Majority Bootstrap Percolation on $G(n, p)$

Cecilia Holmgren*
Department of Mathematics
Uppsala University, Sweden
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge, U.K.
cecilia.holmgren@math.uu.se

Tomas Juškevičius
Department of Mathematical Sciences
University of Memphis, U.S.A.
tomas.juskevicius@gmail.com

Nathan Kettle
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge, U.K.
nathan.kettle@cantab.net

Submitted: Mar 9, 2016; Accepted: Dec 12, 2016; Published: Jan 20, 2017
Mathematics Subject Classifications: 60C05; 05C80; 60K35

Abstract

Majority bootstrap percolation on a graph G is an epidemic process defined in the following manner. Firstly, an initially infected set of vertices is selected. Then step by step the vertices that have at least half of its neighbours infected become infected. We say that percolation occurs if eventually all vertices in G become infected.

In this paper we provide sharp bounds for the critical size of the initially infected set in majority bootstrap percolation on the Erdős-Rényi random graph $G(n, p)$. This answers an open question by Janson, Luczak, Turova and Vallier (2012). Our results obtained for $p = c \log(n)/n$ are close to the results obtained by Balogh, Bollobás and Morris (2009) for majority bootstrap percolation on the hypercube. We conjecture that similar results will be true for all regular-like graphs with the same density and sufficiently strong expansion properties.

Keywords: bootstrap percolation, Erdős-Rényi random graph, threshold

*Partly supported by the Swedish Research Council.

1 Introduction

The classical bootstrap percolation, called r-neighbour bootstrap percolation, concerns a deterministic process on a graph. Firstly, a subset of the vertices of a graph G is initially infected. Then, at each time step the infection spreads to any vertex with at least r infected neighbours. This process is a cellular automaton, of the type first introduced by von Neumann in [13]. This particular model was introduced by Chalupa, Leith and Reich in [6], where G was taken to be the Bethe lattice.

A standard way of choosing the initially infected vertices is to independently infect each vertex with probability q. The probability that the entire graph eventually becomes infected (i.e., percolates) is increasing with q. It is therefore sensible to study the quantity $q_c = \inf\{q : \mathbb{P}_q(G \text{ percolates}) \geqslant c\}$, in particular the critical probability $q_{1/2}$ and the size of the critical window $q_{1-\epsilon} - q_\epsilon$.

A natural setting for this problem is the finite grid $[n]^d$. Many of the results on bootstrap percolation concern this problem. The first to study this graph were Aizenman and Lebowitz in [1], who showed that in 2-neighbour bootstrap percolation when d is fixed we have that $q_{1/2} = \Theta((\log n)^{1-d})$.

The r-neighbour bootstrap percolation process has also been studied on the random regular graph by Balogh in [3] and on the Erdős-Rényi random graph $G(n,p)$ by Janson, Łuczak, Turova and Vallier in [8].

In majority bootstrap percolation a vertex becomes infected if a majority of its neighbours are infected. In [2] Balogh, Bollobás and Morris studied this process on the hypercube and showed that if the vertices of the n-dimensional hypercube are independently infected with probability

$$q = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\log n}{n}} + \frac{\lambda \log\log n}{\sqrt{n \log n}},$$

then, with high probability, percolation occurs (i.e., all vertices eventually become infected) if $\lambda > \frac{1}{2}$ and does not occur if $\lambda \leqslant -2$.

In this paper we study majority bootstrap percolation on the Erdős-Rényi random graph $G(n,p)$ above the connectivity threshold. To determine the critical probability q_c in the majority bootstrap percolation model on $G(n,p)$ was stated as an open problem by Janson, Łuczak, Turova and Vallier in [8]. We answer this question by giving sharp asymptotics for the critical probability in Corollary 2, and giving a similar result in Theorem 1 for the critical size of the initially infected set (that determines whether or not percolation occurs). The results for $G(2^n, cn/2^n)$ that we establish in Corollary 2 are close to the aforementioned result for the hypercube in [2]. This leads us to believe that the second term in the asymptotic expansion of the infection probability will be of the same magnitude for all regular-like graphs on n vertices and $Cn\log(n)$ edges having sufficiently strong expansion properties. We leave this as an open question.

2 Main Results

In this section we shall state our main results, i.e., Theorem 1 and Corollary 2, and discuss two different ways of selecting the initially infected set. The lemmas that we use to prove Theorem 1 are stated and proved in Section 3 and Section 4; these lemmas also depend on inequalities that are proved separately in Appendix A and Appendix B.

For a graph G with some subset $I_0 \subset V(G)$ of initially infected vertices, the majority bootstrap process on G is defined by setting

$$I_{t+1} = I_t \cup \{v \in V(G) : |I_t \cap \Gamma(v)| \geqslant \frac{|\Gamma(v)|}{2}\},$$

where $\Gamma(v)$ is the neighbourhood of v. For a finite graph G, this process will terminate with $I_{T+1} = I_T$. Denote by $I = I_T$ the set of eventually infected vertices.

We shall look at the case of $G = G(n,p)$, the graph on n vertices, where each edge is included independently with probability p. Often $p := p(n) \to 0$ as $n \to \infty$, but we use the standard notation to just write p also for functions depending on n. Our initial setup is slightly different than for the hypercube mentioned above, instead of infecting each vertex independently with some probability q, we shall infect a random set of vertices of size $m := m(n)$.

In the normal setup for the majority bootstrap process on $G(n,p)$, we would first choose the edges of $G(n,p)$, and then choose an initially infected set I_0 of size m uniformly from $[n]^{(m)}$ (i.e., from all subsets of $[n]$ with m elements). As these two choices are independent we shall equivalently set $I_0 = [m]$, and then choose the edges of $G(n,p)$. This is the $MB(n,p\ ;m)$ process. We say that the $MB(n,p\ ;m)$ process percolates if $I = [n]$, i.e., if all vertices eventually become infected.

We now introduce some notation that shall be used. We use the standard asymptotic little-o notation and this is always taken as n (or N) tends to infinity, i.e., if (b_n) is a sequence of numbers, we say that $b_n = o(a_n)$ if $b_n/a_n \to 0$, as $n \to \infty$; however, sometimes we instead write $b_n \ll a_n$ and similarly we write $b_n \gg a_n$ if $b_n/a_n \to \infty$. We define

$$d = \frac{np}{1-p}, \tag{1}$$

thus d is roughly the average degree in $G(n,p)$ for $p = o(1)$. We denote the binomial distribution with parameters n and p by $B(n,p)$. We shall sometimes abuse the notation and denote by $B(n,p)$ a random variable that has a binomial distribution. We reserve m for the size of I_0 and shall always assume that

$$m = \frac{n}{2} - \frac{n}{2}\sqrt{\frac{\log d}{d}} + \lambda n \frac{\log\log\log d}{\sqrt{d \log d}} + o\left(n\frac{\log\log\log d}{\sqrt{d\log d}}\right), \tag{2}$$

for some constant λ and with d given in (1). We also use the standard notation that an event E_n holds *with high probability*, i.e., for the event E_n it holds that $\mathbb{P}(E_n) \to 1$, as $n \to \infty$. Let $\omega(n)$ denote some arbitrary positive function that is increasing and unbounded, as n tends to infinity.

The inequalities below are only claimed to be true for n large enough. For the $MB(n,p\,;m)$ process, define

$$\mathcal{P}_m\left(G(n,p)\right) = \mathbb{P}\left(I=[n]\right).$$

We shall now state the main result of this paper.

Theorem 1. *Fix some number $\epsilon>0$. Assume that for n large enough, $(1+\epsilon)\log n \leqslant p(1-p)n$ and $p \leqslant 1-\frac{(\log n)^4}{n}$. Let $d=\frac{np}{1-p}$. If the initially infected set I_0 has size*

$$m = \frac{n}{2} - \frac{n}{2}\sqrt{\frac{\log d}{d}} + \lambda n \frac{\log\log\log d}{\sqrt{d\log d}} + o\left(n\frac{\log\log\log d}{\sqrt{d\log d}}\right),$$

then

$$\mathcal{P}_m\left(G(n,p)\right) \overset{n\to\infty}{\longrightarrow} \begin{cases} 1, & \text{if } \lambda > \frac{1}{2},\\ 0, & \text{if } \lambda < 0. \end{cases}$$

Our second result concerns a more natural setup, where each vertex is initially independently infected with probability q, we have that, with high probability, $||I_0|-qn| \leqslant \omega(n)\sqrt{q(1-q)n}$. When $\sqrt{n} \ll \frac{n\log\log\log d}{\sqrt{d\log d}}$, i.e, when $p \ll \frac{(\log\log\log n)^2}{\log n}$, our result above shall still hold in this setting for $q=m/n$.

More formally define the $MB'(n,p\,;q)$ to be the process in which the graph $G(n,p)$ is chosen, and each vertex is initially infected independently with probability q. Then, the infection spreads by the majority bootstrap percolation process. For the process $MB'(n,p\,;q)$ define

$$\mathcal{P}'_q(G(n,p)) = \mathbb{P}(I=[n]).$$

Corollary 2. *Fix some number $\epsilon>0$. Assume that for n large enough, $(1+\epsilon)\log n \leqslant p(1-p)n$ and $p \leqslant 1-\frac{(\log n)^4}{n}$. Let $d=\frac{np}{1-p}$.*

If $p \ll \frac{(\log\log\log n)^2}{\log n}$, then with $q=\frac{1}{2}-\frac{1}{2}\sqrt{\frac{\log d}{d}}+\lambda\frac{\log\log\log d}{\sqrt{d\log d}}$, we have,

$$\mathcal{P}'_q\left(G(n,p)\right) \overset{n\to\infty}{\longrightarrow} \begin{cases} 1, & \text{if } \lambda > \frac{1}{2},\\ 0, & \text{if } \lambda < 0. \end{cases}$$

If $p \gg \frac{(\log\log\log n)^2}{\log n}$, then with $q=\frac{1}{2}-\frac{1}{2}\sqrt{\frac{\log d}{d}}+\theta\frac{1}{\sqrt{n}}$, we have,

$$\mathcal{P}'_q\left(G(n,p)\right) \to \Phi(2\theta),$$

where $\Phi(x)$ denotes the distribution function of the standard normal random variable.

Proof. As each vertex is infected independently, $|I_0|$ has distribution $B(n,q)$. Thus, with high probability, it holds that $||I_0|-qn| \leqslant \omega(n)\sqrt{q(1-q)n}$. If $p \ll \frac{(\log\log\log n)^2}{\log n}$, then $n\frac{\log\log\log d}{\sqrt{d\log d}} \gg \sqrt{n}$ and the result follows from Theorem 1.

If $p \gg \frac{(\log\log\log n)^2}{\log n}$, then for each fixed $\delta > 0$, by the Central Limit Theorem we obtain that

$$\begin{aligned}
\mathcal{P}'_q\left(G(n,p)\right) &= \sum_{m=0}^{n} \mathbb{P}\left(B(n,q)=m\right)\mathcal{P}_m\left(G(n,p)\right) \\
&\geqslant \mathbb{P}\left(B(n,q) \geqslant qn + (\delta-\theta)\sqrt{n}\right)\mathcal{P}_{\lfloor qn+(\delta-\theta)\sqrt{n}\rfloor}\left(G(n,p)\right) \\
&= \mathbb{P}\left(B(n,q)/\sqrt{q(1-q)n} \geqslant (qn+(\delta-\theta)\sqrt{n})/\sqrt{q(1-q)n}\right)(1+o(1)) \\
&\to \Phi(2(\theta-\delta)),
\end{aligned}$$

where the third line follows as

$$\mathcal{P}_{\lfloor qn+(\delta-\theta)\sqrt{n}\rfloor}\left(G(n,p)\right) \to 1$$

for $p \gg \frac{(\log\log\log n)^2}{\log n}$ by Theorem 1 and the fourth line by e.g., the Berry-Esséns inequality which implies that $B(n,q)$ is approximated by
$N(nq, q(1-q)n)$ with error at most $\frac{1}{\sqrt{n}}$. A similar argument shows that

$$1 - \mathcal{P}'_q(G(n,p)) \geqslant \Phi(-2(\theta+\epsilon))(1+o(1)),$$

and so,

$$\mathcal{P}'_q(G(n,p)) \to \Phi(2\theta). \qquad \square$$

Remark 3. If $p = c\frac{(\log\log\log n)^2}{\log n}$ for some positive constant c, then there is a positive probability that the $MB'(n,p\,;q)$ process percolates. This follows by the same arguments as in the proof of Corollary 2.

