$, let*

$$H(W) = \mathrm{Cov}_W\big(f(\mathcal{K}_x^W), \mathbf{1}\{v \leftrightarrow S\}\big)\,,$$

and if $x \notin W$ or $v \notin W$, then let $H(W) = 0$. Then the following hold.

(i) We have $H(W) \geq 0$.

(ii) For every $B \subseteq W$,

$$\mathbb{E}_W\big[H(W \setminus \mathcal{C}_B^W); x \notin \mathcal{C}_B^W\big]\mathbb{P}_W(v \nleftrightarrow S) \leq \mathbb{P}_W(v \nleftrightarrow S \cup B)H(W)\,. \tag{9}$$

Proof. The two variables defining $H(W)$ are increasing functions of the hyperedge states, so $H(W) \geq 0$ by Theorem 2.3. If $\mathbb{P}_W(v \nleftrightarrow S) = 0$, then both sides of (9) are zero. We may therefore suppose that this probability is positive.

Explore all of $\mathcal{K}_B$, and let $\mathcal{F}_B$ be the information revealed. Let

$$G = f(\mathcal{K}_x^W) \qquad J = \mathbf{1}\{v \leftrightarrow S\} \qquad Z = \mathbf{1}\{v \nleftrightarrow S,\ v \leftrightarrow B\} \qquad \widehat{G} = \mathbb{E}_W[G \mid \mathcal{F}_B]\,.$$

The law of total covariance and Lemma 2.12 give

$$H(W) = \mathbb{E}_W\big[H(W \setminus \mathcal{C}_B^W); x \notin \mathcal{C}_B^W\big] + \mathrm{Cov}_W(\widehat{G}, J)\,. \tag{10}$$

On the event $x \notin \mathcal{C}_B^W$, the first conditional covariance is the covariance in the model induced on $W \setminus \mathcal{C}_B^W$. On the complementary event G has been determined, so that conditional covariance is zero.

The variable $J + Z$ is the indicator of the increasing event $\{v \leftrightarrow S \cup B\}$. Lemma 2.16, applied to the exploration of $\mathcal{C}_B^W$, gives

$$\mathrm{Cov}_W(\widehat{G}, J + Z) \geq 0\,.$$

The variable Z is $\mathcal{F}_B$-measurable, so $\mathrm{Cov}_W(\widehat{G}, Z) = \mathrm{Cov}_W(G, Z)$.

Let $D = \{v \nleftrightarrow S\}$, $a = \mathbb{P}_W(D)$, and $r = \mathbb{P}_W(Z)$. Conditional on D, the clusters of S and v are disjoint. Apply Lemma 2.15 to G, viewed as an increasing function of $\mathcal{K}_S$, and to $1 - \mathbf{1}\{v \leftrightarrow B\}$, viewed as a decreasing function of $\mathcal{K}_v$. It follows that

$$\mathrm{Cov}_W(G, Z) \leq \frac{r}{a}\,\mathrm{Cov}_W(G, \mathbf{1}_D) = -\frac{r}{a}H(W)\,.$$

Using this bound in (10) yields

$$\mathbb{E}_W\big[H(W \setminus \mathcal{C}_B^W); x \notin \mathcal{C}_B^W\big] \leq \Big(1 - \frac{r}{a}\Big)\,H(W)\,.$$

Finally, $a - r = \mathbb{P}_W(v \nleftrightarrow S \cup B)$. Multiplication by a proves (9). ☐

The next lemma replaces a connection probability computed in the model that survives a deletion by the corresponding probability computed globally. It is the step where hyperedges have to be tracked by name rather than by the vertices they reach, since by Lemma 2.10 two of them can leave the same trace.

Lemma 2.18. (Comparison after Deleting a Cluster) *Let x, o, v be distinct vertices, let $A \subseteq U$ contain x and contain neither o nor v, and let $F : 2^U \to [0, \infty)$ satisfy $F(W) = 0$ whenever $v \notin W$. For $W \subseteq U$, let*

$$q_A(W) = \begin{cases} \dfrac{\mathbb{P}_W(o \leftrightarrow v,\ v \nleftrightarrow A)}{\mathbb{P}_W(v \nleftrightarrow A)}, & \mathbb{P}_W(v \nleftrightarrow A) > 0, \\ 0, & \mathbb{P}_W(v \nleftrightarrow A) = 0\,. \end{cases}$$

Suppose that, for every $W \subseteq U$ and every $B \subseteq W$,

$$\mathbb{E}_W\big[F(W \setminus \mathcal{C}_B^W); x \notin \mathcal{C}_B^W\big]\mathbb{P}_W(v \nleftrightarrow A) \leq \mathbb{P}_W(v \nleftrightarrow A \cup B)F(W)\,. \tag{11}$$

Then every $Y \subseteq U$ satisfies

$$\begin{aligned}&\mathbb{P}(o \leftrightarrow v,\ v \nleftrightarrow A \cup Y)\mathbb{E}\big[F(U \setminus \mathcal{C}_Y); x \notin \mathcal{C}_Y\big] \\ &\qquad \le \mathbb{P}(v \nleftrightarrow A \cup Y)\mathbb{E}\big[q_A(U \setminus \mathcal{C}_Y)F(U \setminus \mathcal{C}_Y); x \notin \mathcal{C}_Y\big]\,. \end{aligned} \tag{12}$$

Proof. The reason for keeping labels in the following argument is that two hyperedges meeting an exposed set can have the same trace outside that set. Their labels remain distinct even in that case.

Step 1: The stronger assertion. Let L be a finite label set and attach to each $\ell \in L$ a set of vertices $\alpha(\ell) \subseteq U$. For $Q \subseteq L$ and $W \subseteq U$, write

$$[Q]_W = W \cap \bigcup_{\ell \in Q} \alpha(\ell) \qquad R^W_Q = W \setminus \mathcal{C}^W_{[Q]_W}\,.$$

Define

$$\begin{aligned} e_W(Q) &= \mathbb{P}_W(o \leftrightarrow v,\ v \nleftrightarrow A \cup [Q]_W) \\ m_W(Q) &= \mathbb{P}_W(v \nleftrightarrow A \cup [Q]_W) \\ y_W(Q) &= \mathbb{E}_W\big[F(R^W_Q); x \notin \mathcal{C}^W_{[Q]_W}\big] \\ z_W(Q) &= \mathbb{E}_W\big[q_A(R^W_Q)F(R^W_Q); x \notin \mathcal{C}^W_{[Q]_W}\big]\,. \end{aligned}$$

We prove, simultaneously for every finite label set and every such attachment, that

$$e_W(N_1)y_W(N_2) \le m_W(N_1 \cup N_2)z_W(N_1 \cap N_2) \qquad \text{for all } N_1, N_2 \subseteq L\,. \tag{13}$$

The proof is by strong induction on $|W|$.

If $v \in A \cup [N_1]_W$ or $o \in [N_1]_W$, then $e_W(N_1) = 0$. If $x \in [N_2]_W$, then $y_W(N_2) = 0$. If $v \notin W$, then the condition on F also gives $y_W(N_2) = 0$. These cases establish (13), so exclude them below.

Step 2: Disjoint label sets. Suppose that $N_1 \cap N_2 = \varnothing$. Apply Lemma 2.13, with source v, to the constant function 1 and the increasing indicator $\mathbf{1}\{o \in \mathcal{C}_v\}$. Use the two avoided sets $A \cup [N_1]_W$ and $A \cup [N_2]_W$. Their intersection contains A, even when the sets attached to distinct labels overlap. Monotonicity in the avoided set therefore gives

$$m_W(N_2)e_W(N_1) \le e_W(\varnothing)m_W(N_1 \cup N_2)\,. \tag{14}$$

The assumption (11), with $B = [N_2]_W$, says

$$y_W(N_2)m_W(\varnothing) \le m_W(N_2)F(W)\,.$$

If $m_W(\varnothing) > 0$, then multiply this inequality by $e_W(N_1)$, use (14), and divide by $m_W(\varnothing)$. Since

$$z_W(\varnothing) = q_A(W)F(W) = \frac{e_W(\varnothing)}{m_W(\varnothing)}F(W)\,,$$

the result is (13). If $m_W(\varnothing) = 0$, then $e_W(N_1) = 0$, and the result is immediate.

Step 3: A common label. Suppose that $Z = N_1 \cap N_2$ is nonempty. A label whose attached set misses W can be discarded. If this makes Z empty, then the first case applies. We may thus assume that

$$K = [Z]_W$$

is nonempty. The zero cases already removed imply $x, o, v \notin K$. Expose every hyperedge meeting K, delete K, and let E_K be the set of exposed hyperedge labels. In $W \setminus K$ use the new label set

$$(L \setminus Z) \,\dot\cup\, E_K\,.$$

Attach $\alpha(\ell) \setminus K$ to an old label ℓ and $\iota(e) \setminus K$ to a new label $e \in E_K$.

For $I, J \subseteq E_K$, let

$$\widehat{N}_{1,I} = (N_1 \setminus Z) \,\dot\cup\, I \qquad \widehat{N}_{2,J} = (N_2 \setminus Z) \,\dot\cup\, J\,.$$

The labels form a disjoint union, so

$$\widehat{N}_{1,I} \cup \widehat{N}_{2,J} = ((N_1 \cup N_2) \setminus Z) \mathbin{\dot\cup} (I \cup J), \tag{15}$$

$$\widehat{N}_{1,I} \cap \widehat{N}_{2,J} = I \cap J\,. \tag{16}$$

Let $\pi(I)$ be the product probability that exactly the hyperedges in I are open among those meeting K. Lemma 2.11 and Lemma 2.12 expand the four terms in (13) as

$$e_W(N_1) = \sum_{I\subseteq E_K} \pi(I)e(I) \qquad\qquad y_W(N_2) = \sum_{J\subseteq E_K} \pi(J)y(J)$$

$$m_W(N_1 \cup N_2) = \sum_{I\subseteq E_K} \pi(I)m(I) \qquad\qquad z_W(N_1 \cap N_2) = \sum_{I\subseteq E_K} \pi(I)z(I)\,.$$

Here $e(I)$ and $y(J)$ are the corresponding quantities in $W \setminus K$ with sources $\widehat{N}_{1,I}$ and $\widehat{N}_{2,J}$. The functions $m(H)$ and $z(H)$ use the sources $((N_1 \cup N_2) \setminus Z) \mathbin{\dot\cup} H$ and H, respectively.

The induction hypothesis in $W \setminus K$, applied to $\widehat{N}_{1,I}$ and $\widehat{N}_{2,J}$, and (15)–(16) give

$$e(I)y(J) \le m(I \cup J)z(I \cap J)\,.$$

Independence of the exposed hyperedges gives

$$\pi(I)\pi(J) = \pi(I \cup J)\pi(I \cap J)\,.$$

Thus the four functions $\pi(I)e(I)$, $\pi(I)y(I)$, $\pi(I)m(I)$ and $\pi(I)z(I)$ satisfy the hypothesis of Theorem 2.4. That theorem proves (13). Since K is nonempty, this completes the strong induction.

Step 4: Return to vertex sources. Give each vertex of Y a separate singleton label and take $N_1 = N_2 = L$. Then $[L]_U = Y$. Substitution in (13) gives exactly (12). $\square$

Two disjoint clusters may be exposed in either order. The identity below says that the two orders give the same answer, and it is what lets one vertex be removed from a source set in the last lemma of the six.

