



**HAL**  
open science

# Depth-First Search performance in random digraphs

Philippe Jacquet, Svante Janson

► **To cite this version:**

Philippe Jacquet, Svante Janson. Depth-First Search performance in random digraphs. AofA 2024 - 35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, Jun 2024, Bath (UK), United Kingdom. <10.4230/LIPIcs.AofA.2024>. <hal-04888842>

**HAL Id: hal-04888842**

**<https://hal.science/hal-04888842v1>**

Submitted on 15 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# 1 Depth-First Search performance in random 2 digraphs

3 **Philippe Jacquet** ✉ 🏠 

4 Inria Saclay Ile de France, [ 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau], France

5 **Svante Janson** ✉ 🏠 

6 Department of Mathematics, Uppsala University, [PO Box 480, SE-751 06 Uppsala], Sweden

## 7 — Abstract —

---

8 We present an analysis of the depth-first search algorithm in a random digraph model with independent  
9 outdegrees having an arbitrary distribution with finite variance. The results include asymptotics for  
10 the distribution of the stack index and depths of the search. The search yields a series of trees of  
11 finite size before and after the exploration of a giant tree. Our analysis mainly concerns the giant  
12 tree. Most results are first order. This analysis proposed by Donald Knuth in his next to appear  
13 volume of *The Art of Computer Programming* gives interesting insight in one of the most elegant  
14 and efficient algorithm for graph analysis due to Tarjan.

15 **2012 ACM Subject Classification** Mathematics of computing → Trees

16 **Keywords and phrases** Depth First Search, random digraph, Analysis of Algorithms

17 **Digital Object Identifier** 10.4230/LIPIcs...8

18 **Funding** *Svante Janson*: Supported by the Knut and Alice Wallenberg Foundation and the Swedish  
19 Research Council (VR)

## 20 **1** Introduction

21 This paper is a continuation of our earlier paper [1], which studies a special case using a  
22 simpler method. The motivation of both paper is a new section in Donald Knuth's *The Art*  
23 *of Computer Programming* [3], which is dedicated to Depth-First Search (DFS) in a digraph.  
24 The DFS is an important computing tool dedicated to the exploration of large unstructured  
25 dataset, mostly organised as large directed graphs, and the DFS is the foundation of the daily  
26 crawling process of the graph of the Web performed by Google [2]. Briefly, the DFS starts  
27 with an arbitrary vertex, and explores the arcs from that vertex one by one. When an arc is  
28 found leading to a vertex that has not been seen before, the DFS explores the arcs from it in  
29 the same way, in a recursive fashion, before returning to the next arc from its parent. This  
30 eventually yields a tree containing all descendants of the the first vertex (which is the root of  
31 the tree). If there still are some unseen vertices, the DFS starts again with one of them and  
32 finds a new tree, and so on until all vertices are found. We refer to [3] for details as well as  
33 for historical notes. (See also S1–S2 in Section 2.) Note that the digraphs in [3] and here  
34 are multi-digraphs, where loops and multiple arcs are allowed. (Although in our random  
35 model they are few and usually not important.) The DFS algorithm generates a spanning  
36 forest (the *depth-first forest*) in the digraph, with all arcs in the forest directed away from the  
37 roots. Our main purpose is to study the properties of the depth-first forest, starting with a  
38 random digraph  $G$ ; in particular we study the distribution of the stack index and the depth  
39 of vertices in the depth-first forest.

40 The random digraph model that we consider has  $n$  vertices and a given outdegree  
41 distribution  $\mathbf{P}$ . The outdegrees (number of outgoing arcs) of the  $n$  vertices are independent  
42 random numbers with this distribution. The endpoint of each arc is uniformly selected at  
43 random among the  $n$  vertices, independently of all other arcs. (Therefore, an arc can loop



© Philippe Jacquet and Svante Janson;

licensed under Creative Commons License CC-BY 4.0

Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

back to the starting vertex, and multiple arcs can occur.) We consider asymptotics as  $n \rightarrow \infty$  for a fixed outdegree distribution. In [1], we studied the case when the outdegree distribution is geometric. In the present paper, we generalise this and let  $\mathbf{P}$  be an arbitrary distribution; however, we assume throughout the paper that the outdegree distribution  $\mathbf{P}$  has a finite second moment.

► Remark 1. Related results are proved by (author?) [5] for DFS in an undirected Erdős–Rényi graph  $G(n, \lambda/n)$ ; the case when  $\lambda = 1 + \varepsilon$  for small  $\varepsilon > 0$  is studied further by another argument by (author?) [8]. DFS in the random digraph  $D(n, p)$  has also been considered previously, for example in the proof of [4, Theorem 3]. Although this is for a different random graph model, DFS on  $G(n, \lambda/n)$  is the same as DFS on the Erdős–Rényi digraph  $D(n, \lambda/n)$ , which is essentially the same as the digraph studied in the present paper with outdegree distribution  $\text{Po}(\lambda)$ . Hence, the main result of [5], which shows convergence of the depth profile in the depth-first forest to a certain deterministic limit, is essentially the special case  $\mathbf{P} = \text{Po}(\lambda)$  of our result for the depths (Theorem 6). The proofs are quite different. See also (author?) [6], where related results are given for the (undirected) configuration model.

We analyze the process  $d(t)$  of depths of the vertices, in the order they are found by the DFS. For the geometric outdegree distribution studied in [1],  $d(t)$  is a Markov chain, which was used in our proofs. For general outdegree distributions, this is no longer true. We show in Section 2 that we can use the size  $I(t)$  of the stack of arcs kept by the DFS as a substitute; this is a Markov chain, and we obtain limit results for the stack size with deviation in  $O_{L^2}(n^{1/2})$ . (See Section 1.1 below for notation.) In a second step (Section 3), this is used to derive limit results for the depths  $d(t)$ , but the results obtained are within deviation in  $o_p(n)$  which is close to the order of the result. We give also an alternative approach in Section 4 where the depths  $d(t)$  are analysed directly by a different method; the results there are preliminary and less complete, but it seems that this method yields sharper results than the method in Section 3, with deviation within  $O_p(n^\beta)$  with  $\beta$  arbitrary close to  $4/5$ ; thus the worst case error is now negligible compared to the main order.

Many details and further results will be given in the forthcoming full paper.

## 1.1 Some notation

We denote the given outdegree distribution by  $\mathbf{P}$ . We let  $\eta$  denote random variables with this distribution. In particular, we denote the outdegree of vertex  $v$  by  $\eta(v)$ . Recall that our standing assumption is that these outdegrees are i.i.d. (independent and identically distributed) with  $\eta(v) \sim \mathbf{P}$ . We let  $v_t$  denote the  $t$ -th vertex found by the DFS, and simplify notation by letting  $\eta_t := \eta(v_t)$  be its outdegree. It follows from the construction of the DFS that also the random variables  $\eta_t$ ,  $t = 1, \dots, n$  are i.i.d. with distribution  $\mathbf{P}$ ; this fundamental property will be used repeatedly without further mention.

We assume throughout that the second moment  $\mathbb{E}\eta^2 < \infty$ . This is essential for some results (e.g. results on asymptotic normality), but we conjecture that many results hold assuming only that the first moment  $\mathbb{E}\eta < \infty$ .

