

# Supercritical Site Percolation on the Hypercube: Small Components are Small

Sahar Diskin <sup>\*</sup>      Michael Krivelevich <sup>†</sup>

October 6, 2022

## Abstract

We consider supercritical site percolation on the  $d$ -dimensional hypercube  $Q^d$ . We show that typically all components in the percolated hypercube, besides the giant, are of size  $O(d)$ . This resolves a conjecture of Bollobás, Kohayakawa, and Łuczak from 1994.

## 1 Introduction

The  $d$ -dimensional hypercube  $Q^d$  is the graph with the vertex set  $V(Q^d) = \{0, 1\}^d$ , where two vertices are adjacent if they differ in exactly one coordinate. Throughout this paper, we denote by  $n = 2^d$  the order of the hypercube.

In bond percolation on  $G$ , one considers the subgraph  $G_p$  obtained by including every edge independently with probability  $p$ . Erdős and Spencer [7] conjectured that, similar to the classical  $G(n, p)$  model, there is a phase transition at  $p = \frac{1}{d}$  in  $Q_p^d$ : for  $\epsilon > 0$ , when  $p = \frac{1-\epsilon}{d}$ , **whp** all components have size  $O(d)$ , and when  $p = \frac{1+\epsilon}{d}$ , there exists **whp** a unique giant component in  $Q_p^d$  whose order is linear in  $n$ . Their conjecture was proved by Ajtai, Komlós, and Szemerédi [1]. Bollobás, Kohayakawa, and Łuczak [4] improved upon that result, extending it to a wider range of  $p$ , allowing  $\epsilon = o(1)$ , and giving an asymptotic value of the order of the giant component of  $Q_p^d$ . Furthermore, they proved that the second largest component in  $Q_p^d$  is typically of order  $O(d)$ .

In site percolation on  $G$ , one considers the induced subgraph  $G[R]$ , where  $R$  is a random subset of vertices formed by including each vertex independently with probability  $p$ . In the setting of site percolation on the hypercube, Bollobás, Kohayakawa, and Łuczak [5] proved the following:

**Theorem 1.1** (Theorems 8, 9 of [5]). *Let  $\epsilon > 0$  be a small enough constant, and let  $p = \frac{1+\epsilon}{d}$ . Let  $R$  be a random subset formed by including each vertex of  $Q^d$  independently with probability  $p$ . Then, **whp**, there is a unique giant component of  $Q^d[R]$ , whose asymptotic size is  $\frac{(2\epsilon - O(\epsilon^2))^n}{d}$ . Furthermore, **whp** the size of the other components is at most  $d^{10}$ .*

Motivated by the results in bond percolation on the hypercube, in the same paper [5], and also in their paper on bond percolation on the hypercube [4], Bollobás, Kohayakawa, and Łuczak conjectured that **whp** all the components of  $Q^d[R]$  besides the giant component are of asymptotic size at most  $\gamma d$ , where  $\gamma$  is a constant depending only on  $\epsilon$  (see Conjecture 11 of [5]). We prove this conjecture:

**Theorem 1.2.** *Let  $\epsilon > 0$  be a small enough constant, and let  $p = \frac{1+\epsilon}{d}$ . Let  $R$  be a random subset formed by including each vertex of  $Q^d$  independently with probability  $p$ . Then, **whp**, all the components of  $Q^d[R]$  besides the giant component are of size  $O_\epsilon(d)$ .*

---

<sup>\*</sup>School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: sahardiskin@mail.tau.ac.il.

<sup>†</sup>School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: krivelev@tauex.tau.ac.il. Research supported in part by USA–Israel BSF grant 2018267.

We note that in our proof we obtain an inverse polynomial dependency on  $\epsilon$ , of order  $\frac{1}{\epsilon^5}$ , in the hidden constant in  $O_\epsilon(d)$ .

The text is structured as follows. In Section 2, we show that big components are ‘everywhere dense’ (see the exact definition in Lemma 2.3). In Section 3, we exhibit several structures that **whp** do not appear in the site-percolated hypercube. In Section 4, we prove Theorem 1.2.

