Bootstrap percolation on the random graph $G_{n,p}$

Svante Janson* Tomasz Łuczak†‡ Tatyana Turova§¶ Thomas Vallier‖

15 December, 2010; revised 27 September, 2011

Abstract

Bootstrap percolation on the random graph $G_{n,p}$ is a process of spread of "activation" on a given realization of the graph with a given number of initially active nodes. At each step those vertices which have not been active but have at least $r \geq 2$ active neighbours become active as well.

We study the size A^* of the final active set. The parameters of the model are, besides r (fixed) and n (tending to ∞), the size $a = a(n)$ of the initially active set and the probability $p = p(n)$ of the edges in the graph. We show that the model exhibits a sharp phase transition: depending on the parameters of the model the final size of activation with a high probability is either $n - o(n)$ or it is $o(n)$. We provide a complete description of the phase diagram on the space of the parameters of the model. In particular, we find the phase transition and compute the asymptotics (in probability) for A^*; we also prove a central limit theorem for A^* in some ranges. Furthermore, we provide the asymptotics for the number of steps until the process stops.

1 Introduction

Bootstrap percolation on a graph G is defined as the spread of *activation* or *infection* according to the following rule, with a given threshold $r \geq 2$: We start with a set $\mathcal{A}(0) \subseteq V(G)$ of *active* vertices. Each inactive vertex that has at least r active neigbours becomes active. This is repeated until no more vertices become active, i.e., when no inactive vertex has r or more active neigbours. Active vertices never become inactive, so the set of active vertices grows monotonously.

To avoid confusion, we will use the terminology that each active vertex *infects* all its neighbours, so that a vertex that is infected (at least) r times becomes active.

We are mainly interested in the final size A^* of the active set, and in particular whether eventually all vertices will be active or not. If they are, we say that the initial set $\mathcal{A}(0)$ *percolates* (completely). We will study a sequence of graphs of order $n \to \infty$; we then also say that (a

Research mainly done while the authors all visited Institut Mittag-Leffler (Djursholm, Sweden), during the program 'Discrete Probability' 2009.

*Department of Mathematics, Uppsala University, Uppsala, Sweden

†Department of Discrete Mathematics, Adam Mickiewicz University, Poznań, Poland

‡Partially supported by the Foundation for Polish Science

§Center for Mathematical Sciences, Lund University, Lund, Sweden

¶Partially supported by the Swedish Science Foundation

‖Institut Mittag-Leffler, Djursholm, Sweden

sequence of) $\mathcal{A}(0)$ *almost percolates* if the number of vertices that remain inactive is $o(n)$, i.e., if $A^* = n - o(n)$.

Bootstrap percolation on a lattice (which is a special example of a cellular automata) was introduced in 1979 by Chalupa, Leath and Reich [25] as a simplified model of some magnetic systems. Since then bootstrap percolation has been studied on various graphs, both deterministic and random. One can study either a random initial set or the deterministic problem of choosing an initial set that is optimal in some sense. A simple example of the latter is the classical folklore problem to find the minimal percolating set in a two-dimensional grid (i.e., a finite square $[n]^2$ in the square lattice); see Balogh and Pete [14] and Bollobás [19]. (These references also treat higher-dimensional grids $[n]^d$.) Another extremal problem is studied by Morris [40]. The problem with a random initial set was introduced by Chalupa, Leath and Reich [25] (lattices and regular infinite tree), and further studied on lattices by Schonmann [43]; it has in particular been studied on finite grids (in two dimensions or more), see Aizenman and Lebowitz [1], Balogh and Pete [14], Cerf and Cirillo [21], Cerf and Manzo [22], Holroyd [30], Balogh, Bollobás and Morris [10], Gravner, Holroyd and Morris [28]. In a recent paper Balogh, Bollobás, Duminil-Copin and Morris [7] derived a sharp asymptotic for the critical density (i.e., the critical size of a random initial set) for bootstrap percolation on grids of any dimension, generalizing results of Balogh, Bollobás, and Morris [9]. Grids with a different edge set where studied by Holroyd, Liggett and Romik [31]. The study of bootstrap percolation on lattices is partly explained by its origin in statistical physics, and the bootstrap process is being successfully used in studies of the Ising model, see [23], [24], [27], [41]. Lately bootstrap percolation has also been studied on varieties of graphs different from lattices and grids, see e.g., Balogh and Bollobás [6] (hypercube); Balogh, Peres and Pete [13] (infinite trees); Balogh and Pittel [15], Janson [33] (random regular graphs); an extension where the threshold may vary between the vertices is studied by Amini [2]. An anisotropic bootstrap percolation was studied by Duminil-Copin and van Enter [26]. Further, a graph bootstrap percolation model introduced by Bollobás [18] already in 1968, where edges are infected instead of vertices, was analysed recently by Balogh, Bollobás and Morris [11], and Balogh, Bollobás, Morris and Riordan [12].

In the present paper, we study bootstrap percolation on the Erdös–Rényi random graph $G_{n,p}$ with an initial set $\mathcal{A}(0)$ consisting of a given number a of vertices chosen at random. (By symmetry, we obtain the same results for any deterministic set of a vertices.) Recall that $G_{n,p}$ is the random graph on the set of vertices $V_n = \{1, \dots, n\}$ where all possible edges between pairs of different vertices are present independently and with the same probability p. As usual, we let $p = p(n)$ depend on n.

A problem equivalent to bootstrap percolation on $G_{n,p}$ in the case $p = \lambda/n$ was studied by Scalia-Tomba [42], although he used a different formulation as an epidemic. (Ball and Britton [4, 5] study a more general model with different degrees of severity of infection.) Otherwise, bootstrap percolation on $G_{n,p}$ was first studied by Vallier [48]; we here use a simple method (the same as [42]) that allows us to both simplify the proofs and improve the results. We will state the results for a general fixed $r \geq 2$ (the case $r = 1$ is much different, see Remark 5.9); the reader may for simplicity consider the case $r = 2$ only, since there are no essential differences for higher r.

We will see that there is a threshold phenomenon: typically, either the final size A^* is small (at most twice the initial size a), or it is large (sometimes exactly n, but if p is so small that there are vertices of degree less than r, these can never become active except initially so eventually at most $n - o(n)$ will become infected). We can study the threshold in two ways: In the first version, we keep n and p fixed and vary a. In the second version, we fix n and a and vary p. We will state some results for both versions and some for the former version only; these too can easily be translated to the second version. We will also study dynamical versions, where we add new external infections or activations or new edges until we reach the threshold, see Section 4.

Apart from the final size A^*, we will also study the time τ the bootstrap process takes until completion. We count the time in *generations*: generation 0 is $\mathcal{A}(0)$, generation 1 is the set of other vertices that have at least r neighbours in generation 0, and so on. The process stops as soon as there is an empty generation, and we let τ be the number of (non-empty) generations. Thus, if we let $\mathcal{G}_k$ be the set of vertices activated in generation k, then

$$\tau := \max\{k \geq 0 : \mathcal{G}_k \neq \emptyset\} = \min\{k \geq 1 : \mathcal{G}_k = \emptyset\} - 1. \tag{1.1}$$

Remark 1.1. Bootstrap percolation does not seem to be a good model for usual infectious diseases. (See, however, Ball and Britton [4].) It might be a better model for the spread of rumors or other ideas or beliefs. (Cf. the well-known rule "What I tell you three times is true." in Carroll [20].)

Bootstrap percolation can be also viewed as a simplified model for propagation of activity in a neural network. Although related neuronal models are too involved for a rigorous analysis (see, e.g., [37], [45], [47]) they inspired study of bootstrap percolation on $G_{n,p}$ by Vallier [48]. There is a further discussion on the application of bootstrap percolation on $G_{n,p}$ to neuromodelling in [46].

Remark 1.2. Instead of $G_{n,p}$, one might study the random graph $G(n,m)$, with a given number $m = m(n)$ of edges. It is easy to obtain result for $G(n,m)$ from our results for $G_{n,p}$, using monotonicity, but we usually leave this to the reader. (In the dynamical model in Subsection 4.3, we consider $G(n,m)$, however.)

Remark 1.3. An alternative to starting with an initial active set of fixed size a is to let each vertex be initially activated with probability $\alpha = \alpha(n) > 0$, with different vertices activated independently. Note that this is the same as taking the initial size a random with $a \in \mathrm{Bin}(n,\alpha)$. For most results the resulting random variation in a in negligible, and we obtain the same results as for $a = n\alpha$, but for the Gaussian limit in Theorems 3.6(iii) and 4.5, the asymptotic variances are changed by constant factors. We leave the details to the reader.

Some open problems arise from our study. In [10], Balogh, Bollobás and Morris determine the critical probability for bootstrap percolation on grids when the dimension $d = d(n) \to \infty$. A similar idea translated to the $G(n,p)$ graph would be to study what happens when $r = r(n) \to \infty$. This problem is not treated here although our methods might be useful also for such problems. The problem of majority percolation where a vertex becomes activated if at least half of its neighbours are active ($r(v) = d(v)/2$) has been studied on the hypercube by Balogh, Bollobás and Morris [8]. On the d-dimensional grid $d(v)/2 = d$ but on the $G(n,p)$ graph, this problem is completely different and still open. (We thank the referee for these suggestions.)

The method is described in Section 2. The main results are stated in Section 3, with further results in Sections 4–5. Proofs are given in Sections 6–11.

1.1 Notation

All unspecified limits are as $n \to \infty$. We use $\xrightarrow{\mathrm{d}}$ for convergence in distribution and $\xrightarrow{\mathrm{p}}$ for convergence in probability of random variables; we further use O_{p} and o_{p} in the standard sense (see e.g. [35] and [34]), and we use w.h.p. (with high probability) for events with probability tending to 1 as $n \to \infty$. Note that, for example, '$= o(1)$ w.h.p.' is equivalent to '$= o_p(1)$' and to '$\xrightarrow{\mathrm{p}} 0$', and that '$\sim a_n$ w.h.p.' is equivalent to '$= (1 + o_p(1))a_n$' (see [34]). A statement of the type "when $\mathcal{P}$, then w.h.p. $\mathcal{Q}$" (or similar wording), where $\mathcal{P}$ and $\mathcal{Q}$ are two events, means that $\mathbb{P}(\mathcal{P}$ and (not $\mathcal{Q})) \to 0$, i.e. that w.h.p. "(not $\mathcal{P}$) or $\mathcal{Q}$" holds. (See e.g. Theorem 3.2 and Proposition 10.10.) If $\mathbb{P}(\mathcal{P})$ is bounded away from 0, this is equivalent to "conditioned on $\mathcal{P}$, $\mathcal{Q}$ holds w.h.p.".

If X_n is a sequence of random variables, and μ_n and σ_n^2 are sequences of real numbers, with $\sigma_n^2 > 0$, we say that $X_n \in \mathrm{AsN}(\mu_n, \sigma_n^2)$ if $(X_n - \mu_n)/\sigma_n \xrightarrow{\mathrm{d}} N(0,1)$.

Occasionally we use the subsubsequence principle [35, p. 12], which says that to prove a limit result (for example for real numbers, or for random variables in probability or in distribution), it is sufficient to show that every subsequence has a subsubsequence where the result holds. We may thus without loss of generality select convenient subsequences in a proof, for example such that another given sequence either converges or tends to ∞.

2 A useful reformulation

In order to analyse the bootstrap percolation process on $G_{n,p}$, we change the time scale; we forget the generations and consider at each time step the infections from one vertex only. Choose $u_1 \in \mathcal{A}(0)$ and give each of its neighbours a *mark*; we then say that u_1 is *used*, and let $\mathcal{Z}(1) := \{u_1\}$ be the set of used vertices at time 1. We continue recursively: At time t, choose a vertex $u_t \in \mathcal{A}(t-1) \setminus \mathcal{Z}(t-1)$. We give each neighbour of u_t a new mark. Let $\Delta\mathcal{A}(t)$ be the set of inactive vertices with r marks; these now become active and we let $\mathcal{A}(t) = \mathcal{A}(t-1) \cup \Delta\mathcal{A}(t)$ be the set of active vertices at time t. We finally set $\mathcal{Z}(t) = \mathcal{Z}(t-1) \cup \{u_t\} = \{u_s : s \le t\}$, the set of used vertices. (We start with $\mathcal{Z}(0) = \emptyset$, and note that necessarily $\Delta\mathcal{A}(t) = \emptyset$ for $t < r$.)

The process stops when $\mathcal{A}(t) \setminus \mathcal{Z}(t) = \emptyset$, i.e., when all active vertices are used. We denote this time by T;

$$T := \min\{t \ge 0 : \mathcal{A}(t) \setminus \mathcal{Z}(t) = \emptyset\}. \tag{2.1}$$

Clearly, $T \le n$; in particular, T is finite. The final active set is $\mathcal{A}(T)$; it is clear that this is the same set as the one produced by the bootstrap percolation process defined in the introduction; only the time development differs. Hence we may as well study the version just described. (This is true for any choice of the vertices u_t. For definiteness, we may assume that we keep the unused, active vertices in a queue and choose u_t as the first vertex in the queue, and that the vertices in $\Delta\mathcal{A}(t)$ are added at the end of the queue in order of their labels. Thus u_t will always be one of the oldest unused, active vertices, which will enable us to recover the generations; see further Section 10. In Section 4, we consider other ways of choosing u_t.) This reformulation was used already by Scalia-Tomba [42] (for a more general model). It is related to the (continuous-time) construction by Sellke [44] for an epidemic process.

Let $A(t) := |\mathcal{A}(t)|$, the number of active vertices at time t. Since $|\mathcal{Z}(t)| = t$ and $\mathcal{Z}(t) \subseteq \mathcal{A}(t)$ for $t = 0, \dots, T$, we also have

$$T = \min\{t \ge 0 : A(t) = t\} = \min\{t \ge 0 : A(t) \le t\}. \tag{2.2}$$

Moreover, since the final active set is $\mathcal{A}(T) = \mathcal{Z}(T)$, its size A^* is

$$A^* := A(T) = |\mathcal{A}(T)| = |\mathcal{Z}(T)| = T. \tag{2.3}$$

Hence, the set $\mathcal{A}(0)$ percolates if and only if $T = n$, and $\mathcal{A}(0)$ almost percolates if and only if $T = n - o(n)$.

We analyse this process by the standard method of revealing the edges of the graph $G_{n,p}$ only on a need-to-know basis. We thus begin by choosing u_1 as above and then reveal its neighbours; we then find u_2 and reveal its neighbours, and so on. Let, for $i \notin \mathcal{Z}(s)$, $I_i(s)$ be the indicator that there is an edge between the vertices u_s and i. This is also the indicator that i gets a mark at time s, so if $M_i(t)$ is the number of marks i has at time t, then

$$M_i(t) = \sum_{s=1}^{t} I_i(s), \tag{2.4}$$

at least until i is activated (and what happens later does not matter). Note that if $i \notin \mathcal{A}(0)$, then, for every $t \leq T$, $i \in \mathcal{A}(t)$ if and only if $M_i(t) \geq r$. The crucial feature of this description of the process, which makes the analysis simple, is that the random variables $I_i(s)$ are i.i.d. $\mathrm{Be}(p)$.

We have defined $I_i(s)$ only for $s \leq T$ and $i \notin \mathcal{Z}(s)$, but it is convenient to add further (redundant) variables so that $I_i(s)$ are defined, and i.i.d., for all $i \in V_n$ and all $s \geq 1$. One way to do this formally is to reverse the procedure above. We start with i.i.d. $I_i(s) \in \mathrm{Be}(p)$, for $i \in V_n$ and $s \geq 1$, and a set $\mathcal{A}(0) \subseteq V_n$. We let $\mathcal{Z}(0) := \emptyset$ and start with an empty graph on V_n. We then, as above, for $t = 1, \dots, n$ select $u_t \in \mathcal{A}(t-1) \setminus \mathcal{Z}(t-1)$ if this set is non-empty; otherwise we select $u_t \in V_n \setminus \mathcal{Z}(t-1)$ (taking, for example, the smallest such vertex). We define $M_i(t)$ by (2.4) for all $i \in V_n$ and $t \geq 0$, and update $\mathcal{A}(t) := \mathcal{A}(0) \cup \{i : M_i(t) \geq r\}$ and $\mathcal{Z}(t) := \mathcal{Z}(t-1) \cup \{u_t\} = \{u_s : s \leq t\}$. Furthermore, add an edge $u_t i$ to the graph for each vertex $i \notin \mathcal{Z}(t)$ such that $I_i(t) = 1$. Finally, define T by (2.1) or (2.2).

It is easy to see that this constructs a random graph $G_{n,p}$ and that $\mathcal{A}(t)$, $t \leq T$, is as above for this graph, so the final active set of the bootstrap percolation on the graph is $\mathcal{A}(T)$.

Define also, for $i \in V_n \setminus \mathcal{A}(0)$,

$$Y_i := \min\{t : M_i(t) \geq r\}. \tag{2.5}$$

If $Y_i \leq T$, then Y_i is the time vertex i becomes active, but if $Y_i > T$, then i never becomes active. Thus, for $t \leq T$,

$$\mathcal{A}(t) = \mathcal{A}(0) \cup \{i \notin \mathcal{A}(0) : Y_i \leq t\}. \tag{2.6}$$

By (2.4), each $M_i(t)$ has a binomial distribution $\mathrm{Bin}(t, p)$. Further, by (2.4) and (2.5), each Y_i has a negative binomial distribution $\mathrm{NegBin}(r, p)$:

$$\mathbb{P}(Y_i = k) = \mathbb{P}\big(M_i(k-1) = r-1,\, I_i(k) = 1\big) = \binom{k-1}{r-1} p^r (1-p)^{k-r}; \tag{2.7}$$

moreover, these random variables Y_i are i.i.d.

We let, for $t = 0, 1, 2, \dots$,

$$S(t) := |\{i \notin \mathcal{A}(0) : Y_i \leq t\}| = \sum_{i \notin \mathcal{A}(0)} \mathbf{1}\{Y_i \leq t\}, \tag{2.8}$$

so, by (2.6), and our notation $A(0) = a$,

$$A(t) = A(0) + S(t) = S(t) + a. \tag{2.9}$$

By (2.9), (2.2) and (2.3), it suffices to study the stochastic process $S(t)$. Note that $S(t)$ is a sum of $n-a$ i.i.d. processes $\mathbf{1}\{t \geq Y_i\}$, each of which is 0/1-valued and jumps from 0 to 1 at time Y_i, where Y_i has the distribution $\mathrm{NegBin}(r,p)$ in (2.7). We write $S(t) = S_{n-a}(t)$ when we want to emphasize the number of summands in $S(t)$; more generally we define $S_m(t) := \sum_{i=1}^m \mathbf{1}\{Y_i \leq t\}$ for any $m \leq n$ (assuming for consistency that $\mathcal{A}(0) = \{n-a+1, \dots, n\}$).

The fact that $S(t)$, and thus $A(t)$, is a sum of i.i.d. processes makes the analysis easy; in particular, for any given t,

$$S(t) \in \mathrm{Bin}\bigl(n-a, \pi(t)\bigr), \tag{2.10}$$

where

$$\pi(t) := \mathbb{P}(Y_1 \leq t) = \mathbb{P}(M_1(t) \geq r) = \mathbb{P}\bigl(\mathrm{Bin}(t,p) \geq r\bigr). \tag{2.11}$$

In particular, we have

$$\mathbb{E}\, S(t) = (n-a)\pi(t), \tag{2.12}$$

$$\mathrm{Var}\, S(t) = (n-a)\pi(t)\bigl(1-\pi(t)\bigr) \leq \mathbb{E}\, S(t) \leq n\pi(t). \tag{2.13}$$

To avoid rounding to integers sometimes below, we define $S(t) := S(\lfloor t \rfloor)$ and $\pi(t) := \pi(\lfloor t \rfloor)$ for all real $t \geq 0$. We also sometimes (when it is obviously harmless) ignore rounding to simplify notation.

3 Main results

3.1 Limits in probability

For given r, n, and p we define, for reasons that will be seen later,

$$t_{\mathsf{c}} := \left(\frac{(r-1)!}{np^r}\right)^{1/(r-1)}, \tag{3.1}$$

$$a_{\mathsf{c}} := \left(1 - \frac{1}{r}\right) t_{\mathsf{c}}, \tag{3.2}$$

$$b_{\mathsf{c}} := n \frac{(pn)^{r-1}}{(r-1)!} e^{-pn}. \tag{3.3}$$

In particular, for $r=2$, $t_{\mathsf{c}} := 1/(np^2)$ and $a_{\mathsf{c}} := 1/(2np^2)$. For future use, note also that (3.1) can be written

$$n \frac{(pt_{\mathsf{c}})^r}{r!} = \frac{t_{\mathsf{c}}}{r}. \tag{3.4}$$

Our standard assumptions $p \ll n^{-1/r}$ and $p \gg n^{-1}$ imply that

$$t_{\mathsf{c}} \to \infty, \qquad pt_{\mathsf{c}} \to 0, \qquad t_{\mathsf{c}}/n \to 0, \qquad a_{\mathsf{c}} \to \infty, \qquad a_{\mathsf{c}}/n \to 0, \qquad b_{\mathsf{c}}/n \to 0, \qquad pb_{\mathsf{c}} \to 0, \tag{3.5}$$

and further

$$\mathbb{E}\, S_n(t_{\mathsf{c}}) = n\pi(t_{\mathsf{c}}) \sim n \frac{(pt_{\mathsf{c}})^r}{r!} = \frac{t_{\mathsf{c}}}{r}, \tag{3.6}$$

$$
\begin{aligned}
n - \mathbb{E}\, S_n(n) = n\bigl(1-\pi(n)\bigr) &= n\,\mathbb{P}\bigl(\mathrm{Bin}(n,p) \le r-1\bigr) \sim n\,\mathbb{P}\bigl(\mathrm{Bin}(n,p) = r-1\bigr) \\
&\sim b'_{\mathsf{c}} := n\frac{(pn)^{r-1}}{(r-1)!}(1-p)^n.
\end{aligned} \tag{3.7}
$$

If $p \ll n^{-1/2}$, then $(1-p)^n \sim e^{-np}$ and (3.7) yields

$$
n - \mathbb{E}\, S_n(n) = n\bigl(1-\pi(n)\bigr) \sim b'_{\mathsf{c}} \sim b_{\mathsf{c}}; \tag{3.8}
$$

if p is larger ($p = \Omega(n^{-1/2})$, i.e., $n^{-1/2} = O(p)$), this is not quite true, but in this case both b'_{c} and b_{c} decrease to 0 very fast; in all cases

$$
n - \mathbb{E}\, S_n(n) = n\bigl(1-\pi(n)\bigr) = b_{\mathsf{c}} + o(b_{\mathsf{c}}+1). \tag{3.9}
$$

Recall that our main interest is in $S(t) = S_{n-a}(t)$ rather than $S_n(t)$, see (2.10); for $S(t)$ we obviously have similar results, with additional error terms depending on a, see (2.12) and, e.g., (8.2).

Note further that by (3.3), for any $\beta \in (-\infty,\infty)$,

$$
np - \bigl(\log n + (r-1)\log(np)\bigr) \to \begin{cases} -\infty \\ \beta \\ \infty \end{cases} \iff b_{\mathsf{c}} \to \begin{cases} \infty, \\ (r-1)!^{-1}e^{-\beta}, \\ 0, \end{cases}
$$

which by simple calculations yields, provided $p \ge n^{-1}$,

$$
np - \bigl(\log n + (r-1)\log\log n\bigr) \to \begin{cases} -\infty \\ \beta \\ \infty \end{cases} \iff b_{\mathsf{c}} \to \begin{cases} \infty, \\ (r-1)!^{-1}e^{-\beta}, \\ 0. \end{cases} \tag{3.10}
$$

Our first result, to be refined later, shows that the threshold for almost percolation is $a = a_{\mathsf{c}}$. The proof of the theorems in this section are given later (Sections 8–10). Let us recall that A^* is the final size of the active set, and that $A^* = T = A(T) = a + S_{n-a}(T)$.

Theorem 3.1. *Suppose that $r \ge 2$ and $n^{-1} \ll p \ll n^{-1/r}$.*

(i) *If $a/a_{\mathsf{c}} \to \alpha < 1$, then $A^* = (\varphi(\alpha) + o_{\mathrm{p}}(1))t_{\mathsf{c}}$, where $\varphi(\alpha)$ is the unique root in $[0,1]$ of*

$$
r\varphi(\alpha) - \varphi(\alpha)^r = (r-1)\alpha. \tag{3.11}
$$

(For $r=2$, $\varphi(\alpha) = 1 - \sqrt{1-\alpha}$.) Further, $A^*/a \xrightarrow{\mathrm{p}} \varphi_1(\alpha) := \frac{r}{r-1}\varphi(\alpha)/\alpha$, with $\varphi_1(0) := 1$.

(ii) *If $a/a_{\mathsf{c}} \ge 1+\delta$, for some $\delta > 0$, then $A^* = n - o_{\mathrm{p}}(n)$; in other words, we have w.h.p. almost percolation. More precisely, $A^* = n - O_{\mathrm{p}}(b_{\mathsf{c}})$.*

(iii) *In case* (ii), *if further $a \le n/2$, say, we further have complete percolation, i.e. $A^* = n$ w.h.p., if and only if $b_{\mathsf{c}} \to 0$, i.e., if and only if $np - (\log n + (r-1)\log\log n) \to \infty$.*

It is easily verified that φ_1 is a continuous, strictly increasing function $[0,1] \to [1, r/(r-1)]$. In particular, in the subcritical case (i), we thus have w.h.p. $A^* < (r/(r-1))a \le 2a$, so the activation will not spread to many more than the originally active nodes.

In the supercritical case (ii), we have the following more detailed result.

Theorem 3.2. *Suppose that $r \ge 2$, $n^{-1} \ll p \ll n^{-1/r}$ and $a = o(n)$, and that $A^* = n - o_p(n)$ as for example in Theorem 3.1*(ii). *Then,*

(i) *If $np - (\log n + (r-1)\log\log n) \to -\infty$, so $b_c \to \infty$ by* (3.10), *then $A^* = n - b_c(1 + o_p(1))$. In particular, w.h.p. we do not have complete percolation.*

(ii) *If $np - (\log n + (r-1)\log\log n) \to \infty$, so $b_c \to 0$ by* (3.10), *then w.h.p. $A^* = n$, so we have complete percolation.*

(iii) *If $np - (\log n + (r-1)\log\log n) \to \beta \in (-\infty, \infty)$, so $b_c \to b > 0$ by* (3.10), *then $n - A^* \xrightarrow{\mathrm{d}}$ $\mathrm{Po}(b)$; in particular, $\mathbb{P}(A^* = n) \to \exp(-b) \in (0,1)$.*

More generally, even if we do not have almost percolation w.h.p., the result holds w.h.p. provided $A^* \ge 3t_c$.

By the last statement we mean that $\mathbb{P}(\text{the result fails and } A^* \ge 3t_c) \to 0$. In particular, it holds w.h.p. conditioned on $A^* \ge 3t_c$, provided $\liminf \mathbb{P}(A^* \ge 3t_c) > 0$.