The mean outdegree, *i.e.*, the expectation  $\mathbb{E}\eta$  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  $\lambda > 1$ .

Let  $d(v)$  be the depth of vertex  $v$ , and let  $d(t) := d(v_t)$ .

As usual, w.h.p. means *with high probability*, *i.e.*, with probability  $1 - o(1)$  as  $n \rightarrow \infty$ .

For random variables  $X_n$  and positive numbers  $a_n$ , we write  $X_n = o_p(a_n)$  if, as  $n \rightarrow \infty$ ,  $X_n/a_n \xrightarrow{P} 0$ , *i.e.*, if for every  $\varepsilon > 0$ , we have  $\mathbb{P}(|X_n| > \varepsilon a_n) \rightarrow 0$ . Furthermore,  $X_n = O_p(a_n)$

90 means that the family  $\{X_n/a_n\}$  is bounded in probability, *i.e.*, if for every  $\varepsilon > 0$  there exists  
 91  $C$  such that  $\mathbb{P}(|X_n| > Ca_n) < \varepsilon$  for all  $n$ .  $X_n = o_{L^2}(a_n)$  means  $\mathbb{E}[|X_n/a_n|^2] \rightarrow 0$ , and  
 92  $X_n = O_{L^2}(a_n)$  means  $\mathbb{E}[|X_n/a_n|^2] = O(1)$ .

93  $\text{Ge}(p)$  denotes the geometric distribution with parameter  $p \in (0, 1]$ ; thus  $\xi \sim \text{Ge}(p)$  means  
 94 that  $\xi$  is a random variable with

$$95 \quad \mathbb{P}(\xi = k) = (1 - p)^k p, \quad k = 0, 1, 2, \dots \quad (1.1)$$

96 We write  $x^+ := \max(x, 0)$ .

## 97 **2 Stack index analysis**

98 As said above, unlike the special case  $\mathbf{P} = \text{Ge}(1 - p)$  studied in [1], in general the depth  
 99  $d(t)$  does not follow a Markov chain. Therefore, we instead first consider the stack index  
 100  $I(t)$  defined below, which does evolve as a Markov chain; many arguments for  $I(t)$  below are  
 101 similar to arguments for  $d(t)$  in [1, Section 2].

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

104 **S1.** If the stack is empty, pick a new vertex  $v$  that has not been seen before (if there is no such  
 105 vertex, we have finished). Otherwise, pop the last arc from the stack, and reveal its endpoint  
 106  $v$  (which is uniformly random over all vertices). If  $v$  already is seen, repeat.

107 **S2.** ( $v$  is now a new vertex) Reveal the outdegree  $\eta(v)$  of  $v$  and add to the stack  $\eta(v)$  new arcs  
 108 from  $v$ , with unspecified endpoints. GOTO S1.

109 Let  $I(t)$  be the size of the stack when  $v_t$  is found (but before we add the arcs from  $v_t$ ).  
 110 After  $v_t$  is found, and the  $\eta_t$  arcs from  $v_t$  have been added to the stack, the stack size is thus  
 111  $I(t) + \eta_t$ . We next perform step S1 one or several times. As long as the stack is not empty,  
 112 we find each time an already seen vertex with probability  $t/n$ , and in this case we repeat S1.  
 113 Hence, conditioned on  $I(t)$  and  $\eta_t$ , for  $k \geq 1$ , the probability that S1 is performed exactly  $k$   
 114 times is

$$115 \quad \left(\frac{t}{n}\right)^{k-1} \left(1 - \frac{t}{n}\right), \quad (2.1)$$

116 provided  $1 \leq k \leq I(t) + \eta_t$ , and if none of these events occur, then S1 is repeated a last  
 117 time and a new vertex is picked that will be the root of a new tree in the depth-first forest  
 118 (unless we have finished the DFS). Note that (2.1) equals the probability  $\mathbb{P}(\xi = k)$  with  
 119  $\xi_t \sim \text{Ge}(1 - t/n)$ , see (1.1). Thus, we can write

$$120 \quad I(t + 1) = (I(t) + \eta_t - 1 - \xi_t)^+, \quad 1 \leq t < n, \quad (2.2)$$

121 where  $\xi_t \sim \text{Ge}(1 - t/n)$  is a random variable independent of the history; more precisely,  
 122 the random variables  $\eta_t$  ( $1 \leq t \leq n$ ) and  $\xi_t$  ( $1 \leq t < n$ ) are all independent. We start the  
 123 stochastic recursion (2.2) with  $I(1) = 0$ .

124 We define

$$125 \quad \zeta_t := \eta_t - \xi_t - 1. \quad (2.3)$$

126 Thus, (2.2) can be written  $I(t + 1) = (I(t) + \zeta_t)^+$ . We see also that, for  $1 \leq t < n$ ,

$$127 \quad v_{t+1} \text{ is a root} \iff I(t) + \zeta_t < 0 \iff I(t + 1) > I(t) + \zeta_t. \quad (2.4)$$

## 8:4 Depth-First Search performance in random digraphs

128 Note that the random variables  $\zeta_t$  are independent, but have different distributions. We  
129 define also

$$130 \quad \tilde{I}(t) := \sum_{i=1}^{t-1} (\eta_i - 1 - \xi_i) = \sum_{i=1}^{t-1} \zeta_i \quad (2.5)$$

131 and

$$132 \quad I_*(t) := \min_{1 \leq j \leq t} \tilde{I}(j). \quad (2.6)$$

133 It follows from the recursion (2.2) and induction that

$$134 \quad I(t) = \tilde{I}(t) - I_*(t). \quad (2.7)$$

135 By (2.5) and (2.7),

$$136 \quad \tilde{I}(t+1) = \tilde{I}(t) + \zeta_t = I_*(t) + I(t) + \zeta_t \quad (2.8)$$

137 and thus it follows from (2.4) and (2.6) that, for  $1 \leq t < n$ ,

$$138 \quad v_{t+1} \text{ is a root} \iff I(t) + \zeta_t < 0 \iff \tilde{I}(t+1) < I_*(t) \iff I_*(t+1) < I_*(t). \quad (2.9)$$

139 Obviously,  $v_1$  is also a root, with  $I(1) = \tilde{I}(1) = I_*(1) = 0$ .

140 We have

$$141 \quad \mathbb{E} \zeta_t = \mathbb{E} \eta_t - \mathbb{E} \xi_t - 1 = \lambda - \frac{t/n}{1 - t/n} - 1 = \lambda - \frac{1}{1 - t/n}. \quad (2.10)$$

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

$$143 \quad \mathbb{E} \tilde{I}(t) = \sum_{i=1}^{t-1} \mathbb{E} \zeta_i = (t-1)\lambda - \sum_{i=1}^{t-1} \frac{1}{1 - i/n} = n\tilde{\iota}(t/n) + O(1), \quad (2.11)$$

144 where

$$145 \quad \tilde{\iota}(\theta) := \int_0^\theta \left( \lambda - \frac{1}{1-x} \right) dx = \lambda\theta + \log(1-\theta). \quad (2.12)$$

146 Let, as in [1],

$$147 \quad \theta_0 := (1 - \lambda^{-1})^+ = \begin{cases} 1 - \lambda^{-1}, & \lambda > 1, \\ 0, & \lambda \leq 1, \end{cases} \quad (2.13)$$

