

# DEPTH-FIRST SEARCH PERFORMANCE IN A RANDOM DIGRAPH WITH GEOMETRIC OUTDEGREE DISTRIBUTION

PHILIPPE JACQUET AND SVANTE JANSON

ABSTRACT. We present an analysis of the depth-first search algorithm in a random digraph model with independent outdegrees having a geometric distribution.

The results include asymptotic results for the depth profile of vertices, the height (maximum depth) and average depth, the number of trees in the forest, the size of the largest and second-largest trees, and the numbers of arcs of different types in the depth-first jungle. Most results are first order. For the height we show an asymptotic normal distribution.

This analysis proposed by Donald Knuth in his next to appear volume of *The Art of Computer Programming* gives interesting insight in one of the most elegant and efficient algorithm for graph analysis due to Tarjan.

## 1. INTRODUCTION

The motivation of this paper is a new section in Donald Knuth's *The Art of Computer Programming* [12], which is dedicated to Depth-First Search (DFS) in a digraph. Briefly, the DFS starts with an arbitrary vertex, and explores the arcs from that vertex one by one. When an arc is found leading to a vertex that has not been seen before, the DFS explores the arcs from that vertex in the same way, in a recursive fashion, before returning to the next arc from its parent. This eventually yields a tree containing all descendants of the the first vertex (which is the root of the tree). If there still are some unseen vertices, the DFS starts again with one of them and finds a new tree, and so on until all vertices are found. We refer to [12] for details as well as for historical notes. (See also S1–S2 in Section 4.) Note that the digraphs in [12] and here are multi-digraphs, where loops and multiple arcs are allowed. (Although in our random model these are few and usually not important.) The DFS algorithm generates a spanning forest (the *depth-first forest*) in the digraph, with all arcs in the forest directed away from the roots. Our main purpose is to study the properties of the depth-first forest, starting with a random digraph  $G$ ; in particular we study the distribution of the depth of vertices in the depth-first forest.

The random digraph model that we consider (following Knuth [12]) has  $n$  vertices and a given outdegree distribution  $\mathbf{P}$ , which in the main part of the present paper is a geometric distribution  $\text{Ge}(1 - p)$  for some fixed  $0 < p < 1$ . The outdegrees (number of outgoing arcs) of the  $n$  vertices are independent random numbers with this distribution. The endpoint of each arc is uniformly selected at random among the  $n$  vertices, independently of all other arcs. (Therefore, an arc can loop back

---

*Date:* 30 December, 2022.

*Key words and phrases.* depth-first search, depth-first forest, random digraph, search depth, depth profile, height.

Supported by the Knut and Alice Wallenberg Foundation.

We thank Donald Knuth for posing us questions and conjectures that led to the present paper.

to the starting vertex, and multiple arcs can occur.) We consider asymptotics as  $n \rightarrow \infty$  for a fixed outdegree distribution.

In the present paper, we study the case of a geometric outdegree distribution in detail; we also (in Section 3) briefly give corresponding results for the shifted geometric outdegree distribution  $\text{Ge}_1(1-p)$ , and discuss the similarities and differences between the two cases. The case of a general outdegree distribution (with finite variance) will be studied in a forthcoming paper [10], where we use a somewhat different method which allows us to extend many (but not all) of the results in the present paper and obtain similar, but partly weaker, results; see also Section 4. One reason for studying the geometric case separately is that its lack-of-memory property leads to interesting features and simplifications not present for general outdegree distributions; this is seen both in [12] and in the proofs and results below. In particular, the depth process studied in Section 2 will be a Markov chain, which is the basis of our analysis.

In addition to studying the depth-first forest, we give also (in Section 2.5) some results on the number of edges of different types in the *depth-first jungle*; this is defined in [12] as the original digraph with arcs classified by the DFS algorithm into the following five types, see Figure 1 for examples:

- *loops*;
- *tree arcs*, the arcs in the resulting depth-first forest;
- *back arcs*, the arcs which point to an ancestor of the current vertex in the current tree;
- *forward arcs*, the arcs which point to an already discovered descendant of the current vertex in the current tree;
- *cross arcs*, all other arcs (these point to an already discovered vertex which is neither a descendant nor an ancestor of the current vertex, and might be in another tree).

(See further the exercises in [12].)

**Remark 1.1.** Some related results for DFS in an undirected Erdős–Rényi graph  $G(n, \lambda/n)$  are proved by Enriquez, Faraud and Ménard [5] and Diskin and Krivelevich [4], and DFS in a random Erdős–Rényi digraph has been studied for example in the proof of [13, Theorem 3]. These models are closely related to our model with a Poisson outdegree distribution  $\mathbf{P}$ ; they will therefore be further discussed in [10].

**Remark 1.2.** We consider only the case of a fixed outdegree distribution  $\mathbf{P}$ . The results can be extended to distributions  $\mathbf{P}_n$  depending on  $n$ , under suitable conditions. This is particularly interesting in the critical case, with expectations  $\lambda_n \rightarrow 1$ ; however, this is out of the scope of the present paper.

The main results for a geometric out-degree distribution are stated and proved in Section 2. We analyze the process  $d(t)$  of depths of the vertices, in the order they are found by the DFS. For a geometric outdegree distribution (but not in general),  $d(t)$  is a Markov chain, and we find its first-order limit by a martingale argument; moreover, we show Gaussian fluctuations. We also find results for the numbers of different types of arcs defined above; this includes verifying some conjectures from previous versions of [12].

In Section 3 we study briefly the case of a shifted geometric outdegree distribution. The same method as in Section 2 works in this case too, but the explicit results are somewhat different. One motivation for this section is to show that some of the

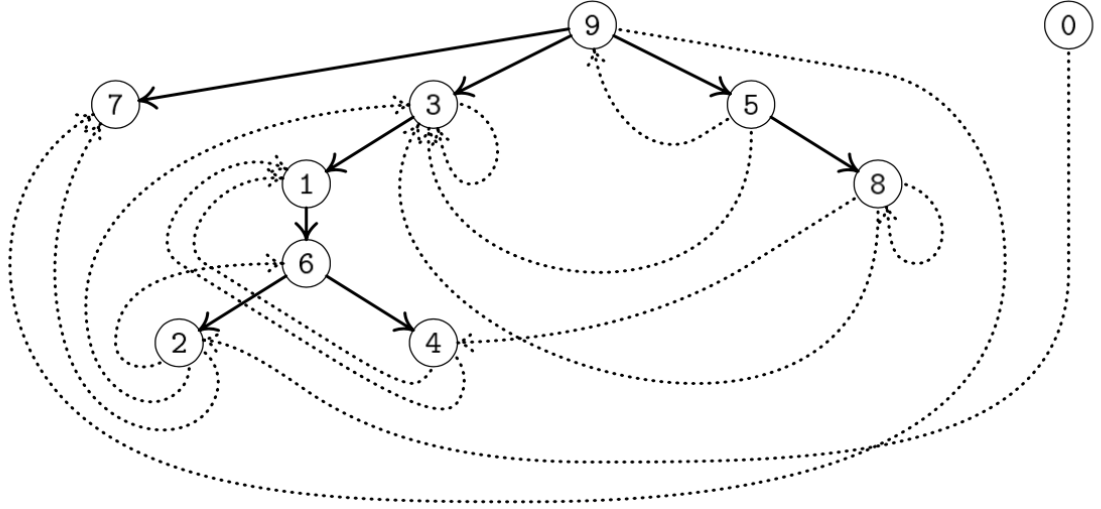


FIGURE 1. Example of a depth-first forest (jungle) from [12], by courtesy of Donald Knuth. Tree arcs are solid (*e.g.*  $9 \rightarrow 3$ ). For example,  $3 \dashrightarrow 3$  is a loop,  $2 \dashrightarrow 3$  is a back arc,  $9 \dashrightarrow 7$  is a forward arc,  $8 \dashrightarrow 4$  and  $0 \dashrightarrow 2$  are cross arcs.

relations found in Section 2 for a geometric outdegree distribution do not hold for arbitrary distributions.

We end in Section 4 with some comments on the case of general outdegree distributions.

**1.1. Some notation.** We denote the given outdegree distribution by  $\mathbf{P}$ . Recall that our standing assumption is that the outdegrees of the vertices are i.i.d. (independent and identically distributed).

The mean outdegree, *i.e.*, the expectation of  $\mathbf{P}$ , is denoted by  $\lambda$ . In analogy with branching processes, we say that the random digraph is *subcritical* if  $\lambda < 1$ , *critical* if  $\lambda = 1$ , and *supercritical* if  $\lambda > 1$ .

As usual, w.h.p. means *with high probability*, *i.e.*, with probability  $1 - o(1)$  as  $n \rightarrow \infty$ . We use  $\xrightarrow{\mathbf{P}}$  for convergence in probability, and  $\xrightarrow{\mathbf{d}}$  for convergence in distribution of random variables.

Moreover, let  $(a_n)$  be a sequence of positive numbers, and  $X_n$  a sequence of random variables. We write  $X_n = o_{\mathbf{P}}(a_n)$  if, as  $n \rightarrow \infty$ ,  $X_n/a_n \xrightarrow{\mathbf{P}} 0$ , *i.e.*, if for every  $\varepsilon > 0$ , we have  $\mathbb{P}(|X_n| > \varepsilon a_n) \rightarrow 0$ . Note that this is equivalent to the existence of a sequence  $\varepsilon_n \rightarrow 0$  such that  $\mathbb{P}(|X_n| > \varepsilon_n a_n) \rightarrow 0$ , or in other words  $|X_n| \leq \varepsilon_n a_n$  w.h.p. (This is sometimes denoted “ $X_n = o(a_n)$  w.h.p.”, but we will not use this notation.)

Furthermore,  $X_n = O_{L^2}(a_n)$  means  $\mathbb{E}[|X_n/a_n|^2] = O(1)$ . Note that  $X_n = O_{L^2}(a_n)$  implies  $X_n = o_{\mathbf{P}}(\omega_n a_n)$ , for any sequence  $\omega_n \rightarrow \infty$ . Note also that  $X_n = O_{L^2}(a_n)$  implies  $\mathbb{E} X_n = O(a_n)$ ; thus error terms of this type implies immediately estimates for expectations and second moments. In particular, for the most common case below,  $X_n = O_{L^2}(n^{1/2})$  is equivalent to  $\mathbb{E} X_n = O(n^{1/2})$  and  $\text{Var} X_n = O(n)$ .

$\text{Ge}(1 - p)$  denotes the geometric distribution on  $\{0, 1, \dots\}$ ; thus  $\eta \sim \text{Ge}(1 - p)$  means that  $\eta$  is a random variable with  $\mathbb{P}(\eta = k) = p^k(1 - p)$ ,  $k \geq 0$ . Similarly,

$\text{Ge}_1(1-p)$  denotes the shifted geometric distribution on  $\{1, 2, \dots\}$ ; thus  $\eta \sim \text{Ge}_1(1-p)$  means  $\mathbb{P}(\eta = k) = p^{k-1}(1-p)$ ,  $k \geq 1$ .  $\text{Po}(\lambda)$  denotes a Poisson distribution with mean  $\lambda$ .

