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

Cecilia Holmgren\(^2\), Tomas Juškevičius\(^†\), and Nathan Kettle\(^‡\)

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 more infected than non-infected neighbours are infected. We say that percolation occurs if eventually all vertices in \( G \) become infected.

In this paper we study majority bootstrap percolation on the Erdős-Rényi random graph \( G(n, p) \) above the connectivity threshold. Perhaps surprisingly, the results obtained for small \( p \) are comparable to the results for the hypercube obtained by Balogh, Bollobás and Morris \[2\].

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 \( p \). The probability that the entire graph eventually becomes infected is increasing with \( p \). It is therefore sensible to study the quantity \( p_c = \inf \{ p : \mathbb{P}_p(G \text{ infected}) \geq c \} \), in particular the critical probability \( p_{1/2} \) and the size of the critical window \( p_{1-\epsilon} - p_{\epsilon} \).

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 \( p_{1/2} = \Theta((\log n)^{1-d}) \).

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, Luczak, Turova and Vallier in \[8\].

In majority bootstrap percolation a vertex becomes infected if a majority of its neighbours are. 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

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

Date: August 12, 2015.

2010 Mathematics Subject Classification. 60C05; 05C80; 60K35.

Key words and phrases. bootstrap percolation, Erdős-Rényi random graph, threshold.
then, with high probability, percolation occurs (i.e., all vertices eventually become infected) if \( \lambda > \frac{1}{2} \) and does not occur if \( \lambda \leq -2 \).

In this paper we shall study majority bootstrap percolation on the Erdős-Rényi random graph \( G(n, p) \) above the connectivity threshold. We will see that for small \( p \) our results are in fact comparable to the results for the hypercube in [2], noting that the degree for each vertex in the \( n \)-dimensional hypercube (with \( 2^n \) vertices) is equal to \( n \).

2. Main Results

In this section we shall state our main results and discuss two different ways of selecting the initially infected set. The proofs of these theorems (in Section 3 and Section 4) use inequalities that are described separately in Section 5.

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 \( I_t = I_t \cup \{ v \in V(G) : |I_t \cap \Gamma(v)| \geq \frac{1}{2} |\Gamma(v)| \} \), 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 \( p = p(n) \to \infty \) 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 \( I_0 \) uniformly from \([n](m)\). 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 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 \( n \to \infty \). We set \( d = \frac{np}{1-p} \), 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

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

for some constant \( \lambda \). 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

\[
\mathbb{P}_m(G(n, p)) = \mathbb{P}(I = [n]).
\]
We shall now state the main result of this paper.

**Theorem 1.** Fix some number $\epsilon > 0$. Assume that for $n$ large enough,

$$(1 + \epsilon) \log n \leq p(1 - p)n.$$  

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

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

Our second result concerns a more natural setup, where each vertex is initially independently infected with probability $q$, we have that, with high probability, $|I_0| - qn \leq \omega(n) \sqrt{q(1-q)n}$. When $\sqrt{n} \ll \left( \frac{\log \log \log n}{\log n} \right)^2$, i.e., when $p \ll \left( \frac{\log \log \log n}{\log n} \right)^2$, 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

$$P_q'(G(n,p)) = P(I = [n]).$$

**Corollary 2.** Fix some number $\epsilon > 0$. Assume that for $n$ large enough,

$$(1 + \epsilon) \log n \leq p(1 - p)n.$$  

If $p \ll \left( \frac{\log \log \log n}{\log n} \right)^2$, 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

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

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

$$P_q'(G(n,p)) \to \Phi(2\theta),$$

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 $|I_0| - qn \leq \omega(n) \sqrt{q(1-q)n}$. If $p \ll \left( \frac{\log \log \log n}{\log n} \right)^2$, then $n \frac{\log \log \log d}{\sqrt{d \log d}} \gg \sqrt{n}$ and the result follows from Theorem 1.
If \( p \gg \frac{(\log \log \log n)^2}{n} \), then for each fixed \( \delta > 0 \) by the Central Limit Theorem we obtain
\[
\mathcal{P}_q'(G(n,p)) = \sum_{m=0}^{n} \mathbb{P}(B(n,q) = m) \mathcal{P}_m(G(n,p)) \\
\geq \mathbb{P}(B(n,q) \geq qn + (\delta - \theta)\sqrt{n}) \mathcal{P}_{qn+(\delta-\theta)\sqrt{n}}(G(n,p)) \\
= \mathbb{P}\left(B(n,q)/\sqrt{q(1-q)n} \geq (qn + (\delta - \theta)\sqrt{n})/\sqrt{q(1-q)n}\right) (1 + o(1)) \\
\to \Phi(2(\theta - \delta)),
\]
where the fourth line follows as \( \mathcal{P}_{qn+(\delta-\theta)\sqrt{n}}(G(n,p)) \to 1 \) for \( p \gg \frac{(\log \log \log n)^2}{\log n} \) by Theorem 1. A similar argument shows that
\[
1 - \mathcal{P}_q'(G(n,p)) \geq \Phi(-2(\theta + \epsilon))(1 + o(1)),
\]
and so
\[
\mathcal{P}_q'(G(n,p)) \to \Phi(2\theta).
\]

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 \( \frac{\log n}{\sqrt{n}} \), 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 3.** 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 \( I \). We shall call a proper subset \( S \subset [n] \) closed if for all \( v \in [n] \setminus S \) we have \( |\Gamma(v) \cap S| < \frac{|\Gamma(v)|}{2} \). Recall that \( I_0 \) is the set of initial 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 majority bootstrap process does not percolate, let \( T \) be such that the process has stabilized, i.e., \( I = IT = IT_{T+1} \neq [n] \). 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. 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 5 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. Recall that
\[ m = \frac{n}{2} - \frac{n}{2} \sqrt{\frac{\log d}{d}} + \lambda n \frac{\log \log \log d}{\sqrt{d \log d}} + o \left( n \frac{\log \log \log d}{\sqrt{d \log d}} \right). \]

We assume that for some fixed $\epsilon > 0$ it holds that for $n$ large enough $(1 + \epsilon) \log n \leq p(1 - p)n$. However, for some of the results below it is enough to assume that $p(1 - p)n = \omega(n)$.

**Lemma 4.** In the $MB(n, p; m)$ process,
\[ |I_1 \setminus I_0| \geq n \frac{(\log \log d)^{2\lambda}}{e^8 \sqrt{d \log d}}, \]
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 $X = |I_1 \setminus I_0| = \sum 1(A_i)$. 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 $\mu_1$ and $\mu_2$, respectively. We have that
\[ r = \mathbb{P}(\Gamma(i) \cap I_0) \geq \mathbb{P}(\Gamma(i) \cap ([n] \setminus I_0)) = \mathbb{P}(B(m, p) \geq B(n - m - 1, p)) = \mathbb{P}(B(m, p) + B(n - m - 1, 1 - p) \geq \mu_1 + \mu_2 + p(n - 2m - 1)). \]

For $p \gg \frac{1}{n}$, we have $p(n - 2m - 1) = \omega(n) \sqrt{p(1 - p)n}$ and $p(n - 2m - 1)^2 = o(n \sqrt{p(1 - p)n})$. Applying the bound from Proposition 21 to the last equality with $N = \frac{n - 1}{2}$, $S = \frac{n - 2m - 1}{2}$ and $h = p(n - 2m - 1)$, we obtain
\begin{align*}
  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) \\
  &> \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{align*}
where in the second line we have used the asymptotic relation
\[ d(n - 2m - 1)^2 = n^2 \log d - 4\lambda n^2 \log \log \log d + o(n^2). \]