Remark 3.3. Let $\mathcal{B}$ be the set of vertices in $G_{n,p}$ with degrees less than r. These are never activated unless they happen to be in the initially active set $\mathcal{A}(0)$, and for each of the vertices, this has probability $a/n \to 0$ if $a = o(n)$; hence trivially $A^* \le n - |\mathcal{B}|(1 - o_p(1))$. We have, cf. (3.7) and (3.9),

$$\mathbb{E}\,|\mathcal{B}| = n\,\mathbb{P}\big(\mathrm{Bin}(n-1,p) \le r-1\big) \sim b_c + o(b_c + 1),$$

with concentration of $|\mathcal{B}|$ around its mean if $b_c \to \infty$ and a limiting Poisson distribution if $b_c \to b < \infty$, see [35, Sections 6.2–6.3] and [16]. Comparing this with Theorem 3.2 we see that in the supercritical case, and with $a = o(n)$, the final inactive set $V_n \setminus \mathcal{A}(T)$ differs from $\mathcal{B}$ by $o_p(|\mathcal{B}|)$ vertices only, and in the case $b_c = O(1)$ (combining cases (ii) and (iii) in Theorem 3.2), w.h.p. $V_n \setminus \mathcal{A}(T) = \mathcal{B}$. In other words, when we get a large active set, the vertices that remain inactive are mainly the ones with degrees less than r, and if further $b_c = O(1)$, they are w.h.p. exactly the vertices with degrees less than r.

We can, as discussed earlier, also consider thresholds for p for a given a.

Theorem 3.4. *Suppose that $r \ge 2$ and that $a \to \infty$ with $a = o(n)$. Then the threshold for p for almost percolation is*

$$p_c := \left(\frac{(r-1)^{r-1}(r-1)!}{r^{r-1}}\right)^{1/r} \left(na^{r-1}\right)^{-1/r} \tag{3.12}$$

in the sense that if, for some $\delta > 0$, $p \le (1-\delta)p_c$ then $A^* \le 2a = o(n)$ w.h.p., while if $p \ge (1+\delta)p_c$, then $A^* = n - o(n)$ w.h.p. In the latter case, further $A^* = n$ w.h.p. if and only if $p = (\log n + (r-1)\log\log n + \omega(n))/n$ for some $\omega(n) \to \infty$.

Note that $n^{-1} \ll p_{\mathsf{c}} \ll n^{-1/r}$. The equation (3.12) is the inverse to (3.2) in the sense that the functions $a \mapsto p_{\mathsf{c}}$ and $p \mapsto a_{\mathsf{c}}$ that they define are the inverses of each other. For $r = 2$, (3.12) simplifies to $p_{\mathsf{c}} = (2na)^{-1/2}$.

Remark 3.5. Note that the thresholds for complete and almost percolation are different only for large a. Indeed, for such a case the threshold p_{c} for almost percolation can be so small that the graph $G(n, p_{\mathsf{c}})$ may not be even connected. Then, besides of p_{c}, we have the second threshold for the complete percolation, e.g. if $a = n/\log n$ and $r = 2$ there are two thresholds: $p_{\mathsf{c}} = \Theta(\sqrt{\log n}/n)$ for almost percolation, and $\Theta(\log(n)/n)$ for complete percolation. If a is small enough so that $G(n, p_{\mathsf{c}})$ is dense enough (e.g., if $a \le 0.49\, n/\log^2 n$ when $r = 2$), these two thresholds coincide.

3.2 Gaussian limits

To study the threshold at a_{c} more precisely, we approximate $\pi(t)$ in (2.11) by the corresponding Poisson probability

$$\tilde{\pi}(t) := \mathbb{P}\big(\mathrm{Po}(tp) \ge r\big) = \psi(tp) := \sum_{j=r}^{\infty} \frac{(pt)^j}{j!} e^{-pt}. \tag{3.13}$$

Note that ψ is a differentiable, increasing function on $(0, \infty)$, and that

$$\frac{\mathrm{d}}{\mathrm{d}t} \tilde{\pi}(t) = p\psi'(pt) = p \frac{(pt)^{r-1}}{(r-1)!} e^{-pt} = \frac{p^r t^{r-1}}{(r-1)!} e^{-pt}. \tag{3.14}$$

By a standard estimate for Poisson approximation of a binomial distribution, see e.g. [16, Theorem 2.M],

$$|\pi(t) - \tilde{\pi}(t)| \le d_{\mathrm{TV}}\big(\mathrm{Bin}(t, p), \mathrm{Po}(tp)\big) < p, \tag{3.15}$$

where d_{TV} denotes the total variation distance. A sharper estimate for small t will be given in Lemma 9.4.

We define, for given n and p,

$$a_{\mathsf{c}}^* := -\min_{t \le 3t_{\mathsf{c}}} \frac{n\tilde{\pi}(t) - t}{1 - \tilde{\pi}(t)}, \tag{3.16}$$

and let $t_{\mathsf{c}}^* \in [0, 3t_{\mathsf{c}}]$ be the point where the minimum is attained. Under our standard assumptions $n^{-1} \ll p \ll n^{-1/r}$, for $t \le 3t_{\mathsf{c}}$, when $pt \to 0$ by (3.5), we have by (3.13) and (3.4)

$$n\tilde{\pi}(t) \sim n \frac{(pt)^r}{r!} = \left(\frac{t}{t_{\mathsf{c}}}\right)^r \frac{t_{\mathsf{c}}}{r} \tag{3.17}$$

and thus $\tilde{\pi}(t) \to 0$ and $1 - \tilde{\pi}(t) \sim 1$; it follows easily that $a_{\mathsf{c}}^* \sim a_{\mathsf{c}}$ and $t_{\mathsf{c}}^* \sim t_{\mathsf{c}}$. More precise estimates are given in Lemma 9.5, where it also is shown that t_{c}^* is unique (for large n, at least). Furthermore, by Lemma 9.1 below, for large n, the minimum in (3.16) could as well be taken over $t \le n/2$, say, since $n\tilde{\pi}(t) - t \ge 0$ for $t \in [3t_{\mathsf{c}}, n/2]$.

The following theorem shows that the precise threshold for a is $a_{\mathsf{c}}^* \pm O(\sqrt{a_{\mathsf{c}}})$, with a width of the threshold of the order $\sqrt{a_{\mathsf{c}}} \sim \sqrt{a_{\mathsf{c}}^*}$. Φ denotes the standard normal distribution function. Note that Theorem 3.2 applies, provided $a = o(n)$, and provides more detailed information on A^* when A^* is large (i.e., in (ii) and in (iii) conditioned on, say, $A^* \ge 3t_{\mathsf{c}}$).

Theorem 3.6. *Suppose that $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$.*

(i) *If $(a - a_{\mathsf{c}}^*)/\sqrt{a_{\mathsf{c}}} \to -\infty$, then for every $\varepsilon > 0$, w.h.p. $A^* \leq t_{\mathsf{c}}^* \leq t_{\mathsf{c}}(1+\varepsilon)$. If further $a/a_{\mathsf{c}}^* \to 1$, then $A^* = (1 + o_{\mathrm{p}}(1))t_{\mathsf{c}}$.*

(ii) *If $(a - a_{\mathsf{c}}^*)/\sqrt{a_{\mathsf{c}}} \to +\infty$, then $A^* = n - O_{\mathrm{p}}(b_{\mathsf{c}})$.*

(iii) *If $(a - a_{\mathsf{c}}^*)/\sqrt{a_{\mathsf{c}}} \to y \in (-\infty, \infty)$, then for every $\varepsilon > 0$ and every $b^* \gg b_{\mathsf{c}}$ with $b^* = o(n)$,*

$$\mathbb{P}(A^* > n - b^*) \to \Phi\big((r-1)^{1/2} y\big),$$
$$\mathbb{P}\big(A^* \in [(1-\varepsilon)t_{\mathsf{c}}, (1+\varepsilon)t_{\mathsf{c}}]\big) \to 1 - \Phi\big((r-1)^{1/2} y\big).$$

For the corresponding result when we keep a fixed and change p, we define, for given n and a,

$$\gamma(p) := \inf_{t \leq n/2} \big\{(n-a)\tilde{\pi}(t) - t\big\}. \tag{3.18}$$

Since $\tilde{\pi}$ is an increasing function of p, $\gamma(p)$ is increasing, with $\gamma(0) = -n/2$ and, provided e.g. $a = o(n)$, $\gamma(1) = o(1)$ (attained at $t = o(1)$). Given $a = a(n) \to \infty$ with $a = o(n)$, there is thus (for large n) a unique p_{c}^* such that

$$\gamma(p_{\mathsf{c}}^*) = -a. \tag{3.19}$$

We will see in Lemma 9.2 that $p_{\mathsf{c}}^* \sim p_{\mathsf{c}}$. It is easily verified that, for large n at least, $a_{\mathsf{c}}^* = a \iff \gamma(p) = -a$, and thus $p \mapsto a_{\mathsf{c}}^*$ and $a \mapsto p_{\mathsf{c}}^*$ are the inverses of each other.

Theorem 3.7. *Suppose $r \geq 2$ and $a \to \infty$ with $a = o(n)$.*

(i) *If $(p/p_{\mathsf{c}}^* - 1)a^{1/2} \to -\infty$, then $A^* \leq \big((r/(r-1) + o_{\mathrm{p}}(1)\big)a$. If further $p/p_{\mathsf{c}}^* \to 1$, then $A^* = \big((r/(r-1) + o_{\mathrm{p}}(1)\big)a$.*

(ii) *If $(p/p_{\mathsf{c}}^* - 1)a^{1/2} \to +\infty$, then $A^* = n - o_{\mathrm{p}}(n)$; if further $np - (\log n + (r-1)\log\log n) \to \infty$, then $A^* = n$ w.h.p.*

(iii) *If $(p/p_{\mathsf{c}}^* - 1)a^{1/2} \to \lambda \in (-\infty, \infty)$, then for every $\varepsilon > 0$,*

$$\mathbb{P}(A^* > (1-\varepsilon)n) \to \Phi(r(r-1)^{-1/2}\lambda), \tag{3.20}$$
$$\mathbb{P}(A^* \in [(\tfrac{r}{r-1} - \varepsilon)a, (\tfrac{r}{r-1} + \varepsilon)a] \to 1 - \Phi(r(r-1)^{-1/2}\lambda). \tag{3.21}$$

If further $np - (\log n + (r-1)\log\log n) \to \infty$, then (3.20) can be replaced by $\mathbb{P}(A^* = n) \to \Phi(r(r-1)^{-1/2}\lambda)$.

In the subcritical cases in Theorems 3.1(i) and 3.6(i), we also obtain a Gaussian limit for the size of the final active set.

Theorem 3.8. *Suppose $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Let t_* be the smallest positive root of*

$$(n-a)\tilde{\pi}(t_*) + a - t_* = 0. \tag{3.22}$$

(i) *If* $a/a_c \to \alpha \in (0,1)$, *then* $t_* \sim \varphi(\alpha)t_c$ *with* $\varphi(\alpha) \in (0,1)$ *given by* (3.11), *and* $A^* \in \mathrm{AsN}(t_*, \varphi_2(\alpha)t_c)$, *where* $\varphi_2(\alpha) := \varphi(\alpha)^r(1-\varphi(\alpha)^{r-1})^{-2}/r$.

(ii) *If* $a/a_c \to 1$ *and also* $(a-a_c^*)/\sqrt{a_c} \to -\infty$, *then* $t_* \sim t_c$, *more precisely*

$$t_* = t_c^* - \bigl(1+o(1)\bigr)\sqrt{\frac{2t_c}{r-1}(a_c^*-a)}, \tag{3.23}$$

and

$$A^* \in \mathrm{AsN}\left(t_*, \frac{t_c}{2(r-1)^2(1-a/a_c^*)}\right).$$

Remark 3.9. It follows from the proof that in both cases, for large n at least, t_* is the unique root of (3.22) in $[0, t_c^*]$. In (i), also $t_* < t_c$, so t_* is the unique root in $[0, t_c]$. In (ii), this is not always true. By Lemma 9.5, still for large n, $t_c^* > t_c$ and $t_c^* - t_c \sim pt_c^2/(r-1)$. If, for example, $r = 2$ and $p = \log n/n$, then $t_c^* - t_c \sim n/\log^3 n$, while $a = a_c^* - \sqrt{n}$ yields $t_c^* - t_* \sim \sqrt{2}n^{3/4}/\log n \ll t_c^* - t_c$.

3.3 The number of generations

In the supercritical case $a - a_c^* \gg \sqrt{a_c}$, when $\mathcal{A}(0)$ w.h.p. almost percolates by Theorem 3.6, we have the following asymptotic formula for the number of generations until the bootstrap percolation process stops.

Theorem 3.10. *Suppose that* $r \ge 2$, $n^{-1} \ll p \ll n^{-1/r}$ *and* $a = o(n)$. *Assume than* $a - a_c^* \gg \sqrt{a_c}$ *(so that* $\mathcal{A}(0)$ *w.h.p. almost percolates). Then, w.h.p.,*

$$\tau \sim \frac{\pi\sqrt{2}}{\sqrt{r-1}}\Bigl(\frac{t_c}{a-a_c^*}\Bigr)^{1/2} + \frac{1}{\log r}\left(\log\log(np) - \log_+ \log\frac{a}{a_c}\right) + \frac{\log n}{np} + O_p(1). \tag{3.24}$$

This theorem is an immediate consequence of Propositions 10.1, 10.4, 10.7 and 10.10 in Section 10. Moreover, these propositions show that the three terms (excepting the error term $O_p(1)$) in the formula (3.24) are the numbers of generations required for three distinct phases of the evolution: the beginning including (possibly) a bottleneck when the size is about t_c; a period of doubly exponential growth; and a final phase where the last vertices are activated. Note that each of the three terms may be the largest one.

Example 3.11. Let $p = n^{-\alpha}$, with $1/r < \alpha < 1$, and suppose $a = O(a_c)$. Then the third term in (3.24) is $O(1)$ and can be ignored while the second term is $\log\log n/\log r + O(1)$. If we are safely supercritical, say $a = 2a_c$, then the first term too is $O(1)$ and the result is $\tau \sim \log\log n/\log r$ w.h.p., dominated by the second term.

If instead the process is only barely supercritical, with $a = a_c^* + a_c^\beta$ say, with $1/2 < \beta < 1$, then the first term in (3.24) is Cn^γ with $C > 0$ and the exponent $\gamma = \frac{1-\beta}{2} \cdot \frac{r\alpha-1}{r-1}$, which dominates the other terms. Note that the exponent here can be any positive number in $(0, 1/4)$ (with $\gamma \approx 1/4$ if $\alpha \approx 1$ and $\beta \approx 1/2$, so the graph is very sparse and the initial set is minimal).

Finally, if $p = \log\log n/n$, say, so the graph is very sparse, and $a = 2a_c$, then again the first term in (3.24) is $O(1)$, the second is $O(\log\log\log\log n)$, while the third is $\log n/\log\log n$, which thus dominates the sum.

Note that the second term is $O(\log\log n)$ and the third is $o(\log n)$ (and in many cases $O(1)$ so it can be ignored), while the first term may be as large as $n^{1/4-o(1)}$ (although it too in many cases is $O(1)$).

Remark 3.12. In the subcritical case, one could presumably obtain similar results for the number of generations until the process stops, but we have not pursued this topic here.

4 Dynamical models

We usually assume, as above, that a and p are given, but we can also consider dynamical models where one of them grows with time.

4.1 Adding external activations

In the first dynamical model, we let n and p be given, and consider a realization of $G_{n,p}$. We start with all vertices inactive (and completely uninfected). We then activate the vertices (from the outside) one by one, in random order. After each external activation, the bootstrap percolation mechanism works as before, activating all vertices that have at least r active neighbours until no such vertices remain; this is done instantaneously (or very rapidly) so that this is completed before the next external activation. Let A_0 be the number of externally activated vertices the first time that the active set $\mathcal{A}$ is 'big' in some sense. For example, for definiteness, we may define 'big' as $|\mathcal{A}| > n/2$. (It follows from Theorem 3.1 that any threshold $|\mathcal{A}| > cn$ for a constant $c \in (0,1)$ will give the same asymptotic results, as well as thresholds tending to 0 or n sufficiently slowly. If $np - (\log n + (r-1)\log\log n) \to \infty$, we may also choose the condition $|\mathcal{A}| = n$, i.e., complete percolation $\mathcal{A} = V_n$.) Then A_0 is a random variable (depending both on the realization of $G_{n,p}$ and on the order of external activations). In this formulation, the threshold result in Theorem 3.1 may be stated as follows.

Theorem 4.1. *Suppose that $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Then $A_0/a_\mathsf{c} \xrightarrow{\mathrm{p}} 1$.*

Proof. The active set after a external activations is the same as the final active set $\mathcal{A}(T)$ in the static model considered in the rest of this paper with these vertices chosen to be active initially. Hence, for any given a, $A_0 \leq a$ if and only if bootstrap percolation with a initially active yields a big final active set. In particular, if $\delta > 0$, then Theorem 3.1(i) implies that $\mathbb{P}(A_0 \leq (1-\delta)a_\mathsf{c}) \to 0$, while Theorem 3.1(ii)(iii) imply that $\mathbb{P}(A_0 \leq (1+\delta)a_\mathsf{c}) \to 1$. ☐

More precisely, Theorem 3.6 yields a Gaussian limit.

Theorem 4.2. *Suppose that $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Then $A_0 \in \mathrm{AsN}\big(a_\mathsf{c}^*, a_\mathsf{c}/(r-1)\big)$.*

Proof. Let $x \in (-\infty, \infty)$. Then, arguing as in the proof of Theorem 4.1 but now using Theorem 3.6(iii) (with $y = x/\sqrt{r-1}$), we find

$$\mathbb{P}\Big(\frac{A_0 - a_\mathsf{c}^*}{\sqrt{a_\mathsf{c}/(r-1)}} \leq x\Big) = \mathbb{P}\big(A_0 \leq a_\mathsf{c}^* + x\sqrt{a_\mathsf{c}/(r-1)}\big) \to \Phi(x). \qquad \square$$

We have here for simplicity assumed that the external activations are done by sampling without replacement, but otherwise independently of whether the vertices already are (internally) activated. A natural variation is to only activate vertices that are inactive. Let A_0' be the number of externally activated vertices when the active set becomes big in this version. Since a new activation of an already active vertex does not matter at all, A_0' equals in the version above the number of externally active vertices among the first A_0 that are not already internally activated. Thus $A_0 - A_0'$ is the number of external activations that hit an already active vertex. It is easily verified that this is $o_p(a_c)$, and thus Theorem 4.1 holds for A_0' as well; we omit the details. It seems likely that it is possible to derive a version of the Gaussian limit in Theorem 4.2 for A_0' too, but that would require a more careful estimate of $A_0 - A_0'$ (and in particular its variance), which we have not done, so we leave this possibility as an open problem.

Remark 4.3. One way to think about this dynamical model, where we add new active vertices successively and may think of these as being initially active, is to see it as a sequence of bootstrap percolation processes, one for each $a = 0, 1, \dots, n$; the processes live on the same graph $G_{n,p}$ but have different numbers of initially active vertices, and they are coupled in a natural way. In order to really have the same realization of $G_{n,p}$ for different a, we have to be careful in the choice of the order in which we explore the vertex neighbourhoods, i.e., the choice of u_t. (Recall that $G_{n,p}$ is constructed from the indicators $I_i(s)$ and the sequence (u_t), see Section 2.) We can achieve this by first making a list L of all vertices in the (random) order in which they are externally activated. We then at each time t choose u_t as an unused internally activated vertex (for example the most recent one) if there is any such vertex, and otherwise as the next unused vertex in the list L.

This model makes it possible to pose now questions about the bootstrap percolation. For example, we may consider the critical process starting with exactly A_0 initially active vertices (i.e., the first process that grows beyond the bottleneck and becomes big) and ask for the number of generations until the process dies out. Alternatively, we may consider the process starting with exactly $A_0 - 1$ initially active vertices (i.e., the last process that does not become big) and ask for its final size.

Such questions will not be treated in the present paper, but we mention that it is easily seen that the final size with $A_0 - 1$ initially active vertices is $t_c(1 + o_p(1))$ so that the final size jumps from about t_c to about n with the addition of a single additional initial vertex. Furthermore, we conjecture that, under suitable conditions, the number of generations for the process with A_0 initially active vertices is of order $a_c^{1/3}$ (which is much larger than the number of generations for any fixed a, see Subsection 3.3).

4.2 Adding external infections

An alternative to external activations is external infections, where we again start with all vertices inactive and uninfected, and infect vertices one by one from the outside, choosing the infected vertices at random (independently and with replacement); as before, r infections (external or internal) are needed for activation, and active vertices infect their neighbours. Let J_0 be the number of external infections when the active set first becomes 'big' (as in Subsection 4.1). (Thus, J_0 is a random variable.)

In the original model, each initially active vertex infects about np other vertices so the total number of initial infections is about npa; it is thus easy to guess that $J_0 \approx npa_c$. Indeed, this is the case as is shown by the next theorem. We cannot (as far as we know) directly derive this from our previous results, since the dependencies between infections in the two versions are slightly different, but it follows by a minor variation of our method, see Section 8. We believe that the result could be sharpened to a Gaussian limit as in Theorem 4.2, but we leave this to the reader.

Theorem 4.4. *Suppose that $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Then $J_0/(npa_\mathsf{c}) \xrightarrow{\mathrm{p}} 1$.*

In particular, for $r = 2$, we thus have $J_0 \overset{\mathrm{p}}{\sim} npa_\mathsf{c} = 1/(2p)$.

4.3 Adding edges

In the second dynamical model, n and a are given; we start with n vertices of which a are active, but no edges. We then add the edges of the complete graph K_n one by one, in random order. As in the previous dynamical model, bootstrap percolation takes place instantaneously after each new edge is added.

It is convenient to use the standard method of adding the edges at random times (as in, e.g., [32]). Thus, each edge e in K_n is added at a time U_e, where U_e are independent and uniformly distributed on $[0, 1]$. Then, at a time $u \in [0, 1]$, the resulting graph is $G_{n,p}$ with $p = u$. (We use u to denote this time variable, in order not to confuse it with the time t used to describe the bootstrap percolation process.)

Let the random variable M be the number of edges required to obtain a big active set $\mathcal{A}$, where 'big' is defined as in Subsection 4.1.

Theorem 4.5. *Suppose $r \geq 2$ and $a \to \infty$ with $a = o(n)$. Then $M = \binom{n}{2}p_\mathsf{c}(1 + o_\mathrm{p}(1)) = \frac{1}{2}n^2 p_\mathsf{c}(1 + o_\mathrm{p}(1))$. More precisely,*

$$M \in \mathrm{AsN}\left(\binom{n}{2}p_\mathsf{c}^*, \frac{r-1}{4r^2}\frac{(n^2p_\mathsf{c})^2}{a}\right).$$

The proof is given in Section 9.

Remark 4.6 (Coupling different p)**.** The proof of Theorem 4.5 is based on using our earlier results for a single p. We might also want to study the bootstrap percolation process for all p at once (or equivalently, in $G(n, m)$ for all m at once), i.e. with a coupling of the models for different p, for given n and a. As in Remark 4.3, this requires a careful choice of the order in which the vertices are inspected. We can achieve this by modifying the formulation in Section 2 as follows:

When we have chosen a vertex u_t, we reveal the times U_{tj} that the edges from it appear; this tells us the neigbourhood of u_t at any time u. We begin by choosing $u_1, \dots, u_a$ as the initially active vertices. We then, after each choice of u_t, $t \geq a$, calculate for each of the remaining $n - t$ vertices the time when it acquires the r:th edge to $\{u_1, \dots, u_t\}$, and let u_{t+1} be the vertex such that this time is minimal. Then, fixing any time $u = p$, the chosen vertices u_t will all be active until the first time that no unused active vertices remain and the process stops. In this manner, we have found a choice of $u_1, u_2, \dots$ that satisfies the description in Section 2 for all $p \in [0, 1]$ simultaneously.

As in Remark 4.3, we can use this model to study for example the last 'small' or the first 'big' bootstrap percolation process, when we add edges one by one with a given set of initially active vertices. Again, we will not consider such questions in the present paper.

5 Boundary cases

We have above assumed $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$. In this section we treat the cases when these assumptions do not hold. Proofs are given in Section 11.

We begin with the sparse case $p(n) \sim c/n$ when t_{c} and a_{c} defined by (3.1)–(3.2) are of order n. (The exact values are no longer relevant, since they are based on approximations no longer valid.) This suggests that the interesting case is when $a \asymp n$, i.e., when a positive fraction of all vertices are initially active. Indeed, Theorem 5.2 shows that, w.h.p., if we start with a positive fraction of the graph, then the activation spreads to a larger part of the graph but does not reach almost all vertices; if, on the contrary, the size of the original set of activated vertices is negligible with respect to the size of the graph, then the activation does not spread to a positive fraction of the graph. Provided c is large enough, there is, as found by Scalia-Tomba [42], also in this case a dichotomy, or 'phase transition', similar to Theorem 3.1, with a sudden jump from a 'small' to a 'large' final active set, although in this case all sets are of order n so the jump is not as dramatic as for larger p.

Define, for $x \geq 0$, $c \geq 0$ and $\theta \in [0, 1]$,

$$f(x, c, \theta) := (1 - \theta)\,\mathbb{P}(\mathrm{Po}(cx) \geq r) + \theta - x = (1 - \theta) \sum_{j=r}^{\infty} \frac{(cx)^j}{j!} e^{-cx} - x + \theta \tag{5.1}$$

$$= 1 - x - (1 - \theta)\,\mathbb{P}(\mathrm{Po}(cx) \leq r - 1) = 1 - x - (1 - \theta) \sum_{j=0}^{r-1} \frac{(cx)^j}{j!} e^{-cx}, \tag{5.2}$$

and let $x_0(\theta)$ be the smallest root $x \geq 0$ of

$$f(x, c, \theta) = 0; \tag{5.3}$$

similarly, let $x_1(\theta)$ be the largest root in $[0, 1]$ of this equation. Since $f(0, c, \theta) = \theta \geq 0$ and $f(1, c, \theta) = -(1-\theta)\,\mathbb{P}(\mathrm{Po}(c) \leq r-1) \leq 0$, there is always at least one root in $[0, 1]$, and $0 \leq x_0(\theta) \leq x_1(\theta) \leq 1$; further $0 < x_0(\theta) \leq x_1(\theta) < 1$ when $0 < \theta < 1$ while $x_0(0) = 0$ and $x_0(1) = x_1(1) = 1$. We also define

$$c_{\mathsf{c}} = c_{\mathsf{c}}(r) := r + \frac{\mathbb{P}\big(\mathrm{Po}(r - 1) \leq r - 2\big)}{\mathbb{P}\big(\mathrm{Po}(r - 1) = r - 1\big)} = r + \frac{\sum_{j=0}^{r-2} (r - 1)^j / j!}{(r - 1)^{r-1} / (r - 1)!}. \tag{5.4}$$

Thus $c_{\mathsf{c}}(2) = 3$, $c_{\mathsf{c}}(3) = 9/2$, $c_{\mathsf{c}}(4) = 53/9$.