Our notation is fairly standard. Throughout the rest of the text, unless specifically stated otherwise, we set  $\epsilon$  to be a small enough constant,  $p = \frac{1+\epsilon}{d}$ , and  $R$  to be a random subset of  $V(Q^d)$  formed by including each vertex independently with probability  $p$ . Furthermore, we denote by  $L_1$  the largest connected component of  $Q^d[R]$ . We denote by  $N_G(S)$  the external neighbourhood of a set  $S$  in the graph  $G$ . When  $T$  is a subset of  $V(G)$ , we write  $N_T(S) = N_G(S) \cap T$ . We omit rounding signs for the sake of clarity of presentation.

## 2 Big components are everywhere dense

We begin with the following claim, which holds for any  $d$ -regular graph:

**Claim 2.1.** *Let  $G = (V, E)$  be a  $d$ -regular graph on  $n$  vertices. Let  $\epsilon$  be a small enough constant, and let  $p = \frac{1+\epsilon}{d}$ . Form  $R \subseteq V$  by including each vertex of  $V$  independently with probability  $p$ . Then, **whp**, any connected component  $S$  of  $G[R]$  with  $|S| = k > 300 \ln n$  has  $|N_G(S)| \geq \frac{9kd}{10}$ .*

*Proof.* We run the Depth First Search (DFS) algorithm with  $n$  random bits  $X_i$ , discovering the connected components while generating  $G[R]$  (see [8]). If there is a connected component  $S$  of size  $k$ , then there is an interval of random bits whose length is  $k + |N_G(S)|$  in the DFS where we receive  $k$  positive answers. Thus, the probability of having a component violating the claim, whose discovery by the DFS starts with a given bit  $X_i$  is at most:

$$\mathbb{P} \left[ \text{Bin} \left( \frac{9kd}{10} + k, \frac{1+\epsilon}{d} \right) \geq k \right]$$

Hence, by a Chernoff-type bound (see Appendix A in [2]), the probability of having such a component is at most

$$\exp \left( -\frac{k}{100} \right) \leq o \left( \frac{1}{n^2} \right),$$

where the inequality follows from our bound on  $k$ . We complete the proof with a union bound on the  $< n$  possible values of  $k$  and  $< n$  possible starting points of the interval in the DFS.  $\square$

We further require the following simple claim:

**Claim 2.2.** *Let  $S \subseteq V(Q^d)$  such that  $|S| = k \leq d$ . Then, there exist pairwise disjoint subcubes, each of dimension at least  $d - k + 1$ , such that every  $v \in S$  is in exactly one of these subcubes.*

*Proof.* We prove by induction on  $k$ . The case  $k = 1$  is trivial.

Assume that the statement holds for all  $k' < k$ . Choose any two vertices of  $S$ . There is at least one coordinate on which they do not agree. Consider the two pairwise disjoint (and complementary) subcubes of dimension  $d - 1$ : one where we fix this coordinate to be 0 and let the other coordinates vary, and the other where we fix this coordinate to be 1 and let the other coordinates vary. Clearly, each of the two vertices is in exactly one of these subcubes. Since these subcubes are disjoint, no vertex from the other  $k - 2$  vertices is in both of them. We can then apply the induction hypothesis on each of these subcubes, giving rise to pairwise disjoint subcubes of dimension at least  $d - 1 - (k - 2) = d - k + 1$ , each having exactly one of the vertices of  $S$ .  $\square$

We conclude this section with a lemma, bounding the probability that a fixed small set of vertices has no vertex at Hamming distance 1 from many vertices in large components or in their neighbourhoods. The proof here borrows several ideas used in [1, 5, 6].