148 and let  $\theta_1$  is the largest root in  $[0, 1)$  of  $\tilde{\iota}(\theta_1) = 0$ ; thus

$$149 \quad \log(1 - \theta_1) = -\lambda\theta_1 \quad (2.14)$$

150 and  $\theta_1$  equals the survival probability of a Bienayme–Galton–Watson (BGW) process with  
151  $\text{Po}(\lambda)$  offspring distribution. Define

$$152 \quad \tilde{\iota}^+(\theta) := [\tilde{\iota}(\theta)]^+ = \begin{cases} \lambda\theta + \log(1 - \theta), & 0 \leq \theta \leq \theta_1, \\ 0, & \theta_1 \leq \theta \leq 1, \end{cases} \quad (2.15)$$

153 It is easy to see that if  $\lambda \leq 1$ , then  $\tilde{\iota}^+(\theta) = 0$  for all  $\theta \in [0, 1]$ , while if  $\lambda > 1$ , then  $\tilde{\iota}^+(\theta) > 0$   
154 for  $0 < \theta < \theta_1$  (where  $0 < \theta_1 < 1$ ), with a maximum at  $\theta_0 = 1 - \lambda^{-1}$ .

155 We now can argue similarly to [1], using  $I(t)$  instead of  $d(t)$ . The proof of [1, Theorem  
156 2.4] applies with very minor differences, and yields:

157 ▶ **Theorem 2.** *Suppose that the outdegree distribution has finite variance. Then*

$$158 \quad \max_{1 \leq t \leq n} |I(t) - n\tilde{\iota}^+(t/n)| = O_{L^2}(n^{1/2}). \quad (2.16)$$

159 *Furthermore, for every  $\varepsilon > 0$ , we also have*

$$160 \quad \max_{1 \leq t \leq (1-\varepsilon)n} |\tilde{I}(t) - n\tilde{\iota}(t/n)| = O_{L^2}(n^{1/2}). \quad (2.17)$$

161 We can use Theorem 2 to show the following results, extending [1, Theorems 4.1 and 4.3]  
 162 to general outdegree distributions. The proofs are similar to the ones in [1]; details will be  
 163 given in the full paper.

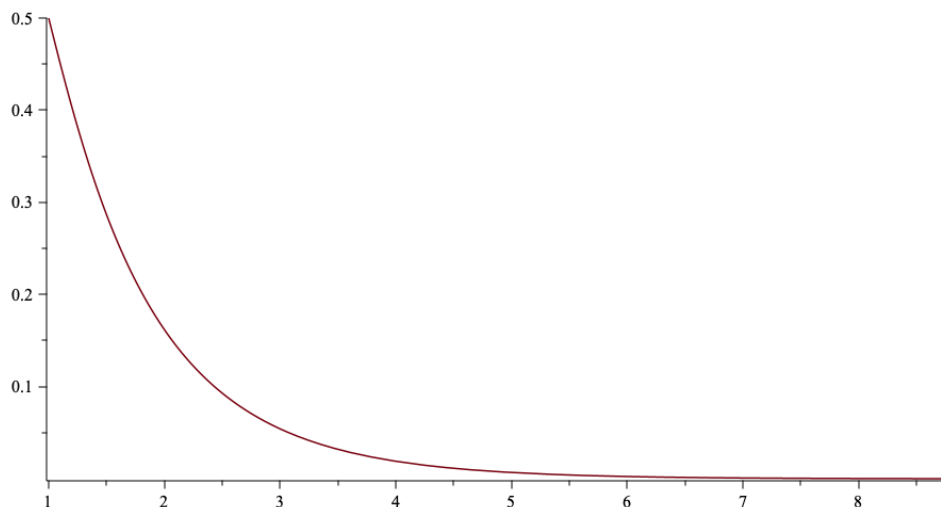
164 ▶ **Theorem 3.** *Suppose that the outdegree distribution has finite variance. Let  $N$  be the*  
 165 *number of trees in the depth-first forest. Then*

$$166 \quad N = \psi n + O_{L^2}(n^{1/2}), \quad (2.18)$$

167 *where*

$$168 \quad \psi := 1 - \theta_1 - \frac{\lambda}{2}(1 - \theta_1)^2. \quad (2.19)$$

Figure 1 shows the parameter  $\psi$  as a function of the average degree  $\lambda$ .



■ **Figure 1**  $\psi$ , as function of  $\lambda$ .

169

170 ▶ **Theorem 4.** *Suppose that the outdegree distribution has finite variance. Let  $\mathbf{T}_1$  be the*  
 171 *largest tree in the depth-first forest.*

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

175 More precise results will be given in the full paper.

176 **3 Stack size and depth**

177 We can recover the depth of the vertices from the stack size process, or more precisely, from  
 178 the process  $\tilde{I}(t)$  defined above; this uses a method that has been used by several authors in  
 179 the study of other random trees, see for example [9, (1.1)] and [7, Proposition 4].

180 When vertex  $v_t$  is found by the DFS, the stack consist of all future arcs from the ancestors  
 181 of  $v_t$  in the depth-first forest. Hence, the stack size  $I(t)$  is the total number of future arcs from  
 182 the ancestors of  $v_t$ . As the DFS continues, it first explores the descendants of  $v_t$ . During this  
 183 period, the ancestors of  $v_t$  are still ancestors of the current vertex, and their future arcs remain  
 184 the same as when  $v_t$  was found. Hence, if  $v_s$  is a descendant of  $v_t$ , then  $I(s) \geq I(t)$ . Moreover,  
 185 we then also have  $I_*(s) = I_*(t)$  by (2.9), since no new root has been found, and thus, by  
 186 (2.7),  $\tilde{I}(s) \geq \tilde{I}(t)$ . On the other hand, when the DFS has finished exploring descendants of  $v_t$ ,  
 187 then it has explored all arcs from  $v_t$  and the DFS next backtracks to the ancestors of  $v_t$  by  
 188 exploring the future edges of the ancestors of  $v_t$  one by one. If another vertex  $v_u$  is found  
 189 in this way, then  $I(u) < I(t)$ ; furthermore,  $I_*(u) = I_*(t)$  as above, and thus  $\tilde{I}(u) < \tilde{I}(t)$ .  
 190 Finally, if no further new vertex is found in the same tree as  $v_t$ , then the DFS next finds a  
 191 new root  $v_u$ ; in this case,  $I(u) = 0 \leq I(t)$ , and, by (2.9),  $I_*(u) < I_*(t)$ . Hence,  $\tilde{I}(u) < \tilde{I}(t)$   
 192 holds in this case too. This leads to the following characterisation. (Details will be given in  
 193 the full paper.)

194 **► Lemma 5.** (i) For any  $s, t \in [n]$ ,  $v_s$  is a descendant of  $v_t$  if and only if  $s > t$  and

195 
$$\min_{t \leq i \leq s} \tilde{I}(i) = \tilde{I}(t). \tag{3.1}$$

196 (ii) For any  $t \geq 1$ , the ancestors of  $v_t$  are the vertices  $v_s$  corresponding to the weak  
 197 right-to-left minima of  $(\tilde{I}(s))_{1 \leq s \leq t}$ , excluding  $s = t$ . Hence the depth  $d(t)$  of  $v_t$  is the number  
 198 of such right-to-left minima.