Lemma 5.1. (i) *If $0 \leq c \leq c_{\mathsf{c}}$, then (5.3) has a unique root $x = x_0(\theta) \in [0, 1]$ for every $\theta \in [0, 1]$, and $x_0(\theta)$ is a continuous strictly increasing function of θ.*

(ii) *If $c > c_{\mathsf{c}}$, then there exists $\theta_{\mathsf{c}}^- = \theta_{\mathsf{c}}^-(c)$ and $\theta_{\mathsf{c}} = \theta_{\mathsf{c}}(c)$ with $0 \leq \theta_{\mathsf{c}}^- < \theta_{\mathsf{c}} < 1$ such that (5.3) has three roots in $[0, 1]$ when $\theta \in (\theta_{\mathsf{c}}^-, \theta_{\mathsf{c}})$, but a unique root when $\theta \in [0, \theta_{\mathsf{c}}^-)$ or $\theta \in (\theta_{\mathsf{c}}, 1]$; if $\theta = \theta_{\mathsf{c}}^- > 0$ or $\theta = \theta_{\mathsf{c}}$, there are two roots, one of them double. The smallest root $x_0(\theta)$ is strictly increasing and continuous on $[0, 1]$ except at θ_{c} where it has a jump from $x_0(\theta_{\mathsf{c}})$ to $x_1(\theta_{\mathsf{c}}) > x_0(\theta_{\mathsf{c}})$, where $x_1(\theta_{\mathsf{c}}) = x_0(\theta_{\mathsf{c}}+) := \lim_{\theta \searrow \theta_{\mathsf{c}}} x_0(\theta)$ is the other root for $\theta = \theta_{\mathsf{c}}$. Furthermore, if $\theta = \theta_{\mathsf{c}}$, then $f(x, c, \theta) \geq 0$ for $x \in [0, x_1(\theta)]$, and $x_0(\theta)$ is a double root.*

Theorem 5.2. *Suppose that $r \geq 2$, $p \sim c/n$ and $a \sim \theta n$ for some constants $c \geq 0$ and $\theta \geq 0$.*

(i) *If $\theta = 0$, i.e., if $a = o(n)$, then $A^*/a \xrightarrow{\mathrm{p}} 1$.*

(ii) *If $c = 0$, i.e., if $p = o(1/n)$, then $A^*/a \xrightarrow{\mathrm{p}} 1$.*

(iii) *If $0 \leq c \leq c_{\mathsf{c}}$, then $A^*/n \xrightarrow{\mathrm{p}} x_0(\theta)$, where $x_0(\theta)$ is the unique non-negative root of* (5.3).

(iv) *If $c > c_{\mathsf{c}}$ and $\theta \neq \theta_{\mathsf{c}}(c)$ given by Lemma 5.1, then $A^*/n \xrightarrow{\mathrm{p}} x_0(\theta)$, where $x_0(\theta)$ is the smallest non-negative root of* (5.3).

There is thus a jump in the final size at $a = \theta_{\mathsf{c}} n$. Remark 11.1 shows how to find θ_{c}.

Remark 5.3. $\theta_{\mathsf{c}}(c)$ and $\theta_{\mathsf{c}}^-(c)$ are decreasing functions of c. ($\theta_{\mathsf{c}}(c)$ is strictly decreasing, while $\theta_{\mathsf{c}}^-(c)$ is constant 0 for large c.) Hence their largest value is, by the calculation in Remark 11.1,

$$\theta_{\mathsf{c}}^* = \theta_{\mathsf{c}}^*(r) := \theta_{\mathsf{c}}(c_{\mathsf{c}}) = \theta_{\mathsf{c}}^-(c_{\mathsf{c}}) = 1 - \frac{1}{r\,\mathbb{P}(\mathrm{Po}(r-1) = r-1) + \mathbb{P}(\mathrm{Po}(r-1) \leq r-2)}.$$

Thus $\theta_{\mathsf{c}}^*(2) = 1 - e/3$, $\theta_{\mathsf{c}}^*(3) = 1 - e^2/9$, $\theta_{\mathsf{c}}^*(4) = 1 - 2e^3/53$.

The threshold for $\theta_{\mathsf{c}}^- = 0$ can be calculated too. For $r = 2$, $\theta_{\mathsf{c}}^-(c) = 0$ for $c \geq e^y/y$, where $e^y = 1 + y + y^2$; numerically, this is $c \geq 3.35091\dots$.

Remark 5.4. We have here considered a given $p \sim c/n$ and varied $a \sim \theta n$. If we instead, as in Theorem 3.4, take a given $a \sim \theta n$ for a fixed θ and vary $c = pn$, we have a similar phenomenon. Lemma 5.1 and Theorem 5.2 apply for every combination of θ and c, and by considering the set of $(c, \theta) \in \mathbb{R}_+^2$ such that (5.3) has two or three roots, it follows from Remark 5.3 that if $\theta \geq \theta_{\mathsf{c}}^*$, then $A^*/n \xrightarrow{\mathrm{p}} x_0(c)$, where $x_0(c)$ is the unique root of (5.3) and thus a continuous function of c, while if $\theta < \theta_{\mathsf{c}}^*$, then there is a range of c where (5.3) has three roots, and one value of c where the limit value $x_0(c)$ jumps from a 'small' to a 'large' value. Thus there is, again, a kind of phase transition.

The following theorem shows that if we for simplicity take $p = c/n$, then the precise threshold for a in Theorem 5.2(iv) is $\theta_{\mathsf{c}} n \pm O(\sqrt{n})$, with a width of the threshold of the order $\sqrt{n}$.

Theorem 5.5. *Suppose that $r \geq 2$ and $p = \frac{c}{n}$ with $c > c_{\mathsf{c}}$ fixed. Let $\theta_{\mathsf{c}} = \theta_{\mathsf{c}}(c)$, $x_0 = x_0(\theta_{\mathsf{c}})$ and $x_1 = x_1(\theta_{\mathsf{c}})$ be as in Lemma 5.1; thus x_0 and x_1 are the two roots in $[0,1]$ of $f(x, c, \theta_{\mathsf{c}}) = 0$, with $x_0 < x_1$.*

(i) *If $(a - \theta_{\mathsf{c}} n)/\sqrt{n} \to -\infty$, then for any $\varepsilon > 0$, w.h.p. $A^* \leq (1+\varepsilon)x_0 n$. If further $a \sim \theta_{\mathsf{c}} n$, then $A^* = (1 + o_{\mathrm{p}}(1))x_0 n$.*

(ii) *If $(a - \theta_{\mathsf{c}} n)/\sqrt{n} \to +\infty$, then for any $\varepsilon > 0$, w.h.p. $A^* \geq (1-\varepsilon)x_1 n$. If further $a \sim \theta_{\mathsf{c}} n$, then $A^* = (1 + o_{\mathrm{p}}(1))x_1 n$.*

(iii) *If $(a - \theta_{\mathsf{c}} n)/\sqrt{n} \to y \in (-\infty, \infty)$, then there exists a sequence $\varepsilon_n \to 0$ such that*

$$\mathbb{P}\big(A^* \in [(1-\varepsilon_n)x_1 n, (1+\varepsilon_n)x_1 n]\big) \to \Phi\big(y/\sigma\big),$$
$$\mathbb{P}\big(A^* \in [(1-\varepsilon_n)x_0 n, (1+\varepsilon_n)x_0 n]\big) \to 1 - \Phi\big(y/\sigma\big),$$

where $\sigma^2 = (1-\theta_{\mathsf{c}})\psi(cx_0)/(1 - \psi(cx_0)) > 0$.

At the other, dense, endpoint of our range we have $p(n) \sim cn^{-1/r}$. Then t_c and a_c in (3.1)–(3.2) are of order constant. (Again the exact values are irrelevant.) This suggests, and the following theorem makes more precise, that the process will either die out or grow very quickly, with the outcome determined by the first few steps, and that the activation can spread from a set of constant size to the entire graph with a positive probability, which however is bounded away from 1.

Theorem 5.6. *Suppose $r \geq 2$ and $p \sim cn^{-1/r}$ for a constant $c > 0$.*

(i) *If a is fixed with $a \geq r$, then*
$$\mathbb{P}(A^* = n) \to \zeta(a, c) \tag{5.5}$$
for some $\zeta(a,c) \in (0,1)$. Furthermore, there exist numbers $\zeta(a,c,k) > 0$ for $k \geq a$ such that $\mathbb{P}(A^* = k) \to \zeta(a,c,k)$ for each fixed $k \geq a$, and $\sum_{k=a}^{\infty} \zeta(a,c,k) + \zeta(a,c) = 1$.

(ii) *If $a \to \infty$, then $\mathbb{P}(A^* = n) \to 1$, i.e., $A^* = n$ w.h.p.*

Remark 5.7. The limiting probabilities in Theorem 5.6 can be expressed as hitting probabilities of an inhomogeneous random walk. Let $\xi_k \in \mathrm{Po}\big(\binom{k-1}{r-1} c^r\big)$, $k \geq 1$, be independent, and let $\widetilde{S}_k := \sum_{j=1}^{k} (\xi_j - 1)$ and $\widetilde{T} := \min\{k : a + \widetilde{S}_k = 0\} \in \mathbb{N} \cup \{\infty\}$. Then
$$\zeta(a,c) = \mathbb{P}(\widetilde{T} = \infty) = \mathbb{P}(a + \widetilde{S}_k \geq 1 \text{ for all } k \geq 1) \tag{5.6}$$
and $\zeta(a,c,k) = \mathbb{P}(\widetilde{T} = k)$. Consequently, Theorem 5.6(i) can also be written as
$$d_{\mathrm{TV}}\big(A^*, \min(\widetilde{T}, n)\big) \to 0,$$
where d_{TV} is the total variation distance.

If the probability of connections p is even larger, $p \gg n^{-1/r}$, then the initial set percolates as long as $a \geq r$.

Theorem 5.8. *Let $r \geq 2$. If $p \gg n^{-1/r}$ and $a \geq r$, then $A^* = n$ w.h.p.*

Remark 5.9. The case $r = 1$ is different. In this case, infection is equivalent to activation, and spreads to every vertex connected to an active vertex. Thus the final active set is the union of the components of the graph that contain at least one initially active vertex. It is well-known that this is equivalent to the Reed–Frost model for epidemics, where each infected person infects everyone else with probability p, all infections being independent. (This equivalence is easily seen by the argument in Section 2.) The Reed–Frost model has been much studied, see for example von Bahr and Martin-Löf [3], Martin-Löf [38, 39]. We state some known result for comparison with our results for $r \geq 2$; proofs can be found in [3, 38], where also further details are given (including central limit theorems as in Subsection 3.2), or by modifying the proofs of the results above. Many results follow also easily from known results on the component structure of $G_{n,p}$.

If $p = \log n/n + \omega(n)/n$, with $\omega(n) \to \infty$, then w.h.p. $G_{n,p}$ is connected and thus $A^* = n$ as soon as $a \neq 0$. More generally, if $p \gg n^{-1}$ and $a \geq 1$, then w.h.p. $A^* = n - o(n)$. (Cf. Theorem 5.8.)

The case $p = c/n$ is perhaps more interesting. There are many (w.h.p. $\geq c'n$) isolated vertices, so we cannot have percolation or almost percolation unless $a/n \to 1$. If $c > 1$, there is a single giant component of size $\rho n + o_{\mathrm{p}}(n)$, with $\rho = \rho(c) > 0$, and thus, if $a \geq 1$ is fixed, then there is a dichotomy, with either $A^* = o(n)$ or $A^* = \rho n + o(n)$ w.h.p., with probabilities converging to the positive $(1-\rho)^a$ and $1-(1-\rho)^a$, respectively. (Cf. Theorem 5.6.)

If $c \leq 1$ and a is fixed, then, by the same argument, A^* converges to the total size of a Galton–Watson process with $\mathrm{Po}(c)$ offspring distribution and a initial individuals (a Borel–Tanner distribution). Thus A^*/a is stochastically bounded but does not converge in probability to a constant. (Cf. Theorem 5.2(i).)

If $c < 1$ and $a \to \infty$ but $a = o(n)$, then $A^*/a \xrightarrow{\mathrm{p}} 1/(1-c)$. (Cf. Theorem 5.2(i).)

If $p \sim c/n$ with any $c > 0$ and $a \sim \theta n$ with $\theta > 0$, then $A^*/n \xrightarrow{\mathrm{p}} x_0(\theta)$, where x_0 is the unique positive root of (5.3), where now $f(x, c, \theta) = 1 - x - (1-\theta)e^{-cx}$ by (5.2) with $r = 1$. In other words, Theorem 5.2(iii) holds for $r = 1$ too with $c_{\mathsf{c}} = \infty$, and there is no threshold. (For $\theta = 0$ there is the well-known threshold at $c = 1$, but note that also in this case, x_0 is continuous in both c and θ and there is no jump as in Theorem 5.2(iv).)

6 Overview of the proofs

By (2.3), (2.2) and (2.9), for $u = 1, 2, \dots$,

$$A^* \geq u \iff T \geq u \iff \min_{t<u}\bigl(A(t) - t\bigr) > 0 \iff a + \min_{t<u}\bigl(S(t) - t\bigr) > 0. \tag{6.1}$$

Hence, $A^* = T$ is completely determined by the stochastic process $\min_{t<u}\bigl(S(t) - t\bigr)$, $u \geq 1$. In particular, $\mathcal{A}(0)$ percolates if and only if $a > -\min_{t<n}\bigl(S(t) - t\bigr) = \max_{t<n}\bigl(t - S(t)\bigr)$.

Note that (6.1) is an exact representation of A^*; we have not yet made any approximations. To obtain asymptotic results, we introduce some simple approximations. We give an informal overview of the argument here; details will follow in later sections.

First, $S(t) \approx \mathbb{E}\, S(t)$ by the law of large numbers. A simple calculation will show that $f(t) := \mathbb{E}\, S(t) - t$ starts at 0 for $t = 0$, then decreases to a minimum at $t \approx t_{\mathsf{c}}$ given in (3.1), and then increases until $\mathbb{E}\, S(t) \approx n$ and thus $f(t) \approx n - t$; then $f(t) \approx n - t$ holds until $t = n$, so $f(t)$ decreases again in this range to a final value $f(n) = \mathbb{E}\, S(n) - n \approx 0$.

There are thus two candidates for the minimum point of $S(t) - t$: either $t \approx t_{\mathsf{c}}$ or $t \approx n$. What happens at $t \approx n$ makes the difference between almost percolation and complete percolation; we will study this too in detail later, but for the moment we ignore it and concentrate on whether we have almost percolation or not, and we see that, roughly, this is determined by whether $a > -(S(t_{\mathsf{c}}) - t_{\mathsf{c}})$ or not, which can be approximated by $a > -(\mathbb{E}\, S(t_{\mathsf{c}}) - t_{\mathsf{c}})$. A simple calculation yields $\mathbb{E}\, S(t_{\mathsf{c}}) - t_{\mathsf{c}} \approx -a_{\mathsf{c}}$, which establishes the threshold at a_{c}.

This argument also gives the following picture of the course of the activation $\mathcal{A}(t)$ in the critical case $a \approx a_{\mathsf{c}}$. (We leave the modifications in the subcritical and supercritical cases to the reader.) We start with $A(0) = a$. At first, there are very few new vertices that reach the threshold of r infections, and the number $A(t) - t$ of unused vertices goes down, and approaches 0 as t approaches t_{c}. However, the rate of activation of new vertices is increasing, because a pool of vertices with $r - 1$ infections has accumulated, and as $t \to t_{\mathsf{c}}$, new vertices are activated at about the same rate as they are used. There are now two possibilities: either the activation dies out at this point, with a total size about $t_{\mathsf{c}} = r/(r-1)a_c$, or it survives this bottleneck, and it then rapidly grows after time t_c until almost all vertices are active. In the latter case there are again two possibilities: either all remaining vertices are finally active (complete percolation), or a few are not.

7 Approximation of $S(t)$ by its mean

For notational convenience, we assume that $V_n \setminus \mathcal{A}(0) = \{1, \dots, n-a\}$. Note first that $(n-a)^{-1}S(t) = (n-a)^{-1}\sum_{i=1}^{n-a} \mathbf{1}\{Y_i \le t\}$ is the empirical distribution function of $\{Y_i\}_1^{n-a}$. By the law of large numbers for the binomial distribution (2.10), for every $t = t(n)$, $S(t) = (n-a)\pi(t) + o_{\mathrm{p}}(n)$. Moreover, by the Glivenko–Cantelli theorem [36, Proposition 4.24], this holds uniformly for all t:

Lemma 7.1. $\sup_{t\ge 0} |S(t) - \mathbb{E}\, S(t)| = o_{\mathrm{p}}(n)$.

Proof. If $n - a \ge \sqrt{n}$, say, this is a weaker version of [36, Proposition 4.24]. For smaller $n - a$, the result is trivial, since $0 \le S(t) \le n - a$. □

For small t, the uniform error bound in Lemma 7.1 is not good enough. (It can be improved to $O_{\mathrm{p}}(n^{1/2})$, see Lemma 7.3, but this too is too large for our purposes.) For each t, (2.13) gives a bound $O_{\mathrm{p}}((n\pi(t))^{1/2})$. We extend this to a uniform bound for a range of t by a martingale argument. We begin by introducing a pair of well-known martingales for empirical distribution functions. (See [32, Lemma 2.1] for a continuous time version.)

Lemma 7.2. *The stochastic process*

$$\frac{S(t) - \mathbb{E}\, S(t)}{1 - \pi(t)}, \qquad t = 0, 1, \dots, \tag{7.1}$$

is a martingale, and

$$\frac{S(t) - \mathbb{E}\, S(t)}{\pi(t)}, \qquad t = r, r+1, \dots, \tag{7.2}$$

is a reverse martingale.

Proof. Since $S(t)$ is a sum of $n - a$ i.i.d. processes $\mathbf{1}\{Y_i \le t\}$, it suffices to treat each of these separately, i.e., for the first part to show that, for each i,

$$X(t) = X_i(t) := \frac{\mathbf{1}\{Y_i \le t\} - \mathbb{P}(Y_i \le t)}{1 - \mathbb{P}(Y_i \le t)} = 1 - \frac{\mathbf{1}\{Y_i > t\}}{\mathbb{P}(Y_i > t)}$$

is a martingale. This is elementary: If $Y_i \le t$, then $X(t) = X(t+1) = 1$. If $Y_i > t$, then $X(t) = -\pi(t)/(1-\pi(t))$ either jumps to $X(t+1) = 1$ or decreases to $X(t+1) = -\pi(t+1)/(1-\pi(t+1))$, and the conditional probability of these events are $(\pi(t+1)-\pi(t))/(1-\pi(t))$ and $(1-\pi(t+1))/(1-\pi(t))$, respectively, so a simple calculation yields $\mathbb{E}\big(X(t+1) \mid Y_i > t\big) = -\pi(t)/(1-\pi(t))$. (Alternatively, this follows from the case $X(t) = 1$ and the fact that $\mathbb{E}\, X(t+1) = \mathbb{E}\, X(t)$.) Hence, $\mathbb{E}\big(X(t+1) \mid X(1), \dots, X(t)\big) = X(t)$.

For the second part, we similarly find that $\tilde{X}(t) := \mathbf{1}\{Y_i \le t\}/\pi(t)$ is a reverse martingale, i.e., that $\mathbb{E}(\tilde{X}(t) \mid \tilde{X}(t+1), \dots) = \tilde{X}(t+1)$. □

Lemma 7.3. *For any t_0,*

$$\mathbb{E}\Big(\sup_{t\le t_0}|S(t)-\mathbb{E}\,S(t)|\Big)^2\le 16n\pi(t_0), \tag{7.3}$$

$$\mathbb{E}\Big(\sup_{t\ge t_0}|S(t)-\mathbb{E}\,S(t)|\Big)^2\le 16n(1-\pi(t_0)). \tag{7.4}$$

Proof. Assume first $\pi(t_0)\le 1/2$. Let $\xi(t):=\big(S(t)-\mathbb{E}\,S(t)\big)/(1-\pi(t))$. This is a martingale by Lemma 7.2, and Doob's inequality [36, Proposition 7.16] yields, using (2.13),

$$\mathbb{E}\Big(\sup_{t\le t_0}|S(t)-\mathbb{E}\,S(t)|\Big)^2\le \mathbb{E}\sup_{t\le t_0}|\xi(t)|^2\le 4\,\mathbb{E}\,|\xi(t_0)|^2=4\frac{\operatorname{Var}S(t_0)}{(1-\pi(t_0))^2}\le 8n\pi(t_0), \tag{7.5}$$

which proves (7.3) in this case. Similarly, if $\pi(t_0)\ge 1/2$, then we obtain using the reverse martingale (7.2)

$$\mathbb{E}\Big(\sup_{t\ge t_0}|S(t)-\mathbb{E}\,S(t)|\Big)^2\le 4\frac{\operatorname{Var}S(t_0)}{\pi(t_0)^2}\le 8n(1-\pi(t_0)). \tag{7.6}$$

Now, let t_1 be the largest integer such that $\pi(t_1)\le 1/2$. We can apply (7.5) with $t_0=t_1$ and (7.6) with $t_0=t_1+1$, and thus

$$\mathbb{E}\Big(\sup_{t\ge 0}|S(t)-\mathbb{E}\,S(t)|\Big)^2\le \mathbb{E}\Big(\sup_{t\le t_1}|S(t)-\mathbb{E}\,S(t)|\Big)^2+\mathbb{E}\Big(\sup_{t\ge t_1+1}|S(t)-\mathbb{E}\,S(t)|\Big)^2\le 8n.$$

This immediately implies (7.3) for $\pi(t_0)>1/2$ and (7.4) for $\pi(t_0)<1/2$. $\square$

8 Approximation of $\mathbb{E}\,S(t)$ and proofs of Theorems 3.1–3.4

For (real) $t>0$ and $pt\le 1$, say, by (2.11),

$$\pi(t)=\sum_{j=r}^{\lfloor t\rfloor}\binom{\lfloor t\rfloor}{j}p^j(1-p)^{\lfloor t\rfloor-j}=\binom{\lfloor t\rfloor}{r}p^r\big(1+O(pt)\big)=\frac{t^rp^r}{r!}\big(1+O(pt+t^{-1})\big) \tag{8.1}$$

cf. (3.17), and thus, by (2.12),

$$\mathbb{E}\,S(t)-n\frac{t^rp^r}{r!}=O\big(nt^rp^r(pt+t^{-1}+a/n)\big). \tag{8.2}$$

It thus makes sense to approximate $f(t):=\mathbb{E}\,S(t)-t$ by $\bar f(t):=n(tp)^r/r!-t$. An elementary calculation shows that $\bar f$ has, on $[0,\infty)$ a unique, global minimum at t_{c} given by (3.1), and that the minimum value is $\bar f(t_{\mathsf{c}})=-a_{\mathsf{c}}$. We obtain, for example, the following estimate.

Lemma 8.1. *Suppose that $r\ge 2$, $n^{-1}\ll p\ll n^{-1/r}$ and $a=o(n)$. Then*

$$\sup_{0\le x\le 10r}\big|S(xt_{\mathsf{c}})-\tfrac1r x^rt_{\mathsf{c}}\big|=o_{\mathrm{p}}(t_{\mathsf{c}}).$$

Proof. First, (8.2) and (3.4) yield, recalling (3.5), uniformly for $x \le 10r$,

$$\mathbb{E}\, S(xt_{\mathsf{c}}) = nx^r \frac{t_{\mathsf{c}}^r p^r}{r!}(1+o(1/x)) = x^r \frac{t_{\mathsf{c}}}{r}(1+o(1/x)) = \frac{x^r}{r} t_{\mathsf{c}} + o(t_{\mathsf{c}}).$$

Further, Lemma 7.3 yields by (8.1) and (3.4),

$$\sup_{0\le x\le 10r} |S(xt_{\mathsf{c}}) - \mathbb{E}\, S(xt_{\mathsf{c}})|^2 = O_{\mathrm{p}}\bigl(n\pi(10rt_{\mathsf{c}})\bigr) = O_{\mathrm{p}}(nt_{\mathsf{c}}^r p^r) = O_{\mathrm{p}}(t_{\mathsf{c}}) = o_{\mathrm{p}}(t_{\mathsf{c}}^2),$$

and the result follows. $\square$

We shall use Lemma 8.1 to prove now that in the subcritical case ($a \sim \alpha a_c$ with $\alpha < 1$) there exists $t < t_c$ such that w.h.p. $A(t) \le t$ and then determine the precise value of $A(T) = T$.

Proof of Theorem 3.1(i). The assumption on a may be written, by (3.2),

$$a = (\alpha + o(1))a_{\mathsf{c}} = \bigl(\alpha(1-r^{-1}) + o(1)\bigr)t_{\mathsf{c}}. \tag{8.3}$$

Hence, (2.9) and Lemma 8.1, taking $x = 1$, yield

$$A(t_{\mathsf{c}}) - t_{\mathsf{c}} = S(t_{\mathsf{c}}) + a - t_{\mathsf{c}} = t_{\mathsf{c}}/r + o_{\mathrm{p}}(t_{\mathsf{c}}) + a - t_{\mathsf{c}} = t_{\mathsf{c}}\bigl(r^{-1} + \alpha(1-r^{-1}) - 1 + o_{\mathrm{p}}(1)\bigr).$$

Since $\alpha(1-r^{-1}) < 1 - r^{-1}$, w.h.p. $A(t_{\mathsf{c}}) - t_{\mathsf{c}} < 0$, and thus, by (2.2), $T < t_{\mathsf{c}}$.

We apply Lemma 8.1 again, now taking $x = T/t_{\mathsf{c}}$, and see that $S(T) = (T/t_{\mathsf{c}})^r t_{\mathsf{c}}/r + o_{\mathrm{p}}(t_{\mathsf{c}})$. Since $S(T) = A(T) - a = T - a$, we find using (8.3) that

$$T - \alpha(1-r^{-1})t_{\mathsf{c}} = S(T) + o(t_{\mathsf{c}}) = \left(\frac{T}{t_{\mathsf{c}}}\right)^r \frac{t_{\mathsf{c}}}{r} + o_{\mathrm{p}}(t_{\mathsf{c}})$$

and thus

$$r\frac{T}{t_{\mathsf{c}}} - (r-1)\alpha = \left(\frac{T}{t_{\mathsf{c}}}\right)^r + o_{\mathrm{p}}(1). \tag{8.4}$$

Since the function $h(x) := rx - x^r$ is strictly increasing from 0 to $r-1$ on $[0,1]$, (8.4) implies (using the fact just shown that $T/t_{\mathsf{c}} < 1$ w.h.p.) that $T/t_{\mathsf{c}} \xrightarrow{\mathrm{p}} y$, where y is the unique root in $[0,1]$ of $h(y) = (r-1)\alpha$, i.e., $y = \varphi(\alpha)$ given by (3.11).