We define  $\rho_0(x)$ , for  $x \geq 0$ , as the largest solution in  $[0, 1)$  to

$$1 - \rho_0 = e^{-x\rho_0}. \quad (1.1)$$

As is well known,  $\rho_0(x)$  is the survival probability of a Galton–Watson process with a Poisson offspring distribution  $\text{Po}(x)$  with mean  $x$ . We have  $\rho_0(x) = 0$  for  $x \leq 1$  and  $0 < \rho_0(x) < 1$  for  $x > 1$ . (See e.g. [2, Theorem I.5.1].)

For a real number  $x$ , we write  $x^+ := \max\{x, 0\}$ .  $[n] := \{1, \dots, n\}$ . All logarithms are natural.  $C$  and  $c$  are sometimes used for positive constants.

**Remark 1.3.** We state many results with error estimates in  $L^2$ , which means estimates on the second moment; we conjecture that the results extend to higher moment and estimates in  $L^p$  for any  $p < \infty$ , but we have not pursued this.

## 2. DEPTH ANALYSIS WITH GEOMETRIC OUTDEGREE DISTRIBUTION

In this section we assume that the outdegree distribution is geometric  $\text{Ge}(1-p)$  for some fixed  $0 < p < 1$ , and thus has mean

$$\lambda := \frac{p}{1-p}. \quad (2.1)$$

When doing the DFS on a random digraph of the type studied in this paper, it is natural to reveal the outdegree of a vertex as soon as we find it. (See S1–S2 in Section 4.) However, for a geometric outdegree distribution, because of its lack-of-memory property, we do not have to immediately reveal the outdegree when we find a new vertex  $v$ . Instead, we only check whether there is at least one outgoing arc (probability  $p$ ), and if so, we find its endpoint and explore this endpoint if it has not already been visited; eventually, we return to  $v$ , and then we check whether there is another outgoing arc (again probability  $p$ , by the lack-of-memory property), and so on. This will yield the important Markov property in the construction in the next subsection.

In the following, by a *future arc* from some vertex, we mean an arc that at the current time has not yet been seen by the DFS.

**2.1. Depth Markov chain.** Our aim is to track the evolution of the search depth as a function of the number  $t$  of discovered vertices. Let  $v_t$  be the  $t$ -th vertex discovered by the DFS ( $t = 1, \dots, n$ ), and let  $d(t)$  be the depth of  $v_t$  in the resulting depth-first forest, *i.e.*, the number of tree edges that connect the root of the current tree to  $v_t$ . The first found vertex  $v_1$  is a root, and thus  $d(1) = 0$ .

The quantity  $d(t)$  follows a Markov chain with transitions ( $1 \leq t < n$ ):

(i)  $d(t+1) = d(t) + 1$ .

This happens if, for some  $k \geq 1$ ,  $v_t$  has at least  $k$  outgoing arcs, the first  $k-1$  arcs lead to vertices already visited, and the  $k$ th arc leads to a new vertex (which then becomes  $v_{t+1}$ ). The probability of this is

$$\sum_{k=1}^{\infty} p^k \left(\frac{t}{n}\right)^{k-1} \left(1 - \frac{t}{n}\right) = \frac{(1-t/n)p}{1-pt/n}. \quad (2.2)$$

(ii)  $d(t+1) = d(t)$ , assuming  $d(t) > 0$ .

This holds if all arcs from  $v_t$  lead to already visited vertices, *i.e.*, (i) does not happen, and furthermore, the parent of  $v_t$  has at least one future arc leading to an unvisited vertex. These two events are independent. Moreover, by the lack-of-memory property, the number of future arcs from the parent of  $v_t$  has also the distribution  $\text{Ge}(1-p)$ . Hence, the probability that one of these future arcs leads to an unvisited vertex equals the probability in (2.2). The probability of (ii) is thus

$$\left(1 - \frac{(1-t/n)p}{1-pt/n}\right) \frac{(1-t/n)p}{1-pt/n}. \quad (2.3)$$

(iii)  $d(t+1) = d(t) - \ell$ , assuming  $d(t) > \ell \geq 1$ .

This happens if all arcs from  $v_t$  lead to already visited vertices, and so do all future arcs from the  $\ell$  nearest ancestors of  $v_t$ , while the  $(\ell+1)$ th ancestor has at least one future arc leading to an unvisited vertex. The argument in (ii) generalizes and shows that this has probability

$$\left(1 - \frac{(1-t/n)p}{1-pt/n}\right)^{\ell+1} \frac{(1-t/n)p}{1-pt/n}. \quad (2.4)$$

(iv)  $d(t+1) = d(t) - \ell$ , assuming  $d(t) = \ell \geq 0$ .

By the same argument as in (ii) and (iii), except that the  $(\ell+1)$ th ancestor does not exist and we ignore it, we obtain the probability

$$\left(1 - \frac{(1-t/n)p}{1-pt/n}\right)^{\ell+1}. \quad (2.5)$$

Note that (iv) is the case when  $d(t+1) = 0$  and thus  $v_{t+1}$  is the root of a new tree in the depth-first forest.

We can summarize (i)–(iv) in the formula

$$d(t+1) = (d(t) + 1 - \xi_t)^+, \quad (2.6)$$

where  $\xi_t$  is a random variable, independent of the history, with the distribution

$$\mathbb{P}(\xi_t = k) = (1 - \pi_t)^k \pi_t, \quad k \geq 0, \quad (2.7)$$

where

$$\pi_t := \frac{(1-t/n)p}{1-pt/n} = 1 - \frac{1-p}{1-pt/n}. \quad (2.8)$$

In other words,  $\xi_t$  has the geometric distribution  $\text{Ge}(\pi_t)$ . Define

$$\tilde{d}(t) := \sum_{i=1}^{t-1} (1 - \xi_i), \quad (2.9)$$

and note that (2.9) is a sum of independent random variables. Then  $d(t)$  can be recovered from the simpler process  $\tilde{d}(t)$  as follows.

**Lemma 2.1.** *We have*

$$d(t) = \tilde{d}(t) - \min_{1 \leq j \leq t} \tilde{d}(j), \quad 1 \leq t \leq n. \quad (2.10)$$

*Proof.* We use induction on  $t$ . Evidently, (2.10) holds for  $t = 1$  since  $d(1) = \tilde{d}(1) = 0$ .

Suppose that (2.10) holds for some  $t < n$ . Then (2.9) yields

$$\tilde{d}(t+1) = \tilde{d}(t) + 1 - \xi_t = d(t) + 1 - \xi_t + \min_{1 \leq j \leq t} \tilde{d}(j). \quad (2.11)$$

If  $d(t) + 1 - \xi_t \geq 0$ , then (2.11) shows that  $\tilde{d}(t+1) \geq \min_{1 \leq j \leq t} \tilde{d}(j)$ , and thus  $\min_{1 \leq j \leq t+1} \tilde{d}(j) = \min_{1 \leq j \leq t} \tilde{d}(j)$ ; furthermore,  $d(t+1) = d(t) + 1 - \xi_t$  by (2.6), and it follows that (2.10) holds for  $t+1$ .

On the other hand, if  $d(t) + 1 - \xi_t < 0$ , then (2.11) shows that  $\tilde{d}(t+1) < \min_{1 \leq j \leq t} \tilde{d}(j)$ , and thus  $\min_{1 \leq j \leq t+1} \tilde{d}(j) = \tilde{d}(t+1)$ . In this case,  $d(t+1) = 0$  by (2.6), and it follows that (2.10) holds for  $t+1$  in this case too.  $\square$

**Remark 2.2.** Similar formulas have been used for other, related, problems with random graphs and trees, where trees have been coded as walks, see for example [1, Section 1.3]. Note that in our case, unlike e.g. [1],  $\tilde{d}(t)$  may have negative jumps of arbitrary size.

**2.2. Main result for depth analysis.** Note first that (2.8) implies that, using  $\lambda = p/(1-p)$ ,

$$\mu_t := \mathbb{E} \xi_t = \frac{1 - \pi_t}{\pi_t} = \frac{1 - p}{p(1 - t/n)} = \frac{1}{\lambda(1 - t/n)}. \quad (2.12)$$

Hence, (2.9) implies that the expectation of  $\tilde{d}(t)$  is

$$\mathbb{E}[\tilde{d}(t)] = \sum_{i=1}^{t-1} (1 - \mathbb{E} \xi_i) = \sum_{i=1}^{t-1} (1 - \mu_i) = \sum_{i=1}^{t-1} \left(1 - \frac{1}{\lambda(1 - i/n)}\right). \quad (2.13)$$

Let  $\theta := t/n$ . We fix a  $\theta^* < 1$  and obtain that, uniformly for  $\theta \leq \theta^*$ ,

$$\mathbb{E}[\tilde{d}(t)] = \int_0^\theta \left(1 - \frac{1}{\lambda(1 - x/n)}\right) dx + O(1) = n\tilde{\ell}(\theta) + O(1), \quad (2.14)$$

where

$$\tilde{\ell}(\theta) := \int_0^\theta \left(1 - \frac{1}{\lambda(1 - x)}\right) dx = \theta + \lambda^{-1} \log(1 - \theta). \quad (2.15)$$

Note that the derivative  $\tilde{\ell}'(\theta) = 1 - \lambda^{-1}/(1 - \theta)$  is (strictly) decreasing on  $(0, 1)$ , *i.e.*,  $\tilde{\ell}$  is concave. Moreover, if  $\lambda > 1$  (*i.e.*,  $p > \frac{1}{2}$ ) (the supercritical case), then  $\tilde{\ell}'(0) > 0$ , and (2.15) shows that  $\tilde{\ell}(\theta)$  is positive and increasing for  $\theta < \theta_0 := 1 - \lambda^{-1} = (2p - 1)/p$ . After the maximum at  $\theta_0$ ,  $\tilde{\ell}(\theta)$  decreases and tends to  $-\infty$  as  $\theta \nearrow 1$ . Hence, there exists a  $\theta_0 < \theta_1 < 1$  such that  $\tilde{\ell}(\theta_1) = 0$ ; we then have  $\tilde{\ell}(\theta) > 0$  for  $0 < \theta < \theta_1$  and  $\tilde{\ell}(\theta) < 0$  for  $\theta > \theta_1$ . We will see that in this case the depth-first forest w.h.p. contains a giant tree, of order and height both linear in  $n$ , while all other trees are small.

On the other hand, if  $\lambda \leq 1$  (*i.e.*,  $p \leq \frac{1}{2}$ ) (the subcritical and critical cases), then  $\tilde{\ell}'(0) \leq 0$  and  $\tilde{\ell}(\theta)$  is negative and decreasing for all  $\theta \in (0, 1)$ . In this case, we define  $\theta_0 := \theta_1 := 0$  and note that the properties just stated for  $\tilde{\ell}$  still hold (rather trivially). We will see that in this case w.h.p. all trees in the depth-first forest are small.