199 Let  $G_\eta(z) := \sum_k \mathbb{P}(\eta = k)z^k$  be the probability generating function of  $\eta$ . For  $\theta \in [0, 1]$ ,  
 200 let  $\eta_\theta$  be a random variable obtained by thinning the outdegree  $\eta$  and keeping each arc with  
 201 probability  $1 - \theta$ , independently; thus  $\eta_\theta$  has probability generating function

202 
$$G_{\eta_\theta}(z) = G_\eta(\theta + (1 - \theta)z) = G_\eta(1 - (1 - \theta)(1 - z)). \tag{3.2}$$

203 Note that the number of outgoing arcs from  $v_t$  to vertices that are not discovered earlier  
 204 has the distribution of  $\eta_\theta$  with  $\theta = t/n$ . Let  $\rho(\theta)$  be the survivability probability that the  
 205 Bienayme–Galton–Watson (BGW) process with offspring distribution  $\eta_\theta$  has infinite size. It  
 206 satisfies the identity:

207 
$$1 - \rho(\theta) = G_{\eta_\theta}(1 - \rho(\theta)) = G_\eta(\theta + (1 - \theta)(1 - \rho(\theta))) = G_\eta(1 - (1 - \theta)\rho(\theta)). \tag{3.3}$$

208 We have  $\mathbb{E}\eta_\theta = (1 - \theta)\lambda$ . Hence, if  $(1 - \theta)\lambda > 1$ , then  $\rho(\theta) > 0$ ; otherwise  $\rho(\theta) = 0$ .

209 Define

210 
$$\tilde{\ell}(\theta) := \int_0^\theta \rho(x) dx, \quad 0 \leq \theta \leq \theta_0, \tag{3.4}$$

211 and

212 
$$\tilde{\ell}^+(\theta) := \begin{cases} \tilde{\ell}(\theta) = \int_0^\theta \rho(x) dx, & 0 \leq \theta \leq \theta_0, \\ \tilde{\ell}(\tilde{\theta}), \text{ where } \tilde{\theta} \in (0, \theta_0) \text{ and } \tilde{\ell}(\tilde{\theta}) = \tilde{\ell}(\theta), & \theta_0 < \theta < \theta_1, \\ 0, & \theta_1 \leq \theta \leq 1. \end{cases} \tag{3.5}$$

213 If  $\lambda \leq 1$ , this simply means  $\tilde{\ell}(\theta) := 0$  for all  $\theta \in [0, 1]$ .

214 Using Lemma 5 and Theorem 2 we can obtain the following result; the detailed argument  
215 (given in the full paper) is unfortunately rather long.

216 ▶ **Theorem 6.** *We have*

$$217 \quad \max_{1 \leq t \leq n} |d(t) - n\tilde{\ell}^+(t/n)| = o_{L^2}(n). \quad (3.6)$$

## 218 **4 Depth analysis**

219 In this section we sketch an alternative approach where we study the depth directly, without  
220 using the stack index. Details will be given in the full paper. In particular, for simplicity  
221 we approximate below generating functions and other quantities by their limits as  $n \rightarrow \infty$ ,  
222 sometimes omitting careful estimates of the errors. We consider only the case  $\lambda > 1$ , since  
223 otherwise there is no giant tree and the depth  $d_t$  is small for all  $t$ .

224 Recall that the outdegree  $\eta$  has the p.g.f.  $G_\eta(z) = \sum_k \mathbb{P}(\eta = k)z^k$ , and that at time  $t$ , the  
225 number of outgoing arcs from  $v_t$  that lead to vertices not already seen has the distribution  
226  $\eta_\theta$  with p.g.f.  $G_{\eta_\theta}(z) = \eta(\theta + (1 - \theta)z)$ , see (3.2).

### 227 **4.1 Splitting the giant tree**

228 Recall the definitions of  $\theta_0$  and  $\theta_1$  from (2.13)–(2.14). It is easily seen that these are,  
229 roughly, the proportions of discovered nodes when we find the top of the giant tree, and  
230 when we leave the giant tree, respectively. Let  $\varepsilon \in (0, 1)$  be a fixed number which will be  
231 specified later. We divide the interval  $[0, n\theta_1]$  into three intervals:  $\mathcal{U}_n = [0, \theta_0 n - n^{1-\varepsilon}]$   
232 called the *climbing zone*,  $\mathcal{D}_n = [\theta_0 n + n^{1-\varepsilon}, \theta_1 n]$ , called the *descending zone*, and the band  
233  $\mathcal{B}_n = [\theta_0 n - n^{1-\varepsilon}, \theta_0 n + n^{1-\varepsilon}]$ , called the *neutral zone*.

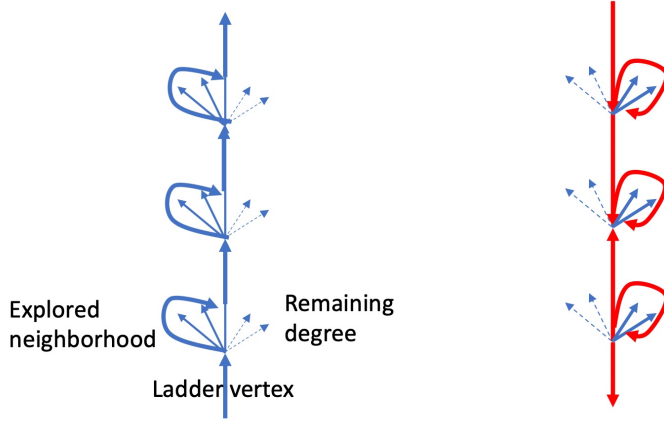
234 For  $t \in \mathcal{U}_n$ , we say that  $v_t$  is a *ladder vertex* if for all  $t' \in \mathcal{U}_n$ :  $t' > t \Rightarrow d(t') \geq d(t)$ . In  
235 other words,  $t$  is a right minimum of the function  $d(t)$  in  $\mathcal{U}$ . For an integer  $i$ , we define  $d_1^{-1}(i)$   
236 as the last integer  $t$  in  $\mathcal{U}_n$  such that  $d(t) = i$  (provided such a  $t$  exists); note that then  $v_t$  is a  
237 ladder vertex. Also, we say that  $t$  is an *up-time* if  $v_t$  is a ladder vertex and  $t$  is the first time  
238 it is visited.

239 As a property of DFS, the ladder vertices are also revisited in the descending zone. Let  $v$   
240 be a ladder vertex visited at time  $t$  in the climbing zone. When revisited the depth would be  
241 back at  $d(t)$ , see Figure 2.

