

# MAJORITY BOOTSTRAP PERCOLATION ON THE HYPERCUBE

JÓZSEF BALOGH, BÉLA BOLLOBÁS, AND ROBERT MORRIS

ABSTRACT. In majority bootstrap percolation on a graph  $G$ , an infection spreads according to the following deterministic rule: if at least half of the neighbours of a vertex  $v$  are already infected, then  $v$  is also infected, and infected vertices remain infected forever. Percolation occurs if eventually every vertex is infected.

The elements of the set of initially infected vertices,  $A \subset V(G)$ , are normally chosen independently at random, each with probability  $p$ , say. This process has been extensively studied on the sequence of torus graphs  $[n]^d$ , for  $n = 1, 2, \dots$ , where  $d = d(n)$  is either fixed or a very slowly growing function of  $n$ . For example, Cerf and Manzo [14] showed that the critical probability is  $o(1)$  if  $d(n) \leq \log_* n$ , i.e., if  $p = p(n)$  is bounded away from zero then the probability of percolation on  $[n]^d$  tends to one as  $n \rightarrow \infty$ .

In this paper we study the case when the growth of  $d$  to  $\infty$  is not excessively slow; in particular, we show that the critical probability is  $1/2 + o(1)$  if  $d \geq (\log \log n)^2 \log \log \log n$ , and give much stronger bounds in the case that  $G$  is the hypercube,  $[2]^d$ .

## 1. INTRODUCTION

Consider a finite graph  $G$  with two parameters,  $q_v$  and  $r_v$ , attached to each vertex  $v$ , with  $q_v + r_v$  greater than the degree of  $v$ . Suppose each of the vertices may take either one of two states, ‘active’ and ‘inactive’, say. At each instant, some of the vertices may ‘wake up’ at random; whenever a vertex  $v$  does so, if at least  $q_v$  of its neighbours are active then it becomes active, if at least  $r_v$  of its neighbours are inactive then it becomes inactive, and if neither of these cases holds then  $v$  keeps its state. Given an initial distribution of active sites, one can then ask what happens to the system in the long run.

For example, we may take  $G$  to be a  $d$ -regular graph with  $d$  odd, and parameters  $q_v = r_v = (d + 1)/2$  for every vertex  $v$ : in this case,

---

The first author was supported during this research by OTKA grant T049398 and NSF grants DMS-0302804, DMS-0603769 and DMS 0600303, and UIUC Campus Research Board 06139 and 07048, the second by ITR grant CCR-0225610 and ARO grant W911NF-06-1-0076, and the third by PRONEX CNPq/FAPERJ grant E-26/171.167/2003-APQ1 and by MCT grant PCI EV-8C.

whenever a vertex wakes up, its state becomes that of the majority of its neighbours. For a  $d$ -regular graph with  $d$  even, we may take  $q_v = r_v = d/2 + 1$  for every vertex  $v$ .

A special (and very well-studied) example of this process is the zero-temperature Ising model, where the process occurs on the lattice  $\mathbb{Z}^d$ . Despite all this interest, however, only rather weak (though far from easy) results have so far been proven (see [17], [23] and [24], for example). Another example is the celebrated model of the brain, introduced over 60 years ago by McCulloch and Pitts [22], and shown by them to be complex enough to include a universal Turing machine as a particular case.

In this paper we shall study a simpler model, introduced by Chalupa, Leith and Reich [15] in 1979, in which the vertices may only change state in one direction, from inactive to active, say. In order to make this easier to remember, let us refer instead to ‘healthy’ and ‘infected’ vertices, so that a healthy vertex may be infected, but infected vertices never recover.

More generally (and more precisely), let  $G$  be a finite graph, let  $r \in \mathbb{N}$ , and let  $A \subset V(G)$  be a set of initially infected vertices. The set of infected vertices is updated as follows: if a healthy vertex has at least  $r$  infected neighbours, he becomes infected; otherwise he remains healthy. In other words, we have a sequence of sets

$$A = A^{(0)} \subset A^{(1)} \subset \dots \subset A^{(m)} \subset \dots$$

where  $A^{(m+1)} = A^{(m)} \cup \{v \in V(G) : |\Gamma(v) \cap A^{(m)}| \geq r\}$ . If the entire graph is eventually infected, i.e.,  $A^{(m)} = V(G)$  for some  $m \in \mathbb{N}$ , then the set  $A$  is said to *percolate* on  $G$ . This process is known as the  *$r$ -neighbour bootstrap percolation on  $G$* ; if  $G$  is  $d$ -regular with  $d$  odd and  $r = \lceil d/2 \rceil$ , then this is the *majority* bootstrap percolation. We remark that this process is very different from the *random majority* bootstrap process studied in [4], and also the *biased majority* process studied in [26], since in each of those processes vertices may change states in both directions, i.e., may be infected and then later healed.

The bootstrap process has been well-studied in the case that  $G$  is the torus  $[n]^d$ , where  $r$  and  $d$  are both fixed (with  $r \leq d$ ), the elements of the set  $A$  are chosen independently at random, and  $n \rightarrow \infty$ . (In fact the results below were proved for the grid (i.e., the subgraph of  $\mathbb{Z}^d$  induced by the vertices of  $[n]^d$ ) but the proofs also apply to the corresponding tori. Note that for  $n = 2$  the grid is  $d$ -regular, but for  $n \geq 3$  the degrees vary from  $d$  to  $2d$ .)

Let  $p = p(n) = \mathbb{P}(v \in A)$  for each  $v \in V(G)$ , and write  $\mathbb{P}_p$  for the corresponding (product) probability measure. Clearly the probability

that  $A$  percolates is monotone in  $p$ , since extra initial infections can only make percolation more likely. Hence there exists a (unique) value  $p_c \in (0, 1)$ , depending on  $n$ ,  $d$  and  $r$ , for which  $\mathbb{P}(A \text{ percolates}) = 1/2$ . We call this value the *critical probability*, and in general define

$$p_c(G, r) = \sup \left\{ p \in (0, 1) : \mathbb{P}_p(A \text{ percolates on } G) \leq \frac{1}{2} \right\}.$$

The challenge is to determine the value of  $p_c$ , and also the size of the *critical window*, i.e., the range  $p_{1-\varepsilon} - p_\varepsilon$ , where  $p_\alpha$  satisfies  $\mathbb{P}_{p_\alpha}(A \text{ percolates}) = \alpha$  for each  $\alpha \in (0, 1)$ , and  $\varepsilon \rightarrow 0$ . In general we would like to show that the window is small, and hence that the threshold for percolation is ‘sharp’.

The first rigorous results for bootstrap percolation on finite graphs were obtained by Aizenman and Lebowitz [3], who showed that

$$\frac{c(d)}{(\log n)^{d-1}} \leq p_c([n]^d, 2) \leq \frac{C(d)}{(\log n)^{d-1}}$$

for some functions  $0 < c(d) < C(d)$ , and moreover that the size of the critical window for  $[n]^d$  is  $O(p_c)$  when  $d$  is fixed and  $r = 2$ . The problem when  $3 \leq r \leq d$  seems to be more difficult (when  $r > d$  we have  $p_c = 1 - o(1)$ , so the problem is less interesting). Despite this, Cerf and Manzo [14], building on work of Cerf and Cirillo [13] (as well as the older work of Schonmann [25] on the lattice  $\mathbb{Z}^d$ ), were able to prove the corresponding result for all fixed  $d$  and  $r$ . They proved that, if  $r \leq d$ , then

$$\frac{c(d, r)}{(\log_{r-1} n)^{d-r+1}} \leq p_c([n]^d, r) \leq \frac{C(d, r)}{(\log_{r-1} n)^{d-r+1}},$$

for some functions  $0 < c(d, r) < C(d, r)$ , where  $\log_{r-1} n$  is the  $(r-1)$ -times iterated logarithm, i.e.,  $\log_1 n = \log n$  and  $\log_{k+1} n = \log(\log_k n)$  for each  $k \in \mathbb{N}$ . In particular, their proof implies that  $p_c([n]^d, d) = o(1)$  if  $1 \ll d \leq \log_* n$ , where  $\log_* n$  is the number  $k$  such that  $\log_k n \geq 1 > \log_{k+1} n$ . In particular, note that  $\log_* n \ll \log_k n$  for every  $k \in \mathbb{N}$ .

Finally, we remark that in the case  $d = r = 2$ , Holroyd [20] was able to prove a much sharper result: that in fact

$$p_c([n]^2, 2) = \frac{\pi^2}{18 \log n} + o\left(\frac{1}{\log n}\right).$$

(Here, and throughout,  $\log$  is taken to the base  $e$ ). The reader who is interested in bootstrap percolation on other types of graphs should also see the work of Balogh and Pittel [8] on random  $d$ -regular graphs, and of Balogh, Peres and Pete [7] on infinite trees. An application of the techniques of bootstrap percolation to the zero-temperature Ising

model may be found in [17], and for a brief survey of the physical applications of the bootstrap process, see [1].

In this paper we shall be interested in majority bootstrap percolation on very high-dimensional lattices, and more generally on arbitrary  $d$ -regular ‘lattice-like’ graphs. The fundamental result of Cerf and Manzo, stated above, has two drawbacks. The first is that the known bounds on  $c(d, r)$  and  $C(d, r)$  are rather far apart for large  $d$  and  $r$ , and the second that the theorem is only useful when  $n$  is extremely large (so that  $\log_{r-1} n > c(d, r)$ ). We shall show that the latter problem is unavoidable;

indeed when  $n$  is not quite so large (at most  $2^{2\sqrt{\frac{d}{\log d}}}$ , say) then the behaviour of  $p_c$  is quite different (for sufficiently large  $d$ ). We shall also study in more detail majority bootstrap percolation on one particular graph, the hypercube  $Q_d = [2]^d$ , which may be thought of as an extreme point of the family  $\{[n]^d\}$ , where  $n = n(t)$  and  $d = d(t)$  are arbitrary functions. The hypercube is a very well-studied combinatorial object; for example, see the work (relating to a different percolation problem on the hypercube) of [16], [2] and [11], and the recent improvements of [12], [18] and [19]. In a subsequent paper [6] we shall investigate 2-neighbour bootstrap percolation on high-dimensional lattices, and show that the problem essentially reduces to the equivalent question for the hypercube. As the reader will discover, a similar phenomenon also occurs for majority percolation.

The structure of the remainder of the paper is as follows. In the next section we shall state our main results; in Section 3 we shall describe some of the simple tools that we shall use later; in Sections 4 and 5 we shall prove fairly strong bounds on  $p_c(Q_n, n/2)$ ; in Section 6 we shall study more general  $d$ -regular graphs; and in Section 7 we shall describe some open problems and conjectures. The paper ends with an appendix in which the tools from Section 3 are proved.

## 2. MAIN RESULTS

In this section we shall state our main results. We begin with a theorem which gives bounds on the critical probability for majority bootstrap percolation on the hypercube. Recall that the  $n$ -dimensional hypercube,  $Q_n = [2]^n$ , is the  $n$ -regular graph with vertex set  $V(Q_n) = \{0, 1\}^n$ , and edge set  $E(Q_n)$ , where

$$xy \in E(Q_n) \text{ if and only if } |\{i : x_i \neq y_i\}| = 1.$$

Observe that the vertices of  $Q_n$  may also be thought of as subsets of  $[n]$ , in which case  $xy \in E(Q_n)$  if and only if  $|x \Delta y| = 1$ .

We shall assume throughout that  $n$  is even, so that we may write  $n/2$  instead of  $\lceil n/2 \rceil$ . However, all our proofs are also valid for  $n$  odd.

**Theorem 1.** *Let  $n \in \mathbb{N}$ ,  $\lambda \in \mathbb{R}$ ,*

$$p = p(n) = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\lambda \log \log n}{\sqrt{n \log n}},$$

*and let the elements of  $A \subset V(Q_n)$  be chosen independently at random, each with probability  $p$ . Then in majority bootstrap percolation,*

$$\mathbb{P}(A \text{ percolates on } Q_n) \rightarrow \begin{cases} 0 & \text{if } \lambda \leq -2 \\ 1 & \text{if } \lambda > 1/2 \end{cases}$$

*as  $n \rightarrow \infty$ . In particular,*

$$p_c(Q_n, n/2) \geq \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} - \frac{2 \log \log n}{\sqrt{n \log n}}$$

*for sufficiently large  $n$ , and*

$$p_c(Q_n, n/2) \leq \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\log \log n}{2\sqrt{n \log n}} + o\left(\frac{\log \log n}{2\sqrt{n \log n}}\right)$$

*as  $n \rightarrow \infty$ .*

Note that Theorem 1 determines the first two terms in the expansion of  $p_c(Q_n, n/2)$ , but not even the order of the third, since it does not tell us whether or not  $A$  is likely to percolate when  $\lambda = 0$ . In Section 5, we shall discuss how these bounds might be further improved, and explain the limitations of our method.

We next turn to more general  $d$ -regular graphs on  $N$  vertices. As it turns out, the method of the proof of Theorem 1 can be adapted, in a slightly simpler form, to deal with a wide range of such graphs. As usual, we write  $\Gamma(u)$  for the set of neighbours of a vertex  $u \in V(G)$ . Furthermore, for each  $u \in V(G)$  and  $k \in \mathbb{N}$  let us define

$$S(u, k) = \{v \in V(G) : d(u, v) = k\},$$

so that  $S(u, 1) = \Gamma(u)$ , and also let  $B(u, k) = \{v : d(u, v) \leq k\}$ . The property of the hypercube which we shall need in order to prove the theorem below, is that the set  $S(u, k) \cap \Gamma(v)$  is not ‘too big’ for any  $u \in V(G)$ ,  $v \in V(G) \setminus B(u, k-1)$ , and ‘sufficiently large’  $k$  (the exact size needed depends on  $N$ ).

The following theorem is somewhat technical, so on a first reading the reader may wish to assume that  $k$  and the functions  $f_i$  are all constant.

**Theorem 2.** *For each  $d \in \mathbb{N}$ , let  $N = N(d), k = k(d) \in \mathbb{N}$ , and let  $G = G(d)$  be a  $d$ -regular graph on  $N$  vertices. Furthermore, let  $\omega, f, f_1, \dots, f_k : \mathbb{N} \rightarrow \mathbb{N}$  be arbitrary functions satisfying*

$$1 \leq f_i(d) \leq f(d) = o\left(\frac{d}{k \log d}\right)$$

and  $\omega(d) \rightarrow \infty$  as  $d \rightarrow \infty$ . Suppose that, for every  $d \in \mathbb{N}$  and  $i \in [k]$

$$|S(x, i) \cap \Gamma(y)| \leq f_i(d)$$

for every  $x \in V(G)$  and  $y \in V(G) \setminus B(x, i-1)$ , and

$$N \leq \exp\left(\frac{d^k}{(\omega(d)k)^k (f_{k-1}(d) + f_k(d)) \prod_{i=1}^{k-1} f_i(d)}\right)$$

for every  $d \in \mathbb{N}$ . Then

$$p_c(G, d/2) = \frac{1}{2} + o(1)$$

as  $d \rightarrow \infty$ .

We remark that the bound on  $N$  in Theorem 2 cannot be improved substantially. An example showing this will be described in Section 6. Other extensions are possible though; in particular, the bound on  $|S(x, i) \cap \Gamma(y)|$  does not have to hold for every  $x \in V(G)$ : rather, a (small) exceptional set is permissible, since with high probability these vertices will be infected in the first round anyway. We shall not need this generalization however, and do not wish to further complicate the statement of the theorem unnecessarily.

Our main motivation for proving Theorem 2 is the following, almost immediate corollary. Let  $[n]^d$  denote the  $d$ -dimensional  $n \times \dots \times n$  torus, and note that  $[n]^d$  is  $2d$ -regular, and has  $n^d$  vertices. Moreover,  $[n]^d$  satisfies the conditions of Theorem 2 with  $f_i(d) = i + 1$  for each  $i \in \mathbb{N}$  (see the ‘Proof of Corollary 3’ in Section 6). Thus, we may apply the theorem as long as  $k^2 = o\left(\frac{d}{\log d}\right)$ , and  $n^d \leq 2^{2k}$ ; doing so gives the following result.

**Corollary 3.** *Let  $n = n(t)$  and  $d = d(t)$  be functions satisfying*

$$3 \leq n = 2^{2^{o\left(\sqrt{\frac{d}{\log d}}\right)}},$$

or equivalently,  $d \geq \varepsilon(\log \log n)^2 \log \log \log n$  for some  $\varepsilon > 0$ . Then

$$p_c([n]^d, d) = \frac{1}{2} + o(1)$$

as  $t \rightarrow \infty$ .