Note that in all cases,

$$\theta_0 := (1 - \lambda^{-1})^+ = \begin{cases} 1 - \lambda^{-1}, & \lambda > 1, \\ 0, & \lambda \leq 1, \end{cases} \quad (2.16)$$

and that  $\theta_1$  is the largest solution in  $[0, 1)$  to

$$\log(1 - \theta_1) = -\lambda\theta_1. \quad (2.17)$$

**Remark 2.3.** The equation (2.17) may also be written  $1 - \theta_1 = \exp(-\lambda\theta_1)$ , which shows that

$$\theta_1 = \rho_0(\lambda), \quad (2.18)$$

the survival probability of a Galton–Watson process with  $\text{Po}(\lambda)$  offspring distribution defined in (1.1).

We define  $\tilde{\ell}^+(\theta) := [\tilde{\ell}(\theta)]^+$ . Thus, by (2.15) and the comments above,

$$\tilde{\ell}^+(\theta) = \begin{cases} \theta + \lambda^{-1} \log(1 - \theta), & 0 \leq \theta \leq \theta_1, \\ 0, & \theta_1 \leq \theta \leq 1. \end{cases} \quad (2.19)$$

We can now state one of our main results.

**Theorem 2.4.** *We have*

$$\max_{1 \leq t \leq n} |d(t) - n\tilde{\ell}^+(t/n)| = O_{L^2}(n^{1/2}). \quad (2.20)$$

*Proof.* Since (2.9) is a sum of independent random variables,  $\tilde{d}(t) - \mathbb{E}\tilde{d}(t)$  ( $t = 1, \dots, n$ ) is a martingale, and Doob's inequality [6, Theorem 10.9.4] yields, for all  $T \leq n$ ,

$$\mathbb{E}[\max_{t \leq T} |\tilde{d}(t) - \mathbb{E}\tilde{d}(t)|^2] \leq 4 \mathbb{E}[|\tilde{d}(T) - \mathbb{E}\tilde{d}(T)|^2] = 4 \sum_{i=1}^{T-1} \text{Var}(\xi_i). \quad (2.21)$$

As above, fix  $\theta^* < 1$ , and assume, as we may, that  $\theta^* > \theta_1$ . Let  $T^* := \lfloor n\theta^* \rfloor$ , and consider first  $t \leq T^*$ . For  $i < T^*$ , we have  $\text{Var} \xi_i = O(1)$ , and thus, for  $T = T^*$ , the sum in (2.21) is  $O(T^*) = O(n)$ . Consequently, (2.21) yields

$$\max_{t \leq T^*} |\tilde{d}(t) - \mathbb{E}\tilde{d}(t)| = O_{L^2}(n^{1/2}). \quad (2.22)$$

Hence, by (2.14),

$$M^* := \max_{t \leq T^*} |\tilde{d}(t) - n\tilde{\ell}(t/n)| = O_{L^2}(n^{1/2}). \quad (2.23)$$

(Note that  $T^*$  and  $M^*$  depend on the choice of  $\theta^*$ .) For  $t \leq T^*$ , the definition of  $M^*$  in (2.23) implies

$$\left| \min_{1 \leq j \leq t} \tilde{d}(j) - n \min_{1 \leq j \leq t} \tilde{\ell}(j/n) \right| \leq M^*. \quad (2.24)$$

Moreover, for  $t/n \leq \theta_1$ , we have  $\min_{1 \leq j \leq t} \tilde{\ell}(j/n) = O(1/n)$ , while for  $t/n \geq \theta_1$ , we have  $\min_{1 \leq j \leq t} \tilde{\ell}(j/n) = \tilde{\ell}(t/n)$ . Hence, for all  $t \leq T^*$ ,

$$\min_{1 \leq j \leq t} \tilde{\ell}(j/n) = \tilde{\ell}(t/n) - \tilde{\ell}^+(t/n) + O(1/n), \quad (2.25)$$

and thus, by (2.24),

$$\left| \min_{1 \leq j \leq t} \tilde{d}(j) - n\tilde{\ell}(t/n) + n\tilde{\ell}^+(t/n) \right| \leq M^* + O(1/n). \quad (2.26)$$

Finally, by (2.10), (2.23) and (2.26),

$$|d(t) - n\tilde{\ell}^+(t/n)| \leq 2M^* + O(1/n). \quad (2.27)$$

This holds uniformly for  $t \leq T^*$ , and thus, by (2.23),

$$\max_{1 \leq t \leq T^*} |d(t) - n\tilde{\ell}^+(t/n)| = O_{L^2}(n^{1/2}). \quad (2.28)$$

It remains to consider  $T^* < t \leq n$ . Then the argument above does not quite work, because  $\pi_t \searrow 0$  and thus  $\text{Var} \xi_t \nearrow \infty$  as  $t \nearrow n$ . We therefore modify  $\xi_t$ . We define  $\hat{\pi}_t := \max\{\pi_t, \pi_{T^*}\}$ ; thus  $\hat{\pi}_t = \pi_t$  for  $t \leq T^*$  and  $\hat{\pi}_t > \pi_t$  for  $t > T^*$ . We may then define independent random variables  $\hat{\xi}_t$  such that  $\hat{\xi}_t \sim \text{Ge}(\hat{\pi}_t)$  and  $\hat{\xi}_t \leq \xi_t$  for all  $t < n$ . (Thus,  $\hat{\xi}_t = \xi_t$  for  $t \leq T^*$ .)

In analogy with (2.9)–(2.10), we further define

$$\hat{\tilde{d}}(t) := \sum_{i=1}^{t-1} (1 - \hat{\xi}_i), \quad (2.29)$$

$$\hat{d}(t) := \hat{\tilde{d}}(t) - \min_{1 \leq j \leq t} \hat{\tilde{d}}(j) = \max_{1 \leq j \leq t} \sum_{i=j}^{t-1} (1 - \hat{\xi}_i). \quad (2.30)$$

Since  $\hat{\xi}_i \leq \xi_i$ , (2.30) implies that  $\hat{d}(t) \geq d(t)$  for all  $t$ .

We have  $\text{Var}[\hat{\xi}_t] = O(1)$ , uniformly for all  $t < n$ , and thus the argument above yields

$$\max_{1 \leq t \leq n} |\hat{d}(t) - n[\hat{\ell}(t/n)]^+| = O_{L^2}(n^{1/2}), \quad (2.31)$$

where

$$\hat{\ell}(\theta) := \int_0^\theta \min\left\{ \left(1 - \frac{1}{\lambda(1-x)}\right), \left(1 - \frac{1}{\lambda(1-\theta^*)}\right) \right\} dx. \quad (2.32)$$

We have  $\hat{\ell}(\theta) = \tilde{\ell}(\theta)$  for  $\theta \leq \theta^*$ , and for  $\theta \geq \theta^*$ ,  $\hat{\ell}(\theta)$  is negative and decreasing (since  $\theta^* > \theta_1$ ). Hence,  $[\hat{\ell}(\theta)]^+ = \tilde{\ell}^+(\theta)$  for all  $0 < \theta \leq 1$ . In particular,  $[\hat{\ell}(\theta)]^+ = \tilde{\ell}^+(\theta) = 0$  for all  $\theta \geq \theta^*$ , and (2.31) implies

$$\max_{T^* < t \leq n} \hat{d}(t) = O_{L^2}(n^{1/2}). \quad (2.33)$$

Recalling  $0 \leq d(t) \leq \hat{d}(t)$ , we thus have

$$\max_{T^* < t \leq n} |d(t) - n\tilde{\ell}^+(t/n)| = \max_{T^* < t \leq n} d(t) \leq \max_{T^* < t \leq n} \hat{d}(t) = O_{L^2}(n^{1/2}), \quad (2.34)$$

which completes the proof.  $\square$

**Corollary 2.5.** *The height  $\Upsilon$  of the depth-first forest is*

$$\Upsilon := \max_{1 \leq t \leq n} d(t) = vn + O_{L^2}(n^{1/2}), \quad (2.35)$$

where

$$v = v(p) := \tilde{\ell}^+(\theta_0) = \begin{cases} 0, & 0 < \lambda \leq 1, \\ 1 - \lambda^{-1} - \lambda^{-1} \log \lambda, & \lambda > 1. \end{cases} \quad (2.36)$$



*Proof.* Immediate from Theorem 2.4 and (2.15), since we have  $\max_t \tilde{\ell}^+(t/n) = \max_\theta \tilde{\ell}^+(\theta) + O(1/n)$  and  $\max_\theta \tilde{\ell}^+(\theta) = \tilde{\ell}^+(\theta_0) = \tilde{\ell}(\theta_0)$ .  $\square$

In Section 2.3 we will improve this when  $\lambda > 1$ , and show that then the height  $\Upsilon$  is asymptotically normally distributed (Theorem 2.11).

**Corollary 2.6.** *The average depth  $\bar{d}$  in the depth-first forest is*

$$\bar{d} := \frac{1}{n} \sum_{t=1}^n d(t) = \alpha n + O_{L^2}(n^{1/2}), \quad (2.37)$$

where  $\alpha = 0$  if  $\lambda \leq 1$ , and, in general,

$$\alpha = \alpha(p) := \frac{1}{2}\theta_1^2 - \frac{1}{\lambda} \left( (1 - \theta_1) \log(1 - \theta_1) + \theta_1 \right) = \frac{\lambda - 1}{\lambda} \theta_1 - \frac{1}{2}\theta_1^2. \quad (2.38)$$

*Proof.* By Theorem 2.4,

$$\frac{1}{n} \sum_{t=1}^n d(t) = \sum_{t=1}^n \tilde{\ell}^+(t/n) + O_{L^2}(n^{1/2}) = n\alpha + O_{L^2}(n^{1/2}), \quad (2.39)$$

where

$$\begin{aligned} \alpha &:= \int_0^1 \tilde{\ell}^+(x) dx = \int_0^{\theta_1} \tilde{\ell}(x) dx = \int_0^{\theta_1} \left( x + \lambda^{-1} \log(1 - x) \right) dx \\ &= \frac{1}{2}\theta_1^2 - \lambda^{-1} \left( (1 - \theta_1) \log(1 - \theta_1) + \theta_1 \right), \end{aligned} \quad (2.40)$$

which yields (2.38), using (2.17).  $\square$

**Remark 2.7.** When  $\lambda > 1$ , the height  $\Upsilon$  and average depth  $\bar{d}$  are thus linear in  $n$ , unlike many other types of random trees. This might imply a rather slow performance of algorithms that operate on the depth-first forest if it is built explicitly in a computer's memory.

**2.3. Asymptotic normality.** In this subsection, we show that in the supercritical case  $\lambda > 1$ , Theorem 2.4 can be improved to yield convergence of  $d(t)$  (after rescaling) to a Gaussian process, at least on  $[0, \theta_1]$ . As a consequence, we show that the height  $\Upsilon$  is asymptotically normal.