Let us calculate the variance of $X$. Let
\[ r' = \mathbb{P}(A_j | A_i) - r, \]
this being the same for any $i \neq j$. We have
\[
\text{Var}(X) = \sum_{i,j \in [n]\setminus[m]} (\mathbb{P}(A_j|A_i) - \mathbb{P}(A_j))\mathbb{P}(A_i)
\]
\[
= (1-r)r(n-m) + r' r(n-m)(n-m-1),
\]  
(2)

where the first term in (2) 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. Note that
\[
\mathbb{P}(A_j|B_{ij}) = \mathbb{P}(\overline{B}(m,p) \geq \overline{B}(n-m-2,p) + 1)
\]  
(3)

and
\[
\mathbb{P}(A_j|\overline{B}_{ij}) = \mathbb{P}(B(m,p) \geq B(n-m-2,p)),
\]  
(4)

where $B(m,p)$ and $B(n-m-2,p)$ are independent random variables. Note that $\mathbb{P}(A_j|B_{ij}) \leq \mathbb{P}(A_j|\overline{B}_{ij})$, hence we may bound $r'$ by
\[
r' = \mathbb{P}(A_j|A_i) - \mathbb{P}(A_j) = \mathbb{P}(A_j|B_{ij})\mathbb{P}(B_{ij}|A_i) + \mathbb{P}(A_j|\overline{B}_{ij})\mathbb{P}(\overline{B}_{ij}|A_i) - \mathbb{P}(A_j)
\]
\[
\leq \mathbb{P}(A_j|\overline{B}_{ij}) - \mathbb{P}(A_j)
\]
\[
= p \mathbb{P}(A_j|\overline{B}_{ij})
\]
\[
= p\mathbb{P}(B(m,p) = B(n-m-2,p)) ,
\]

where the last equality follows from (3) and (4).

As $p\left(\frac{n}{2} - m - 1\right) = \omega(n)\sqrt{p(1-p)n}$ (which is true for $p \gg \frac{1}{n}$) we get from Proposition 23 applied with $(N, S, T) = (\frac{n}{2} - 1, \frac{n}{2} - m - 1, 0)$, that $r'$ is at most
\[
\frac{p\left(\frac{n}{2} - m - 1\right)}{2\pi(1-p)(\frac{n}{2} - 1)} \exp\left(-\frac{p\left(\frac{n}{2} - m - 1\right)^2}{(1-p)(\frac{n}{2} - 1)} + o(1)\right)
\]
\[
+ \frac{3}{\pi(\frac{n}{2} - m - 1)} \exp\left(-\frac{9p\left(\frac{n}{2} - m - 1\right)^2}{8(1-p)(\frac{n}{2} - 1)}\right)
\]
\[
< \frac{p\sqrt{\log d}}{2\pi(1-p)\sqrt{d}} \exp\left(-\frac{\log d}{2} + 2\lambda \log \log d + o(1)\right)
\]
\[
+ \frac{6\sqrt{d}}{\pi n \sqrt{\log d}} \exp\left(-\frac{9 \log d}{16} + \frac{9 \lambda \log \log d}{4} + o(1)\right).
\]

The second term is much smaller than the first term, and so (for $n$ large enough)
\[
r' < \left(\frac{\sqrt{\log d (\log \log d)^{2\lambda}}}{\pi n}\right).
\]  
(5)
We are now able to bound the probability that $X$ is small. From (2) and Chebyshev’s inequality we get
\[
\mathbb{P}\left(X \leq \frac{(n-m)r}{2}\right) \leq \mathbb{P}\left(|X - (n-m)r| \geq \frac{(n-m)r}{2}\right) \leq \frac{4\text{Var}(X)}{(n-m)r^2} = 4((1-r) + (n-m-1)r') \leq \frac{4r'}{r} + o(1).
\]

From (5) and (1) this is at most
\[
\left(\frac{2e^4 \log d \sqrt{d}}{n}\right)(1 + o(1)) + o(1) = o(1),
\]
and so we have, with high probability, that $|I_1 \setminus I_0|$ is at least $\frac{(n-m)r}{2}$. By using (1) we get that for large $n$,
\[
\frac{(n-m)r}{2} \geq \frac{n(\log \log d)^2}{e^8 \sqrt{d \log d}},
\]
which completes the proof. \(\square\)

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

**Lemma 5.** Suppose that for some fixed $\epsilon > 0$ we have
\[
p(1-p)n \geq (1+\epsilon) \log n,
\]
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 $|\Gamma(v) \cap ([n] \setminus S)| > |\Gamma(v) \setminus S|$. 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 $[n] \setminus S$.

If $\frac{n}{2} + \frac{7n}{2\sqrt{d}} < s < \frac{4n}{3}$, then $p(2s-n) \geq 7\sqrt{p(1-p)n}$, and so
\[
ps(n-s) - 3(n-s)\sqrt{p(1-p)s} > 2p\left(\frac{n-s}{2}\right) + 4(n-s)\sqrt{p(1-p)(n-s)}.
\]

By Proposition 24 every set of size $n-s$ has at most
\[
p\left(\frac{n-s}{2}\right) + 2(n-s)\sqrt{p(1-p)(n-s)}
\]
edges with probability at least $1 - \frac{1}{q}$, and by Proposition 25 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 $1 - \frac{1}{4n^2}$. Therefore, with high probability, every set $S$ of size

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

is not closed.

If $s \geq \frac{4n}{5} + 4\sqrt{n}$ and $p(1 - p)n \geq 4\log n$, then we know from Proposition 26 that with probability at least $1 - \frac{1}{4n^2}$ there does not exist a closed set of size $s$ in $G(n, p)$. The result follows as $\sum_{i \geq 1} n^{-\frac{1}{4n^2}} = o(1)$.

If $n - n^{\frac{5}{12}} \geq s \geq \frac{4n}{5} + 4\sqrt{n}$ and $5\log n \geq p(1 - p)n \geq (1 + \epsilon)\log n$, then we know from Corollary 27 that with probability at least $1 - \frac{1}{4n^2}$ there does not exist a closed set of size $s$ in $G(n, p)$.

If $s \geq n - n^{\frac{5}{12}}$ and $5\log n \geq p(1 - p)n \geq (1 + \epsilon)\log n$, then we know from Proposition 29 that with probability at least $1 - \frac{1}{4n^2}$ every set $S^C := [n] \setminus S$ of size $n - s$ has at most $2(n - s)$ edges, and so has a vertex $v_{SC}$ of degree at most 4. By Proposition 28 we have that, with high probability, the minimum degree of $G(n, p)$ is at least 9, and so $v_{SC}$ 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 $T$ of a suitable size and bound the probability that none of the vertices in $T$ are infected by $S$.