The proofs of Theorems 1 and 2 are not short, but the main ideas may be summarised in a few sentences. For the upper bound in Theorem 1, we study the first two steps of the process in detail, and show that with high probability, at least (about)  $3/4$  of the vertices will be infected by this stage; the rest of the proof is then simply a matter of battling the weak dependence between the events  $\{x \in A^{(2)}\}_{x \in V(Q_n)}$ .

The upper bound is somewhat harder, since the process may continue running for many steps. We overcome this by introducing a new, ‘more generous’ process, which nonetheless stops quickly; a simple coupling then shows that the original process must also stop eventually, before it has infected the entire vertex set. The proof of Theorem 2 is similar, though in this case the process must be allowed to run for many more steps, since we are dealing with much larger vertex sets. However, since we only wish to prove a much weaker result (that  $p_c(G, d) = 1/2 + o(1)$ ), the details of the calculations become much simpler.

### 3. TOOLS

Since all the results in this section will either be well known, or simple approximations of binomials, we shall postpone the proofs until the appendix. We begin by recalling the standard Chernoff bound (see [10] for example).

**Lemma 4.** *Let  $n \in \mathbb{N}$ ,  $0 < p < 1$ ,  $t \geq 0$  and  $S(n) \sim \text{Bin}(n, p)$ . Then*

$$(a) \quad \mathbb{P}(S(n) \geq np + t) \leq \exp\left(-\frac{2t^2}{n}\right)$$

and similarly,

$$(b) \quad \mathbb{P}(S(n) \leq np - t) \leq \exp\left(-\frac{2t^2}{n}\right)$$

The following lemma gives an almost matching lower bound in the case in which we shall be interested.

**Lemma 5.** *Let  $C \geq 0$  be a constant, and  $n \in \mathbb{N}$  be sufficiently large. Let  $p = \frac{1}{2} - \delta$ , where  $0 \leq 8\delta^4 n \leq 1$ , and let  $S(n) \sim B(n, p)$ . Then*

$$\mathbb{P}\left(S(n) \geq \frac{n}{2} + C\right) \geq \exp\left(-2\delta^2 n - 4\delta\sqrt{\frac{n}{\log n}} - \frac{\log \log n}{2} - 6\right).$$

We next state a simple generalization of Lemma 4, which will be a key tool in the proofs of the lower bounds in Theorems 1 and 2.

**Lemma 6.** *Let  $t, k, d_1, \dots, d_k \in \mathbb{N}$  and  $p \in (0, 1)$ . Let  $X_i \sim \text{Bin}(d_i, p)$  for each  $i \in [k]$ , let  $Y_k = \sum_{i=1}^k iX_i$ , and let  $D(k) = \sum_{i=1}^k i^2 d_i$ . Then*

$$\mathbb{P}(Y_k \geq \mathbb{E}(Y_k) + t) \leq (2t)^{k-1} \exp\left(-\frac{2t^2}{D(k)}\right).$$

Now an easy approximation, which we shall use to prove the upper bounds in Theorems 1 and 2.

**Lemma 7.** *Let  $p \in (0, 1)$  and  $n \in \mathbb{N}$  satisfy  $pn^2 \leq 1$ , and let  $S(n) \sim \text{Bin}(n, p)$ . Then*

$$\mathbb{P}(S(n) \geq m) \leq 2p^{m/2}$$

for every  $m \in [n]$ . In particular, if  $c, \varepsilon > 0$  and  $p \leq e^{-cn}$ , then, for some  $b = b(c, \varepsilon) > 0$  not depending on  $n$ ,

$$\mathbb{P}(S(n) \geq \varepsilon n) \leq e^{-bn^2}.$$

In order to apply Lemma 7, we shall make frequent use of the following simple lemma.

**Lemma 8.** *Let  $G$  be a graph, let  $k, m \in \mathbb{N}$ , and suppose that for each  $x \in V(G)$ ,*

$$|B(x, k)| = |\{y \in V(G) : d(x, y) \leq k\}| \leq m.$$

Then there exists a partition

$$V(G) = B_1 \cup \dots \cup B_m$$

of  $V(G)$ , such that if  $y, z \in B_i$ , then  $d(y, z) \geq k + 1$ .

Lemma 8 immediately implies the following result for hypercubes. We remark that below, and throughout the paper, we shall often write  $x \in Q_n$  to mean  $x \in V(Q_n)$ .

**Lemma 9.** *Let  $n, k \in \mathbb{N}$ , and  $x \in Q_n$ . Then there exists a partition*

$$S(x, k) = B_1 \cup \dots \cup B_m$$

of  $S(x, k)$  into  $m \leq k \binom{n}{k-1} \leq 2n^{k-1}$  sets, such that if  $x, y \in B_j$  for some  $j$ , then  $d(x, y) \geq 2k$ .

We remark that Lemma 9 is a simple special case of the following old result, due to Baranyai [9].

**Lemma (Baranyai, 1973).** *Let  $K_n^{(h)}$  denote the complete  $h$ -uniform hypergraph on  $n$  vertices. If  $h$  divides  $n$ , then  $K_n^{(h)}$  can be partitioned into 1-regular hypergraphs.*

We shall also need the following two, rather easy lemmas, and one trivial observation, which follows by the convexity of  $\binom{x}{2}$ .

**Lemma 10.** *Let  $\delta = \delta(n) \rightarrow 0$  as  $n \rightarrow \infty$ , and let  $p = \frac{1}{2} - \delta$ . Let  $S(n) \sim B(n, p)$  and let  $S'(n) \sim 1 + B(n - 1, p)$ . Then, for any  $0 \leq m = m(n) \leq n/2$ ,*

$$\mathbb{P}(S'(n) \geq m) = (1 + o(1)) \mathbb{P}(S(n) \geq m)$$

as  $n \rightarrow \infty$ .

**Lemma 11.** *Let  $\delta = \delta(n) \rightarrow 0$  as  $n \rightarrow \infty$ , and let  $p = \frac{1}{2} - \delta$ . Let  $S(n) = X + Y(n)$ , where  $X \sim \text{Bin}(1, p)$  and  $Y(n) \sim \text{Bin}(n - 1, p)$ . Then, for any  $0 \leq m = m(n) \leq n/2$ ,*

$$\mathbb{P}(X(1) = 1 \mid S(n) \geq m) = (1 + o(1)) \mathbb{P}(X(1) = 1)$$

as  $n \rightarrow \infty$ .

**Observation 12.** *Let  $k, a_1, \dots, a_k, A \in \mathbb{N}$ , and suppose  $\max_i \{a_i\} \leq A$ . Then*

$$\sum_i \binom{a_i}{2} \leq \frac{\sum_i a_i}{A} \binom{A}{2}.$$

Finally, we shall use the following, probably well-known lemma (see [21], Lemma B.7), and we recall the standard second moment method (see [10], for example).

**Lemma 13.** *Let  $p \in (0, 1)$ , let  $n \in \mathbb{N}$ , and let  $S(n) \sim \text{Bin}(n, p)$ . Then*

$$\mathbb{P}(S(n) \leq \lfloor np \rfloor - 1) \leq 1/2 \leq \mathbb{P}(S(n) \leq \lceil np \rceil).$$

*In particular, if  $\varepsilon > 0$  and  $p = p(n) \in [\varepsilon, 1 - \varepsilon]$  for every  $n \in \mathbb{N}$ , then*

$$\mathbb{P}(S(n) \geq np) = 1/2 + o(1)$$

as  $n \rightarrow \infty$ .

**Lemma 14.** *For any random variable  $X$ , and any  $\alpha > 0$ , we have*

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \alpha |\mathbb{E}(X)|\right) \leq \frac{\text{Var}(X)}{\alpha^2 \mathbb{E}(X)^2}.$$

*In particular, if  $X_n$  is a sequence of non-negative random variables, such that  $\text{Var}(X_n) = o(\mathbb{E}(X_n)^2)$ , then*

$$\mathbb{P}\left(\frac{\mathbb{E}(X_n)}{2} \leq X_n \leq 2\mathbb{E}(X_n)\right) = 1 - o(1).$$

## 4. PROOF OF THE UPPER BOUND IN THEOREM 1

Let  $n \in \mathbb{N}$  be sufficiently large, and let

$$p(n) = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\lambda \log \log n}{\sqrt{n \log n}},$$

for some  $\lambda \in \mathbb{R}$ . We shall show that if  $\lambda > 1/2$ , then for some constant  $c > 0$  and each vertex  $x \in V$ ,  $\mathbb{P}(x \notin A^{(11)}) < e^{-cn^2}$ , and thus that  $\mathbb{P}(A^{(11)} = V) = 1 - o(1)$ .

The proof comes in three stages: first we shall show that  $\mathbb{P}(x \in A^{(2)}) > 3/4$  (Lemma 15); next we shall show that  $\mathbb{P}(x \notin A^{(5)}) < e^{-cn}$  for some constant  $c$  (Lemma 16); and finally, we shall show that  $\mathbb{P}(x \notin A^{(11)}) < e^{-cn^2}$  (Lemma 17).

In all that follows, we shall assume that the elements of  $A^{(0)}$  are chosen independently, each with probability  $p$ , and that  $A^{(i)}$  are the infected vertices after  $i$  rounds.

In fact, we need a slightly more general concept. Let  $A_r^{(i)}$  denote the set of infected vertices after round  $i$  if the infection threshold is  $r$ . Thus  $A^{(i)} = A_{n/2}^{(i)}$ .

**Lemma 15.** *Let  $x \in V(Q_n)$ , and  $p$  and  $A^{(0)}$  be defined as above. Let  $r \leq n/2 + 100$ . If  $\lambda > 1/2$ , then  $\mathbb{P}(x \in A_r^{(2)}) \geq 3/4 + o(1)$  as  $n \rightarrow \infty$ .*

*Proof.* Let  $x \in V(Q_n)$ , where  $n$  is large. With probability  $p = 1/2 + o(1)$ , we have  $x \in A^{(0)}$ . For those  $x \notin A^{(0)}$ , we shall use Lemma 5 to show that  $\mathbb{P}(|\Gamma(x) \cap A_r^{(1)}| \geq r) \geq 1/2 + o(1)$ .

So let us assume that  $x \notin A^{(0)}$ , and let  $R = \Gamma(x) \cap A^{(0)}$  and  $S = \Gamma(x) \cap A_r^{(1)} \setminus A^{(0)}$ . We want to show that  $\mathbb{P}(|R| + |S| \geq r) \geq 1/2 + o(1)$ . This follows easily from the following claim.

Claim: Let  $100 \leq m \leq n/2$ . Then

$$\mathbb{P}\left(|S| \leq \frac{\sqrt{n}(\log n)^{(4\lambda-1)/2}}{e^{11}} \mid |R| \geq r - m\right) = o(1).$$

*Proof of claim.* We use the standard second moment method (Lemma 14). Throughout the proof of the claim we assume that  $|R| \geq r - m$ ; in particular, we shall sometimes write  $\mathbb{P}(\cdot)$  for  $\mathbb{P}(\cdot \mid |R| \geq r - m)$ . First we must bound the expected size of  $S$ . By Lemma 5 we have, for each vertex  $y \in \Gamma(x)$ ,

$$\mathbb{P}(|\Gamma(y) \cap A^{(0)}| \geq r) \geq \exp\left(-2\delta^2 n - 4\delta \sqrt{\frac{n}{\log n}} - \frac{\log \log n}{2} - 6\right),$$

where  $\delta = \frac{1}{2} - p = \frac{1}{2} \sqrt{\frac{\log n}{n}} - \frac{\lambda \log \log n}{\sqrt{n \log n}}$ . (Note that we must use Lemma 5 with  $C = 101$  here, since we assume  $x \notin A^{(0)}$ .) Observe that

$$2\delta^2 n + 4\delta \sqrt{\frac{n}{\log n}} = \frac{\log n}{2} - 2\lambda \log \log n + 2 + O\left(\frac{(\log \log n)^2}{\log n}\right),$$

and note that, by Lemma 11,  $\mathbb{P}(y \notin A^{(0)} \mid |R| \geq r - m) \geq 1/3$  if  $n$  is sufficiently large. Thus

$$\begin{aligned} \mathbb{E}(|S| \mid |R| \geq r - m) &= n \mathbb{P}(y \notin A^{(0)} \mid |R| \geq r - m) \mathbb{P}(|\Gamma(y) \cap A^{(0)}| \geq r) \\ &\geq \frac{n}{3e^6} \exp\left(-2\delta^2 n - 4\delta \sqrt{\frac{n}{\log n}} - \frac{\log \log n}{2}\right) \\ &\geq \frac{\sqrt{n} (\log n)^{(4\lambda-1)/2}}{3e^9} \end{aligned}$$

since  $n$  is sufficiently large, so we can assume the term  $O\left(\frac{(\log \log n)^2}{\log n}\right)$  is at most 1.

Now, we need to show that the variance is not too big. Consider two distinct vertices  $y, z \in \Gamma(x)$ , and note that  $|\Gamma(y) \cap \Gamma(z)| = 2$ , and that  $\mathbb{P}(y \in S \mid z \in S) \geq \mathbb{P}(y \in S)$ . Let  $\Gamma(y) \cap \Gamma(z) = \{w, x\}$ . Let  $p' = \mathbb{P}(y \in A^{(0)} \mid |R| \geq r - m)$ , and note that, by Lemma 11,  $p' = (1 + o(1))p$ . Then

$$\begin{aligned} \mathbb{P}(y \in S \mid z \in S, |R| \geq r - m) &\leq \mathbb{P}(y \in S \mid w \in A^{(0)}, |R| \geq r - m) \\ &= \mathbb{P}(y \notin A^{(0)} \mid |R| \geq r - m) \mathbb{P}(|(\Gamma(y) \cap A^{(0)}) \setminus \{x, w\}| \geq r - 1) \\ &= (1 - p') \mathbb{P}(S'(n) \geq r), \end{aligned}$$

where  $S'(n) \sim 1 + \text{Bin}(n - 2, p)$ . But

$$\mathbb{P}(y \in S \mid |R| \geq r - m) = (1 - p') \mathbb{P}(S(n) \geq r),$$

where  $S(n) \sim \text{Bin}(n - 1, p)$ . Thus, by Lemma 10,

$$\mathbb{P}(y \in S \mid z \in S) = (1 + o(1))\mathbb{P}(y \in S).$$

Finally note that  $\mathbb{E}(|S|) \rightarrow \infty$ , and so  $\text{Var}(|S|) = o(\mathbb{E}(|S|)^2)$ .

By Lemma 14, it follows that for any  $\varepsilon > 0$ ,  $|S| \geq (1 - \varepsilon)\mathbb{E}(|S|)$  with high probability. In particular,

$$\mathbb{P}\left(|S| \leq \frac{\sqrt{n} (\log n)^{(4\lambda-1)/2}}{e^{11}}\right) \leq \mathbb{P}\left(|S| \leq \frac{\mathbb{E}(|S|)}{2}\right) = o(1),$$

as claimed.  $\square$

Now, let  $m = \frac{\sqrt{n}(\log n)^{(4\lambda-1)/2}}{e^{11}}$ , let  $\widehat{R}$  denote the event that  $|R| \geq r - m$ , and let  $\widehat{S}$  denote the event that  $|S| \geq m$ . Since  $\lambda > 1/2$ , we have  $m \gg \sqrt{n \log n}$ , and so  $r - m < np = \mathbb{E}(|R|)$ . Therefore, by Lemma 13,

$$\mathbb{P}(\widehat{R}) \geq 1/2 + o(1)$$

as  $n \rightarrow \infty$ . Moreover, by the claim we have

$$\mathbb{P}(\widehat{S} \mid (x \notin A^{(0)}) \wedge \widehat{R}) = 1 - o(1).$$

Thus

$$\begin{aligned} \mathbb{P}(x \in A_r^{(2)}) &\geq \mathbb{P}(x \in A^{(0)}) + \mathbb{P}(\widehat{R} \wedge \widehat{S} \mid x \notin A^{(0)}) \mathbb{P}(x \notin A^{(0)}) \\ &= p + (1 - p) \mathbb{P}(\widehat{R}) \mathbb{P}(\widehat{S} \mid (x \notin A^{(0)}) \wedge \widehat{R}) \\ &\geq \frac{3}{4} + o(1), \end{aligned}$$

as required.  $\square$

**Remark 1.** Note that we only needed  $\lambda > 1/2$  in order to show that

$$m = \frac{\sqrt{n}(\log n)^{(4\lambda-1)/2}}{e^{11}} \geq \sqrt{n \log n} \geq \left(\frac{1}{2} - p\right)n + 100.$$

Thus, our proof will actually give

$$p_c(Q_n, n/2) \leq \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\log \log n}{2\sqrt{n \log n}} + O\left(\frac{1}{\sqrt{n \log n}}\right).$$

On first sight, it would seem that we are basically done, since in the third round essentially all of the remaining healthy vertices should be infected. Unfortunately, and crucially, however, we have lost independence. The next step, in which we go from most of the vertices to almost all of them, in fact turns out to be the most problematic.