Remark 4. When p is smaller than the connectivity threshold, $G(n,p)$ contains isolated vertices. Due to the way we define the $MB(n,p\,;m)$ process, any uninfected isolated vertex becomes infected in the first time step, so this is not an obstruction to complete percolation. However, once p drops to below $\frac{\log n}{2n}$, then, with high probability, $G(n,p)$ contains isolated edges and neither endpoint of an isolated edge becomes infected if both endpoints are initially uninfected. This means that $\mathcal{P}_m(G(n,p)) \to 0$ unless $m = n - o(n)$.

Remark 5. Preliminary versions of this paper (including the same results) were included in the Phd thesis by Kettle [11] and in the PhD thesis by Juškevičius [9]. There is also a recent study by Stefánsson and Vallier [12] on this subject using completely different methods than those that are used in this paper (but using similar methods as was used by Janson, Łuczak, Turova and Vallier in [8]), where they show the first asymptotics of the thresholds $m \sim \frac{n}{2}$ in Theorem 1 above, and similarly thus the first asymptotics of the threshold $q \sim \frac{1}{2}$ in Corollary 2 above.

3 Upper Bound

As G is finite the $MB(n,p\,;m)$ process will eventually terminate with some set $I \subset [n]$ of infected vertices. If we do not infect the whole graph, or, equivalently, we have that $I \neq [n]$, then we can say something about the structure of I. In the definition below we introduce the notion of a closed set, which is important in our proof of the upper bound of Theorem 1.

Definition 6. *We shall call a proper subset S of $[n]$* closed *if for all $v \in [n] \setminus S$ we have that $|\Gamma(v) \cap S| < \frac{|\Gamma(v)|}{2}$.*

Recall that I_0 is the set of initially infected vertices and that a vertex $v \in I_{t+1}$, if either $v \in I_t$ or if at least half of its neighbours lies in I_t. In particular $I_t \subseteq I_{t+1}$. If the $MB(n,p\,;m)$ process does not percolate, let T be such that the process has stabilized, i.e., $I = I_T = I_{T+1} \neq [n]$. Then, I is a closed set, and thus we must have that the initially infected vertices I_0 is a subset of a closed set. Recall that m is the size of I_0 and that we assume that m satisfies the equality in (2). We shall show that, if $\lambda > \frac{1}{2}$, then, with high probability, I_0 is contained in no closed sets in three stages. Using Lemma 8 will allow us that, with high probability, the graph $G(n,p)$ has no "large" closed sets. After that we shall bound the expected number of medium sized closed sets that I_0 is contained in, hence by the Markov inequality it will follow that, with high probability, there are no medium sized closed sets containing I_0. But before we proceed with proving these two facts, we shall show that, with high probability, the number of infected vertices after one time step, $|I_1|$, is large, and so I_0 can rarely be contained in a small closed set. We assume that for some fixed $\epsilon > 0$ and for n large enough, it holds that $(1+\epsilon)\log n \leqslant p(1-p)n$ and that $p \leqslant 1 - \frac{(\log n)^4}{n}$. However, for some of the results below it is enough to assume that $p(1-p)n \to \infty$.

Lemma 7. *In the $MB(n,p\,;m)$ process, with m given in (2), d given in (1) and with $\lambda > \frac{1}{2}$, we have that*

$$|I_1 \setminus I_0| \geqslant \frac{n(\log\log d)^{2\lambda}}{e^8\sqrt{d\log d}},$$

with high probability.

Proof. For $i \in [n] \setminus I_0$, denote by A_i the event that vertex i is infected at time one, that is the event that i has fewer neighbours in $[n] \setminus I_0$ than it does in I_0. The events A_i are identical and very weakly correlated but not independent. Let X be the number of vertices infected at the first step of the process. Then, $X = |I_1 \setminus I_0| = \sum \mathbf{1}(A_i)$. We shall use Chebyshev's inequality to bound the probability that X is small.

As the events A_i are identical we shall set $r = \mathbb{P}(A_i)$, so $\mathbb{E}(X) = (n-m)r$. Let $B(m,p)$ and $B(n-m-1, 1-p)$ be independent random variables with means μ_1 and μ_2, respectively. We have that

$$\begin{aligned} r &= \mathbb{P}\left(|\Gamma(i) \cap I_0| \geqslant |\Gamma(i) \setminus I_0|\right) \\ &= \mathbb{P}\left(B(m,p) \geqslant B(n-m-1,p)\right) \\ &= \mathbb{P}\left(B(m,p) + B(n-m-1,1-p) \geqslant \mu_1 + \mu_2 + p(n-2m-1)\right). \end{aligned} \tag{3}$$

We will now apply the bound from Proposition 24 to the last equality in (3) with the parameters $N = \frac{n-1}{2}$, $S = \frac{n-1-2m}{2}$ and $h = p(n-2m-1)$. We first check that the conditions on these parameters hold true.

Recalling from (1) that $d = \frac{np}{1-p}$ and using the expression of m in (2) we note that

$$2p(n-2m-1) \sim np\sqrt{\frac{\log d}{d}} = \sqrt{np(1-p)}\sqrt{\log(\frac{np}{1-p})}, \tag{4}$$

$$2p(n-2m-1)^2 \sim n^2p^2\frac{\log d}{d} = n(1-p)\log(\frac{np}{1-p}). \tag{5}$$

Hence, for $p \gg \frac{1}{n}$, it follows from (4) that $h = \omega(n)\sqrt{p(1-p)n}$ and from (5) it follows that $hS = o(n\sqrt{p(1-p)n})$. Furthermore, for $p \leqslant 1 - \frac{(\log n)^4}{n}$, it follows from (4) that $h = o((p(1-p)n)^{\frac{2}{3}})$, since this holds when $\log(\frac{np}{1-p}) = o((p(1-p)n)^{\frac{1}{3}})$. Hence, the conditions on the parameters in Proposition 24 hold. Applying the bound from Proposition 24 to the last equality in (3) (with the parameters $N = \frac{n-1}{2}$, $S = \frac{n-1-2m}{2}$ and $h = p(n-2m-1)$), we obtain that

$$r > \frac{\sqrt{p(1-p)(n-1)}}{2\pi p(n-2m-1)}\exp\left(-\frac{p(n-2m-1)^2}{2(1-p)(n-1)} - 4 - o(1)\right). \tag{6}$$

Again recalling from (1) that $d = \frac{np}{1-p}$ we can simplify the expression in (6) by observing from (2) that we have the asymptotic relation

$$d(n-2m-1)^2 = n^2\log d - 4\lambda n^2 \log\log\log d + o(n^2)$$

and applying (4) to deduce that

$$\begin{aligned} r &> \frac{1}{2\pi\sqrt{\log d}}\exp\left(-\frac{\log d}{2} + 2\lambda\log\log\log d - 4 + o(1)\right) \\ &> \left(\frac{(\log\log d)^{2\lambda}}{2\pi e^4\sqrt{d\log d}}\right)(1+o(1)). \end{aligned} \tag{7}$$

The variance of X is equal to

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{i,j\in[n]\setminus[m]} (\mathbb{P}(A_j|A_i) - \mathbb{P}(A_j))\mathbb{P}(A_i) \\ &= (1-r)r(n-m) + r'r(n-m)(n-m-1), \end{aligned} \tag{8}$$

where $r' = \mathbb{P}(A_j|A_i) - r$, this being the same for any $i \neq j$. In (8), the first term is the sum over $i = j$ and the second term is the sum over $i \neq j$. Let B_{ij} and $\overline{B}_{ij}$ be the events that ij is, or is not, an edge in G respectively; hence, $\mathbb{P}(B_{ij}) = p$ and $\mathbb{P}(\overline{B}_{ij}) = 1-p$. Note that

$$\mathbb{P}(A_j|B_{ij}) = \mathbb{P}\left(B(m,p) \geqslant B(n-m-2,p) + 1\right) \tag{9}$$

and

$$\mathbb{P}\left(A_j|\overline{B}_{ij}\right) = \mathbb{P}\left(B(m,p) \geqslant B(n-m-2,p)\right), \tag{10}$$

where $B(m,p)$ and $B(n-m-2,p)$ are independent random variables. Note that

$$\mathbb{P}(A_j|B_{ij}) \leqslant \mathbb{P}\left(A_j|\overline{B}_{ij}\right),$$

hence we may bound r' by

$$\begin{aligned} r' = \mathbb{P}(A_j|A_i) - \mathbb{P}(A_j) &= \mathbb{P}(A_j|B_{ij})\mathbb{P}(B_{ij}|A_i) + \mathbb{P}\left(A_j|\overline{B}_{ij}\right)\mathbb{P}\left(\overline{B}_{ij}|A_i\right) - \mathbb{P}\left(A_j\right) \\ &\leqslant \mathbb{P}\left(A_j|\overline{B}_{ij}\right) - \mathbb{P}\left(A_j\right) \\ &= p\left(\mathbb{P}\left(A_j|\overline{B}_{ij}\right) - \mathbb{P}\left(A_j|B_{ij}\right)\right) \\ &= p\mathbb{P}\left(B(m,p) = B(n-m-2,p)\right), \end{aligned} \tag{11}$$

where the second last equality follows from the law of total probability, i.e.,

$$P(A_j) = \mathbb{P}\left(A_j|B_{ij}\right)\mathbb{P}\left(B_{ij}\right) + \mathbb{P}\left(A_j|\overline{B}_{ij}\right)\mathbb{P}\left(\overline{B}_{ij}\right),$$

and the last equality follows from (9) and (10).

In order to compute an upperbound of r' in (11) we apply Proposition 26 with the parameters $N = \frac{n}{2} - 1$, $S = \frac{n}{2} - m - 1$ and $T = 0$ to the last equality in (11). By using similar calculations as in (4)–(5) we can check that the conditions on these parameters are fulfilled, and thus by similar calculations as in (6)–(7) we get from Proposition 26, that r' is at most

$$\begin{aligned} &\frac{p(\frac{n}{2}-m-1)}{2\pi(1-p)(\frac{n}{2}-1)}\exp\left(-\frac{p(\frac{n}{2}-m-1)^2}{(1-p)(\frac{n}{2}-1)} + o(1)\right) \\ &+ \frac{3}{\pi(\frac{n}{2}-m-1)}\exp\left(-\frac{9p(\frac{n}{2}-m-1)^2}{8(1-p)(\frac{n}{2}-1)}\right) \\ &< \frac{p\sqrt{\log d}}{2\pi(1-p)\sqrt{d}}\exp\left(-\frac{\log d}{2} + 2\lambda\log\log\log d + o(1)\right) \\ &+ \frac{6\sqrt{d}}{\pi n\sqrt{\log d}}\exp\left(-\frac{9\log d}{16} + \frac{9\lambda\log\log\log d}{4} + o(1)\right). \end{aligned}$$

The second term is much smaller than the first term, and so (for n large enough),

$$r' < \left(\frac{\sqrt{\log d}(\log\log d)^{2\lambda}}{\pi n}\right). \tag{12}$$

We are now able to bound the probability that X is small. From (8) and Chebyshev's inequality, we get that

$$\begin{aligned}\mathbb{P}\left(X \leqslant \frac{(n-m)r}{2}\right) &\leqslant \mathbb{P}\left(|X-(n-m)r| \geqslant \frac{(n-m)r}{2}\right)\\ &\leqslant \frac{4\mathrm{Var}(X)}{((n-m)r)^2}\\ &= \frac{4((1-r)+(n-m-1)r')}{(n-m)r}\\ &< \frac{4r'}{r}+o(1).\end{aligned}$$

From (12) and (7) this is at most

$$\left(\frac{2e^4\sqrt{d}\log d}{n}\right)(1+o(1))+o(1)=o(1);$$

recalling that we assume that $p \leqslant 1-\frac{\log^4 n}{n}$.