### 242 **4.2 Average interval between two up-times**

243 Assume that  $t \in \mathcal{U}_n$  is an up-time of a ladder vertex  $v_t$  with depth  $d(t) = i$ . Then (assuming  
244 that  $v_t$  is not the last ladder vertex) the next up-time  $t'$  finds the next ladder vertex  $v_{t'}$ .  
245 Clearly,  $v_{t'}$  must be a child of  $v_t$  in the DFS, so its depth  $d(t') = i + 1$ , and it is the first child  
246 such that the DFS does not return to  $v_t$  before the end of  $\mathcal{U}_n$ . Given an up-time  $t$ , the future  
247 events in  $\mathcal{U}_n$ , including the value of  $t'$ , are independent of the past history of the DFS. If we  
248 ignore the variation of  $\theta$  in the interval  $[t, t']$ , then the DFS there can be regarded as a BGW  
249 tree  $\mathcal{T}_\theta$  (rooted at  $v_t$ ) with offspring distribution  $\eta_\theta$ ; since we assume that  $\theta < \theta_0$ , the BGW  
250 tree  $\mathcal{T}_\theta$  is supercritical. We denote by  $g_\theta(z)$  the generating function of the total number of  
251 vertices in  $\mathcal{T}_\theta$  when the number is finite; it satisfies the fixed point equation:

$$252 \quad g_\theta(z) = zG_{\eta_\theta}(g_\theta(z)). \quad (4.1)$$



■ **Figure 2** The DFS in the giant tree: the climbing phase (left) and the descending phase (right) through ladder vertices.

253 thus  $g_\theta(1) = 1 - \rho(\theta)$ , the BGW tree extinction probability in (3.3).

254 We denote by  $H(z, \theta)$  the probability generating function of the number of visited nodes  
 255 in the subtrees of the root of  $\mathcal{T}_\theta$  before hitting an infinite subtree, under the condition that  
 256 the root has indeed an infinite subtree. Let  $H'$  denote the derivative with respect to the first  
 257 variable  $z$ .

258 ▶ **Theorem 7.** *We have the identity*

$$259 \quad H(z, \theta) = 1 + \frac{z - 1}{1 - g_\theta(z)} \tag{4.2}$$

260 and thus the expected number of visited nodes in the BGW tree is

$$261 \quad H'(1, \theta) = \frac{1}{1 - g_\theta(1)} = \frac{1}{\rho(\theta)} = O((\theta_0 - \theta)^{-1}). \tag{4.3}$$

262 The p.g.f. of the number of visited vertices between two consecutive up-times  $t$  and  $t'$  is thus  
 263 asymptotically, with  $\theta = t/n$

$$264 \quad H\left(z, \theta \pm O\left(\frac{1}{n(\theta_0 - \theta)^2}\right)\right) \tag{4.4}$$

265 and the average number of new visited vertices is  $H'(1, \theta) + O\left(\frac{1}{n(\theta_0 - \theta)^4}\right)$ .

266 **Proof.** If  $k$  is the number of children of the root in the BGW tree  $\mathcal{T}_\theta$ , then the generating  
 267 function of the number of subtrees before finding an infinite one is  $\sum_{\ell < k} z^\ell (1 - \rho(\theta))^\ell \rho(\theta)$ .  
 268 The generating function of the number of visited vertices in the BGW tree before finding an  
 269 infinite subtree is thus  $z\rho(\theta) \sum_{\ell < k} (g_\theta(z))^\ell$ . Summing over the offspring distribution we get  
 270 the unconditional generating function

$$271 \quad z \frac{\rho(\theta)}{1 - g_\theta(z)} (G_{\eta_\theta}(1) - G_{\eta_\theta}(g_\theta(z))) = \frac{\rho(\theta)}{1 - g_\theta(z)} (z - g_\theta(z)) \\ 272 \quad = \rho(\theta) \left( \frac{z - 1}{1 - g_\theta(z)} + 1 \right). \tag{4.5}$$

273 To get the p.g.f.  $H(z, \theta)$  of the number of discovered nodes conditioned on  $\mathcal{T}_\theta$  having at least  
 274 one infinite subtree, one must divide by  $\rho(\theta)$ , which is the probability that the BGW tree is

275 infinite. The error term comes from the fact that the fraction of discovered vertices varies  
 276 during the interval in the proportion of the size of the interval divided by  $n$ . If we look at  
 277 the generating function  $H(z, \theta)$  it turns out (see the appendix) that the interval multiplied  
 278 by  $(\theta_0 - \theta)^2$  is bounded in probability, thus a variation of  $\theta$  of order  $O((\theta_0 - \theta)^{-2}/n)$ . ◀

279 Let  $\Delta_i$  be the time interval between two consecutive up-times  $t$  and  $t'$  such that  $d(t) = i$ .  
 280 To get the number  $u(t, t')$  of up-times in an interval  $(t, t']$  one must find the probability that  
 281  $\Delta_i + \Delta_{i+1} + \dots + \Delta_{i+k} \leq t' - t$ . Since the up-times are renewal points in the interval  $\mathcal{U}_n$ , we  
 282 have, given that  $t$  is an up-time,

$$283 \quad \begin{cases} \mathbb{P}(\Delta_i + \Delta_{i+1} + \dots + \Delta_{i+k} > t' - t) &= \mathbb{P}(u(t, t') < k) \\ \mathbb{P}(\Delta_i + \Delta_{i+1} + \dots + \Delta_{i+k} \leq t' - t) &= \mathbb{P}(u(t, t') \geq k). \end{cases} \quad (4.6)$$

284 Thus by Chebychev's inequality, we have the Chernoff type bounds

$$285 \quad \forall x > 0 : \begin{cases} \mathbb{P}(u(t, t') < k) &\leq H(e^x, t/n)^k e^{-(t'-t)x} \\ \mathbb{P}(u(t, t') \geq k) &\leq H(e^{-x}, t'/n)^k e^{(t'-t)x} \end{cases} \quad (4.7)$$

286 The estimate of these probabilities will come from an easy application of generating  
 287 functions.

288 ▶ **Lemma 8.** *When the quantity  $\theta$  is smaller than but close to  $\theta_0$ , we have  $\rho(\theta) = 1 - g_\theta(1) =$   
 289  $2\frac{\lambda^3}{\lambda_2}(\theta_0 - \theta) + O((\theta_0 - \theta)^2)$  where  $\lambda_2 = G''_\eta(1)$ . Furthermore the generating function  $g_\theta(z)$   
 290 has a radius of convergence which is  $1 + \frac{\lambda^4}{2\lambda_2}(\theta_0 - \theta)^2 + O((\theta_0 - \theta)^3)$ , and for  $z$  such that  
 291  $|z - 1| \ll (\theta_0 - \theta)^2$  we have*

$$292 \quad g_\theta(z) = 1 - 2\frac{\lambda^3}{\lambda_2}(\theta_0 - \theta) + \frac{z - 1}{\lambda(\theta_0 - \theta)} + O\left(\frac{(z - 1)^2}{(\theta_0 - \theta)^3} + (\theta_0 - \theta)^2\right). \quad (4.8)$$

$$293 \quad H(z, \theta) = 1 + \frac{\lambda_2(z - 1)}{2\lambda^3(\theta_0 - \theta)} + O\left(\frac{(z - 1)^2}{(\theta_0 - \theta)^3} + (\theta_0 - \theta)^2\right) \quad (4.9)$$

$$294 \quad H'(1, \theta) = \frac{\lambda_2}{2\lambda^3(\theta_0 - \theta)} + O(1). \quad (4.10)$$