Recall that the elements of  $A$  are always chosen independently at random with probability  $p$ .

**Lemma 16.** *For each  $\delta > 0$  there exists a constant  $c = c(\delta) > 0$  such that the following holds. Let  $n \in \mathbb{N}$ , and let  $p \in \left(\frac{1}{2} - \sqrt{\frac{\log n}{n}}, \frac{1}{2}\right)$  be large enough so that,*

$$\mathbb{P}\left(x \in A_{n/2+3}^{(2)}\right) \geq \frac{1}{2} + \delta.$$

*Then  $\mathbb{P}(x \in A^{(5)}) \geq 1 - e^{-cn}$ .*

*Proof.* We wish to estimate  $\mathbb{P}(x \notin A^{(5)})$ ; in order to do so, we must show that the probability that  $|\Gamma(x) \setminus A^{(4)}| \geq n/2$  is small. We begin with an important fact:

If  $S \subset Q_n$  satisfies  $d(y, z) \geq 5$  for every  $y, z \in S$  with  $y \neq z$ , then, for any  $r$ , the events  $\{y \in A_r^{(2)}\}_{y \in S}$  are independent.

Let  $Y = S(x, 3) = \{y \in Q_n : d(x, y) = 3\}$ , so  $|Y| = \binom{n}{3}$ , and let  $m = 3 \binom{n}{2} \leq 2n^2$ . By Lemma 9 there exist disjoint sets  $B_1, \dots, B_m$  such that  $\bigcup B_j = Y$ , and for each  $j \in [m]$ , the events  $\{y \in A^{(2)}\}_{y \in B_j}$  are independent.

Now, let  $\varepsilon > 0$ , and with foresight, observe that by Lemma 4,

$$\mathbb{P}\left(|\Gamma(x) \cap A^{(0)}| < \frac{n}{2} - \delta^2 \varepsilon^2 n\right) \leq \exp(-\delta^4 \varepsilon^4 n), \quad (1)$$

since  $\left(\frac{1}{2} - p\right)n \ll \delta^2 \varepsilon^2 n$ . Let  $J(x)$  denote the event that this does not happen, i.e., that  $|\Gamma(x) \cap A^{(0)}| \geq n/2 - \delta^2 \varepsilon^2 n$ , and assume from now on that  $J(x)$  holds. Now choose  $S \subset \Gamma(x) \setminus A^{(0)}$  with  $|S| = n/2$ . We thus have at most  $\binom{n/2 + \delta^2 \varepsilon^2 n}{n/2}$  choices for  $S$ . We shall show that it is extremely unlikely that  $S \cap A^{(4)} = \emptyset$ .

Indeed, let  $T = \Gamma(S) \setminus \{x\}$ , so  $d(x, y) = 2$  for every  $y \in T$ , and note that  $T = T_1 \cup T_2$ , where  $T_i = \{y \in T : |\Gamma(y) \cap S| = i\}$  for  $i = 1, 2$ . Note further that  $|T_1| = n^2/4$  and  $|T_2| = \binom{n/2}{2}$ .

Let  $a = |T_1 \cap A^{(3)}|$  and  $b = |T_2 \cap A^{(3)}|$ , and suppose that  $S \cap A^{(4)} = \emptyset$ . Then  $a + 2b \leq n^2/4$ . (This follows by simply counting edges – each vertex of  $S$  sends at least  $n/2$  edges to vertices not in  $A^{(3)}$ .) But, by

Lemma 4, since  $p \in \left(\frac{1}{2} - \sqrt{\frac{\log n}{n}}, \frac{1}{2}\right)$ , we have

$$\mathbb{P}\left(\left||T_1 \cap A^{(0)}| - \frac{n^2}{8}\right| \geq 2\sqrt{n^3 \log n}\right) \leq \exp(-n \log n), \quad (2)$$

and similarly

$$\mathbb{P}\left(\left||T_2 \cap A^{(0)}| - \frac{n^2}{16}\right| \geq 2\sqrt{n^3 \log n}\right) \leq \exp(-n \log n). \quad (3)$$

Let  $F(S)$  denote the event that neither of these events occurs, i.e., that  $\left||T_i \cap A^{(0)}| - n^2/8i\right| \leq 2\sqrt{n^3 \log n}$  for  $i = 1, 2$ , so  $\mathbb{P}(F(S)) \geq 1 - 2n^{-n}$ , and assume that the event  $F(S)$  holds for every  $S$ .

So, with very high probability, if  $|T \cap A^{(3)} \setminus A^{(0)}| > 6\sqrt{n^3 \log n}$  then  $a + 2b > n^2/4$ , and thus  $S \cap A^{(4)} \neq \emptyset$ . Hence the following claim will (essentially) complete the proof.

Claim:

$$\mathbb{P}\left(|T \cap A^{(3)} \setminus A^{(0)}| = O\left(\sqrt{n^3 \log n}\right)\right) \leq 6n^2 \exp\left(-\frac{\delta^2 \varepsilon n}{2}\right). \quad (4)$$

*Proof of claim.* Consider the bipartite graph  $H$ , with vertex set  $W \cup Y$ , where  $W = T \setminus A^{(0)}$  and  $Y = S(x, 3)$ , and edge set  $\{wy : wy \in E(Q_n)\}$ . Furthermore, let us colour the edges of  $H$  red and blue, according to whether or not the endpoint in  $Y$  is also in  $A^{(2)}$ , i.e.,  $c(wy) = \text{red}$  if and only if  $y \in A^{(2)}$ .

Now,  $e(H) = (n-2)|W|$ , and  $|W| = \frac{3n^2}{16} + O(\sqrt{n^3 \log n})$ , since we assumed that  $F(S)$  holds. Suppose  $|T \cap A^{(3)} \setminus A^{(0)}| = O(\sqrt{n^3 \log n})$ . Then only  $O(\sqrt{n^3 \log n})$  vertices of  $W$  have at least  $n/2$  neighbours in  $A^{(2)}$ , and thus the number of red edges in  $H$ ,  $e_R(H)$ , satisfies

$$e_R(H) \leq \frac{n|W|}{2} + O(\sqrt{n^5 \log n}) = \frac{3n^3}{32} + O(\sqrt{n^5 \log n}),$$

and thus

$$e_B(H) \geq \frac{n|W|}{2} - O(\sqrt{n^5 \log n}) = \frac{3n^3}{32} - O(\sqrt{n^5 \log n})$$

also. Now, recall the partition  $B_1, \dots, B_m$  of  $Y$  into independent sets, and define a refinement of it by setting

$$B_j(i) = \{z \in B_j : |\Gamma(z) \cap T| = i\}$$

for each  $i \in [3]$  and each  $j \in [m]$ . Note that the sets  $B_j(i)$  are determined by the set  $S$  and the partition  $B_1, \dots, B_m$ .

Subclaim: If  $\varepsilon > 0$  (chosen above) is small enough, then there exists an  $i \in [3]$  and a  $j \in [m]$  such that

- (a)  $\ell \geq 3\varepsilon n$  edges of  $H$  are incident with  $B_j(i)$ , and
- (b) at most  $(1/2 + \delta/2)\ell$  of these edges are red.

*Proof of subclaim.* Suppose the subclaim is false. Let us count the total number of blue edges. From those  $B_j(i)$  with at most  $3\varepsilon n$  incident edges, we get at most  $9\varepsilon mn \leq 18\varepsilon n^3$  edges. From the others, we get at most  $e(H)(1/2 - \delta/2) \leq n|W|(1/2 - \delta/2)$  blue edges. If  $\varepsilon$  is small enough, this contradicts the bound on  $e_B(H)$  above.  $\square$

Now, recall that the events  $\{y \in A^{(2)}\}_{y \in B_j(i)}$  are independent, and that we chose  $p$  so that  $\mathbb{P}(y \in A_{n/2+3}^{(2)}) \geq 1/2 + \delta$  for each  $y \in Q_n$ . Unfortunately, this event is not independent of which members of  $B(x, 2)$  are

in  $A^{(0)}$ ; however, for each vertex  $y \in B_j(i)$ ,  $y$  has at most three neighbours in  $B(x, 2)$ . Considering the  $(n - 3)$ -dimensional sub-hypercube containing  $y$  but none of these neighbours, we see that, for any set  $B(x, 2) \cap A^{(0)}$ , we have  $\mathbb{P}(y \in A_{n/2}^{(2)}) \geq 1/2 + \delta$ .

Let  $E_j(i)$  denote the event that the set  $B_j(i)$  satisfies the conditions (a) and (b) of the subclaim, and note again that each vertex in  $B_j(i)$  sends  $i \leq 3$  red edges into  $T$ . Thus, by Lemma 4,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i,j} E_j(i)\right) &\leq \sum_{i,j} \mathbb{P}\left(\text{Bin}\left(\frac{3\varepsilon n}{i}, \frac{1}{2} + \delta\right) \leq \left(\frac{1}{2} + \frac{\delta}{2}\right) \frac{3\varepsilon n}{i}\right) \\ &\leq 6n^2 \exp\left(-\frac{\delta^2 \varepsilon n}{2}\right). \end{aligned}$$

However, we showed that if  $|T \cap A^{(3)} \setminus A^{(0)}| = O\left(\sqrt{n^3 \log n}\right)$ , then one of the events  $E_j(i)$  occurs. Hence this proves the claim.  $\square$

Let  $M(S)$  denote the event that  $|T \cap A^{(3)} \setminus A^{(0)}| > 6\sqrt{n^3 \log n}$ , so by the claim,  $\mathbb{P}(M(S)) \geq 1 - 6n^2 \exp\left(-\frac{\delta^2 \varepsilon n}{2}\right)$ . Finally, recall that, assuming the event  $J(x)$  holds, we had at most

$$\binom{n/2 + \delta^2 \varepsilon^2 n}{\delta^2 \varepsilon^2 n} \leq \left(\frac{e}{\delta^2 \varepsilon^2}\right)^{\delta^2 \varepsilon^2 n} \leq \exp\left(\frac{\delta^2 \varepsilon n}{10}\right) \quad (5)$$

choices for the set  $S$  if  $\varepsilon$  is sufficiently small, since  $(\frac{e}{x^2})^x < e^{\delta/10}$  if  $x$  is sufficiently small.

Now, we claim that if each of the events  $J(x)$ ,  $F(S)$  and  $M(S)$  holds (for all  $S$  as above), then  $x \in A^{(5)}$ . Indeed,  $F(S) \wedge M(S)$  implies that  $a + 2b > n^2/4$ , and therefore that  $S \cap A^{(4)} \neq \emptyset$ , as explained above. If this holds for every  $S$ , then it follows that  $x \in A^{(5)}$ . Therefore, by (1), (2), (3), (4) and (5),

$$\begin{aligned} \mathbb{P}(x \notin A^{(5)}) &\leq \mathbb{P}(J(x)^c) + \sum_S \mathbb{P}(F(S)^c) + \sum_S \mathbb{P}(M(S)^c) \\ &\leq \exp(-\delta^4 \varepsilon^4 n) + \exp\left(\frac{\delta^2 \varepsilon n}{10}\right) \left(6n^2 \exp\left(-\frac{\delta^2 \varepsilon n}{2}\right) + 2n^{-n}\right) \\ &\leq \exp(-cn) \end{aligned}$$

for some  $c = c(\delta) > 0$ , as required.  $\square$

Finally, we make the (easier) jump from exponential to super-exponential probability; this step takes us from round  $k$ , to round  $2k + 1$ .

**Lemma 17.** *Let  $k \in \mathbb{N}$ , and let  $n$  be sufficiently large (in particular,  $n \geq 4k$ ). Let  $x \in Q_n$  and  $c > 0$ . Suppose that  $p$  is chosen such that*

$$\mathbb{P}(y \in A^{(k)}) < e^{-cn}$$

for each  $y \in Q_n$ . Then there exists a constant  $d = d(c, k) > 0$  such that

$$\mathbb{P}(x \in A^{(2k+1)}) \leq e^{-dn^2}.$$

*Proof.* As in Lemma 16, note first that

if  $S \subset Q_n$  satisfies  $d(y, z) \geq 2k + 1$  for every  $y, z \in S$  with  $y \neq z$ , then the events  $\{y \in A^{(k)}\}_{y \in S}$  are independent.

Let  $x \in Q_n$ , and let  $m = (k + 1) \binom{n}{k} \leq 2n^k$ . By Lemma 9, there exist disjoint sets  $B_1, \dots, B_m$  such that  $\bigcup B_j = S(x, k + 1)$ , and for each  $j \in [m]$ , the events  $\{y \in A^{(k)}\}_{y \in B_j}$  are independent.

The argument is now very simple. Suppose  $x \notin A^{(2k+1)}$ ; we claim that for each  $t \in [0, k + 1]$  there exist a set  $T(t) \subset S(x, t)$  such that  $T(t) \cap A^{(2k+1-t)} = \emptyset$ , and

$$|T(t)| \geq \frac{(n/2)!}{(n/2 - t)!t!} \geq \frac{n^t}{4^t t!}.$$

Indeed, let  $T(0) = \{x\}$ , and note that since  $x \notin A^{(2k+1)}$ ,  $T(0)$  satisfies the conditions. Now, suppose we have found  $T(t)$  as required. Then each  $y \in T(t)$  has at most  $n/2$  neighbours in  $S(x, t + 1) \cap A^{(2k-t)}$ , and so at least  $n/2 - t$  neighbours in  $S(x, t + 1) \setminus A^{(2k-t)}$ . Moreover, each element of  $S(x, t + 1)$  has exactly  $t + 1$  neighbours in  $S(x, t)$ . Thus, by counting edges, there must exist a set  $T(t + 1) \subset S(x, t + 1)$  with  $T(t + 1) \cap A^{(2k-t)} = \emptyset$ , and  $|T(t + 1)| \geq (n/2 - t)|T(t)|/(t + 1)$ , as required. The second inequality holds since  $n \geq 4k$ .

Consider  $T(k + 1)$ , and note that it has at least  $\alpha n^{k+1}$  elements, where  $\alpha > 0$  does not depend on  $n$ . Thus, there must exist a  $j \in [m]$  and an absolute constant  $\varepsilon > 0$  (not depending on  $n$ ) such that  $|B_j| \geq \varepsilon n$ , and  $|T(k + 1) \cap B_j| \geq \varepsilon |B_j|$ . Indeed, if no such  $j$  exists, then we would have  $|T(k + 1)| \leq \varepsilon mn + \varepsilon \binom{n}{k+1} < \alpha n^{k+1}$  if  $\varepsilon$  is sufficiently small (recall that  $m \leq 2n^k$ ).

Now, recall that  $\mathbb{P}(y \notin A^{(k)}) < e^{-cn}$  for each  $y \in B_j$ , that the events  $\{y \in A^{(k)}\}_{y \in B_j}$  are independent, and that  $T(k + 1) \cap A^{(k)} = \emptyset$ . Thus, by Lemma 7 we have

$$\mathbb{P}(|T(k + 1) \cap B_j| \geq \varepsilon |B_j| \mid |B_j| \geq \varepsilon n) \leq e^{-b|B_j|^2} \leq e^{-b\varepsilon^2 n^2}$$

for some  $b = b(c, \varepsilon) > 0$ , and so

$$\begin{aligned} \mathbb{P}(x \notin A^{(2k+1)}) &\leq \mathbb{P}(\exists j \text{ with } |B_j| \geq \varepsilon n \text{ and } |T(k+1) \cap B_j| \geq \varepsilon |B_j|) \\ &\leq m e^{-b\varepsilon^2 n^2} \leq e^{-dn^2} \end{aligned}$$

for some  $d = d(c, \varepsilon) > 0$ . Since  $\varepsilon = \varepsilon(k)$ , the lemma follows.  $\square$

Before moving on to the lower bound, let us put together the pieces from this section.