Hence, with high probability, we have that $|I_1 \setminus I_0|$ is at least $\frac{(n-m)r}{2}$. By using (7) we get that for large n,

$$\frac{(n-m)r}{2} \geqslant \frac{n(\log\log d)^{2\lambda}}{e^8\sqrt{d\log d}},$$

which completes the proof. $\square$

We now show that $G(n,p)$ contains no large closed sets by a simple edge set comparison.

Lemma 8. *Suppose that for some fixed $\epsilon > 0$ we have,*

$$p(1-p)n \geqslant (1+\epsilon)\log n,$$

for n large enough. Then, with high probability, $G(n,p)$ contains no closed set of size greater than $\frac{n}{2}+\frac{7n}{2\sqrt{d}}$.

Proof. Let us write s for the size of the set S i.e., $s=|S|$. In order for the set S to be closed, each vertex $v \in [n]\setminus S$ has to have the majority of its neighbours outside S. In other words, we must have $|\Gamma(v)\cap([n]\setminus S)| > |\Gamma(v)\cap S|$. Summing over the vertices in $[n]\setminus S$, we have that the number of edges from S to $[n]\setminus S$ must be fewer than twice the number of edges in $[n]\setminus S$. We will split the proof into a few cases. We will first consider the case when S is of size s such that $\frac{n}{2}+\frac{7n}{2\sqrt{d}} < s < \frac{4n}{5}$.

If $\frac{n}{2}+\frac{7n}{2\sqrt{d}} < s < \frac{4n}{5}$, by Proposition 27 every set of size $n-s$ has at most

$$p\binom{n-s}{2}+2(n-s)\sqrt{p(1-p)(n-s)}$$

edges with probability at least $1 - \frac{1}{4^{n-s}}$, and by Proposition 28 every set S of size s has at least

$$ps(n-s) - 3(n-s)\sqrt{p(1-p)s}$$

edges between it and its complement with probability at least $1 - \frac{1}{4^{n-s}}$.

Recall that $d = \frac{np}{1-p}$. If $\frac{n}{2} + \frac{7n}{2\sqrt{d}} < s < \frac{4n}{5}$, then $p(2s-n) \geqslant 7\sqrt{p(1-p)n}$, and so,

$$ps(n-s) - 3(n-s)\sqrt{p(1-p)s} > 2p\binom{n-s}{2} + 4(n-s)\sqrt{p(1-p)(n-s)}.$$

Therefore, with high probability, every set S of size s, such that

$$\frac{n}{2} + \frac{7n}{2\sqrt{d}} < s < \frac{4n}{5}$$

is not closed.

If $s \geqslant \frac{4n}{5}$ and $p(1-p)n \geqslant 4\log n$, then we know from Proposition 29 that with probability at least $1 - n^{-\frac{n-s}{120}}$ there does not exist a closed set of size s in $G(n,p)$. The result follows as $\sum_{i\geqslant 1} n^{-\frac{i}{120}} = o(1)$.

If $n - n^{\frac{27}{28}} \geqslant s \geqslant \frac{4n}{5}$ and $5\log n \geqslant p(1-p)n \geqslant (1+\epsilon)\log n$, then we know from Corollary 30 that with probability at least $1 - n^{-\frac{n-s}{120}}$ there does not exist a closed set of size s in $G(n,p)$.

If $s \geqslant n - n^{\frac{27}{28}}$ and $5\log n \geqslant p(1-p)n \geqslant (1+\epsilon)\log n$, then we know from Proposition 32 that with probability at least $1 - n^{-\frac{n-s}{120}}$ every set $S^C := [n] \setminus S$ of size $n-s$ has at most $2(n-s)$ edges, and so the subgraph of G induced by S^C has a vertex v_{S^C} of degree at most 4. By Proposition 31 we have that, with high probability, the minimum degree of $G(n,p)$ is at least 9, and so v_{S^C} will become infected if all of S is infected, and so S is not closed. □

Lastly, we turn to bounding the expected number of medium sized closed sets I_0 is contained in. We shall therefore want a bound on the probability that a set S of size at least s in a particular range of s is closed. To do this we shall pick a test set Q of a suitable size and bound the probability that none of the vertices in Q are infected by S.

Lemma 9. *Fix $\epsilon > 0$ and define*

$$s = \left\lfloor \frac{n}{2} - \frac{n\sqrt{\log d}}{2\sqrt{d}} + \frac{n(\log\log d)^{1+\epsilon}}{\sqrt{d\log d}} \right\rfloor. \tag{13}$$

Take any set of vertices S in $G(n,p)$, such that $s \leqslant |S| < \frac{2n}{3}$. Then, for n large enough,

$$\mathbb{P}(S \textit{ is closed}) \leqslant \exp\left(-\frac{n(\log d)^{(\log\log d)^{\epsilon}-2}}{e^7\sqrt{d}}\right). \tag{14}$$

Proof. Let S be a set of vertices, such that $s \leqslant |S| \leqslant \frac{2n}{3}$. Consider a set $Q \subset V(G) \setminus S$ of size $|Q| = t = \left\lfloor \frac{n}{(\log d)^2} \right\rfloor$. We shall condition on the edge set of Q since then the events F_v, that v is not infected by S for each vertex $v \in Q$, are independent.

Denote by $E = E(Q)$ the edge set of Q, and set $d_E(v)$ to be the degree of vertex $v \in Q$, when Q has edge set E. We have that

$$\mathbb{P}(F_v|E) = \mathbb{P}(|\Gamma(v) \cap S| < d_E(v) + |\Gamma(v) \setminus (S \cup Q)|).$$

Therefore,

$$\begin{aligned}\mathbb{P}(S \text{ is closed}) &\leqslant \sum_E \mathbb{P}(E) \prod_{v \in Q} \mathbb{P}(F_v|E) \\ &= \sum_E \mathbb{P}(E) \prod_{v \in Q} \mathbb{P}(B(|S|, p) < B(n - |S| - t, p) + d_E(v)),\end{aligned}$$

where $\mathbb{P}(E)$ is the probability of a particular edge set $E \subset \{0,1\}^{\binom{t}{2}}$ and is equal to $p^{|E|}(1-p)^{\binom{t}{2} - |E|}$.

The function $f_{|S|}(x) = \mathbb{P}(B(|S|, p) < B(n - |S| - t, p) + x)$ (for independent binomial random variables $B(|S|, p)$ and $B(n - |S| - t, p)$) is decreasing in $|S|$, so we have $f_s(x) \geqslant f_{|S|}(x)$. Let us suppress the dependency on s by writing $f(x)$ instead of $f_s(x)$. Thus, we write

$$f(x) = \mathbb{P}(B(s, p) < B(n - s - t, p) + x). \tag{15}$$

We have,

$$\mathbb{P}(S \text{ is closed}) \leqslant \sum_E \mathbb{P}(E) \prod_{v \in Q} f(d_E(v)). \tag{16}$$

The rest of the proof shall be spent bounding (16). The degrees of the vertices in Q are heavily concentrated around pt, and we shall expand f around pt to show that (16) is not much larger than $f(pt)^t$.

We have by Corollary 16 that f is log-concave, and so for any x and y with $f(y) \neq 0$,

$$f(x) \leqslant f(y) \left(\frac{f(y+1)}{f(y)} \right)^{x-y}. \tag{17}$$

Setting $y = \lceil pt \rceil \in \mathbb{N}$, we get from (16) and (17) that

$$\begin{aligned}\mathbb{P}(S \text{ is closed}) &\leqslant \sum_E \mathbb{P}(E) \prod_{v \in Q} f(y) \left(\frac{f(y+1)}{f(y)} \right)^{d_E(v)-y} &(18) \\ &= \sum_E \mathbb{P}(E) f(y)^t \left(\frac{f(y+1)}{f(y)} \right)^{2|E|-ty}. &(19)\end{aligned}$$

There is no dependence on E other than its size, and so,

$$\mathbb{P}(S \text{ is closed}) \leqslant \sum_{i=0}^{\binom{t}{2}} \binom{\binom{t}{2}}{i} p^i (1-p)^{\binom{t}{2}-i} f(y)^t \left(\frac{f(y+1)}{f(y)}\right)^{2i-ty}$$
$$= \left(1-p+p\left(\frac{f(y+1)}{f(y)}\right)^2\right)^{\binom{t}{2}} \left(\frac{f(y)}{f(y+1)}\right)^{ty} f(y)^t. \tag{20}$$

Recall that $y = \lceil pt \rceil$. Setting $\frac{f(y+1)}{f(y)} = 1+a$, we bound (20) using the inequalities $1+w \leqslant e^w$ and $(1+x)^{-1} \leqslant 1-x+x^2$ for $x \geqslant 0$ to get

$$\mathbb{P}(S \text{ is closed}) \leqslant \left(1+2ap+a^2p\right)^{\frac{t^2}{2}} \left(\frac{1}{1+a}\right)^{pt^2} f(y)^t$$
$$\leqslant \exp\left((2ap+a^2p)\frac{t^2}{2} + (a^2-a)pt^2\right) f(y)^t$$
$$= \exp\left(\frac{3pa^2t^2}{2}\right) f(y)^t. \tag{21}$$

We have that

$$f(y+1) = f(y) + \mathbb{P}\left(B(s,p) = B(n-s-t,p)+y\right).$$

Let us write $z = \mathbb{P}\left(B(s,p) = B(n-s-t,p)+y\right)$ to ease up the notation. Thus, $f(y+1) = f(y)+z$. Note that $a = \frac{z}{f(y)}$. We will now bound z by applying Proposition 26 with the parameters $N = \frac{n-t+T}{2}$, $S = \frac{n-2s-t+T}{2}$ and $T = \frac{\lceil pt \rceil}{p}$. Note that $0 \leqslant T-t < p^{-1}$ and recall the definition of s in (13). That the conditions on the parameters are satisfied again follow by the same calculations as in (4)–(5). By using similar calculations as in (6)–(7) we get

$$z < \frac{n-2s+\frac{1}{p}}{2\pi(1-p)n} \exp\left(-\frac{2p(\frac{n}{2}-s)^2}{(1-p)(n-t)} + o(1)\right)$$
$$+ \frac{6}{\pi(n-2s)} \exp\left(-\frac{9p(n-2s)^2}{16(1-p)(n+\frac{1}{p})}\right)$$
$$= \frac{\sqrt{\log d}}{2\pi(1-p)\sqrt{d}} \exp\left((-\frac{\log d}{2} + 2(\log\log d)^{1+\epsilon})(1+\frac{t}{n}) + o(1)\right)$$
$$+ \frac{6\sqrt{d}}{\pi n\sqrt{\log d}} \exp\left(-\frac{9\log d}{16} + \frac{9(\log\log d)^{1+\epsilon}}{4} + o(1)\right). \tag{22}$$

The second term in (22) is much smaller than the first so as $6 < 2\pi$ and $t\log d = \left\lfloor \frac{n}{(\log d)^2} \right\rfloor \log d = o(n)$, we get (for n large enough) that

$$z < \frac{\sqrt{\log d}(\log d)^{2(\log\log d)^\epsilon}}{6(1-p)d}. \tag{23}$$

Recalling the definition of f in (15), we can rewrite $f(y)$ as

$$f(y) = 1 - \mathbb{P}\left(B(s,p) + B(n-s-t,(1-p)) \geqslant n-s-t+y\right).$$

We will now bound $f(y)$ by applying Proposition 24 with the parameters $N = \frac{n-t}{2}$, $S = \frac{n-2s-t}{2}$ and $h = p(n-2s) + y - pt$.