**Lemma 6.** Fix $\epsilon > 0$ and define

$$s = \left[ \frac{n}{2} - \frac{n\sqrt{\log d}}{2\sqrt{d}} + \frac{n(\log \log d)^{1+\epsilon}}{\sqrt{d\log d}} \right].$$

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

$$\mathbb{P}(S \text{ is closed}) \leq \exp \left( -\frac{n(\log d)^{\log(\log d) - 2}}{e^\epsilon \sqrt{d}} \right).$$

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

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

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

Therefore,

$$\mathbb{P}(S \text{ is closed}) \leq \sum_{E} \mathbb{P}(E) \prod_{v \in T} \mathbb{P}(F_v | E)$$

$$= \sum_{E} \mathbb{P}(E) \prod_{v \in T} \mathbb{P}(B(|S|, p) < B(n - |S| - t, p) + d_E(v)),$$
where $\mathbb{P}(E)$ is the probability of a particular edge set $E \subset \{0,1\}^{[t]}$ and is equal to $p^{\left|E\right|} (1-p)^{\left|\{1\}^{[t]} - E\right|}$.

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

$$\mathbb{P}(S \text{ is closed}) \leq \sum_{E} \mathbb{P}(E) \prod_{v \in T} f(d_{E}(v)). \quad (6)$$

The rest of the proof shall be spent bounding (6). The degree of vertices in $T$ is heavily concentrated around $pt$, and we shall expand $f$ around $pt$ to show that (6) is not much larger than $f(pt)^t$.

We have by Corollary 13 that $f$ is log-concave, and so for any $x$ and $y$ with $f(y) \neq 0$,

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

Setting $y = \lfloor pt \rfloor \in \mathbb{N}$ we get

$$\mathbb{P}(S \text{ is closed}) \leq \sum_{E} \mathbb{P}(E) \prod_{v \in T} f(y) \left( \frac{f(y + 1)}{f(y)} \right)^{d_{E}(v) - y}$$
$$= \sum_{E} \mathbb{P}(E) f(y)^t \left( \frac{f(y + 1)}{f(y)} \right)^{2\left|E\right| - ty}.$$

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

$$\mathbb{P}(S \text{ is closed}) \leq \sum_{i=0}^{\left|\{1\}^{[t]}\right|} \left( \frac{\left|\{1\}^{[t]}\right|}{i} \right) p^i (1-p)^{\left|\{1\}^{[t]}\right|-i} f(y)^t \left( \frac{f(y + 1)}{f(y)} \right)^{2\left|\{1\}^{[t]}\right| - ty}$$
$$= \left( 1 - p + p \left( \frac{f(y + 1)}{f(y)} \right)^2 \right)^{\left|\{1\}^{[t]}\right|} \left( \frac{f(y)}{f(y + 1)} \right)^{ty} f(y)^t. \quad (7)$$

Setting $\frac{f(y + 1)}{f(y)} = 1 + a$, we bound (7) using the inequalities $1 + w \leq e^w$ and $(1 + x)^{-1} \leq 1 - x + x^2$ for $x \geq 0$ to get

$$\mathbb{P}(S \text{ is closed}) \leq \left( 1 + 2ap + a^2p \right)^{\frac{2}{t}} \left( \frac{1}{1+a} \right)^{pt^2} f(y)^t$$
$$\leq \exp \left( 2ap + a^2p \frac{t^2}{2} + (a^2 - a)pt^2 \right) f(y)^t$$
$$= \exp \left( \frac{3pa^2t^2}{2} \right) f(y)^t. \quad (8)$$

We have that

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

Let us write $z = \mathbb{P} (B(s, p) = B(n - s - t, p) + y)$ to ease up the notation. Thus, $f(y + 1) = f(y) + z$. By Proposition 23 applied with $N = \frac{a + 1 + t}{2}$.
\[ S = \frac{n-2s-t+T}{2} \text{ and } T = \left\lfloor \frac{p|t|}{p} \right\rfloor \text{ and noting that } 0 \leq T - t < p^{-1}, \text{ we have} \]

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

The second term in (9) is much smaller than the first so as \( 6 < 2\pi \) and \( t \log d = o(n) \) we get (for \( n \) large enough)

\[
z < \frac{\sqrt{\log d}(\log d)^{2(\log \log d)^{\epsilon}}}{6(1-p)d}. \]

We can rewrite \( f(y) \) as

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

We have for \( p \gg \frac{1}{n} \) the asymptotic relation

\[
(p(n-2s) + 1)(t + 2s - n) = o(n\sqrt{np(1-p)}),
\]

and so using Proposition 21 with \((N, S, h) = (\frac{n-t}{2}, \frac{n-2s-t}{2}, p(n-2s)+y-pt)\) we obtain (for \( n \) large enough) that

\[
f(y) < 1 - \frac{\sqrt{p(1-p)(n-t)}}{2\pi p(n-2s)+1} \exp \left( \frac{(p(n-2s)+1)^2}{2p(1-p)(n-t)} - 4 - o(1) \right) < 1 - \frac{(\log d)^{2(\log \log d)^\epsilon}}{e^6 \sqrt{d \log d}} < \exp \left( -\frac{(\log d)^{(\log \log d)^\epsilon}}{e^6 \sqrt{d}} \right), \tag{10} \]

the second inequality follows from the same reasoning used in (9) and that \( e^6 > 2\pi e^4 \).

We can also apply Proposition 22 to get a lower bound on \( f(y) \) (for \( n \) large enough) of

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

here the bound on \( 1 - f(y) \) is actually \( o(1) \), being within a constant factor of the bound in (10).

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 (8) we get (for $n$ large enough)

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

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

$$P(S \text{ is closed}) < \exp \left( - \frac{n(\log d)^{2(\log \log d)^\epsilon}}{e^7 \sqrt{d}} \right),$$

as $t > \frac{2n}{e^{(\log d)^\epsilon}}$. \qed

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

**Proposition 7.** Assume that

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

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

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

that contain $I_0$ is $o(1)$.

**Proof.** Let $S$ be a set of size $s$ in our range, $s$ can have at most $\left\lfloor \frac{n \sqrt{\log d}}{\sqrt{d}} \right\rfloor$ different values. For each possible value of $s$ and $n$ large enough (using Stirling’s formula) there are at most

$$\binom{n-m}{s-m} \leq \left( \frac{n}{s-m} \right)^{\frac{n \sqrt{\log d}}{\sqrt{d}}} \leq \left( \frac{e \sqrt{d}}{\sqrt{\log d}} \right)^{\frac{n \sqrt{\log d}}{\sqrt{d}}} \leq \exp \left( \frac{n(\log d)^{\frac{3}{2}}}{e^3 \sqrt{d}} \right)$$

possible closed sets that can contain $I_0$. By Lemma 6 the expected number of closed sets is (for $n$ large enough) less than

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

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

**Corollary 8.** Fix some number $\epsilon > 0$. Assume that for $n$ large enough,

$$(1 + \epsilon) \log n \leq p(1-p)n.$$

If the initially infected set $I_0$ has size

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

then for $\lambda > 0$, with high probability, the $MB(n,p;m)$ process percolates.
Proof. We have from Lemma 4 that, with high probability, \( I_0 = \lfloor m \rfloor \) is contained in no 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 7 applied to \( \epsilon = \lambda - \frac{1}{2} \) that, with high probability, \( I_0 \) is contained in no 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 5 that, with high probability, \( I_0 \) is contained in no 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 9.** Fix some number \( \epsilon > 0 \). Assume that for \( n \) large enough,
\[
(1 + \epsilon) \log n \leq p(1 - p)n.
\]
If the initially infected set \( I_0 \) has size
\[
m = \frac{n}{2} - \frac{n \sqrt{\log d}}{2 \sqrt{d}} + \lambda n \frac{\log \log d}{\sqrt{d \log d}} + o \left( \frac{n \log \log d}{\sqrt{d \log d}} \right),
\]
then, for \( \lambda < 0 \), with high probability, the \( MB(n,p;m) \) process does not percolate.