**Corollary 18.** *Let  $\lambda > 1/2$ , let  $n \in \mathbb{N}$  be sufficiently large, and let*

$$p(n) = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\lambda \log \log n}{\sqrt{n \log n}}.$$

*Then, in majority bootstrap percolation on  $Q_n$ , with initial set  $A$  of density  $p$ ,*

$$\mathbb{P}(A \text{ percolates}) \rightarrow 1$$

*as  $n \rightarrow \infty$ .*

*Proof.* Let  $n$  and  $p$  be as given. By Lemma 15 it follows that  $\mathbb{P}(x \in A^{(2)}) \geq 2/3$ , for sufficiently large  $n$  (depending on  $\lambda$ ). So, by Lemma 16, it follows that  $\mathbb{P}(x \notin A^{(5)}) \leq e^{-cn}$  for some  $c = c(\lambda) > 0$ , and thus by Lemma 17, it follows that  $\mathbb{P}(x \notin A^{(11)}) \leq e^{-dn^2}$  for some  $d = d(\lambda) > 0$ . Finally, the union bound gives

$$\begin{aligned} \mathbb{P}(A \text{ does not percolate}) &\leq \mathbb{P}\left(\bigcup_{x \in Q_n} (x \notin A^{(11)})\right) \\ &\leq \sum_{x \in Q_n} \mathbb{P}(x \notin A^{(11)}) \leq 2^n e^{-dn^2} = o(1), \end{aligned}$$

as required.  $\square$

## 5. PROOF OF THE LOWER BOUND IN THEOREM 1

Let  $n \in \mathbb{N}$  be sufficiently large, and let

$$p(n) = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\lambda \log \log n}{\sqrt{n \log n}},$$

for some  $\lambda \in \mathbb{R}$ , as in the previous section. We shall couple the bootstrap process with a modified process in which, if  $\lambda \leq -2$ , then  $A^{(4)} = A^{(3)} \neq Q_n$  with high probability.

We shall start slowly, and build up to the full result. First, we consider just a two step process, and show that if  $p \leq 1/2 - \varepsilon$  then the original process does not percolate. Let us refer to the original (majority) process as **Boot**, and let  $t = t(n) \geq 0$  be any non-negative

function. We define the process **Boot1**( $t$ ) as follows.

- The elements of  $A^{(0)}$  are chosen independently at random, each with probability  $p$ .
- $x \in A^{(1)}$  if  $x \in A^{(0)}$  or  $|\Gamma(x) \cap A^{(0)}| \geq n/2 - t$ .
- If  $i \geq 1$ , then  $x \in A^{(i+1)}$  if  $x \in A^{(i)}$  or  $|\Gamma(x) \cap A^{(i)}| \geq n/2$ .

Note that **Boot** = **Boot**(0), and that the process **Boot1**( $t$ ) dominates the process **Boot**, in the sense that given the same initial set  $A^{(0)}$ , then for each  $k \in \mathbb{N}$ , the set  $A^{(k)}$  given by **Boot1**( $t$ ) contains that given by **Boot**.

The following simple result, together with Corollary 18, implies that the critical probability for percolation in the hypercube is  $1/2 + o(1)$ .

**Proposition 19.** *Let  $\varepsilon > 0$ , and suppose  $p = 1/2 - \varepsilon$  and  $t = \varepsilon n/4$ . Then in **Boot1**( $t$ ),  $A^{(2)} = A^{(1)}$  with high probability.*

*Proof.* Let  $x \in Q_n$ , and suppose  $x \in A^{(2)} \setminus A^{(1)}$ . Then there must exist a set  $T \subset \Gamma(x)$ , with  $|T| = t$ , and  $T \subset A^{(1)} \setminus A^{(0)}$ . This is because  $x \in A^{(2)} \setminus A^{(1)}$  implies  $|\Gamma(x) \cap A^{(1)}| \geq n/2$ , and  $x \notin A^{(1)}$  implies  $|\Gamma(x) \cap A^{(0)}| < n/2 - t$ . We shall show that  $\mathbb{P}(T \text{ exists}) < e^{-cn^2}$  for some  $c > 0$ .

Indeed, recall that  $S(x, k) = \{y \in Q_n : d(x, y) = k\}$  for each  $k \in \mathbb{N}$ , and consider the set  $\Gamma(T) \cap S(x, 2)$ . It has  $\binom{t}{2}$  elements with two neighbours in  $T$ , and  $t(n-t)$  elements with one neighbour in  $T$ . Denote these two sets  $B$  and  $C$  respectively. Now, we claim that since  $T \subset A^{(1)} \setminus A^{(0)}$ , we have

$$|C \cap A^{(0)}| \geq t \left( \frac{n}{2} - t \right) - 2 \binom{t}{2} \geq \frac{nt}{2} - 2t^2. \quad (6)$$

This follows by counting edges. Indeed, note that each member of  $T$  has at least  $n/2 - t$  neighbours in  $A^{(0)}$ , and we assumed that  $x \notin A^{(0)}$ . Each vertex of  $C$  has only one neighbour in  $T$ , so even if every member of  $B$  is in  $A^{(0)}$  we still get the bound (6). But

$$\begin{aligned} \mathbb{E}(|C \cap A^{(0)}|) &= pt(n-t) = \frac{nt}{2} - \varepsilon nt - \left( \frac{1}{2} - \varepsilon \right) t^2 \\ &\leq \frac{nt}{2} - \varepsilon nt, \end{aligned}$$

and  $|C \cap A^{(0)}| \sim \text{Bin}(t(n-t), p)$ , so, since  $t = \varepsilon n/4$ ,

$$\mathbb{P} \left( |C \cap A^{(0)}| \geq \frac{nt}{2} - 2t^2 \right) \leq \exp \left( - \frac{2(\varepsilon nt - 2t^2)^2}{t(n-t)} \right) \leq \exp \left( - \frac{\varepsilon^3 n^2}{8} \right)$$

by Lemma 4. But we have only at most  $2^n$  choices for the set  $T$ , so

$$\mathbb{P}(x \in A^{(2)} \setminus A^{(1)}) \leq 2^n \exp\left(-\frac{\varepsilon^3 n^2}{8}\right) < e^{-cn^2}$$

for some  $c > 0$ , and so, since  $|V(Q_n)| = 2^n$ ,

$$\mathbb{P}(A^{(2)} \setminus A^{(1)} \neq \emptyset) \leq 2^n \mathbb{P}(x \in A^{(2)} \setminus A^{(1)}) = o(1),$$

as required.  $\square$

The following corollary is immediate from the proposition and Corollary 18.

**Corollary 20.**  $p_c(Q_n, n/2) = \frac{1}{2} + o(1)$ .

*Proof.* By Corollary 18 we have  $p_c(Q_n, n/2) \leq 1/2 + o(1)$ , so let  $\varepsilon > 0$ ,  $t = \varepsilon n/4$  and  $p = 1/2 - \varepsilon$ , and consider the **Boot1**( $t$ ) process on  $Q_n$ . First note that  $A^{(1)} \neq V(Q_n)$  with high probability, since, by Lemma 4,  $\mathbb{P}(|A^{(1)} \setminus A^{(0)}| \geq 2^n/100) = o(1)$ . Thus, by Proposition 19, the process **Boot1**( $t$ ) does not percolate with high probability.

Now, coupling **Boot** with **Boot1**( $t$ ) in the obvious way, we see that also **Boot** does not percolate whp, so  $p_c(Q_n, n/2) \geq 1/2 - \varepsilon$ .  $\square$

**Remark 2.** In fact, letting  $\varepsilon = \alpha n^{-1/3}$  for some  $\alpha \in \mathbb{R}$  with  $\alpha^3 > 32$ , the same proof in fact gives

$$p_c(Q_n, n/2) > \frac{1}{2} - \frac{\alpha}{n^{1/3}},$$

since  $\mathbb{P}(x \in A^{(2)} \setminus A^{(1)}) \leq 2^n e^{-\alpha^3 n/8}$ , and so  $\mathbb{P}(A^{(2)} \setminus A^{(1)} \neq \emptyset) = o(1)$ .

In order to improve this bound, we have to allow the process to go a little further. We call the following process **Boot3**( $t$ ).

- The elements of  $A^{(0)}$  are chosen independently at random.
- $x \in A^{(1)}$  if  $x \in A^{(0)}$  or  $|\Gamma(x) \cap A^{(0)}| \geq n/2 - 3t$ .
- $x \in A^{(2)}$  if  $x \in A^{(1)}$  or  $|\Gamma(x) \cap A^{(1)}| \geq n/2 - 2t$ .
- $x \in A^{(3)}$  if  $x \in A^{(2)}$  or  $|\Gamma(x) \cap A^{(2)}| \geq n/2 - t$ .
- If  $i \geq 3$ , then  $x \in A^{(i+1)}$  if  $x \in A^{(i)}$  or  $|\Gamma(x) \cap A^{(i)}| \geq n/2$ .

Note that, since we are trying to distinguish between values of  $p$  which differ by  $O\left(\frac{\log \log n}{\sqrt{n \log n}}\right)$ , we should take  $t$  no larger than  $n$  times

this. In fact, we shall show that if  $t = \sqrt{\frac{n}{\log n}}$ ,  $p$  is as above and

$\lambda \leq -2$ , then in the **Boot3**( $t$ ) process we have  $A^{(4)} = A^{(3)} \neq Q_n$  with high probability. We begin by showing that  $A^{(3)} \neq Q_n$  (in fact we only need the slightly weaker condition, that  $\lambda < 1/4$ ).

**Lemma 21.** *Let  $\lambda < 1/4$ , let  $\delta = \frac{1}{2}\sqrt{\frac{\log n}{n}} - \frac{\lambda \log \log n}{\sqrt{n \log n}}$ , and suppose that  $p = \frac{1}{2} - \delta$  and  $t = \sqrt{\frac{n}{\log n}}$ . Then in **Boot3**( $t$ ),  $A^{(3)} \neq Q_n$  whp.*

*Proof.* Let  $x \in Q_n$ . We shall show that  $\mathbb{P}(x \notin A^{(3)}) > e^{-cn}$  for some small  $c > 0$ , and that there is a set  $\{x_1, \dots, x_\ell\}$ , with  $\ell \geq 2^n/n^6$ , for which the statements  $x_i \in Q_n$  are independent. The proof that  $\mathbb{P}(x \in A^{(3)})$  is not too big is similar to that of Lemma 15.

First we show that  $\mathbb{P}(y \in A^{(2)} \mid y \notin A^{(0)}) = o(1)$ , for each  $y \in \Gamma(x)$ . In fact, we shall need a slightly stronger result: that this still holds, given any set  $\Gamma(x) \cap A^{(0)}$ . Clearly the events  $y \in A^{(2)}$  and  $z \in A^{(0)}$  are positively correlated, so let  $y \in \Gamma(x) \setminus A^{(0)}$ , and assume that  $\Gamma(x) \cap A^{(0)} = \Gamma(x) \setminus \{y\}$ . Now, let  $R = \Gamma(y) \cap A^{(0)}$  and  $S = \Gamma(y) \cap A^{(1)} \setminus A^{(0)}$ . We want to show that  $\mathbb{P}(|R| + |S| \geq n/2 - 2t) \rightarrow 0$  as  $n \rightarrow \infty$ . This follows easily from the following claim.

Claim:  $\mathbb{P}(|S| > e^{10} \sqrt{n} (\log n)^{2\lambda}) = o(1)$ .

*Proof of claim.* We use the second moment method (Lemma 14). First we must bound the expected size of  $S$ . By Lemma 4 we have, for each vertex  $z \in \Gamma(y)$ ,

$$\begin{aligned} \mathbb{P}\left(|\Gamma(z) \cap A^{(0)}| \geq \frac{n}{2} - 3t\right) &= \mathbb{P}\left(|\Gamma(z) \cap A^{(0)} \setminus \Gamma(x)| \geq \frac{n}{2} - 3t - 1\right) \\ &\leq \exp\left(-\frac{2(\delta n - 4t)^2}{n}\right), \end{aligned}$$

where  $\delta = \frac{1}{2} - p = \frac{1}{2}\sqrt{\frac{\log n}{n}} - \frac{\lambda \log \log n}{\sqrt{n \log n}}$ , since  $z$  has only one other neighbour in  $\Gamma(x)$ , and  $t \geq 1$ . Observe that

$$2\delta^2 n - 16\delta t + \frac{32t^2}{n} = \frac{\log n}{2} - 2\lambda \log \log n - 8 + O\left(\frac{(\log \log n)^2}{\log n}\right),$$

and so

$$\begin{aligned} \mathbb{E}(|S|) &= n \mathbb{P}(z \notin A^{(0)}) \mathbb{P}(|\Gamma(z) \cap A^{(0)}| \geq n/2 - 3t) \\ &\leq n \exp\left(-\frac{\log n}{2} + 2\lambda \log \log n + 9\right) = e^9 \sqrt{n} (\log n)^{2\lambda} \end{aligned}$$

since  $n$  is sufficiently large, so we can assume the term  $O\left(\frac{(\log \log n)^2}{\log n}\right)$  is at most 1. It is similarly straightforward (using Lemma 5) to show that  $\mathbb{E}(|S|) \rightarrow \infty$  as  $n \rightarrow \infty$ .

Now, we need to show that the variance is not too big. Consider two distinct vertices  $u, v \in \Gamma(y) \setminus \{x\}$ , and note that  $|\Gamma(u) \cap \Gamma(v)| = 2$ , and that  $\mathbb{P}(u \in S | v \in S) \geq \mathbb{P}(u \in S)$ . Let  $\Gamma(u) \cap \Gamma(v) = \{w, y\}$ , and also let  $\Gamma(u) \cap \Gamma(x) = \{a, y\}$ . Then, since we assumed  $y \notin A^{(0)}$ ,

$$\mathbb{P}(u \in S | v \in S) \leq \mathbb{P}(u \in S | w \in A^{(0)}) = (1-p) \mathbb{P}(S'(n) \geq n/2 - 3t),$$

where  $S'(n) = |(\Gamma(u) \cap A^{(0)}) \cup \{a, w\} \setminus \{y\}| \sim 2 + \text{Bin}(n-3, p)$ . But

$$\mathbb{P}(u \in S) = (1-p) \mathbb{P}(S(n) \geq n/2 - 3t),$$

where  $S(n) = |(\Gamma(u) \cap A^{(0)}) \cup \{a\} \setminus \{y\}| \sim 1 + \text{Bin}(n-2, p)$ . Thus, by Lemma 10,

$$\mathbb{P}(u \in S | v \in S) = (1 + o(1)) \mathbb{P}(u \in S),$$

and so, since  $\mathbb{E}(|S|) \rightarrow \infty$  as  $n \rightarrow \infty$ , we have  $\text{Var}(|S|) = o(\mathbb{E}(|S|)^2)$ .

By Lemma 14, it follows that  $|S| \leq 2\mathbb{E}(|S|)$  with high probability. Thus

$$\mathbb{P}(|S| > e^{10} \sqrt{n} (\log n)^{2\lambda}) \leq \mathbb{P}(|S| > 2\mathbb{E}(|S|)) = o(1),$$

as claimed.  $\square$

Now, let  $m = e^{10} \sqrt{n} (\log n)^{2\lambda}$ , let  $\widehat{R}$  denote the event that  $|R| \geq n/2 - \delta n/2$ , and let  $\widehat{S}$  denote the event that  $|S| \geq \delta n/2 - 2t \geq \delta n/3$ . Note that  $\delta n \gg m$  since  $\lambda < 1/4$ . Thus  $\mathbb{P}(\widehat{R}) = o(1)$ , by Lemma 4, and  $\mathbb{P}(\widehat{S}) = o(1)$ , by the claim, and hence

$$\mathbb{P}(y \in A^{(2)} | y \notin A^{(0)}) \leq \mathbb{P}(\widehat{R}) + \mathbb{P}(\widehat{S}) = o(1).$$

To complete the proof of the lemma, we need to use the simple fact that for positively correlated events  $E_1, \dots, E_k$ ,

$$\mathbb{P}\left(\bigcap_{i=1}^k E_i\right) \geq \prod_{i=1}^k \mathbb{P}(E_i).$$

Given  $y \in Q_n$ , let  $T(y)$  denote the event that  $y \notin A^{(2)}$ . Noting that the events  $\{T(y)\}_{y \in \Gamma(x) \setminus A^{(0)}}$  are positively correlated, we have, for any set  $\Gamma(x) \cap A^{(0)}$ , any (small)  $\varepsilon > 0$ , and sufficiently large  $n$ ,

$$\mathbb{P}(\Gamma(x) \cap A^{(2)} \setminus A^{(0)} = \emptyset) = \mathbb{P}\left(\bigcap_{y \in \Gamma(x) \setminus A^{(0)}} T(y)\right) \geq (1 - \varepsilon)^n,$$

and therefore

$$\begin{aligned} \mathbb{P}(x \notin A^{(3)}) &\geq \mathbb{P}(x \notin A^{(0)}) \mathbb{P}\left(|\Gamma(x) \cap A^{(0)}| \leq \frac{n}{2} - 3t\right) \mathbb{P}\left(\bigcap_{y \in \Gamma(x) \setminus A^{(0)}} T(y)\right) \\ &\geq \frac{1}{2} \times (1 - \varepsilon) \times (1 - \varepsilon)^n \geq \exp(-2\varepsilon n). \end{aligned}$$

Finally, if  $d(u, v) \geq 7$  then the events  $u \in A^{(3)}$  and  $v \in A^{(3)}$  are independent, so, by Lemma 8, there exists a set  $K$  of size at least  $2^n/n^6$  for which the events  $\{x \in A^{(3)}\}_{x \in K}$  are all independent. Thus

$$\mathbb{P}(x \in A^{(3)} \text{ for all } x \in K) \leq (1 - e^{-2\varepsilon n})^{|K|} = o(1)$$

if  $\varepsilon$  is chosen to be sufficiently small. This completes the proof.  $\square$

Next, we shall show that the **Boot3**( $t$ ) process stops after at most three steps if  $\lambda \leq -2$ . We shall need the following lemma about counting 3-uniform hypergraphs.