This proves the first assertion, and if $\alpha > 0$, the second follows. If $\alpha = 0$, then $a = o(t_{\mathsf{c}})$, and (8.2) implies, for every fixed $\lambda > 0$, $\mathbb{E}\, S(\lambda a) = O(na^r p^r) = o(ant_{\mathsf{c}}^{r-1}p^r) = o(a)$. Hence, for every fixed $\lambda > 1$, $A(\lambda a) = S(\lambda a) + a = a + o_{\mathrm{p}}(a)$, so w.h.p. $A(\lambda a) < \lambda a$, and thus $a \le T < \lambda a$. Consequently, when $\alpha = 0$, $T/a \xrightarrow{\mathrm{p}} 1$. $\square$

We turn to the proof of the supercritical case in Theorem 3.1. The following lemma shows that if the process of activation can escape the bottleneck at t_c then the process continues until (almost) percolation. The idea is to split the time interval $[3t_c, n]$ into different intervals. Then in the proof of Theorem 3.1(ii)–(iii), it remains to show that if a is supercritical then $A(t) > t$ for $t < 3t_c$.

Let $b^* := b_{\mathsf{c}}\omega(n)$, where $\omega(n) \to \infty$ slowly but is otherwise arbitrary.

Lemma 8.2. *Suppose that $r \ge 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Then, for any a, w.h.p. $A(t) > t$ for all $t \in [3t_{\mathsf{c}}, n - b^*]$.*

Proof. By (2.9), $A(t) = S_{n-a}(t) + a \geq S_n(t)$, so it suffices to show that $S_n(t) > t$ (or equivalently, to take $a = 0$). We separate the proof into a number of different cases for different ranges of t. We assume at some places, without further mention, that n is large enough.

Case 1: $t \in [3t_{\mathsf{c}}, 8rt_{\mathsf{c}}]$. By Lemma 8.1, w.h.p. for all such t,

$$S_n(t) \geq \frac{1}{r}\left(\frac{t}{t_{\mathsf{c}}}\right)^r t_{\mathsf{c}} - t_{\mathsf{c}} \geq \frac{3^{r-1}}{r}t - t_{\mathsf{c}} \geq \frac{3}{2}t - t_{\mathsf{c}} > t.$$

Case 2: $t \in [8rt_{\mathsf{c}}, p^{-1}]$. Let $t_j := 2^j rt_{\mathsf{c}}$, $j \geq 1$, and let $J := \min\{j \geq 1 : pt_j \geq 1\}$. For $t_{\mathsf{c}} \leq t \leq p^{-1}$, using (3.1),

$$\pi(t) \geq \binom{\lfloor t \rfloor}{r} p^r(1-p)^{t-r} = \frac{t^r}{r!}p^r e^{-tp}\big(1+o(1)\big) \geq \frac{1}{3}\frac{t^r p^r}{r!} = \frac{1}{3}t\left(\frac{t}{t_{\mathsf{c}}}\right)^{r-1}\frac{t_{\mathsf{c}}^{r-1}p^r}{r!} = \frac{t}{3rn}\left(\frac{t}{t_{\mathsf{c}}}\right)^{r-1}.$$

Hence, for $3 \leq j \leq J-1$, $\mathbb{E}\, S_n(t_j) = n\pi(t_j) \geq \frac{2^j}{3}t_j \geq \frac{8}{3}t_j$, and thus, using Chebyshev's inequality and (2.13),

$$\mathbb{P}\big(S_n(t_j) \leq 2t_j\big) \leq \mathbb{P}\big(S_n(t_j) \leq \tfrac{3}{4}\,\mathbb{E}\, S_n(t_j)\big) \leq \frac{\operatorname{Var} S_n(t_j)}{\big(\frac{1}{4}\,\mathbb{E}\, S_n(t_j)\big)^2} \leq \frac{16}{n\pi(t_j)} \leq \frac{6}{t_j}.$$

Hence,

$$\mathbb{P}\big(S_n(t) \leq t \text{ for some } t \in [8rt_{\mathsf{c}}, t_J]\big) \leq \sum_{j=3}^{J-1}\mathbb{P}\big(S_n(t_j) \leq 2t_j\big) \leq \sum_{j=3}^{J-1}\frac{6}{t_j} < \frac{12}{t_3} < \frac{2}{rt_{\mathsf{c}}} = o(1).$$

Case 3: $t \in [p^{-1}, c_1 n]$ for a suitable small $c_1 > 0$. Let $t_1' := \lceil p^{-1} \rceil$. Then

$$\pi(t_1') = \mathbb{P}\big(\mathrm{Bin}(t_1', p) \geq r\big) = \mathbb{P}\big(\mathrm{Po}(t_1'p) \geq r\big) + O(p) \geq 2c_1,$$

for some small c_1. Hence w.h.p. $S_n(t_1') > c_1 n$ and thus $S_n(t) \geq S_n(t_1') > c_1 n \geq t$.

Case 4: $t \in [c_1 n, n - p^{-1}]$. Let $t_2' := \lfloor c_1 n \rfloor$ and $t_3' := n - p^{-1}$. Then

$$1 - \pi(t_2') = \mathbb{P}\big(\mathrm{Bin}(t_2', p) < r\big) = O\big((t_2'p)^{r-1}e^{-t_2'p}\big) = O\big((np)^{r-1}e^{-c_1 np}\big) = o\big((np)^{-1}\big).$$

Thus, $\mathbb{E}\big(n - S_n(t_2')\big) = n(1-\pi(t_2')) = o(p^{-1})$, and w.h.p., $n - S_n(t_2') < p^{-1}$, i.e., $S_n(t_2') > n - p^{-1} = t_3'$.

Case 5: $t \in [n - p^{-1}, n - b^*]$. We have $t_3' := n - p^{-1}$. Then

$$1 - \pi(t_3') = \mathbb{P}\big(\mathrm{Bin}(\lfloor t_3' \rfloor, p) < r\big) = O\big((t_3'p)^{r-1}e^{-t_3'p}\big) = O\big((np)^{r-1}e^{-np}\big) = O(b_{\mathsf{c}}/n).$$

Hence, $\mathbb{E}\big(n - S_n(t_3')\big) = n(1 - \pi(t_3')) = O(b_{\mathsf{c}}) = o(b^*)$, and thus w.h.p. $n - S_n(t_3') < b^*$, i.e., $S_n(t_3') > n - b^*$. ∎

Remark 8.3. The proof shows that once we reach at least $1/p$ active vertices, the active set will w.h.p. grow to at least $n - b^*$ in at most 3 generations. (Hence, the size then is $n - O_{\mathrm{p}}(b_{\mathsf{c}})$; see [34, Lemma 3].)

Lemma 8.4.
$$\min_{x\geq 0}\left(\frac{x^r}{r}-x\right)=\frac{1}{r}-1,$$
attained at $x=1$ only.

Proof. Elementary calculus. $\square$

Proof of Theorem 3.1(ii)–(iii). For $0 \leq t \leq 3t_{\mathsf{c}}$, we may assume $a \leq 3t_{\mathsf{c}}$ since otherwise $A(t) > t$ trivially. In this case, Lemmas 8.1 and 8.4 (with $x = t/t_{\mathsf{c}}$) show that w.h.p., uniformly in $t \leq 3t_{\mathsf{c}}$,
$$A(t)=a+S(t)\geq(1+\delta)(1-r^{-1})t_{\mathsf{c}}+\frac{1}{r}\left(\frac{t}{t_{\mathsf{c}}}\right)^r t_{\mathsf{c}}-o(t_{\mathsf{c}})\geq\delta(1-r^{-1})t_{\mathsf{c}}+\frac{t}{t_{\mathsf{c}}}t_{\mathsf{c}}-o(t_{\mathsf{c}})>t.$$
This and Lemma 8.2 show that w.h.p. $A(t) > t$ for all $t \leq n - b^*$, and thus $A^* > n - b^*$.

Hence $n - A^* < b^* = b_{\mathsf{c}}\omega(n)$ w.h.p., for any choice of $\omega(n) \to \infty$, which is equivalent to $n - A^* = O_{\mathrm{p}}(b_{\mathsf{c}})$, see e.g. [34, Lemma 3]. This proves (ii).

If $b_{\mathsf{c}} \to 0$, we may choose $b^* = 1$; then w.h.p. $n - A^* < 1$, so $A^* = n$. Conversely, if $b_{\mathsf{c}} \not\to 0$, then, at least for a subsequence, there exists with probability at least $c > 0$ a vertex with degree $\leq r-1$, and with probability $1 - a/n$, this vertex will never be activated so $A^* < n$, see Remark 3.3. This proves (iii). $\square$

Proof of Theorem 3.2. Choose $b^* := npb_{\mathsf{c}} \gg b_{\mathsf{c}}$. By (3.3), $b^*p = (np)^{r+1}e^{-np}/(r-1)! \to 0$. Hence, $(n - b^*)p = np + o(1) \to \infty$ and
$$1-\pi(n-b^*)=\mathbb{P}\big(\mathrm{Bin}(\lfloor n-b^*\rfloor,p)\leq r-1\big)\sim\frac{(n-b^*)^{r-1}p^{r-1}}{(r-1)!}(1-p)^{n-b^*}\sim 1-\pi(n).$$
Consequently, see (3.7),
$$\mathbb{E}\big(A(n)-A(n-b^*)\big)=\mathbb{E}\big(S(n)-S(n-b^*)\big)\leq n\big(\pi(n)-\pi(n-b^*)\big)=o\big(n(1-\pi(n))\big)=o(b'_{\mathsf{c}}). \tag{8.5}$$

By assumption and Lemma 8.2, w.h.p. $T > n - b^*$, and thus $A(n - b^*) \leq A(T) \leq A(n)$. Hence (8.5) implies
$$A^*=T=A(T)=A(n)+o_{\mathrm{p}}(b'_{\mathsf{c}}). \tag{8.6}$$
Further,
$$n-A(n)=n-a-S(n)\in\mathrm{Bin}(n-a,1-\pi(n)), \tag{8.7}$$
with mean $(n - a)(1 - \pi(n)) \sim b'_{\mathsf{c}}$, see (3.7).

If $b_{\mathsf{c}} \to \infty$, then $b'_{\mathsf{c}} \sim b_{\mathsf{c}}$; thus (8.7) implies $n - A(n) = b_{\mathsf{c}} + o_{\mathrm{p}}(b_{\mathsf{c}})$, and (8.6) yields (i).

In $b_{\mathsf{c}} \to b < \infty$, then $b'_{\mathsf{c}} = b_{\mathsf{c}} + o(1) \to b$; thus (8.6) yields $A^* = A(n) + o_{\mathrm{p}}(1)$, and hence (since the variables are integer valued) $A^* = A(n)$ w.h.p. Further, in this case (8.7) implies $n - A(n) \xrightarrow{\mathrm{d}} \mathrm{Po}(b)$, and (ii) and (iii) follow. $\square$

Proof of Theorem 3.4. An easy consequence of Theorem 3.1. $\square$

We end this section with a proof of Theorem 4.4, where we start with a number of external infections (but no initially active vertices). As said in Subsection 4.2, we do this by a minor variation of our method. We include this proof to show the flexibility of the method, but we omit parts that are identical or almost identical to the proofs above.

Proof of Theorem 4.4. In order to preserve independence between vertices, we consider the model with a Poisson number $W \in \mathrm{Po}(\mu)$ of external infections (independent of everything else). Then each vertex i receives $W_i \in \mathrm{Po}(\mu/n)$ external infections, and these random variables are independent. The analysis in Section 2 becomes slightly modified: the number of infections (marks) at time t now is $M_i^\mu(t) := W_i + M_i(t)$, so Y_i is replaced by $Y_i^\mu := \min\{t : M_i^\mu(t) \geq r\}$ and $S(t)$ is replaced by $S^\mu(t) := \sum_{i=1}^n \mathbf{1}\{Y_i^\mu \leq t\}$. We now have $A(t) = S^\mu(t)$, so $A^* = T = \min\{t \geq 0 : S^\mu(t) = t\}$.

We take $\mu = ynpa_{\mathsf{c}}$ for a fixed $y > 0$ and claim that if $y < 1$, then w.h.p. $A^* < t_{\mathsf{c}}/r$ and thus $J_0 > W$, while if $y > 1$, then w.h.p. $A^* = n - o_{\mathrm{p}}(n)$ and thus $J_0 \leq W$. The result then follows by taking $y = 1 \pm \varepsilon/2$ for small $\varepsilon > 0$.

To prove these claims, we first note that $\mathbb{E}\, S^\mu(t) = n\,\mathbb{P}(M_i^\mu(t) \geq r)$ with, for such μ and $t = O(t_{\mathsf{c}})$,

$$
\begin{aligned}
\mathbb{P}(M_i^\mu(t) \geq r) = \mathbb{P}(W_i + M_i(t) \geq r) &= \sum_{j=0}^{r-1} \mathbb{P}(W_i = j)\,\mathbb{P}(M_i(t) \geq r - j) + \mathbb{P}(W_i \geq r) \\
&\sim \sum_{j=0}^{r} \frac{(\mu/n)^j}{j!} \cdot \frac{(tp)^{r-j}}{(r-j)!} = \frac{(tp + \mu/n)^r}{r!} = \frac{p^r(t + ya_{\mathsf{c}})^r}{r!}.
\end{aligned}
\tag{8.8}
$$

We obtain as in Lemma 8.1, using versions of Lemmas 7.2 and 7.3 for $S^\mu(t)$,

$$
\sup_{0 \leq x \leq 10r} \left| S^\mu(xt_{\mathsf{c}}) - \tfrac{1}{r}(x + ya_{\mathsf{c}}/t_{\mathsf{c}})^r t_{\mathsf{c}} \right| = o_{\mathrm{p}}(t_{\mathsf{c}}).
\tag{8.9}
$$

Recall that $a_{\mathsf{c}}/t_{\mathsf{c}} = 1 - 1/r$ by (3.2). If $y < 1$, then (8.9) with $x = 1/r$ implies that w.h.p.

$$
A\left(\frac{t_{\mathsf{c}}}{r}\right) = S^\mu\left(\frac{t_{\mathsf{c}}}{r}\right) < \frac{1}{r}\left(\frac{1}{r} + \left(1 - \frac{1}{r}\right)\right)^r t_{\mathsf{c}} = \frac{t_{\mathsf{c}}}{r}
$$

and thus $A^* = T < t_{\mathsf{c}}/r$ as claimed.

Conversely, if $y > 1$, then Lemma 8.4 shows that

$$
\frac{(x + ya_{\mathsf{c}}/t_{\mathsf{c}})^r}{r} \geq x + y\frac{a_{\mathsf{c}}}{t_{\mathsf{c}}} + \left(\frac{1}{r} - 1\right) = x + (y-1)\left(1 - \frac{1}{r}\right).
$$

Hence, (8.9) shows that w.h.p. $A(xt_{\mathsf{c}}) = S^\mu(xt_{\mathsf{c}}) > xt_{\mathsf{c}}$ for $x \leq 10r$, and thus $A^* = T > 10rt_{\mathsf{c}}$. Further, since $S^\mu(t) \geq S_n(t)$, Lemma 8.2 implies that w.h.p. $A(t) > t$ for all $t \in [3t_{\mathsf{c}}, n - b^*]$, and thus $A^* \geq n - b^*$ w.h.p., which proves the second claim and completes the proof. $\square$

Note that (8.8) and (8.9) show that $A(t) = S^\mu(t)$ is, to the first order, $\mathbb{E}\, S_n(t)$ shifted horizontally by $\mu/(np) = ya_{\mathsf{c}}$, while in our standard model $A(t)$ is S_t shifted vertically by a. Since we study the hitting time of the linear barrier $A(t) = t$, these are essentially equivalent.

9 Proofs of Theorems 3.6–3.8 and 4.5

We begin with an estimate of $\tilde\pi(t)$ defined in (3.13).

Lemma 9.1. *Suppose that $r \ge 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Then, for large n, $n\tilde\pi(t) \ge 1.4\,t$ for $t \in [3t_{\mathsf{c}}, n/2]$.*

Proof. Assume not. Then we can find, for a subsequence $n = n_k \to \infty$, $t = t_k \in [3t_{\mathsf{c}}, n/2]$ such that $n\tilde\pi(t) < 1.4\,t$. Selecting a subsequence, we may further assume that $pt \to z \in [0,\infty]$. We consider three cases separately.

(i): $z = 0$, i.e., $pt \to 0$. Then, from (3.13) and (3.4),

$$n\tilde\pi(t) \sim n\frac{(pt)^r}{r!} = \frac{t^r}{rt_{\mathsf{c}}^{r-1}} \ge \frac{3^{r-1}}{r}t \ge \frac{3}{2}t.$$

(ii): $pt \to z \in (0,\infty)$. Then $n\tilde\pi(t) \sim n\psi(z)$ with $\psi(z) > 0$, and $t = O(1/p) = o(n) \ll n\tilde\pi(t)$.

(iii): $z = \infty$, i.e., $pt \to \infty$. Then $n\tilde\pi(t) \sim n \ge 2t$.

In all cases we have for large n a contradiction to $n\tilde\pi(t) < 1.4\,t$. □

Lemma 9.2. *Suppose $r \ge 2$ and $a \to \infty$ with $a = o(n)$. Then p_{c} and p^*_{c} defined by* (3.12) *and* (3.19) *satisfy $p^*_{\mathsf{c}} \sim p_{\mathsf{c}}$. In particular, $n^{-1} \ll p^*_{\mathsf{c}} \ll n^{-1/r}$.*

Proof. Let $p = yp_{\mathsf{c}}$ for some fixed $y > 0$ and define t_{c} and a_{c} by (3.1) and (3.2). Then $t_{\mathsf{c}} = y^{-(r-1)/r}(r/(r-1))a$ and $a_{\mathsf{c}} = ay^{-(r-1)/r}$. Further, $n^{-1} \ll p \ll n^{-1/r}$ and, by (3.5), $pt_{\mathsf{c}} \to 0$ and $t_{\mathsf{c}} = o(n)$. Hence, if $x = O(1)$, and $t = xt_{\mathsf{c}}$, then $pt = o(1)$, $t = o(n)$ and, uniformly in bounded x, by (3.13) and (3.4),

$$\tilde\pi(t) = \frac{(pt)^r}{r!} + O\big((pt)^{r+1}\big) = \frac{x^r t_{\mathsf{c}}}{nr}\big(1+o(1)\big).$$

Hence, uniformly in $x \le 3$,

$$(n-a)\tilde\pi(t) - t = \Big(\frac{x^r}{r} + o(1) - x\Big)t_{\mathsf{c}}. \tag{9.1}$$

By Lemma 9.1, for large n, $(n-a)\tilde\pi(t) - t \ge 0$ for $t \in [3t_{\mathsf{c}}, n/2]$, and thus, by (9.1) and Lemma 8.4,

$$\begin{aligned}\gamma(p) &= \inf_{t\le 3t_{\mathsf{c}}}\big\{(n-a)\tilde\pi(t) - t\big\} = \Big(\inf_{x\le 3}\Big(\frac{x^r}{r} - x\Big) + o(1)\Big)\,t_{\mathsf{c}} = \Big(\frac{1}{r} - 1 + o(1)\Big)t_{\mathsf{c}} = -\big(1+o(1)\big)a_{\mathsf{c}} \\ &= -\Big(y^{-(r-1)/r} + o(1)\Big)a.\end{aligned} \tag{9.2}$$

Hence, if $y = 1-\delta < 1$, then $y^{-(r-1)/r} > 1$ and thus, for large n, $\gamma(p) < -a$ so $p^*_{\mathsf{c}} > p = (1-\delta)p_{\mathsf{c}}$. Conversely, if $y = 1+\delta > 1$, then (9.2) yields, for large n, $\gamma(p) > -a$ so $p^*_{\mathsf{c}} < p = (1+\delta)p_{\mathsf{c}}$.

Consequently, $p^*_{\mathsf{c}}/p_{\mathsf{c}} \to 1$. □

We also need more precise estimates of $S(t)$. The following Gaussian process limit is fundamental. $D[0,B]$ denotes the space of right-continuous functions on $[0,B]$, with the Skorohod topology, see e.g. [17] (for $B = 1$; the general case is similar by a change of variables) or [36, Chapter 16].

Lemma 9.3. *Suppose $r \geq 2$ and $a \to \infty$ with $a = o(n)$. Then*

$$Z(x) := \frac{S(xt_{\mathsf{c}}) - \mathbb{E}\, S(xt_{\mathsf{c}})}{\sqrt{t_{\mathsf{c}}}} \xrightarrow{\mathrm{d}} W(x^r/r) \tag{9.3}$$

in $D[0, B]$ for any fixed B, where W is a standard Brownian motion.

The conclusion, convergence in $D[0, B]$ for every fixed B, can also be expressed as convergence in $D[0, \infty)$.

Proof. This is a result on convergence of empirical distribution functions (of $\{Y_i\}$), cf. [17, Theorem 16.4]; we get here a Brownian motion instead of a Brownian bridge as in [17] because we consider for each B only a small initial part of the distribution of Y_i.

For every fixed $x > 0$, by (8.1) and (3.5), $\pi(xt_{\mathsf{c}}) \sim (xt_{\mathsf{c}}p)^r/r! \to 0$, and thus by (2.13) and (3.4)

$$\operatorname{Var} S(xt_{\mathsf{c}}) \sim n\pi(xt_{\mathsf{c}}) \sim \frac{np^r x^r t_{\mathsf{c}}^r}{r!} = \frac{x^r t_{\mathsf{c}}}{r} \to \infty.$$

Hence (2.10) and the Central Limit Theorem yield $Z(x) \xrightarrow{\mathrm{d}} N(0, x^r/r)$ for every $x > 0$, which proves $Z(x) \xrightarrow{\mathrm{d}} W(x^r/r)$ for each fixed x.

This is easily extended to finite-dimensional convergence: Suppose that $0 < x_1 < \cdots < x_\ell$ are fixed, and let $I_{ij} := \mathbf{1}\{Y_i \in (x_{j-1}t_{\mathsf{c}}, x_j t_{\mathsf{c}}]\}$, with $x_0 = 0$. Thus, $S(x_j t_{\mathsf{c}}) - S(x_{j-1}t_{\mathsf{c}}) = \sum_{i=1}^{n-a} I_{ij}$. Then, for $1 \leq j \leq \ell$ and $k \neq j$,

$$\begin{aligned}
\mathbb{E}\, I_{ij} &= \pi(x_j t_{\mathsf{c}}) - \pi(x_{j-1}t_{\mathsf{c}}), \\
\operatorname{Var} I_{ij} &= \mathbb{E}\, I_{ij}(1 - \mathbb{E}\, I_{ij}) \sim \mathbb{E}\, I_{ij} = \pi(x_j t_{\mathsf{c}}) - \pi(x_{j-1}t_{\mathsf{c}}) \sim \Bigl(\frac{x_j^r}{r} - \frac{x_{j-1}^r}{r}\Bigr)\frac{t_{\mathsf{c}}}{n}, \\
\operatorname{Cov}(I_{ij}, I_{ik}) &= -\,\mathbb{E}\, I_{ij}\, \mathbb{E}\, I_{ik} = O\bigl(\pi(x_\ell t_{\mathsf{c}})^2\bigr) = O\bigl((t_{\mathsf{c}}/n)^2\bigr) = o(t_{\mathsf{c}}/n).
\end{aligned}$$

Note that $(I_{ij})_{j=1}^{\ell}$, $i = 1, 2, \dots, n$, are i.i.d. random vectors. The multi-dimensional Central Limit Theorem with (for example) the Lindeberg condition (which follows from the one-dimensional version in e.g. [29, Theorem 7.2.4] or [36, Theorem 5.12] by the Cramér–Wold device) thus shows that $\bigl(Z(x_j) - Z(x_{j-1})\bigr)_{j=1}^{\ell} \xrightarrow{\mathrm{d}} (V_j)_{j=1}^{\ell}$ with V_j jointly normal with $\mathbb{E}\, V_j = 0$, $\operatorname{Var} V_j = x_j^r/r - x_{j-1}^r/r$ and $\operatorname{Cov}(V_j, V_k) = 0$ for $j \neq k$. Hence, $(V_j)_{j=1}^{\ell} \overset{\mathrm{d}}{=} (W(x_j^r/r) - W(x_{j-1}^r/r)_{j=1})^{\ell}$, and thus $\bigl(Z(x_j)\bigr)_{j=1}^{\ell} \xrightarrow{\mathrm{d}} (W(x_j^r/r))_{j=1}^{\ell}$.

To show (9.3), it thus remains to show tightness of $Z(x)$. We use [17, Theorem 15.6] with $\gamma = 2$ and $\alpha = 1$ (an alternative would be to instead use Aldous's tightness criterion [36, Theorem 16.11]); it thus suffices to prove that, for every x_1, x_2, x_3 with $0 \leq x_1 \leq x_2 \leq x_3 \leq B$, and some constant C depending on B but not on n or x_1, x_2, x_3,

$$\mathbb{E}\bigl\{|Z(x_2) - Z(x_1)|^2 |Z(x_3) - Z(x_2)|^2\bigr\} \leq C(x_3 - x_1)^2. \tag{9.4}$$

With the notation above and $I'_{ij} := I_{ij} - \mathbb{E}\, I_{ij}$, the left hand side of (9.4) can be written

$$t_{\mathsf{c}}^{-2}\, \mathbb{E} \sum_{i,j,k,l=1}^{n-a} I'_{i2} I'_{j2} I'_{k3} I'_{l3} = t_{\mathsf{c}}^{-2} \sum_{i,j,k,l=1}^{n-a} \mathbb{E}(I'_{i2} I'_{j2} I'_{k3} I'_{l3}).$$

By independence, the only non-zero terms are those where i, j, k, l either coincide in two pairs, or all four indices coincide, and it follows easily that (for any i)

$$\mathbb{E}\big\{|Z(x_2) - Z(x_1)|^2 |Z(x_3) - Z(x_2)|^2\big\} \le 3t_{\mathsf{c}}^{-2}(n-a)^2\, \mathbb{E}\, I_{i2}\, \mathbb{E}\, I_{i3}. \tag{9.5}$$

Further, since each Y_i is integer-valued, the left-hand side of (9.4) vanishes unless there is at least one integer in each of the intervals $(x_1 t_{\mathsf{c}}, x_2 t_{\mathsf{c}}]$ and $(x_2 t_{\mathsf{c}}, x_3 t_{\mathsf{c}}]$, which implies that $x_3 t_{\mathsf{c}} - x_1 t_{\mathsf{c}} > 1$, so we only have to consider this case. It follows from (2.7) that for $m \le x_3 t_{\mathsf{c}} \le B t_{\mathsf{c}}$,

$$\mathbb{P}(Y_i = m) \le \frac{m^{r-1}}{(r-1)!} p^r \le \frac{B^{r-1} t_{\mathsf{c}}^{r-1} p^r}{(r-1)!} = \frac{B^{r-1}}{n},$$

and thus, assuming $x_3 t_{\mathsf{c}} - x_1 t_{\mathsf{c}} > 1$,

$$\mathbb{E}\, I_{i2} + \mathbb{E}\, I_{i3} \le (\lfloor x_3 t_{\mathsf{c}} \rfloor - \lfloor x_1 t_{\mathsf{c}} \rfloor) \frac{B^{r-1}}{n} \le (x_3 t_{\mathsf{c}} - x_1 t_{\mathsf{c}} + 1) \frac{B^{r-1}}{n} \le 2(x_3 t_{\mathsf{c}} - x_1 t_{\mathsf{c}}) \frac{B^{r-1}}{n}.$$

Consequently, (9.5) yields,

$$\mathbb{E}\big\{|Z(x_2) - Z(x_1)|^2 |Z(x_3) - Z(x_2)|^2\big\} \le 3\frac{n^2}{t_{\mathsf{c}}^2} \Big(2(x_3 t_{\mathsf{c}} - x_1 t_{\mathsf{c}}) \frac{B^{r-1}}{n}\Big)^2 = 12(x_3 - x_1)^2 B^{2(r-1)},$$

which proves (9.4) with $C = 12B^{2(r-1)}$. The proof is complete. ∎

We also need a more careful estimate of $\pi(t)$ than above, and we use the corresponding Poisson probability $\tilde{\pi}(t)$ defined in (3.13).