**Remark 10.** Note that Lemma 9 and Corollary 8 prove Theorem 1.

In fact to prove Lemma 9, 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.** We say that a set of vertices \( T \) percolates if all of its vertices will be infected eventually. For \( T \subset I \setminus I_0 \) we can order the vertices of \( I_0 \cup T \) by the time they get infected. That is, take any order of \( T \) such that a vertex from \( I_j \) is infected before any vertex from \( I_{j'} \) if \( j < j' \). Notice that for each \( v \in T \) the majority of its neighbours (in the whole graph) are in the set of its predecessors in this order. Our strategy will be to show that if \( \lambda < 0 \) then, with high probability, there is no percolating set \( T \) of a particular size and thus the \( MB(n,p;m) \) process does not percolate.

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

$$\mathbb{P}(T \text{ percolates}) \leq \sum_E \mathbb{P}(E) \prod_{i=1}^{t} \mathbb{P}(B(m, p) + d_E(i) \geq B(n - m - t, p)) \quad (11)$$

(for independent random variables $B(m, p)$ and $B(n - m - t, p)$). Denote $g(x) = \mathbb{P}(B(m, p) + x \geq B(n - m - t, p))$. Due to the log-concavity of $g$ (Corollary 13) we have for integers $x, y$, that

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

Using the latter inequality with $x = d_E(i)$ and $y = \lfloor pt \rfloor$, we can bound (11) by

$$\sum_E \mathbb{P}(E) \prod_{i=1}^{t} g(y) \left( \frac{g(y + 1)}{g(y)} \right)^{d_E(i) - y} \leq \sum_E \mathbb{P}(E) g(y)^t \left( \frac{g(y + 1)}{g(y)} \right)^{2 |E| - ty}$$

$$= \sum_{j=0}^{\lfloor \frac{t}{2} \rfloor} \binom{\frac{t}{2}}{j} p^j (1 - p)^{\lfloor \frac{t}{2} \rfloor - j} g(y)^t \left( \frac{g(y + 1)}{g(y)} \right)^{2j - ty}$$

$$= \left( 1 - p + p \left( \frac{g(y + 1)}{g(y)} \right)^2 \right)^{\binom{\frac{t}{2}}{j}} \left( \frac{g(y)}{g(y + 1)} \right)^{pt^2} g(y)^t.$$ 

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

$$\left( 1 - p + p(1 + a)^2 \right)^{\binom{\frac{t}{2}}{j}} (1 - a + a^2)^{ty} g(y)^t$$

$$\leq \exp \left( (2ap + a^2p) \frac{t^2}{2} + (a^2 - a)pt^2 \right) g(y)^t$$

$$= \left( \exp \left( \frac{3pa^2t}{2} \right) \right) g(y)^t. \quad (12)$$

We have by definition that $g(y)$ is equal to

$$g(y) = \mathbb{P}(X_1 + X_2 \geq \mu_1 + \mu_2 + pm - 2pm - pt - \lfloor pt \rfloor),$$

where $X_1 = B(m, p)$ with mean $\mu_1$ and $X_2 = B(n - m - t, (1 - p))$ with mean $\mu_2$. Setting $t = \lceil n(\log \log d)^{\lambda} / \sqrt{d \log d} \rceil$ and using Proposition 22 with $N = \frac{a^2t}{2}$, $S = \frac{a^2 - a}{2}$ and $h = p(n - 2m - t) - y$ to bound $g(y)$, we obtain
(for \( n \) large enough)
\[
g(y) < \frac{\sqrt{p(1-p)(n-t)}}{pn-2pm-2pt-1} \exp \left( -\frac{(pm-2pm-2pt-1)^2}{2p(1-p)(n-t)} + 3 + o(1) \right)
\]
\[
< \frac{e^3}{\sqrt{\log d}} \exp \left( -\frac{\log d}{2} + 2\lambda \log \log d + O((\log \log d)^\lambda) \right)
\]
\[
< \left( \frac{e^4(\log \log d)^{2\lambda}}{\sqrt{d \log d}} \right),
\]
when \( \lambda < 0 \).

We can also bound \( g(y) \) from below by Proposition 21
\[
g(y) > \frac{\sqrt{p(1-p)(n-t)}}{2\pi(pm-2pm-2pt)} \exp \left( -\frac{(pm-2pm-2pt)^2}{2p(1-p)(n-t)} - 4 - o(1) \right)
\]
\[
> \frac{1}{2\pi e^4/\log d} \exp \left( -\frac{\log d}{2} + 2\lambda \log \log d + O((\log \log d)^\lambda) \right)
\]
\[
> \left( \frac{\log \log d)^{2\lambda}}{e^6/\sqrt{d \log d}} \right),
\]
when \( \lambda < 0 \) (and \( n \) is large enough).

By definition of \( g \) we have that
\[
g(y+1) = g(y) + P(B(m,p) + y + 1 = B(n-m-t,p)).
\]

Let us write \( z = P(B(m,p) + y + 1 = B(n-m-t,p)) \) for convenience.

We shall now obtain an upper bound for \( z \). Using Proposition 23 with \( T = -\frac{y+1}{p}, N = \frac{n-t+T}{2} \) and \( S = N - m \), we obtain
\[
z < \frac{n^2 - m - 2t}{2\pi(1-p)(\frac{n^2}{2} - 2t - \frac{2}{p})} \exp \left( -\frac{2p(\frac{n^2}{2} - m - 2t - \frac{2}{p})}{(1-p)(n-t)} + o(1) \right)
\]
\[
+ \frac{3}{\pi p(\frac{n^2}{2} - m - 2t - \frac{2}{p})} \exp \left( -\frac{9p(\frac{n^2}{2} - m - 2t - \frac{2}{p})}{8(1-p)(\frac{n^2}{2} - 2t - \frac{1}{p})} \right)
\]
\[
< \frac{\sqrt{\log d}}{2\pi(1-p)\sqrt{d}} \exp \left( -\frac{\log d}{2} + 2\lambda \log \log d + o(1) \right)
\]
\[
+ \frac{6\sqrt{d}}{\pi pm \sqrt{d}} \exp \left( -\frac{9 \log d}{16} + \frac{9\lambda \log \log d}{4} + o(1) \right).
\]

The first term is much larger than the second, and so we obtain (for \( n \) large enough) the inequality
\[
z < \frac{\sqrt{\log d(\log \log d)^{2\lambda}}}{\pi(1-p)d}.
\]

We have that \( a = \frac{z}{g(y)} \), and so from (13) and (14) (for \( n \) large enough)
\[
a < \frac{e^8 \log d}{\pi(1-p)\sqrt{d}} < \frac{e^5 \log d}{(1-p)\sqrt{d}}.
\]
The expected number of sets of size $t$ that percolates is (for $n$ large enough)

$$\binom{n-m}{t} \mathbb{P}(T \text{ percolates}) < \binom{n}{t} \left( \frac{e^5(\log \log d)^{2\lambda}}{t\sqrt{d\log d}} \right)^t$$

because $\binom{n}{t} \leq \left( \frac{en}{t} \right)^t$. We chose $t = \frac{n(\log \log d)\lambda}{\sqrt{d\log d}}$, 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, with high probability, percolation does not occur for $\lambda < 0$. $\Box$

5. Inequalities

We begin this section with some remarks on the log-concavity of the distribution function of the Binomial distribution. These results are standard, see for example [10], but we prove them for completeness.