Given a 3-uniform hypergraph  $H$ , and  $i, j \in [n]$  with  $i < j$ , we shall write  $d_H(i, j)$  for the degree of the pair  $\{i, j\}$  in  $H$ , i.e.,

$$d_H(i, j) = |\{k : ijk \in E(H)\}|.$$

We write  $H \leq G$  if  $H$  is a (not necessarily induced) sub-hypergraph of  $G$ , and define  $\|H\| = \sum_{i < j} \binom{d_H(i, j)}{2}$ .

**Lemma 22.** *Let  $G$  be a labelled 3-uniform hypergraph with  $n$  vertices. Further, let  $t^2 \leq n \ll t^3$ ,  $s \leq t^3$ ,  $m \in \mathbb{N}$ , and*

$$S(G, m, s) = \{H \leq G : e(H) = s \text{ and } \|H\| \geq m\}.$$

*Then, for sufficiently large  $n$ ,*

$$|S(G, m, s)| \leq 2^n \left(\frac{20nt}{s}\right)^s \binom{e(G)}{s - 2m/n + 2t^5/n}.$$

*Proof.* Let  $H \in S(G, m, s)$ , so  $H \leq G$ ,  $e(H) = s$  and  $\|H\| \geq m$ . We partition the elements  $K = \{(i, j) : i < j\}$  into two sets, which we imaginatively title ‘big’ and ‘little’. To be precise, let  $\tau = \tau(t)$  be a function to be determined later, let

$$K_1 = \{(i, j) : d_H(i, j) \geq \tau\},$$

be the big set, and let  $K_2 = K \setminus K_1$  be the small one. Let  $|K_1| = k_1$  and  $|K_2| = k_2$ , let

$$L = \{e \in E(H) : (i, j) \in K_1 \text{ for some } i, j \in e\},$$

and  $|L| = \ell$ . We shall show that if  $\|H\|$  is large, then  $\ell$  must be large, and hence that there are only few choices for  $H$ .

Indeed, first note that  $\sum_{i<j} d_H(i, j) = 3s$ , since each edge contains exactly three elements of  $K$ , and hence  $k_1 \leq \frac{3s}{\tau}$ . Also, note that

$\sum_{(i,j) \in K_1} d_H(i, j) \leq \ell + \binom{k_1}{2}$ , since  $\ell$  edges intersect  $K_1$ , and each pair of elements of  $K_1$  is contained in at most one edge  $e \in E(H)$ . Thus we gain at most  $\binom{k_1}{2}$  extra in the sum from those edges of  $L$  which contain more than one element of  $K_1$ .

Now we reach the crux. Recall that  $d_H(i, j) \leq n$  for all  $i$  and  $j$ , and  $d_H(i, j) \leq \tau$  if  $(i, j) \in K_2$ . Thus, by the convexity of  $\binom{x}{2}$  (Observation 12), and letting  $\tau = t^2/2$ , we have

$$\begin{aligned} \|H\| &\leq \left( \frac{1}{n} \sum_{(i,j) \in K_1} d_H(i, j) \right) \binom{n}{2} + \left( \frac{1}{\tau} \sum_{(i,j) \in K_2} d_H(i, j) \right) \binom{\tau}{2} \\ &\leq \frac{(\ell + k_1^2)n}{2} + \frac{3s\tau}{2} \leq \frac{n\ell}{2} + t^5 \end{aligned} \quad (7)$$

since  $k_1 \leq \frac{3s}{\tau} \leq 6t$ , so  $k_1^2 n \ll t^5$ .

Now we have only to count. To determine  $H$ , it suffices to choose the  $k_1$  pairs  $(i, j)$  in  $K_1$ , and  $\ell$  edges incident to these pairs, and then  $s - \ell$  other edges. Thus,

$$|S(G, m, s)| \leq \sum_{k_1, \ell} \binom{n^2}{k_1} \binom{k_1 n}{\ell} \binom{e(G)}{s - \ell} \leq 6t \sum_{\ell} n^{12t} \left( \frac{6ent}{s} \right)^s \binom{e(G)}{s - \ell}$$

since  $k_1 \leq 6t$ , and using the trivial bound  $\ell \leq s$ . But  $\binom{e(G)}{s - \ell}$  is

decreasing in  $\ell$ , and  $\ell \geq \frac{2(m - t^5)}{n}$  by (7), so

$$\begin{aligned} |S(G, m, s)| &\leq 6stn^{12t} \left( \frac{6ent}{s} \right)^s \binom{e(G)}{s - 2(m - t^5)/n} \\ &\leq 2^n \left( \frac{20nt}{s} \right)^s \binom{e(G)}{s - 2m/n + 2t^5/n} \end{aligned}$$

for sufficiently large  $n$ , as claimed.  $\square$

We are ready to prove the key lemma.

**Lemma 23.** *Let  $\lambda \leq -2$ , let  $\delta = \frac{1}{2} \sqrt{\frac{\log n}{n}} - \frac{\lambda \log \log n}{\sqrt{n \log n}}$ , and suppose that  $p = \frac{1}{2} - \delta$  and  $t = \sqrt{\frac{n}{\log n}}$ . Then in  $\mathbf{Boot3}(t)$ ,  $A^{(4)} = A^{(3)}$  whp.*

*Proof.* Let  $n \in \mathbb{N}$  be sufficiently large, let  $x \in Q_n$ , and suppose that  $x \in A^{(4)} \setminus A^{(3)}$ . Then there exists a set  $T \subset \Gamma(x)$  with  $|T| = t$ , and such that  $T \subset A^{(3)} \setminus A^{(2)}$ . This is because  $x \in A^{(4)} \setminus A^{(3)}$  implies  $|\Gamma(x) \cap A^{(3)}| \geq n/2$ , and  $x \notin A^{(3)}$  implies  $|\Gamma(x) \cap A^{(2)}| < n/2 - t$ . It is convenient to think of the vertices of  $Q_n$  as subsets of  $[n]$ , and to assume (without loss of generality) that  $x = \emptyset$ , and that  $T = \{\{i\} : i \in [t]\}$ . We have at most  $2^n$  choices for the set  $T$ .

Similarly, each vertex  $y \in T$  must have at least  $t + 1$  neighbours in  $A^{(2)} \setminus A^{(1)}$ . Since  $x \notin A^{(2)}$ , these are in  $S(x, 2)$ , and note that each vertex of  $S(x, 2)$  has at most two neighbours in  $T$ . Thus there must exist a set  $T' \subset (A^{(2)} \setminus A^{(1)}) \cap S(x, 2) \cap \Gamma(T)$  with  $t^2/2 \leq |T'| \leq t^2$ . Let  $|T'| = t'$ . Given  $T$ , we have at most  $t^2 \binom{nt}{t^2}$  choices for  $T'$ .

Using the same logic one more time, each vertex  $y \in T'$  must have at least  $t + 1$  neighbours in  $A^{(1)} \setminus A^{(0)}$ , and at least  $t$  of these are in  $S(x, 3)$ . Each vertex of  $S(x, 3)$  has at most three neighbours in  $T'$ , so there must exist a set  $S \subset (A^{(1)} \setminus A^{(0)}) \cap S(x, 3) \cap \Gamma(T')$  with  $t^3/6 \leq |S| \leq t^3$ . Let  $|S| = s$ , and, considering  $S$  as a 3-uniform hypergraph on  $[n]$ , let  $\|S\| = m$ . If  $m < t^5$ , we could use the trivial upper bound  $\binom{nt^2}{s}$  on the number of choices for  $S$  (given  $T'$ ). For  $m \geq t^5$  however, we shall need the following stronger bound, which follows from Lemma 22.

Claim 1: Given  $T'$ ,  $s$  and  $m$ , there are at most

$$2^n \left( \frac{20nt}{s} \right)^s \binom{nt^2}{s - 2m/n + 2t^5/n}$$

potential sets  $S$  with  $|S| = s$  and  $\|S\| \geq m$ .

*Proof of Claim 1.* Let  $G$  be the 3-uniform hypergraph on  $[n]$  with edge set  $E(G) = \{e \in S(x, 3) \cap \Gamma(T')\}$ , and similarly consider  $S$  to be a 3-uniform hypergraph in the obvious way, i.e.,  $E(S) = \{e \in S(x, 3) \cap S\}$ . Note that  $S \subset E(G)$ ,  $|E(G)| \leq nt' \leq nt^2$ ,  $t^2 \leq n \ll t^3$ , and  $s \leq t^3$ , so the result follows immediately by Lemma 22.  $\square$

Now, consider the neighbourhood  $D$  of  $S$  in  $S(x, 4)$ , and let  $d$  be the number of edges of  $Q_n$  between  $S$  and  $D$ . Next, partition  $D$  into four parts,  $D_1, D_2, D_3$  and  $D_4$ , where each element of  $D_i$  has  $i$  neighbours in  $S$ , and let  $|D_i| = d_i$ . Furthermore, let  $R_i = D_i \cap A^{(0)}$ , and let  $|R_i| = r_i$ . We have

$$d = d_1 + 2d_2 + 3d_3 + 4d_4 = (n - 3)s,$$

and

$$r := r_1 + 2r_2 + 3r_3 + 4r_4 \geq s \left( \frac{n}{2} - 3t - 2 \right),$$

since each vertex in  $S$  has at least  $n/2 - 3t$  neighbours in  $A^{(0)}$ , and at most two neighbours in  $A^{(0)} \setminus S(x, 4)$ . Also,

$$\mathbb{E}(r) = pd = p(n-3)s \leq s \left( \frac{n}{2} - \delta n \right),$$

and  $r_i \sim \text{Bin}(d_i, p)$  for  $i = 1, 2, 3, 4$ .

In order to apply Claim 1, we shall need some bound on  $\|S\|$ . The following claim gives us one.

Claim 2:  $\|S\| \geq d_2 + 3d_3 + 6d_4$ .

*Proof of claim.* If  $ijkl \in D_2$  then  $ijk \in S$  and  $ijl \in S$ , say. Thus we add one pair to  $d_S(i, j)$ . Similarly if  $ijkl \in D_3$ , we add a pair to  $d_S(i, j)$ ,  $d_S(i, k)$  and  $d_S(i, l)$ , say, and if  $ijkl \in D_4$ , we add a pair to each of the six degrees. Finally, each pair of 3-sets are both contained in at most one 4-set.  $\square$

Now we apply Lemma 6. Let  $m = d_2 + 3d_3 + 6d_4$ , so  $\|S\| \geq m$ . By Claim 1, we have at most

$$2^n \left( \frac{20nt}{s} \right)^s \binom{nt^2}{s - 2m/n + 2t^5/n}$$

ways of choosing the elements of  $S$ . Thus, recalling that  $s \geq t^3/6 \gg n$ , so  $2^n(\delta ns)^3 \ll 2^s$ , we get

$$\begin{aligned} \mathbb{P}(\exists S \mid s, T') &\leq \sum_{d_1, d_2, d_3, d_4} 2^n \left( \frac{20nt}{s} \right)^s \binom{nt^2}{s - 2m/n + 2t^5/n} \\ &\quad \times \mathbb{P}(r \geq \mathbb{E}(r) + (\delta n - 4t)s) \\ &\leq \sum_{d_1, d_2, d_3, d_4} \left( \frac{300n}{t^2} \right)^s \left( \frac{ent^2}{s - \frac{2m}{n} + \frac{3t^5}{n}} \right)^{s - \frac{2m}{n} + \frac{3t^5}{n}} \exp \left( -\frac{2(\delta n - 4t)^2 s^2}{D(4)} \right), \end{aligned}$$

where  $D(4) = d_1 + 4d_2 + 9d_3 + 16d_4$ , as in Lemma 6. Note the replacement of  $2t^5/n$  by  $3t^5/n$ , which is motivated by the later computation.

The rest of the proof is just a straightforward calculation. Indeed, consider

$$M(\mathbf{d}) = \left( \frac{300n}{t^2} \right)^s \left( \frac{ent^2}{s - 2m/n + 3t^5/n} \right)^{s - \frac{2m}{n} + \frac{3t^5}{n}} \exp \left( -\frac{2(\delta n - 4t)^2 s^2}{D(4)} \right),$$

where  $(n-3)s = d_1 + 2d_2 + 3d_3 + 4d_4$  and  $m = d_2 + 3d_3 + 6d_4$ , and note that

$$\begin{aligned} D(4) &= d_1 + 4d_2 + 9d_3 + 16d_4 \\ &= (n-3)s + 2d_2 + 6d_3 + 12d_4 = (n-3)s + 2m. \end{aligned} \quad (8)$$

This trivial observation will allow us to bound  $M(\mathbf{d})$  from above for all  $(d_1, d_2, d_3, d_4)$ .

Indeed, recalling that  $\delta = \frac{1}{2}\sqrt{\frac{\log n}{n}} - \frac{\lambda \log \log n}{\sqrt{n \log n}}$  and  $t = \sqrt{\frac{n}{\log n}}$ , first observe that

$$\begin{aligned} n \log n - 4\lambda n \log \log n &\geq 4(\delta n - 4t)^2 &\geq n \log n - 4\lambda n \log \log n - 32\delta t n \\ &\geq n \log n - 3\lambda n \log \log n \end{aligned}$$

since  $t^2 \ll n$  and  $1 \leq 32\delta t \ll \log \log n$ . Thus, if  $m \geq \alpha sn$  for some  $\alpha \geq 0$ , then

$$\begin{aligned} \frac{2(\delta n - 4t)^2 s^2}{sn + 2m} &= \frac{2(\delta n - 4t)^2 s^2}{sn} - \frac{4m(\delta n - 4t)^2 s^2}{sn(sn + 2m)} \\ &\geq \frac{s(\log n - 3\lambda \log \log n)}{2} - \frac{m(\log n - 4\lambda \log \log n)}{(1 + 2\alpha)n}. \end{aligned} \quad (9)$$

Finally, we may assume that  $s - 2m/n + 2t^5/n \geq 0$ , as otherwise  $\binom{e(G)}{s - 2m/n + 2t^5/n} = 0$ . Thus, applying (9) with  $\alpha = 0$ , and using (8), we have

$$\begin{aligned} M(\mathbf{d}) &\leq \left(\frac{300n}{t^2}\right)^s \left(\frac{ent^2}{t^5/n}\right)^{s - \frac{2m}{n} + \frac{3t^5}{n}} \exp\left(-\frac{2(\delta n - 4t)^2 s^2}{sn + 2m}\right) \\ &\leq \left(\frac{900n}{t^2}\right)^s \left(\frac{n^2}{t^3}\right)^{s - \frac{2m}{n} + \frac{3t^5}{n}} n^{-s/2 + m/n} (\log n)^{\lambda(3s/2 - 4m/n)} \\ &= \left(\frac{t^6}{n^3(\log n)^{4\lambda}}\right)^{m/n} \left(\frac{n^2}{t^3}\right)^{3t^5/n} \left(\frac{900n^{5/2}(\log n)^{3\lambda/2}}{t^5}\right)^s. \end{aligned}$$

Now, recall that  $\frac{n}{t^2} = \log n$ , and  $s \geq \frac{t^3}{6}$ . So if  $m \leq \frac{sn}{4}$ , then

$$\begin{aligned} M(\mathbf{d}) &\leq \left( \frac{1}{(\log n)^{4\lambda+3}} \right)^{m/n} n^{3t^3/\log n} (900(\log n)^{3\lambda/2+3/2})^s \\ &\leq \left( \frac{1}{(\log n)^{4\lambda+3}} \right)^{s/4} e^{18s} (900(\log n)^{3\lambda/2+3/2})^s \\ &\leq (900e^{18}(\log n)^{\lambda/2+3/4})^s \leq (\log n)^{-s/6}. \end{aligned}$$

However, if  $m \geq \frac{sn}{4}$ , then we may apply (9) with  $\alpha = 1/4$ , so

$$\begin{aligned} M(\mathbf{d}) &\leq \left( \frac{900n}{t^2} \right)^s \left( \frac{n^2}{t^3} \right)^{s - \frac{2m}{n} + \frac{3t^5}{n}} \exp\left( -\frac{2(\delta n - 4t)^2 s^2}{sn + 2m} \right) \\ &\leq \left( \frac{900n}{t^2} \right)^s \left( \frac{n^2}{t^3} \right)^{s - \frac{2m}{n} + \frac{3t^5}{n}} n^{-s/2+2m/3n} (\log n)^{\lambda(3s/2-8m/3n)} \\ &= \left( \frac{t^{18}}{n^{10}(\log n)^{8\lambda}} \right)^{m/3n} \left( \frac{n^2}{t^3} \right)^{3t^5/n} \left( \frac{900n^{5/2}(\log n)^{3\lambda/2}}{t^5} \right)^s \\ &\leq \left( \frac{Cn^{20}(\log n)^{10\lambda}}{t^{42}} \right)^{s/12} = \left( \frac{C(\log n)^{20+10\lambda}}{t^2} \right)^{s/12} \end{aligned}$$

for some constant  $C$ . (Note that in both calculations we made the substitution  $m = sn/4$ ; we could do this in the first case because  $4\lambda + 3 < 0$ , and in the second because  $n^{10} \gg t^{18}$ .) Since  $t \gg \log n$ , it follows again that  $M(\mathbf{d}) \leq (\log n)^{-s/6}$  for sufficiently large  $n$ .