Lemma 9.4. *Assume $n^{-1} \ll p \ll n^{-1/r}$. Uniformly for $t \ge 1$, $\pi(t) = \tilde{\pi}(t)\big(1 + O(t^{-1})\big)$. In particular, uniformly for $t \le 3t_{\mathsf{c}}$,*

$$\pi(t) = \tilde{\pi}(t) + O\big((pt)^r/t\big) = \tilde{\pi}(t) + O\big(\tilde{\pi}(t_{\mathsf{c}})/t_{\mathsf{c}}\big) = \tilde{\pi}(t) + O\big(n^{-1}\big).$$

Proof. Assume first $pt \le 1$. By (2.11),

$$\begin{aligned}
\pi(t) &= \sum_{j=r}^{t} \mathbb{P}\big(\mathrm{Bin}(t,p) = j\big) = \sum_{j=r}^{t} \frac{t^j}{j!} \left(1 + O\Big(\frac{j^2}{t}\Big)\right) p^j (1-p)^{t+O(j)} \\
&= \sum_{j=r}^{\infty} \frac{(pt)^j}{j!} e^{-pt + O(tp^2)} \Big(1 + O(j^2/t + jp)\Big) \\
&= \tilde{\pi}(t)\big(1 + O(tp^2 + t^{-1} + p)\big) = \tilde{\pi}(t)\big(1 + O(t^{-1})\big).
\end{aligned}$$

For $pt > 1$, $\tilde{\pi}(t)$ is bounded below, and the result follows from (3.15).

If $t \le 3t_{\mathsf{c}}$, then $t = O(t_{\mathsf{c}}) = o(1/p)$ by (3.5), and thus, using (3.13) and (3.4),

$$\pi(t) - \tilde{\pi}(t) = O\big(\tilde{\pi}(t)/t\big) = O\big((pt)^r/t\big) = O\big((pt_{\mathsf{c}})^r/t_{\mathsf{c}}\big) = O\big(1/n\big).$$ ∎

Proof of Theorem 3.6. It suffices to consider a such that $a \sim a_{\mathsf{c}} = (1 - r^{-1})t_{\mathsf{c}}$. It then follows by (2.9) and Lemma 8.1 that, uniformly for $x \le 10r$,

$$A(xt_{\mathsf{c}}) - xt_{\mathsf{c}} = a + S(xt_{\mathsf{c}}) - xt_{\mathsf{c}} = a_{\mathsf{c}} + \tfrac{1}{r}x^r t_{\mathsf{c}} - xt_{\mathsf{c}} + o_{\mathrm{p}}(t_{\mathsf{c}}) = \bigl(1 - r^{-1} + \tfrac{1}{r}x^r - x\bigr)t_{\mathsf{c}} + o_{\mathrm{p}}(t_{\mathsf{c}}). \tag{9.6}$$

By Lemma 8.4, the coefficient $1 - r^{-1} + x^r/r - x$ equals 0 at $x = 1$ but is strictly positive for all other $x \ge 0$. It follows that for every $\delta > 0$, w.h.p. $A(xt_{\mathsf{c}}) - xt_{\mathsf{c}} > 0$ for all $x \in [0, 1-\delta] \cup [1+\delta, 10r]$. By a simple standard argument, there thus exists a sequence $\delta_n \to 0$, where we may further assume that $\delta_n > |t^*_{\mathsf{c}}/t_{\mathsf{c}} - 1|$, such that w.h.p. $A(xt_{\mathsf{c}}) - xt_{\mathsf{c}} > 0$ for all $x \in [0, 1 - \delta_n] \cup [1 + \delta_n, 10r]$.

Hence, w.h.p. either $T \in [(1 - \delta_n)t_{\mathsf{c}}, (1 + \delta_n)t_{\mathsf{c}}]$, or $A(t) > t$ for all $t \le 10rt_{\mathsf{c}}$; in the latter case, for any $b^* \gg b_{\mathsf{c}}$, w.h.p. $A(t) > t$ for all $t \le n - b^*$ by Lemma 8.2, so $T \ge n - b^*$; hence $T = n - O_{\mathrm{p}}(b_{\mathsf{c}})$ and, more precisely, provided $a = o(n)$, Theorem 3.2 applies.

We thus only have to investigate the interval $[(1-\delta_n)t_{\mathsf{c}}, (1+\delta_n)t_{\mathsf{c}}]$ more closely. By the Skorohod coupling theorem [36, Theorem 4.30], we may assume that the processes for different n are coupled such that the limit (9.3) holds a.s., and not just in distribution. Since convergence in $D[0, B]$ to a continuous function is equivalent to uniform convergence, this means that (a.s.) $Z(x) \to W(x^r/r)$ uniformly for $x \le B$; in particular, uniformly for $x \in [1 - \delta_n, 1 + \delta_n]$,

$$S(xt_{\mathsf{c}}) = (n - a)\pi(xt_{\mathsf{c}}) + t_{\mathsf{c}}^{1/2} Z(x) = (n - a)\pi(xt_{\mathsf{c}}) + t_{\mathsf{c}}^{1/2}\bigl(W(1/r) + o(1)\bigr). \tag{9.7}$$

Let $\xi := W(1/r) \in N(0, 1/r)$. Then, by (9.7) and Lemma 9.4, uniformly for $x \in [1 - \delta_n, 1 + \delta_n]$,

$$S(xt_{\mathsf{c}}) = (n - a)\tilde{\pi}(xt_{\mathsf{c}}) + O(1) + t_{\mathsf{c}}^{1/2}\bigl(\xi + o(1)\bigr) = (n - a)\tilde{\pi}(xt_{\mathsf{c}}) + t_{\mathsf{c}}^{1/2}\bigl(\xi + o(1)\bigr), \tag{9.8}$$

and thus, refining (9.6),

$$A(xt_{\mathsf{c}}) - xt_{\mathsf{c}} = a + S(xt_{\mathsf{c}}) - xt_{\mathsf{c}} = a + (n - a)\tilde{\pi}(xt_{\mathsf{c}}) - xt_{\mathsf{c}} + t_{\mathsf{c}}^{1/2}\xi + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}). \tag{9.9}$$

Hence, recalling (3.16) and that the minimum there is attained at $t^*_{\mathsf{c}} \in [(1 - \delta_n)t_{\mathsf{c}}, (1 + \delta_n)t_{\mathsf{c}}]$,

$$\begin{aligned}\min_{t\in[(1-\delta_n)t_{\mathsf{c}},(1+\delta_n)t_{\mathsf{c}}]} \frac{A(t) - \lfloor t\rfloor}{1 - \tilde{\pi}(t)} &= a + \min_{t\in[(1-\delta_n)t_{\mathsf{c}},(1+\delta_n)t_{\mathsf{c}}]} \frac{n\tilde{\pi}(t) - t}{1 - \tilde{\pi}(t)} + t_{\mathsf{c}}^{1/2}\xi + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}) \\ &= a - a^*_{\mathsf{c}} + t_{\mathsf{c}}^{1/2}\xi + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}).\end{aligned} \tag{9.10}$$

We have shown that w.h.p. $A^* = T \le (1 + \delta_n)t_{\mathsf{c}}$ if and only if this minimum is ≤ 0, and otherwise $T = n - O_{\mathrm{p}}(b_{\mathsf{c}})$, and the results follow; for (i) we also observe that (9.9) and (9.10) imply that w.h.p. $A(t^*_{\mathsf{c}}) - t^*_{\mathsf{c}} < 0$ and thus $T < t^*_{\mathsf{c}}$. For example, in (iii) we have

$$a - a^*_{\mathsf{c}} + t_{\mathsf{c}}^{1/2}\xi + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}) = ya_{\mathsf{c}}^{1/2} + t_{\mathsf{c}}^{1/2}\xi + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}) = \bigl((r - 1)^{1/2}y + r^{1/2}\xi + o_{\mathrm{p}}(1)\bigr)(t_{\mathsf{c}}/r)^{1/2}$$

and the probability that this is positive tends to $\mathbb{P}\bigl((r - 1)^{1/2}y + r^{1/2}\xi > 0\bigr) = \Phi((r - 1)^{1/2}y)$, since $r^{1/2}\xi \in N(0, 1)$. $\square$

Proof of Theorem 3.7. It suffices to consider $p \sim p^*_{\mathsf{c}} \sim p_{\mathsf{c}}$, which implies that $a_{\mathsf{c}} = a_{\mathsf{c}}(p) \sim a_{\mathsf{c}}(p_{\mathsf{c}}) = a$. Hence the arguments in the proof of Theorem 3.6 apply. In particular, again it suffices to consider $t \in J = J_n := [(1 - \delta_n)t_{\mathsf{c}}, (1 + \delta_n)t_{\mathsf{c}}]$, where now $t_{\mathsf{c}} = t_{\mathsf{c}}(p_{\mathsf{c}}) = (r/(r - 1))a$. The infimum in (3.18) is attained for some $t = t_{\mathsf{c}}^{**}$, where by Lemma 9.1 $t_{\mathsf{c}}^{**} \le 3t_{\mathsf{c}}$ for large n, and an argument as in (9.6) shows that $t_{\mathsf{c}}^{**} \sim t_{\mathsf{c}}$. We may assume that δ_n is chosen such that $t_{\mathsf{c}}^{**} \in J$. Then, by (9.9),

$$\min_{t\in J}\bigl\{A(t) - \lfloor t\rfloor\bigr\} = a + \min_{t\in J}\bigl\{(n-a)\tilde\pi(t) - t\bigr\} + t_{\mathsf{c}}^{1/2}\xi + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}) \tag{9.11}$$

where, by (3.18) and the comments just made (for large n)

$$a + \min_{t\in J}\bigl\{(n-a)\tilde\pi(t) - t\bigr\} = a + \gamma(p) = -\gamma(p_{\mathsf{c}}^*) + \gamma(p). \tag{9.12}$$

Further, writing (3.18) as $\gamma(p) := \min_t\{F(tp) - t\}$, with $F(x) = (n-a)\psi(x)$, we have at the minimum point $t := t_{\mathsf{c}}^{**}$ the derivative $pF'(pt_{\mathsf{c}}^{**}) - 1 = 0$. Hence, uniformly for $|p_1 - p_{\mathsf{c}}^*| \le \varepsilon p_{\mathsf{c}}^*$ and $|t - t_{\mathsf{c}}| \le \varepsilon t_{\mathsf{c}}$, for any $\varepsilon = \varepsilon_n \to 0$, using (3.14), $F'(p_1 t) = (1+o(1))F'(pt_{\mathsf{c}}^{**}) = (1+o(1))/p_{\mathsf{c}}^*$ and thus by the mean-value theorem, for some p_1 between p and p_{c}^*,

$$F(tp) - F(tp_{\mathsf{c}}^*) = t(p - p_{\mathsf{c}}^*)F'(tp_1) = t_{\mathsf{c}}\frac{p - p_{\mathsf{c}}^*}{p_{\mathsf{c}}^*}\bigl(1 + o(1)\bigr).$$

Since the minimum in (3.18) may be taken over such t only, for suitable ε_n, this yields

$$\gamma(p) - \gamma(p_{\mathsf{c}}^*) = t_{\mathsf{c}}\frac{p - p_{\mathsf{c}}^*}{p_{\mathsf{c}}^*}\bigl(1 + o(1)\bigr).$$

Consequently, (9.11) and (9.12) yield

$$\min_{t\in J}\bigl\{A(t) - \lfloor t\rfloor\bigr\} = t_{\mathsf{c}}\frac{p - p_{\mathsf{c}}^*}{p_{\mathsf{c}}^*}\bigl(1 + o(1)\bigr) + t_{\mathsf{c}}^{1/2}\xi + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}).$$

Hence,

$$\mathbb{P}\bigl(\min_{t\in J}\bigl\{A(t) - \lfloor t\rfloor\bigr\} > 0\bigr) = \mathbb{P}\Bigl(t_{\mathsf{c}}^{1/2}\frac{p - p_{\mathsf{c}}^*}{p_{\mathsf{c}}^*} + \xi > 0\Bigr) + o(1) = \mathbb{P}\Bigl(-r^{1/2}\xi < (rt_{\mathsf{c}})^{1/2}\frac{p - p_{\mathsf{c}}^*}{p_{\mathsf{c}}^*}\Bigr) + o(1),$$

where $r^{1/2}\xi \in N(0,1)$ and $t_{\mathsf{c}} = \frac{r}{r-1}a$, and the different parts of the theorem follow. ∎

Lemma 9.5. *Suppose that $r \ge 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Then, for large n at least, the minimum point t_{c}^* in* (3.16) *is unique, and $t_{\mathsf{c}}^* \sim t_{\mathsf{c}}$, $a_{\mathsf{c}}^* \sim a_{\mathsf{c}}$; more precisely,*

$$t_{\mathsf{c}}^* = \Bigl(1 + \frac{pt_{\mathsf{c}}}{r-1} + o(pt_{\mathsf{c}})\Bigr)t_{\mathsf{c}}, \tag{9.13}$$

$$a_{\mathsf{c}}^* = \Bigl(1 - \frac{1}{r} + \frac{pt_{\mathsf{c}}}{r+1} + o(pt_{\mathsf{c}})\Bigr)t_{\mathsf{c}} = \Bigl(1 + \frac{rpt_{\mathsf{c}}}{r^2-1} + o(pt_{\mathsf{c}})\Bigr)a_{\mathsf{c}}, \tag{9.14}$$

Proof. Let

$$g(t) := \frac{n\tilde\pi(t) - t}{1 - \tilde\pi(t)} = \frac{n - t}{1 - \tilde\pi(t)} - n. \tag{9.15}$$

Then

$$g'(t) = \frac{-(1 - \tilde\pi(t)) + (n-t)\tilde\pi'(t)}{(1 - \tilde\pi(t))^2} = \frac{(n-t)\tilde\pi'(t) + \tilde\pi(t) - 1}{(1 - \tilde\pi(t))^2}.$$

Let

$$h(t) := (1-\tilde\pi(t))^2 g'(t) = (n-t)\tilde\pi'(t) + \tilde\pi(t) - 1. \tag{9.16}$$

Then $h'(t) = (n-t)\tilde\pi''(t) > 0$ for $t < (r-1)/p$, and in particular (for large n) for $t \le 3t_{\mathsf c}$, see (3.5). Further, $h(0) = -1$ and by (3.14) and (3.1), for large n,

$$h(3t_{\mathsf c}) = 3^{r-1} - 1 + o(1) > 0;$$

hence, there is a unique $t_{\mathsf c}^* \in [0, 3t_{\mathsf c}]$ such that $h(t_{\mathsf c}^*) = 0$, or equivalently $g'(t_{\mathsf c}^*) = 0$. Further, $g''(t_{\mathsf c}^*) = h'(t_{\mathsf c}^*)/(1-\tilde\pi(t_{\mathsf c}^*))^2 > 0$ so $t_{\mathsf c}^*$ is the unique minimum point of $g(t)$ in $[0, 3t_{\mathsf c}]$, as we defined $t_{\mathsf c}^*$ after (3.16).

Let $x = t_{\mathsf c}^*/t_{\mathsf c} \in [0,3]$. Then, by (3.14), (3.1) and (3.17),

$$0 = h(t_{\mathsf c}^*) = \Bigl(1 - \frac{t_{\mathsf c}^*}{n}\Bigr)x^{r-1}e^{-pt_{\mathsf c}^*} + O\Bigl(\frac{t_{\mathsf c}^*}{n}\Bigr) - 1 = x^{r-1}e^{-pt_{\mathsf c}^*} - 1 + O\Bigl(\frac{t_{\mathsf c}}{n}\Bigr).$$

Hence, recalling $n^{-1} \ll p$ and $pt_{\mathsf c} \to 0$,

$$x = e^{pt_{\mathsf c}^*/(r-1)}\Bigl(1 + O\Bigl(\frac{t_{\mathsf c}}{n}\Bigr)\Bigr) = 1 + \frac{pt_{\mathsf c}^*}{r-1} + o(pt_{\mathsf c}). \tag{9.17}$$

In particular, $x = 1 + o(1)$, so $t_{\mathsf c}^* \sim t_{\mathsf c}$, and (9.13) follows from (9.17). Finally, substituting (9.13) in (3.16) yields, using (9.15) together with $\tilde\pi(t_{\mathsf c}^*) = O(t_{\mathsf c}/n) = o(pt_{\mathsf c})$ by (3.17) and $n^{-1} \ll p$, and also (3.13) and (3.17),

$$\begin{aligned}
\frac{a_{\mathsf c}^*}{t_{\mathsf c}} &= -\frac{g(t_{\mathsf c}^*)}{t_{\mathsf c}} = \bigl(1 + o(pt_{\mathsf c})\bigr)\frac{t_{\mathsf c}^* - n\tilde\pi(t_{\mathsf c}^*)}{t_{\mathsf c}} = \bigl(1 + o(pt_{\mathsf c})\bigr)\Bigl(x - \frac{n\tilde\pi(t_{\mathsf c}^*)}{t_{\mathsf c}}\Bigr) \\
&= \bigl(1 + o(pt_{\mathsf c})\bigr)\Bigl(x - \frac{n(pt_{\mathsf c}^*)^r}{r!\,t_{\mathsf c}}e^{-pt_{\mathsf c}^*}\Bigl(1 + \frac{pt_{\mathsf c}^*}{r+1} + o(pt_{\mathsf c}^*)\Bigr)\Bigr) \\
&= \bigl(1 + o(pt_{\mathsf c})\bigr)\Bigl(x - \frac{x^r}{r}e^{-pt_{\mathsf c}}\Bigl(1 + \frac{pt_{\mathsf c}}{r+1} + o(pt_{\mathsf c})\Bigr)\Bigr) \\
&= x - \frac{x^r}{r}\Bigl(1 - pt_{\mathsf c} + \frac{pt_{\mathsf c}}{r+1}\Bigr) + o(pt_{\mathsf c}) = x - \frac{x^r}{r} + \frac{1}{r}\Bigl(pt_{\mathsf c} - \frac{pt_{\mathsf c}}{r+1}\Bigr) + o(pt_{\mathsf c}),
\end{aligned}$$

and (9.14) follows by (9.17) and $x - x^r/r = 1 - 1/r + O(x-1)^2$. ☐

Proof of Theorem 3.8. In the case (i), i.e., when $\alpha < 1$, by Theorem 3.1(i),

$$T = A^* = (\varphi(\alpha) + o_{\mathrm p}(1))t_{\mathsf c}. \tag{9.18}$$

By Theorem 3.6(i), this holds as well in case (ii), i.e., when $\alpha = 1$ and, correspondingly, $\varphi(\alpha) = \varphi(1) = 1$. Thus, for any $\alpha \le 1$ there exist $\delta_n \to 0$ such that w.h.p. $T \in I_n := [(\varphi(\alpha) - \delta_n)t_{\mathsf c}, (\varphi(\alpha) + \delta_n)t_{\mathsf c}]$. As in the proof of Theorem 3.6, we may by the Skorohod coupling theorem [36, Theorem 4.30] assume that the limit in (9.3) holds a.s., uniformly in $x \le B$. For $t \in I_n$, $t/t_{\mathsf c} \to \varphi(\alpha)$, and (9.3) then implies that, uniformly for $t \in I_n$,

$$S(t) = \mathbb{E}\, S(t) + t_{\mathsf c}^{1/2} Z(t/t_{\mathsf c}) = (n-a)\pi(t) + t_{\mathsf c}^{1/2} W\bigl(\varphi(\alpha)^r/r\bigr) + o_{\mathrm p}(t_{\mathsf c}^{1/2}). \tag{9.19}$$

Let $\xi := W\big(\varphi(\alpha)^r/r\big) \in N\big(0, \varphi(\alpha)^r/r\big)$. Then, by (9.19) and Lemma 9.4, for $t \in I_n$,

$$S(t) = (n-a)\tilde{\pi}(t) + t_{\mathsf{c}}^{1/2}(\xi + o_{\mathrm{p}}(1)).$$

Since w.h.p. $T \in I_n$, we may here substitute $t = T$, and obtain

$$0 = A(T) - T = a + S(T) - T = a + (n-a)\tilde{\pi}(T) - T + t_{\mathsf{c}}^{1/2}(\xi + o_{\mathrm{p}}(1)). \tag{9.20}$$

Define the function $\tilde{g}(t)$ by

$$\tilde{g}(t) := a + (n-a)\tilde{\pi}(t) - t; \tag{9.21}$$

thus (3.22) is $\tilde{g}(t_*) = 0$. Then we have shown in (9.20),

$$\tilde{g}(T) = -t_{\mathsf{c}}^{1/2}(\xi + o_{\mathrm{p}}(1)). \tag{9.22}$$

The function $\tilde{g}$ is continuous on $[0,\infty)$ with $\tilde{g}(0) = a > 0$. Consider the two cases separately.

(i): When $\alpha < 1$ we have, by (9.21) and (3.17), $\tilde{g}(t_{\mathsf{c}}) = a + (1+o(1))t_{\mathsf{c}}/r - t_{\mathsf{c}} = a - (1+o(1))a_{\mathsf{c}} < 0$ (for large n), since $a \sim \alpha a_{\mathsf{c}}$. Further, on $[0, t_{\mathsf{c}}]$, using (3.14) and (3.1),

$$\tilde{g}'(t) = (n-a)\tilde{\pi}'(t) - 1 = \frac{n-a}{n}\left(\frac{t}{t_{\mathsf{c}}}\right)^{r-1} e^{-pt} - 1 = \left(\frac{t}{t_{\mathsf{c}}}\right)^{r-1} - 1 + o(1); \tag{9.23}$$

this is negative for $t < (1-\varepsilon)t_{\mathsf{c}}$ for any $\varepsilon > 0$ and large n, and it follows that (for large n, at least), $\tilde{g}$ has a unique root t_* in $[0, t_{\mathsf{c}}]$. It follows from (3.17) and (3.11) that $t_*/t_{\mathsf{c}} \to \varphi(\alpha)$.

Since also $T/t_{\mathsf{c}} \xrightarrow{\mathrm{p}} \varphi(\alpha)$, (9.23) implies that $\tilde{g}'(t) = -(1 - \varphi(\alpha)^{r-1}) + o_p(1)$ for all t between t_* and T, and thus the mean value theorem yields

$$\tilde{g}(T) = \tilde{g}(T) - \tilde{g}(t_*) = (T - t_*)\big(-(1-\varphi(\alpha)^{r-1}) + o_p(1)\big),$$

which together with (9.22) yields, recalling $\varphi(\alpha) < 1$,

$$T - t_* = -\big((1-\varphi(\alpha)^{r-1})^{-1} + o_p(1)\big)\tilde{g}(T) = \big((1-\varphi(\alpha)^{r-1})^{-1}\xi + o_p(1)\big)t_{\mathsf{c}}^{1/2}.$$

The result in (i) follows.

(ii): Let $g(t) := \tilde{g}(t)/(1-\tilde{\pi}(t)) - a$ and $h(t)$ be as in the proof of Lemma 9.5, (9.15)–(9.16). We know that $g(t_{\mathsf{c}}^*) = -a_{\mathsf{c}}^*$ and $g'(t_{\mathsf{c}}^*) = 0$. Further, for $t \sim t_{\mathsf{c}}$, we have by (3.5), (3.13), (3.17) and (3.14),

$$\begin{aligned}
\tilde{\pi}(t) &\sim \frac{(tp)^r}{r!} \sim \frac{t_{\mathsf{c}}}{rn} = o(1),\\
\tilde{\pi}'(t) &\sim \frac{r}{t}\tilde{\pi}(t) \sim \frac{1}{n},\\
\tilde{\pi}''(t) &\sim \frac{r-1}{t}\tilde{\pi}'(t) \sim \frac{r-1}{nt_{\mathsf{c}}}.
\end{aligned}$$

Hence, by (9.16), $h(t) = o(1)$, $h'(t) = (n-t)\pi''(t) \sim (r-1)/t_{\mathsf{c}}$ and

$$g''(t) = \frac{h'(t)}{(1-\tilde{\pi}(t))^2} + 2\frac{h(t)\tilde{\pi}'(t)}{(1-\tilde{\pi}(t))^3} = \frac{r-1}{t_{\mathsf{c}}}\big(1+o(1)\big) + o\Big(\frac{1}{n}\Big) = \frac{r-1}{t_{\mathsf{c}}}\big(1+o(1)\big). \tag{9.24}$$

Consequently, a Taylor expansion yields, for $t \sim t_{\mathsf{c}} \sim t^*_{\mathsf{c}}$,

$$g(t) = -a^*_{\mathsf{c}} + \frac{r-1}{2t_{\mathsf{c}}}(t-t^*_{\mathsf{c}})^2\bigl(1+o(1)\bigr). \tag{9.25}$$

We have $\tilde{g}(t_*) = 0$ and thus $g(t_*) = -a$. Further, (3.17) and (3.11) again yield $t_*/t_{\mathsf{c}} \to \varphi(1) = 1$. Hence, (9.25) yields (3.23).