Lemma 2.19. (Exchanging the Order of Two Cluster Explorations) *For $A \subseteq W$ and $K \subseteq W$, let*

$$\mu_A^W(K) = \mathbb{P}_W(\mathcal{C}_A^W = K)\,.$$

Let $A, B \subseteq W$, and let $K, L \subseteq W$ be disjoint sets satisfying $A \subseteq K$ and $B \subseteq L$. Then

$$\mu_A^W(K)\mu_B^{W\setminus K}(L) = \mu_B^W(L)\mu_A^{W\setminus L}(K)\,. \tag{17}$$

Proof. On the event $\mathcal{C}_A^W = K$, every hyperedge meeting both K and $W \setminus K$ is closed. By Lemma 2.12, or by summing that lemma over all values with support K, the hyperedges contained in $W \setminus K$ retain their product law. The left side of (17) is therefore

$$\mathbb{P}_W(\mathcal{C}_A^W = K,\ \mathcal{C}_B^W = L)\,.$$

Exploring the B-cluster first shows that the right side is the same joint probability. $\square$

That last lemma is the exact identity moving the induction from one level to the next.

Lemma 2.20. (Adding One Vertex to a Source) *Let x and z be distinct vertices. Let Y and S be disjoint subsets of U with $x \in S$ and $z \notin Y \cup S$. Let g be a bounded real function of $\mathcal{K}_x$. Let*

$$d_x = \mathbb{P}(x \nleftrightarrow Y) \qquad d_z = \mathbb{P}(z \nleftrightarrow Y \cup S) \qquad \rho(u) = \mathbb{P}(u \in \mathcal{C}_z \mid z \nleftrightarrow Y \cup S)\,,$$

and suppose $d_x d_z > 0$. Conditional on $x \nleftrightarrow Y$, expose $\mathcal{C}_Y$, let $W = U \setminus \mathcal{C}_Y$, and let π be the law of W. For $T \subseteq U$, let

$$r_T^W(u) = \mathrm{Cov}_W\big(g(\mathcal{K}_x^W), \mathbf{1}\{u \leftrightarrow T\}\big)\,.$$

For a set K, first form $\mathcal{C}_Y^{-K}$ in the configuration obtained by deleting K and every hyperedge meeting K. In the expression $\mathcal{K}_x^{-\mathcal{C}_Y^{-K}}$, restore the hyperedges meeting K and then delete $\mathcal{C}_Y^{-K}$ and every hyperedge meeting it. Let

$$\Phi(K) = \mathbb{E}\left[\mathbf{1}\{\mathcal{C}_x^{-K} \cap Y = \varnothing\}\left(g(\mathcal{K}_x^{-\mathcal{C}_Y^{-K}}) - g(\mathcal{K}_x^{-K})\right)\right]. \tag{18}$$

Then, for every vertex u,

$$\begin{aligned}&\mathbb{E}_\pi\left[r_S^W(u) - r_{S\cup\{z\}}^W(u) - \rho(u) r_S^W(z)\right]\\ &\qquad = \frac{d_z}{d_x}\operatorname{Cov}\left(\Phi(\mathcal{C}_z), \mathbf{1}\{u \in \mathcal{C}_z\} \mid z \not\leftrightarrow Y \cup S\right).\end{aligned} \tag{19}$$

Proof. For an induced vertex set R containing x, let

$$a(R) = \mathbb{E}_R[g(\mathcal{K}_x^R)].$$

Step 1: A fixed induced hypergraph. We first prove an identity with W fixed. Suppose W contains x and z, let γ be a real number, and set

$$D = \{z \not\leftrightarrow S\} \qquad J_u = D \cap \{u \in \mathcal{C}_z^W\}.$$

The disjoint identities

$$\mathbf{1}\{u \leftrightarrow S \cup \{z\}\} = \mathbf{1}\{u \leftrightarrow S\} + \mathbf{1}_{J_u} \qquad \mathbf{1}\{z \leftrightarrow S\} = 1 - \mathbf{1}_D$$

give

$$\begin{aligned}&r_S^W(u) - r_{S\cup\{z\}}^W(u) - \gamma r_S^W(z)\\ &\qquad = -\operatorname{Cov}_W(g(\mathcal{K}_x^W), \mathbf{1}_{J_u}) + \gamma \operatorname{Cov}_W(g(\mathcal{K}_x^W), \mathbf{1}_D).\end{aligned}$$

On D, the set $K = \mathcal{C}_z^W$ is disjoint from S. Summing Lemma 2.12 over the sets $\mathcal{K}_z^W$ whose incidence sets together with z make up K shows that, conditional on $\mathcal{C}_z^W = K$, the hyperedges contained in $W \setminus K$ still have their product law, so

$$\mathbb{E}_W[g(\mathcal{K}_x^W) \mid \mathcal{C}_z^W = K] = a(W \setminus K).$$

Expanding the two covariances therefore yields

$$\begin{aligned}&r_S^W(u) - r_{S\cup\{z\}}^W(u) - \gamma r_S^W(z)\\ &\qquad = \sum_{\substack{K\subseteq W\\ K\cap S=\varnothing}} \mu_{\{z\}}^W(K)\big(a(W) - a(W\setminus K)\big)\big(\mathbf{1}\{u\in K\} - \gamma\big).\end{aligned} \tag{20}$$

If $z \notin W$, then both sides are interpreted as zero.

Step 2: Expanding Φ. We next expand (18). For every K disjoint from $Y \cup S$, condition on $\mathcal{C}_Y^{U\setminus K} = Q$. Exact cluster conditioning gives

$$\Phi(K) = \sum_{\substack{Q\subseteq U\setminus K\\ x\notin Q}} \mu_Y^{U\setminus K}(Q)\left[a(U\setminus Q) - a(U\setminus(K\cup Q))\right]. \tag{21}$$

For the first term in the brackets, the hyperedges meeting K were absent from the exploration of $\mathcal{C}_Y^{U\setminus K}$ and retain their independent states. Together with the hyperedges disjoint from $K \cup Q$, they give the product model on $U \setminus Q$. For the second term, the boundary of Q is closed in $U \setminus K$, leaving the product model on $U \setminus (K \cup Q)$. The condition $x \notin Q$ is precisely the indicator in (18).

Step 3: Exchanging the order. Take $\gamma = \rho(u)$ in (20). Multiplying the conditional covariance on the right side of (19) by d_z and using (21) gives

$$\sum_{\substack{K\subseteq U\\ K\cap(Y\cup S)=\varnothing}} \mu^U_{\{z\}}(K)\Phi(K)\big(\mathbf{1}\{u\in K\}-\rho(u)\big)\,.$$

Insert the sum over Q from (21). Lemma 2.19 changes each product of cluster probabilities according to

$$\mu^U_{\{z\}}(K)\mu^{U\setminus K}_Y(Q) = \mu^U_Y(Q)\mu^{U\setminus Q}_{\{z\}}(K)\,.$$

After interchanging the two finite sums, the preceding display becomes

$$\sum_{\substack{Q\subseteq U\\ x\notin Q}} \mu^U_Y(Q) \sum_{\substack{K\subseteq U\setminus Q\\ K\cap S=\varnothing}} \mu^{U\setminus Q}_{\{z\}}(K)\,[a(U\setminus Q)-a(U\setminus(K\cup Q))]$$
$$\cdot\big(\mathbf{1}\{u\in K\}-\rho(u)\big)\,.$$

Terms with $z\in Q$ are zero and have only been added. By (20), the inner sum is

$$r^{U\setminus Q}_S(u) - r^{U\setminus Q}_{S\cup\{z\}}(u) - \rho(u)r^{U\setminus Q}_S(z)\,.$$

Finally,

$$\frac{\mu^U_Y(Q)}{d_x}\qquad \text{for } x\notin Q\,,$$

is the law of $U\setminus\mathcal{C}_Y$ conditional on $x\nleftrightarrow Y$. Dividing by d_x proves (19). □

3. Finite Connection Inequalities and the Critical Probability

Suppose we know two things: that a source vertex reaches some vertex of a given set with high probability, and that each vertex of that set reaches a target with high probability. Can we conclude that the source reaches the target? When the set has one element the union bound suffices. The difficulty is that the lattice argument of this section produces sets whose size grows without bound, and the union bound degrades linearly in that size. What is needed is a bound that does not.

Kozma and Nitzan [2024] proposed such statements as conjectures in 2024 and proved that one of them implies the vanishing of the percolation probability at the critical probability in every dimension. Their reduction replaces a lattice question not covered by duality or the lace expansion with a combinatorial statement about finite weighted graphs.

Section 3.1 sets up the problem of connecting through a set and shows how the union bound fails. Section 3.2 states the three finite connection statements used below. Section 3.3 states the finite criterion and describes its two cases. Section 3.4 shows that the additive bound suffices for bond percolation.

The multiplicative inequality below is Conjecture 1 of Kozma and Nitzan [2024]. The uniform near-one statement is their Conjecture 3; they recorded that they could neither prove nor disprove Conjecture 1. The intermediate additive bound does not appear in their paper. All three statements are proved here. Section 5 proves the multiplicative inequality; the additive and uniform near-one bounds follow from it, and the bond-percolation reduction uses the uniform near-one bound.

3.1. Connecting through a set.

Throughout this section the model is the one of Definition 2.7: a finite vertex set, finitely many hyperedges, each open independently with its own probability. Bond percolation on a finite graph is the case in which every incidence set is a pair, and site percolation is another case, as Section 6.1 shows. Thus each finite statement in this section specializes to both models. Write o and b for vertices, A for a nonempty set of vertices, and $o\leftrightarrow A$ for the event that o is connected to at least one vertex of A.

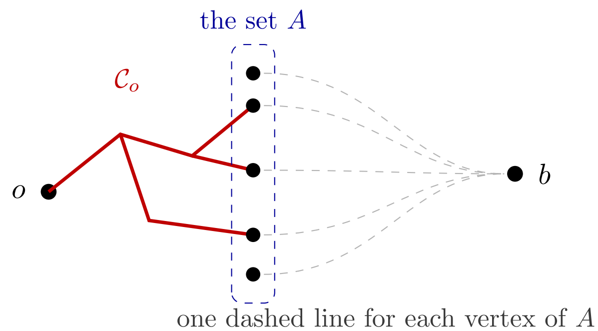


FIGURE 4. Connecting through a set. The source on the left reaches the set in the middle with high probability, and each vertex of that set reaches the target on the right with high probability. The union bound incurs one term for every dashed line, whereas the inequalities of Section 3.2 give a bound independent of the size of the set.

Assume

$$\mathbb{P}(o \leftrightarrow A) \geq 1-\delta \qquad \text{and} \qquad \mathbb{P}(a \leftrightarrow b) \geq 1-\delta \quad \text{for every } a \in A\,, \tag{22}$$

and we want a lower bound on $\mathbb{P}(o \leftrightarrow b)$. A direct estimate uses a union bound over the vertices of the set.