Recall that for an interval  $I \subseteq \mathbb{R}$ ,  $D(I)$  is the space of functions  $I \rightarrow \mathbb{R}$  that are right-continuous with left limits (*càdlàg*) equipped with the Skorohod topology. For definitions of the topology see e.g. [3], [7], [11, Appendix A.2], or [8]; for our purposes it is enough to know that convergence in  $D(I)$  to a continuous limit is equivalent to uniform convergence on compact subsets of  $I$ . (Note that it thus matters if the endpoints are included in  $I$  or not; for example, convergence in  $D[0, 1)$  and  $D(0, 1]$  mean different things.)

We define  $d(0) := \bar{d}(0) := 0$ .

**Lemma 2.8.** *Assume  $\lambda > 1$ . Then*

$$n^{-1/2} (\bar{d}(\lfloor n\theta \rfloor) - n\tilde{\ell}(\theta)) \xrightarrow{d} Z(\theta) \quad \text{in } D[0, 1], \quad (2.41)$$

where  $Z(\theta)$  is a continuous Gaussian process on  $[0, 1)$  with mean  $\mathbb{E} Z(\theta) = 0$  and covariance  $\text{Cov}(Z(x), Z(y)) = v(\min\{x, y\})$ , where

$$v(\theta) := \frac{(1-p)^2\theta}{p^2(1-\theta)} - \frac{1-p}{p} \log(1-\theta) = \lambda^{-2} \frac{\theta}{1-\theta} - \lambda^{-1} \log(1-\theta). \quad (2.42)$$

Equivalently,  $Z(\theta) = B(v(\theta))$  for a Brownian motion  $B(x)$ .

*Proof.* Since the random variables  $\xi_t$  are independent, (2.9) and (2.7)–(2.8) yield, similarly to (2.13),

$$\mathrm{Var}[\tilde{d}(t)] = \sum_{i=1}^{t-1} \mathrm{Var} \xi_i = \sum_{i=1}^{t-1} \frac{1 - \pi_i}{\pi_i^2} = \sum_{i=1}^{t-1} \frac{(1-p)(1-pi/n)}{p^2(1-i/n)^2}. \quad (2.43)$$

Hence, uniformly for  $t/n \leq \theta^*$  for any  $\theta^* < 1$ ,

$$\mathrm{Var}[\tilde{d}(t)] = nv(t/n) + O(1), \quad (2.44)$$

with

$$v(\theta) := \frac{1-p}{p^2} \int_0^\theta \frac{1-px}{(1-x)^2} dx = \frac{(1-p)^2\theta}{p^2(1-\theta)} - \frac{1-p}{p} \log(1-\theta), \quad (2.45)$$

in agreement with (2.42). Since also  $\mathbb{E} \tilde{d}(\lfloor n\theta \rfloor) = n\tilde{\ell}(\theta) + O(1)$  by (2.14), the marginal convergence for a fixed  $\theta$  in (2.41) follows by the classical central limit theorem for independent (not identically distributed) variables, e.g. using Lyapounov's condition [6, Theorem 7.2.2].

The functional limit (2.41) is thus a version of Donsker's theorem [3, Theorem 16.1], extended from the i.i.d. case to the non-identically distributed variables  $\xi_i$ . We expect that such generalizations of Donsker's theorem exist in the literature, but we do not know any specific reference; however, standard proofs extend without problems. (For example, the proof of [3, Theorem 16.1]. Alternatively, finite-dimensional convergence follows from the classical central limit theorem, and tightness in  $D[0,1]$  can easily be shown by e.g. Aldous's criterion [11, Theorem 16.11].) Moreover, (2.41) follows also directly by general results on convergence of martingales, for example [8, Proposition 2.6] or [9, Proposition 9.1], which both are based on [7, Theorem VIII.3.12].  $\square$

**Lemma 2.9.** *Assume  $\lambda > 1$  and let  $0 < \theta^* < \theta_1$ . Then*

$$\min_{1 \leq j \leq \lfloor n\theta^* \rfloor} \tilde{d}(j) = o_{\mathbb{P}}(n^{1/2}). \quad (2.46)$$

*Proof.* Let  $t_n := \lfloor n^{2/3} \rfloor$ . If  $n$  is large enough, then  $t_n < n\theta^*$ , and, since  $\tilde{\ell}'(0) = 1 - \lambda^{-1} > 0$  by (2.15),

$$\min_{t_n/n \leq \theta \leq \theta^*} \tilde{\ell}(\theta) = \tilde{\ell}(t_n/n) \geq ct_n/n \geq cn^{-1/3} \quad (2.47)$$

for some constant  $c > 0$ . Furthermore, (2.23) implies

$$\max_{t_n \leq t \leq n\theta^*} |\tilde{d}(t) - n\tilde{\ell}(t/n)| = O_{L^2}(n^{1/2}) = o_{\mathbb{P}}(n^{2/3}) \quad (2.48)$$

(recall that  $O_{L^2}(a_n)$  implies  $o_{\mathbb{P}}(\omega_n a_n)$  for any  $a_n$  and any  $\omega_n \rightarrow \infty$ ). It follows from (2.47)–(2.48) that w.h.p.  $|\tilde{d}(t) - n\tilde{\ell}(t/n)| < n\tilde{\ell}(t/n)$  for all  $t \in [t_n, n\theta^*]$ . Hence, w.h.p.,  $\tilde{d}(t) > 0 = \tilde{d}(1)$  for all  $t \in [t_n, n\theta^*]$ . Consequently, w.h.p.,

$$\min_{1 \leq t \leq n\theta^*} \tilde{d}(t) = \min_{1 \leq t \leq t_n} \tilde{d}(t). \quad (2.49)$$

For  $t \leq t_n$ , we use Doob's inequality in the form (2.21) again. Since  $\min_{t \leq t_n} \tilde{d}(t) \leq \tilde{d}(1) = 0$  and  $\mathbb{E} \tilde{d}(t) \geq 0$  for  $t \leq t_n$  (for  $n$  large), we have  $|\min_{t \leq t_n} \tilde{d}(t)| \leq \max_{t \leq t_n} |\tilde{d}(t)|$ –

$\mathbb{E} \tilde{d}(t)$  and thus (2.21) yields

$$\mathbb{E} \left| \min_{t \leq t_n} \tilde{d}(t) \right|^2 \leq 4 \sum_{i=1}^{t_n-1} \text{Var}(\xi_i) = O(t_n) = O(n^{2/3}). \quad (2.50)$$

Hence,

$$\min_{t \leq t_n} \tilde{d}(t) = O_{L^2}(n^{1/3}) = o_p(n^{1/2}). \quad (2.51)$$

The proof is completed by combining (2.49) and (2.51).  $\square$

**Theorem 2.10.** *Assume  $\lambda > 1$ . Then*

$$n^{-1/2}(d(\lfloor n\theta \rfloor) - n\tilde{\ell}(\theta)) \xrightarrow{d} Z(\theta) \quad \text{in } D[0, \theta_1] \quad (2.52)$$

where  $Z(\theta)$  is the continuous Gaussian process defined in Lemma 2.8.

*Proof.* By (2.10) and Lemma 2.9, for any  $\theta^* < \theta_1$ ,

$$\max_{0 \leq t \leq \lfloor n\theta^* \rfloor} |d(t) - \tilde{d}(t)| = \left| \min_{1 \leq j \leq \lfloor n\theta^* \rfloor} \tilde{d}(j) \right| = o_p(n^{1/2}). \quad (2.53)$$

The theorem now follows from Lemma 2.8.  $\square$

**Theorem 2.11.** *Let  $\lambda > 1$ . Then the height  $\Upsilon$  of the depth-first forest has an asymptotic normal distribution:*

$$\frac{\Upsilon - vn}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2) \quad (2.54)$$

with  $v$  given by (2.36), and

$$\sigma^2 := \lambda^{-1} - \lambda^{-2} + \lambda^{-1} \log \lambda. \quad (2.55)$$

*Proof.* Fix some  $\theta^* \in (\theta_0, \theta_1)$ . By Theorem 2.10 and the Skorohod coupling theorem [11, Theorem 4.30], we may assume that the random variables for different  $n$  are coupled such that (2.52) holds (almost) surely. Since  $Z(\theta)$  is continuous, this implies uniform convergence on  $[0, \theta^*]$ , i.e.,

$$d(\lfloor n\theta \rfloor) = n\tilde{\ell}(\theta) + n^{1/2}Z(\theta) + o(n^{1/2}), \quad (2.56)$$

uniformly on  $[0, \theta^*]$ . (The  $o(n^{1/2})$  here are random, but uniform in  $\theta$ .) For  $|\theta - \theta_0| \leq n^{-1/6}$ , we have  $Z(\theta) = Z(\theta_0) + o(1)$ , since  $Z$  is continuous, and thus (2.56) yields, almost surely,

$$d(\lfloor n\theta \rfloor) = n\tilde{\ell}(\theta) + n^{1/2}Z(\theta_0) + o(n^{1/2}), \quad |\theta - \theta_0| \leq n^{-1/6}. \quad (2.57)$$

Since  $\max_{\theta} \tilde{\ell}(\theta) = \tilde{\ell}(\theta_0)$ , it follows that

$$\max_{|\theta - \theta_0| \leq n^{-1/6}} d(\lfloor n\theta \rfloor) = n\tilde{\ell}(\theta_0) + n^{1/2}Z(\theta_0) + o(n^{1/2}). \quad (2.58)$$

On the other hand, for  $|\theta - \theta_0| \geq n^{-1/6}$ , we have by a Taylor expansion, for some  $c > 0$ ,

$$\tilde{\ell}(\theta) \leq \tilde{\ell}(\theta_0) - c(\theta - \theta_0)^2 \leq \tilde{\ell}(\theta_0) - cn^{-1/3}. \quad (2.59)$$

Hence, (2.20) implies

$$\max_{|\theta - \theta_0| \geq n^{-1/6}} d(t) \leq n \max_{|\theta - \theta_0| \geq n^{-1/6}} \tilde{\ell}(\theta) + O_{L^2}(n^{1/2}) \leq n\tilde{\ell}(\theta_0) - cn^{2/3} + O_{L^2}(n^{1/2}). \quad (2.60)$$

Comparing (2.58) and (2.60), we see that w.h.p. the maximum in (2.58) is larger than the one in (2.60), and thus

$$\Upsilon = \max_{0 \leq \theta \leq 1} d(\lfloor n\theta \rfloor) = n\tilde{\ell}(\theta_0) + n^{1/2}Z(\theta_0) + o(n^{1/2}). \quad (2.61)$$

Hence, w.h.p.,

$$\frac{\Upsilon - \tilde{\ell}(\theta_0)n}{\sqrt{n}} = Z(\theta_0) + o(1), \quad (2.62)$$

which implies

$$\frac{\Upsilon - \tilde{\ell}(\theta_0)n}{\sqrt{n}} \xrightarrow{d} Z(\theta_0) \sim N(0, v(\theta_0)). \quad (2.63)$$

Since  $\tilde{\ell}(\theta_0) = v$  by (2.36), this shows (2.54) with  $\sigma^2 := v(\theta_0)$ , which gives (2.55) by (2.42) and  $\theta_0 := 1 - \lambda^{-1}$ .  $\square$

#### 2.4. The trees in the forest.

**Theorem 2.12.** *Let  $N$  be the number of trees in the depth-first forest. Then*

$$N = \psi n + O_{L^2}(n^{1/2}), \quad (2.64)$$

where

$$\psi = \psi(p) := 1 - \theta_1 - \frac{\lambda}{2}(1 - \theta_1)^2. \quad (2.65)$$

Figure 2 shows the parameter  $\psi$  as a function of the average degree  $\lambda$ .