Since Theorem 3.6 yields $T = t_{\mathsf{c}}\bigl(1+o_{\mathrm{p}}(1)\bigr)$ and $T < t^*_{\mathsf{c}}$ w.h.p., (9.22) yields $g(T) = \tilde{g}(T)/(1-\tilde{\pi}(T)) - a = -a - t_{\mathsf{c}}^{1/2}(\xi + o_{\mathrm{p}}(1))$; thus, similarly, (9.25) yields, using $a^*_{\mathsf{c}} - a \gg t_{\mathsf{c}}^{1/2}$,

$$\begin{aligned} t^*_{\mathsf{c}} - T &= \bigl(1+o_{\mathrm{p}}(1)\bigr)\sqrt{\frac{2t_{\mathsf{c}}}{r-1}(a^*_{\mathsf{c}} - a - t_{\mathsf{c}}^{1/2}(\xi + o_{\mathrm{p}}(1)))} = \bigl(1+o_{\mathrm{p}}(1)\bigr)\sqrt{\frac{2t_{\mathsf{c}}}{r-1}(a^*_{\mathsf{c}} - a)} \\ &= \bigl(1+o_{\mathrm{p}}(1)\bigr)(t^*_{\mathsf{c}} - t_*). \end{aligned}$$

Hence, w.h.p., every t between T and t_* satisfies $t^*_{\mathsf{c}} - t = \bigl(1+o(1)\bigr)(t^*_{\mathsf{c}} - t_*)$, and then by (9.24),

$$g'(t) = \bigl(1+o(1)\bigr)\frac{r-1}{t_{\mathsf{c}}}(t - t^*_{\mathsf{c}}) = -\bigl(1+o(1)\bigr)\frac{r-1}{t_{\mathsf{c}}}(t^*_{\mathsf{c}} - t_*).$$

Finally, the mean value theorem yields, similarly to the case (i),

$$T - t_* = \frac{g(T) - g(t_*)}{-\bigl(1+o(1)\bigr)\frac{r-1}{t_{\mathsf{c}}}(t^*_{\mathsf{c}} - t_*)} = \frac{t_{\mathsf{c}}^{1/2}(\xi + o_{\mathrm{p}}(1))}{\bigl(2(r-1)t_{\mathsf{c}}^{-1}(a^*_{\mathsf{c}} - a)\bigr)^{1/2}},$$

and the result in (ii) follows, recalling $\xi \in N(0, 1/r)$ and $a^*_{\mathsf{c}} \sim a_{\mathsf{c}} = \frac{r-1}{r}t_{\mathsf{c}}$. ☐

Proof of Theorem 4.5. We use the version described in Subsection 4.3 where edges are added at random times. Let $\widehat{U}$ be the time the active set becomes big, i.e., the time the M:th edge is added. For any given p, then $\widehat{U} \le p$ if and only if at time p, the active set is big, which is the same as saying that there is a big active set in $G_{n,p}$. Fix $x \in (-\infty, \infty)$ and choose $p = p^*_{\mathsf{c}} + (r-1)^{1/2}r^{-1}xa^{-1/2}p_{\mathsf{c}}$. Then Theorem 3.7 (with $\lambda = (r-1)^{1/2}r^{-1}x$) yields $\mathbb{P}(\widehat{U} \le p) \to \Phi(x)$. In other words, $(\widehat{U} - p^*_{\mathsf{c}})/((r-1)^{1/2}r^{-1}a^{-1/2}p_{\mathsf{c}}) \xrightarrow{\mathrm{d}} N(0,1)$, or

$$\widehat{U} \in \mathrm{AsN}(p^*_{\mathsf{c}}, (r-1)r^{-2}a^{-1}p_{\mathsf{c}}^2). \tag{9.26}$$

Let $N(u)$ be the number of edges at time u. Then $N(0) = 0$ and, in analogy with Lemma 7.2, $\bigl(N(u) - \binom{n}{2}u\bigr)/(1-u)$ is a martingale on $[0,1)$. Thus, Doob's inequality yields, as in the proof of Lemma 7.3, for any $u_0 \in [0,1]$, $\mathbb{E}\bigl(\sup_{u \le u_0} |N(u) - \binom{n}{2}u|^2\bigr) \le 16\binom{n}{2}u_0 = O(n^2u_0)$, cf. [32, Lemma 3.2]. Hence,

$$\sup_{u \le u_0}\left|N(u) - \binom{n}{2}u\right| = O_{\mathrm{p}}(nu_0^{1/2}).$$

Choosing $u_0 = 2p_{\mathsf{c}}$ we thus obtain, since $\widehat{U} \le 2p_{\mathsf{c}}$ w.h.p.,

$$M = N(\widehat{U}) = \binom{n}{2}\widehat{U} + O_{\mathrm{p}}(np_{\mathsf{c}}^{1/2}). \tag{9.27}$$

We have by (3.12), for some constant $c = c(r)$,

$$\frac{np_{\mathsf{c}}^{1/2}}{n^2 p_{\mathsf{c}}/a^{1/2}} = \frac{a^{1/2}}{np_{\mathsf{c}}^{1/2}} = \frac{a^{1/2} c\big(na^{r-1}\big)^{1/(2r)}}{n} = c\Big(\frac{a}{n}\Big)^{1-1/2r} = o(1).$$

Consequently, the error term in (9.27) is $o_{\mathrm{p}}(n^2 p_{\mathsf{c}}/a^{1/2})$, and the result follows from (9.27) and (9.26). ∎

10 The number of generations

Let $T_0 := 0$ and define inductively

$$T_{j+1} := A(T_j), \qquad j \geq 0. \tag{10.1}$$

Thus $A(T_0) = A(0) = |\mathcal{A}(0)| = |\mathcal{G}_0|$, the size of generation 0 (the initially active vertices). Further, by our choice of u_t as one of the oldest unused, active vertices, $\mathcal{Z}(T_1) = \mathcal{Z}(A(0)) = \mathcal{A}(0) = \mathcal{G}_0$ and $\mathcal{Z}(T_2) = \mathcal{Z}(A(T_1)) = \mathcal{A}(T_1) = \mathcal{G}_0 \cup \mathcal{G}_1$; in general, by induction, all vertices in generation k (and earlier) have been found and declared active at time T_k, and they have been used at time $T_{k+1} = A(T_k)$. In other words,

$$\bigcup_{j=0}^{k} \mathcal{G}_j = \mathcal{A}(T_k) = \mathcal{Z}(T_{k+1}), \qquad k \geq 0.$$

In particular, the size of generation k equals

$$|\mathcal{G}_k| = |\mathcal{Z}(T_{k+1}) \setminus \mathcal{Z}(T_k)| = T_{k+1} - T_k, \qquad k \geq 0,$$

and the number of generations τ defined by (1.1) is

$$\tau = \max\{k \geq 0 : T_{k+1} > T_k\} = \min\{k \geq 1 : T_{k+1} = T_k\} - 1.$$

We begin by considering the supercritical case. We then consider the spread of activation in the bootstrap percolation process in three different stages in each of the following subsections. We first consider the bottleneck when the size is close to t_{c}; we know that this is where the activation will stop in the critical case, and in the slightly supercritical case, the activation will grow slowly here, and this will dominate the total time. Then follows a period of doubly exponential growth, and finally, when there are only $o(n)$ vertices remaining, it may take some time to sweep up the last of them. Recall that Example 3.11 shows that each of the three phases may dominate the two others.

We define, for any $m \leq n$,

$$\tau(m) := \inf\{j : T_j \geq m\}, \tag{10.2}$$

with the interpretation that $\tau(m) = \infty$ if this set of j is empty, i.e., if $m > A^* = T$.

10.1 The bottleneck

We consider first $\tau(3t_{\mathsf{c}})$, i.e., the number of generations required to achieve at least $3t_{\mathsf{c}}$ active vertices. (The constant 3 is chosen for convenience; any constant > 1 would give the same result within $O(1)$ w.h.p.) In the really supercritical case, this is achieved quickly.

Proposition 10.1. *Suppose that $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Assume $a \geq (1+\delta)a_{\mathsf{c}}$ for some $\delta > 0$. Then, w.h.p. $\tau(3t_{\mathsf{c}}) = O(1)$.*

Proof. Lemmas 8.1 and 8.4 imply that uniformly for $0 \leq t \leq 3t_{\mathsf{c}}$, with $x = t/t_{\mathsf{c}}$,

$$A(t) - t = S(t) - t + a = \left(\tfrac{1}{r}x^r - x\right) t_{\mathsf{c}} + a + o_{\mathrm{p}}(t_{\mathsf{c}}) \geq -a_{\mathsf{c}} + (1+\delta)a_{\mathsf{c}} + o_{\mathrm{p}}(a_{\mathsf{c}})$$

and thus w.h.p.

$$A(t) - t \geq \frac{\delta}{2}a_{\mathsf{c}} \geq \frac{\delta}{4}t_{\mathsf{c}}.$$

Hence, in this range, w.h.p. each generation has size at least $(\delta/4)t_{\mathsf{c}}$, and the numbers of generations $\tau(3t_{\mathsf{c}})$ required to reach $3t_{\mathsf{c}}$ is thus w.h.p. bounded by $12/\delta$. $\square$

In the slightly supercritical case when $a \sim a_{\mathsf{c}}$, this part may be a real bottleneck, however. We will approximate $A(t)$ by deterministic functions and begin with a definition: Given a function $F : [0,\infty) \to [0,\infty)$, define the iterates $T^F_{j+1} := F(T^F_j)$ with $T^F_0 := 0$. Thus $T_j = T^A_j$.

Lemma 10.2. *If $A \leq F$, then $T_j \leq T^F_j$ for every j. If $A \geq F$, then $T_j \geq T^F_j$ for every j.*

Proof. By induction. Assume for example $A \leq F$ and $T_j \leq T^F_j$. Then, since A is (weakly) increasing,

$$T_{j+1} = A(T_j) \leq A(T^F_j) \leq F(T^F_j) = T^F_{j+1}.$$

$\square$

We next prove a deterministic lemma.

Lemma 10.3. *Let $a, b, t_0 > 0$, and let $F(t) := t + a + b(t-t_0)^2$. Assume $a \leq t_0$ and $bt_0 \leq 1$. Let N be the smallest integer such that $T^F_N > 2t_0$. Then*

$$N = \bigl(1 + O(bt_0)\bigr) \int_{-t_0}^{t_0} \frac{1}{a + bx^2}\,\mathrm{d}x + O(1).$$

Proof. Assume that $t \in [0, 2t_0]$ and let $\Delta := F(t) - t$. The assumptions on a and b imply $0 < \Delta \leq a + bt_0^2 \leq 2t_0$. For $s \in [t, t+\Delta]$ we have $|F'(s) - 1| = |2b(s-t_0)| \leq 6bt_0$, and thus, by the mean-value theorem, $|F(s) - s - (F(t) - t)| \leq 6bt_0\Delta = 6bt_0(F(t) - t)$. Thus, uniformly for such s, $F(s) - s = (F(t) - t)(1 + O(bt_0))$ and thus $(F(t) - t)^{-1} = (F(s) - s)^{-1}(1 + O(bt_0))$. Consequently,

$$1 = \int_t^{t+\Delta} \frac{1}{F(t) - t}\,\mathrm{d}s = \bigl(1 + O(bt_0)\bigr) \int_t^{t+\Delta} \frac{1}{F(s) - s}\,\mathrm{d}s = \bigl(1 + O(bt_0)\bigr) \int_t^{t+\Delta} \frac{1}{a + b(s-t_0)^2}\,\mathrm{d}s.$$

If $t = T^F_j$, then $t + \Delta = F(t) = T^F_{j+1}$. Summing for $j = 0, \dots, N-1$ we thus obtain

$$N = \bigl(1 + O(bt_0)\bigr) \int_0^{T^F_N} \frac{1}{a + b(s-t_0)^2}\,\mathrm{d}s \geq \bigl(1 + O(bt_0)\bigr) \int_0^{2t_0} \frac{1}{a + b(s-t_0)^2}\,\mathrm{d}s$$

and similarly, omitting $j = N-1$,

$$N-1 \le \bigl(1+O(bt_0)\bigr)\int_0^{2t_0} \frac{1}{a+b(s-t_0)^2}\,\mathrm{d}s.$$

The result follows, using the change of variable $s = x + t_0$. $\square$

Proposition 10.4. *Suppose that $r \ge 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Assume $a/a_{\mathsf{c}} \to 1$ and $a - a^*_{\mathsf{c}} \gg \sqrt{a_{\mathsf{c}}}$. Then,*

$$\tau(3t_{\mathsf{c}}) = \frac{\pi\sqrt{2}+o_{\mathrm{p}}(1)}{\sqrt{r-1}}\Bigl(\frac{t_{\mathsf{c}}}{a-a^*_{\mathsf{c}}}\Bigr)^{1/2}.$$

Proof. By (8.1) and (3.4), $n\pi(3t_{\mathsf{c}}) = O(t_{\mathsf{c}})$. Hence, by Lemmas 7.3 and 9.4, for $t \le 3t_{\mathsf{c}}$,

$$S(t) = \mathbb{E}\,S(t) + O_{\mathrm{p}}\bigl(t_{\mathsf{c}}^{1/2}\bigr) = (n-a)\tilde\pi(t) + O_{\mathrm{p}}\bigl(t_{\mathsf{c}}^{1/2}\bigr). \tag{10.5}$$

Let $H(t) := a + (n-a)\tilde\pi(t) - t$ and define $h := \inf_{t\le 3t_{\mathsf{c}}} H(t)$. Let the infimum be attained at t_*; it follows from (3.17) and Lemma 8.4 that $t_* \sim t_{\mathsf{c}}$, cf. (9.2). We have $H(t_*) = h$, $H'(t_*) = 0$ and, uniformly for $t \le 3t_{\mathsf{c}}$, using (3.14), (3.5) and (3.4),

$$\begin{aligned}H''(t) &= (n-a)\tilde\pi''(t) = (n-a)p^r\frac{t^{r-1}}{(r-1)!}\Bigl(\frac{r-1}{t}-p\Bigr)e^{-pt} = np^r\frac{t^{r-2}}{(r-1)!}\bigl(r-1+o(1)\bigr)\\ &= \left(\frac{t}{t_{\mathsf{c}}}\right)^{r-2}\frac{r-1+o(1)}{t_{\mathsf{c}}} = \frac{r-1}{t_{\mathsf{c}}}\left(1+o(1)+O\Bigl(\frac{|t-t_{\mathsf{c}}|}{t_{\mathsf{c}}}\Bigr)\right).\end{aligned}$$

Hence, by a Taylor expansion, for $0 \le t \le 3t_{\mathsf{c}}$,

$$H(t) = h + \frac{r-1}{2t_{\mathsf{c}}}(t-t_*)^2\left(1+o(1)+O\Bigl(\frac{|t-t_{\mathsf{c}}|}{t_{\mathsf{c}}}\Bigr)\right). \tag{10.6}$$

Notice that in the last two formulas, the term $o(1)$ tends to 0 as $n\to\infty$, uniformly in $t \le 3t_{\mathsf{c}}$, and $O(\dots)$ is uniform in n; these uniformities allow us to combine the two terms in a meaningful way.

On the interval $[0, 3t_{\mathsf{c}}]$, $\tilde\pi(t) = o(1)$ by (3.17) and (3.5), and thus by (3.16)

$$h \sim \inf_{t\le 3t_{\mathsf{c}}}\frac{H(t)}{1-\tilde\pi(t)} = \inf_{t\le 3t_{\mathsf{c}}}\frac{a+(n-a)\tilde\pi(t)-t}{1-\tilde\pi(t)} = a + \inf_{t\le 3t_{\mathsf{c}}}\frac{n\tilde\pi(t)-t}{1-\tilde\pi(t)} = a - a^*_{\mathsf{c}}. \tag{10.7}$$

In particular, by our assumption, $h \gg a_{\mathsf{c}}^{1/2}$. Consequently, by (10.5) and (10.6), for any fixed small $\varepsilon > 0$ and $|t - t_{\mathsf{c}}| \le 2\varepsilon t_{\mathsf{c}}$, w.h.p.

$$A(t) - t = a + S(t) - t = H(t) + o_{\mathrm{p}}(h) = (1+O(\varepsilon))\left(h + \frac{r-1}{2t_{\mathsf{c}}}(t-t_*)^2\right). \tag{10.8}$$

Let $t_1 := (1-\varepsilon)t_* \sim (1-\varepsilon)t_{\mathsf{c}}$ and $t_2 := (1+\varepsilon)t_* \sim (1+\varepsilon)t_{\mathsf{c}}$. For $0 \le t \le t_1$ and $t_2 \le t \le 3t_{\mathsf{c}}$, Lemmas 8.1 and 8.4 imply that w.h.p. $A(t) - t \ge ct_{\mathsf{c}}$, for some constant $c = c(\varepsilon) > 0$. The numbers of generations required to cover the intervals $[0, t_1]$ and $[t_2, 3t_{\mathsf{c}}]$ are thus $O(1/c(\varepsilon))$, so $\tau(3t_{\mathsf{c}}) = \tau'_\varepsilon + O(1/c(\varepsilon))$, where τ'_ε is the number of generations needed to increase the size from at least t_1 to at least t_2. To find τ'_ε, we may redefine T_n by starting with $T_0 := t_1$ and iterate as in (10.1) until we reach t_2. (Note that since A is increasing, if we start with a larger T_0, then every T_n will be larger. Hence, to start with exactly t_1 can only affect τ'_ε by at most 1.) By (10.8) and Lemma 10.2, we may on the interval $[t_1, t_2]$ w.h.p. obtain upper and lower bounds from $F_\pm(t) = t + (1 \pm C\varepsilon)(h + b(t - t_*)^2)$, where $b := (r-1)/(2t_{\mathsf{c}}) > 0$ and C is some constant. Let $t_0 := t_* - t_1 = \varepsilon t_* > 0$. We have $a^*_{\mathsf{c}} \sim a_{\mathsf{c}}$ and by assumption $a \sim a_{\mathsf{c}}$, so by (10.7), $h = o(a_{\mathsf{c}}) = o(t_{\mathsf{c}})$ and thus $h < t_0/2$ for large n. Furthermore, $bt_0 = O(\varepsilon t_*/t_{\mathsf{c}}) = O(\varepsilon)$. If ε is small enough, we thus have $bt_0 \le 1/2$ and, by a translation $t \mapsto t - t_1$, Lemma 10.3 applies to both F_+ and F_- and yields, w.h.p., using (10.7),

$$
\begin{aligned}
\tau'_\varepsilon &= \bigl(1 + O(\varepsilon)\bigr) \int_{-\varepsilon t_*}^{\varepsilon t_*} \frac{\mathrm{d}x}{h + bx^2} + O(1) = \bigl(1 + O(\varepsilon)\bigr) \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{h + bx^2} + O\left(\frac{1}{b\varepsilon t_*}\right) + O(1) \\
&= \bigl(1 + O(\varepsilon)\bigr) \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{h + bx^2} + O\left(\frac{1}{\varepsilon}\right) = \bigl(1 + O(\varepsilon)\bigr) \frac{\pi}{(hb)^{1/2}} + O\left(\frac{1}{\varepsilon}\right) \\
&= \bigl(1 + O(\varepsilon)\bigr) \Bigl(\frac{2t_{\mathsf{c}}}{(r-1)(a - a^*_{\mathsf{c}})}\Bigr)^{1/2} \pi + O(1/\varepsilon).
\end{aligned}
$$

Since $t_{\mathsf{c}}/(a - a^*_{\mathsf{c}}) \to \infty$, it follows that for every $\varepsilon > 0$, w.h.p., with $c'(\varepsilon) := \min(c(\varepsilon), \varepsilon) > 0$,

$$
\tau(3t_{\mathsf{c}}) = \tau'_\varepsilon + O(1/c(\varepsilon)) = \bigl(1 + O(\varepsilon)\bigr) \frac{\pi\sqrt{2}}{\sqrt{r-1}} \Bigl(\frac{t_{\mathsf{c}}}{a - a^*_{\mathsf{c}}}\Bigr)^{1/2} + O(1/c'(\varepsilon)) = \frac{\pi\sqrt{2} + O(\varepsilon)}{\sqrt{r-1}} \Bigl(\frac{t_{\mathsf{c}}}{a - a^*_{\mathsf{c}}}\Bigr)^{1/2}.
$$

The result follows since $\varepsilon > 0$ is arbitrary. $\square$

Remark 10.5. In the critical case $(a - a^*_{\mathsf{c}})/\sqrt{a_{\mathsf{c}}} \to y \in (-\infty, \infty)$, we can use a minor variation of the same argument, now using Lemma 9.3, where $h \sim a - a_{\mathsf{c}}$ above is replaced by the random

$$
h' = a - a_{\mathsf{c}} + t_{\mathsf{c}}^{1/2} W(1/r) + o_{\mathrm{p}}(t_{\mathsf{c}}^{1/2}) = r^{-1/2} t_{\mathsf{c}}^{1/2} \bigl(y\sqrt{r-1} + \xi + o_{\mathrm{p}}(1)\bigr),
$$

where $\xi \sim N(0,1)$. We have $\tau(3t_{\mathsf{c}}) < \infty \iff A^* \ge 3t_{\mathsf{c}} \iff h' > 0$; this is w.h.p. equivalent to $y\sqrt{r-1} + \xi > 0$. (This thus happens with probability $\Phi((r-1)^{1/2} y) + o(1)$, as stated in Theorem 3.6(iii).) The argument above then shows that conditioned on $\tau(3t_{\mathsf{c}}) < \infty$ (i.e., on $A^* \ge 3t_{\mathsf{c}}$),

$$
\tau(3t_{\mathsf{c}})/t_{\mathsf{c}}^{1/4} \xrightarrow{\mathrm{d}} \Bigl(\frac{2^{1/2}\pi r^{1/4}}{\sqrt{r-1}} \bigl(\xi + y\sqrt{r-1}\bigr)^{-1/2} \Bigm| \xi + y\sqrt{r-1} > 0\Bigr).
$$

In particular, then $\tau(3t_{\mathsf{c}}) = \Theta_p\bigl(t_{\mathsf{c}}^{1/4}\bigr)$.

Note that in the supercritical case in Proposition 10.4, the time $\tau(3t_{\mathsf{c}})$ is always smaller than $t_{\mathsf{c}}^{1/4}$, but that it approaches the order $t_{\mathsf{c}}^{1/4}$ when $a - a^*_{\mathsf{c}}$ grows only a little faster than the critical value $a_{\mathsf{c}}^{1/2}$. Hence, we can say that the worst possible number of generations to pass the bottleneck at t_{c} is of the order $t_{\mathsf{c}}^{1/4}$.

10.2 The doubly exponential growth

We next consider the growth from size $3t_{\mathsf{c}}$ up to $1/p$. We will show that in this range, the growth is doubly exponential. Again, we approximate $A(t)$ by deterministic functions.

Define for any $\delta \in \mathbb{R}$, cf. (3.4),

$$F_\delta(t) := n\frac{(tp)^r}{r!}(1+\delta) = \left(\frac{t}{t_{\mathsf{c}}}\right)^{r-1}\frac{t}{r}(1+\delta). \tag{10.9}$$

Lemma 10.6. *For every $\delta > 0$, there are positive constants ε and K such that w.h.p. $F_{-\delta}(t) \le A(t) \le F_\delta(t)$ for all $t \in [K(t_{\mathsf{c}}+a), \varepsilon/p]$.*

Proof. By (8.1) and (10.9), for $Kt_{\mathsf{c}} \le t \le \varepsilon/p$ (with $\varepsilon \le 1$), if n is large enough so $t_{\mathsf{c}} \ge 1$,

$$\pi(t) = \frac{(tp)^r}{r!}\bigl(1+O(\varepsilon+K^{-1})\bigr) = \frac{1}{n}F_0(t)\bigl(1+O(\varepsilon+K^{-1})\bigr).$$

We may thus choose ε and K such that for all such t (and large n)

$$F_{-\delta/4}(t) \le n\pi(t) \le F_{\delta/4}(t). \tag{10.10}$$

For $t \ge K(t_{\mathsf{c}}+a)$, (10.9) implies

$$F_0(t) = \left(\frac{t}{t_{\mathsf{c}}}\right)^{r-1}\frac{t}{r} \ge K^r\frac{a}{r},$$

so choosing K large enough, we have $a \le (\delta/4)F_0(t)$ for all $t \in [K(t_{\mathsf{c}}+a), \varepsilon/n]$, and thus by (10.10)

$$F_{-\delta/4}(t) - a \le \mathbb{E}\,S(t) = (n-a)\pi(t) \le F_{\delta/4}(t) \le F_{\delta/2}(t) - a.$$

Hence, by Chebyshev's inequality, using (2.13) and (10.10),

$$\begin{aligned}\mathbb{P}\bigl\{A(t) \notin [F_{-3\delta/4}(t), F_{3\delta/4}(t)]\bigr\} &= \mathbb{P}\bigl\{S(t) \notin [F_{-3\delta/4}(t)-a, F_{3\delta/4}(t)-a]\bigr\} \\ &\le \frac{n\pi(t)}{(\delta F_0(t)/4)^2} \le \frac{F_{\delta/4}(t)}{(\delta F_0(t)/4)^2} = \frac{16(1+\delta/4)}{\delta^2 F_0(t)}.\end{aligned} \tag{10.11}$$

Define $t_j := (1+\delta/5)^{j/r}K(t_{\mathsf{c}}+a)$. Then, (10.11) and (10.9) show that, assuming as we may $\delta \le 1$,

$$\begin{aligned}\sum_{j\ge 0:\, t_j \le \varepsilon/p}\mathbb{P}\bigl\{A(t_j) \notin [F_{-3\delta/4}(t_j), F_{3\delta/4}(t_j)]\bigr\} &\le \sum_{j\ge 0}\frac{20}{\delta^2 F_0(t_j)} = \sum_{j\ge 0}\frac{20}{\delta^2 F_0(t_0)}(1+\delta/5)^{-j} \\ &= \frac{100(1+\delta/5)}{\delta^3 F_0(t_0)} \to 0,\end{aligned}$$

since, using (10.9) again and (3.5),

$$F_0(t_0) = F_0(K(t_{\mathsf{c}}+a)) \ge F_0(t_{\mathsf{c}}) = \frac{t_{\mathsf{c}}}{r} \to \infty.$$

Consequently, w.h.p. $A(t_j) \in [F_{-3\delta/4}(t_j), F_{3\delta/4}(t_j)]$ for all $j \ge 0$ with $t_j \le \varepsilon/p$. However, if $t_j \le t \le t_{j+1}$, then $F_0(t_j) \le F_0(t) \le F_0(t_{j+1}) = (1+\delta/5)F_0(t_j)$, and it follows that, since both $A(t)$ and $F_0(t)$ are monotone, w.h.p.

$$(1+\delta/5)^{-1}F_{-3\delta/4}(t) \le A(t) \le (1+\delta/5)F_{3\delta/4}(t)$$

for all $t \in [K(t_{\mathsf{c}}+a), (1+\delta/5)^{-1/r}\varepsilon/p]$, which, provided δ is small and ε is replaced by $\varepsilon/2$, say, yields the result. $\square$

Proposition 10.7. *Suppose that $r \geq 2$ and $n^{-1} \ll p \ll n^{-1/r}$. Then w.h.p., when $A^* \geq 3t_{\mathsf{c}}$,*

$$\tau(1/p) - \tau(3t_{\mathsf{c}}) = \frac{1}{\log r}\left(\log\log(np) - \log_+ \log\frac{a}{a_{\mathsf{c}}}\right) + O(1).$$

Proof. Choose a fixed $0 < \delta < 1$, and choose ε and K as in Lemma 10.6. (In this proof, we do not have to let $\delta \to 0$, so we can take $\delta = 1/2$, say.) First, $\tau(K(t_{\mathsf{c}} + a)) - \tau(3t_{\mathsf{c}})$, the number of generations from $3t_{\mathsf{c}}$ to $K(t_{\mathsf{c}} + a)$, is w.h.p. $O(1)$. Indeed, after $\tau(3t_{\mathsf{c}})$ generations we have at least $\max(3t_{\mathsf{c}}, a)$ active vertices, and in each of the following generations until well beyond $K(t_{\mathsf{c}} + a)$, the number is w.h.p. multiplied by at least 1.3, say, by the proof of Lemma 8.2 or by Lemmas 9.1, 9.4 and 7.3. Similarly, $\tau(1/p) - \tau(\varepsilon/p) \leq 1$ w.h.p., arguing as in Case 3 of the proof of Lemma 8.2.