Thus, since  $d_1, \dots, d_4 \leq sn$ ,

$$\mathbb{P}(\exists S \mid s, T') \leq \sum_{d_1, d_2, d_3, d_4} (\log n)^{-s/6} \leq (sn)^4 (\log n)^{-t^3/36},$$

and we have at most  $t^3$  choices for  $s$ , and at most  $2^{2n} t^2 \binom{nt}{t^2}$  choices for  $x$ ,  $T$  and  $T'$ . Thus, summing over all of these, we obtain

$$\mathbb{P}(\exists x \in A^{(4)} \setminus A^{(3)}) \leq t^2 2^{2n} \left( \frac{en}{t} \right)^{t^2} (nt^3)^4 \left( \frac{1}{\log n} \right)^{t^3/36} = o(1),$$

as required.  $\square$

At last, we are ready to prove Theorem 1.

*Proof of Theorem 1.* The upper bound in the theorem is exactly Corollary 18, so let  $\lambda \leq -2$ , let  $n \in \mathbb{N}$  be sufficiently large, let  $t = \frac{n}{\log n}$ , let

$$p(n) = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\lambda \log \log n}{\sqrt{n \log n}},$$

and consider the **Boot3**( $t$ ) process on  $Q_n$ .

By Lemma 21 we have  $\mathbb{P}(A^{(3)} = V(Q_n)) = o(1)$ , and by Lemma 23 we have  $\mathbb{P}(A^{(4)} \neq A^{(3)}) = o(1)$ . Therefore,  $\mathbb{P}(A \text{ percolates}) = o(1)$  in the **Boot3**( $t$ ) process. The obvious coupling of **Boot** with **Boot3**( $t$ ) now shows that  $\mathbb{P}(A \text{ percolates}) = o(1)$  in the original process also, as required.  $\square$

We conclude the section by briefly discussing ways in which Theorem 1 could be strengthened, and the limitations of our method. The alert reader will no doubt have noticed that the constant  $\lambda = -2$  is not sharp; indeed, with a little more care (and no extra ideas) we could have proved that  $A$  is unlikely to percolate whenever  $\lambda < -3/4$ .

However, our method, as it stands, cannot prove the result for any  $\lambda > -1/4$ . To see this, consider an ideal set  $S$ , with  $|S| = t^3$  and  $\|S\|$  small. Using our method (based on the random variable  $r$ ), and Lemmas 4 and 5, it has probability at most about

$$\exp(-2\delta^2 n t^3)$$

of being contained in  $A^{(1)} \setminus A^{(0)}$ . There are about  $\binom{nt^2}{t^3}$  such sets, and so, writing  $X$  for the number of suitable sets  $S \subset A^{(1)} \setminus A^{(0)}$ , we get

$$\begin{aligned} \mathbb{P}(X \geq 1) &\leq \mathbb{E}(X) \approx \binom{nt^2}{t^3} \exp(-2\delta^2 n t^3) \\ &\approx \left( \frac{n}{t} \exp\left(-\frac{\log n}{2} + 2\lambda \log \log n\right) \right)^{t^3} \approx (\log n)^{(2\lambda+1/2)t^3}, \end{aligned}$$

which is small only if  $\lambda < -1/4$ .

There are two obvious places in which we could potentially be leaking probability. The first is in our estimation (using  $r$ ) of the probability that  $S$  is contained in  $A^{(1)} \setminus A^{(0)}$ ; the second is in the inequality  $\mathbb{P}(X \geq 1) \leq \mathbb{E}(X)$ . A heuristic calculation suggests that  $\text{Var}(X)$  is not too big, and so we suspect that the first of these is in fact the problem.

Finally, we point out that two simpler changes, which one might think would improve the result, in fact do not help. Firstly, we could take the process one (or more) step(s) further (i.e., consider a **Boot4**( $t$ ) process), but we would just run into the same calculation, with  $t^3$  replaced by  $t^4$ . Alternatively, we could increase  $t$ ; however, we cannot

do so significantly, since we need the inequality  $\delta t \ll \log \log n$  in order to prove inequality (9).

## 6. PERCOLATION ON $d$ -REGULAR GRAPHS

In this section we shall prove Theorem 2, which uses the ideas of the previous two sections, and deduce Corollary 3. Throughout this section, let  $G$  be a graph as in the statement of Theorem 2, so  $G$  is a  $d$ -regular graph on  $N$  vertices satisfying

$$|S(x, i) \cap \Gamma(y)| \leq f_i(d)$$

for every  $x, y \in V(G)$  with  $y \in V(G) \setminus B(x, i - 1)$ , and every  $i \in [k]$ , where  $d, N, k \in \mathbb{N}$ , and  $f_1, \dots, f_k : \mathbb{N} \rightarrow \mathbb{N}$  are functions satisfying

$$1 \leq f_i(d) \leq \frac{\eta d}{k \log d}$$

for some small constant  $\eta > 0$ , to be chosen later. Since percolation on a 1-regular graph is not very interesting, let us assume that  $d \geq 2$ . In fact, when  $\eta$  is small, the inequality above implies that  $d \geq 1/\eta$ .

We begin with the upper bound; the idea is that, by Chernoff's inequality (Lemma 4), all but exponentially few of the vertices are in  $A^{(1)}$ , and therefore that  $\mathbb{P}(x \notin A^{(k)}) \ll 1/N$ . First we use Lemma 8 to prove a version of Lemma 9 applicable to the graph  $G$ .

**Lemma 24.** *Let  $d, k \in \mathbb{N}$  and  $G$  be as described above, and let  $x \in V(G)$ . Then there exists a partition*

$$S(x, k) = B_1 \cup \dots \cup B_m$$

of  $S(x, k)$  into  $m \leq d(f_{k-1}(d) + f_k(d)) + 1$  sets, such that if  $y, z \in B_j$  for some  $j$ , then  $d(y, z) \geq 3$ .

*Proof.* Let  $y \in S(x, k)$ , and consider the set

$$Y := B(y, 2) \cap S(x, k) = \{z \in S(x, k) : d(y, z) \leq 2\}.$$

We claim that  $|Y| \leq d(f_{k-1}(d) + f_k(d)) + 1$ . Indeed,  $y$  has  $d$  neighbours, of which none are in  $B(x, k - 2)$ , at most  $f_{k-1}(d)$  are in  $S(x, k - 1)$ , and at most  $f_k(d)$  are in  $Y$ . But if  $z \notin B(x, k - 1)$  then it has at most  $f_k(d)$  neighbours in  $S(x, k)$ , by assumption. Thus

$$\begin{aligned} |Y| &\leq 1 + f_k(d) + df_{k-1}(d) + (d - f_{k-1}(d))f_k(d) \\ &\leq d(f_{k-1}(d) + f_k(d)) + 1 \end{aligned}$$

as claimed. Now, by Lemma 8 applied to the graph  $G[S(x, k)]$ , it follows that the claimed partition exists.  $\square$

We are ready to prove the main step in the upper bound.

**Lemma 25.** *Let  $d, k \in \mathbb{N}$ ,  $0 < \eta \leq 1/6$  and  $G$  be as described above. Let  $x \in V(G)$  and  $c > 0$ . Suppose that  $2\eta \leq c$ , and that  $p$  is chosen such that, in the majority bootstrap process on  $G$ ,*

$$\mathbb{P}(y \notin A^{(1)}) < e^{-cd}.$$

Then,

$$\mathbb{P}(x \notin A^{(k+1)}) \leq d^2 \exp\left(-\frac{cd^k}{3^{k+1}(f_{k-1}(d) + f_k(d)) \prod_{i=1}^{k-1} f_i(d)}\right).$$

*Proof.* The proof is very similar to that of Lemma 17, and so we shall give only a sketch. Let  $x \in V(G)$ , and let  $m = d(f_{k-1}(d) + f_k(d)) + 1$ . By Lemma 24, there exist sets  $B_1, \dots, B_m$  such that  $\bigcup B_j = S(x, k)$ , and for each  $j \in [m]$ , the events  $\{y \in A^{(1)}\}_{y \in B_j}$  are independent.

We proceed as in the proof of Lemma 17: suppose  $x \notin A^{(k+1)}$ ; then for each  $t \in [0, k]$  there exists a set  $T(t) \subset S(x, t)$  such that  $T(t) \cap A^{(k+1-t)} = \emptyset$ , and

$$|T(t)| \geq \frac{d^t}{3^t \prod_{i=1}^{t-1} f_i(d)}.$$

Indeed, let  $T(0) = \{x\}$ , and note that since  $x \notin A^{(k+1)}$ ,  $T(0)$  satisfies the conditions. Now, suppose we have found  $T(t)$  as required. Then, each  $y \in T(t)$  has at most  $d/2$  neighbours in  $S(x, t+1) \cap A^{(k-t)}$  (since  $y \notin A^{(k+1-t)}$ ), and at most  $f_{t-1}(d) + f_t(d) \leq d/6$  neighbours in  $B(x, t)$  (by the properties of  $G$ , and since  $\eta \leq 1/6$ ), and thus it has at least  $d/3$  neighbours in  $S(x, t+1) \setminus A^{(k-t)}$ . Moreover, each element of  $S(x, t+1)$  has at most  $f_t(d)$  neighbours in  $S(x, t)$ . Thus, by counting edges, there must exist a set  $T(t+1) \subset S(x, t+1) \setminus A^{(2k-t)}$  such that  $|T(t+1)| \geq d|T(t)|/3f_t(d)$ , as required.

Now, define

$$N(k) = \frac{d^k}{m3^k \prod_{i=1}^{k-1} f_i(d)},$$

and note that  $|T(k)| \geq mN(k)$  and that therefore, by the pigeonhole principle, there must exist  $j \in [m]$  such that  $|T(k) \cap B_j| \geq \lceil N(k) \rceil \geq 1$ .

Now, recall that  $\mathbb{P}(y \notin A^{(1)}) < e^{-cd}$  for each  $y \in B_j$ , that the events  $\{y \in A^{(1)}\}_{y \in B_j}$  are independent, and that  $T(k) \cap A^{(1)} = \emptyset$ . Observe also that, since  $k \log d \leq \eta d$ ,

$$e^{-cd} |B_j|^2 \leq d^{2k} e^{-cd} = \exp(-cd + 2\eta d) \leq 1.$$

Thus, by Lemma 7 we have

$$\mathbb{P}(|T(k) \cap B_j| \geq N(k)) \leq 2(e^{-cd})^{N(k)/2} = 2 \exp\left(-\frac{cdN(k)}{2}\right),$$

and so

$$\begin{aligned} \mathbb{P}(x \notin A^{(2k+1)}) &\leq \mathbb{P}(\exists j \text{ with } |T(k) \cap B_j| \geq N(k)) \leq 2m \exp\left(-\frac{cdN(k)}{2}\right) \\ &\leq d^2 \exp\left(-\frac{cd^k}{3^{k+1}(f_{k-1}(d) + f_k(d)) \prod_{i=1}^{k-1} f_i(d)}\right) \end{aligned}$$

as required.  $\square$

The upper bound in Theorem 2 will follow easily from Lemma 25 (see ‘Proof of Theorem 2’, below), and so we now turn to the lower bound. The method is based on that of Section 5; we begin by defining the natural generalization of the **Boot1**( $t$ ) and **Boot3**( $t$ ) processes. Given  $k, t \in \mathbb{N}$ , we call the following process **Bootk**( $t$ ).

- The elements of  $A^{(0)}$  are chosen independently at random, each with probability  $p$ .
- If  $0 \leq m \leq k-1$ , then
 
$$x \in A^{(m+1)} \text{ if } x \in A^{(m)} \text{ or } |\Gamma(x) \cap A^{(m)}| \geq n/2 - (k-m)t.$$
- If  $m \geq k$ , then  $x \in A^{(m+1)}$  if  $x \in A^{(m)}$  or  $|\Gamma(x) \cap A^{(m)}| \geq n/2$ .

We shall show that, if  $p = \frac{1}{2} - \varepsilon$  and  $t = \frac{\varepsilon d}{3k}$ , then in the **Bootk**( $t$ ) process we have  $A^{(k+1)} = A^{(k)} \neq V(G)$  with high probability. The following lemma is the key step.

**Lemma 26.** *Let  $d, k \in \mathbb{N}$ ,  $\eta > 0$  and  $G$  be as described above. Let  $\varepsilon > 0$ ,  $p = \frac{1}{2} - \varepsilon$ ,  $t = \frac{\varepsilon d}{3k}$  and  $m \in [k]$ , and suppose that  $12\eta \leq \varepsilon^2$ . Then, in the **Bootk**( $t$ ) process, for every  $x \in V(G)$ ,*

$$\mathbb{P}(x \in A^{(m+1)} \setminus A^{(m)}) \leq \exp\left(-\frac{\varepsilon^{m+2} d^{m+1}}{6^{m+1} k^m \prod_{i=1}^m f_i(d)}\right).$$

*Proof.* Let  $x \in V(G)$ , and suppose that  $x \in A^{(m+1)} \setminus A^{(m)}$ . Then we claim that, for each  $\ell \in [0, m]$  there exists a set  $T(\ell) \subset S(x, \ell)$  such that  $T(\ell) \subset A^{(m-\ell+1)} \setminus A^{(m-\ell)}$ , and

$$\frac{t^\ell}{2^\ell \prod_{i=1}^{\ell-1} f_i(d)} \leq |T(\ell)| \leq t^\ell.$$

Indeed, let  $T(0) = \{x\}$ , let  $t \in [0, m-1]$ , and assume that we have found  $T(\ell)$  as required. Now, let  $y \in T(\ell)$ , and note that, as in the proofs of Proposition 19 and Lemma 23, since  $Y \in A^{(m-\ell+1)} \setminus A^{(m-\ell)}$ , it follows that  $y$  has at least  $t$  neighbours in  $A^{(m-\ell)} \setminus A^{(m-\ell-1)}$ .

Recall that  $y \in S(x, \ell)$ , so  $y$  has at most

$$f_{\ell-1}(d) + f_{\ell}(d) \leq \frac{2\eta d}{k \log d} \leq \frac{t}{2}$$

neighbours in  $B(x, \ell)$ , since  $12\eta \leq \varepsilon \log d$ . Thus  $y$  has at least  $t/2$  neighbours in  $S(x, \ell+1) \cap A^{(m-\ell)} \setminus A^{(m-\ell-1)}$ , and each vertex of  $S(x, \ell+1)$  has at most  $f_{\ell}(d)$  neighbours in  $T(\ell)$ , so, by counting edges,

$$|S(x, \ell+1) \cap A^{(m-\ell)} \setminus A^{(m-\ell-1)}| \geq \frac{|T(\ell)|t}{2f_{\ell}(d)}$$

as required.

Choose a set  $T(m)$  as described, let  $s = |T(m)|$ , note that  $T(m) \subset S(x, m) \cap A^{(1)} \setminus A^{(0)}$ , and consider the neighbourhood  $W$  of  $T(m)$  in  $S(x, m+1)$ . As in the proof of Lemma 23, we shall consider the number  $r$  of edges between  $T(m)$  and  $W$  whose end-point in  $W$  lies in  $A^{(0)}$ , and show that it is far from the expected number.