Lemma 3.1. (The Union Bound over A) *Let o and b be vertices of a finite family of independent hyperedges, and let A be a nonempty set of vertices. Then*

$$\mathbb{P}(o \nleftrightarrow b) \leq \mathbb{P}(o \nleftrightarrow A) + \sum_{a\in A} \mathbb{P}(a \nleftrightarrow b)\,.$$

Proof. If o is connected to b then no term on the right is needed, so assume the event $\{o \nleftrightarrow b\}$ occurs. Either o fails to reach A, which is the first term, or o reaches some vertex $a \in A$; in the latter case that a cannot reach b, since otherwise o would reach b by concatenating the two open paths. Hence

$$\{o \nleftrightarrow b\} \subseteq \{o \nleftrightarrow A\} \cup \bigcup_{a\in A} \{a \nleftrightarrow b\}\,,$$

and the claim follows from countable subadditivity. $\square$

Under the hypotheses (22) the lemma gives $\mathbb{P}(o \nleftrightarrow b) \leq \delta + |A|\delta$, which is not uniform in $|A|$ and gives no near-one estimate when $|A| \geq 1/\delta$. The limitation is not an artifact of a crude proof: the lemma discards the information that the vertices of the set are alternatives to one another rather than requirements, and a bound that treats them as requirements must pay for each one.

The size of the set cannot be controlled in advance, so a bound uniform in that size is required. In the reduction of Section 3.3, the set is the set of vertices through which a crossing of one block can enter the next, and its size grows with the block scale. There is no bound on it available in advance.

3.2. **The Kozma–Nitzan inequality and two consequences.** Three statements must be kept apart. Each is an assertion about every finite family of independent hyperedges, every o and b, and every nonempty finite set A; that quantifier is not repeated below. They differ in how the loss accumulates. Kozma and Nitzan [2024] stated the first and the third as their Conjectures 1 and 3.

Definition 3.2 (Kozma–Nitzan Conjecture 1) The first inequality is

$$\mathbb{P}(o \leftrightarrow b) \geq \mathbb{P}(o \leftrightarrow A) \cdot \min_{a\in A} \mathbb{P}(a \leftrightarrow b)\,.$$

The loss is a factor, so it compounds when the inequality is applied repeatedly.

Definition 3.3 (An Additive Consequence) The additive bound asserts that for every $t \geq 0$, if $\mathbb{P}(a \leftrightarrow b) \geq 1 - t$ for every $a \in A$, then

$$\mathbb{P}(o \leftrightarrow b) \geq \mathbb{P}(o \leftrightarrow A) - t\,.$$

Here the loss is a summand, and it does not depend on the size of A.

Definition 3.4 (A Uniform Near-One Consequence) The uniform near-one statement asserts that for every $\varepsilon > 0$ there is $\delta > 0$, depending on ε alone, such that if $\mathbb{P}(o \leftrightarrow A) > 1 - \delta$ and $\mathbb{P}(a \leftrightarrow b) > 1 - \delta$ for every $a \in A$, then $\mathbb{P}(o \leftrightarrow b) > 1 - \varepsilon$.

The three are listed in decreasing order of strength. The implications are recorded next.

Lemma 3.5. (The Kozma–Nitzan Inequality Implies the Additive Bound) *The inequality of Definition 3.2 implies the additive bound of Definition 3.3.*

Proof. Assume the first inequality, and let $t \geq 0$ with $\mathbb{P}(a \leftrightarrow b) \geq 1 - t$ for every $a \in A$. If $t \geq 1$ then the conclusion is immediate because $\mathbb{P}(o \leftrightarrow b) \geq 0 \geq \mathbb{P}(o \leftrightarrow A) - t$. So assume $t < 1$. Then

$$\mathbb{P}(o \leftrightarrow b) \geq \mathbb{P}(o \leftrightarrow A)(1 - t) = \mathbb{P}(o \leftrightarrow A) - t\mathbb{P}(o \leftrightarrow A) \geq \mathbb{P}(o \leftrightarrow A) - t\,,$$

using $\mathbb{P}(o \leftrightarrow A) \leq 1$ in the last step. $\square$

Lemma 3.6. (The Additive Bound Implies the Uniform Near-One Statement) *The additive bound implies the uniform near-one statement, with $\delta = \varepsilon/2$.*

Proof. Let $\varepsilon > 0$ and set $\delta = \varepsilon/2$. Suppose $\mathbb{P}(o \leftrightarrow A) > 1 - \delta$ and $\mathbb{P}(a \leftrightarrow b) > 1 - \delta$ for every $a \in A$. Applying the additive bound with $t = \delta$,

$$\mathbb{P}(o \leftrightarrow b) \geq \mathbb{P}(o \leftrightarrow A) - \delta > 1 - 2\delta = 1 - \varepsilon\,.$$

The number δ depends only on ε, as required. $\square$

Sections 4 and 5 prove the multiplicative statement, and Lemmas 3.5 and 3.6 then give the other two. The uniform near-one statement is used in the next subsection.

3.3. **A finite criterion for bond criticality.** The required reduction is Theorem 6 of Kozma and Nitzan [2024].

Theorem 3.7. (Kozma–Nitzan's Uniform Near-One Criterion) *Suppose the uniform near-one statement of Definition 3.4 holds. Then for every integer $d \geq 2$ the percolation probability of Bernoulli bond percolation on $\mathbb{Z}^d$ satisfies $\theta(p_c(d)) = 0$.*

The accompanying Lean development includes a formalization of Theorem 3.7. Its proof is not reproduced here; the following paragraphs summarize its two cases.

In two dimensions the conclusion is classical and the finite connection hypothesis is not needed: the critical probability equals $1/2$ and the percolation probability vanishes there, by Harris [1960] and Kesten [1980].

For $d \geq 3$ the argument is a renormalization. Suppose for a contradiction that $\theta(p_c) > 0$. Partition $\mathbb{Z}^d$ into blocks of a large side length and declare a block good when the configuration inside it contains a crossing cluster joining all its faces. Percolation of good blocks would give an infinite open cluster, so the aim is to show that blocks are good with probability close enough to 1. What has to be checked is that the crossing cluster of one block connects to the crossing cluster of the next, and this is exactly the problem of connecting through a set: the source is the crossing cluster of the first block, the set is the set of vertices on the shared face through which the connection can pass, and the target is the crossing cluster of the second block. The set is the face of a block, so its size grows with the block scale, which is precisely why the union bound of

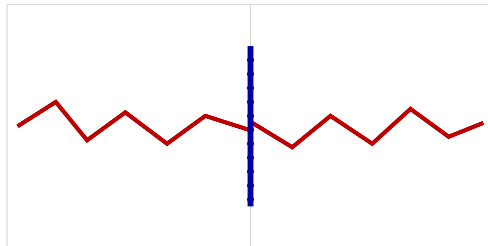


FIGURE 5. The set in the renormalization of Theorem 3.7. The crossing cluster of the left block must be joined to the crossing cluster of the right block, and the connection may cross the shared face, drawn in bold, at a vertex that is not fixed in advance. Since the number of such vertices grows with the side length of the blocks, the union bound over them does not give a scale-uniform estimate.

Lemma 3.1 cannot be used. The uniform near-one statement supplies a bound independent of the face size. The block percolation so obtained yields percolation in a slab, and the strict inequality for critical probabilities under an essential enhancement, proved by Aizenman and Grimmett [1991], then forces $p > p_c$, which is the contradiction. One-step renormalization of this kind originates with Grimmett and Marstrand [1990]; the absence of percolation at the critical probability in half-spaces is due to Barsky, Grimmett, and Newman [1991], and the corresponding statement inside a slab to Duminil-Copin, Sidoravicius, and Tassion [2016]. A textbook account of block renormalization is Grimmett [1999, §7.1–7.2].

The hypothesis of Theorem 3.7 is a statement about arbitrary finite families of independent hyperedges and is independent of the dimension, the lattice, and translation invariance. The number δ must also be independent of the intermediate set, which allows the block scale to be chosen after δ is fixed.

3.4. An additive bound sufficient for bond criticality.

Combining the results of this section reduces the lattice problem to a single finite inequality.

Corollary 3.8. (The Additive Bound Suffices for Bond Criticality) *Suppose the additive bound of Definition 3.3 holds. Then for every integer $d \geq 2$ the percolation probability of Bernoulli bond percolation on $\mathbb{Z}^d$ satisfies $\theta(p_c(d)) = 0$.*

Proof. The additive bound implies the uniform near-one statement by Lemma 3.6, and Theorem 3.7 gives the conclusion. □

By Theorem 1.15 the conclusion of Corollary 3.8 is equivalent to the continuity of the percolation probability on the whole parameter interval. Thus the additive bound is sufficient for the bond-percolation conclusion in every dimension.

The next two sections prove the required finite inequalities by an indirect route. Section 4 constructs a family of inequalities comparing conditional covariances at different vertices, and Section 5 converts that family into the additive bound by an induction on the size of the set. The estimate uses one selected vertex rather than a sum over every vertex.

4. Conditional Covariance Inequalities

Section 3 reduced the vanishing of the percolation probability at the critical probability to one finite inequality. This section develops a family of inequalities comparing conditional covariances at two vertices. Its bond-percolation form yields the additive and uniform near-one bounds used in the earlier formalization. The formulation here retains labeled hyperedges so that distinct hyperedges remain distinguishable when their traces coincide.

Theorem 4.9, in Section 4.4, states that a certain alternating sum of conditional covariances is nonnegative. The data are a finite family of independent hyperedges, a vertex x, a set Y that the cluster of x must avoid, two further vertices o and v, and a finite list of further vertices $d_1, \ldots, d_\ell$. Take Y empty, the list empty, and f the indicator that x is connected to v; the resulting statement is the Harris inequality of Section 2. The theorem extends positive correlation in two directions: it remains valid under avoidance conditioning and after successive subtraction of contributions from other vertices. The second property is what the induction in Section 5 needs, and it is why the vertices $d_1, \ldots, d_\ell$ are carried at all.

Section 4.1 introduces the conditional covariance and records what Section 2 already gives about it, namely that it is nonnegative. Section 4.2 states the comparison between o and v without auxiliary vertices and identifies the Harris inequality as a special case. Section 4.3 introduces the vertices $d_1, \ldots, d_\ell$ and the alternating sum, and proves the algebraic identity used in the induction. Section 4.4 states the theorem. Section 4.5 records a consequence used in Section 5: for fixed data the whole family reduces to finitely many inequalities in the edge parameters.

Throughout this section the model is the one of Definition 2.7, x is a vertex, and Y is a set of vertices with $x \notin Y$. Every conditioning is on the event $\{x \nleftrightarrow Y\}$ that the cluster of x reaches no vertex of Y, and that event is assumed to have positive probability. For bond percolation the set $\mathcal{K}_x$ of Definition 2.8 is the set of open edges with an endpoint in $\mathcal{C}_x$.

4.1. **Conditional covariance under avoidance.** **Definition 4.1** (Conditional Covariance Under Avoidance) Let f be an increasing real function of $\mathcal{K}_x$. For a vertex u define

$$\kappa(u) := \operatorname{Cov}\Big(f(\mathcal{K}_x), \mathbf{1}\{u \in \mathcal{C}_x\} \;\Big|\; \{x \nleftrightarrow Y\}\Big)\,. \tag{23}$$

Both arguments of the covariance are increasing functions of $\mathcal{K}_x$, so Corollary 2.14 applies and gives the following at once.

Lemma 4.2. (The Conditional Covariance Is Nonnegative) *In the setting of Definition 4.1, the following hold.*

(i) We have $\kappa(u) \geq 0$ for every vertex u.
(ii) If $u \in Y$, then $\kappa(u) = 0$.

Proof. Both $f(\mathcal{K}_x)$ and $\mathbf{1}\{u \in \mathcal{C}_x\}$ are increasing functions of $\mathcal{K}_x$, so their conditional covariance is nonnegative by Corollary 2.14.