Consequently it suffices to consider $\tau(\varepsilon/p) - \tau(K(t_{\mathsf{c}} + a))$. We define iterates $T_j^{F_\delta}$ as in Subsection 10.1 by $T_{j+1}^{F_\delta} := F_\delta(T_j^{F_\delta})$, $j \geq 0$, but now starting with $T_0^{F_\delta} := K(t_{\mathsf{c}} + a)$. Further, let

$$N_\delta := \min\{j \geq 0 : T_j^{F_\delta} \geq \varepsilon/p\}. \tag{10.12}$$

By Lemma 10.6 we may assume that $F_{-\delta}(t) \leq A(t) \leq F_\delta(t)$ for all $t \in [K(t_{\mathsf{c}}+a), \varepsilon/p]$, and then, by induction as in Lemma 10.2, $T_j^{F_{-\delta}} \leq T_{j+\tau(K(t_{\mathsf{c}}+a))} \leq T_{j+1}^{F_\delta}$ for all $j \geq 0$ with $T_{j-1+\tau(K(t_{\mathsf{c}}+a))} \leq \varepsilon/p$. Consequently, w.h.p.

$$N_{-\delta} \geq \tau(\varepsilon/p) - \tau(K(t_{\mathsf{c}} + a)) \geq N_\delta - 1. \tag{10.13}$$

To find N_δ, rewrite (10.9) as

$$\frac{F_\delta(t)}{c_\delta t_{\mathsf{c}}} = \left(\frac{t}{c_\delta t_{\mathsf{c}}}\right)^r,$$

where $c_\delta := (r/(1+\delta))^{1/(r-1)}$. Iterating we see that, for $j \geq 0$,

$$\frac{T_j^{F_\delta}}{c_\delta t_{\mathsf{c}}} = \left(\frac{T_0^{F_\delta}}{c_\delta t_{\mathsf{c}}}\right)^{r^j} = \left(\frac{K(t_{\mathsf{c}} + a)}{c_\delta t_{\mathsf{c}}}\right)^{r^j},$$

and thus

$$\log\left(\frac{T_j^{F_\delta}}{c_\delta t_{\mathsf{c}}}\right) = r^j \log\left(\frac{K(t_{\mathsf{c}} + a)}{c_\delta t_{\mathsf{c}}}\right),$$

and

$$j \log r = \log\log\left(\frac{T_j^{F_\delta}}{c_\delta t_{\mathsf{c}}}\right) - \log\log\left(\frac{K(t_{\mathsf{c}} + a)}{c_\delta t_{\mathsf{c}}}\right).$$

Consequently,

$$N_\delta = \left\lceil \left(\log\log\left(\frac{\varepsilon/p}{c_\delta t_{\mathsf{c}}}\right) - \log\log\left(\frac{K(t_{\mathsf{c}} + a)}{c_\delta t_{\mathsf{c}}}\right)\right) \Big/ \log r \right\rceil. \tag{10.14}$$

In order to simplify this, note that, using (3.1),

$$\log\left(\frac{\varepsilon/p}{c_\delta t_{\mathsf{c}}}\right) = \log\left(\frac{1}{pt_{\mathsf{c}}}\right) + O(1) = \frac{1}{r-1}\log(np) + O(1), \tag{10.15}$$

and thus

$$\log\log\left(\frac{\varepsilon/p}{c_\delta t_{\mathsf{c}}}\right) = \log\log(np) + O(1). \tag{10.16}$$

Further, we may assume that $a \ge a_{\mathsf{c}}/2 \ge t_{\mathsf{c}}/4$, since otherwise the process is subcritical and $A^* < 3t_{\mathsf{c}}$ w.h.p. by Theorem 3.1. Hence, $\log(K(t_{\mathsf{c}} + a)) = \log a + O(1)$ and thus, since also $\log(c_\delta t_{\mathsf{c}}) = \log a_{\mathsf{c}} + O(1)$,

$$\log\left(\frac{K(t_{\mathsf{c}} + a)}{c_\delta t_{\mathsf{c}}}\right) = \log a - \log a_{\mathsf{c}} + O(1) = \log\frac{a}{a_{\mathsf{c}}} + O(1). \tag{10.17}$$

We may assume that $K \ge ec_\delta$, so $\log\bigl(K(t_{\mathsf{c}} + a)/(c_\delta t_{\mathsf{c}})\bigr) \ge 1$, and then (10.17) yields

$$\log\log\left(\frac{K(t_{\mathsf{c}} + a)}{c_\delta t_{\mathsf{c}}}\right) = \log_+\log\frac{a}{a_{\mathsf{c}}} + O(1). \tag{10.18}$$

Finally, (10.14), (10.16) and (10.18) yield

$$N_\delta \log r = \log\log(np) - \log_+\log\frac{a}{a_{\mathsf{c}}} + O(1).$$

Note that the right-hand side depends on δ only in the error term $O(1)$. Hence, we have the same result for $N_{-\delta}$, and the result follows by (10.13) and the comments at the beginning of the proof. $\square$

10.3 The final stage

We finally consider the evolution after $1/p$ vertices have become active. We let, as in Section 8, $b^* := b_{\mathsf{c}}\omega(n)$ where $\omega(n) \to \infty$ slowly; we assume that $b^* \ll 1/p$ (which is possible since $pb_{\mathsf{c}} \to 0$ by (3.5)). By Remark 8.3, $\tau(n - b^*) \le \tau(1/p) + 3$ w.h.p., so it suffices to consider the evolution when less than b^* vertices remain.

Let $\mathcal{F}_t := \sigma\{I_i(s) : 1 \le i \le n,\ 1 \le s \le t\}$ be the σ-field describing the evolution up to time t.

Lemma 10.8. *For any t and u with $0 \le t \le t + u \le n$, the conditional distribution of $A(t + u) - A(t) = S(t + u) - S(t)$ given $\mathcal{F}_t$ is $\mathrm{Bin}\bigl(n - A(t), \pi(t; u)\bigr)$, where*

$$\pi(t; u) := \frac{\pi(t + u) - \pi(t)}{1 - \pi(t)}. \tag{10.19}$$

If further $n - b^* \le t \le t + u \le n$, then, uniformly in all such t and u,

$$\pi(t; u) = pu\bigl(1 + o(1)\bigr). \tag{10.20}$$

Proof. Conditioned on $\mathcal{F}_t$, $A(t)$ is a given number, and of the $n-a$ summands in (2.8), $n-a-S(t) = n - A(t)$ are zero. For any of these terms, the probability that it changes from 0 at time t to 1 at time $t + u$ is, by (2.11),

$$\mathbb{P}(Y_i \le t + u \mid Y_i > t) = \frac{\mathbb{P}(t < Y_i \le t + u)}{\mathbb{P}(Y_i > t)} = \frac{\pi(t + u) - \pi(t)}{1 - \pi(t)} = \pi(t; u).$$

Hence, the conditional distribution of $S(t+u)-S(t)$ is $\mathrm{Bin}\big(n-A(t), \pi(t;u)\big)$.

To see the approximation (10.20), note first that for $n-b^* \le t \le n$, since we assume $pb^* \to 0$, we have $b^* \ll 1/p \ll n$ so $t \sim n$. Hence, using again $pb^* \to 0$ and recalling the notation b'_{c} from (3.7),

$$\pi(t+1)-\pi(t) = \mathbb{P}(Y_1 = t+1) = \binom{t}{r-1} p^r (1-p)^{t+1-r} \sim \frac{n^{r-1}}{(r-1)!} p^r (1-p)^n = \frac{pb'_{\mathsf{c}}}{n}. \tag{10.21}$$

Furthermore, cf. (3.7), still for $n-b^* \le t \le n$,

$$1-\pi(t) = \mathbb{P}\big(\mathrm{Bin}(t,p) \le r-1\big) \sim \mathbb{P}\big(\mathrm{Bin}(t,p) = r-1\big) \sim \frac{n^{r-1}}{(r-1)!} p^{r-1} (1-p)^n = \frac{b'_{\mathsf{c}}}{n}. \tag{10.22}$$

Consequently, $\pi(t+u)-\pi(t) = \big(1+o(1)\big) upb'_{\mathsf{c}}/n$ and

$$\pi(t;u) = \frac{\pi(t+u)-\pi(t)}{1-\pi(t)} = \big(1+o(1)\big) \frac{upb'_{\mathsf{c}}/n}{b'_{\mathsf{c}}/n} = \big(1+o(1)\big) up. \qquad \square$$

Lemma 10.9. *Suppose that $r \ge 2$, $n^{-1} \ll p \ll n^{-1/r}$ and $a = o(n)$. If $b_{\mathsf{c}} \to \infty$ and $n-b^* \le t \le n$, then $A(t) = n - b_{\mathsf{c}}\big(1+o_{\mathrm{p}}(1)\big)$; in particular, $n - A(t) < 2b_{\mathsf{c}}$ w.h.p.*

Proof. We have, using (2.12) and (10.22), since $b_{\mathsf{c}} \to \infty$ implies $b'_{\mathsf{c}} \sim b_{\mathsf{c}}$,

$$\mathbb{E}\big(n-A(t)\big) = n - a - \mathbb{E}\, S(t) = (n-a)(1-\pi(t)) \sim (n-a)\frac{b'_{\mathsf{c}}}{n} \sim b_{\mathsf{c}}$$

and similarly, using (2.13),

$$\mathrm{Var}\big(n-A(t)\big) = \mathrm{Var}\, S(t) \le (n-a)(1-\pi(t)) \sim b_{\mathsf{c}}.$$

Thus, by Chebyshev's inequality, since $b_{\mathsf{c}} \to \infty$,

$$n - A(t) = \big(1+o(1)\big) b_{\mathsf{c}} + O_{\mathrm{p}}\big(b_{\mathsf{c}}^{1/2}\big) = \big(1+o_{\mathrm{p}}(1)\big) b_{\mathsf{c}}. \qquad \square$$

Proposition 10.10. *Suppose that $r \ge 2$, $n^{-1} \ll p \ll n^{-1/r}$ and $a = o(n)$. Then, when $A^* \ge 3t_{\mathsf{c}}$,*

$$\tau - \tau(1/p) = \big(1+o(1)\big) \frac{\log n}{np} + O_{\mathrm{p}}(1).$$

In particular, if further $p \ge c\log(n)/n$ for some $n \ge 0$, then $\tau - \tau(1/p) = O_{\mathrm{p}}(1)$.

Furthermore, when $A^* = n$, w.h.p. $\tau - \tau(1/p) \le 3$.

Proof. By Remark 8.3, after $\tau(1/p)+3$ generations, the active size is $T_{\tau(1/p)+3} \ge n - b^*$ w.h.p.

If $b_{\mathsf{c}} \to 0$, we can choose $b^* = 1/2$, so w.h.p. $T_{\tau(1/p)+3} = n$ and $\tau \le \tau(1/p) + 3$.

More generally, if $b_{\mathsf{c}} = O(1)$, we have by (10.21)

$$\mathbb{E}\big(S(n) - S(n-b^*)\big) \le n\big(\pi(n) - \pi(n-b^*)\big) \sim nb^* \frac{pb'_{\mathsf{c}}}{n} = pb^* b_{\mathsf{c}} = O(pb^*) = o(1).$$

Hence, w.h.p. $S(n) = S(n-b^*)$, which means that no further activations occur after $n-b^*$. Consequently, in this case too w.h.p. $\tau = \tau(n-b^*) \le \tau(1/p)+3$. In particular, this proves that $\tau \le \tau(1/p)+3$ w.h.p. when $A^* = n$, since w.h.p. $A^* < n$ if $b_{\mathsf{c}} \to \infty$ by Theorem 3.2.

Further, when $b_{\mathsf{c}} = O(1)$, (3.10) implies that $np \ge \log n$ for large n, so $\log n/(np) \le 1$ and the result holds in this case.

Now assume that $b_{\mathsf{c}} \to \infty$. For convenience, we modify the counting of generations and start at $t = n-b^*$, regarding the active but unused vertices at $n-b^*$ as "generation 0". (We may assume that b^* is an integer.) Thus define, recursively,

$$\begin{aligned}
T'_0 &:= n-b^*,\\
T'_{j+1} &:= A(T'_j), \qquad j \ge 0,\\
\Delta_j &:= T'_{j+1} - T'_j = A(T'_j) - T'_j,\\
\tau' &:= \max\{j \ge 0 : \Delta_j > 0\}.
\end{aligned}$$

Since w.h.p. $T_{\tau(1/p)-1} \le \max(1/p, a) < n-b^* \le T_{\tau(1/p)+3}$, it follows by induction that $T_{\tau(1/p)-1+j} \le T'_j \le T_{\tau(1/p)+3+j}$, $j \ge 0$, and thus w.h.p.

$$\tau' + \tau(1/p) - 1 \le \tau \le \tau' + \tau(1/p) + 3. \tag{10.23}$$

Consequently, it suffices to estimate τ'.

By Lemma 10.8, conditioned on $\mathcal{F}_{T'_j}$ (i.e., on T'_j and the evolution up to T'_j, which in particular specifies $A(T'_j)$), for large n,

$$\mathbb{E}\big(\Delta_{j+1} \mid \mathcal{F}_{T'_j}\big) = \big(n - A(T'_j)\big)\pi(T'_j;\Delta_j) \le \big(n - A(T'_0)\big)2p\Delta_j$$

and thus, by induction, since $\Delta_0 \le n - T'_0 = b^*$,

$$\mathbb{E}\big(\Delta_j \mid \mathcal{F}_{T'_0}\big) \le \big(2\big(n-A(T'_0)\big)p\big)^j\Delta_0 \le \big(2\big(n-A(T'_0)\big)p\big)^j b^*. \tag{10.24}$$

Further, Lemma 10.9 yields $n - A(T'_0) = n - A(n-b^*) < 2b_{\mathsf{c}}$ w.h.p. Consequently, (10.24) implies, w.h.p. for all $j \ge 0$ (simultaneously),

$$\mathbb{E}\big(\Delta_j \mid \mathcal{F}_{T'_0}\big) \le \big(4pb_{\mathsf{c}}\big)^j b^*. \tag{10.25}$$

Recall that $pb_{\mathsf{c}} \to 0$ by (3.5), so we may assume $4pb_{\mathsf{c}} < 1$. If j is chosen such that $(4pb_{\mathsf{c}})^j b^* \to 0$, then (10.25) implies that w.h.p. $\Delta_j = 0$ and thus $\tau' < j$. Hence, for any $\omega' = \omega'(n) \to \infty$, w.h.p.

$$\tau' \le \frac{\log b^*}{|\log(pb_{\mathsf{c}}) + \log 4|} + \omega'(n),$$

which is another way of saying [34, Lemma 3],

$$\tau' \le \frac{\log b^*}{|\log(pb_{\mathsf{c}}) + \log 4|} + O_{\mathrm{p}}(1) = \frac{\log b^*}{|\log(pb_{\mathsf{c}})|}\big(1+o(1)\big) + O_{\mathrm{p}}(1). \tag{10.26}$$

For a lower bound, fix ε with $0 < \varepsilon < 1$ and define the deterministic numbers Δ_j^- by

$$\Delta_j^- := (1-\varepsilon)^{j+1}(pb_{\mathsf{c}})^j b^*. \tag{10.27}$$

Let $\omega'' := 1/(pb_{\mathsf{c}}) \to \infty$. We claim that w.h.p.

$$\Delta_j \geq \Delta_j^- \qquad \text{for all } j \geq 0 \text{ such that } \quad \Delta_j^- \geq \omega''. \tag{10.28}$$

By our assumption $4pb_{\mathsf{c}} < 1$, we have $\Delta_{j+1}^-/\Delta_j^- < 1/4$, so $\Delta_j^- \to 0$ geometrically fast.

By Lemma 10.9 and $b_{\mathsf{c}}/b^* \to 0$, w.h.p.

$$\Delta_0 = A(T_0') - (n - b^*) = A(T_0') - n + b^* \geq b^* - 2b_{\mathsf{c}} \geq (1-\varepsilon)b^* = \Delta_0^-,$$

so (10.28) holds w.h.p. for $j = 0$.

Say that $j \geq 0$ is *good* if $\Delta_j \geq \Delta_j^-$ and *fat* if $A(T_j') > n - (1 - \varepsilon/4)b_{\mathsf{c}}$. Let $j \geq 0$. At time T_j' we have $A(T_j') - T_j' = \Delta_j$ active but unused vertices. Further, by Lemma 10.8 we have, conditioned on $\mathcal{F}_{T_j'}$ (which specifies both T_j' and Δ_j)

$$\Delta_{j+1} = T_{j+2}' - T_{j+1}' = A(T_j' + \Delta_j) - A(T_j') \in \operatorname{Bin}\bigl(n - A(T_j'), \pi(T_j'; \Delta_j)\bigr).$$

By Lemma 10.8, $\pi(T_j'; \Delta_j) = p\Delta_j\bigl(1 + o(1)\bigr) \geq p\Delta_j(1 - \varepsilon/4)$ for n large, so if j is good but not fat,

$$\mathbb{E}(\Delta_{j+1} \mid \mathcal{F}_{T_j'}) = \bigl(n - A(T_j')\bigr)\pi(T_j'; \Delta_j) \geq (1-\varepsilon/4)^2 b_{\mathsf{c}} p \Delta_j \geq (1-\varepsilon/2) b_{\mathsf{c}} p \Delta_j^- \geq (1+\varepsilon/2)\Delta_{j+1}^-$$

and Chebyshev's inequality yields, since $x \mapsto x/(x-a)^2$ is decreasing for $x > a$,

$$\begin{aligned}\mathbb{P}\bigl(\Delta_{j+1} < \Delta_{j+1}^- \mid \mathcal{F}_{T_j'}\bigr) &\leq \frac{\operatorname{Var}(\Delta_{j+1} \mid \mathcal{F}_{T_j'})}{\bigl(\mathbb{E}(\Delta_{j+1} \mid \mathcal{F}_{T_j'}) - \Delta_{j+1}^-\bigr)^2} \leq \frac{\mathbb{E}(\Delta_{j+1} \mid \mathcal{F}_{T_j'})}{\bigl(\mathbb{E}(\Delta_{j+1} \mid \mathcal{F}_{T_j'}) - \Delta_{j+1}^-\bigr)^2} \\ &\leq \frac{(1+\varepsilon/2)\Delta_{j+1}^-}{\bigl(\varepsilon\Delta_{j+1}^-/2\bigr)^2} = O\left(\frac{1}{\Delta_{j+1}^-}\right).\end{aligned}$$

Say that j is *bad* if j is not good and that j *fails* if j is fat or bad. Then, by stopping at the first j that fails we see that

$$\begin{aligned}\mathbb{P}(\text{some } j \leq \omega'' \text{ fails}) &\leq \mathbb{P}(\text{some } j \leq \omega'' \text{ is fat}) + \mathbb{P}(0 \text{ is bad}) \\ &\qquad + \sum_{j>0\colon \Delta_j^- \geq \omega''} \mathbb{P}(j \text{ is bad} \mid j-1 \text{ is good and not fat}) \\ &\leq \mathbb{P}\bigl(A(n) > n - (1-\varepsilon/4)b_{\mathsf{c}}\bigr) + o(1) + \sum_{j\colon \Delta_j^- \geq \omega''} O\left(\frac{1}{\Delta_j^-}\right) = o(1),\end{aligned}$$

since $A(n) < n - (1-\varepsilon/4)b_{\mathsf{c}}$ w.h.p. by Lemma 10.9 and the final sum is $O(1/\omega'') = o(1)$ because the terms $1/\Delta_j^-$ increase geometrically so the sum is dominated by its largest (and last) term.

We have shown that w.h.p., if $\Delta_j^- \geq \omega''$, then $\Delta_j \geq \Delta_j^- > 0$ and thus $\tau' \geq j$. Hence, by (10.28) and (10.27), w.h.p.

$$\tau' \geq \left\lfloor \frac{\log((1-\varepsilon)b^*/\omega'')}{|\log((1-\varepsilon)pb_{\mathsf{c}})|} \right\rfloor = \frac{\log b^*}{|\log(pb_{\mathsf{c}})|}\bigl(1 + o(1)\bigr) + O(1). \tag{10.29}$$

Combining the upper bound (10.26) and the lower bound (10.29), we find

$$\tau' = \frac{\log b^*}{|\log(pb_{\mathsf{c}})|}\bigl(1+o(1)\bigr) + O_{\mathrm{p}}(1). \tag{10.30}$$

By (3.3), $\log(pb_{\mathsf{c}}) = -(np - r\log(np) + O(1))$ and

$$\log n \geq \log b^* \geq \log b_{\mathsf{c}} \geq \log n - pn - O(1).$$

Hence, finally (10.30) yields

$$\tau' = \frac{\log n + O(np)}{np - r\log(np) + O(1)}\bigl(1+o(1)\bigr) + O_{\mathrm{p}}(1) = \frac{\log n}{np}\bigl(1+o(1)\bigr) + O_{\mathrm{p}}(1).$$

The result now follows from (10.23). □

11 Proofs of Theorems 5.2, 5.6, 5.8

We prove in this section Theorems 5.2, 5.6 and 5.8 related to the boundary cases. We consider first the case $p \sim c/n$.

Proof of Lemma 5.1. By the implicit function theorem, at least locally, the root $x_0(\theta)$ is smooth except at points where $f(x,c,\theta) = \frac{\partial}{\partial x} f(x,c,\theta) = 0$. We begin by studying such *critical points.*

Let $g(y) := \mathbb{P}(\mathrm{Po}(y) \leq r-1) = 1 - \psi(y)$, cf. (3.13). Differentiations yield

$$g'(y) = -\mathbb{P}\bigl(\mathrm{Po}(y) = r-1\bigr) = -\frac{y^{r-1}}{(r-1)!}e^{-y}, \tag{11.1}$$

$$g''(y) = \Bigl(\frac{r-1}{y} - 1\Bigr)g'(y) = \frac{r-1-y}{y}g'(y). \tag{11.2}$$

We have, see (5.2), $f(x,c,\theta) = 1 - x - (1-\theta)g(cx)$ and thus $\frac{\partial}{\partial x} f(x,c,\theta) = -1 - c(1-\theta)g'(cx)$. Hence, $f(x,c,\theta) = \frac{\partial}{\partial x} f(x,c,\theta) = 0$ if and only if

$$\begin{cases} (1-\theta)g(cx) = 1-x, \\ c(1-\theta)g'(cx) = -1, \end{cases}$$

which imply $g(cx) = -c(1-x)g'(cx)$ and thus

$$c = cx - \frac{g(cx)}{g'(cx)}. \tag{11.3}$$

Let $h(y) := y - g(y)/g'(y)$, $y > 0$, so (11.3) says $c = h(cx)$. Then, by (11.2),

$$h'(y) = 1 - \frac{g'(y)}{g'(y)} + \frac{g(y)g''(y)}{g'(y)^2} = \frac{r-1-y}{y} \cdot \frac{g(y)}{g'(y)}.$$

Since $g(y) > 0$ and $g'(y) < 0$ for $y > 0$, h has a global minimum at $y = r - 1$, and the minimum value is

$$\min_{y>0} h(y) = h(r-1) = r - 1 - \frac{g(r-1)}{g'(r-1)} = r + \frac{\mathbb{P}\big(\mathrm{Po}(r-1) \le r-2\big)}{\mathbb{P}\big(\mathrm{Po}(r-1) = r-1\big)} = c_{\mathsf{c}}.$$

Furthermore, $h(y) > y \to \infty$ as $y \to \infty$, and $h(y) \to \infty$ as $y \to 0$ too, because then $g(y) \to 1$ and $g'(y) \to 0$.

Consequently, if $0 \le c < c_{\mathsf{c}}$, then (11.3) has no solution $x > 0$, and thus there is no critical point. If $c = c_{\mathsf{c}}$, there is exactly one $x > 0$ satisfying (11.3) (viz., $x = (r-1)/c_{\mathsf{c}}$), and if $c > c_{\mathsf{c}}$, there are two. Since (11.3) implies $c > cx$, these roots are in $(0, 1)$.

To complete the proof, it is perhaps simplest to rewrite (5.3) as $\theta = \vartheta(x)$, with

$$\vartheta(x) := 1 - (1 - x)/g(cx). \tag{11.4}$$

Since $g(y) > 0$ for $y \ge 0$, ϑ is a smooth function on $[0, 1]$, with $\vartheta(0) = 0$ and $\vartheta(1) = 1$. Moreover, $f(x, c, \theta) = g(cx)\big(\theta - \vartheta(x)\big)$, which implies that

$$f(x, c, \theta) = \frac{\partial}{\partial x} f(x, c, \theta) = 0 \iff \theta = \vartheta(x) \text{ and } \vartheta'(x) = 0.$$

Consequently, by the results above, if $c < c_{\mathsf{c}}$, then $\vartheta' \ne 0$ so $\vartheta'(x) > 0$ for $x \ge 0$. In this case, ϑ is strictly increasing and thus a bijection $[0, 1] \to [0, 1]$, and x_0 is its inverse.

If $c = c_{\mathsf{c}}$, then $\vartheta' = 0$ only at a single point, and it follows again that ϑ is a strictly increasing function and x_0 is its inverse.

If $c > c_{\mathsf{c}}$, then $\vartheta'(x) = 0$ at two values x_1 and x_2 with $0 < x_1 < x_2 < 1$ and $cx_1 < r - 1 < cx_2$. It can be seen, e.g. using (11.2), that $\vartheta''(x_1) < 0 < \vartheta''(x_2)$, and thus ϑ is *decreasing* on the interval $[x_1, x_2]$. The result follows, with $\theta_{\mathsf{c}} = \vartheta(x_1)$, $\theta_{\mathsf{c}}^- = \max(\vartheta(x_2), 0)$ and $x_0(\theta_{\mathsf{c}}) = x_1$. (Note that $\vartheta(x_2) = \min_{x\in[0,1]} \vartheta(x) < 0$ if c is large enough.) ☐

Remark 11.1. If $c > c_{\mathsf{c}}$, then thus $x_0(\theta_{\mathsf{c}}) = x_1$ is the smallest root of $\vartheta'(x) = 0$, or equivalently $x_1 = y_1/c$ where y_1 is the smallest root of $h(y) = c$; further, $\theta_{\mathsf{c}} = \vartheta(x_1)$ while $x_0(\theta_{\mathsf{c}}+)$ is the other root of $\vartheta(x) = \vartheta(x_1)$.