**Lemma 2.3.** *Fix  $S \subseteq V(Q^d)$  of cardinality  $|S| = k \leq \frac{\epsilon d}{10}$ . Let  $X$  be the set of vertices in components of  $Q^d[R]$  of order at least  $n^{1-\epsilon}$ . Then, the probability that every  $v \in S$  has less than  $\frac{\epsilon^2 d}{40}$  neighbours (in  $Q^d$ ) in  $X \cup N_{Q^d}(X)$  is at most  $\exp\left(-\frac{\epsilon^2 dk}{40}\right)$ .*

*Proof.* Let  $S = \{v_1, \dots, v_k\}$ . By Claim 2.2, we can consider pairwise disjoint subcubes of dimension at least  $(1 - \frac{\epsilon}{10})d$ , each containing exactly one of the vertices of  $S$ . We denote these subcubes by  $Q_1, \dots, Q_k$ , where  $v_i \in V(Q_i)$ .

For each  $i$ , consider  $Q_i$ , and assume, without loss of generality, that  $v_i$  is the origin of  $Q_i$ . We then create  $\frac{\epsilon d}{10}$  pairwise disjoint subcubes of  $Q_i$  of dimension at least  $(1 - \frac{\epsilon}{5})d$ , each at distance 1 from  $v_i$ , by fixing one of the first  $\frac{\epsilon d}{10}$  coordinates of  $Q_i$  to be 1, the rest of the first  $\frac{\epsilon d}{10}$  coordinates to be 0, and letting the other coordinates vary. Denote these subcubes by  $Q_i(1), \dots, Q_i(\frac{\epsilon d}{10})$ , and the vertex at distance 1 from  $v_i$  in  $Q_i(j)$  by  $v_i(j)$ . We denote by  $n'$  the order of each  $Q_i(j)$ , and note that  $n' \geq 2^{(1-\frac{\epsilon}{5})d} = n^{1-\frac{\epsilon}{5}}$ .

Fix  $i$ . Observe that  $p$  is still supercritical at every  $Q_i(j)$  since  $(1+\epsilon)(1-\frac{\epsilon}{5}) \geq 1 + \frac{3\epsilon}{5}$ . Denote by  $L_1(i, j)$  the largest component of  $Q_i(j)[R]$ . Then, by Theorem 1.1, **whp**  $|L_1(i, j)| \geq \frac{7\epsilon n'}{6d}$ .

Furthermore, by Claim 2.1, **whp**  $|N_{Q_i(j)}(L_1(i, j))| \geq \frac{21\epsilon n'}{20}$ . Thus, setting  $\mathcal{A}_{(i,j)}$  to be the event

that  $|L_1(i, j) \cup N_{Q_i(j)}(L_1(i, j))| \geq \frac{21\epsilon n'}{20}$ , we have that  $P(\mathcal{A}_{(i,j)}) = 1 - o(1)$ . Now, let  $\mathcal{B}_{(i,j)}$

be the event that  $v_i(j) \in L_1(i, j) \cup N_{Q_i(j)}(L_1(i, j))$ . Since the hypercube is transitive, we have that  $P(\mathcal{B}_{(i,j)} | \mathcal{A}_{(i,j)}) \geq \frac{21\epsilon n' / 20}{n'} = \frac{21\epsilon}{20}$ . Therefore,  $P(\mathcal{B}_{(i,j)} \cap \mathcal{A}_{(i,j)}) \geq (1 - o(1)) \frac{21\epsilon}{20} \geq \epsilon$ . Since the subcubes  $Q_i(j)$  are pairwise disjoint, the events  $\mathcal{B}_{(i,j)} \cap \mathcal{A}_{(i,j)}$  are independent for different  $j$ . Thus, by a typical Chernoff-type bound, with probability at least  $1 - \exp\left(-\frac{\epsilon^2 d}{40}\right)$ , at least

$\frac{\epsilon^2 d}{40}$  of the  $v_i(j)$  are in  $L_1(i, j) \cup N_{Q_i(j)}(L_1(i, j))$  with  $|L_1(i, j) \cup N_{Q_i(j)}(L_1(i, j))| \geq \frac{21\epsilon n'}{20}$ . Thus,

with the same probability,  $v_i$  is at distance 1 from at least  $\frac{\epsilon^2 d}{40}$  vertices in components whose size is at least  $\frac{\epsilon n'}{d} \geq \frac{\epsilon n^{1-\frac{\epsilon}{5}}}{d} \geq n^{1-\epsilon}$  or in their neighbourhoods, that is, in  $X \cup N_{Q^d}(X)$ .