295

296 **Proof.** In Appendix A. ◀

297 ▶ **Lemma 9.** *Let  $\varepsilon > 0$  and  $\frac{1-\varepsilon}{2} < \alpha < 1 - \varepsilon$ . For  $a, A > 0$  there exist  $B > 0$  and  $C > 0$  such  
 298 that for  $t < t' \in \mathcal{U}_n$  with  $an^{1-\varepsilon} \leq t' - t \leq An^{1-\varepsilon}$  we have*

$$299 \quad \mathbb{P}\left(u(t, t') \geq \frac{1}{H'(1, \theta)}(t' - t + Bn^\alpha)\right) < \exp(-Cn^{2\alpha-1-\varepsilon}), \quad (4.11)$$

$$300 \quad \mathbb{P}\left(u(t, t') < \frac{1}{H'(1, \theta')}(t' - t - Bn^\alpha)\right) < \exp(-Cn^{2\alpha-1-\varepsilon}). \quad (4.12)$$

301

302 **Proof.** Now  $t$  and  $t'$  can be within  $n^{1-\varepsilon}$  from  $n\theta_0$  which has an impact on the estimate of  
 303  $\log H(e^x, \theta)$ . For the super-critical BGW tree, we know from Lemma 8 and its proof that  
 304  $H'(1, \theta) = O((\theta_0 - \theta)^{-1})$  and that, for  $|x| \ll (\theta_0 - \theta)^2$ ,

$$305 \quad H(e^x, \theta) \leq \exp\left(xH'(1, \theta) + D\frac{x^2}{(\theta_0 - \theta)^3}\right) \quad (4.13)$$

306 for some  $D > 0$ .

## 8:10 Depth-First Search performance in random digraphs

307 For (4.11), let  $t' - t = kH'(1, \theta) - Bn^\alpha$  for some  $B > 0$  and  $0 < \alpha < 1 - \varepsilon$ . Thus  
 308  $k = \frac{1}{H'(1, \theta)}(t' - t - Bn^\alpha)$  which makes  $k$  of the order  $n^{1-\varepsilon}(\theta_0 - \theta) \gtrsim n^{1-2\varepsilon}$ . We use the  
 309 estimates (4.7). The quantity  $H(e^x, \theta)^k e^{-(t'-t)x}$  in (4.7) is smaller than  $\exp(\frac{Dkx^2}{(\theta_0 - \theta)^3} - Bxn^\alpha)$ ,  
 310 since the terms in  $kxH'(1, \theta)$  cancel. The minimal value of this quantity is reached for  
 311  $x = \frac{B}{2Dk}n^\alpha(\theta_0 - \theta)^3$  which is of order  $\frac{B}{2D(t'-t)}n^\alpha(\theta_0 - \theta)^3 H'(1, \theta) \lesssim n^{\alpha-1+\varepsilon}(\theta_0 - \theta)^2 \ll (\theta_0 - \theta)^2$ .  
 312 This minimal value is  $\exp(-\frac{B^2}{4kD}(\theta_0 - \theta)^3 H'(1, \theta)n^{2\alpha}) \leq \exp(-Cn^{2\alpha-1-\varepsilon})$  for some  $C > 0$ .

313 The proof of (4.12) is essentially the same.  $\blacktriangleleft$

314 At this point it is tempting to take the optimal value for  $\alpha$  to be  $(1 - \varepsilon)/2$  with  $\varepsilon$  close  
 315 to 0. But in fact this would ignore that the quantities  $H'(1, \theta)$  and  $H'(1, \theta')$  may differ by  
 316 a factor bounded away from 1, which would introduce an error greater than the term  $Bn^\alpha$ .  
 317 This in fact reduce the possibilities for  $1 - \varepsilon$  and  $\alpha$  for a consistent deviation probability. The  
 318 values  $\varepsilon = 1/5$  and  $\alpha = 3/5$  seem to be the limiting values that we can achieve as we show  
 319 next.

320 **► Theorem 10.** For all  $\beta > \frac{4}{5}$  for all  $t \in \mathcal{U}_n$  we have  $d(t) = n \int_0^{t/n} \rho(\theta) d\theta + o_p(n^\beta)$ .

321 **Proof.** Assume first that  $t$  is an up-time, and that there is an increasing sequence of up-  
 322 times (not necessarily consecutive)  $t_0, t_1, \dots, t_\ell$  with  $t_0 = 0$  and  $t_\ell = t$  with the constraint  
 323 that for all  $i < \ell$ :  $t_{i+1} - t_i \leq An^{1-\varepsilon}$ . Since  $u(0, t) = \sum_i u(t_i, i+1)$ , it turns out that  
 324 the probability that there exists  $i$  such that  $u(t_i, t_{i+1}) > \frac{1}{H'(1, t_i/n)}(t_{i+1} - t_i + Bn^\alpha)$  is  
 325 smaller than  $n \exp(-Cn^{2\alpha-1-\varepsilon})$  as shown in Lemma 9. Therefore with high probability  
 326  $\sum_i u(t_i, t_{i+1}) \leq \sum_i \frac{t_{i+1} - t_i}{H'(1, t_i/n)} + O(n^{\alpha+\varepsilon})$ , provided that  $2\alpha - 1 - \varepsilon > 0$ .

327 Similarly with high probability  $\sum_i u(t_i, t_{i+1}) \geq \sum_i \frac{t_{i+1} - t_i}{H'(1, t_{i+1}/n)} - O(n^{\alpha+\varepsilon})$ . Since both

$$328 \quad \sum_i \frac{t_{i+1} - t_i}{H'(1, t_i/n)} = n \int_0^{t/n} \frac{d\theta}{H'(1, \theta)} + O(n^{1-\varepsilon}) \quad (4.14)$$

$$329 \quad \sum_i \frac{t_{i+1} - t_i}{H'(1, t_{i+1}/n)} = n \int_0^{t/n} \frac{d\theta}{H'(1, \theta)} - O(n^{1-\varepsilon}) \quad (4.15)$$

330 it follows that with high probability  $u(0, t) = n \int_0^{t/n} \frac{d\theta}{H'(1, \theta)} + O(n^{1-\varepsilon}) + O(n^{\alpha+\varepsilon})$ , provided that  
 331  $2\alpha - 1 - \varepsilon > 0$ . The smallest error order is when  $1 - \varepsilon = \alpha + \varepsilon$ , thus when  $1 - \varepsilon = \alpha + \varepsilon > 4/5$ .

332 When  $t$  is not an up-time, the same result holds since the extra depth explored after the  
 333 last up-time is  $O_p(n^{2\varepsilon})$ , well within the error term  $O(n^\beta)$ .  $\blacktriangleleft$

### 334 4.3 Average interval between two down times

335 During the descent phase of the giant tree (*i.e.* the values of  $t$  in  $\mathcal{D}_n$ ), the DFS revisits  
 336 all ladder vertices met during the ascending phase. The process consists of exploring the  
 337 remaining outgoing neighbours of the ladder vertices that have not been visited during the  
 338 ascending phase.

339 We notice that given the ladder times, the remaining outdegrees of the ladder vertices  
 340 form a sequence of independent random variable.

341 Let  $t$  the time at which a ladder vertex is again revisited. Let  $t'$  be the first time at  
 342 which it was visited during the ascending period, thus  $d(t') = d(t)$ . Let  $\theta = t'/n \in (\theta_0, \theta_1)$  and  
 343  $u(\theta) = t'/n$ . We have the following theorem.