Note that $0 \leqslant y - pt < 1$. Recall the definition of s in (13). For $p \gg \frac{1}{n}$ it holds that

$$(p(n-2s)+1)(t+2s-n) = o(n\sqrt{np(1-p)});$$

hence, $hS = o(n\sqrt{np(1-p)})$. Again by applying the same calculations as in (4) it follows that the conditions on h hold. Thus, we obtain, for n large enough, that

$$\begin{aligned} f(y) &< 1 - \frac{\sqrt{p(1-p)(n-t)}}{2\pi(p(n-2s)+1)} \exp\left(-\frac{(p(n-2s)+1)^2}{2p(1-p)(n-t)} - 4 - o(1)\right) \\ &< 1 - \frac{(\log d)^{2(\log\log d)^\epsilon}}{e^6\sqrt{d\log d}} \\ &< \exp\left(-\frac{(\log d)^{(\log\log d)^\epsilon}}{e^6\sqrt{d}}\right). \end{aligned} \tag{24}$$

The second inequality in (24) follows from the same reasoning used in (22)–(23) and that $e^6 > 2\pi e^4$.

We can also apply Proposition 25 (with the same parameters as above) to get a lower bound on $f(y)$ (for n large enough) of

$$f(y) > 1 - \frac{\sqrt{p(1-p)(n-t)}}{p(n-2s)} \exp\left(-\frac{p(n-2s)^2}{2(1-p)(n-t)} + 3 + o(1)\right) > \frac{1}{2}, \tag{25}$$

here the bound on $1 - f(y)$ is actually $o(1)$, being within a constant factor of the bound in (24).

Applying (23), (24) and (25) we are now able to get a good upper bound on a,

$$a = \frac{z}{f(y)} < \frac{\sqrt{\log d}(\log d)^{2(\log\log d)^\epsilon}}{3(1-p)d}.$$

Substituting these bounds into (21) we get (for n large enough) that

$$\mathbb{P}(S \text{ is closed}) < \exp\left(\frac{p(\log d)^{4(\log\log d)^\epsilon} n}{6(1-p)^2 d^2 \log d} - \frac{(\log d)^{(\log\log d)^\epsilon}}{e^6\sqrt{d}}\right)^t.$$

The second term in the exponential is much larger than the first term, and so (for n large enough),

$$\begin{aligned}\mathbb{P}(S \text{ is closed}) &< \exp\left(-\frac{(\log d)^{(\log\log d)^\epsilon}}{2e^6\sqrt{d}}\right)^t \\ &< \exp\left(-\frac{n(\log d)^{(\log\log d)^\epsilon-2}}{e^7\sqrt{d}}\right),\end{aligned}$$

as $t > \frac{2n}{e(\log d)^2}$. □

We shall now bound the expected number of closed sets in the medium sized range that contains I_0, this is also a bound on the probability that I_0 is contained in such a medium sized closed set.

Proposition 10. *Assume that*

$$m = \frac{n}{2} - \frac{n\sqrt{\log d}}{2\sqrt{d}} + \frac{n\lambda\log\log\log d}{\sqrt{d\log d}} + o\left(n\frac{\log\log\log d}{\sqrt{d\log d}}\right),$$

and choose some $\epsilon > 0$. Then, the expected number of closed sets in $G(n,p)$ of size between

$$\frac{n}{2} - \frac{n\sqrt{\log d}}{2\sqrt{d}} + \frac{n(\log\log d)^{1+\epsilon}}{\sqrt{d\log d}} \quad \text{and} \quad \frac{n}{2} + \frac{4n}{\sqrt{d}}$$

that contain $I_0 = [m]$ is $o(1)$.

Proof. Let S be a set of size s in our range. For each possible value of s and n large enough (using Stirling's formula) there are at most

$$\binom{n-m}{s-m} < \binom{n}{\lfloor\frac{n\sqrt{\log d}}{\sqrt{d}}\rfloor} < \left(\frac{e\sqrt{d}}{\sqrt{\log d}}\right)^{\frac{n\sqrt{\log d}}{\sqrt{d}}} < \exp\left(\frac{n(\log d)^{\frac{3}{2}}}{\sqrt{d}}\right) \tag{26}$$

possible closed sets that can contain I_0. Note that s can have at most $\left\lfloor\frac{n\sqrt{\log d}}{\sqrt{d}}\right\rfloor$ different values. Hence, using this fact together with the bound in (26) and the bound of $\mathbb{P}(S \text{ is closed})$ in (14) of Lemma 9, we get that the expected number of closed sets is (for n large enough) less than

$$\frac{n\sqrt{\log d}}{\sqrt{d}}\exp\left(\frac{n(\log d)^{\frac{3}{2}}}{\sqrt{d}} - \frac{n(\log d)^{(\log\log d)^\epsilon-2}}{e^7\sqrt{d}}\right),$$

and this is $o(1)$ as $(\log\log d)^\epsilon$ is unbounded. □

Corollary 11. *Fix some number $\epsilon > 0$. Assume that for n large enough, $(1+\epsilon)\log n \leqslant p(1-p)n$ and $p \leqslant 1 - \frac{(\log n)^4}{n}$. Let $d = \frac{np}{1-p}$. If the initially infected set I_0 has size*

$$m = \frac{n}{2} - \frac{n}{2}\sqrt{\frac{\log d}{d}} + \lambda n\frac{\log\log\log d}{\sqrt{d\log d}} + o\left(n\frac{\log\log\log d}{\sqrt{d\log d}}\right),$$

then for $\lambda > \frac{1}{2}$, with high probability, the $MB(n,p\,;m)$ process percolates.

Proof. We have from Lemma 7 that, with high probability, $I_0 = [m]$ is not contained in a closed set of size less than

$$\frac{n}{2} - \frac{n\sqrt{\log d}}{2\sqrt{d}} + \frac{n(\log\log d)^{2\lambda}}{e^8\sqrt{d\log d}}.$$

Using the Markov inequality it follows from Proposition 10 applied to $\epsilon = \lambda - \frac{1}{2}$ that, with high probability, I_0 is not contained in a closed set of size between

$$\frac{n}{2} - \frac{n\sqrt{\log d}}{2\sqrt{d}} + \frac{n(\log\log d)^{\lambda+\frac{1}{2}}}{\sqrt{d\log d}} \quad \text{and} \quad \frac{n}{2} + \frac{4n}{\sqrt{d}}.$$

We have from Lemma 8 that, with high probability, I_0 is not contained in a closed set of size greater than

$$\frac{n}{2} + \frac{7n}{2\sqrt{d}},$$

and so, for $\lambda > \frac{1}{2}$, with high probability, I_0 is not contained in any closed set in $G(n,p)$ and hence percolates. $\square$

4 Lower Bound

In this section we shall show the lower bound of Theorem 1. We show the following result.

Lemma 12. *Fix some number $\epsilon > 0$. Assume that for n large enough, $(1+\epsilon)\log n \leqslant p(1-p)n$ and $p \leqslant 1 - \frac{(\log n)^4}{n}$. Let $d = \frac{np}{1-p}$. If the initially infected set I_0 has size*

$$m = \frac{n}{2} - \frac{n}{2}\sqrt{\frac{\log d}{d}} + \lambda n\frac{\log\log\log d}{\sqrt{d\log d}} + o\left(n\frac{\log\log\log d}{\sqrt{d\log d}}\right),$$

then, for $\lambda < 0$, with high probability, the $MB(n,p\ ;m)$ process does not percolate.

Remark 13. Note that Lemma 12 and Corollary 11 prove Theorem 1.

In fact to prove Lemma 12, as might be expected, we shall show that, with high probability, the $MB(n,p\ ;m)$ process terminates with I (the set of eventual infected vertices) only slightly larger than $|I_0| = m$. We shall do this by bounding the expected number of sets of some size that could be the first vertices to be infected.

Proof. Recall that I_j is the set of vertices that are infected in time step j.

Let t be a natural number. We want to construct a set $Q \subset I \setminus I_0$ such that $|Q| = t$ in the following manner. We pick elements from I_1 until we get t elements. If $t > |I_1|$ we take all elements from I_1 and continue to take elements from I_2 in the same manner until we get t elements in total. In general we continue to pick elements from I_j until we get t elements in total, if we need more elements than I_j can provide we take all of them and continue to I_{j+1}. The set Q will then have the property that for each infected node, all the nodes which helped infect it (except the nodes from I_0) are also included in Q. If a set $Q \in V(G) \setminus I_0$ has this property we say that Q percolates. Note that if the entire graph percolates (i.e. the $MB(n,p\ ;m)$ process percolates) we can always find such a percolating set Q for every value of t between 0 and $n - |I_0|$.

Our strategy will be to show that if $\lambda < 0$, then, with high probability, there is no percolating set Q of a particular size and thus the $MB(n,p\,;m)$ process does not percolate.

Set $t = |Q|$, and denote by $E = E(Q)$ the edge set of Q. Write $d_E(i)$ for the degree within Q of a vertex $i \in Q$. We want to bound the probability that Q percolates. To do so, we modify the infection rule within Q so that the vertices inside Q consider their neighbours in Q to be already infected, regardless of their real state at any particular time step. The latter assumption only increases the probability and, more importantly (after conditioning on E), makes the events for vertices in Q to be infected independent. This is because these events now only depend on how many edges each vertex has to I_0 and $V(G) \setminus (I_0 \cup Q)$. Thus, if we consider the conditional probability $\mathbb{P}(Q \text{ percolates}|E)$, and then take the expectation we get

$$\mathbb{P}(Q \text{ percolates}) \leqslant \sum_E \mathbb{P}(E) \prod_{i=1}^{t} \mathbb{P}(B(m,p) + d_E(i) \geqslant B(n-m-t,p)) \tag{27}$$

(for independent random variables $B(m,p)$ and $B(n-m-t,p)$). Note that $B(m,p)+d_E(i)$ in (27) counts the number of neighbours between the vertex $i \in Q$ to the rest of the vertices in Q and to the set I_0, while $B(n-m-t,p)$ counts the number of neighbours between the vertex $i \in Q$ to $V(G) \setminus (I_0 \cup Q)$. This means that $\mathbb{P}(B(m,p) + d_E(i) \geqslant B(n-m-t,p))$ is an upper bound on the probability that i gets infected. Denote

$$g(x) = \mathbb{P}\left(B(m,p) + x \geqslant B(n-m-t,p)\right). \tag{28}$$

Due to the log-concavity of g (Corollary 16) we have for integers x, y,

$$g(x) \leqslant g(y) \left(\frac{g(y+1)}{g(y)}\right)^{x-y}.$$

By applying the latter inequality with $x = d_E(i)$ and $y = \lceil pt \rceil$ and using similar calculations as in (18) and (20), we can bound (27) by

$$\begin{aligned}
\sum_E \mathbb{P}(E) \prod_{i=1}^{t} g(y) \left(\frac{g(y+1)}{g(y)}\right)^{d_E(i)-y} &= \sum_E \mathbb{P}(E) g(y)^t \left(\frac{g(y+1)}{g(y)}\right)^{2|E|-ty} \\
= \sum_{j=0}^{\binom{t}{2}} \binom{\binom{t}{2}}{j} p^j (1-p)^{\binom{t}{2}-j} g(y)^t \left(\frac{g(y+1)}{g(y)}\right)^{2j-ty} \\
= \left(1 - p + p\left(\frac{g(y+1)}{g(y)}\right)^2\right)^{\binom{t}{2}} \left(\frac{g(y)}{g(y+1)}\right)^{ty} g(y)^t.
\end{aligned}$$