Since  $v_i$ 's are in pairwise disjoint subcubes, these events are also independent for each  $v_i$ . Hence, the probability that none of the  $v_i$  are at distance 1 from at least  $\frac{\epsilon^2 d}{40}$  vertices in  $X \cup N_{Q^d}(X)$  is at most  $\exp\left(-\frac{\epsilon^2 dk}{40}\right)$ .  $\square$

### 3 Unlikely structures in the percolated hypercube

Denote by  $N_G^2(v)$  the set of vertices in  $G$  at distance exactly 2 from  $v$ . The following lemma shows that, typically, there are no large sections of a sphere of radius 2 in  $Q^d[R]$ .

**Lemma 3.1.** *Whp, there is no  $v \in Q^d$  such that  $|N_{Q^d}^2(v) \cap R| \geq 2d$ .*

*Proof.* We have  $n$  ways to choose  $v$ , and  $\binom{d}{2d}$  ways to choose a subset of  $2d$  vertices from  $N_{Q^d}^2(v)$ . We include them in  $R$  with probability  $p^{2d}$ . Hence, by the union bound, the probability of violating the lemma is at most:

$$2^d \binom{\binom{d}{2d}}{2d} \left(\frac{1+\epsilon}{d}\right)^{2d} \leq 2^d \left(\frac{\epsilon d}{4} \cdot \frac{1+\epsilon}{d}\right)^{2d} \leq \left(\frac{14}{15}\right)^d = o(1).$$

$\square$

From Lemma 3.1, we are able to derive the following lemma:

**Lemma 3.2.** *Whp there are no  $S \subseteq R$  and  $W \subseteq V(Q^d)$  disjoint from  $S$  such that  $|W| \leq \frac{\epsilon^4 d |S|}{9 \cdot 200^2}$  and every  $v \in S$  has  $d_{Q^d}(v, W) \geq \frac{\epsilon^2 d}{200}$ .*

*Proof.* Assume the contrary. By our assumption on  $S$ , there are at least  $\frac{\epsilon^2 d |S|}{200}$  edges between  $S$  and  $W$ . Thus, the average degree from  $W$  to  $S$  is at least  $\frac{\epsilon^2 d |S|}{200 |W|}$ . Now, let us count the number of cherries with the vertex of degree 2 in  $W$ . By Jensen's inequality, we have at least  $\binom{\epsilon^2 d |S| / 200}{2} |W| \geq \frac{\epsilon^4 d^2 |S|^2}{4 \cdot 200^2 |W|} \geq \frac{9d |S|}{4}$  such cherries, where we used our assumption on  $|W|$ . Thus, by the pigeonhole principle, there is a vertex  $v \in S$  that is in  $\frac{2}{|S|} \cdot \frac{9d |S|}{4} = \frac{9d}{2}$  cherries. Since every pair of vertices in  $S$  is connected by at most two paths of length 2 in  $Q^d$ , we obtain that  $v$  is at Hamming distance 2 from at least  $\frac{9d}{2} = \frac{9d}{4} > 2d$  vertices in  $S \subseteq R$ . On the other hand, by Lemma 3.1, **whp** there is no  $v \in Q^d$  such that  $|N_{Q^d}^2(v) \cap R| \geq 2d$ , completing the proof.  $\square$

The next lemma bounds the number of subtrees of a given order in a  $d$ -regular graph.

**Lemma 3.3.** *Let  $t_k(G)$  denote the number of  $k$ -vertex trees contained in a  $d$ -regular graph  $G$  on  $n$  vertices. Then,*

$$t_k(G) \leq n(ed)^{k-1}.$$

This follows directly from Lemma 2.1 of [3].