**Proposition 11.** 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,

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

**Proof.** We proceed by induction on $n$, with the base case $n = 1$ being trivial as one of the terms on the left hand side of the inequality is zero. Otherwise conditioning on $X_{n+1}$, and writing $f_{n,k} = \mathbb{P}(\sum_{i=1}^{n} X_i = k)$ we get,

$$f_{n+1,k-1}f_{n+1,k+1} = (p_{n+1}f_{n,k-2} + (1-p_{n+1})f_{n,k-1})(p_{n+1}f_{n,k} + (1-p_{n+1})f_{n,k+1})$$

$$\leq (p_{n+1}f_{n,k-1} + (1-p_{n+1})f_{n,k})^2$$

$$= (f_{n+1,k})^2$$

The inequality follows as $f_{n,k-2}f_{n,k+1} \leq f_{n,k-1}f_{n,k}$ is implied by the induction hypothesis. $\Box$

**Proposition 12.** The cumulative distribution of a discrete non-negative log-concave random variable $X$ is log-concave, that is for all $k$,

$$\mathbb{P}(X \leq k - 1) \mathbb{P}(X \leq k + 1) \leq (\mathbb{P}(X \leq k))^2.$$
Proof. Setting \( r_i = \mathbb{P}(X = i) \) we get by Proposition 11,
\[
(r_0 + \ldots + r_{k-1})r_{k+1} \leq (r_1 + \ldots + r_k)r_k + r_k r_0,
\]
and so,
\[
(r_0 + \ldots + r_{k-1})(r_0 + \ldots + r_{k+1}) \leq (r_0 + \ldots + r_k)^2.
\]
\[\square\]

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 12 is still true if we replace \( \leq \), with \(<\), \(>\), or \(\geq\).

**Corollary 13.** 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. \[\square\]

A substantial part of this section is now taken up with providing tight bounds, up to a constant factor, on binomial probabilities and their sums.

**Proposition 14.** Suppose \( pn \geq 1 \) 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. \[\square\]

**Corollary 15.** Suppose \( p(1-p)n = \omega(n) \) and \( k = pn + h \), where \( 0 < h = o\left(\left(p(1-p)n\right)^{\frac{3}{2}}\right) \), 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 \( k = \omega(n) \) and \( n - k = \omega(n) \), and so the inequality follows from Proposition 14. \[\square\]

**Proposition 16.** Suppose \( pn \geq 1 \) and \( k \geq pn + h \), where \( h(1-p)n \geq 3 \). 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. \[\square\]
Corollary 17. Suppose \( p(1-p)n = \omega(n) \) and \( k \geq pn + h \), where
\[
1 < h = o((p(1-p)n)^{\frac{3}{2}}),
\]
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 [16] which can be applied as \( h(1-p)n = \omega(n) \).

Proposition 18. Suppose \( p(1-p)n = \omega(n) \) and \( 0 < h = o((p(1-p)n)^{\frac{3}{2}}) \), then
\[
\mathbb{P}(B(n, p) \geq pn + h) < \frac{\sqrt{p(1-p)n}}{\sqrt{2\pi h}} \exp \left( - \frac{h^2}{2p(1-p)n} + o(1) \right).
\]

Proof. This proof follows that of Theorem 1.3 in [5]. For \( m \geq pn + h \), we have
\[
\mathbb{P}(B(n, p) = m + 1) \leq \frac{m + (1-p)n}{m} = \lambda.
\]

Hence,
\[
\mathbb{P}(B(n, p) \geq pn + h) \leq \frac{1}{1-\lambda} \mathbb{P}(B(n, p) = \lceil pn + h \rceil).
\]

As \( (1-\lambda)^{-1} < \frac{\sqrt{p(1-p)n}}{h} \left( 1 + \frac{h}{pn} \right) < \frac{\sqrt{(1-p)n}}{h} \frac{h}{n} \), we get from Proposition [16] that
\[
\mathbb{P}(B(n, p) \geq pn + h) < \frac{\sqrt{p(1-p)n}}{h\sqrt{2\pi}} \exp \left( - \frac{h^2}{2p(1-p)n} + \frac{h}{p(1-p)n} + \frac{h^3}{p^2n^2} \right),
\]
the last two terms in the exponent being \( o(1) \), for \( h = o(p(1-p)n)^{\frac{3}{2}} \).

Proposition 19. Suppose \( p(1-p)n = \omega(n) \) and
\[
(p(1-p)n)^{\frac{3}{2}} < h = o((p(1-p)n)^{\frac{3}{2}}),
\]
then
\[
\mathbb{P}(B(n, p) \geq pn + h) > \frac{\sqrt{p(1-p)n}}{h\sqrt{2\pi}} \exp \left( - \frac{h^2}{2p(1-p)n} - \frac{3}{2} - o(1) \right).
\]

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

We can apply Corollary [13] as \( h + \frac{p(1-p)n}{h} = o((p(1-p)n)^{\frac{3}{2}}) \), and so it follows that
\[
\mathbb{P}(B(n, p) \geq pn + h) > \frac{\sqrt{p(1-p)n}}{h\sqrt{2\pi}} \exp \left( - \left( \frac{h + \frac{p(1-p)n}{h}}{2p(1-p)n} - o(1) \right) \right).
\]
This is greater than the stated bound because
\[
(h + \frac{p(1-p)n}{h})^2 \leq h^2 + 3p(1-p)n.
\]
\[\square\]

We shall also want a weaker but more general bound than Proposition \[18\] due to Bernstein in \[4\].

**Lemma 20.** Let \(X_1, \ldots, X_n\) be independent zero-mean random variables. Suppose that \(|X_i| \leq M\), then for all positive \(t\),
\[
\Pr\left(\sum_{i=1}^{n} X_i > t\right) \leq \exp\left(\frac{-t^2}{2\sum E(X_i^2) + \frac{2}{3}Mt}\right).
\]

**Proof.** For a proof see \[7\]. \[\square\]

We have in this section, 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 21.** Suppose that \(p(1-p)N = \omega(N)\), the inequality
\[
2(2p(1-p)N)^{\frac{1}{2}} < h = o((p(1-p)N)^{\frac{1}{2}})
\]
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 \(\mu_1\) and variance \(\sigma_1^2\); and \(X_2 = B(N+S, (1-p))\) with mean \(\mu_2\) and variance \(\sigma_2^2\), we have
\[
\Pr(X_1 + X_2 \geq \mu_1 + \mu_2 + h) > \frac{\sqrt{2p(1-p)N}}{2\pi h} \exp\left(\frac{-h^2}{4p(1-p)N} - 4 - o(1)\right).
\]