Substituting $\frac{g(y+1)}{g(y)} = 1+a$ and the elementary inequality $1/(1+a) \leqslant 1-a+a^2$, we bound the latter expression by

$$\begin{aligned}
&\left(1-p+p(1+a)^2\right)^{\binom{t}{2}} \left(1-a+a^2\right)^{ty} g(y)^t \\
&\leqslant \exp\left((2ap+a^2p)\frac{t^2}{2} + (a^2-a)pt^2\right) g(y)^t \\
&= \left(\exp\left(\frac{3pa^2t}{2}\right) g(y)\right)^t .
\end{aligned}$$

Hence,

$$\mathbb{P}(Q \text{ percolates}) \leqslant \left(\exp\left(\frac{3pa^2t}{2}\right) g(y)\right)^t . \tag{29}$$

We have by the definition of g in (28) (and recalling that $y = \lceil pt \rceil$) that $g(y)$ is equal to

$$g(y) = \mathbb{P}\left(X_1 + X_2 \geqslant \mu_1 + \mu_2 + pn - 2pm - pt - y\right),$$

where $X_1 = B(m,p)$ with mean μ_1 and $X_2 = B(n-m-t,(1-p))$ with mean μ_2. Setting $t = \left\lfloor n(\log\log d)^\lambda/\sqrt{d\log d}\right\rfloor$ and using Proposition 25 (with the parameters $N = \frac{n-t}{2}$, $S = \frac{n-2m-t}{2}$ and $h = p(n-2m-t) - y$) to bound $g(y)$ (again the conditions on the parameters N, S and h are fulfilled by the calculations in (4)– (5)). Thus, we obtain (for n large enough) and $\lambda < 0$ that

$$\begin{aligned}
g(y) &< \frac{\sqrt{p(1-p)(n-t)}}{pn-2pm-2pt-1} \exp\left(-\frac{(pn-2pm-2pt-1)^2}{2p(1-p)(n-t)} + 3 + o(1)\right) \\
&< \frac{e^3}{\sqrt{\log d}} \exp\left(-\frac{\log d}{2} + 2\lambda \log\log\log d + O((\log\log d)^\lambda)\right) \\
&< \left(\frac{e^4(\log\log d)^{2\lambda}}{\sqrt{d\log d}}\right).
\end{aligned} \tag{30}$$

We can also bound $g(y)$ in (28) from below by Proposition 24,

$$\begin{aligned}
g(y) &> \frac{\sqrt{p(1-p)(n-t)}}{2\pi(pn-2pm-2pt)} \exp\left(-\frac{(pn-2pm-2pt)^2}{2p(1-p)(n-t)} - 4 - o(1)\right) \\
&> \frac{1}{2\pi e^4\sqrt{\log d}} \exp\left(-\frac{\log d}{2} + 2\lambda \log\log\log d + O((\log\log d)^\lambda)\right) \\
&> \left(\frac{(\log\log d)^{2\lambda}}{e^6\sqrt{d\log d}}\right),
\end{aligned} \tag{31}$$

when $\lambda < 0$ (and n is large enough).

By the definition of g in (28) we have that

$$g(y+1) = g(y) + \mathbb{P}(B(m,p) + y + 1 = B(n-m-t,p)).$$

Let us write $z = \mathbb{P}(B(m,p) + y + 1 = B(n-m-t,p))$ for convenience. We shall now obtain an upper bound for z. Using Proposition 26 (with the parameters $T = -\frac{y+1}{p}$, $N = \frac{n-t+T}{2}$ and $S = N - m$), we obtain that

$$\begin{aligned}
z &< \frac{\frac{n}{2} - m - 2t}{2\pi(1-p)(\frac{n}{2} - 2t - \frac{2}{p})} \exp\left(-\frac{2p(\frac{n}{2} - m - 2t - \frac{2}{p})^2}{(1-p)(n-t)} + o(1) \right) \\
&+ \frac{3}{\pi p(\frac{n}{2} - m - 2t - \frac{2}{p})} \exp\left(-\frac{9p(\frac{n}{2} - m - 2t - \frac{2}{p})^2}{8(1-p)(\frac{n}{2} - 2t - \frac{1}{p})} \right) \\
&< \frac{\sqrt{\log d}}{2\pi(1-p)\sqrt{d}} \exp\left(-\frac{\log d}{2} + 2\lambda \log\log\log d + o(1) \right) \\
&+ \frac{6\sqrt{d}}{\pi p n \sqrt{\log d}} \exp\left(-\frac{9 \log d}{16} + \frac{9\lambda \log\log\log d}{4} + o(1) \right).
\end{aligned}$$

The first term is much larger than the second, and so we obtain (for n large enough) the inequality

$$z < \frac{\sqrt{\log d}(\log\log d)^{2\lambda}}{\pi(1-p)d}. \tag{32}$$

We have that $a = \frac{z}{g(y)}$, and so from (31) and (32) (for n large enough),

$$a < \frac{e^6 \log d}{\pi(1-p)\sqrt{d}} < \frac{e^5 \log d}{(1-p)\sqrt{d}}. \tag{33}$$

Using the fact that $t = \left\lfloor n(\log\log d)^\lambda / \sqrt{d \log d} \right\rfloor$ and applying (33) and (30) we can now bound the expression in (29) (for n large enough) by

$$\begin{aligned}
\mathbb{P}(Q \text{ percolates}) &< \left(\exp\left(\frac{3pe^{10}(\log d)^2 n (\log\log d)^\lambda}{2(1-p)^2 d\sqrt{d\log d}} \right) \frac{e^4(\log\log d)^{2\lambda}}{\sqrt{d \log d}} \right)^t \\
&< \left(\frac{e^5(\log\log d)^{2\lambda}}{\sqrt{d\log d}} \right)^t
\end{aligned}$$

(where the second inequality follows since the exponent in the exponential is $o(1)$ when $p \leqslant 1 - \frac{\log^4 n}{n}$ by the calculations in (4)).

The expected number of sets of size t that percolates is (for n large enough) bounded by

$$\begin{aligned}
\binom{n-m}{t} \mathbb{P}(Q \text{ percolates}) &< \binom{n}{t} \left(\frac{e^5(\log\log d)^{2\lambda}}{\sqrt{d\log d}} \right)^t \\
&< \left(\frac{e^6 n(\log\log d)^{2\lambda}}{t\sqrt{d\log d}} \right)^t,
\end{aligned}$$

because $\binom{n}{t} \leqslant \left(\frac{en}{t}\right)^t$. We chose $t = \left\lfloor \frac{n(\log\log d)^\lambda}{\sqrt{d \log d}} \right\rfloor$, with $\lambda < 0$, and so the expected number of sets of size t that percolates is bounded above by

$$(e^6(\log\log d)^\lambda)^t = o(1).$$

Therefore, by the Markov inequality, it follows that with high probability, percolation does not occur for $\lambda < 0$. □

A Inequalities for Binomial Distributions

In this appendix we collect the inequalities for binomial distributions that we have applied in the proof of Theorem 1. We first present some well-known inequalities for binomial distributions and proceed by proving tight bounds for sums of binomial distributions (with different parameters) that are less standard.

We begin with some remarks on the log-concavity of the distribution function of the binomial distribution. We say that a sequence $a_k : k \in Z$ of positive numbers is log-concave if for all k we have that $p_k^2 \geqslant p_{k-1}p_{k+1}$. These results are standard, see for example [10].

Proposition 14. *The sum of independent Bernoulli random variables is log-concave, that is if X_i are independent Bernoulli random variables with means p_i, then for any k we have,*

$$\mathbb{P}(\sum_{i=1}^n X_i = k-1)\mathbb{P}(\sum_{i=1}^n X_i = k+1) \leqslant (\mathbb{P}(\sum_{i=1}^n X_i = k))^2.$$

Proposition 15. *The cumulative distribution of a discrete non-negative log-concave random variable X is log-concave, that is for all k,*

$$\mathbb{P}(X \leqslant k-1)\mathbb{P}(X \leqslant k+1) \leqslant (\mathbb{P}(X \leqslant k))^2.$$

When X is the sum of n independent Bernoulli random variables, we can rewrite $X = n - Y$, where Y is also the sum of n independent Bernoulli random variables, and so Proposition 15 is still true if we replace $\leqslant$, with $<$, $>$ or $\geqslant$.

Corollary 16. *The cumulative distribution of the sum or difference of independent binomial random variables is log-concave.*

Proof. Sums and differences of independent binomial random variables are also sums of independent Bernoulli random variables plus a constant, and so are log-concave. □

The remaining part of Appendix A is taken up with providing tight bounds, up to a constant factor, on binomial probabilities and their sums.

Proposition 17. *Suppose that $pn \geqslant 1$ and $k = pn + h < n$, where $h > 0$. Set*

$$\beta = \frac{1}{12k} + \frac{1}{12(n-k)},$$

then $\mathbb{P}(B(n,p) = k)$ is at least

$$\frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{h^2}{2p(1-p)n} - \frac{h^3}{2(1-p)^2n^2} - \frac{h^4}{3p^3n^3} - \frac{h}{2pn} - \beta\right).$$

Proof. This is Theorem 1.5 in [5], p. 12. ☐

Corollary 18. *Suppose that $p(1-p)n \to \infty$, and $k = pn + h$, where*

$$0 < h = o\left((p(1-p)n)^{\frac{2}{3}}\right).$$

Then

$$\mathbb{P}(B(n,p) = k) > \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{h^2}{2p(1-p)n} - o(1)\right).$$

Proof. For h in this range we have,

$$\frac{h^3}{2(1-p)^2n^2} + \frac{h^4}{3p^3n^3} + \frac{h}{2pn} = o(1).$$

We also have that $k \to \infty$, and $n - k \to \infty$, as $n \to \infty$, and so the inequality follows from Proposition 17. ☐

Proposition 19. *Suppose that $pn \geqslant 1$ and $k \geqslant pn + h$, where $h(1-p)n \geqslant 3$. Then,*

$$\mathbb{P}(B(n,p) = k) < \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{h^2}{2p(1-p)n} + \frac{h^3}{p^2n^2} + \frac{h}{(1-p)n}\right).$$

Proof. This is Theorem 1.2 of [5], p. 10. ☐

Corollary 20. *Suppose that $p(1-p)n \to \infty$, and $k \geqslant pn + h$, where*

$$1 < h = o((p(1-p)n)^{\frac{2}{3}}),$$

then

$$\mathbb{P}(B(n,p) = k) < \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{h^2}{2p(1-p)n} + o(1)\right).$$

Proof. For h in this range we have,

$$\frac{h^3}{p^2n^2} + \frac{h}{(1-p)n} = o(1),$$

and so the inequality follows from Proposition 19, which can be applied as $h(1-p)n \to \infty$, as $n \to \infty$. ☐

Proposition 21. *Suppose that $p(1-p)n \to \infty$, and $0 < h = o((p(1-p)n)^{\frac{2}{3}})$, then*

$$\mathbb{P}(B(n,p) \geqslant pn+h) < \frac{\sqrt{p(1-p)n}}{\sqrt{2\pi}h} \exp\left(-\frac{h^2}{2p(1-p)n} + o(1)\right).$$