If $u \in Y$ then on the event $\{x \nleftrightarrow Y\}$ the cluster of x does not reach u, so the indicator $\mathbf{1}\{u \in \mathcal{C}_x\}$ vanishes identically under the conditioned measure, and a covariance with a constant is zero. ☐

The argument of Section 5 requires a comparison between the values of κ at two distinct vertices; nonnegativity alone does not provide one.

4.2. **The comparison without auxiliary vertices.** Fix two distinct vertices o and v, neither in Y and neither equal to x. The comparison we need transports the value of κ at v to its value at o, at a price given by the probability that the two vertices o and v are joined to each other.

Definition 4.3 (The Number Lambda) Let o and v be distinct vertices outside $Y \cup \{x\}$, and let $d_1, \ldots, d_\ell$ be finitely many further vertices, distinct from one another and from x, o and v, and none of them in Y. Assume the event that v avoids $Y \cup \{x\} \cup \{d_1, \ldots, d_\ell\}$ has positive probability. The number λ is

$$\lambda := \mathbb{P}\Big(o \in \mathcal{C}_v \;\Big|\; \{v \nleftrightarrow Y \cup \{x\} \cup \{d_1, \ldots, d_\ell\}\}\Big)\,, \tag{24}$$

the probability that o lies in the open cluster of v, given that v reaches none of x, of Y, and of $d_1, \ldots, d_\ell$. When there are no vertices d_j the conditioning is on v avoiding $Y \cup \{x\}$ alone.

The conditioning in (24) must include every d_j. Omitting them gives a larger number, and the comparison below then fails: on the star with center d_1 and leaves x, o and v, all three edges open with probability $1/2$, taking Y empty and $f(\mathcal{K}_x) = \mathbf{1}\{v \in \mathcal{C}_x\}$, the inequality with the smaller conditioning set would read $0 \geq 1/48$.

Two features of (24) matter. The conditioning is on an avoidance event for v, not for x, so λ is computed under a different conditioning from the one defining κ. And λ does not depend on the function f.

Definition 4.4 (The Inequality Without Auxiliary Vertices) In the setting above, the inequality without auxiliary vertices is

$$\kappa(o) \geq \lambda\kappa(v) \tag{25}$$

for every increasing real function f of $\mathcal{K}_x$.

In its simplest case (25) reduces to a statement already available.

Lemma 4.5. (The Harris Inequality as a Special Case) *Let x, o and v be distinct vertices, take Y empty, and take $f(\mathcal{K}_x) = \mathbf{1}\{v \in \mathcal{C}_x\}$. Then the inequality without auxiliary vertices (25) is equivalent to the inequality*

$$\mathbb{P}\big(o \leftrightarrow \{x, v\} \text{ and } x \leftrightarrow v\big) \geq \mathbb{P}\big(o \leftrightarrow \{x, v\}\big)\mathbb{P}\big(x \leftrightarrow v\big)\,, \tag{26}$$

which is an instance of the Harris inequality, Theorem 2.3.

Proof. With Y empty there is no conditioning, and with $f(\mathcal{K}_x) = \mathbf{1}\{v \in \mathcal{C}_x\}$ the covariance (23) becomes the unconditioned covariance of the two indicators $\mathbf{1}\{v \in \mathcal{C}_x\}$ and $\mathbf{1}\{o \in \mathcal{C}_x\}$. Both are increasing events, namely $\{x \leftrightarrow v\}$ and $\{x \leftrightarrow o\}$, so (25) compares two covariances of pairs of increasing events. Expanding the covariances and collecting the terms in which the two clusters meet gives (26), and the events $\{o \leftrightarrow \{x, v\}\}$ and $\{x \leftrightarrow v\}$ are both increasing, so Theorem 2.3 applies. $\square$

The general case of (25), with Y nonempty and f an arbitrary increasing function, does not follow from the Harris inequality. The conditioning destroys the product structure used in the proof of Theorem 2.3, as Example 2.6 showed.

4.3. **Successive removal of auxiliary vertices.** The induction of Section 5 removes vertices from a set one at a time, and a removed vertex does not leave the argument: it stays as a vertex whose contribution must be subtracted. The family therefore has to carry a list of such vertices from the start.

Definition 4.6 (The Vertices d_j and Their Constants) Let $d_1, \ldots, d_\ell$ be distinct vertices, none equal to x, to o or to v, and none in Y. For $1 \leq j \leq \ell$ define

$$\rho_j(u) \coloneqq \mathbb{P}\big(u \in \mathcal{C}_{d_j} \;\big|\; \{d_j \nleftrightarrow Y \cup \{x\} \cup \{d_1, \ldots, d_{j-1}\}\}\big)\,, \tag{27}$$

the probability that u lies in the open cluster of d_j, given that d_j reaches neither x, nor Y, nor an earlier vertex of the list; the conditioning event is assumed to have positive probability.

Every hyperedge probability strictly between 0 and 1 makes all of these events have positive probability, which is the situation of Theorem 4.9. The set avoided in (27) grows along the list, so the constants are computed in a definite order and ρ_j depends on $d_1, \ldots, d_{j-1}$. This ordering is not a convenience; it is what makes the induction close.

Given these vertices and their constants, define a recursive subtraction that removes one contribution at a time. The operation applies to an arbitrary real function of the vertices, not only to κ; this generality is used in Section 5.

Definition 4.7 (Recursive Subtraction and the Functional M) Let $\rho_1, \dots, \rho_\ell$ be as in (27) and let λ be as in (24). For a real function q of the vertices let $q^{[0]} = q$ and

$$q^{[j]}(u) = q^{[j-1]}(u) - \rho_j(u) q^{[j-1]}(d_j) \qquad (1 \le j \le \ell)\,, \tag{28}$$

and let

$$M[q] = q^{[\ell]}(o) - \lambda q^{[\ell]}(v)\,. \tag{29}$$

The j-th step subtracts from the value at u the value at d_j, weighted by the conditional connection probability $\rho_j(u)$. Both operations are linear.

Lemma 4.8. (Recursive Subtraction and M Are Linear)

(i) For each j with $0 \le j \le \ell$, the map $q \mapsto q^{[j]}$ is linear.
(ii) The map $q \mapsto M[q]$ is linear.

Proof. We induct on j. For $j = 0$ the map is the identity. For the step, $q^{[j]}$ is obtained from $q^{[j-1]}$ by subtracting the number $q^{[j-1]}(d_j)$ times the fixed function ρ_j, and $\rho_1, \dots, \rho_\ell$ do not depend on q; a difference of linear maps is linear. The functional M is a difference of two evaluations of $q^{[\ell]}$, so it is linear as well. □

4.4. **Conditional covariance inequalities with auxiliary vertices.** Recall that κ depends on the increasing function f through (23).

Theorem 4.9. (Conditional Covariance Inequalities) *Let every hyperedge probability lie strictly between 0 and 1. Let x be a vertex, let Y be a set of vertices with $x \notin Y$, let o and v be distinct vertices outside $Y \cup \{x\}$, and let $d_1, \dots, d_\ell$ be distinct vertices, none equal to x, to o or to v, and none in Y. Then for every increasing real function f of $\mathcal{K}_x$,*

$$M[\kappa] \ge 0\,. \tag{30}$$

Written out, the case $\ell = 0$ of (30) is (25), and the case $\ell = 1$ reads

$$\kappa(o) - \rho_1(o)\kappa(d_1) \ge \lambda\big(\kappa(v) - \rho_1(v)\kappa(d_1)\big)\,.$$

Each successive recursive step compares the same two vertices after removing one more contribution from $d_1, \dots, d_\ell$.

The proof uses the six results of Section 2.6 together with the preceding definitions and identities.

Proof. Since every hyperedge probability is less than one, isolating each named vertex has positive probability. Thus every conditional probability used below has a positive denominator. Adding a constant to f changes no covariance, so we may and do suppose that f is nonnegative.

For a vertex function h, write

$$(L_j h)(u) = h(u) - \rho_j(u) h(d_j)\,.$$

Then $h^{[j]} = L_j \cdots L_1 h$. We prove the theorem by strong induction on ℓ. The argument below includes the case $\ell = 0$, for which every sum indexed by j is empty.

Step 1: Remove the conditioning cluster. Let

$$\nu = \mathbb{P}(\cdot \mid x \nleftrightarrow Y)\,.$$

Under ν, expose $\mathcal{K}_Y$ and let

$$W = U \setminus \mathcal{C}_Y\,.$$

This set contains x. Lemma 2.12 says that, conditional on what was exposed, the hyperedges of the model induced on W have their original independent laws. Let π be the law of W under ν. For an increasing function g of $\mathcal{K}_x$, define

$$\kappa_g^W(u) = \mathrm{Cov}_W\big(g(\mathcal{K}_x^W), \mathbf{1}\{u \in \mathcal{C}_x^W\}\big)$$

and

$$\Delta_\ell(g) = \mathbb{E}_\pi\left[(L_\ell \cdots L_1 \kappa_g^W)(o) - \lambda(L_\ell \cdots L_1 \kappa_g^W)(v)\right] .$$

The constants ρ_j and λ in this formula are those of the theorem.

We next show that it is enough to prove

$$\Delta_\ell(g) \geq 0 \tag{31}$$

for every nonnegative increasing g. By linearity there are constants β_u, indexed by o, v and $d_1, \ldots, d_\ell$, such that

$$M[h] = \sum_u \beta_u h(u) .$$

For a set A of hyperedges, set

$$H(A) = \sum_u \beta_u \mathbf{1}\{u \in \mathcal{C}_x(A)\} ,$$

where $\mathcal{C}_x(A)$ is the component of x in the open hyperedges of A. If

$$\mathfrak{M}_\ell(g) = M[\kappa_g] ,$$

then

$$\mathfrak{M}_\ell(g) = \mathrm{Cov}_\nu(g(\mathcal{K}_x), H(\mathcal{K}_x)) .$$

Write $A = \mathcal{K}_x$ and $B = \mathcal{K}_Y$ under ν. Let P_A and P_B denote conditional expectation given A and B, respectively, and define the two-step resampling operator

$$\mathcal{A} = P_A P_B \qquad (\mathcal{A}g)(A) = \mathbb{E}_\nu\left[\mathbb{E}_\nu[g(A) \mid B] \mid A\right] .$$

The law of total covariance, first conditional on B, gives

$$\mathfrak{M}_\ell(g) = \Delta_\ell(g) + \mathfrak{M}_\ell(\mathcal{A}g) . \tag{32}$$

Here is the calculation. Let $h(B) = \mathbb{E}_\nu[g(A) \mid B]$. The first term in the law of total covariance is

$$\mathbb{E}_\nu[\mathrm{Cov}_\nu(g(A), H(A) \mid B)] = \Delta_\ell(g)$$

by Lemma 2.12. Since $H(A)$ is measurable with respect to A, the second term is

$$\begin{aligned}\mathrm{Cov}_\nu(h(B), \mathbb{E}_\nu[H(A) \mid B]) &= \mathrm{Cov}_\nu(h(B), H(A)) \\ &= \mathrm{Cov}_\nu(\mathbb{E}_\nu[h(B) \mid A], H(A)) = \mathfrak{M}_\ell(\mathcal{A}g) .\end{aligned}$$

We record two properties of $\mathcal{A}$. Given $B = J$, exact cluster conditioning makes the x-cluster an independent cluster in the graph obtained by deleting the support of J. Hence

$$J \longmapsto \mathbb{E}_\nu[g(A) \mid B = J]$$

is decreasing when g is increasing. Given $A = I$, the Y-cluster has the product law off the support of I. If $I_1 \subseteq I_2$, then couple the two laws using the same hyperedge states. The set $\mathcal{K}_Y$ obtained after deleting the larger support is contained in the other one. Applying the preceding decreasing conditional mean shows

$$(\mathcal{A}g)(I_1) \leq (\mathcal{A}g)(I_2) .$$

Thus $\mathcal{A}$ preserves increasing functions. A function defined only on the values it can take is extended, when needed, by

$$\overline{g}(J) = \max\bigl(\{g(I) : I \subseteq J \text{ and } g \text{ is defined at } I\} \cup \{0\}\bigr) .$$

This agrees with g wherever g was defined and is increasing on the entire Boolean lattice. The operator $\mathcal{A}$ also preserves nonnegativity and constants.