**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 \(\frac{h}{\sqrt{2p(1-p)N}}\) respectively. We can bound
\[
\Pr(X_1 + X_2 \geq \mu_1 + \mu_2 + h)
\]
from below by summing over the disjoint regions
\[
\sum_{i=-l}^{l-1} \Pr\left(X_1 < \mu_1 + \frac{h}{2} - iz, X_2 < \mu_2 + \frac{h}{2} + (i+1)z, X_1 + X_2 \geq \mu_1 + \mu_2 + h\right).
\]
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 \(|z|(|z| - 1)/2\) pairs of integer values \(x_1, x_2\), which \(X_1, X_2\) can take while still satisfying all three relations in \[13\]. We have that \(h > 2lz\), and so if \(X_1, X_2\) satisfy all three relations in \[13\], then \(X_1 \geq \mu_1\) and \(X_2 \geq \mu_2\).
As we are only considering the region in which \(X_1, X_2\) are larger than their means we can bound the sum in (15) from below by
\[
\sum_{i=-l}^{l-1} \frac{|z|(|z|-1)}{2} \mathbb{P} \left( X_1 = \left\lfloor \mu_1 + \frac{h}{2} - iz \right\rfloor \right) \mathbb{P} \left( X_2 = \left\lfloor \mu_2 + \frac{h}{2} + (i+1)z \right\rfloor \right).
\] (16)

We have that \(p(1-p)(N-S) = \omega(N)\) and \(h+iz = o(p(1-p)(N-S))^\frac{1}{2}\), and so we can apply Corollary 15 to get that the quantity in (16) is at least
\[
\sum_{i=-l}^{l-1} \frac{|z|(|z|-1)}{4\pi \sigma_1 \sigma_2} \exp \left( -\frac{\left(4i^2 + 2\right)(N+S)+(i+1)^2(N-S) - 2p(1-p)(N^2 - S^2)}{4p(1-p)(N^2 - S^2)} \right) - o(1).
\] (17)

Expanding this out, and noticing \(|z| = z(1 + o(1))\) and
\[(N-S)(N+S) = N^2(1 + o(1))\]
we get that the sum in (16) is at least
\[
\sum_{i=-l}^{l-1} \frac{z^2}{4\pi p(1-p)N} \exp \left( -\frac{\left(h^2 N + 2hzN + 4i^2z^2N + 4(1-p)N \right)}{4p(1-p)(N^2 - S^2)} \right) - o(1),
\] where the approximations for \(|z|\) and \(\sigma_1, \sigma_2\) have been taken care of in the \(o(1)\) in the exponential term. We have that \(4i^2 + 2i + 2 \leq 6i^2\) and \(i^2z^2 \leq 2p(1-p)N\), and so using the bounds in the statement of the proposition, the sum in (17) is at least
\[
\sum_{i=-l}^{l-1} \frac{z^2}{4\pi p(1-p)N} \exp \left( -\frac{\left(h^2 N + 16(1-p)N^2 \right)}{4p(1-p)(N^2 - S^2)} \right) - o(1)
\]
\[
\frac{lz^2}{2\pi p(1-p)N} \exp \left( -\frac{h^2 N}{4p(1-p)(N^2 - S^2)} - 4 - o(1) \right)
\]
\[
> \frac{\sqrt{2p(1-p)N}}{2\pi h} \exp \left( -\frac{h^2}{4p(1-p)N} - 4 - o(1) \right).
\]

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

**Proposition 22.** Suppose that \(p(1-p)N = \omega(N)\). Furthermore assume that
\[2(2p(1-p)N)^{\frac{1}{4}} < h = o((p(1-p)N)^{\frac{1}{2}})\] and \(Sh = o(N(p(1-p)N)^{\frac{1}{2}})\).

Then we have
\[
\mathbb{P}(X_1 + X_2 \geq \mu_1 + \mu_2 + h) < \frac{\sqrt{2p(1-p)N}}{h} \exp \left( -\frac{h^2}{4p(1-p)N} + 3 + o(1) \right),
\]
for independent random variables \(X_1 = B(N-S,p)\) with mean \(\mu_1\) and variance \(\sigma_1^2\), and \(X_2 = B(N+S,(1-p))\) with mean \(\mu_2\) and variance \(\sigma_2^2\).
Proof. The conditions on $S$ and $h$ imply that $S = o(N)$. Set $z = \frac{2Np(1-p)}{h}$, and $l = \left\lceil \frac{h^2}{4Np(1-p)} \right\rceil$. We bound $\mathbb{P}(X_1 + X_2 \geq \mu_1 + \mu_2 + h)$ from below by covering the region where this inequality holds by

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

$$\sum_{i+j \geq -1}^{l \leq i, j < -l} \left( \mathbb{P} \left( 0 \leq X_1 - \mu_1 - \frac{h}{2} - iz < z \right) \mathbb{P} \left( 0 \leq X_2 - \mu_2 - \frac{h}{2} - jz < z \right) \right)$$

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

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

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 $[z]^2$ points inside the specified region, and so the product inside the sum of (19) is at most

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

We have that $p(1-p)(N \pm S) = o(N)$ and

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

and so we can apply Corollary 17 to get that the sum in (19) is at most

$$\sum_{i+j \geq -1}^{l \leq i, j < -l} \frac{[z]^2}{2\pi p(1-p)\sqrt{N^2 - S^2}} \cdot \exp \left( -\frac{\left( \frac{h}{2} + iz \right)^2 (N + S) + \left( \frac{h}{2} + jz \right)^2 (N - S)}{2p(1-p)(N^2 - S^2)} + o(1) \right).$$

This is equal to

$$\frac{z^2}{2\pi p(1-p)N} \exp \left( -\frac{h^2 N}{4p(1-p)(N^2 - S^2)} + o(1) \right) \sum_{i+j \geq -1}^{l \leq i, j < -l} \exp \left( -\frac{h(i + j)zN + hzS(i - j) + z^2 N(i^2 + j^2) + z^2 S(i^2 - j^2)}{2p(1-p)(N^2 - S^2)} \right).$$

We can bound the above by noting that $|i - j| \leq \sqrt{2i^2 + j^2}$ and $|i^2 - j^2| \leq i^2 + j^2$. As we also have that $Np(1-p)/2 \leq z^2 l \leq Np(1-p)$, the inner sum appearing in (22) is at most

$$\sum_{i+j \geq -1}^{l \leq i, j < -l} \exp \left( -(i+j) + \sqrt{\frac{i^2 + j^2}{4l^2}} - \frac{i^2 + j^2}{4l^2} + o(1) \right).$$

(23)
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 \(|i - j| \geq \sqrt{21tl}\), then \(i^2 + j^2 \geq \frac{21tl}{4}\), and so
\[
\frac{i^2 + j^2}{4l} - \sqrt{\frac{i^2 + j^2}{4l}} \geq \left(\frac{21}{8} - \frac{21}{8}\right)t.
\]
Therefore the number of points \((i,j)\) in the plane with integer coordinates
and satisfying both \(i^2 + j^2 \geq \frac{21tl}{4}\) and \(|i - j| < \sqrt{21tl}\)
is at most 
\[
2\left(t + 1\right)\sqrt{t} \exp\left(-(t-1)\right).
\]
The latter sum is less than \(50\sqrt{l}\), and so the sum in (19) is bounded above by
\[
\frac{50\sqrt{p(1-p)N}}{h\pi} \exp\left(-\frac{h^2}{4p(1-p)N} + o(1)\right).
\]
Secondly we bound the probability (20). As \(l > \frac{h^2}{8Np(1-p)}\) we have that
\[
\mathbb{P}\left(X_1 \geq \mu_1 + \frac{h}{2} + lz\right) < \mathbb{P}\left(X_1 \geq \mu_1 + \frac{3h}{4}\right).
\]
By Proposition 18 we get that the quantity in (20) is at most
\[
\frac{4\sqrt{p(1-p)(N-S)}}{3h\sqrt{2\pi}} \exp\left(-\frac{9h^2}{32p(1-p)(N-S)} + o(1)\right) < \frac{2}{3\sqrt{\pi}} \frac{\sqrt{2p(1-p)N}}{h} \exp\left(-\frac{h^2}{4p(1-p)N} + o(1)\right).
\]
Similarly, the probability in (21) is at most
\[
\frac{2}{3\sqrt{\pi}} \frac{\sqrt{2p(1-p)N}}{h} \exp\left(-\frac{h^2}{4p(1-p)N} + o(1)\right).
\]
As \(\frac{50}{\sqrt{2\pi}} + \frac{2}{3\sqrt{\pi}} < e^3\) we get that the sum of our three bounds, (24), (25), and (26) is at most the stated bound. 