Proof. This proof follows that of Theorem 1.3 in [5]. For $m \geqslant pn+h$, we have,

$$\frac{\mathbb{P}(B(n,p)=m+1)}{\mathbb{P}(B(n,p)=m)} \leqslant 1 - \frac{h+(1-p)}{(1-p)(pn+h+1)} = \lambda.$$

Hence,

$$\mathbb{P}(B(n,p) \geqslant pn+h) \leqslant \frac{1}{1-\lambda}\mathbb{P}(B(n,p) = \lceil pn+h \rceil).$$

As $(1-\lambda)^{-1} < \frac{p(1-p)n}{h}(1+\frac{h}{pn}) < \frac{p(1-p)n}{h}e^{\frac{h}{pm}}$, we get from Proposition 19 that

$$\mathbb{P}(B(n,p) \geqslant pn+h) < \frac{\sqrt{p(1-p)n}}{h\sqrt{2\pi}} \exp\left(-\frac{h^2}{2p(1-p)n} + \frac{h}{p(1-p)n} + \frac{h^3}{p^2n^2}\right)$$

the last two terms in the exponent being $o(1)$, for $h = o(p(1-p)n)^{\frac{2}{3}}$. □

Proposition 22. *Suppose that $p(1-p)n \to \infty$, and*

$$(p(1-p)n)^{\frac{1}{2}} < h = o((p(1-p)n)^{\frac{2}{3}}),$$

then

$$\mathbb{P}(B(n,p) \geqslant pn+h) > \frac{\sqrt{p(1-p)n}}{h\sqrt{2\pi}} \exp\left(-\frac{h^2}{2p(1-p)n} - \frac{3}{2} - o(1)\right).$$

Proof. Due to the unimodality of the binomial distribution, we have that the probability density function of the binomial distribution is decreasing away from its mean, and so,

$$\mathbb{P}(B(n,p) \geqslant pn+h) > \frac{p(1-p)n}{h}\mathbb{P}(B(n,p) = pn+h+\frac{p(1-p)n}{h}).$$

We can apply Corollary 18 as $h + \frac{p(1-p)n}{h} = o((p(1-p)n)^{\frac{2}{3}})$, and so it follows that

$$\mathbb{P}(B(n,p) \geqslant pn+h) > \frac{\sqrt{p(1-p)n}}{h\sqrt{2\pi}} \exp\left(-\frac{(h+\frac{p(1-p)n}{h})^2}{2p(1-p)n} - o(1)\right).$$

This is greater than the stated bound because

$$(h+\frac{p(1-p)n}{h})^2 \leqslant h^2 + 3p(1-p)n.$$ □

We shall also want a weaker but more general bound than Proposition 21 due to Bernstein in [4].

Lemma 23. *Let $X_1, \ldots, X_n$ be independent zero-mean random variables. Suppose that $|X_i| \leqslant M$, then for all positive t,*

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > t\right) \leqslant \exp\left(-\frac{t^2}{2\sum \mathbb{E}(X_j^2) + \frac{2}{3}Mt}\right).$$

Proof. For a proof see [7]. □

We have so far discussed well-known deviation inequalities for standard binomial distributions. We will now proceed to present some analogous results for sums of binomial distributions with different parameters p, that we have not been able to find in the litterature.

Proposition 24. *Suppose that $p(1-p)N \to \infty$, the inequality*

$$2(2p(1-p)N)^{\frac{1}{2}} < h = o((p(1-p)N)^{\frac{2}{3}})$$

holds and

$$hS = o(N((p(1-p)N)^{\frac{1}{2}})).$$

For the independent random variables; $X_1 = B(N-S, p)$, with mean μ_1 and variance σ_1^2,; and $X_2 = B(N+S, (1-p))$ with mean μ_2 and variance σ_2^2, we have,

$$\mathbb{P}(X_1 + X_2 \geqslant \mu_1 + \mu_2 + h) > \frac{\sqrt{2p(1-p)N}}{2\pi h} \exp\left(-\frac{h^2}{4p(1-p)N} - 4 - o(1)\right).$$

Proof. The conditions on S and h imply that $S = o(N)$. Set z and l equal to $\frac{2p(1-p)N}{h}$ and $\left\lfloor \frac{h}{\sqrt{2p(1-p)N}} \right\rfloor$ respectively. We can bound

$$\mathbb{P}(X_1 + X_2 \geqslant \mu_1 + \mu_2 + h)$$

from below by summing over the disjoint regions

$$\sum_{i=-l}^{l-1} \mathbb{P}\left(X_1 < \mu_1 + \frac{h}{2} - iz, X_2 < \mu_2 + \frac{h}{2} + (i+1)z, X_1 + X_2 \geqslant \mu_1 + \mu_2 + h\right). \tag{34}$$

These regions are disjoint as if $X_1 < \mu_1 + \frac{h}{2} - (i+1)z$ and $X_2 < \mu_2 + \frac{h}{2} + (i+1)z$, then $X_1 + X_2 < \mu_1 + \mu_2 + h$. For each i the region specified is an isosceles right angled triangle with axis-parallel legs of length z, and so there are at least $\lfloor z \rfloor (\lfloor z \rfloor - 1)/2$ pairs of integer values x_1, x_2, which X_1, X_2 can take while still satisfying all three relations in (34). We have that $h > 2lz$, and so if X_1, X_2 satisfy all three relations in (34), then $X_1 \geqslant \mu_1$ and $X_2 \geqslant \mu_2$. As we are only considering the region in which X_1, X_2 are larger than their means we can bound the sum in (34) from below by

$$\sum_{i=-l}^{l-1} \frac{\lfloor z \rfloor (\lfloor z \rfloor - 1)}{2} \mathbb{P}\left(X_1 = \left\lceil \mu_1 + \frac{h}{2} - iz \right\rceil\right) \mathbb{P}\left(X_2 = \left\lceil \mu_2 + \frac{h}{2} + (i+1)z \right\rceil\right). \tag{35}$$

We have that $p(1-p)(N-S) \to \infty$, as $N \to \infty$, and $h + lz = o(p(1-p)(N-S))^{\frac{2}{3}}$, and so we can apply Corollary 18 to get that the quantity in (35) is at least

$$\sum_{i=-1}^{l-1} \frac{\lfloor z \rfloor(\lfloor z \rfloor - 1)}{4\pi\sigma_1\sigma_2} \cdot$$
$$\exp\left(-\frac{(\frac{h}{2}-iz+1)^2(N+S)+(\frac{h}{2}+(i+1)z+1)^2(N-S)}{2p(1-p)(N^2-S^2)} - o(1)\right).$$

Expanding this out, and noticing $\lfloor z \rfloor = z(1+o(1))$ and

$$(N-S)(N+S) = N^2(1+o(1))$$

we get that the sum in (35) is at least

$$\sum_{i=-l}^{l-1} \frac{z^2}{4\pi p(1-p)N} \cdot$$
$$\exp\left(-\frac{h^2N+2hzN+4i^2z^2N+(4i+2)z^2N+o(p(1-p)N^2)}{4p(1-p)(N^2-S^2)} - o(1)\right), \tag{36}$$

where the approximations for $\lfloor z \rfloor$ and σ_1, σ_2 have been taken care of in the $o(1)$ in the exponential term. We have that $4i^2 + 4i + 2 \leqslant 6l^2$ and $l^2z^2 \leqslant 2p(1-p)N$, and so using the bounds in the statement of the proposition, the sum in (36) is at least

$$\sum_{i=-l}^{l-1} \frac{z^2}{4\pi p(1-p)N} \exp\left(-\frac{h^2N+16p(1-p)N^2}{4p(1-p)(N^2-S^2)} - o(1)\right)$$
$$\frac{lz^2}{2\pi p(1-p)N} \exp\left(-\frac{h^2N}{4p(1-p)(N^2-S^2)} - 4 - o(1)\right)$$
$$> \frac{\sqrt{2p(1-p)N}}{2\pi h} \exp\left(-\frac{h^2}{4p(1-p)N} - 4 - o(1)\right).$$

The last inequality following because $l > h/(2\sqrt{2p(1-p)N})$ and $hS = o(N(p(1-p)N)^{\frac{1}{2}})$. ☐

Proposition 25. *Suppose that $p(1-p)N \to \infty$. Furthermore assume that*

$$2(2p(1-p)N)^{\frac{1}{2}} < h = o((p(1-p)N)^{\frac{2}{3}}) \quad \textit{and} \quad Sh = o(N(p(1-p)N)^{\frac{1}{2}}).$$

Then, we have,

$$\mathbb{P}(X_1 + X_2 \geqslant \mu_1 + \mu_2 + h) < \frac{\sqrt{2p(1-p)N}}{h} \exp\left(-\frac{h^2}{4p(1-p)N} + 3 + o(1)\right),$$

for independent random variables $X_1 = B(N-S, p)$ with mean μ_1 and variance σ_1^2, and $X_2 = B(N+S, (1-p))$ with mean μ_2 and variance σ_2^2.

Proof. The conditions on S and h imply that $S = o(N)$. Set $z = \frac{2Np(1-p)}{h}$, and $l = \left\lfloor \frac{h^2}{4Np(1-p)} \right\rfloor$. We bound $\mathbb{P}(X_1 + X_2 \geqslant \mu_1 + \mu_2 + h)$ from below by covering the region where this inequality holds by

$$\mathbb{P}(X_1 + X_2 \geqslant \mu_1 + \mu_2 + h) < \tag{37}$$

$$\sum_{\substack{-l \leqslant i,j \leqslant l-1 \\ i+j \geqslant -1}} \left(\mathbb{P}\left(0 \leqslant X_1 - \mu_1 - \frac{h}{2} - iz < z\right) \mathbb{P}\left(0 \leqslant X_2 - \mu_2 - \frac{h}{2} - jz < z\right)\right) \tag{38}$$

$$+ \mathbb{P}\left(X_1 \geqslant \mu_1 + \frac{h}{2} + lz\right) \tag{39}$$

$$+ \mathbb{P}\left(X_2 \geqslant \mu_2 + \frac{h}{2} + lz\right). \tag{40}$$

We shall bound these three summands separately. Again because $h > 2lz$ we are only considering the range in which X_1 and X_2 are greater than their means. Firstly for each i, j pair there are at most $\lceil z \rceil^2$ points inside the specified region, and so the product inside the sum of (38) is at most

$$\lceil z \rceil^2 \mathbb{P}\left(X_1 = \left\lceil \mu_1 + \frac{h}{2} + iz \right\rceil\right) \mathbb{P}\left(X_2 = \left\lceil \mu_2 + \frac{h}{2} + jz \right\rceil\right).$$

We have that $p(1-p)(N \pm S) \to \infty$, as $N \to \infty$, and

$$1 < h \pm lz = o(p(1-p)(N \pm S))^{\frac{2}{3}},$$

and so we can apply Corollary 20 to get that the sum in (38) is at most

$$\sum_{\substack{-l \leqslant i,j \leqslant l-1 \\ i+j \geqslant -1}} \frac{\lceil z \rceil^2}{2\pi p(1-p)\sqrt{N^2 - S^2}}$$

$$\cdot \exp\left(-\frac{\left(\frac{h}{2} + iz\right)^2 (N+S) + \left(\frac{h}{2} + jz\right)^2 (N-S)}{2p(1-p)(N^2 - S^2)} + o(1)\right).$$