There is a number $\epsilon > 0$ such that the transition represented by $\mathcal{A}$ has a fixed component of mass at least ϵ. Indeed, during the resampling of B, close every still available edge incident to Y. This has probability at least

$$\epsilon = \prod_{e:\iota(e)\cap Y\neq\varnothing}(1-p_e) > 0$$

and forces $\mathcal{C}_Y = Y$. The subsequent resampling of A then has a fixed law, independent of its previous value. Consequently

$$\operatorname{osc}(\mathcal{A}^n g) \leq (1-\epsilon)^n \operatorname{osc}(g)\,. \tag{33}$$

If $Y = \varnothing$, then the first application of $\mathcal{A}$ is already constant. Otherwise, every averaging operator raises the minimum and lowers the maximum. Equation (33) shows that these bounds approach the same number, so $\mathcal{A}^n g$ converges uniformly to a constant.

Iterating (32) gives

$$\mathfrak{M}_\ell(f) = \sum_{i=0}^{n-1}\Delta_\ell(\mathcal{A}^i f) + \mathfrak{M}_\ell(\mathcal{A}^n f)\,.$$

The last term tends to zero because $\mathfrak{M}_\ell$ is a linear functional on a finite state space and vanishes on constants. Therefore (31) for nonnegative increasing functions implies $\mathfrak{M}_\ell(f) \geq 0$.

Step 2: Add the vertices d_j one at a time. Fix a nonnegative increasing function g, and let

$$S_j = \{x, d_1, \ldots, d_j\} \qquad S_0 = \{x\} \qquad X = S_\ell\,.$$

For $T \subseteq U$ and a vertex u, define

$$r_T^W(u) = \operatorname{Cov}_W\big(g(\mathcal{K}_x^W), \mathbf{1}\{u \leftrightarrow T\}\big) \qquad R_j(u) = \mathbb{E}_\pi[r_{S_j}^W(u)]\,.$$

Thus

$$R_0(u) = \mathbb{E}_\pi[\kappa_g^W(u)]$$

and, with

$$h_u(W) = \operatorname{Cov}_W\big(g(\mathcal{K}_x^W), \mathbf{1}\{u \leftrightarrow X\}\big)\,,$$

we have $R_\ell(u) = \mathbb{E}_\pi[h_u(W)]$.

Use the deletion notation of Lemma 2.20 and define

$$\Phi(K) = \mathbb{E}\left[\mathbf{1}\{\mathcal{C}_x^{-K}\cap Y = \varnothing\}\left(g(\mathcal{K}_x^{-\mathcal{C}_Y^{-K}}) - g(\mathcal{K}_x^{-K})\right)\right]\,. \tag{34}$$

We need the fact that Φ is nonnegative and increasing on sets that avoid Y. This follows pointwise. For a fixed configuration, write

$$A_K = \{\mathcal{C}_x^{-K}\cap Y = \varnothing\} \qquad U_K = \mathcal{K}_x^{-\mathcal{C}_Y^{-K}} \qquad V_K = \mathcal{K}_x^{-K}\,.$$

On A_K, the x-cluster and the Y-cluster in the graph with K deleted are disjoint. Restoring the edges meeting K and then deleting the edges meeting $\mathcal{C}_Y^{-K}$ cannot remove an edge of the former x-cluster. Thus

$$V_K \subseteq U_K \quad \text{on the event } A_K\,. \tag{35}$$

The integrand in (34) is therefore nonnegative.

Now let $K \subseteq N$ and suppose $N \cap Y = \varnothing$. Deleting more vertices gives

$$\mathcal{C}_Y^{-N} \subseteq \mathcal{C}_Y^{-K} \qquad V_N \subseteq V_K \qquad U_K \subseteq U_N\,,$$

and $A_K \subseteq A_N$. On A_K, monotonicity of g gives

$$g(U_K) - g(V_K) \leq g(U_N) - g(V_N)\,.$$

On $A_N \setminus A_K$, the expression on the right is nonnegative by (35). Hence $\Phi(K) \leq \Phi(N)$. Extend Φ to all subsets of U by

$$\overline{\Phi}(K) = \max\{\Phi(I) : I \subseteq K,\ I \cap Y = \varnothing\}\,.$$

The extension is nonnegative and increasing, and it agrees with Φ whenever $K \cap Y = \varnothing$. Every cluster to which it is applied below avoids Y, so we write Φ for the extension.

For $1 \leq j \leq \ell$, set

$$Y_j = Y \cup \{x, d_1, \ldots, d_{j-1}\} \qquad \alpha_j = \frac{\mathbb{P}(d_j \not\leftrightarrow Y_j)}{\mathbb{P}(x \not\leftrightarrow Y)}$$

and define

$$\eta_j(u) = \operatorname{Cov}\left(\Phi(\mathcal{C}_{d_j}), \mathbf{1}\{u \in \mathcal{C}_{d_j}\} \;\middle|\; d_j \not\leftrightarrow Y_j\right) .$$

Apply Lemma 2.20 with $z = d_j$ and $S = S_{j-1}$. Its constant $\rho(u)$ is precisely $\rho_j(u)$, and it gives

$$L_j R_{j-1} - R_j = \alpha_j \eta_j . \tag{36}$$

Set

$$D_j = L_j \cdots L_1 R_0 - R_j \qquad D_0 = 0 .$$

Equation (36) gives

$$D_j = L_j D_{j-1} + \alpha_j \eta_j .$$

Iteration yields

$$D_\ell = \sum_{j=1}^{\ell} \alpha_j L_\ell L_{\ell-1} \cdots L_{j+1} \eta_j ,$$

where the operator following α_ℓ is the identity. Since the operators L_j commute with expectation,

$$\begin{aligned} \Delta_\ell(g) &= \mathbb{E}_\pi[h_o(W) - \lambda h_v(W)] \\ &\quad + \sum_{j=1}^{\ell} \alpha_j \left[(L_\ell \cdots L_{j+1}\eta_j)(o) - \lambda(L_\ell \cdots L_{j+1}\eta_j)(v)\right] . \end{aligned} \tag{37}$$

The expression in square brackets is an instance of the theorem with $\ell - j$ vertices. Its distinguished vertex is d_j, its avoided set is Y_j, and its ordered list is $d_{j+1}, \ldots, d_\ell$. The constants in its successive operators are the original $\rho_{j+1}, \ldots, \rho_\ell$, because

$$Y_j \cup \{d_j, d_{j+1}, \ldots, d_{\ell-1}\} = Y \cup \{x, d_1, \ldots, d_{\ell-1}\} .$$

Its final connection constant is λ, since its final avoided set is

$$Y_j \cup \{d_j, d_{j+1}, \ldots, d_\ell\} = Y \cup X .$$

The function $\Phi(\mathcal{C}_{d_j})$ is a nonnegative increasing function of the open hyperedges of the cluster of d_j. The strong induction hypothesis therefore says that every term in the sum in (37) is nonnegative.

Step 3: Compare o and v in each induced hypergraph. For a fixed W, define

$$q(W) = \begin{cases} \mathbb{P}_W(o \leftrightarrow v \mid v \not\leftrightarrow X), & v \in W, \\ 0, & v \notin W . \end{cases}$$

We first prove

$$h_o(W) \geq q(W) h_v(W) . \tag{38}$$

If $v \notin W$, then $q(W) = h_v(W) = 0$ and the assertion is the Harris inequality for $h_o(W)$. If $o \notin W$, then both $q(W)$ and $h_o(W)$ are zero. We may therefore suppose $o, v \in W$.

Let

$$m = \mathbb{E}_W[g(\mathcal{K}_x^W)] \qquad D = \{v \not\leftrightarrow X\} \qquad J = \{o \leftrightarrow v,\ v \not\leftrightarrow X\} .$$

The disjoint union

$$\{o \leftrightarrow X \cup \{v\}\} = \{o \leftrightarrow X\} \,\dot\cup\, J$$

and the Harris inequality give

$$h_o(W) \geq -\mathbb{E}_W[(g(\mathcal{K}_x^W) - m); J] . \tag{39}$$

Conditional on D, apply Lemma 2.15 to the source sets X and $\{v\}$. The function $g(\mathcal{K}_x^W)$ is increasing in $\mathcal{K}_X^W$, and $1-\mathbf{1}\{o\in\mathcal{C}_v^W\}$ is decreasing in $\mathcal{K}_v^W$. Their positive correlation is equivalent to

$$\mathbb{P}_W(D)\mathbb{E}_W[(g(\mathcal{K}_x^W)-m);J]\le\mathbb{P}_W(J)\mathbb{E}_W[(g(\mathcal{K}_x^W)-m);D]\,.$$

Since

$$h_v(W)=-\mathbb{E}_W[(g(\mathcal{K}_x^W)-m);D]\qquad q(W)=\frac{\mathbb{P}_W(J)}{\mathbb{P}_W(D)}\,,$$

this inequality and (39) prove (38).

It remains to compare the varying number $q(W)$ with λ. Apply Lemma 2.18 with $A=X$ and $F(W)=h_v(W)$. The Harris inequality gives $F(W)\ge 0$, and if $v\notin W$ then $F(W)=0$. Lemma 2.17, with source $S=X$ and cluster function g, is exactly the assumption (11).

The conclusion (12) gives

$$\begin{aligned}&\mathbb{P}(o\leftrightarrow v,\ v\nleftrightarrow X\cup Y)\mathbb{E}[h_v(U\setminus\mathcal{C}_Y);x\notin\mathcal{C}_Y]\\&\qquad\le\mathbb{P}(v\nleftrightarrow X\cup Y)\mathbb{E}[q(U\setminus\mathcal{C}_Y)h_v(U\setminus\mathcal{C}_Y);x\notin\mathcal{C}_Y]\,.\end{aligned}$$

The first probability is λ times the second. The second is positive because v can be isolated, so cancellation gives

$$\mathbb{E}[q(W)h_v(W);x\nleftrightarrow Y]\ge\lambda\mathbb{E}[h_v(W);x\nleftrightarrow Y]\,. \tag{40}$$

Divide by $\mathbb{P}(x\nleftrightarrow Y)$ and average (38). Equation (40) then shows that

$$\mathbb{E}_\pi[h_o(W)-\lambda h_v(W)]\ge 0\,.$$

This proves that the first term in (37) is nonnegative. Step 2 and the induction hypothesis show that all its other terms are nonnegative as well. Hence $\Delta_\ell(g)\ge 0$ for every nonnegative increasing g. Step 1 now gives $M[\kappa]=\mathfrak{M}_\ell(f)\ge 0$, as required. ☐

4.5. **Reduction to indicator functions.** One structural consequence of Theorem 4.9 is used repeatedly in Section 5, and it follows from linearity alone.