344 **► Theorem 11.** The p.g.f. of the remaining degree is asymptotically

$$345 \quad \tilde{\eta}_\theta(z) = (1 - u(\theta)) \frac{G_\eta(z) - 1 + \rho(u(\theta))}{z - 1 + (1 - u(\theta))\rho(u(\theta))}. \quad (4.16)$$

346 **Proof.** Let  $\tilde{\eta}_{\theta,k}(z)$  be the generating function of the remaining degree of the vertex  $v_t$   
 347 assuming the latter has outdegree  $k$ . Then, asymptotically,

$$348 \quad \tilde{\eta}_{\theta,k}(z) = \sum_{\ell < k} (1 - a(u(\theta)))^\ell a(u(\theta)) z^{k-1-\ell} \quad (4.17)$$

$$349 \quad = a(u(\theta)) \frac{z^k - (1 - a(u(\theta)))^k}{z - 1 + a(u(\theta))} \quad (4.18)$$

350 where  $a(\theta) = (1 - \theta)\rho(\theta)$  is the probability that a new vertex has not yet been visited and is  
 351 the root of an infinite BGW tree.

352 Summing over all outdegrees weighted by their probabilities we get the unconditional  
 353 generating function of the remaining outdegree:  $a(u(\theta)) \frac{G_\eta(z) - G_\eta(1 - a(u(\theta)))}{z - 1 + a(u(\theta))}$  which is equal to  
 354  $a(u(\theta)) \frac{G_\eta(z) - 1 + \rho(u(\theta))}{z - 1 + a(u(\theta))}$  since (3.3) yields  $G_\eta(1 - a(u(\theta))) = G_{\eta_{u(\theta)}}(1 - \rho(u(\theta))) = 1 - \rho(u(\theta))$ .  
 355 The above expression has value  $\rho(u(\theta))$  at  $z = 1$  and therefore must be divided by this factor  
 356 in order to get the p.g.f.  $\tilde{\eta}_\theta(z)$ . ◀

357 The “down-times” in  $\mathcal{D}_n$  bear some symmetries with the up-times in  $\mathcal{U}_n$ , with time  
 358 reversed.

359 ▶ **Theorem 12.** *The p.g.f. of the number of visited vertices between two consecutive down-times*  
 360 *is asymptotically*

$$361 \quad \bar{H}_\theta(z) = (1 - u(\theta)) \frac{g_\theta(z) - 1 + \rho(u(\theta))}{(1 - \theta)(g_\theta(z) - 1) + (1 - u(\theta))\rho(u(\theta))} \quad (4.19)$$

362 and the average number of visited vertices is  $\bar{H}'_\theta(1) = \frac{1}{\rho(u(\theta))} \frac{\lambda - \frac{1}{1-u(\theta)}}{\frac{1}{1-\theta} - \lambda}$ .

363 **Proof.** We have  $\bar{H}_\theta(z) = \tilde{\eta}_\theta(\theta + (1 - \theta)g_\theta(z))$ , which simplifies using (3.2) and (4.1). ◀

364 Let  $\bar{\Delta}_i$  be the number of new visited nodes between two consecutive down-times  $t$  and  $t'$  such  
 365 that  $d(t') = i$ . To get the number  $d(t, t')$  of down-times between  $t$  and  $t'$  we use the renewal  
 366 property of the sequence of down-times. With a similar reasoning as in (4.6)–(4.7) we get

$$367 \quad \begin{cases} \mathbb{P}(\bar{\Delta}_i + \bar{\Delta}_{i-1} + \cdots + \bar{\Delta}_{i-k} > t' - t) & = \mathbb{P}(d(t, t') < k) \\ \mathbb{P}(\bar{\Delta}_i + \bar{\Delta}_{i-1} + \cdots + \bar{\Delta}_{i-k} \leq t' - t) & = \mathbb{P}(d(t, t') \geq k). \end{cases} \quad (4.20)$$

368 and

$$369 \quad \forall x > 0 : \begin{cases} \mathbb{P}(d(t, t') < k) & \leq (\bar{\Delta}_{t'/n}(e^x))^k e^{-(t'-t)x} \\ \mathbb{P}(d(t, t') \geq k) & \leq (\bar{\Delta}_{t/n}(e^{-x}))^k e^{(t'-t)x}. \end{cases} \quad (4.21)$$

370 We have a theorem whose proof is absolutely similar to the proof of Theorem 10:

371 ▶ **Theorem 13.** *For all  $\beta > \frac{4}{5}$  with high probability for all  $t \in \mathcal{D}_n$  we have  $d(t) = n \int_{t/n}^{\theta_1} \frac{d\theta}{\bar{H}'_\theta(1)} +$   
 372  $O(n^\beta)$ .*

373 For  $\theta \in [\theta_0, \theta_1]$ , define  $\hat{\ell}(\theta) = \int_\theta^{\theta_1} \frac{d\tau}{\bar{H}'_\tau(1)}$ . Since  $d(t) = d(t')$ , where asymptotically  $t/n =$   
 374  $\theta \in [\theta_0, \theta_1]$  and  $t'/n = u(\theta) \in [0, \theta_0]$ , we obtain from Theorems 10 and 13, recalling (3.4),

$$375 \quad \hat{\ell}(\theta) = \int_0^{u(\theta)} \rho(x) dx = \tilde{\ell}(u(\theta)). \quad (4.22)$$

376 ▶ **Theorem 14.** *We have the identity  $u(\theta) = \check{\theta}$  in (3.5).*

## 8:12 Depth-First Search performance in random digraphs

377 **Proof.** Theorem 12 yields

$$378 \quad \widehat{\ell}'(\theta) = -\frac{1}{\widehat{H}'_\theta(1)} = -\frac{\rho(u(\theta))}{\lambda - \frac{1}{1-u(\theta)}} \frac{1 - (1-\theta)\lambda}{1-\theta}. \quad (4.23)$$

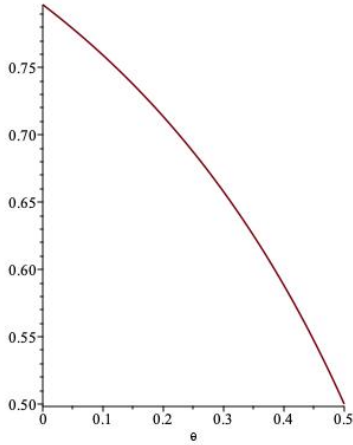
379 Furthermore  $\widetilde{\ell}'(\theta) = \rho(\theta)$  by (3.4). Hence, (4.22) implies

$$380 \quad \frac{du(\theta)}{d\theta} = \frac{\widehat{\ell}'(\theta)}{\widetilde{\ell}'(u(\theta))} = \frac{\frac{1}{1-\theta} - \lambda}{\frac{1}{1-u(\theta)} - \lambda}. \quad (4.24)$$

381 Similarly, the identity  $\widetilde{\ell}(\check{\theta}) = \widetilde{\ell}(\theta)$  in (3.5) yields, recalling (2.12),

$$382 \quad \frac{d\check{\theta}}{d\theta} = \frac{\check{\ell}'(\theta)}{\check{\ell}'(\check{\theta})} = \frac{\frac{1}{1-\theta} - \lambda}{\frac{1}{1-\check{\theta}} - \lambda}. \quad (4.25)$$

383 Since  $u(\theta_0) = \check{\theta}_0 = \theta_0$ , the differential equations (4.24) and (4.25) have the same solution.  
 384 We call the function  $\theta \mapsto \check{\theta}$  the mirror function; note that it only depends of average  
 385 outdegree without any further consideration on the details of the outdegree distribution. All  
 386 computations done we have  $\check{\theta} = 1 + \frac{1}{\lambda} W_{-1}(-(1-\theta)e^{(\theta-1)\lambda})$  using the branch  $W_{-1}(\cdot)$  of the  
 387 Lambert  $W$  function, see Figure 3 for  $\lambda = 2$ . ◀



■ **Figure 3** The mirror function for  $\lambda = 2$ .