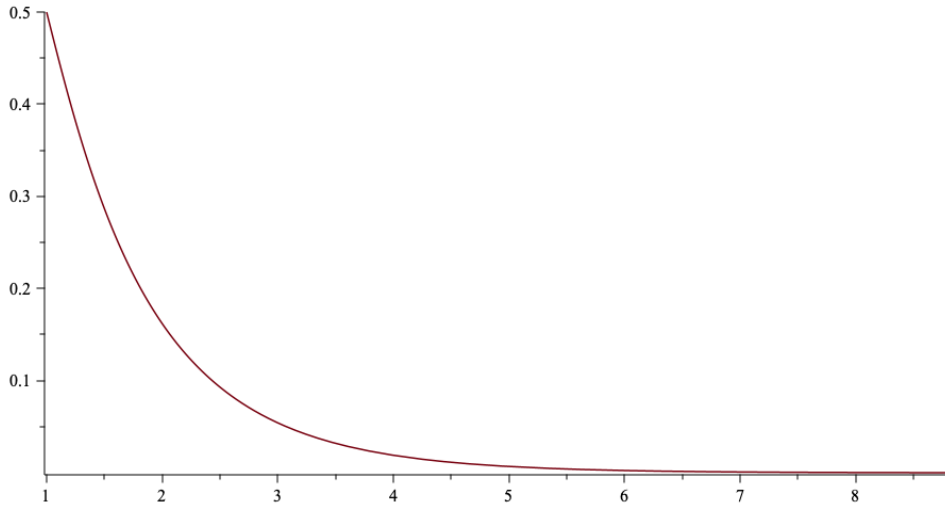


FIGURE 2.  $\psi$ , as function of  $\lambda$ .

*Proof.* Let  $J_t := \mathbf{1}\{d(t) = 0\}$ , the indicator that vertex  $t$  is a root and thus starts a new tree. Thus  $N = \sum_1^n J_t$ .

If  $\theta_1 > 0$  (i.e.,  $\lambda > 1$ ), then Theorem 2.4 shows that w.h.p.  $d(t) > 0$  in the interval  $(1, n\theta_1)$ , except possibly close to the endpoints. Thus the DFS will find one giant tree of order  $\approx \theta_1 n$ , possibly preceded by a few small trees, and, as we will see later in the proof, followed by many small trees. To obtain a precise estimate, we note

that there exists a constant  $c > 0$  such that  $\tilde{\ell}(\theta) \geq \min\{c\theta, c(\theta_1 - \theta)\}$  for  $\theta \in [0, \theta_1]$ . Hence, if  $t \leq n\theta_1$  and  $d(t) = 0$ , then  $\tilde{d}(t) \leq d(t) = 0$  by (2.10) and, recalling (2.23),

$$M^* \geq n\tilde{\ell}(t/n) \geq c \min\{t, n\theta_1 - t\}. \quad (2.66)$$

Consequently,  $d(t) = 0$  with  $t \leq n\theta_1$  implies  $t \in [1, c^{-1}M^*] \cup [n\theta_1 - c^{-1}M^*, n\theta_1]$ . The number of such  $t$  is thus  $O(M^* + 1) = O_{L^2}(n^{1/2})$ , using (2.23).

Let  $T_1 := \lceil n\theta_1 \rceil$ . We have just shown that (the case  $\theta_1 = 0$  is trivial)

$$\sum_{t=1}^{T_1-1} J_t = O_{L^2}(n^{1/2}). \quad (2.67)$$

It remains to consider  $t \geq T_1$ . For any integer  $k \geq 0$ , the conditional distribution of  $\xi_t - k$  given  $\xi_t \geq k$  equals the distribution of  $\xi_t$ . Hence, recalling (2.12),

$$\mathbb{E}[(\xi_t - k)^+] = \mathbb{E}[\xi_t - k \mid \xi_t \geq k] \mathbb{P}(\xi_t \geq k) = \mu_t \mathbb{P}(\xi_t - k \geq 0). \quad (2.68)$$

We use again the stochastic recursion (2.6). Let  $\mathcal{F}_t$  be the  $\sigma$ -field generated by  $\xi_1, \dots, \xi_{t-1}$ . Then  $d(t)$  is  $\mathcal{F}_t$ -measurable, while  $\xi_t$  is independent of  $\mathcal{F}_t$ . Hence, (2.6) and (2.68) yield

$$\begin{aligned} \mathbb{E}[d(t+1) \mid \mathcal{F}_t] &= \mathbb{E}[d(t) + 1 - \xi_t \mid \mathcal{F}_t] + \mathbb{E}[(\xi_t - 1 - d(t))^+ \mid \mathcal{F}_t] \\ &= d(t) + 1 - \mu_t + \mu_t \mathbb{P}[\xi_t - 1 - d(t) \geq 0 \mid \mathcal{F}_t] \\ &= d(t) + 1 - \mu_t + \mu_t \mathbb{P}[d(t+1) = 0 \mid \mathcal{F}_t] \\ &= d(t) + 1 - \mu_t + \mu_t \mathbb{E}[J_{t+1} \mid \mathcal{F}_t]. \end{aligned} \quad (2.69)$$

We write  $\Delta d(t) := d(t+1) - d(t)$  and  $\bar{J}_t := 1 - J_t$ . Then (2.69) yields

$$\mathbb{E}[\Delta d(t) - 1 + \mu_t \bar{J}_{t+1} \mid \mathcal{F}_t] = 0. \quad (2.70)$$

Define

$$\mathcal{M}_t := \sum_{i=1}^{t-1} \mu_i^{-1} (\Delta d(i) - 1 + \mu_i \bar{J}_{i+1}) = \sum_{i=1}^{t-1} (\mu_i^{-1} \Delta d(i) - \mu_i^{-1} + \bar{J}_{i+1}). \quad (2.71)$$

Then  $\mathcal{M}_t$  is  $\mathcal{F}_t$ -measurable, and (2.70) shows that  $\mathcal{M}_t$  is a martingale. We have, with  $\Delta \mathcal{M}_t := \mathcal{M}_{t+1} - \mathcal{M}_t$ , using (2.6),

$$|\Delta \mathcal{M}_t| \leq \mu_t^{-1} |d(t+1) - d(t) - 1| + \bar{J}_{t+1} \leq \mu_t^{-1} \xi_t + 1, \quad (2.72)$$

and thus, since  $\pi_t \leq p < 1$  for all  $t$  by (2.8),

$$\mathbb{E}|\Delta \mathcal{M}_t|^2 \leq 2\mu_t^{-2} \mathbb{E} \xi_t^2 + 2 = 2 \left( \frac{\pi_t}{1 - \pi_t} \right)^2 \frac{1 - \pi_t + (1 - \pi_t)^2}{\pi_t^2} + 2 = O(1). \quad (2.73)$$

Hence, uniformly for all  $T \leq n$ ,

$$\mathbb{E} \mathcal{M}_T^2 = \sum_{t=1}^{T-1} \mathbb{E} |\Delta \mathcal{M}_t|^2 = O(T) = O(n). \quad (2.74)$$

The definition (2.71) yields

$$\mathcal{M}_n - \mathcal{M}_{T_1} = \sum_{t=T_1}^{n-1} \mu_t^{-1} \Delta d(t) - \sum_{t=T_1}^{n-1} \mu_t^{-1} + \sum_{t=T_1}^{n-1} \bar{J}_{t+1}. \quad (2.75)$$

By a summation by parts, and interpreting  $\mu_n^{-1} := 0$ ,

$$\sum_{t=T_1}^{n-1} \mu_t^{-1} \Delta d(t) = \sum_{t=T_1+1}^n (\mu_{t-1}^{-1} - \mu_t^{-1}) d(t) - \mu_{T_1}^{-1} d(T_1). \quad (2.76)$$

As  $t$  increases,  $\mu_t$  increases by (2.12), and thus  $\mu_{t-1}^{-1} - \mu_t^{-1} > 0$ . Hence, (2.76) implies

$$\begin{aligned} \left| \sum_{t=T_1}^{n-1} \mu_t^{-1} \Delta d(t) \right| &\leq \sum_{t=T_1+1}^n (\mu_{t-1}^{-1} - \mu_t^{-1}) \sup_{i>T_1} |d(i)| + \mu_{T_1}^{-1} |d(T_1)| \leq 2\mu_{T_1}^{-1} \sup_{i\geq T_1} |d(i)| \\ &= O_{L^2}(n^{1/2}) \end{aligned} \quad (2.77)$$

by (2.20), since  $\tilde{\ell}^+(t/n) = 0$  for  $t \geq T_1 \geq n\theta_1$ . Furthermore, (2.74) shows that  $\mathcal{M}_n, \mathcal{M}_{T_1} = O_{L^2}(n^{1/2})$ . Hence, (2.75) yields, using (2.12),

$$\begin{aligned} \sum_{t=T_1+1}^n J_t &= n - T_1 - \sum_{t=T_1+1}^n \bar{J}_t = n - T_1 - \sum_{t=T_1}^{n-1} \mu_t^{-1} + O_{L^2}(n^{1/2}) \\ &= n - T_1 - \sum_{t=T_1}^{n-1} \lambda(1 - t/n) + O_{L^2}(n^{1/2}) = n\psi + O_{L^2}(n^{1/2}), \end{aligned} \quad (2.78)$$

where

$$\psi := 1 - \theta_1 - \int_{\theta_1}^1 \lambda(1 - x) dx = 1 - \theta_1 - \frac{\lambda}{2}(1 - \theta_1)^2. \quad (2.79)$$

The result follows by (2.78) and (2.67).  $\square$

The arguments in the proof of Theorem 2.12 show that in the supercritical case  $\lambda > 1$ , the DFS w.h.p. find first possibly a few small trees, then a giant tree containing all  $v_t$  with  $O_{L^2}(n^{1/2}) \leq t \leq \theta_1 n + O_{L^2}(n^{1/2})$ , and then a large number of small trees. We give some details in the following lemma and theorem.