**Proposition 23.** Suppose that \(p(1-p)N = \omega(N)\), that
\[
\omega(N)(p(1-p)N)^{\frac{1}{2}} \leq pS = o((p(1-p)N)^{\frac{1}{2}})
\]
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 \(\mu_1\) and variance \(\sigma_1^2\) and \(Z_2 = B(N+S-T,p)\) with mean \(\mu_2\) and variance \(\sigma_2^2\).
Proof. Let $\phi(i)$ be the probability that $Z_1 = Z_2 + pT = pN + i$, then

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

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

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

Hence, we get

$$\psi(i) = \left(1 - \frac{S + i}{(1-p)N}\right) \left(1 + \frac{S - i}{(1-p)(N - T)}\right), \tag{27}$$

and so $\psi$ is a decreasing function of $i$. By noting that $e^{x - x^2} \leq (1 + x) \leq e^x$, for $x \geq -\frac{1}{2}$, we can bound $\psi$ for $i = o(p(1-p)N)$ (when $N$ is large enough). We apply $e^{x - x^2} \leq (1 + x)$ for the terms in the numerator of (27) and $\leq (1 + x) \leq e^x$ for the terms in the denominator of (27) to get the following lower bound of $\psi$

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

and we apply $\leq (1 + x) \leq e^x$ for the terms in the numerator of (27) and $e^{x - x^2} \leq (1 + x)$ for the terms in the denominator of (27) to get the following upper bound of $\psi$

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

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

$$\psi\left(\frac{pS}{2}\right) < \exp\left( - \left(\frac{(2N - 3T)S}{2(1-p)N(N - T)}\right) (1 + o(1)) \right)$$

$$< \exp\left( - \frac{(2N - 3T)S}{4(1-p)N(N - T)} \right)$$

$$< 1 - \frac{S}{3(1-p)N} \tag{28}$$

and

$$\psi\left(-\frac{pS}{2}\right) > \exp\left( \left(\frac{(2N + T)S}{2(1-p)N(N - T)}\right) (1 + o(1)) \right)$$

$$> \exp\left( \frac{(2N + t)S}{4(1-p)N(N - T)} \right)$$

$$> 1 + \frac{S}{3(1-p)N}. \tag{29}$$

Therefore $\psi$ 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 $\phi$ 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 \( \mu_2' \) and variance \((\sigma_2')^2\). By Corollary \[17\] 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 \(|i| \leq \frac{pS}{2}\). This is maximized when \( i = \frac{pST - 2pS^2}{2N - T} \) and there takes the value
\[ \frac{1}{2\pi p(1 - p)N} \exp \left( -\frac{pS^2((2N - T)^2 - (T - 2S)^2)}{2(1 - p)(N - S)(N + S - T)(2N - T)} + o(1) \right) = \frac{1}{2\pi p(1 - p)N} \exp \left( -\frac{2pS^2}{(1 - p)(2N - T)} + o(1) \right). \]

We also obtain the bounds (for \( N \) large enough)
\[ \phi\left(\frac{pS}{2}\right) < \frac{1}{2p(1 - p)\pi N} \exp \left( -\frac{pS^2(10N + 8S - 9T)}{8(1 - p)(N - S)(N + S - T)} + o(1) \right) \]
\[ < \frac{1}{2p(1 - p)\pi N} \exp \left( -\frac{9pS^2}{8(1 - p)N} \right) \]
and
\[ \phi\left(-\frac{pS}{2}\right) < \frac{1}{2p(1 - p)\pi N} \exp \left( -\frac{pS^2(10N - 8S - T)}{8(1 - p)(N - S)(N + S - T)} + o(1) \right) \]
\[ < \frac{1}{2p(1 - p)\pi N} \exp \left( -\frac{9pS^2}{8(1 - p)N} \right). \]

Putting this all together and applying \[28\] and \[29\], we obtain (for \( N \) large enough)
\[ \mathbb{P}(Z_1 = Z_2 = pT) < pS \max_i \phi(i) + \frac{1}{1 - \psi\left(\frac{pS}{2}\right)} \phi\left(\frac{pS}{2}\right) + \frac{\psi\left(\frac{pS}{2}\right)}{\psi\left(\frac{pS}{2}\right) - 1} \phi\left(-\frac{pS}{2}\right) \]
\[ < \frac{S}{2\pi (1 - p)N} \exp \left( -\frac{2pS^2}{(1 - p)(2N - T)} + o(1) \right) + 3 \frac{pS}{p\pi S} \exp \left( -\frac{9pS^2}{8(1 - p)N} \right). \]

□

We end with some propositions about the number of edges in and between sets in \( G(n, p) \).

**Proposition 24.** Suppose that \( p(1 - p)n = \omega(n) \). If \( n \) is large enough, then for all \( t > \frac{4}{3} \), we have that with probability at least \( 1 - 4^{-t} \) every set in \( G(n, p) \) of size \( t \) has at most \( p(l)^t + 2t \sqrt{p(1 - p)t} \) edges.
Proof. The expected number of sets of size $t$ with more than 
\[ p\left(\frac{t}{2}\right) + 2t\sqrt{p(1-p)t} \]
edges is 
\[ \binom{n}{t} \mathbb{P}\left( B\left(\frac{t}{2}, p\right) \geq p\left(\frac{t}{2}\right) + 2t\sqrt{p(1-p)t} \right). \]

By Lemma 20 and the fact that $\binom{n}{t} \leq \left(\frac{en}{t}\right)^t$, this expectation is at most 
\[ (5e)^t \exp \left( -\frac{4p(1-p)t^3}{2p(1-p)\binom{t}{2} + \frac{4t\sqrt{p(1-p)t}}{3}} \right). \]

As $\sqrt{p(1-p)t} = \omega(n)$, we have that if $n$ is large enough, then for all $t > \frac{n}{5}$ we have
\[ 2p(1-p)\binom{t}{2} + \frac{4t\sqrt{p(1-p)t}}{3} \leq 1.001p(1-p)t^2. \]

Substituting this in 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,
\[ \exp \left( t(\log 5 + 1) - \frac{4p(1-p)t^3}{1.001p(1-p)t^2} \right) < 4^{-t}. \]

\[ \square \]