We are now ready to state and prove the final lemma of this section:

**Lemma 3.4.** *Let  $C > 0$  be a constant, and let  $\delta > 0$  be a small enough constant. Let  $q = \frac{1+\delta}{d}$ , and form  $R_q$  by including each vertex  $v \in V(Q^d)$  independently with probability  $q$ . Let  $L_1$  the largest component of  $Q^d[R_q]$ . Then, **whp**, there is no  $S \subseteq V(Q^d)$ , such that  $|S| = Cd$ ,  $S$  is connected in  $Q^d$ , and at least  $\frac{\delta d}{10}$  vertices  $v \in S$  have that  $|N_{L_1 \cup N_{Q^d}(L_1)}(v)| < \frac{\delta^2 d}{40}$ .*

*Proof.* By Theorem 1.1, **whp** there is a unique giant component whose size is larger than  $d^{10}$ , which we denote by  $L_1$ . Thus, it suffices to show that **whp** there is no such  $S$ , where  $\frac{\delta d}{10}$  of its vertices have less than  $\frac{\delta^2 d}{40}$  vertices at Hamming distance 1 from components of size at least  $n^{1-\delta}$  or their neighbourhoods, since **whp** these components and their neighbourhoods are in fact  $L_1 \cup N_{Q^d}(L_1)$ .

By Lemma 3.3, we have  $n(ed)^{Cd}$  ways to choose  $S$ . We have  $\binom{Cd}{\frac{\delta d}{10}}$  ways to choose the vertices in  $S$  which do not have at least  $\frac{\delta^2 d}{40}$  vertices at Hamming distance 1 from components of size at least  $n^{1-\delta}$  or their neighbourhoods. By Lemma 2.3, the probability that no vertex in a given set of  $\frac{\delta d}{10}$  vertices in  $S$  has at least  $\frac{\delta^2 d}{40}$  vertices at Hamming distance 1 from components of size at least  $n^{1-\delta}$  or their neighbourhoods is at most  $\exp\left(-\frac{\delta^4 d^2}{400}\right)$ . Hence, the probability of the event violating the statement of the lemma is at most:

$$n(ed)^{Cd} \binom{Cd}{\frac{\delta d}{10}} \exp\left(-\frac{\delta^4 d^2}{400}\right) \leq n \left( (ed)^C \left(\frac{10eC}{\delta}\right)^{\frac{\delta}{10}} \exp\left(-\frac{\delta^4 d}{400}\right) \right)^d = o(1).$$

$\square$

## 4 Proof of Theorem 1.2

Let  $p_1 = \frac{1+\epsilon/2}{d}$ . Form  $R_1$  by including each vertex  $v \in V(Q^d)$  independently with probability  $p_1$ . By Theorem 1.1, **whp** there is a unique giant component, denote it by  $L'_1$ . We can thus split the vertices of the hypercube into the following three disjoint sets:  $T = L'_1 \cup N_{Q^d}(L'_1)$ ,  $M$  is the set of vertices outside  $T$  with at least  $\frac{\epsilon^2 d}{200}$  neighbours in  $T$ , and  $S = V(Q^d) \setminus (T \cup M)$ .

Let  $p_2 = \frac{\epsilon}{2d-2-\epsilon}$ . Form  $R_2$  by including each vertex  $v \in V(Q^d)$  independently with probability  $p_2$ . Note that since  $1 - p = (1 - p_1)(1 - p_2)$ , the random set  $R$  has the same distribution as  $R_1 \cup R_2$ . Thus, with a slight abuse of notation, we write  $R = R_1 \cup R_2$ . We now perform the second exposure (that is, generate  $R_2$ ) on  $S \cup M$  first, and only afterwards on  $T$ . Note that once we show that a connected set (that is, a connected subset of  $Q^d[(S \cup M) \cap R_2]$ ) has a neighbour in  $T \cap R_2$ , it means that it merges into  $L'_1$  (whereas by Theorem 1.1, **whp**  $L'_1 \subseteq L_1$ ). Furthermore, note that all the components of  $Q^d[R]$  outside  $L_1$  are subsets of  $S \cup M$ .