First note that each vertex in  $T(m)$  has at least  $\frac{d}{2} - mt$  neighbours in  $A^{(0)}$ , and at most  $f_{m-1}(d) + f_m(d)$  neighbours in  $B(x, m)$ , so

$$r \geq s \left( \frac{d}{2} - mt - f_{m-1}(d) - f_m(d) \right) \geq \frac{sd}{2} - \frac{\varepsilon sd}{2},$$

since  $mt \leq \frac{\varepsilon d}{3}$  and  $f_i(d) \leq \frac{\eta d}{k \log d} \leq \frac{\varepsilon d}{12}$ . Next, observe that

$$\mathbb{E}(r) \leq psd = \frac{sd}{2} - \varepsilon sd.$$

Now, each vertex of  $W$  has at most  $f_m(d)$  neighbours in  $T(m)$ ; let  $d_i$  be the number of vertices with exactly  $i$  neighbours, and recall (from Lemma 6) that  $D(k) = \sum_{i=1}^k i^2 d_i$  for each  $k \in \mathbb{N}$ . By Lemma 6, we have

$$\mathbb{P} \left( r \geq \mathbb{E}(r) + \frac{\varepsilon sd}{2} \right) \leq (\varepsilon sd)^{f_m(d)} \exp \left( -\frac{(\varepsilon sd)^2}{2D(f_m(d))} \right).$$

Observe that  $D(f_m(d)) \leq f_m(d) \sum_{i=1}^{f_m(d)} i d_i \leq f_m(d)sd$ , note that  $\frac{s}{f_m(d)} \geq \frac{t}{2f_m(d)} \geq \frac{\varepsilon \log d}{6\eta}$ , and recall that  $48\eta^2 \leq \varepsilon^3 \log d$ . It then

follows that  $\frac{\varepsilon^2 sd}{2f_m(d)} \geq \frac{\varepsilon^3 d \log d}{12\eta} \geq 4\eta d$ , and thus

$$\begin{aligned} \mathbb{P}\left(r \geq \mathbb{E}(r) + \frac{\varepsilon sd}{2}\right) &\leq \exp\left(\frac{\eta d}{k \log d} \log(d^{k+1})\right) \exp\left(-\frac{\varepsilon^2 sd}{2f_m(d)}\right) \\ &\leq \exp\left(2\eta d - \frac{\varepsilon^2 sd}{2f_m(d)}\right) \leq \exp\left(-\frac{\varepsilon^2 sd}{4f_m(d)}\right) \end{aligned}$$

since  $s \leq t^m \leq d^k$ .

We have shown that the probability of a particular  $s$ -set being contained in  $A^{(1)} \setminus A^{(0)}$  is small; now we simply sum over all possible  $s$ -sets. There are at most  $\binom{d^m}{s}$  choices for the set  $T(m)$ , and so

$$\mathbb{P}(x \in A^{(m+1)} \setminus A^{(m)}) \leq \binom{d^m}{s} \exp\left(-\frac{\varepsilon^2 sd}{4f_m(d)}\right).$$

But  $\frac{\varepsilon^2 sd}{4f_m(d)} \geq \frac{\varepsilon^2 ks \log d}{4\eta} \geq 3ms \log d$ , and  $\log \binom{d^m}{s} \leq ms \log d$ , so

$$\binom{d^m}{s} \exp\left(-\frac{\varepsilon^2 sd}{4f_m(d)}\right) \leq \exp\left(-\frac{\varepsilon^2 sd}{6f_m(d)}\right).$$

Thus, using our lower bound on  $s$ , and recalling that  $t = \frac{\varepsilon d}{3k}$ , we get

$$\begin{aligned} \mathbb{P}(x \in A^{(m+1)} \setminus A^{(m)}) &\leq \exp\left(-\left(\frac{\varepsilon^2 d}{6f_m(d)}\right) \left(\frac{t^m}{2^m \prod_{i=1}^{m-1} f_i(d)}\right)\right) \\ &\leq \exp\left(-\frac{\varepsilon^{m+2} d^{m+1}}{6^{m+1} k^m \prod_{i=1}^m f_i(d)}\right), \end{aligned}$$

as required.  $\square$

We are ready to prove Theorem 2. The crucial property of  $G = G(d)$ , which we have as yet not used, will be that  $N = |V(G)|$  satisfies

$$N \leq \exp\left(\frac{d^k}{(\omega(d)k)^k (f_{k-1}(d) + f_k(d)) \prod_{i=1}^{k-1} f_i(d)}\right)$$

for every  $d \in \mathbb{N}$ .

*Proof of Theorem 2.* We shall deduce Theorem 2 from Lemmas 25 and 26. In order to apply these lemmas, we need  $\eta$  to be sufficiently small,

so first let  $\varepsilon > 0$ , recall that  $f_i(d) \leq f(d)$  for each  $i \in [k]$ , and let  $d$  be sufficiently large so that

$$f(d) \leq \frac{\varepsilon^2 d}{12k \log d}.$$

We begin with the upper bound. Let  $p = 1/2 + \varepsilon$ , and choose the elements of  $A^{(0)} \subset V(G)$  independently at random with probability  $p$ . Let  $x \in V(G)$ , and recall that  $N = |V(G)|$ ; we shall show that

$$\mathbb{P}(A^{(k+1)} \neq V(G)) \leq N \mathbb{P}(x \notin A^{(k+1)}) = o(1)$$

as  $d \rightarrow \infty$ .

Indeed,  $x$  has  $d$  neighbours, and so, by Lemma 4,

$$\mathbb{P}(x \notin A^{(1)}) \leq \mathbb{P}(\text{Bin}(d, p) < d/2) \leq \exp(-\varepsilon^2 d).$$

Thus, by Lemma 25, applied with  $c = \varepsilon^2$ ,

$$\begin{aligned} N \mathbb{P}(x \notin A^{(k+1)}) &\leq N d^2 \exp\left(-\frac{\varepsilon^2 d^k}{3^{k+1}(f_{k-1}(d) + f_k(d)) \prod_{i=1}^{k-1} f_i(d)}\right) \\ &\leq \exp\left((k \log d)^k \left(\frac{1}{(\omega(d)k)^k} - \frac{\varepsilon^2}{3^{k+1}}\right) + 2 \log d\right) = o(1) \end{aligned}$$

as  $d \rightarrow \infty$ , as required.

Now we turn to the lower bound. Let  $p = 1/2 - \varepsilon$ , and again choose the elements of  $A^{(0)} \subset V(G)$  independently at random with probability  $p$ . Let  $t = \frac{\varepsilon d}{3k}$ , and recall the **Bootk**( $t$ ) process, defined above. We shall show that, in this ‘more generous’ process, we have  $A^{(k+1)} = A^{(k)} \neq V(G)$  with high probability. The result then follows by a straightforward coupling of the two processes.

The first part, that  $A^{(k+1)} = A^{(k)}$  in **Bootk**( $t$ ), follows immediately from Lemma 26, since

$$\begin{aligned} \mathbb{P}(A^{(k+1)} \neq A^{(k)}) &\leq \mathbb{E}(|A^{(k+1)} \setminus A^{(k)}|) = N \mathbb{P}(x \in A^{(k+1)} \setminus A^{(k)}) \\ &\leq N \exp\left(-\frac{\varepsilon^{k+2} d^{k+1}}{6^{k+1} k^k \prod_{i=1}^k f_i(d)}\right) \\ &\leq \exp\left(\frac{d^k}{k^k \prod_{i=1}^k f_i(d)} \left(\frac{1}{\omega(d)^k} - \frac{\varepsilon^{k+2} d}{6^{k+1}}\right)\right) = o(1) \end{aligned}$$

as  $d \rightarrow \infty$ .

For the second part, that  $A^{(k)} \neq V(G)$ , we again use Lemma 26. First note that  $\mathbb{P}(x \in A^{(1)} \setminus A^{(0)}) \leq \exp(-\varepsilon^2 d)$ , by Lemma 4. Now,

recall that, by Lemma 26,

$$\begin{aligned} \mathbb{P}(x \in A^{(m+1)} \setminus A^{(m)}) &\leq \exp\left(-\frac{\varepsilon^{m+2}d^{m+1}}{6^{m+1}k^m \prod_{i=1}^m f_i(d)}\right) \\ &\leq \exp(-d(\log d)^m) \leq \exp(-\varepsilon^2 d) \end{aligned}$$

for every  $m \in [k]$ . Thus, by Markov's inequality,

$$\begin{aligned} \mathbb{P}\left(|A^{(k)} \setminus A^{(0)}| \geq \frac{N}{4}\right) &\leq \frac{4}{N} \mathbb{E}(|A^{(k)} \setminus A^{(0)}|) \\ &\leq 4 \sum_{m=0}^{k-1} \mathbb{P}(x \in A^{(m+1)} \setminus A^{(m)}) \\ &\leq 4k \exp(-\varepsilon^2 d) = o(1) \end{aligned} \quad (10)$$

as  $d \rightarrow \infty$ . Finally, again by Lemma 4,

$$\mathbb{P}\left(|A^{(0)}| \geq \frac{3N}{4}\right) \leq \exp\left(-\frac{N}{8}\right) = o(1). \quad (11)$$

Combining (10) and (11) gives  $\mathbb{P}(A^{(k)} = V(G)) = o(1)$ , as required.  $\square$

Before proving Corollary 3, let us note that, in one direction at least, Theorem 2 cannot be improved substantially. Suppose  $k, f(d) = O(1)$ , so the bound on  $N$  becomes  $N = \exp(o(d^k))$ . The following example shows that the theorem is false if this bound is replaced by  $N = \exp(O(d^{k+3}))$ .

**Example 1.** Given two graphs  $H_1$  and  $H_2$ , we write  $H_1 \cup H_2$  to mean the graph with vertex set  $V(H_1) \cup V(H_2)$  and edge set  $E(H_1) \cup E(H_2)$ , i.e., the graph obtained by putting the graphs side by side.

Let  $d, k \in \mathbb{N}$  (with  $d$  large) and  $C \in \mathbb{R}$ , let  $M = \exp(Cd^{k+3})$ , let  $N = d^{k+3}M$ , and let  $H$  be a  $d$ -regular graph on  $d^{k+3}$  vertices which satisfies the conditions of the theorem. Then the graph

$$G = H_1 \cup \dots \cup H_M,$$

where the graphs  $H_i$  are disjoint copies of  $H$ , is a  $d$ -regular graph on  $N$  vertices which satisfies the conditions of the theorem, but

$$p_c(G, d/2) \geq 1 - \varepsilon,$$

where  $\varepsilon = \varepsilon(C) \rightarrow 0$  as  $C \rightarrow \infty$ . This follows simply because if  $p < 1 - \varepsilon$ , then with high probability at least one of the sets  $V(H_i) \cap A$  is empty, and so no vertex of  $H_i$  is ever infected.

To show that such a graph  $H$  exists, we have only to consider a random  $d$ -regular graph on  $L = d^{k+3}$  vertices. With high probability, such a graph satisfies the conditions of Theorem 2, with  $f(d) = k + 3$ .

To see this, let  $H$  be such a graph, and note that  $S(x, k) \leq d^k$  for each  $x \in V(H)$ , so (heuristically) we have

$$\mathbb{P}(|S(x, k) \cap \Gamma(y)| \geq k + 4) \leq \binom{d}{k + 4} \left(\frac{d^k}{L}\right)^{k+4} \leq \left(\frac{d^{k+1}}{L}\right)^{k+4}$$

for each  $y \in V(H) \setminus B(x, k - 1)$ . Thus

$$\mathbb{P}(\exists x, y : |S(x, k) \cap \Gamma(y)| \geq k + 4) \leq L^2 \left(\frac{d^{k+1}}{L}\right)^{k+4} = o(1)$$

as required. It is straightforward to make this rough argument rigorous.

Finally, let us deduce Corollary 3 from Theorem 2. Recall that  $[n]^d$  denotes the  $d$ -dimensional torus, i.e., the graph with vertex set  $\{1, \dots, n\}^d$  and edge set  $\{xy : \sum_i |x_i - y_i \pmod n| = 1\}$ .

*Proof of Corollary 3.* Let  $n = n(t)$  and  $d = d(t)$  be functions satisfying the given inequalities, and let  $G = [n]^d$ , so  $G$  is a  $2d$ -regular graph on  $N = n^d$  vertices. Let  $\omega$  be any function satisfying  $1 \ll \omega(d) \ll \log \log d$ . We claim that  $G$  satisfies the conditions of Theorem 2 with  $\omega(d)k = \sqrt{\frac{d}{\log d}}$  and  $f(d) = k + 1$ .

Indeed, let  $x \in V(G)$ , and observe that for each  $m \leq k$ ,

$$S(x, m) = \{y \in V(G) : \sum_{i=1}^d \|x_i - y_i\|_{\mathbb{Z}_n} = m\},$$

where  $\|x_i - y_i\|_{\mathbb{Z}_n}$  denotes the distance between  $x_i$  and  $y_i$  in  $\mathbb{Z}_n$ , the integers modulo  $n$ . Thus, given  $y \notin B(x, m - 1)$ , we have

$$|S(x, m) \cap \Gamma(y)| = |\{i : x_i \neq y_i\}| \leq m + 1$$

if  $y \in S(x, m + 1)$ , and  $S(x, m) \cap \Gamma(y) = \emptyset$  otherwise.

All that remains is to observe that, for some  $C \in \mathbb{R}$ ,

$$\log N = d \log n \leq 2^C \sqrt{\frac{d}{\log d}} \leq (\log d)^k = \left(\frac{d}{(\omega(d)k)^2}\right)^k$$

and hence that

$$N \leq \exp\left(\frac{d^k}{(\omega(d)k)^k (f_{k-1}(d) + f_k(d)) \prod_{i=1}^{k-1} f_i(d)}\right)$$

as required.  $\square$

## 7. FURTHER QUESTIONS AND CONJECTURES

In this section we shall briefly discuss various ways in which the work in this paper could be extended. We begin by conjecturing that the upper bound in Theorem 1 is sharp.

**Conjecture 1.**

$$p_c(Q_n, n/2) = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\log n}{n}} + \frac{\log \log n}{2\sqrt{n \log n}} + o\left(\frac{\log \log n}{2\sqrt{n \log n}}\right)$$

as  $n \rightarrow \infty$ .

Next, recall once again the result of Cerf and Manzo [14], that

$$p_c([n]^d, d) = o(1)$$

when  $d \leq \log_* n$ , and Corollary 3, which states that

$$p_c([n]^d, d) = \frac{1}{2} + o(1)$$

when  $d \geq \varepsilon(\log \log n)^2 \log \log \log n$ . The obvious question poses itself: What happens in between?

**Problem 1.** *Determine  $\lim_{t \rightarrow \infty} p_c([n]^d, d)$  for every pair of functions  $n = n(t)$  and  $d = d(t)$  for which the limit exists.*

*In particular, determine the nature of the phase transition between those pairs of functions for which the limit is zero, and those for which it is non-zero.*

Another entirely natural question asks what happens to the critical probability if one changes the threshold function  $r = r(d)$ ? The proof of Theorem 2 extends easily to the case  $r = \alpha d + o(d)$  for some constant  $0 < \alpha < 1$ , and implies that

$$p_c([n]^d, r) = \alpha + o(1)$$

for the same functions  $n(t)$  and  $d(t)$  as in Corollary 3. However, the following problem is likely to be more challenging.

**Problem 2.** *Determine  $p_c([n]^d, r)$  for all functions  $2 \leq r(d) \ll d$ . In particular, characterize the pairs  $(n(d), r(d))$  of functions for which*

$$\frac{d}{r} p_c([n]^d, r) \rightarrow 1$$

as  $d \rightarrow \infty$ .

We remark that in [5], Balogh and Bollobás proved that  $p_c([2]^d, 2)$  is very far from  $\frac{r}{d}$ ; in fact it is (up to a constant factor) equal to  $\frac{2^{-2\sqrt{d}}}{d^2}$ . For even sharper results on a wider class of graphs, see [6].

Finally, there are many other  $d$ -regular graphs which Theorem 2 fails to cover; for example, those with more than  $2^{2^d}$  vertices. Our final question asks, rather vaguely, for a version of the theorem which applies to such graphs.

**Question 1.** *Does there exist a set of ‘local’ conditions which allow one to determine (or bound) the critical probability for an arbitrary  $d$ -regular graph on  $N \geq 2^{2^d}$  vertices?*

Unfortunately, it appears out of reach to prove anything (in general) for fixed  $d$  and  $N \rightarrow \infty$ .

## 8. APPENDIX: PROOFS OF THE TOOLS IN SECTION 3

In this appendix we shall prove the simple tools used earlier. We begin by recalling Stirling’s formula,

$$\sqrt{\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq 2\sqrt{\pi n} \left(\frac{n}{e}\right)^n,$$

and by making the following basic observations.