Proposition 4.10. (Linearity in f) *Fix the model, the vertices x, o and v, the set Y and the list $d_1,\dots,d_\ell$. Then the following hold.*

- *(i) The quantity $M[\kappa]$ is a linear function of f.*
- *(ii) If f is constant, then $M[\kappa]=0$.*

Proof. The map sending f to κ is linear, because a covariance is linear in each of its arguments and the second argument of (23) does not involve f. The map sending κ to $M[\kappa]$ is linear by Lemma 4.8, since λ and the constants ρ_j do not depend on f. A composition of linear maps is linear.

If f is constant then $f(\mathcal{K}_x)$ is a constant random variable, so every covariance (23) vanishes, hence κ is identically zero and so is the displayed quantity. ☐

A consequence of Proposition 4.10 is that it suffices to verify (30) when f is the indicator of an increasing family of sets of hyperedges, since every increasing function on a finite family of sets is a nonnegative combination of such indicators together with a constant. For fixed data Theorem 4.9 is therefore a finite family of inequalities, one for each such indicator. The quantities appearing in them are rational functions of the hyperedge probabilities rather than polynomials, since λ and the constants ρ_j are ratios of two probabilities; clearing the positive denominators turns each inequality into a polynomial one.

5. The Kozma–Nitzan Inequality for Independent Hyperedges

Section 4 produced a family of inequalities comparing conditional covariances at two vertices. This section converts that family into the Kozma–Nitzan inequality stated in Section 3.

Order the intermediate vertices by their conditional expectations m_x. In the application, m_x is the probability that x reaches the target. In each configuration where the source reaches the set, select the reached vertex with the smallest order value. The events selecting the individual vertices partition the event that the source reaches the set. This replaces the size-dependent union bound of Lemma 3.1 by a single expectation. The argument directly gives the Kozma–Nitzan inequality; the additive and uniform near-one statements follow from it.

Section 5.1 defines the ordering and the quantity $D_u(T)$. Section 5.2 compares this quantity at two vertices. Section 5.3 iterates that comparison over the set and explains why the auxiliary vertices of Section 4 are needed for the iteration to close. Section 5.4 completes the derivation, and Section 5.5 gives the bond-percolation consequence.

5.1. **Ordering the intermediate vertices.** Throughout this section the model is the one of Definition 2.7, Y is a set of vertices, F is an increasing real function of sets of vertices, and T is a finite set of vertices disjoint from Y with $\mathbb{P}(\{x \nleftrightarrow Y\}) > 0$ for every $x \in T$. Including the set Y is necessary for the induction below to close. For $x \in T$ write

$$m_x = \frac{\mathbb{E}\big[F(\mathcal{C}_x); \{x \nleftrightarrow Y\}\big]}{\mathbb{P}(\{x \nleftrightarrow Y\})} \tag{41}$$

for the conditional expectation of $F(\mathcal{C}_x)$ given that x avoids Y.

Definition 5.1 (An Ordering Compatible with the Means) A function r from the vertices to the real numbers orders T compatibly with the means when it is injective on T and $m_x \le m_y$ for all $x, y \in T$ with $r(x) < r(y)$.

Such a function exists: order T by the values m_x and break ties arbitrarily.

Definition 5.2 (The Events J_x and the Quantity $D_u(T)$) Let r satisfy Definition 5.1, and let u be a vertex. For $x \in T$ let

$$J_x = \{u \nleftrightarrow Y\} \cap \{u \leftrightarrow x\} \cap \bigcap_{y \in T,\ r(y) < r(x)} \{u \nleftrightarrow y\},$$

the event that u avoids Y and that x has the smallest r-value in $T \cap \mathcal{C}_u$, and set

$$D_u(T) = \mathbb{E}\big[F(\mathcal{C}_u); \{u \nleftrightarrow Y\},\ u \leftrightarrow T\big] - \sum_{x \in T} \mathbb{P}(J_x) m_x\,. \tag{42}$$

Lemma 5.3. (The Events J_x Partition, and $D_u(T)$ Is Independent of the Ordering)

(i) *The events J_x for $x \in T$ are pairwise disjoint and their union is $\{u \nleftrightarrow Y\} \cap \{u \leftrightarrow T\}$.*

(ii) *The quantity $D_u(T)$ satisfies*

$$D_u(T) = \mathbb{E}\Big[\mathbf{1}\{\{u \nleftrightarrow Y\},\ u \leftrightarrow T\}\Big(F(\mathcal{C}_u) - \min_{x \in T \cap \mathcal{C}_u} m_x\Big)\Big], \tag{43}$$

so $D_u(T)$ does not depend on the choice of r satisfying Definition 5.1.

Proof. If J_x and J_y both occur with $x \neq y$, then we may assume $r(x) < r(y)$ since r is injective on T; then J_y asserts $u \nleftrightarrow x$ while J_x asserts $u \leftrightarrow x$, a contradiction. If u avoids Y and reaches T, then $T \cap \mathcal{C}_u$ is nonempty and finite, so it has a unique element with the smallest r-value, and J_x occurs for that element. Conversely each J_x is contained in $\{u \nleftrightarrow Y\} \cap \{u \leftrightarrow T\}$.

For (43), note that on J_x the vertex x has the smallest r-value among the vertices of T in $\mathcal{C}_u$, hence by compatibility the smallest value of m, so $m_x = \min_{y \in T \cap \mathcal{C}_u} m_y$ there. Summing over the partition turns $\sum_x \mathbb{P}(J_x)m_x$ into the expectation of that minimum over $\{u \nleftrightarrow Y\} \cap \{u \leftrightarrow T\}$, which gives (43). The right side of (43) does not mention r. ☐

Two cases are immediate and are used below. If $u \in Y$ then $\{u \nleftrightarrow Y\}$ is empty and $D_u(T) = 0$. If $u \in T$ then $u \in T \cap \mathcal{C}_u$ on $\{u \nleftrightarrow Y\}$, so the minimum in (43) is at most m_u, whence $D_u(T) \geq \mathbb{E}[F(\mathcal{C}_u) - m_u; \{u \nleftrightarrow Y\}] = 0$ by (41).

The next identity is what lets one vertex be removed from T.

Lemma 5.4. (Removing the Final Vertex in the Ordering) *Let T be nonempty, let $k \in T$ have the largest r-value, and let $T_0 = T \setminus \{k\}$. For a vertex u let*

$$p(u) = \mathbb{P}\big(\{k \nleftrightarrow Y \cup T_0\},\ u \leftrightarrow k\big) \qquad \mu(u) = \mathbb{E}\big[F(\mathcal{C}_k); \{k \nleftrightarrow Y \cup T_0\},\ u \leftrightarrow k\big]\,.$$

Then

$$D_u(T) = D_u(T_0) + \mu(u) - p(u)m_k\,. \tag{44}$$

Proof. Since k has the largest r-value, adjoining k to T_0 leaves each event J_x with $x \in T_0$ unchanged and creates one new event, J_k. On $\{u \leftrightarrow k\}$ the clusters of u and of k coincide, so on that event the conditions defining J_k read $\{k \nleftrightarrow Y \cup T_0\}$, and $F(\mathcal{C}_u) = F(\mathcal{C}_k)$ there. Comparing (42) for T and for T_0 therefore adds the term $\mu(u)$ to the expectation and subtracts $p(u)m_k$ from the sum. ☐

5.2. **Comparison at two vertices.** The following is where the family of inequalities of Section 4 enters.

Proposition 5.5. (Comparing $D_u(T)$ at Two Vertices) *In the setting above, let $v \notin Y \cup T$. Then for every vertex o,*

$$\mathbb{P}\big(\{v \nleftrightarrow Y \cup T\} \cap \{o \leftrightarrow v\}\big) D_v(T) \leq \mathbb{P}\big(\{v \nleftrightarrow Y \cup T\}\big) D_o(T)\,. \tag{45}$$

Proof. Assume first that every hyperedge probability lies strictly between 0 and 1, that $o \notin Y \cup T$ and that $o \neq v$. Let $d_1, \ldots, d_\ell$ be distinct vertices outside $Y \cup T \cup \{o, v\}$, and let M be the functional (29) formed from the set $Y \cup T$, that list, and the two vertices o and v. We prove by induction on the size of T, and simultaneously for every such list and every pair o and v, that

$$M\big[u \mapsto D_u(T)\big] \geq 0\,. \tag{46}$$

Step 1: Base case. If T is empty then $D_u(\varnothing) = 0$ for every u by (42), so (46) holds by Lemma 4.8.

Step 2: Inductive step. Let T be nonempty, let $k \in T$ have the largest r-value, let $T_0 = T \setminus \{k\}$, and let p and μ be as in Lemma 5.4. Connection is reflexive, so $p(k) = \mathbb{P}(\{k \nleftrightarrow Y \cup T_0\})$ and $\mu(k) = \mathbb{E}[F(\mathcal{C}_k); \{k \nleftrightarrow Y \cup T_0\}]$. Write

$$\chi(u) = p(k)\mu(u) - \mu(k)p(u) \qquad \gamma = m_k p(k) - \mu(k)\,.$$

Expanding both sides gives the identity

$$p(k)\big(\mu(u) - p(u)m_k\big) = \chi(u) - \gamma p(u)\,. \tag{47}$$

We show that $M[\chi] \geq 0$ and $M[p] \geq 0$. For the first, let κ be the conditional covariance (23) formed with the vertex k in place of x, the set $Y \cup T_0$ in place of Y, and the increasing function sending a set of hyperedges to the value of F at the vertex set it spans together with k. Because $(Y \cup T_0) \cup \{k\} = Y \cup T$, the constants (27) and the number (24) attached to this κ are the ones M was formed from. Then $\chi = p(k)^2\,\kappa$, so $M[\chi] = p(k)^2\,M[\kappa]$ by linearity, and $M[\kappa] \geq 0$ is Theorem 4.9. For the second, apply the same theorem with the increasing function that is 1 on nonempty sets of hyperedges and 0 on the empty one. Writing q_0 for the probability that k avoids $Y \cup T_0$ and lies on no open hyperedge, the corresponding covariance is $q_0\,p(u)$ up to the positive factor $p(k)^2$ at every vertex u other than k, which is all that M evaluates, so $q_0\,M[p] \geq 0$. Closing every hyperedge at k has positive probability, so $q_0 > 0$ and $M[p] \geq 0$.

Next we show that

$$\gamma \le D_k(T_0)\,. \tag{48}$$

By (41) we have $\mathbb{E}[F(\mathcal{C}_k) - m_k; \{k \nleftrightarrow Y\}] = 0$. Splitting $\{k \nleftrightarrow Y\}$ into the part where k avoids T_0 as well and the part where it reaches T_0 turns this into $\gamma = \mathbb{E}[F(\mathcal{C}_k) - m_k; \{k \nleftrightarrow Y\},\ k \leftrightarrow T_0]$. Expanding that restricted expectation over the partition of Lemma 5.3, taken with k in place of u, gives $\gamma = D_k(T_0) + \sum_{x\in T_0} \mathbb{P}(J_x)(m_x - m_k)$, and every summand is at most zero because $r(x) < r(k)$ and r is compatible.