We will now show that **whp** there are no components contained in  $(S \cup M) \cap R$  of size at least  $2C_1d$  for an appropriate choice of a constant  $C_1 > 0$ , possibly depending  $\epsilon$ . We first consider the case where this connected component  $B$  has at least  $C_1d$  vertices in  $M$ , and show that in that case, **whp**  $B$  merges with  $L'_1$  after the second exposure on  $T$ . By construction, every  $v \in M$  has at least  $\frac{\epsilon^2 d}{200}$  neighbours in  $T$ , and  $T \cap M = \emptyset$ . Thus, by Lemma 3.2, we have that **whp**  $|N_T(B \cap M)| \geq \frac{C_1 \epsilon^4 d^2}{9 \cdot 200^2}$ . Hence,

$$\mathbb{P}[|N_T(B) \cap R_2| = 0] \leq \left(1 - \frac{\epsilon}{2d}\right)^{\frac{C_1 \epsilon^4 d^2}{9 \cdot 200^2}} \leq \exp\left(-\frac{C_1 \epsilon^5 d}{18 \cdot 200^2}\right) = o(1/n),$$

by choosing  $C_1 \geq \frac{18 \cdot 200^2}{\epsilon^5}$ . We conclude with the union bound over the  $< n$  connected components in  $(S \cup M) \cap R$ .

Thus, we are left to deal with connected components  $B \subseteq (S \cup M) \cap R$  of size at least  $2C_1d$  with less than  $C_1d$  vertices in  $B \cap M$ . Taking a connected subset  $B' \subseteq B$  of size exactly  $Cd := 2C_1d$ , we have that  $|B' \cap S| \geq |B'| - |B' \cap M| \geq C_1d \geq \frac{\epsilon d}{20}$ . Recalling that  $S, M, T$  and  $L'_1$  were formed according to the supercritical percolation with  $p_1 = \frac{1+\epsilon/2}{d}$ , by Lemma 3.4 applied with  $\delta = \frac{\epsilon}{2}$ , **whp** there is no connected set of size  $Cd$  in  $Q^d$  with at least  $\frac{\epsilon d}{10} = \frac{\epsilon d}{20}$  vertices which do not have at least  $\frac{(\frac{\epsilon}{2})^2 d}{40} > \frac{\epsilon^2 d}{200}$  vertices at Hamming distance 1 from  $L'_1 \cup N_{Q^d}(L'_1)$ . Thus, **whp** there is no such connected set  $B'$ . All in all, **whp**, there is no connected component  $B$  in  $Q^d[R]$  of size at least  $Cd$  outside of  $L_1$ .  $\square$

**Acknowledgements.** The first author wishes to thank Arnon Chor and Dor Elboim for fruitful discussions.

## References

- [1] M. Ajtai, J. Komlós, and E. Szemerédi, *Largest random component of a  $k$ -cube*, *Combinatorica* 2 (1982), 1–7.
- [2] N. Alon and J. H. Spencer, **The probabilistic method**, 4th Ed., Wiley, New York, 2016.
- [3] A. Beveridge, A. M. Frieze and C. McDiarmid, *Random minimum length spanning trees in regular graphs*, *Combinatorica* 18 (1998), 311–333.
- [4] B. Bollobás, Y. Kohayakawa, and T. Łuczak, *The evolution of random subgraphs of the cube*, *Random Structures & Algorithms* 3 (1992), 55–90.
- [5] B. Bollobás, Y. Kohayakawa, and T. Łuczak, *On the evolution of random Boolean functions*, *Extremal problems for finite sets (Visegrád, 1991)*, Volume 3 of *Bolyai Society of Mathematical Studies*, 137–156. János Bolyai Mathematical Society, Budapest, 1994.
- [6] J. Erde, M. Kang and M. Krivelevich, *Expansion in supercritical random subgraphs of the hypercube and its consequences*, *Annals of Probability*, to appear, arXiv:2111.06752.

- [7] P. Erdős and J. Spencer, *Evolution of the  $n$ -cube*, Computers & Mathematics with Applications, 5(1) (1979), 33–39.
- [8] M. Krivelevich, *The phase transition in site percolation on pseudo-random graphs*, Electronic Journal of Combinatorics 23(1) (2016), P1.12.