**Observation 27.**

(a) *Let  $x \leq 1/4$ . Then  $e^{-x-x^2} \leq 1-x \leq e^{-x}$ .*

(b) *Let  $0 \leq \delta \leq 1/8$ . Then  $e^{-2\delta-4\delta^2} \leq \frac{1-\delta}{1+\delta} \leq e^{-2\delta+2\delta^2}$ .*

(c) *Let  $m, n \in \mathbb{N}$  satisfy  $8m \leq n$  and  $4m^3 \leq n^2$ . Then*

$$\binom{n}{n/2+m} \geq \frac{2^{n-1}}{\sqrt{\pi n}} \exp\left(-\frac{2m^2}{n} - 1\right).$$

(d) *Let  $X \sim \text{Bin}(n, p)$ , where  $n \in \mathbb{N}$  and  $p = p(n) \geq \frac{1}{2} - \delta$ , where  $\delta = \delta(n) \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $m = m(n) \in [n/2]$ . Then*

$$\mathbb{P}(X = m) = o(\mathbb{P}(X \geq m))$$

*as  $n \rightarrow \infty$ .*

*Proof.* Part (a) is straightforward. For (b), note that

$$1 - 2\delta \leq \frac{1-\delta}{1+\delta} \leq 1 - 2\delta + 2\delta^2$$

and apply part (a). For (c), note that by Stirling's Formula (applied to  $n!$  and  $n/2!$ ) and part (a), we have

$$\begin{aligned} \binom{n}{n/2+m} &\geq \binom{n}{n/2} \left(\frac{n-2m}{n}\right)^m \geq \frac{2^{n-1}}{\sqrt{\pi n}} \left(1 - \frac{2m}{n}\right)^m \\ &\geq \frac{2^{n-1}}{\sqrt{\pi n}} \exp\left(-\frac{2m^2}{n} - \frac{4m^3}{n^2}\right) \end{aligned}$$

as claimed. For (d), let  $n \in \mathbb{N}$  be sufficiently large, and observe that  $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$  for any  $k \in [n]$ . Thus, for any integer  $0 \leq t \leq \sqrt{n}$ , we have

$$\frac{\mathbb{P}(X = m+t)}{\mathbb{P}(X = m)} \geq \left(\frac{n-m-t}{m+t}\right)^t \left(\frac{p}{1-p}\right)^t \geq \left(\frac{n/2-t}{n/2+t}\right)^t \left(\frac{1-2\delta}{1+2\delta}\right)^t,$$

and so, applying part (b) twice, for  $\frac{2t}{n}$  and for  $2\delta$ ,

$$\frac{\mathbb{P}(X = m+t)}{\mathbb{P}(X = m)} \geq \exp\left(-\frac{4t^2}{n} - 5\delta t - 1\right),$$

since  $t^3 \ll n^2$  and  $\delta^2 t \ll \delta t$ .

Finally we sum from  $t = 0$  to  $1/\delta$ , to obtain

$$\mathbb{P}(X \geq m) \geq \mathbb{P}(X = m) \sum_{t=0}^{1/\delta} \exp\left(-\frac{4t^2}{n} - 5\delta t - 1\right) \geq \frac{e^{-10}}{\delta} \rightarrow \infty$$

as  $n \rightarrow \infty$ . □

We are now ready to prove the results in Section 3. We begin with our reverse Chernoff bound, Lemma 5.

*Proof of Lemma 5.* First let  $C = 0$ . We have

$$\mathbb{P}\left(S(n) \geq \frac{n}{2}\right) = \sum_{k=n/2}^n \binom{n}{k} p^k (1-p)^{n-k}.$$

Note that  $p(1-p) = \frac{1}{4} - \delta^2$ , and observe that since  $p \leq 1/2$ , the function  $g(k) = \binom{n}{k} p^k (1-p)^{n-k}$  is decreasing on  $[n/2, n]$ . Thus, for any  $m \geq 0$ , we have

$$\mathbb{P}\left(S(n) \geq \frac{n}{2}\right) \geq mg\left(\frac{n}{2} + m\right) = m \binom{n}{n/2+m} \left(\frac{1}{4} - \delta^2\right)^{n/2} \left(\frac{p}{1-p}\right)^m.$$

Let  $m = \sqrt{\frac{n}{\log n}}$ . Now, by Observation 27(c) we have

$$\binom{n}{n/2 + m} \geq \frac{2^{n-1}}{\sqrt{\pi n}} \exp\left(-\frac{2m^2}{n} - 1\right),$$

by Observation 27(a), and recalling that  $8\delta^4 n \leq 1$ ,

$$\left(\frac{1}{4} - \delta^2\right)^{n/2} \geq 2^{-n} \exp(-2\delta^2 n - 8\delta^4 n) \geq 2^{-n} \exp(-2\delta^2 n - 1),$$

and by Observation 27(b),

$$\left(\frac{p}{1-p}\right)^m = \left(\frac{1-2\delta}{1+2\delta}\right)^m \geq \exp(-4\delta m - 16\delta^2 m) \geq \exp(-4\delta m - 1).$$

Therefore,

$$\begin{aligned} \mathbb{P}\left(S(n) \geq \frac{n}{2}\right) &\geq \frac{m}{2\sqrt{\pi n}} \exp\left(-\frac{2m^2}{n} - 2\delta^2 n - 4\delta m - 3\right) \\ &\geq \frac{1}{2e^4 \sqrt{\pi \log n}} \exp\left(-2\delta^2 n - 4\delta \sqrt{\frac{n}{\log n}}\right), \end{aligned}$$

as required, since  $2m^2 \leq n$  and  $2e^4 \sqrt{\pi} < e^6$ . For general  $C$  the proof is the same, since  $C \leq m/2$  for large  $n$ , so

$$\mathbb{P}\left(S(n) \geq \frac{n}{2} + C\right) \geq \frac{m}{2} g\left(\frac{n}{2} + m\right),$$

and  $4e^4 \sqrt{\pi} < e^6$ .  $\square$

Next we prove Lemma 6, which generalizes Lemma 4 to a weighted binomial distribution.

*Proof of Lemma 6.* We prove the lemma by induction on  $k$ . For  $k = 1$ , it is exactly Lemma 4(a), so let  $k \geq 2$ , and assume it is true for  $k - 1$ . Recall that  $Y_k = Y_{k-1} + kX_k$ . Thus, by the induction hypothesis and Lemma 4,

$$\begin{aligned} \mathbb{P}\left(Y_k \geq \mathbb{E}(Y_k) + t\right) &\leq \mathbb{P}\left(Y_{k-1} \geq \mathbb{E}(Y_{k-1}) + t\right) + \mathbb{P}\left(X_k \geq \mathbb{E}(X_k) + \frac{t}{k}\right) \\ &\quad + \sum_{m=1}^{t-1} \mathbb{P}\left(Y_{k-1} \geq \mathbb{E}(Y_{k-1}) + m\right) \mathbb{P}\left(X_k \geq \mathbb{E}(X_k) + \frac{t-m}{k}\right) \\ &\leq \exp\left(-\frac{2t^2}{D(k-1)}\right) + \exp\left(-\frac{2t^2}{k^2 d_k}\right) \\ &\quad + \sum_{m=1}^{t-1} (2t)^{k-2} \exp\left(-\frac{2m^2}{D(k-1)} - \frac{2(t-m)^2}{k^2 d_k}\right). \end{aligned}$$

Now, simple calculus gives the maximum at  $m = \frac{tD(k-1)}{D(k-1) + k^2d_k}$ .

Thus

$$\frac{m^2}{D(k-1)} + \frac{(t-m)^2}{k^2d_k} \geq \frac{t^2k^2d_kD(k-1) + t^2(k^2d_k)^2}{k^2d_k(D(k-1) + k^2d_k)^2} = \frac{t^2}{D(k)},$$

since  $D(k-1) + k^2d_k = D(k)$ , and so

$$\begin{aligned} \mathbb{P}\left(Y_k \geq \mathbb{E}(Y_k) + t\right) &\leq 2 \exp\left(-\frac{2t^2}{D(k)}\right) + (t-1)(2t)^{k-2} \exp\left(-\frac{2t^2}{D(k)}\right) \\ &\leq (2t)^{k-1} \exp\left(-\frac{2t^2}{D(k)}\right), \end{aligned}$$

and the induction step is complete.  $\square$

The remaining lemmas are even more straightforward.

*Proof of Lemma 7.* Recall that  $pn^2 \leq 1$ , and that  $S(n) \sim \text{Bin}(n, p)$ . Thus

$$\mathbb{P}(S(n) = m) \leq \binom{n}{m} p^m \leq (np)^m \leq p^{m/2},$$

and

$$\frac{\mathbb{P}(S(n) = m+1)}{\mathbb{P}(S(n) = m)} \leq \sqrt{p} < \frac{1}{2}$$

for every  $m \in \mathbb{N}$ . Therefore  $\mathbb{P}(S(n) \geq m) \leq 2p^{m/2}$ , and the second part follows immediately.  $\square$

*Proof of Lemma 8.* We apply a straightforward greedy algorithm. Taking the vertices one by one, we claim that there is some set  $B_j$ , such that all vertices already in  $B_j$  are distance at least  $k+1$  from the vertex in question. Indeed, this follows immediately from the condition

$$|B(x, k) \setminus \{x\}| = |\{y \in V(G) : d(x, y) \leq k\}| - 1 \leq m - 1.$$

The greedy algorithm thus gives the required partition of  $V(G)$ .  $\square$

Lemma 9 is an immediate consequence of Lemma 8.

*Proof of Lemma 9.* Let us again consider the vertices of  $Q_n$  as subsets of  $[n]$ , and let  $x = \emptyset$ . Given a vertex  $u \in S(x, k)$ , observe that the set

$$\{v \in S(x, k) : d(u, v) \leq 2k - 1\}$$

is exactly the set of  $k$ -subsets of  $[n]$  which intersect  $v$ . There are at most  $k \binom{n}{k-1}$  such sets, and so the result follows by Lemma 8.  $\square$

Finally, we prove the two easy lemmas.

*Proof of Lemma 10.* Let  $S'' = S' - 1 \sim \text{Bin}(n - 1, p)$ , and note that we may choose a coupling so that  $S'' \leq S \leq S'$ . Now

$$\mathbb{P}(S'(n) \geq m) = \mathbb{P}(S''(n) \geq m - 1) = (1 + o(1))\mathbb{P}(S''(n) \geq m),$$

since  $\mathbb{P}(S''(n) = m) = o(\mathbb{P}(S''(n) \geq m))$  by Observation 27(d). Thus

$$\mathbb{P}(S(n) \geq m) \leq \mathbb{P}(S'(n) \geq m) \leq (1 + o(1))\mathbb{P}(S(n) \geq m).$$

□

*Proof of Lemma 11.* We have

$$\begin{aligned} \mathbb{P}(X = 1 | S(n) \geq m) &= \frac{\mathbb{P}((X = 1) \wedge (S(n) \geq m))}{\mathbb{P}(S(n) \geq m)} \\ &= \frac{\mathbb{P}(S(n) \geq m | X = 1)\mathbb{P}(X = 1)}{\mathbb{P}(S(n) \geq m)}, \end{aligned}$$

and  $\mathbb{P}(S(n) \geq m | X = 1) = (1 + o(1))\mathbb{P}(S(n) \geq m)$  by Lemma 10, so the lemma follows. □

## 9. ACKNOWLEDGEMENTS

The authors would like to thank the Institute for Mathematical Sciences at the National University of Singapore for their hospitality during June 2006, when a large part of this research was carried out.

## REFERENCES

- [1] J. Adler and U. Lev, Bootstrap Percolation: visualizations and applications, *Braz. J. Phys.*, **33** (2003), 641–644.
- [2] M. Ajtai, J. Komlós and E. Szemerédi, Largest random component of a  $k$ -cube, *Combinatorica*, **2** (1982), 1–7.
- [3] M. Aizenman and J.L. Lebowitz, Metastability effects in bootstrap percolation, *J. Phys. A.*, **21** (1988) 3801–3813.
- [4] P. Balister, B. Bollobás, R. Johnson and M. Walters, Random Majority Percolation, submitted (2003).
- [5] J. Balogh and B. Bollobas, Bootstrap percolation on the hypercube, *Prob. Rel. Fields*, **134** (2006), 624–648.
- [6] J. Balogh, B. Bollobás and R. Morris, Bootstrap percolation on  $[n]^d$ , in preparation.
- [7] J. Balogh, Y. Peres and G. Pete, Bootstrap percolation on infinite trees and non-amenable groups, *Combinatorics, Probability and Computing*, **15** (2006), 715–730.
- [8] J. Balogh and B. Pittel, Bootstrap percolation on random regular graphs, to appear in *Random Structures and Algorithms*.
- [9] Z. Baranyai, On the factorization of the complete uniform hypergraph, *Infinite and finite sets* (Colloq. Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, 91–108.

- [10] B. Bollobás, *Random Graphs*, Second Edition, Cambridge University Press, 2001.
- [11] B. Bollobás, Y. Kohayakawa and T. Łuczak, The evolution of random subgraphs of the cube, *Random Structures and Algorithms*, **3** (1992), 55–90.
- [12] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: III. The phase transition for the  $n$ -cube, *Combinatorica*, **26** (2006), 395–410.
- [13] R. Cerf and E. N. M. Cirillo, Finite size scaling in three-dimensional bootstrap percolation, *Ann. Prob.*, **27** (1999), 1837–1850.
- [14] R. Cerf and F. Manzo, The threshold regime of finite volume bootstrap percolation, *Stochastic Proc. Appl.*, **101** (2002), 69–82.
- [15] J. Chalupa, P. L. Leath and G. R. Reich, Bootstrap percolation on a Bethe lattice, *J. Phys. C.*, **12** (1979), L31–L35.
- [16] P. Erdős and J. Spencer, Evolution of the  $n$ -cube, *Comput. Math. Appl.*, **5** (1979), 33–39.
- [17] L. R. Fontes, R. H. Schonmann and V. Sidoravicius, Stretched Exponential Fixation in Stochastic Ising Models at Zero Temperature, *Commun. Math. Phys.*, **228** (2002), 495–518.
- [18] R. van der Hofstad and G. Slade. Asymptotic expansion in  $n^{-1}$  for percolation critical values on the  $n$ -cube and  $\mathbb{Z}^n$ , *Random Structures and Algorithms*, **27** (2005), 331–357.
- [19] R. van der Hofstad and G. Slade. Expansion in  $n^{-1}$  for percolation critical values on the  $n$ -cube and  $\mathbb{Z}^n$ : the first three terms. *Combinatorics, Probability and Computing*, **15** (2006), 695–713.
- [20] A. Holroyd, Sharp Metastability Threshold for Two-Dimensional Bootstrap Percolation, *Prob. Th. Rel. Fields*, **125** (2003), 195–224.
- [21] T. Leighton and C.G. Plaxton, Hypercubic sorting networks. *SIAM J. Comput.*, **27** (1998), 1–47.
- [22] W. S. McCulloch and W. Pitts, *A logical calculus of ideas immanent in nervous activity*, *Bulletin of Mathematical Biophysics* **5** (1943), 115–133.
- [23] S. Nanda, C.M. Newman and D. Stein, Dynamics of Ising spin systems at zero temperature, In *On Dobrushin's way (From Probability Theory to Statistical Mechanics)*, eds. R. Minlos, S. Shlosman and Y. Suhov, *Am. Math. Soc. Transl.* (2) **198** (2000), 183–194.
- [24] C.M. Newman and D. Stein, Zero-temperature dynamics of Ising spin systems following a deep quench: results and open problems, *Physica A*, **279** (2000), 159–168.
- [25] R.H. Schonmann, On the behaviour of some cellular automata related to bootstrap percolation, *Ann. Prob.*, **20** (1992), 174–193.
- [26] R.H. Schonmann, Finite size scaling behavior of a biased majority rule cellular automaton, *Physica A*, **167** (1990), 619–627.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, 1409 W. GREEN STREET, URBANA, IL 61801, USA

*E-mail address:* jobal@math.uiuc.edu

TRINITY COLLEGE, CAMBRIDGE CB2 1TQ, ENGLAND, AND, DEPARTMENT OF MATHEMATICAL SCIENCES, THE UNIVERSITY OF MEMPHIS, MEMPHIS, TN 38152, USA

*E-mail address:* B.Bollobas@dpmms.cam.ac.uk

INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA, ESTRADA DONA CASTORINA, 110, JARDIM BOTÂNICO, RIO DE JANEIRO, BRAZIL

*E-mail address:* rdmorris1@impa.br