**Proposition 25.** Suppose that $p(1-p)n = \omega(n)$. If $n$ is large enough then for all $t$ in the range $\frac{n}{9} < t \leq \frac{n}{7}$, 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
\[ \binom{n}{t} \mathbb{P}\left( B(t(n-t), (1-p)) \geq (1-p)t(n-t) + 3t\sqrt{p(1-p)(n-t)} \right). \]

By Lemma 20 and the fact that $\binom{n}{t} \leq \left(\frac{en}{t}\right)^t$, this expectation is at most
\[ (5e)^t \exp \left( -\frac{9p(1-p)t^2(n-t)}{2p(1-p)t(n-t) + 2t\sqrt{p(1-p)(n-t)}} \right). \]

As $\sqrt{(n-t)p(1-p)} = \omega(n)$, we have that if $n$ is large enough, then for all $t$ in the range $\frac{n}{9} < t \leq \frac{n}{7}$,
\[ 2p(1-p)t(n-t) + 2t\sqrt{p(1-p)(n-t)} \leq \frac{9}{4}p(1-p)t(n-t). \]

Substituting this in 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
\[ \exp \left( t(\log 5 + 1) - 4t \right) < 4^{-t}. \]

\[ \square \]
Proposition 26. Suppose that \( (1-p)n \geq 4 \log n \). If \( n \) is large enough, then for all \( t \leq \frac{n}{5} \) we have that with probability at least \( 1 - \frac{1}{150} \), 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

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

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

We can rewrite this as,

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

By Lemma 20, this is at most

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

For \( t < \frac{n}{25} \), using the inequality \( \binom{n}{t} \leq n^t \) we have that the quantity in (30) is (for \( n \) large enough) less than

\[
n^t \exp \left( -\frac{pt(\frac{11n}{15})^2}{4p(1-p)t(n+t-2) + \frac{4p(n-2t-1)}{3}} \right) < n^t \exp \left( -\frac{4t \log n}{n} \cdot \frac{121n}{480} \right) = n^{-\frac{t}{120}}.
\]

For \( \frac{n}{25} \leq t \leq \frac{n}{5} \), using the inequality \( \binom{n}{t} \leq \left( \frac{en}{t} \right)^t \) we have that the quantity in (30) is (for \( n \) large enough) less than,

\[
\left( \frac{en}{t} \right)^t \exp \left( -\frac{pt(\frac{3n}{5})^2}{4p(1-p)t(n+t-2) + \frac{4p(n-2t-1)}{3}} \right) < \left( \frac{24e}{n^\frac{1}{120}} \right)^t < n^{-\frac{t}{120}}.
\]

Corollary 27. Suppose that \( pn \geq \log n \). If \( n \) is large enough, then for all \( t \) satisfying \( n^{\frac{24}{55}} \leq t \leq \frac{n}{5} \), we have that with probability at least \( 1 - \frac{1}{150} \), 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 26 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

\[
\left( \frac{en}{t} \right)^t \exp \left( -\frac{pt(\frac{3n}{5})^2}{4p(1-p)t(n+t-2) + \frac{4p(n-2t-1)}{3}} \right) < \left( \frac{e}{n^\frac{1}{240}} \right)^t < n^{-\frac{t}{120}}.
\]

Proposition 28. For every fixed \( \epsilon > 0 \) and \( p \geq \frac{(1+\epsilon) \log n}{n} \), 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

\[ n \mathbb{P}(B(n-1, p) \leq 8) = n \sum_{i=0}^{8} \binom{n-1}{i} p^i (1-p)^{n-1-i} \leq n \left( \frac{1}{8} \binom{n-1}{8} p^8 (1-p)^{n-9} \left( 1 + \frac{9(1-p)}{p(n-9)} + \left( \frac{9(1-p)}{p(n-9)} \right)^2 + \ldots \right) \right) \leq \frac{9n^9}{8!} p^8 (1-p)^{n-9}. \]  

(31)

These inequalities follow as \( \max_{i \leq 8} \mathbb{P}(B(n-1, p) = i) \) occurs (for \( n \) large enough) when \( i = 8 \), and so \( \mathbb{P}(B(n-1, p) \leq 8) \leq \mathbb{P}(B(n-1, p) = 8) \). The last line of (31) is maximised over \( 0 \leq p \leq 1 \) when \( p = \frac{1}{8} \), that is when \( p = \frac{8}{n-1} \). So for \( p \) in our range, (31) is maximised when \( p = \frac{(1+\epsilon) \log n}{n} \).

Therefore (for \( n \) large enough)

\[ n \mathbb{P}(B(n-1, p) \leq 8) \leq \frac{9n^9 (1+\epsilon)^8 (\log n)^8}{8!n^8} e^{-\left( \frac{n-9(1+\epsilon) \log n}{n} \right)} \leq \frac{(\log n)^8}{n^2}. \]

□

Proposition 29. Suppose that \( (1+\epsilon) \log n \leq pn \leq 5 \log n \). If \( n \) is large enough, then for all \( t \) satisfying \( t \leq \frac{n^{20}}{12} \), we have that with probability at least \( 1 - \frac{t}{120} \), 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

\[ \binom{n}{t} \mathbb{P}(B\left( \left( \begin{array}{c} t \\ 2 \end{array} \right), p \right) \geq 2t) = \binom{n}{t} \sum_{i=2t}^{n} \binom{\left( \begin{array}{c} t \\ 2 \end{array} \right)}{i} p^i (1-p)^{\left( \begin{array}{c} t \\ 2 \end{array} \right) - i}. \]  

(32)

By carefully bounding the summands in (32) for \( i = 2t \) and \( i = 2t+1 \), we shall get a good bound on the total sum. We have that

\[ \left( \frac{\left( \begin{array}{c} t \\ 2 \end{array} \right)}{2t} \right) p^{2t} (1-p)^{\left( \begin{array}{c} t \\ 2 \end{array} \right) - 2t} < \left( \frac{ep(t-1)}{4} \right)^{2t} < \left( \frac{5et \log n}{4n} \right)^{2t}. \]

We also get

\[ \frac{\left( \frac{t}{2t+1} \right) p^{2t+1} (1-p)^{\left( \begin{array}{c} t \\ 2 \end{array} \right) - 2t-1}}{\left( \frac{\left( \begin{array}{c} t \\ 2 \end{array} \right)}{2t} \right) p^{2t} (1-p)^{\left( \begin{array}{c} t \\ 2 \end{array} \right) - 2t}} = \frac{p((\frac{t}{2} - 2t)}{(1-p)(2t+1)} \leq pt < \frac{1}{2}. \]

Because the ratio between consecutive terms in the sum in (32) decreases as \( i \) increases, we have from above that the total sum is at most
twice the first term, therefore
\[
\binom{n}{t} \mathbb{P}(B(\binom{t}{2}, p) \geq 2t) \leq \binom{n}{t} 2 \left( \frac{5et \log n}{4n^2} \right)^{2t} \\
\leq 2^{\frac{3t}{16t^2}} e^{\frac{3}{2} (\log n)^2 t^2} \\
\leq \left( \frac{C (\log n)^2}{n^{\frac{1}{10}}} \right)^{t},
\]
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 \(n^{-\frac{1}{10}}\).

\[\square\]

Acknowledgement. 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, Sur une modification de l’inégalité de tchebychef, 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. 5, 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 Órn 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).

\[\diamond\]

\[\diamond\]

\[\diamond\]

\[\diamond\]

\[\diamond\]

\[\diamond\]

\[\diamond\]

\[\diamond\]