Now apply M to (44), multiply by $p(k) > 0$, and use (47) together with linearity:

$$p(k)M\big[u \mapsto D_u(T)\big] = p(k)M\big[u \mapsto D_u(T_0)\big] + M[\chi] - \gamma M[p]\,.$$

The middle term is nonnegative. Since $M[p] \ge 0$, replacing γ by the larger quantity $D_k(T_0)$ of (48) only decreases the right side. Writing $p(u) = p(k)\,\rho(u)$ with $\rho(u) = \mathbb{P}(u \in \mathcal{C}_k \mid \{k \nleftrightarrow Y \cup T_0\})$, the resulting expression is $p(k)$ times the value of the functional obtained by inserting k, with the constant ρ, at the front of the list $d_1, \dots, d_\ell$, applied to $u \mapsto D_u(T_0)$; this is exactly the recursion (28). That value is nonnegative by the induction hypothesis, applied to the smaller set T_0 with the lengthened list. Dividing by $p(k)$ gives (46).

Step 3: From (46) *to the statement.* Take the empty list, so that $M[q] = q(o) - \lambda q(v)$ with $\lambda = \mathbb{P}(o \in \mathcal{C}_v \mid \{v \nleftrightarrow Y \cup T\})$. Then (46) reads $D_o(T) \ge \lambda D_v(T)$, and multiplying by $\mathbb{P}(\{v \nleftrightarrow Y \cup T\}) > 0$ gives (45). If $o = v$ then the two sides of (45) are equal. If o lies in Y or in T, then the event $\{v \nleftrightarrow Y \cup T\} \cap \{o \leftrightarrow v\}$ is empty, so the left side vanishes while the right side is nonnegative by the two cases recorded after Lemma 5.3.

Step 4: Parameters in $[0,1]$. By (43) each $D_u(T)$ is a finite sum over configurations, each summand being a polynomial in the hyperedge probabilities times a value depending on them only through the finitely many numbers m_x; and each m_x is a ratio of polynomials whose denominator is positive at the given parameters. Both sides of (45) are therefore continuous there. Approximate the given parameters by ones lying strictly inside $(0,1)$, choosing for each an ordering r satisfying Definition 5.1; by Lemma 5.3 the quantity $D_u(T)$ does not depend on that choice. The case already proved applies to each approximation, and continuity gives (45) in general. □

5.3. **Nonnegativity.**

Theorem 5.6. (Nonnegativity of $D_o(T)$) *In the setting of Section 5.1, $D_o(T) \ge 0$ for every vertex o.*

Proof. We induct on the size of T. If T is empty then $D_o(\varnothing) = 0$.

Let T be nonempty, let $k \in T$ have the largest r-value, let $T_0 = T \setminus \{k\}$, and keep p, μ and γ from the proof of Proposition 5.5. Identity (44) with $u = o$ reads $D_o(T) = D_o(T_0) + \mu(o) - m_k\, p(o)$.

Apply the conditional association of Corollary 2.14 with the vertex k, the set $Y \cup T_0$, and the two increasing functions of $\mathcal{K}_k$ given by F evaluated at the spanned vertex set together with k, and by the indicator that o lies in the cluster of k. This gives $p(o)\,\mu(k) \le p(k)\,\mu(o)$, which rearranges to

$$p(k)\big(m_k p(o) - \mu(o)\big) \le p(o)\big(m_k p(k) - \mu(k)\big) = p(o)\gamma\,.$$

By (48) and $p(o) \ge 0$ the right side is at most $p(o)\,D_k(T_0)$, and by Proposition 5.5 applied to T_0 with $v = k$ that is at most $p(k)\,D_o(T_0)$. Hence

$$p(k)D_o(T) = p(k)D_o(T_0) + p(k)\big(\mu(o) - m_k p(o)\big) \ge 0\,.$$

If $p(k) > 0$ then this gives $D_o(T) \ge 0$. If $p(k) = 0$ then $p(o) = 0$ and $\mu(o) = 0$, so $D_o(T) = D_o(T_0)$, which is nonnegative by the induction hypothesis. □

5.4. **The Kozma–Nitzan inequality.** We now prove the inequality stated as Conjecture 1 by Kozma and Nitzan [2024]. The proof selects the reached vertex with smallest r-value.

Theorem 5.7. (Kozma–Nitzan Conjecture 1) *Let o and b be vertices and let A be a nonempty finite set of vertices. Then*

$$\mathbb{P}(o \leftrightarrow b) \geq \mathbb{P}(o \leftrightarrow A) \cdot \min_{a \in A} \mathbb{P}(a \leftrightarrow b)\,.$$

Proof. We apply Theorem 5.6 to one particular function and then estimate the resulting sum from below. The proof has three steps.

Step 1: Choose the data. Take $Y = \varnothing$ and $T = A$ in Section 5.1, and let $F(S) = \mathbf{1}\{b \in S\}$ for a set S of vertices. This function is nonnegative, and it is increasing because enlarging a set can only help it contain b. Its mean at a vertex a is

$$m_a = \mathbb{E}\big[\mathbf{1}\{b \in \mathcal{C}_a\}\big] = \mathbb{P}(a \leftrightarrow b)\,.$$

Choose an ordering r for A satisfying Definition 5.1; such an ordering exists by the paragraph following that definition.

Step 2: Apply the nonnegativity. Theorem 5.6 applied with $T = A$ gives $D_o(A) \geq 0$, which by Definition 5.2 reads

$$\sum_{a \in A} \mathbb{P}(J_a) m_a \leq \mathbb{E}\big[F(\mathcal{C}_o); o \leftrightarrow A\big]\,. \tag{49}$$

The right side of (49) equals $\mathbb{P}(o \leftrightarrow b \text{ and } o \leftrightarrow A)$, since $F(\mathcal{C}_o)$ is the indicator that o is connected to b.

Step 3: Estimate the left side and conclude. Bounding every mean below by the smallest one,

$$\sum_{a \in A} \mathbb{P}(J_a) m_a \geq \Big(\min_{a \in A} m_a\Big) \sum_{a \in A} \mathbb{P}(J_a) = \Big(\min_{a \in A} \mathbb{P}(a \leftrightarrow b)\Big) \mathbb{P}(o \leftrightarrow A)\,,$$

where the last equality is Lemma 5.3. Combining this with (49) and with $\mathbb{P}(o \leftrightarrow b \text{ and } o \leftrightarrow A) \leq \mathbb{P}(o \leftrightarrow b)$ gives the claim. ☐

Step 3 is where the argument does not degrade, for the following reason. The numbers $\mathbb{P}(J_a)$ sum to $\mathbb{P}(o \leftrightarrow A)$ and not to $|A|$ times anything, because Lemma 5.3 makes them a partition. A union bound over the vertices of A would instead produce $\sum_a \mathbb{P}(a \nleftrightarrow b)$, with one term for each vertex of A and no cancellation. Selecting the reached vertex with smallest r-value converts a sum over the vertices of A into a single expectation.

The other two forms follow from Lemmas 3.5 and 3.6.

Corollary 5.8. (The Additive and Uniform Near-One Consequences) *The additive bound of Definition 3.3 and the uniform near-one statement of Definition 3.4 both hold.*

Proof. The inequality of Definition 3.2 is Theorem 5.7. The additive bound follows by Lemma 3.5, and the uniform near-one statement then follows by Lemma 3.6. ☐

5.5. **Bond-percolation criticality.** The preceding finite inequalities give the bond-percolation theorem.

Theorem 5.9. (The Percolation Probability Vanishes at the Critical Probability) *For every integer $d \geq 2$, Bernoulli bond percolation on $\mathbb{Z}^d$ satisfies $\theta(p_c(d)) = 0$. Equivalently, the percolation probability is continuous on the whole parameter interval.*

Proof. Section 5.2 proved Proposition 5.5 from Theorem 4.9, and Section 5.3 proved Theorem 5.6 from that. Theorem 5.7 and Corollary 5.8 then give all three finite inequalities, and the additive bound gives the conclusion by Corollary 3.8. The equivalence with continuity is Theorem 1.15. ☐

The two-dimensional result was known from Harris [1960] and Kesten [1980], and dimensions eleven and above from Hara and Slade [1990] and Fitzner and van der Hofstad [2017]. The argument above covers the previously unresolved range $3 \le d \le 10$ and also applies in the other dimensions.

6. Bond and Site Percolation

The preceding finite results concern independent labeled hyperedges. This section identifies bond and site percolation as two specializations, records the bond lattice consequence, and states how the companion site argument uses the finite inequality.

For bond percolation, the earlier conditional covariance argument established an additive bound, and hence a uniform near-one bound, sufficient for the lattice reduction. The present treatment extends that argument to labeled hyperedges, including the use of $\mathcal{K}_S$ to retain hyperedge identities, and applies the resulting finite inequalities to site percolation.

Section 6.1 gives the two identifications and shows why the one for site percolation needs an extra vertex on each edge, and Section 6.2 records the bond consequence and summarizes the endpoint of the separate site-percolation block argument.

6.1. **Bond and site percolation as special cases.** **Example 6.1** (Bond and Site Percolation)

Bond percolation on a graph G is the case of Definition 2.7 in which every incidence set is the pair of endpoints of an edge.

Site percolation on G needs one extra vertex for each edge, and we take G to have no isolated vertex. Take as vertices those of G together with a new vertex m_e for every edge e. Take one hyperedge for each vertex of G, the one at u having incidence set $\{u\} \cup \{ m_e : e \ni u \}$ and being open with the probability that u is open. Since u lies on an edge, that incidence set has at least two elements, as Definition 2.7 requires. Two distinct vertices of G are connected in this model exactly when they are connected in site percolation on G.

Only distinct vertices are covered, and they have to be: connection in Definition 2.7 is reflexive, so u is connected to itself whether or not u is open. For distinct u and v the claim is checked in one step each way. If u and v are neighbors in G and both open, then the two hyperedges are open and share m_e, so u and v are connected; concatenating along an open path of G gives the general case. Conversely a chain of open hyperedges from u to v passes through added vertices only, and m_e lies on the hyperedges of the two endpoints of e and on no others, so each such passage crosses an edge of G with both endpoints open.

The extra vertices are needed. Giving the hyperedge of u the incidence set $\{u\}$ together with the neighbors of u does not work: that hyperedge alone would make u adjacent to a neighbor whose own hyperedge is closed, and on a star it would make distinct neighbors of the center adjacent to one another.

Site percolation is not itself a case of bond percolation. Gladkov and Zimin [2026] proved that bond percolation cannot simulate site percolation even approximately, already near a vertex of degree three. In the other direction, bond percolation on a graph is site percolation on its line graph, an observation of Fisher and of Fisher and Essam recorded in Kesten [1982, §2.5], where the line graph is called the covering graph. Thus the two models have different direct representations; the labeled-hyperedge formulation provides a common setting for the argument used here.

6.2. **The finite inequality and site-percolation criticality.**

Specializing to the two models of Example 6.1 gives the finite inequality for both, from one proof.

Corollary 6.2. (The Kozma–Nitzan Inequality for Bond and Site Models) *Suppose the inequality of Definition 3.2 holds in the model of Definition 2.7. Then the following hold.*

(i) The inequality holds for bond percolation on every finite graph with independent edge parameters.