**Lemma 2.13.** *Let  $(a, b)$  be a fixed interval with  $0 \leq a < b \leq 1$  and  $b > \theta_1$ . Then w.h.p. there exists a root  $v_t$  in the depth-first forest with  $t/n \in (a, b)$ .*

*Proof.* By increasing  $a$ , we may assume that  $\theta_1 < a < b \leq 1$ . Then, cf. (2.15),  $\tilde{\ell}'(a) < \tilde{\ell}'(\theta_1) \leq 0$  and thus  $\lambda(1 - a) < 1$ . Hence, the argument yielding (2.78) in the proof of Theorem 2.12 yields also

$$\sum_{t=\lceil an \rceil}^{\lfloor bn \rfloor} J_t = bn - an - \sum_{t=\lceil an \rceil}^{\lfloor bn \rfloor} \lambda(1 - t/n) + O_{L^2}(n^{1/2}) = cn + O_{L^2}(n^{1/2}), \quad (2.80)$$

with  $c > (b - a)(1 - \lambda(1 - a)) > 0$ . Hence, w.h.p. there are many roots  $v_t$  with  $t \in (an, bn)$ .  $\square$

**Theorem 2.14.** *Let  $\mathbf{T}_1$  be the largest tree in the depth-first forest.*

- (i) *If  $\lambda \leq 1$ , then  $|\mathbf{T}_1| = o_p(n)$ .*
- (ii) *If  $\lambda > 1$ , then  $|\mathbf{T}_1| = \theta_1 n + O_{L^2}(n^{1/2})$ . Furthermore, the second largest tree has order  $|\mathbf{T}_2| = o_p(n)$ .*

*Proof.* Let  $\varepsilon > 0$ . By covering  $[\theta_1, 1]$  with a finite number of intervals of length  $< \varepsilon/2$ , it follows from Lemma 2.13 that w.h.p. every tree  $\mathbf{T}$  having a root  $v_t$  with  $t > (\theta_1 - \varepsilon/2)n$  has  $|\mathbf{T}| \leq \varepsilon n$ .

In particular, if  $\lambda \leq 1$ , so  $\theta_1 = 0$ , this applies to all trees, and thus w.h.p.  $|\mathbf{T}_1| \leq \varepsilon n$ , which proves (i).

Suppose now  $\lambda > 1$ . Consider the tree  $\mathbf{T}$  in the depth-first forest that contains  $v_{\lfloor n\theta_0 \rfloor}$ , denote its root by  $v_r$  and let  $v_s$  be its last vertex. By the proof of Theorem 2.12,  $d(t) > 0$  for  $O_{L^2}(n^{1/2}) \leq t \leq \theta_1 n - O_{L^2}(n^{1/2})$ , and thus  $r = O_{L^2}(n^{1/2})$  and  $s \geq \theta_1 n - O_{L^2}(n^{1/2})$ .

On the other hand, let  $\theta^* \in (\theta_1, 1)$ . If  $s \geq \theta_1 n$ , let  $u := \min\{s, \lfloor \theta^* n \rfloor\}$ . Since  $r/n \leq \theta_0$ , we have  $\tilde{\ell}(r/n) \geq 0$ . Furthermore, (2.10) implies that  $\tilde{d}(t) > \min_{j \leq t} \tilde{d}(j) = \tilde{d}(r)$  for  $t \in (r, s]$ , and thus  $\tilde{d}(u) > \tilde{d}(r)$ . Hence, by (2.23),

$$-n\tilde{\ell}(u/n) \leq n\tilde{\ell}(r/n) - n\tilde{\ell}(u/n) \leq 2M^* + \tilde{d}(r) - \tilde{d}(u) \leq 2M^* = O_{L^2}(n^{1/2}). \quad (2.81)$$

Since  $\tilde{\ell}(\theta_1) = 0$  and  $\tilde{\ell}'(\theta) \leq -c < 0$  for  $\theta \geq \theta_1$ , it follows that  $u \leq \theta_1 n + O_{L^2}(n^{1/2})$ , and thus  $s \leq \theta_1 n + O_{L^2}(n^{1/2})$ .

Consequently,  $s = \theta_1 n + O_{L^2}(n^{1/2})$ , and thus  $|\mathbf{T}| = s - r + 1 = \theta_1 n + O_{L^2}(n^{1/2})$ . Furthermore, any tree found before  $\mathbf{T}$  has order  $\leq r = o_p(n)$ . The first part of the proof now shows that for every  $\varepsilon > 0$ , there is w.h.p. no tree other than  $|\mathbf{T}|$  of order  $> \varepsilon n$ . Hence, w.h.p.  $\mathbf{T}$  is the largest tree, and (ii) follows.  $\square$

**Remark 2.15.** As said in Remark 2.3,  $\theta_1$ , the asymptotic fraction of vertices in the giant tree, equals the survival probability of a Galton–Watson process with  $\text{Po}(\lambda)$  offspring distribution. Heuristically, this may be explained by the following argument, well known from similar situations. Start at a random vertex and follow the arcs backwards. The indegree of a given vertex is asymptotically  $\text{Po}(\lambda)$ , and the process of exploring backwards from a vertex may be approximated by a Galton–Watson process with this offspring distribution. Hence, the probability of a “large” backwards process converges to the survival probability  $\theta_1$ . It seems reasonable that most vertices in the giant tree have a large backwards process, while most vertices outside the giant have a small backwards process.

Note also that the asymptotic size of the giant tree thus equals the asymptotic size of the giant component in an undirected Erdős–Rényi random graph  $G(n, \lambda/n)$ , which heuristically is given by the same argument. (See also Remark 1.1 and [10].)

**2.5. Types of arcs.** Recall from the introduction the classification of the arcs in the digraph  $G$ . Since we assume that the outdegrees are  $\text{Ge}(1-p)$  and independent, the total number of arcs,  $M$  say, has a negative binomial distribution with mean  $\lambda n$ , and, by a weak version of the law of large numbers,

$$M = \lambda n + O_{L^2}(n^{1/2}). \quad (2.82)$$

In the following theorem, we give the asymptotics of the number of arcs of each type.

**Theorem 2.16.** *Let  $L, T, B, F$  and  $C$  be the numbers of loops, tree arcs, back arcs, forward arcs, and cross arcs in the random digraph. Then*

$$L = O_{L^2}(1), \quad (2.83)$$

$$T = \tau n + O_{L^2}(n^{1/2}), \quad (2.84)$$

$$B = \beta n + O_{L^2}(n^{1/2}), \quad (2.85)$$

$$F = \varphi n + O_{L^2}(n^{1/2}), \quad (2.86)$$

$$C = \chi n + O_{L^2}(n^{1/2}), \quad (2.87)$$

where

$$\tau := \chi := 1 - \psi = \theta_1 + \frac{\lambda}{2}(1 - \theta_1)^2, \quad (2.88)$$

$$\beta := \varphi := \lambda\alpha = (\lambda - 1)\theta_1 - \frac{\lambda}{2}\theta_1^2. \quad (2.89)$$

*Proof.* Let  $\eta_t$  be the number of arcs from  $v_t$ , and let  $\eta_t^<, \eta_t^=, \eta_t^>$  be the numbers of these arcs that lead to some  $v_u$  with  $u < t$ ,  $u = t$  and  $u > t$ , respectively. Then

$$L = \sum_{t=1}^n \eta_t^=. \quad (2.90)$$

Furthermore, an arc  $v_t v_u$  with  $u > t$  is either a tree arc or a forward arc; conversely, every tree arc or forward arc is of this type. Consequently,

$$T + F = \sum_{t=1}^n \eta_t^>. \quad (2.91)$$

Similarly, or by (2.90) and (2.91),

$$B + C = \sum_{t=1}^n \eta_t^<. \quad (2.92)$$

Conditioned on  $\eta_t$ ,  $\eta_t^<$  has a binomial distribution  $\text{Bin}(\eta_t, (t-1)/n)$ , since each arc has probability  $(t-1)/n$  to go to a vertex  $v_u$  with  $u < t$ . In general, if  $X \sim \text{Bin}(m, p)$ , then  $\mathbb{E} X = mp$  and  $\mathbb{E} X^2 = \text{Var} X + (\mathbb{E} X)^2 = mp(1-p) + (mp)^2$ . Hence, by first conditioning on  $\eta_t$ ,

$$\mathbb{E} \eta_t^< = \mathbb{E} \left[ \eta_t \frac{t-1}{n} \right] = \lambda \frac{t-1}{n}, \quad (2.93)$$

$$\text{Var} \eta_t^< \leq \mathbb{E} [(\eta_t^<)^2] = \mathbb{E} \left[ \eta_t \frac{t-1}{n} \left( 1 - \frac{t-1}{n} \right) + \eta_t^2 \left( \frac{t-1}{n} \right)^2 \right] = O(1). \quad (2.94)$$

Furthermore, the random variables  $\eta_t^<$ ,  $t = 1, \dots, n$ , are independent. Hence, (2.92) yields

$$\mathbb{E} B + \mathbb{E} C = \sum_{t=1}^n \mathbb{E} \eta_t^< = \lambda \sum_{t=1}^n \frac{t-1}{n} = \frac{\lambda}{2}(n-1), \quad (2.95)$$

$$\text{Var}[B + C] = \sum_{t=1}^n \text{Var} \eta_t^< = O(n), \quad (2.96)$$

and thus

$$B + C = \mathbb{E}[B + C] + O_{L^2}(n^{1/2}) = \frac{\lambda}{2}n + O_{L^2}(n^{1/2}). \quad (2.97)$$



The same argument with (2.91) yields

$$\mathbb{E} T + \mathbb{E} F = \frac{\lambda}{2}(n-1), \quad (2.98)$$

$$\text{Var}[F + T] = O(n), \quad (2.99)$$

$$F + T = \frac{\lambda}{2}n + O_{L^2}(n^{1/2}). \quad (2.100)$$

Similarly, conditioned on  $\eta_t$ , we have  $\eta_t^- \sim \text{Bin}(\eta_t, 1/n)$ , and we find

$$\mathbb{E} L = \sum_{t=1}^n \mathbb{E} \eta_t^- = n\lambda \frac{1}{n} = \lambda, \quad (2.101)$$

$$\text{Var} L = \sum_{t=1}^n \text{Var} \eta_t^- = O(1), \quad (2.102)$$

$$L = \lambda + O_{L^2}(1) = O_{L^2}(1). \quad (2.103)$$

This proves (2.83). We prove (2.84)–(2.87) one by one.

*T*: In any forest, the number of vertices equals the number of edges + the number of trees. Hence,  $T = n - N$ , where  $n$  is the number of trees in the depth-first forest, and thus Theorem 2.12 implies (2.84) with  $\tau = 1 - \psi$  given by (2.88).