This is equal to

$$\frac{z^2}{2\pi p(1-p)N} \exp\left(-\frac{h^2 N}{4p(1-p)(N^2 - S^2)} + o(1)\right) \sum_{\substack{-l \leqslant i,j < l \\ i+j \geqslant -1}}$$

$$\exp\left(-\frac{h(i+j)zN + hzS(i-j) + z^2 N(i^2 + j^2) + z^2 S(i^2 - j^2)}{2p(1-p)(N^2 - S^2)}\right). \tag{41}$$

We can bound the above by noting that $|i - j| \leqslant \sqrt{2}\sqrt{i^2 + j^2}$ and $|i^2 - j^2| \leqslant i^2 + j^2$. As we also have that $Np(1-p)/2 < z^2 l \leqslant Np(1-p)$, the inner sum appearing in (41) is at most

$$\sum_{\substack{-l \leqslant i,j < l \\ i+j \geqslant -1}} \exp\left(-(i+j) + \sqrt{\frac{i^2 + j^2}{4l}} - \frac{i^2 + j^2}{4l} + o(1)\right). \tag{42}$$

A point (i,j) in the plane with integer coordinates and

$$\frac{i^2+j^2}{4l} - \sqrt{\frac{i^2+j^2}{4l}} < t,$$

also satisfies $|i-j| < \sqrt{21tl}$, as if $|i-j| \geqslant \sqrt{21tl}$, then $i^2+j^2 \geqslant \frac{21tl}{2}$, and so

$$\frac{i^2+j^2}{4l} - \sqrt{\frac{i^2+j^2}{4l}} \geqslant \left(\frac{21}{8} - \sqrt{\frac{21}{8}}\right) t.$$

Therefore the number of points (i,j) in the plane with integer coordinates and satisfying both $\frac{i^2+j^2}{4l} - \sqrt{\frac{i^2+j^2}{4l}} < t$, and $-1 \leqslant i+j < t$ is at most $2(t+1)\sqrt{21lt}$. This allows us crudely bound (42) by

$$2\sqrt{21l}\sum_{t=1}^{\infty}(t+1)\sqrt{t}\exp\left(-(t-1)\right).$$

The latter sum is less than $50\sqrt{l}$, and so the sum in (38) is bounded above by

$$\frac{50\sqrt{p(1-p)N}}{h\pi}\exp\left(-\frac{h^2}{4p(1-p)N} + o(1)\right). \tag{43}$$

Secondly we bound the probability in (39). As $l > \frac{h^2}{8Np(1-p)}$ we have that

$$\mathbb{P}\left(X_1 \geqslant \mu_1 + \frac{h}{2} + lz\right) < \mathbb{P}\left(X_1 \geqslant \mu_1 + \frac{3h}{4}\right).$$

By Proposition 21 we get that the quantity in (39) is at most

$$\begin{aligned} &\frac{4\sqrt{p(1-p)(N-S)}}{3h\sqrt{2\pi}}\exp\left(-\frac{9h^2}{32p(1-p)(N-S)} + o(1)\right) \\ &< \frac{2}{3\sqrt{\pi}}\frac{\sqrt{2p(1-p)N}}{h}\exp\left(-\frac{h^2}{4p(1-p)N} + o(1)\right). \end{aligned} \tag{44}$$

Similarly, the probability in (40) is at most

$$\frac{2}{3\sqrt{\pi}}\frac{\sqrt{2p(1-p)N}}{h}\exp\left(-\frac{h^2}{4p(1-p)N} + o(1)\right). \tag{45}$$

As $\frac{50}{\sqrt{2\pi}} + \frac{4}{3\sqrt{\pi}} < e^3$ we get that the sum of our three bounds, (43), (44), and (45) is at most the stated bound. $\square$

Proposition 26. *Suppose that* $p(1-p)N \to \infty$, *that*

$$\omega(N)(p(1-p)N)^{\frac{1}{2}} \leqslant pS = o((p(1-p)N)^{\frac{2}{3}})$$

and that $T = o(N)$, *then, for* N *large enough,*

$$\mathbb{P}(Z_1 = Z_2 + pT) < \frac{S}{2\pi(1-p)N} \exp\left(-\frac{2pS^2}{(1-p)(2N-T)} + o(1)\right) + \frac{3}{\pi pS} \exp\left(-\frac{9pS^2}{8(1-p)N}\right),$$

for independent random variables $Z_1 = B(N-S, p)$ *with mean* μ_1 *and variance* σ_1^2 *and* $Z_2 = B(N+S-T, p)$ *with mean* μ_2 *and variance* σ_2^2.

Proof. Let $\phi(i)$ be the probability that $Z_1 = Z_2 + pT = pN + i$, then

$$\phi(i) = \binom{N-S}{pN+i}\binom{N+S-T}{pN-pT+i} p^{p(2N-T)+2i}(1-p)^{(1-p)(2N-T)-2i}.$$

Denote the ratio between successive values of $\phi(i)$ by $\psi(i)$. We obtain that

$$\psi(i) = \frac{\phi(i+1)}{\phi(i)} = \frac{p^2((1-p)N - S - i)((1-p)(N-T) + S - i)}{(1-p)^2(pN+i+1)(p(N-T)+i+1)}.$$

Hence, we get that

$$\psi(i) = \frac{\left(1 - \frac{S+i}{(1-p)N}\right)\left(1 + \frac{S-i}{(1-p)(N-T)}\right)}{\left(1 + \frac{i+1}{pN}\right)\left(1 + \frac{i+1}{p(N-T)}\right)}, \tag{46}$$

and so ψ is a decreasing function of i. By noting that $e^{x-x^2} \leqslant (1+x) \leqslant e^x$, for $x \geqslant -\frac{1}{2}$, we can bound ψ for $i = o(p(1-p)N)$. We apply $e^{x-x^2} \leqslant (1+x)$ for the terms in the numerator of (46) and $(1+x) \leqslant e^x$ for the terms in the denominator of (46) to get the following lower bound of ψ for $i = o(p(1-p)N)$,

$$\exp\left(\frac{pST - (2N-T)(i+1-p)}{p(1-p)N(N-T)} - \left(\frac{S+i}{(1-p)N}\right)^2 - \left(\frac{S-i}{(1-p)(N-T)}\right)^2\right).$$

We apply $(1+x) \leqslant e^x$ for the terms in the numerator of (46) and $e^{x-x^2} \leqslant (1+x)$ for the terms in the denominator of (46) to get the following upper bound of ψ for $i = o(p(1-p)N)$,

$$\exp\left(\frac{pST - (2N-T)(i+1-p)}{p(1-p)N(N-T)} + \left(\frac{i+1}{pN}\right)^2 + \left(\frac{i+1}{p(N-T)}\right)^2\right).$$

Substituting in $i = \pm\frac{pS}{2}$, we get (for N large enough) that

$$\begin{aligned}
\psi(\frac{pS}{2}) &< \exp\left(-\left(\frac{(2N-3T)S}{2(1-p)N(N-T)}\right)(1+o(1))\right)\\
&< \exp\left(-\frac{(2N-3T)S}{4(1-p)N(N-T)}\right)\\
&< 1-\frac{S}{3(1-p)N} \qquad (47)
\end{aligned}$$

and

$$\begin{aligned}
\psi(-\frac{pS}{2}) &> \exp\left(\left(\frac{(2N+T)S}{2(1-p)N(N-T)}\right)(1+o(1))\right)\\
&> \exp\left(\frac{(2N+t)S}{4(1-p)N(N-T)}\right)\\
&> 1+\frac{S}{3(1-p)N}. \qquad (48)
\end{aligned}$$

Therefore ψ is greater than 1 at $i = pN - \frac{pS}{2}$ and less than 1 at $i = pN + \frac{pS}{2}$. Consequently (for N large enough), the maximum value of ϕ occurs between these two values.

We have that

$$\phi(i) = \mathbb{P}(Z_1 = \mu_1 + pS + i)\mathbb{P}(Z_2' = \mu_2' + pS - i),$$

where

$$Z_2' = N + S - T - Z_2 = B(N+S-T, (1-p)),$$

with mean μ_2' and variance $(\sigma_2')^2$. By Corollary 20 (applied with the parameter $h = pS+i$ and $h = pS - i$, respectively) we get that

$$\phi(i) < \frac{1}{2\pi\sigma_1\sigma_2'}\exp\left(-\frac{(pS+i)^2(N+S-T)+(pS-i)^2(N-S)}{2p(1-p)(N-S)(N+S-T)}+o(1)\right),$$

for $|i| \leqslant \frac{pS}{2}$. This is maximized when $i = \frac{pST-2pS^2}{2N-T}$ and as $\sigma_1\sigma_2' = N(1+o(1))$, there takes the value

$$\begin{aligned}
&\frac{1}{2\pi p(1-p)N}\exp\left(-\frac{pS^2((2N-T)^2-(T-2S)^2)}{2(1-p)(N-S)(N+S-T)(2N-T)}+o(1)\right)\\
&= \frac{1}{2\pi p(1-p)N}\exp\left(-\frac{2pS^2}{(1-p)(2N-T)}+o(1)\right).
\end{aligned}$$

We also obtain the bounds (for N large enough)

$$\begin{aligned}
\phi(\frac{pS}{2}) &< \frac{1}{2p(1-p)\pi N}\exp\left(-\frac{pS^2(10N+8S-9T)}{8(1-p)(N-S)(N+S-T)}+o(1)\right)\\
&< \frac{1}{2p(1-p)\pi N}\exp\left(-\frac{9pS^2}{8(1-p)N}\right)
\end{aligned}$$

and

$$\phi(\frac{-pS}{2}) < \frac{1}{2p(1-p)\pi N}\exp\left(-\frac{pS^2(10N-8S-T)}{8(1-p)(N-S)(N+S-T)}+o(1)\right)$$
$$< \frac{1}{2p(1-p)\pi N}\exp\left(-\frac{9pS^2}{8(1-p)N}\right).$$

Putting this all together and applying (47) and (48), we obtain (for N large enough) that

$$\mathbb{P}(Z_1 = Z_2 = pT) < pS\max_i \phi(i) + \frac{1}{1-\psi(\frac{pS}{2})}\phi(\frac{pS}{2}) + \frac{\psi(\frac{-pS}{2})}{\psi(\frac{-pS}{2})-1}\phi(\frac{-pS}{2})$$
$$< \frac{S}{2\pi(1-p)N}\exp\left(-\frac{2pS^2}{(1-p)(2N-T)}+o(1)\right)$$
$$+ \frac{3}{p\pi S}\exp\left(-\frac{9pS^2}{8(1-p)N}\right). \qquad \square$$

B Inequalities for Edge-Sets in $G(n,p)$

In this appendix we present the propositions about the number of edges in and between sets in $G(n,p)$ that we have applied in the proof of Theorem 1.