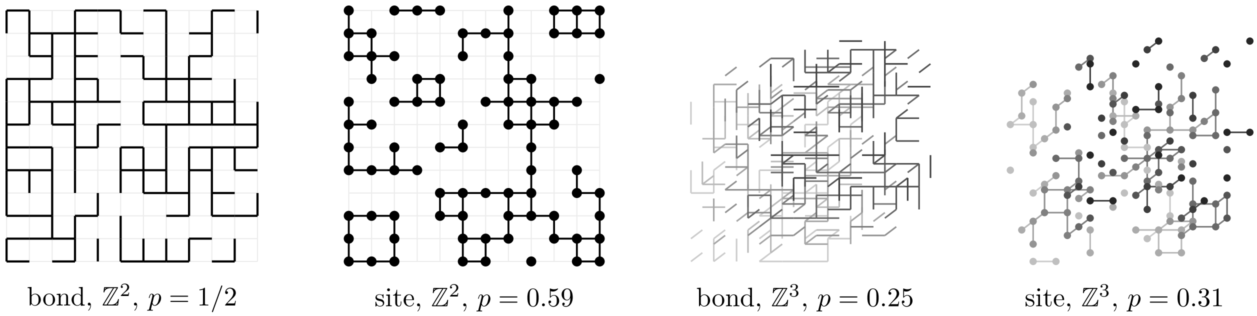


FIGURE 6. The four models these notes cover, each near its critical parameter. Bond percolation opens edges, site percolation opens vertices and joins two open neighbors. In three dimensions the pictures are drawn in projection, with depth shown by shading. The critical parameter is exactly $1/2$ for bond percolation in two dimensions; the other three values are not known in closed form, and the numbers used here are the usual numerical estimates.

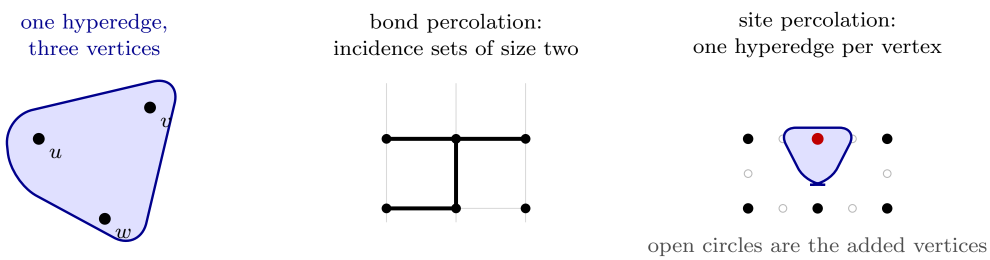


FIGURE 7. A hyperedge and the two models. On the left, a hyperedge incident to three vertices makes all three mutually adjacent when it is open. In the middle, incidence sets of size two give bond percolation. On the right, the hyperedge of a vertex of G reaches that vertex and the added vertex on each incident edge, so two vertices of G are joined exactly when both hyperedges are open.

(ii) The inequality holds for site percolation on every finite graph with no isolated vertex and independent vertex parameters, whenever o and b are distinct and neither of them lies in A.

Proof. Bond percolation is the instance in which every incidence set has two elements, so the first statement is the specialization of the inequality to that instance.

For the second, take the instance of Example 6.1. The quantities the inequality compares are $\mathbb{P}(o \leftrightarrow b)$, $\mathbb{P}(o \leftrightarrow A)$ and $\mathbb{P}(a \leftrightarrow b)$ for $a \in A$, and under the stated hypotheses each of them compares two distinct vertices of G. By Example 6.1 each is therefore the same number in the two models, so the inequality for the instance is the inequality for site percolation. □

For bond percolation the passage from the finite inequality to the lattice is Theorem 3.7: the additive bound follows from the Kozma–Nitzan inequality by Lemma 3.5, and the conclusion is Corollary 3.8. This is Theorem 5.9.

The companion manuscript gives the full site-specific finite block construction. Its endpoint is summarized here. Starting from $\theta_{\text{site}}(p) > 0$, it fixes the target, corridor, and stopped exploration data at p. Each required probability concerns finitely many sites and exceeds its threshold strictly. A common parameter $q < p$ can therefore be chosen so that every one of these finite inequalities remains valid. Running the same finite block exploration at q produces a percolating planar comparison process, and the recorded open-path witnesses lift that process to site percolation on $\mathbb{Z}^d$. Thus, if $\theta_{\text{site}}(p) > 0$, then there is a parameter $q < p$ with $\theta_{\text{site}}(q) > 0$. Applying this implication at $p = p_c$ contradicts the definition of the critical probability. No half-space criticality theorem is used in this endpoint.

The accompanying Lean development proves, for $d \geq 3$, the declaration

`KNAll.Site.site_no_percolation_at_critical`.

Its only arguments are d and a proof of $3 \leq d$. The present implementation uses a third coordinate to map the thin region produced by the exploration into the standard slab family. For $d = 2$, continuity of the site-percolation probability, and hence its vanishing at the critical probability, is the separate classical result of Russo [1981]; it is not part of the Lean declaration above.

References

Rudolf Ahlswede and David E. Daykin. An inequality for the weights of two families of sets, their unions and intersections. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 43 (3):183–185, 1978. ISSN 1432-2064. doi: 10.1007/bf00536201. URL http://dx.doi.org/10.1007/bf00536201.

M. Aizenman, H. Kesten, and C. M. Newman. Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. *Communications in Mathematical Physics*, 111:505–531, 1987. doi: 10.1007/BF01219071.

Michael Aizenman and David J. Barsky. Sharpness of the phase transition in percolation models. *Communications in Mathematical Physics*, 108(3):489–526, 1987. doi: 10.1007/BF01212322.

Michael Aizenman and Geoffrey Grimmett. Strict monotonicity for critical points in percolation and ferromagnetic models. *Journal of Statistical Physics*, 63(5-6):817–835, June 1991. ISSN 1572-9613. doi: 10.1007/bf01029985. URL http://dx.doi.org/10.1007/bf01029985.

David J. Barsky, Geoffrey R. Grimmett, and Charles M. Newman. Percolation in half-spaces: equality of critical densities and continuity of the percolation probability. *Probability Theory and Related Fields*, 90(1):111–148, Mar 1991. ISSN 1432-2064. doi: 10.1007/bf01321136. URL http://dx.doi.org/10.1007/bf01321136.

Béla Bollobás and Oliver Riordan. *Percolation*. Cambridge University Press, Cambridge, 2006. doi: 10.1017/CBO9781139167383.

S. R. Broadbent and J. M. Hammersley. Percolation processes. I. Crystals and mazes. *Proceedings of the Cambridge Philosophical Society*, 53:629–641, 1957. doi: 10.1017/S0305004100032680.

R. M. Burton and M. Keane. Density and uniqueness in percolation. *Communications in Mathematical Physics*, 121(3):501–505, Sept 1989. ISSN 1432-0916. doi: 10.1007/bf01217735. URL http://dx.doi.org/10.1007/bf01217735.

Raphaël Cerf. $\theta(p_c, \mathbb{Z}^d) = 0$?, 2026. Preprint, arXiv:2608.23661.

Hugo Duminil-Copin. Sixty years of percolation. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited lectures*, pages 2829–2856. World Scientific, Hackensack, NJ, 2018. doi: 10.1142/9789813272880_0162.

Hugo Duminil-Copin and Vincent Tassion. A new proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model. *Communications in Mathematical Physics*, 343 (2):725–745, 2016a. doi: 10.1007/s00220-015-2480-z.

Hugo Duminil-Copin and Vincent Tassion. A new proof of the sharpness of the phase transition for Bernoulli percolation on $\mathbb{Z}^d$. *L'Enseignement Mathématique*, 62(1-2):199–206, 2016b. doi: 10.4171/lem/62-1/2-12.

Hugo Duminil-Copin, Vladas Sidoravicius, and Vincent Tassion. Absence of infinite cluster for critical Bernoulli percolation on slabs. *Communications on Pure and Applied Mathematics*, 69(7):1397–1411, 2016. doi: 10.1002/cpa.21641. URL https://arxiv.org/abs/1401.7130. arXiv:1401.7130.

Robert Fitzner and Remco van der Hofstad. Mean-field behavior for nearest-neighbor percolation in $d > 10$. *Electronic Journal of Probability*, 22(43):1–65, 2017. doi: 10.1214/17-EJP56.

C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22:89–103, 1971. doi: 10.1007/BF01651330.

Sacha Friedli and Yvan Velenik. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press, Cambridge, 2017. doi: 10.1017/9781316882603.

Nikita Gladkov. Percolation Inequalities and Decision Trees, 2024. Preprint, arXiv:2408.08457.

Nikita Gladkov and Aleksandr Zimin. Bond percolation does not simulate site percolation. *Electronic Communications in Probability*, 2026. doi: 10.1214/26-ecp760.

Geoffrey Grimmett. *Percolation*, volume 321 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2 edition, 1999. doi: 10.1007/978-3-662-03981-6.

Geoffrey Richard Grimmett and J. M. Marstrand. The supercritical phase of percolation is well behaved. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 430(1879):439–457, Aug 1990. ISSN 2053-9177. doi: 10.1098/rspa.1990.0100. URL http://dx.doi.org/10.1098/rspa.1990.0100.

Takashi Hara and Gordon Slade. Mean-field critical behaviour for percolation in high dimensions. *Communications in Mathematical Physics*, 128(2):333–391, 1990. doi: 10.1007/BF02108785.

Takashi Hara and Gordon Slade. Mean-field behaviour and the lace expansion. In Geoffrey Grimmett, editor, *Probability and phase transition (Cambridge, 1993)*, volume 420 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 87–122. Kluwer Academic Publishers, Dordrecht, 1994. doi: 10.1007/978-94-015-8326-8_6.

T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proceedings of the Cambridge Philosophical Society*, 56:13–20, 1960. doi: 10.1017/S0305004100034241.

Harry Kesten. The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Communications in Mathematical Physics*, 74:41–59, 1980. doi: 10.1007/BF01197577.

Harry Kesten. *Percolation Theory for Mathematicians*, volume 2 of *Progress in Probability and Statistics*. Birkhäuser, Boston, MA, 1982. doi: 10.1007/978-1-4899-2730-9.

Daniel J. Kleitman. Families of non-disjoint subsets. *Journal of Combinatorial Theory*, 1:153–155, 1966. doi: 10.1016/S0021-9800(66)80012-1.

Gady Kozma and Shahaf Nitzan. A reduction of the $\theta(p_c) = 0$ problem to a conjectured inequality, 2024. Preprint, arXiv:2401.12397v1.

M. V. Menshikov. Coincidence of critical points in percolation problems. *Doklady Akademii Nauk SSSR*, 288(6):1308–1311, 1986. English translation: Soviet Math. Dokl. **33** (1986), 856–859.

Lucio Russo. A note on percolation. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 43(1):39–48, 1978. doi: 10.1007/BF00535274.

Lucio Russo. On the critical percolation probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 56:229–237, 1981. doi: 10.1007/BF00535742.

P. D. Seymour and D. J. A. Welsh. Percolation probabilities on the square lattice. *Annals of Discrete Mathematics*, 3:227–245, 1978. doi: 10.1016/S0167-5060(08)70509-0.

J. van den Berg and M. Keane. On the continuity of the percolation probability function. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemporary Mathematics*, pages 61–65. American Mathematical Society, Providence, RI, 1984. doi: 10.1090/conm/026/737388.

J. van den Berg, O. Häggström, and J. Kahn. Some conditional correlation inequalities for percolation and related processes. *Random Structures & Algorithms*, 29(4):417–435, Dec 2006. ISSN 1098-2418. doi: 10.1002/rsa.20102. URL http://dx.doi.org/10.1002/rsa.20102.