*B*: Let  $B_t$  be the number of back arcs from  $v_t$ ; thus  $B = \sum_1^n B_t$ . Let  $\mathcal{F}_t$  be the  $\sigma$ -field generated by all arcs from  $v_i$ ,  $i \leq t$  (*i.e.*, by the outdegrees  $\eta_i$  and the endpoints of all these arcs); note that this includes complete information on the DFS until  $v_{t+1}$  is found, but also on some further arcs (the future arcs from the ancestors of  $v_{t+1}$ ). Then  $d(t)$  is  $\mathcal{F}_{t-1}$ -measurable and  $B_t$  is  $\mathcal{F}_t$ -measurable. Moreover,  $\eta_t$  is independent of  $\mathcal{F}_{t-1}$ . Thus, conditioned on  $\mathcal{F}_{t-1}$ , we still have  $\eta_t \sim \mathbf{P}$ ; we also know  $d(t)$ , and each arc from  $v_t$  is a back arc with probability  $d(t)/n$ . Hence,  $\mathbb{E}[B_t | \mathcal{F}_{t-1}, \eta_t] = \eta_t d(t)/n$ , and consequently

$$\mathbb{E}[B_t | \mathcal{F}_{t-1}] = \mathbb{E}\left[\eta_t \frac{d(t)}{n} | \mathcal{F}_{t-1}\right] = \lambda \frac{d(t)}{n}. \quad (2.104)$$

Similarly, since, as above,  $X \sim \text{Bin}(m, p)$  implies  $\mathbb{E} X^2 = mp(1-p) + (mp)^2$ ,

$$\mathbb{E}[B_t^2 | \mathcal{F}_{t-1}] = \mathbb{E}\left[\eta_t \frac{d(t)}{n} \left(1 - \frac{d(t)}{n}\right) + \left(\eta_t \frac{d(t)}{n}\right)^2 | \mathcal{F}_{t-1}\right] = O(1). \quad (2.105)$$

Define  $\Delta Z_t := B_t - \lambda d(t)/n$  and  $Z_t := \sum_1^t \Delta Z_i$ . Then (2.104) shows that  $\mathbb{E}[\Delta Z_t | \mathcal{F}_t] = 0$ , and thus  $(Z_i)_0^n$  is a martingale, with  $Z_0 = 0$ . Hence,  $\mathbb{E} Z_n = 0$ . Furthermore, (2.104) implies  $\mathbb{E}[(\Delta Z_t)^2] = \mathbb{E}[(B_t - \mathbb{E}[B_t | \mathcal{F}_{t-1}])^2] \leq \mathbb{E}[B_t^2]$ , and thus by (2.105),

$$\mathbb{E}[Z_n^2] = \text{Var}[Z_n] = \sum_{t=1}^n \mathbb{E}[(\Delta Z_t)^2] \leq \sum_{t=1}^n \mathbb{E}[B_t^2] = O(n). \quad (2.106)$$

Consequently,  $Z_n = O_{L^2}(n^{1/2})$ , and thus

$$B = \sum_{t=1}^n B_t = Z_n + \sum_{t=1}^n \lambda \frac{d(t)}{n} = \lambda \bar{d} + Z_n = \lambda \bar{d} + O_{L^2}(n^{1/2}). \quad (2.107)$$

Finally, (2.107) and Corollary 2.6 yield

$$B = \lambda \bar{d} + O_{L^2}(n^{1/2}) = \lambda \alpha n + O_{L^2}(n^{1/2}), \quad (2.108)$$

which shows (2.85) with  $\beta = \lambda \alpha$  as in (2.89), recalling (2.38).

*F*: By (2.100) and (2.84), we have (2.86) with  $\varphi = \frac{\lambda}{2} - \tau$ , which agrees with (2.89) by (2.88) and a simple calculation. In particular,  $\varphi = \beta$ .

*C*: Similarly, it follows from (2.97) and (2.85) that we have (2.87) with  $\chi := \lambda/2 - \beta$ . Since we have found  $\beta = \varphi$ , and we always have  $\tau + \varphi = \lambda/2 = \beta + \chi$ , see (2.97) and (2.100), we thus have  $\chi = \tau$ , and thus (2.88) holds.  $\square$

Note that  $T + F$  and  $B + C$  are asymptotically normal; this follows immediately from (2.91) and (2.92) by the central limit theorem.

**Conjecture 2.17.** *All four variables  $T, B, F, C$  are (jointly) asymptotically normal.*

The equalities  $\tau = \chi$  and  $\beta = \varphi$  mean asymptotic equality of the corresponding expectations of numbers of arcs. In fact, there are exact equalities.

**Theorem 2.18.** *For any  $n$ ,  $\mathbb{E}T = \mathbb{E}C$  and  $\mathbb{E}B = \mathbb{E}F = \lambda \mathbb{E}\bar{d}$ .*

*Proof.* Let  $v, w$  be two distinct vertices. If the DFS finds  $w$  as a descendant of  $v$ , then there will later be  $\text{Ge}(1-p)$  arcs from  $w$ , and each has probability  $1/n$  of being a back arc to  $v$ . Similarly, there will be  $\text{Ge}(1-p)$  future arcs from  $v$ , and each has probability  $1/n$  of being a forward arc to  $w$ . Hence, if  $I_{vw}$  is the indicator that  $w$  is a descendant of  $v$ , and  $B_{vw}$  [ $F_{vw}$ ] is the number of back [forward] arcs  $vw$ , then

$$\mathbb{E}B_{vw} = \mathbb{E}F_{vw} = \frac{\lambda}{n} \mathbb{E}I_{vw}. \quad (2.109)$$

Summing over all pairs of distinct  $v$  and  $w$ , we obtain

$$\mathbb{E}B = \mathbb{E}F = \frac{\lambda}{n} \mathbb{E} \sum_w \sum_{v \neq w} I_{vw} = \frac{\lambda}{n} \mathbb{E} \sum_w d(w) = \lambda \mathbb{E}\bar{d}. \quad (2.110)$$

Finally,  $\mathbb{E}T + \mathbb{E}F = \mathbb{E}C + \mathbb{E}B$  by (2.98) and (2.95), and thus (2.110) implies  $\mathbb{E}T = \mathbb{E}C$ .  $\square$

**Remark 2.19.** Knuth [12] conjectures, based on exact calculation of generating functions for small  $n$ , that, much more strongly,  $B$  and  $F$  have the same distribution for every  $n$ . (Note that  $T$  and  $C$  do not have the same distribution; we have  $T \leq n-1$ , while  $C$  may take arbitrarily large values.)

**Remark 2.20.** A simple argument with generating functions shows that the number of loops at a given vertex  $v$  is  $\text{Ge}(1-p/(n-np+p))$ ; these numbers are independent, and thus  $L \sim \text{NegBin}(n, 1-p/(n-np+p))$  with  $\mathbb{E}L = p/(1-p) = \lambda = O(1)$  and  $\text{Var}(L) = p(1-p+p/n)/(1-p)^2 = \lambda(1+\lambda/n) = O(1)$  [12]. Moreover, it is easily seen that asymptotically,  $L$  has a Poisson distribution,  $L \xrightarrow{d} \text{Po}(\lambda)$  as  $n \rightarrow \infty$ .

### 3. DEPTH, TREES AND ARC ANALYSIS IN THE SHIFTED GEOMETRIC OUTDEGREE DISTRIBUTION

In this section, the outdegree distribution is  $\text{Ge}_1(1-p) = 1 + \text{Ge}(1-p)$ . Thus we now have the mean

$$\lambda = \frac{1}{1-p}. \quad (3.1)$$

Thus  $\lambda > 1$ , and only the supercritical case occurs. As in Section 2, the depth  $d(t)$  is a Markov chain given by (2.6), but the distribution of  $\xi_t$  is now different. The

probability (2.2) is replaced by  $(1 - t/n)/(1 - pt/n)$ , but the number of future arcs from an ancestor is still  $\text{Ge}(1 - p)$ , and, with  $\theta := t/n$ ,

$$\mathbb{P}(\xi_t = k) = \begin{cases} \bar{\pi}_t := \frac{1-\theta}{1-p\theta}, & k = 0, \\ (1 - \bar{\pi}_t)(1 - \pi_t)^{k-1}\pi_t, & k \geq 1, \end{cases} \quad (3.2)$$

where  $\pi_t = p\bar{\pi}_t$  is as in (2.8). The rest of the analysis does not change, and the results in Theorems 2.4–2.16 still hold, but we get different values for many of the constants.

We now have

$$\mathbb{E} \xi_t = \bar{\mu}_t := \frac{(1-p)\theta}{p(1-\theta)} \quad (3.3)$$

and instead of (2.14) we have  $\mathbb{E} \tilde{d}(t) = n\tilde{\ell}(\theta) + O(1)$  where now  $\tilde{\ell}(\theta)$  takes the new value

$$\begin{aligned} \tilde{\ell}(\theta) &:= \int_0^\theta \left(1 - \frac{(1-p)x}{p(1-x)}\right) dx = \frac{1}{p}\theta + \frac{1-p}{p} \log(1-\theta) \\ &= \frac{1}{p}(\theta + \lambda^{-1} \log(1-\theta)). \end{aligned} \quad (3.4)$$

Note that  $\tilde{\ell}(\theta)$  in (3.4) is proportional to (2.15) for the (unshifted) geometric distribution with the same  $\lambda$ , but larger by a factor  $1/p$ . Figures 3 and 4 show  $\tilde{\ell}(\theta)$  for both geometric distributions with the same  $p$  (0.6) and the same  $\lambda$  (2.0), respectively.

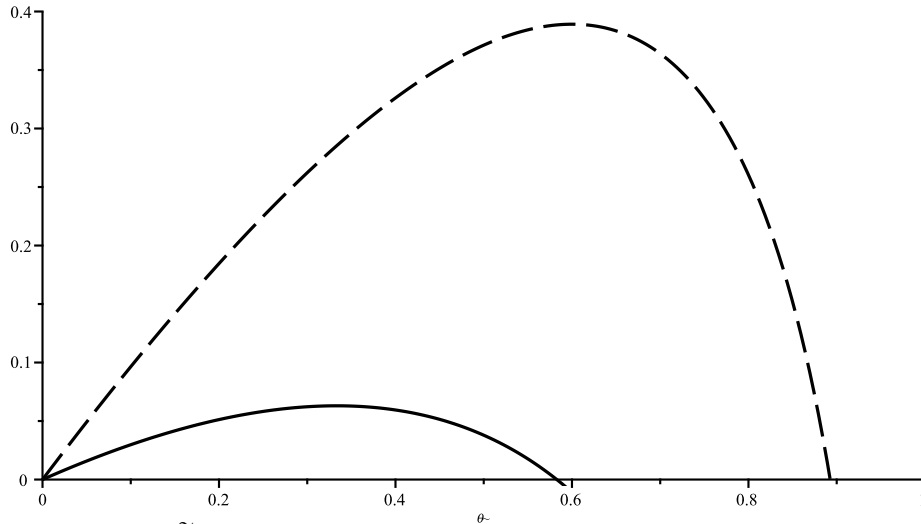


FIGURE 3.  $\tilde{\ell}(\theta)$ , the asymptotic search depth, for geometric distribution  $\text{Ge}(1 - p)$  (solid) and shifted geometric distribution  $\text{Ge}_1(1 - p)$  (dashed) with  $p = 0.6$  and thus  $\lambda = 1.5$  and  $\lambda = 2.5$ , respectively.