If $c = c_{\mathsf{c}}$, we have $y_1 = r - 1$ and thus $x_1 = (r - 1)/c_{\mathsf{c}}$ and, using (11.4) and (5.4),

$$\theta_{\mathsf{c}}(c_{\mathsf{c}}) = \vartheta\Big(\frac{r-1}{c_{\mathsf{c}}}\Big) = 1 - \frac{1 - (r-1)/c_{\mathsf{c}}}{g(r-1)} = 1 - \frac{1}{r\,\mathbb{P}(\mathrm{Po}(r-1) = r-1) + \mathbb{P}(\mathrm{Po}(r-1) \le r-2)}.$$

For $c > c_{\mathsf{c}}$, the two roots $x_1(c)$ and $x_2(c)$ of $\vartheta'(x) = 0$ are smooth functions of c, and thus

$$\frac{\mathrm{d}\theta_{\mathsf{c}}}{\mathrm{d}c} = \frac{\mathrm{d}}{\mathrm{d}c}\vartheta(x_1(c)) = \frac{\partial\vartheta}{\partial c}(x_1(c)) + \frac{\partial\vartheta}{\partial x}(x_1(c))x_1'(c) = \frac{\partial\vartheta}{\partial c}(x_1(c)) < 0,$$

where the last inequality follows from (11.4), and similarly $\mathrm{d}\theta_{\mathsf{c}}^-/\mathrm{d}c < 0$. Hence, $\theta_{\mathsf{c}}(c)$ and $\theta_{\mathsf{c}}^-(c)$ are decreasing functions of c, as claimed in Remark 5.3.

Lemma 11.2. *Suppose that $r \ge 2$, $p = O(1/n)$ and $tp = o(1)$. Then $S(t) = o_{\mathrm{p}}(t)$.*

Proof. We may assume $1 \le t \le 1/p$. (Note that $S(t) = 0$ for $t < r$.) Then $\pi(t) = O(t^r p^r) = o(tp)$ by (8.1), and thus the expected number of activated vertices is $\mathbb{E}\, S(t) = (n-a)\pi(t) = o(npt) = o(t)$. ☐

Proof of Theorem 5.2. First, in (i) and (ii), $ap \to \theta c = 0$. Let $\varepsilon > 0$. Taking $t = (1+\varepsilon)a$ in Lemma 11.2, we find w.h.p. $S((1+\varepsilon)a) < \varepsilon a$ and thus

$$A\bigl((1+\varepsilon)a\bigr) = a + S\bigl((1+\varepsilon)a\bigr) < (1+\varepsilon)a,$$

whence $A^* = T < (1+\varepsilon)a$. Consequently, $1 \le A^*/a < 1+\varepsilon$ w.h.p., proving (i) and (ii).

Next, by (2.9), Lemma 7.1 and (2.12), uniformly for all $t \ge 0$,

$$A(t) = a + S(t) = a + \mathbb{E}\, S(t) + o_{\mathrm{p}}(n) = (n-a)\pi(t) + a + o_{\mathrm{p}}(n)$$

and thus, using also (3.15),

$$n^{-1}A(t) = (1-\theta)\pi(t) + \theta + o_{\mathrm{p}}(1) = (1-\theta)\tilde{\pi}(t) + \theta + o_{\mathrm{p}}(1).$$

Substituting $t = xn$, we find by (3.13), since $tp = xc + o(x)$, uniformly in all $x \ge 0$,

$$n^{-1}A(xn) = (1-\theta)\,\mathbb{P}\bigl(\mathrm{Po}(tp) \ge r\bigr) + \theta + o_{\mathrm{p}}(1) = (1-\theta)\,\mathbb{P}\bigl(\mathrm{Po}(cx) \ge r\bigr) + \theta + o_{\mathrm{p}}(1).$$

and, recalling (5.1), still uniformly in $x \ge 0$,

$$n^{-1}\bigl(A(xn) - xn\bigr) = f(x,c,\theta) + o_{\mathrm{p}}(1). \tag{11.5}$$

Let $\varepsilon > 0$. Since $f(x,c,\theta) > 0$ for $x \in [0, x_0(\theta))$, and thus by compactness $f(\cdot,c,\theta)$ is bounded from below on $[0, x_0(\theta) - \varepsilon]$, (11.5) implies that w.h.p. $A(xn) - xn > 0$ on $[0, x_0(\theta) - \varepsilon]$, and thus $T > (x_0(\theta) - \varepsilon)n$. Furthermore, both in (iii) and in (iv) with $\theta \ne \theta_{\mathsf{c}}$, we have $\frac{\partial}{\partial x} f(x_0(\theta), c, \theta) \ne 0$ and thus, if $\varepsilon > 0$ is small enough, $f(x_0(\theta) + \varepsilon, c, \theta) < 0$, so (11.5) implies that w.h.p. $A((x_0(\theta) + \varepsilon)n) < (x_0(\theta) + \varepsilon)n$ and thus $T < (x_0(\theta) + \varepsilon)n$. ∎

The proof of Theorem 5.5 is very similar to the one of Theorem 3.6. We first give a more precise estimate of the process $S(t)$, which is the analogue of Lemma 9.3 in the case $p = c/n$. However, in this case, we get a Brownian bridge because here we consider a large part of the distribution of Y_i.

Lemma 11.3. *Suppose $r \ge 2$, $p = c/n$ and $a \sim \theta n$ with $c > 0$ and $0 < \theta < 1$. Then*

$$Z(x) := \frac{S(xn) - \mathbb{E}\, S(xn)}{\sqrt{(1-\theta)n}} \xrightarrow{\mathrm{d}} W_0\bigl(\psi(cx)\bigr) \tag{11.6}$$

in $D[0,1]$, where W_0 is a Brownian bridge and $\psi(y) := \mathbb{P}(\mathrm{Po}(y) \ge r)$ as in (3.13).

Proof. Let $\tilde{S}(u) := \sum_{i=1}^{n-a} \mathbf{1}\{U_i \le u\}$, $0 \le u \le 1$, where $U_i \in U(0,1)$ are i.i.d. By (2.8) and (2.11), we have $S(t) \overset{\mathrm{d}}{=} \tilde{S}(\pi(t))$, jointly for all $t \ge 0$. Further, $\frac{1}{n-a}\tilde{S}(u)$, $u \in [0,1]$, is the empirical distribution function of $U_1, \dots, U_{n-a}$, and thus by [17, Theorem 16.4], in $D[0,1]$,

$$\frac{\tilde{S}(u) - \mathbb{E}\,\tilde{S}(u)}{\sqrt{n-a}} \xrightarrow{\mathrm{d}} W_0(u).$$

Furthermore, by (3.15) and (3.13),

$$\pi(xn) = \tilde{\pi}(xn) + O(1/n) = \psi(xnp) + O(1/n) = \psi(cx) + O(1/n),$$

uniformly for $x \geq 0$, and it follows, using the continuity of W_0, that

$$\frac{S(xn) - \mathbb{E}\, S(xn)}{\sqrt{n-a}} \overset{\mathrm{d}}{=} \frac{\tilde{S}(\pi(xn)) - \mathbb{E}\, \tilde{S}(\pi(xn))}{\sqrt{n-a}} \xrightarrow{\mathrm{d}} W_0\bigl(\psi(cx)\bigr)$$

in $D[0,1]$, which proves the result since $n - a \sim (1-\theta)n$. □

Proof of Theorem 5.5. It suffices to consider a such that $a \sim \theta_{\mathsf{c}} n$. By (2.12), (3.15) and (3.13),

$$\mathbb{E}\, S(xn) = (n-a)\pi(xn) = (n-a)\tilde{\pi}(xn) + O(1) = (n-a)\psi(cx) + O(1). \tag{11.7}$$

By the Skorohod coupling theorem [36, Theorem 4.30], we may assume that the processes for different n are coupled such that the limit (11.6) in Lemma 11.3 holds a.s., and not just in distribution. Since convergence in $D[0,1]$ to a continuous function is equivalent to uniform convergence, this means that a.s. $Z(x) \to W_0(\psi(cx))$ uniformly for $x \in [0,1]$. Hence, we have, using (11.7) and (5.1),

$$\begin{aligned} A(xn) - xn &= a + S(xn) - xn = a + \mathbb{E}\, S(xn) + \sqrt{(1-\theta_{\mathsf{c}})n}\, Z(x) - xn \\ &= a + (n-a)\psi(cx) + \sqrt{(1-\theta_{\mathsf{c}})n}\, Z(x) - xn + O(1) \\ &= (a - \theta_{\mathsf{c}} n)(1-\psi(cx)) + nf(x,c,\theta_{\mathsf{c}}) + \sqrt{(1-\theta_{\mathsf{c}})n}\, Z(x) + O(1) \\ &= nf(x,c,\theta_{\mathsf{c}}) + (a-\theta_{\mathsf{c}} n)(1-\psi(cx)) + \sqrt{(1-\theta_{\mathsf{c}})n}\, W_0(\psi(cx)) + o_{\mathrm{p}}(n^{1/2}), \end{aligned} \tag{11.8}$$

uniformly for $x \in [0,1]$.

We first use (11.8) to derive the simple estimate

$$A(xn) - xn = nf(x,c,\theta_{\mathsf{c}}) + o_{\mathrm{p}}(n), \tag{11.9}$$

uniformly for $x \in [0,1]$. By Lemma 5.1, $f(x,c,\theta_{\mathsf{c}}) = 0$ for $x = x_0$ or $x = x_1$, with $f(x,c,\theta_{\mathsf{c}}) > 0$ for $x \in [0,x_0) \cup (x_0,x_1)$ and $f(x,c,\theta_{\mathsf{c}}) < 0$ for $x \in (x_1,1]$. Hence, for any fixed small $\varepsilon > 0$, (11.9) implies that w.h.p. $A(xn) - xn > 0$ for $x \in [0, x_0 - \varepsilon] \cup [x_0+\varepsilon, x_1-\varepsilon]$ and $A(xn) - xn < 0$ for $x \in [x_1+\varepsilon, 1]$, and hence $T \in [x_0-\varepsilon, x_0+\varepsilon] \cup [x_1-\varepsilon, x_1+\varepsilon]$. It follows by a standard argument that there exists a sequence $\varepsilon_n \searrow 0$ such that w.h.p.

$$A^* = T \in [x_0 - \varepsilon_n, x_0+\varepsilon_n] \cup [x_1-\varepsilon_n, x_1+\varepsilon_n].$$

Moreover, w.h.p. $T \in [x_0-\varepsilon_n, x_0+\varepsilon_n]$ if and only if $\inf_{[x_0-\varepsilon_n, x_0+\varepsilon_n]}\bigl(A(xn) - xn\bigr) < 0$. (We may also assume that ε_n is so small that $\varepsilon_n < x_0$ and $2\varepsilon_n < x_1 - x_0$.)

For $x \in [x_0-\varepsilon_n, x_0+\varepsilon_n]$, we have by (11.8) again, and the continuity of ψ and W_0,

$$A(xn) - xn = nf(x,c,\theta_{\mathsf{c}}) + (a-\theta_{\mathsf{c}} n)\bigl(1-\psi(cx_0) + o(1)\bigr) + \sqrt{(1-\theta_{\mathsf{c}})n}\, W_0(\psi(cx_0)) + o_{\mathrm{p}}(n^{1/2}). \tag{11.10}$$

Further, $f(x_0,c,\theta_{\mathsf{c}}) = 0$ and $f(x,c,\theta_{\mathsf{c}}) \geq 0$ for $x \in [x_0-\varepsilon_n, x_0+\varepsilon_n]$, and thus (11.10) yields

$$\inf_{x\in[x_0-\varepsilon_n, x_0+\varepsilon_n]}\bigl(A(xn) - xn\bigr) = (a-\theta_{\mathsf{c}} n)\bigl(1-\psi(cx_0) + o(1)\bigr) + \sqrt{(1-\theta_{\mathsf{c}})n}\, W_0(\psi(cx_0)) + o_{\mathrm{p}}(n^{1/2}). \tag{11.11}$$

The cases (i) and (ii) are easily derived. We thus focus on (iii). We then have, from (11.11),

$$n^{-1/2} \inf_{x\in[x_0-\varepsilon_n,x_0+\varepsilon_n]} \bigl(A(xn)-xn\bigr) = y\bigl(1-\psi(cx_0)\bigr) + \sqrt{1-\theta_{\mathsf{c}}}\, W_0(\psi(cx_0)) + o_{\mathrm{p}}(1)$$

and thus, since $(1-\psi(cx_0))^{-1}\sqrt{1-\theta_{\mathsf{c}}}\, W_0(\psi(cx_0)) \in N(0,\sigma^2)$, where $\sigma^2 = (1-\theta_{\mathsf{c}})\psi(cx_0)/(1-\psi(cx_0)) > 0$,

$$\begin{aligned}\mathbb{P}\left(\inf_{x\in[x_0-\varepsilon_n,x_0+\varepsilon_n]} \bigl(A(xn)-xn\bigr) < 0\right) &= \mathbb{P}\left(y\bigl(1-\psi(cx_0)\bigr) + \sqrt{1-\theta_{\mathsf{c}}}\, W_0(\psi(cx_0)) < 0\right) + o_{\mathrm{p}}(1)\\ &= 1-\Phi(y/\sigma) + o_{\mathrm{p}}(1).\end{aligned}$$

The result follows. □

To prove Theorem 5.6 ($p \sim cn^{-1/r}$), we first show using the previous results that if we can activate $\omega(n)\to\infty$ vertices, then the activation spreads w.h.p. to the entire graph. It remains to show that starting with a finite number of active vertices, the process activates $\omega(n)$ vertices with a probability bounded away from 0 and 1. This will be done using a branching process argument.

Lemma 11.4. *Suppose that $p \ge cn^{-1/r}$ for some $c>0$. If $\omega(n)\to\infty$, then w.h.p. $A(t)>t$ for all t with $\omega(n)\le t\le n-1$.*

Proof. This is easy to prove directly, but we prefer to view it as a corollary of our estimates for smaller p. Thus, let $p' := \omega(n)^{-1/2r}n^{-1/r}$. We may assume $\omega(n)\le n$ and then $n^{-1} \ll p' \ll n^{-1/r}$, so $p'<p$, at least for large n, and we may assume that $G_{n,p'} \subseteq G_{n,p}$. We may consider bootstrap percolation on $G_{n,p'}$ and $G_{n,p}$ simultaneously, with the same initial set A_0 of size a; we use the description in Section 2, starting with families of i.i.d. random indicators $I'_i(s) \in \mathrm{Be}(p')$ and $I_i(s) \in \mathrm{Be}(p)$ where we may assume $I'_i(s) \le I_i(s)$. Then, using $'$ to denote variables for $G_{n,p'}$, $S'(t)\le S(t)$ and $A'(t)\le A(t)$.

We apply Lemma 8.2 to $G_{n,p'}$. The critical time for $G_{n,p'}$ is, see (3.1),

$$t'_{\mathsf{c}} = O\Bigl(\bigl(n(p')^r\bigr)^{-1/(r-1)}\Bigr) = \omega(n)^{1/2(r-1)} = o(\omega(n)).$$

Further, $p' \ge n^{-3/2r} \ge n^{-3/4}$ so, by (3.3), $b'_{\mathsf{c}} \to 0$ and we may choose $b^{*\prime}$ with $b^{*\prime}\to 0$. Hence, Lemma 8.2 shows that w.h.p. $A(t) \ge A'(t) > t$ for $t\in[3t'_{\mathsf{c}}, n-b^{*\prime}]$, and the result follows since, for large n, $3t'_{\mathsf{c}} \le \omega(n)$ and $n-b^{*\prime} > n-1$. □

Proof of Theorem 5.6. For (ii), we apply Lemma 11.4 (if necessary with a smaller c). Taking $\omega(n)=a$, we see that w.h.p. $A(t)>t$ for all $t\in[a,n-1]$. Since also $A(t)\ge a$, we have $A(t)>t$ for all $t\le n-1$, and thus $A^*=T=n$.

For (i) suppose $r\ge 2$, $p\sim cn^{-1/r}$ and let $a\ge r$ be some constant. The probability that a vertex is activated at a given time k is by (2.7)

$$\mathbb{P}(Y_1=k) = \binom{k-1}{r-1}p^r(1-p)^{k-r} \sim \binom{k-1}{r-1}\frac{c^r}{n}. \tag{11.12}$$

For any fixed K, the random variables

$$X_k := A(k) - A(k-1) = S(k) - S(k-1) = \sum_{i \notin \mathcal{A}(0)} \mathbf{1}\{Y_i = k\}, \qquad k = 1, \dots, K,$$

form together with

$$X_{K+1} := n - a - A(K) = \sum_{i \notin \mathcal{A}(0)} \mathbf{1}\{Y_i > K\}$$

a random vector with the multinomial distribution $\operatorname{Mul}\bigl(n-a, (p_k)_{k=1}^{K+1}\bigr)$ with $p_k = \mathbb{P}(Y_1 = k)$, $k \le K$, and $p_{K+1} = \mathbb{P}(Y_1 \ge K+1)$. By (11.12), $(n-a)p_k \to \binom{k-1}{r-1} c^r$ for $k \le K$, and it follows that X_k for $k \le K$ have a joint Poisson limit:

$$(X_k)_{k=1}^K \xrightarrow{\mathrm{d}} (\xi_k)_{k=1}^K \quad \text{with} \quad \xi_k \in \operatorname{Po}\left(\binom{k-1}{r-1} c^r\right) \quad \text{independent.} \tag{11.13}$$

Using the notation of Remark 5.7 we thus obtain

$$A(k) \xrightarrow{\mathrm{d}} a + \sum_{j=1}^{k} \xi_j = a + k + \widetilde{S}_k \quad \text{for } k = 1, \dots, t, \text{ jointly.}$$

and thus $\mathbb{P}(T = k) \to \mathbb{P}(\widetilde{T} = k)$ for $k \le K$ and $\mathbb{P}(T > K) \to \mathbb{P}(\widetilde{T} > K)$.

Since K is arbitrary, we have shown $\mathbb{P}(A^* = k) = \mathbb{P}(T = k) \to \mathbb{P}(\widetilde{T} = k) = \zeta(a, c, k)$ for every finite $k \ge 1$. Furthermore, $\mathbb{P}(T > K) - \mathbb{P}(\widetilde{T} > K) \to 0$ for any fixed K, and a standard argument shows that there exists a sequence $K_n \to \infty$ such that $\mathbb{P}(T > K_n) - \mathbb{P}(\widetilde{T} > K_n) \to 0$, and thus $\mathbb{P}(T > K_n) \to \mathbb{P}(\widetilde{T} = \infty)$. On the other hand, Lemma 11.4 with $\omega(n) = K_n$ shows that $P(K_n \le T < n) \to 0$. Consequently, $\mathbb{P}(T = n) = \mathbb{P}(T > K_n) + o(1) \to \mathbb{P}(\widetilde{T} = \infty) = \zeta(a, c)$.

It is clear that $\zeta(a, c, k) = \mathbb{P}(\widetilde{T} = k) > 0$ for every $k \ge a$. To see that also $\zeta(a, c) = \mathbb{P}(\widetilde{T} = \infty) > 0$, note that, see (11.13), $\mathbb{E}\,\xi_k = \binom{k-1}{r-1} c^r \to \infty$ as $k \to \infty$. Hence, there is some K_0 such that $\mathbb{E}\,\xi_{K_0} > 1$. Since ξ_k stochastically dominates ξ_{K_0} for $k \ge K_0$, it follows that if the process reaches K_0 without stopping, the continuation dominates (up to a change of time) a Galton–Watson branching process with offspring distribution ξ_{K_0}, which is supercritical and thus has a positive probability of living forever. Hence, $\mathbb{P}(\widetilde{T} = \infty) > 0$. $\square$

Proof of Theorem 5.8. It suffices to consider $a = r$. Thus assume $a = r$, and consider the vertices activated in the first generation, i.e., at time $t = r$. There are $S(r) \in \operatorname{Bin}(n - r, p^r)$ such vertices. (Note that, see (2.11), $\pi(r) = \mathbb{P}(\operatorname{Bin}(r, p) = r) = p^r$.) Consequently, $\mathbb{E}\,S(r) = (n-r)p^r \to \infty$. Let $\omega(n) = \mathbb{E}\,S(r)/2$, so $\omega(n) \to \infty$. It follows from Chebyshev's inequality (or Chernoff's) that w.h.p. $S(r) > \omega(n)$. Hence, w.h.p. for all $t \in [r, \omega(n)]$, $A(t) \ge A(r) > S(r) > \omega(n) \ge t$. Together with the trivial $A(t) \ge a = r > t$ for $t < r$ and Lemma 11.4, this shows that w.h.p. $A(t) > t$ for all $t \le n-1$, and thus $A^* = T = n$. $\square$

References

[1] M. Aizenman and J. L. Lebowitz, Metastability effects in bootstrap percolation. *J. Phys. A* **21** (1988), no. 19, 3801–3813.

[2] H. Amini, Bootstrap percolation and diffusion in random graphs with given vertex degrees. *Electron. J. Combin.* **17** (2010), R25.

[3] B. von Bahr and A. Martin-Löf, Threshold limit theorems for some epidemic processes. *Adv. Appl. Probab.* **12** (1980), no. 2, 319–349.

[4] F. Ball and T. Britton, An epidemic model with exposure-dependent severities. *J. Appl. Probab.* **42** (2005), no. 4, 932–949.

[5] F. Ball and T. Britton, An epidemic model with infector and exposure dependent severity. *Math. Biosci.* **218** (2009), no. 2, 105–120.

[6] J. Balogh and B. Bollobás, Bootstrap percolation on the hypercube. *Probab. Theory Rel. Fields* **134** (2006), no. 4, 624–648.

[7] J. Balogh, B. Bollobás, H. Duminil-Copin and R. Morris, The sharp threshold for bootstrap percolation in all dimensions. *Trans. Amer. Math. Soc.*, to appear. arXiv:1010.3326v2

[8] J. Balogh, B. Bollobás and R. Morris, Majority bootstrap percolation on the hypercube. *Combin. Probab. Computing* **18** (2009), 17–51.

[9] J. Balogh, B. Bollobás and R. Morris, Bootstrap percolation in three dimensions. *Ann. Probab.* **37** (2009), 1329–1380.

[10] J. Balogh, B. Bollobás and R. Morris, Bootstrap percolation in high dimensions. *Combin. Probab. Computing* **19** (2010), 643–692.

[11] J. Balogh, B. Bollobás and R. Morris, Graph bootstrap percolation. Preprint, 2011. arXiv:1107.1381

[12] J. Balogh, B. Bollobás, R. Morris, and O. Riordan, Linear algebra and bootstrap percolation. Preprint, 2011. arXiv:1107.1410

[13] J. Balogh, Y. Peres and G. Pete, Bootstrap percolation on infinite trees and non-amenable groups. *Combin. Probab. Comput.* **15** (2006), no. 5, 715–730.

[14] J. Balogh and G. Pete, Random disease on the square grid. *Random Structures Algorithms* **13** (1998), no. 3-4, 409–422.

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

[16] A.D. Barbour, L. Holst and S. Janson, *Poisson Approximation.* Oxford Univ. Press, Oxford, 1992.

[17] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[18] B. Bollobás, Weakly k-saturated graphs. In *Beiträge zur Graphentheorie (Kolloquium, Manebach, 1967)*, pp. 25–31, Teubner, Leipzig, 1968.

[19] B. Bollobás, *The Art of Mathematics. Coffee Time in Memphis.* Cambridge University Press, New York, 2006.

[20] L. Carroll, *The Hunting of the Snark.* Macmillan, London, 1876.

[21] R. Cerf and E.N.M. Cirillo, Finite size scaling in three-dimensional bootstrap percolation. *Ann. Probab.* **27** (1999), 1837–1850.

[22] R. Cerf and F. Manzo, The threshold regime of finite volume bootstrap percolation. *Stochastic Process. Appl.* **101** (2002), no. 1, 69–82.

[23] R. Cerf and F. Manzo, A d-dimensional nucleation and growth model. Preprint, 2010. arXiv:1001.3990

[24] R. Cerf and F. Manzo, Nucleation and growth for the Ising model in d dimensions at very low temperatures. Preprint, 2011. arXiv:1102.1741

[25] J. Chalupa, P. L. Leath and G. R. Reich, Bootstrap percolation on a Bethe latice. *J. Phys. C.* **12** (1979), L31–L35.

[26] H. Duminil-Copin and A. C. D. Van Enter, Sharp metastability threshold for an anisotropic bootstrap percolation model. Preprint, 2010. arXiv:1010.4691

[27] L. R. Fontes, R. H. Schonmann and V. Sidoravicius, Stretched exponential fixation in stochastic Ising models at zero temperature. *Comm. Math. Phys.* **228** (2002), no. 3, 495–518.

[28] J. Gravner, A.E. Holroyd and R. Morris, A sharper threshold for bootstrap percolation in two dimensions. *Probab. Theory Rel. Fields*, to appear. arXiv:1002.3881v2

[29] A. Gut, *Probability: A Graduate Course.* Springer, New York, 2005.

[30] A. E. Holroyd, Sharp metastability threshold for two-dimensional bootstrap percolation. *Probab. Theory Rel. Fields* **125** (2003), no. 2, 195–224.

[31] A.E. Holroyd, T.M. Liggett and D. Romik. Integrals, partitions, and cellular automata. *Trans. Amer. Math. Soc.* **356** (2004), 3349–3368.

[32] S. Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics.* Mem. Amer. Math. Soc., **111**, no. 534, Amer. Math. Soc., Providence, R.I., 1994.

[33] S. Janson, On percolation in random graphs with given vertex degrees. *Electronic J. Probab.* **14** (2009), Paper 5, 86–118.

[34] S. Janson, Probability asymptotics: notes on notation. Institute Mittag-Leffler Report 12, 2009 spring.

[35] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs.* Wiley, New York, 2000.

[36] O. Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.

[37] R. Kozma, M. Puljic, P. Balister, B. Bollobás, and W. Freeman, Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions. *Biol. Cybernet.* **92** (2005) no. 6, 367–379.

[38] A. Martin-Löf, Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Probab.* **23** (1986), no. 2, 265–282.

[39] A. Martin-Löf, The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *J. Appl. Probab.* **35** (1998), no. 3, 671–682.

[40] R. Morris, Minimal percolating sets in bootstrap percolation. *Electron. J. Combin.* **16** (2009), no. 1, Research Paper 2, 20 pp.

[41] R. Morris, Zero-temperature Glauber dynamics on Z^d. *Probab. Theory Rel. Fields* **149** (2011), 417–434.

[42] G.-P. Scalia-Tomba, Asymptotic final-size distribution for some chain-binomial processes. *Adv. Appl. Probab.* **17** (1985), no. 3, 477–495.

[43] R.H. Schonmann, On the behaviour of some cellular automata related to bootstrap percolation. *Ann. Probab.* **20** (1992), 174–193.

[44] T. Sellke, On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Probab.* **20** (1983), no. 2, 390–394.

[45] T. Tlusty and J.-P. Eckmann, Remarks on bootstrap percolation in metric networks. (2009) *J. Phys. A: Math. Theor.* **42**, 205004.

[46] T. Turova, The emergence of connectivity in neuronal networks: from bootstrap percolation to auto-associative memory. *Brain Research*, to appear.

[47] T. Turova and A. Villa, On a phase diagram for random neural networks with embedded spike timing dependent plasticity. *BioSystems* **89** (2007), 280–286.

[48] T. Vallier, *Random graph models and their applications.* Ph. D. thesis, Lund University, 2007.