Proposition 27. *Suppose that $p(1-p)n \to \infty$. If n is large enough, then for all $t > \frac{n}{5}$, we have that with probability at least $1 - 4^{-t}$ every set in $G(n,p)$ of size t has at most $p\binom{t}{2} + 2t\sqrt{p(1-p)t}$ edges.*

Proof. The expected number of sets of size t with more than

$$p\binom{t}{2} + 2t\sqrt{p(1-p)t}$$

edges is

$$\binom{n}{t}\mathbb{P}\left(B\left(\binom{t}{2}, p\right) \geqslant p\binom{t}{2} + 2t\sqrt{p(1-p)t}\right).$$

By Lemma 23 and the fact that $\binom{n}{t} \leqslant (\frac{en}{t})^t$, this expectation is at most

$$(5e)^t \exp\left(-\frac{4p(1-p)t^3}{2p(1-p)\binom{t}{2} + \frac{4t\sqrt{p(1-p)t}}{3}}\right). \tag{49}$$

If $t > \frac{n}{5}$, then we have that $\sqrt{p(1-p)t} \to \infty$, as $n \to \infty$. Hence, if n is large enough, we have,

$$2p(1-p)\binom{t}{2} + \frac{4t\sqrt{p(1-p)t}}{3} \leqslant 1.001p(1-p)t^2.$$

Substituting this in (49) we have that the expected number of sets of size t with more than $p\binom{t}{2} + 2t\sqrt{p(1-p)t}$ edges is (for n large enough) at most,

$$\exp\left(t(\log 5 + 1) - \frac{4p(1-p)t^3}{1.001p(1-p)t^2}\right) < 4^{-t}. \qquad \square$$

Proposition 28. *Suppose that $p(1-p)n \to \infty$. If n is large enough, then for all t in the range $\frac{n}{5} < t \leqslant \frac{n}{2}$, we have that with probability at least $1 - 4^{-t}$ every set in $G(n,p)$ of size t has at least $pt(n-t) - 3t\sqrt{p(1-p)(n-t)}$ edges between it and its complement.*

Proof. The expected number of sets T of size t with less than $pt(n-t) - 3t\sqrt{p(1-p)(n-t)}$ edges between T and $[n] \setminus T$ is

$$\binom{n}{t}\mathbb{P}\left(B(t(n-t),(1-p)) \geqslant (1-p)t(n-t) + 3t\sqrt{p(1-p)(n-t)}\right).$$

By Lemma 23 and the fact that $\binom{n}{t} \leqslant (\frac{en}{t})^t$, this expectation is at most

$$(5e)^t \exp\left(-\frac{9p(1-p)t^2(n-t)}{2p(1-p)t(n-t) + 2t\sqrt{p(1-p)(n-t)}}\right). \tag{50}$$

As $\sqrt{(n-t)p(1-p)} \to \infty$, we have that if n is large enough, then for all t in the range $\frac{n}{5} < t \leqslant \frac{n}{2}$,

$$2p(1-p)t(n-t) + 2t\sqrt{p(1-p)(n-t)} \leqslant \frac{9}{4}p(1-p)t(n-t).$$

Substituting this in (50) we have that the expected number of sets T with a small number of edges between T and $[n] \setminus T$ is (for n large enough) less than

$$\exp\left(t(\log 5 + 1) - 4t\right) < 4^{-t}. \qquad \square$$

Proposition 29. *Suppose that $p(1-p)n \geqslant 4\log n$. If n is large enough, then for all $t \leqslant \frac{n}{5}$ we have that with probability at least $1 - n^{-\frac{t}{120}}$, for every set T in $G(n,p)$ of size t there are at least twice as many edges between T and $[n] \setminus T$ as there are in T.*

Proof. The expected number of sets T of size t, such that there are less than twice as many edges between T and $[n] \setminus T$ as there are in T is

$$\binom{n}{t}\mathbb{P}\left(B\left(t(n-t),p\right) < 2B\left(\binom{t}{2},p\right)\right),$$

for independent random variables $B\left(t(n-t),p\right)$ and $B\left(\binom{t}{2},p\right)$.

We can rewrite this as,

$$\binom{n}{t}\mathbb{P}\left(2B\left(\binom{t}{2},p\right) - pt(t-1) - B(t(n-t),p) + pt(n-t) > pt(n-2t+1)\right).$$

By Lemma 23, this is at most

$$\binom{n}{t}\exp\left(-\frac{(pt(n-2t+1))^2}{2p(1-p)t(n+t-2)+\frac{4pt(n-2t-1)}{3}}\right). \tag{51}$$

For $t<\frac{n}{24}$, using the inequality $\binom{n}{t}\leqslant n^t$ and the fact that $p(1-p)n\geqslant 4\log n$, we have that the quantity in (51) is (for n large enough) less than

$$n^t\exp\left(-\frac{pt(\frac{11n}{12})^2}{\frac{10n}{3}}\right)<n^t\exp\left(-\frac{4t\log n}{n}\cdot\frac{121n}{480}\right)=n^{-\frac{t}{120}}.$$

For $\frac{n}{24}\leqslant t\leqslant\frac{n}{5}$, using the inequality $\binom{n}{t}\leqslant\left(\frac{en}{t}\right)^t$ we have that the quantity in (51) is (for n large enough) less than,

$$\left(\frac{en}{t}\right)^t\exp\left(-\frac{pt(\frac{3n}{5})^2}{\frac{10n}{3}}\right)<\left(\frac{24e}{n^{\frac{2}{5}}}\right)^t<n^{-\frac{t}{120}}.\qquad\square$$

Corollary 30. *Suppose that $pn\geqslant\log n$. If n is large enough, then for all t satisfying $n^{\frac{24}{25}}\leqslant t\leqslant\frac{n}{5}$, we have that with probability at least $1-n^{\frac{-t}{120}}$, for every set T in $G(n,p)$ of size t, there are at least twice as many edges between T and $[n]\setminus T$ than there are in T.*

Proof. By the exact same reasoning as in Proposition 29 the expected number of sets T of size t with less than twice as many edges between T and $[n]\setminus T$ than there are in T is (for n large enough) at most

$$\left(\frac{en}{t}\right)^t\exp\left(-\frac{pt\left(\frac{3n}{5}\right)^2}{\frac{10n}{3}}\right)<\left(\frac{e}{n^{\frac{17}{250}}}\right)^t<n^{-\frac{t}{120}}.\qquad\square$$

Proposition 31. *For every fixed $\epsilon>0$ and $p\geqslant\frac{(1+\epsilon)\log n}{n}$, with high probability, the minimal degree of $G(n,p)$ is greater than 8.*

Proof. The expected number of vertices with degree at most 8 is bounded by

$$\begin{aligned}
n\mathbb{P}(B(n-1,p)\leqslant 8)&=n\sum_{i=0}^{8}\binom{n-1}{i}p^i(1-p)^{n-1-i}\\
&\leqslant n\left(\binom{n-1}{8}p^8(1-p)^{n-9}\left(1+\frac{9(1-p)}{p(n-9)}+\left(\frac{9(1-p)}{p(n-9)}\right)^2+\dots\right)\right)\\
&\leqslant\frac{9n^9}{8!}p^8(1-p)^{n-9}. \qquad (52)
\end{aligned}$$

These inequalities follow as $\max_{i\leqslant 8}\mathbb{P}(B(n-1,p)=i)$ occurs (for n large enough) when $i=8$, and so $\mathbb{P}(B(n-1,p)\leqslant 8)\leqslant 9\mathbb{P}(B(n-1,p)=8)$. The last line of (52) is maximised over $0 \leqslant p \leqslant 1$ when $\frac{p}{8} = \frac{1-p}{n-9}$, that is when $p = \frac{8}{n-1}$. So for p in our range, (52) is maximised when $p = \frac{(1+\epsilon)\log n}{n}$. Therefore (for n large enough)

$$\begin{aligned} n\mathbb{P}(B(n-1,p) \leqslant 8) &\leqslant \frac{9n^9(1+\epsilon)^8(\log n)^8}{8!n^8} e^{-\frac{(n-9)(1+\epsilon)\log n}{n}} \\ &\leqslant \frac{(\log n)^8}{n^{\frac{\epsilon}{2}}} = o(1); \end{aligned}$$

thus the lemma follows by the Markov inequality. □

Proposition 32. *Suppose that $(1+\epsilon)\log n \leqslant pn \leqslant 5\log n$. If n is large enough, then for all t satisfying $t \leqslant n^{\frac{29}{30}}$, we have that with probability at least $1 - n^{-\frac{t}{120}}$, every set in $G(n,p)$ of size t has at most $2t$ edges.*

Proof. The expected number of sets T in $G(n,p)$ of size t with at least $2t$ edges is

$$\binom{n}{t}\mathbb{P}\left(B\left(\binom{t}{2},p\right) \geqslant 2t\right) = \binom{n}{t}\sum_{i=2t}^{n}\binom{\binom{t}{2}}{i}p^i(1-p)^{\binom{t}{2}-i}. \tag{53}$$

By carefully bounding the summands in (53) for $i = 2t$ and $i = 2t+1$, we shall get a good bound on the total sum. We have that

$$\binom{\binom{t}{2}}{2t}p^{2t}(1-p)^{\binom{t}{2}-2t} < \left(\frac{ep(t-1)}{4}\right)^{2t} < \left(\frac{5et\log n}{4n}\right)^{2t}.$$

We also get that

$$\frac{\binom{\binom{t}{2}}{2t+1}p^{2t+1}(1-p)^{\binom{t}{2}-2t-1}}{\binom{\binom{t}{2}}{2t}p^{2t}(1-p)^{\binom{t}{2}-2t}} = \frac{p(\binom{t}{2}-2t)}{(1-p)(2t+1)} \leqslant pt < \frac{1}{2}.$$

Because the ratio between consecutive terms in the sum in (53) decreases as i increases, we have from above that the total sum is at most twice the first term, therefore

$$\begin{aligned} \binom{n}{t}\mathbb{P}\left(B\left(\binom{t}{2},p\right) \geqslant 2t\right) &\leqslant \binom{n}{t}2\left(\frac{5et\log n}{4n}\right)^{2t} \\ &\leqslant 2\frac{e^{3t}25^t(\log n)^{2t}t^t}{16^t n^t} \\ &\leqslant \left(\frac{C(\log n)^2}{n^{\frac{1}{30}}}\right)^t, \end{aligned}$$

and so the expected number of set T in $G(n,p)$ of size t with at least $2t$ edges is (for n large enough) at most $n^{-\frac{t}{120}}$. □

Acknowledgements

We would like to express our gratitude to B. Bollobás and R. Morris who introduced us to the problem.

References

[1] Michael Aizenman and Joel L Lebowitz, *Metastability effects in bootstrap percolation*, Journal of Physics A: Mathematical and General **21** (1988), no. 19, 3801.

[2] József Balogh, Béla Bollobás, and Robert Morris, *Majority bootstrap percolation on the hypercube*, Combinatorics, Probability and Computing **18** (2009), no. 1-2, 17–51.

[3] József Balogh and Boris G Pittel, *Bootstrap percolation on the random regular graph*, Random Structures & Algorithms **30** (2007), no. 1-2, 257–286.

[4] S Bernstein, *On a modification of Chebyshev's inequality and of the error formula of Laplace.*, Annal. Sci. Inst. Sav. Ukr. Sect. Math. I (1924), 38–49.

[5] Béla Bollobás, *Random graphs*, vol. 73, Cambridge university press, 2001.

[6] J Chalupa, PL Leath, and GR Reich, *Bootstrap percolation on a Bethe lattice*, Journal of Physics C: Solid State Physics **12** (1979), no. 1, L31.

[7] Cecil C Craig, *On the Tchebychef inequality of Bernstein*, The Annals of Mathematical Statistics **4** (1933), no. 2, 94–102.

[8] Svante Janson, Tomasz Łuczak, Tatyana Turova, and Thomas Vallier, *Bootstrap percolation on the random graph G(n,p)*, The Annals of Applied Probability **22** (2012), no. 5, 1989–2047.

[9] Tomas Juškevičius, *Probabilistic inequalities and bootstrap percolation*, PhD thesis, Memphis University (May 2015).

[10] Julian Keilson and Hans Gerber, *Some results for discrete unimodality*, Journal of the American Statistical Association **66** (1971), no. 334, 386–389.

[11] Nathan Kettle, *Vertex disjoint subgraphs and non-repetitive sequences*, PhD thesis, University of Cambridge (February 2014).

[12] Sigurdur Örn Stefánsson and Thomas Vallier, *Majority bootstrap percolation on the random graph G(n,p)*, arXiv:1503.07029v1 (2015).

[13] John Von Neumann, Arthur Walter Burks, et al., *Theory of self-reproducing automata*, (1966).