Note that  $\tilde{\ell}(\theta_1) = 0$  still gives (2.17) and (2.18), now with  $\lambda = 1/(1-p)$  as in (3.1), and that  $\lambda > 1$  so  $\theta_1 > 0$  for every  $p$ . Differentiating (3.4) shows that the maximum

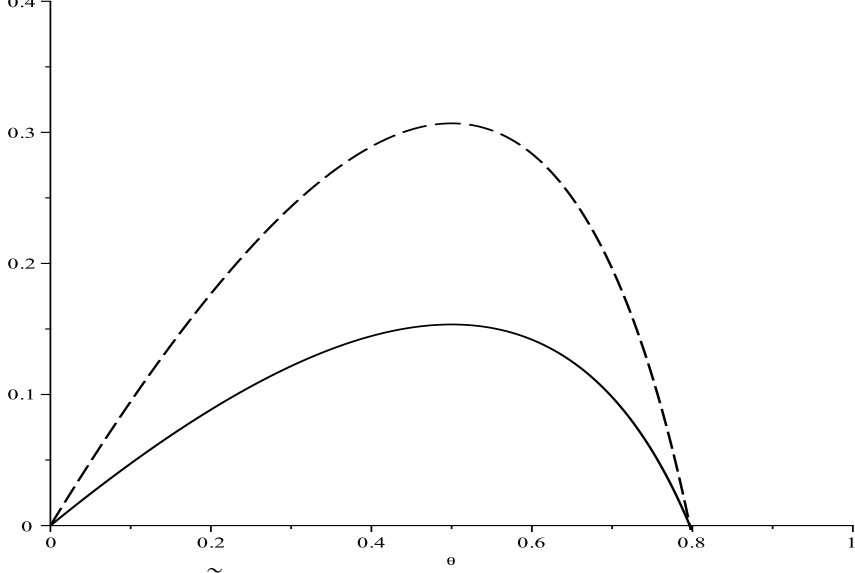


FIGURE 4.  $\tilde{\ell}(\theta)$ , the asymptotic search depth, for geometric distribution  $\text{Ge}(1-p)$  (solid) and shifted geometric distribution  $\text{Ge}_1(1-p)$  (dashed) with  $p = 2/3$  and  $1/2$ , respectively, and thus  $\lambda = 2$ .

point  $\theta_0 = p > 0$ , which again is given by (2.16). Straightforward calculations yield

$$v := \tilde{\ell}(p) = 1 + \frac{1-p}{p} \log(1-p) = 1 - \frac{1}{\lambda-1} \log \lambda, \quad (3.5)$$

$$\alpha := \frac{1}{p} \left( \frac{\theta_1^2}{2} - \frac{1}{\lambda} ((1-\theta_1) \log(1-\theta_1) + \theta_1) \right) = \theta_1 - \frac{\theta_1^2}{2p}. \quad (3.6)$$

Furthermore, (3.2) yields by a simple calculation, with  $\theta := t/n$ ,

$$\text{Var } \xi_t = \frac{(1-p)^2}{p^2(1-\theta)^2} + \frac{1-p}{p(1-\theta)} - \frac{1-p}{p^2}. \quad (3.7)$$

Hence, (2.44) holds with (2.45) replaced by

$$\begin{aligned} v(\theta) &:= \int_0^\theta \left( \frac{(1-p)^2}{p^2(1-x)^2} + \frac{1-p}{p(1-x)} - \frac{1-p}{p^2} \right) dx \\ &= \frac{(1-p)^2 \theta}{p^2(1-\theta)} - \frac{1-p}{p} \log(1-\theta) - \frac{1-p}{p^2} \theta, \end{aligned} \quad (3.8)$$

and then Lemma 2.8 and Theorem 2.10 hold with this  $v(\theta)$ .

Consequently, Theorem 2.11 holds with

$$\sigma^2 := v(p) = -\frac{1-p}{p} \log(1-p). \quad (3.9)$$

In the proof of Theorem 2.12, (2.68) for  $k \geq 1$  still holds with  $\mu_t$  given by (2.12) (except for the formula with  $\lambda$ ), and thus (2.69) is replaced by, using (3.3),

$$\mathbb{E}[d(t+1) \mid \mathcal{F}_t] = d(t) + 1 - \bar{\mu}_t + \mu_t \mathbb{E}[J_{t+1} \mid \mathcal{F}_t]. \quad (3.10)$$

The rest of the proof remains the same with minor modifications, and leads to, instead of (2.78), with  $\theta := t/n$ ,

$$\begin{aligned} \sum_{t=T_1+1}^n J_t &= \sum_{t=T_1}^{n-1} \frac{\bar{\mu}_t - 1}{\mu_t} + O_{L^2}(n^{1/2}) = \sum_{t=T_1}^{n-1} \frac{\frac{(1-p)\theta}{p(1-\theta)} - 1}{\frac{1-p}{p(1-\theta)}} + O_{L^2}(n^{1/2}) \\ &= \sum_{t=T_1}^{n-1} (1 - \lambda(1 - \theta)) + O_{L^2}(n^{1/2}), \end{aligned} \quad (3.11)$$

and thus Theorem 2.12 holds with

$$\psi := \int_{\theta_1}^1 (1 - \lambda(1 - x)) dx = 1 - \theta_1 - \frac{\lambda}{2}(1 - \theta_1)^2, \quad (3.12)$$

just as in (2.65).

In the proof of Theorem 2.16, (2.108) still holds, and we obtain (2.85) with  $\beta = \lambda\alpha$ , and then (2.87) with  $\chi = \lambda/2 - \beta$  just as before (but recall that  $\alpha$  now is different). On the other hand, now the expected numbers of back and forward arcs differ since  $\mathbb{E} B = \lambda \mathbb{E} \bar{d} \sim \lambda\alpha n$  and  $\mathbb{E} F = (\lambda - 1) \mathbb{E} \bar{d} \sim (\lambda - 1)\alpha n$  because the average number of future arcs at a vertex after a descendant has been created is  $\lambda - 1$ . The asymptotic formula (2.84) holds as above with  $\tau := 1 - \psi$ ; hence (2.98) implies that (2.86) holds too, with  $\varphi = \lambda/2 - \tau$ ; as just noted, we now have  $\varphi = (\lambda - 1)\alpha \neq \beta$ . Collecting these constants, we see that Theorem 2.16 holds with

$$\tau := 1 - \psi = \theta_1 + \frac{\lambda}{2}(1 - \theta_1)^2, \quad (3.13)$$

$$\beta := \lambda\alpha = \lambda\theta_1 - \frac{\lambda}{2p}\theta_1^2 = \lambda\theta_1 - \frac{\lambda^2}{2(\lambda - 1)}\theta_1^2, \quad (3.14)$$

$$\varphi := (\lambda - 1)\alpha = (\lambda - 1)\theta_1 - \frac{\lambda}{2}\theta_1^2 = \frac{\lambda}{2} - \tau, \quad (3.15)$$

$$\chi := \frac{\lambda}{2} - \beta = \frac{\lambda}{2}(1 - \theta_1)^2 + \frac{\lambda}{2(\lambda - 1)}\theta_1^2. \quad (3.16)$$

Thus the equality  $\beta = \varphi$  and the equality of the expected number of back and forward arcs in Theorems 2.16 and Theorem 2.18 was an artefact of the geometric degree distribution. Similarly,  $\chi = \lambda/2 - \beta < \lambda/2 - \varphi = \tau$ , and the equality of the expected numbers of tree arcs and cross arcs in Theorem 2.18 also does not hold.

We summarize the results above.

**Theorem 3.1.** *Let the outdegree distribution  $\mathbf{P}$  be the shifted geometric distribution  $\text{Ge}_1(p)$  with  $p \in (0, 1)$ . Then Theorems 2.4–2.16 hold, with the constants now having the values described above (and always  $\lambda > 1$ ), while Theorem 2.18 does not hold.*

#### 4. A GENERAL OUTDEGREE DISTRIBUTION: STACK SIZE

In this section, we consider a general outdegree distribution  $\mathbf{P}$ , with mean  $\lambda$  and finite variance.

When the outdegree distribution is general, the depth does not longer follow a simple Markov chain, since we would have to keep track of the number of children seen so far at each level of the branch of the tree toward the current vertex. Instead we get back a Markov chain if we instead of the depth consider the stack size  $I(t)$ , defined as follows.

The DFS can be regarded as keeping a stack of unexplored arcs, for which we have seen the start vertex but not the endpoint. The stack evolves as follows:

- S1. If the stack is empty, pick a new vertex  $v$  that has not been seen before (if there is no such vertex, we have finished). Otherwise, pop the last arc from the stack, and reveal its endpoint  $v$  (which is uniformly random over all vertices). If  $v$  already is seen, repeat.
- S2. ( $v$  is now a new vertex) Reveal the outdegree  $\eta$  of  $v$  and add to the stack  $\eta$  new arcs from  $v$ , with unspecified endpoints. GOTO S1.

Let again  $v_t$  be the  $t$ -th vertex seen by the DFS, and let  $I(t)$  be the size of the stack when  $v_t$  is found (but before we add the arcs from  $v_t$ ). It can easily be seen that  $I(t)$  will be a Markov chain, similar (but not identical) to the depth process  $d(t)$  in the geometric case studied above. Moreover, it is possible to recover the depth of the vertices from the stack size process, which makes it possible to extend many of the results above, although sometimes with less precision. For details, see [10].

#### REFERENCES

- [1] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25** (1997), no. 2, 812–854.
- [2] Krishna B. Athreya & Peter E. Ney. *Branching Processes*. Springer-Verlag, Berlin, 1972.
- [3] Patrick Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [4] Sahar Diskin & Michael Krivelevich. On the performance of the Depth First Search algorithm in supercritical random graphs. Preprint, 2021. [arXiv:2111.07345](https://arxiv.org/abs/2111.07345)
- [5] Nathanaël Enriquez, Gabriel Faure & Laurent Ménard. Limiting shape of the depth first search tree in an Erdős-Rényi graph. *Random Structures Algorithms* **56** (2020), no. 2, 501–516.
- [6] Allan Gut. *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.
- [7] Jean Jacod & Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 1987.
- [8] Svante 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.
- [9] Svante Janson. Functional limit theorems for multitype branching processes and generalized Polya urns. *Stoch. Process. Appl.* **110** (2004), 177–245.
- [10] Svante Janson & Philippe Jacquet. Depth-First Search performance in random digraphs with general outdegree distributions. In preparation.
- [11] Olav Kallenberg. *Foundations of Modern Probability*. 2nd ed., Springer-Verlag, New York, 2002.
- [12] Donald E. Knuth. *The Art of Computer Programming*, Section 7.4.1.2 (Preliminary draft, 13 February 2022). <http://cs.stanford.edu/~knuth/fasc12a.ps.gz>
- [13] Michael Krivelevich & Benny Sudakov. The phase transition in random graphs: a simple proof. *Random Structures Algorithms* **43** (2013), no. 2, 131–138.

INRIA SACLAY ILE DE FRANCE, FRANCE  
*Email address:* `philippe.jacquet@inria.fr`

DEPARTMENT OF MATHEMATICS, UPPSALA UNIVERSITY, PO Box 480, SE-751 06 UPPSALA,  
SWEDEN  
*Email address:* `svante.janson@math.uu.se`  
*URL:* `http://www2.math.uu.se/~svante/`