388 ▶ **Remark 15.** Theorems 6 and 13 show that  $\widehat{\ell}(\theta) = \widetilde{\ell}^+(\theta)$  for  $\theta \in [\theta_0, \theta_1]$ . Furthermore, the  
 389 identity  $u(\theta) = \check{\theta}$  follows also directly from Theorem 6 and (3.5) together with  $d(t) = d(t')$ .

### 390 **5 An example**

391 We take the example of a constant outdegree. In the case of outdegree 2,  $G_\eta(z) = z^2$ ,  $\lambda = 2$ ,  
 392 and  $\theta_0 = \frac{1}{2}$ . Consider for simplicity only  $\theta \in [0, \theta_0]$ . Then, (3.3) has the explicit solution

393  $\rho(\theta) = 1 - \left(\frac{\theta}{1-\theta}\right)^2$ , and Theorems 2 and 6 show that if  $t/n \rightarrow \theta$ , then

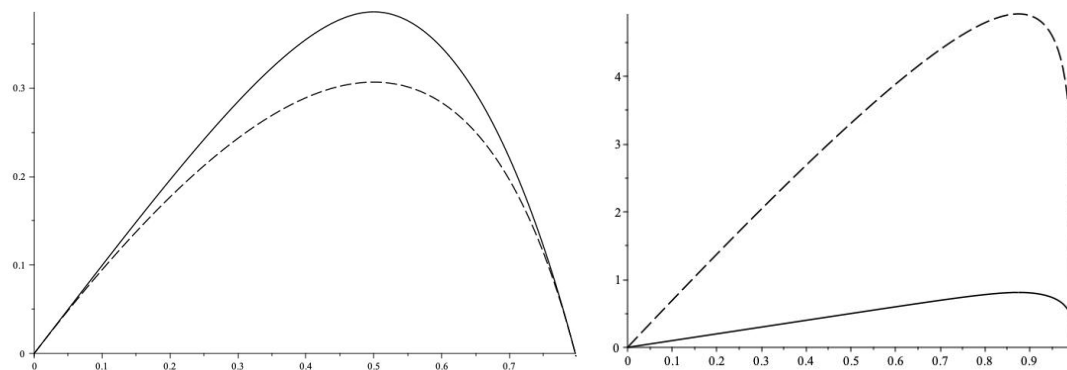
$$394 \quad \frac{1}{n} I(t) \xrightarrow{L^2} \widetilde{\ell}^+(\theta) = 2\theta + \log(1-\theta), \quad (5.1)$$

$$395 \quad \frac{1}{n} d(t) \xrightarrow{L^2} \check{\ell}^+(\theta) = -2\log(1-\theta) - \frac{\theta}{1-\theta}, \quad (5.2)$$

396 In the case of higher constant outdegree, say 8,  $\rho$  must be calculated from the implicit formula  
 397 (3.3), for  $\theta \leq \theta_0 = 1 - \frac{1}{8}$ . Larger outdegrees can be treated similarly. Figure 4 displays the  
 398 average depth and index for constant outdegree 2 and 8. We notice that for  $\eta = 2$ , the stack  
 399 size is smaller than the depth. This might be a surprise because at each new discovered node  
 400 the stack stores the whole set of outgoing arcs, while the depth increases by at most 1. The  
 401 reason is that in this case  $\eta = 2$ , and thus, although the stack keeps all unexplored arcs for  
 402 all ancestors, this is at most one arc for each ancestor and many ancestor have no arc left.  
 403 For  $\eta$  larger this disappears, and the stack size becomes larger than the depth.

## 404 6 Conclusion

405 We have presented an analysis of the Depth-First Search algorithm by Tarjan in a model of  
 406 random graphs recently introduced by Don Knuth. We have presented a version of the stack  
 407 model which can be analyzed as a Markov chain and is much easier to analyse than the real  
 408 depth. The latter requires new insights in a model of Bienaymé–Galton–Watson trees with a  
 409 varying extinction probability, in particular when close to the sub-critical case.



410 **Figure 4** The limit of  $\frac{1}{n}I(t)$  (dashed) and  $\frac{1}{n}d(t)$  (solid) as functions of  $\theta = t/n$  for  $\theta \in [0, \theta_0]$ , for  
 411 the example  $\eta = 2$  (left),  $\eta = 8$  (right).

## 410 References

### 411 References

- 412 1 Philippe Jacquet and Svante Janson. Depth-first search performance in a random  
 413 digraph with geometric outdegree distribution. *La Matematica*, 2024.
- 414 2 Lucja Kot Vignesh Ganapathy Alex Rasmussen Jayant Madhavan, David Ko and Alon  
 415 Halevy. Google’s deep web crawl. *Proceedings of the VLDB Endowment*, (2):1241–1252,  
 416 2208.
- 417 3 Donald E. Knuth. *The Art of Computer Programming*. Preliminary draft,  
 418 <http://cs.stanford.edu/~knuth/fasc12a.ps.gz>, 2022.
- 419 4 Michael Krivelevich and Benny Sudakov. The phase transition in random graphs: a  
 420 simple proof. *Random Structures Algorithms*, (2):131–138, 2013.
- 421 5 Gabriel Faraud Nathanaël Enriquez and Laurent Ménard. Limiting shape of the depth  
 422 first search tree in an erdős-rényi graph. *Random Structures Algorithms*, (2):501–516,  
 423 2020.

## 8:14 REFERENCES

- 424 **6** Laurent Ménard Nathan Noiry Nathanaël Enriquez, Gabriel Faraud. Depth first  
425 exploration of a configuration model. *Electron. J. Probab.*, (53), 2022.
- 426 **7** Jürgen Bennes Götz Kersting. A random walk approach to galton–watson trees. *J.*  
427 *Theoret. Probab.*, (3):777–803, 2000.
- 428 **8** Sahar Diskin Michael Krivelevich. On the performance of the depth first search  
429 algorithm in supercritical random graphs. *Electron. J. Combin.*, (3), 2022.
- 430 **9** Jean-François Le Gall Yves Le Jan. Branching processes in lévy processes: the  
431 exploration process. *Ann. Probab.*, (1):213–252, 1998.

### A

432 **Proof of Lemma 8.** We want first to determine the first fixed point  $g_\theta(1)$  of  $G_{\eta_\theta}$ . We have  
433 the expansion  
434

$$435 \quad G_\eta(y) = 1 + \lambda(y-1) + \frac{\lambda_2}{2}(y-1)^2 + O((y-1)^3) \quad (\text{A.1})$$

436 where  $\lambda_2$  is  $G''_\eta(1)$ . Thus

$$437 \quad G_{\eta_\theta}(y) = G_\eta(\theta + (1-\theta)y) = 1 + (1-\theta)\lambda(y-1) + (1-\theta)^2 \frac{\lambda_2}{2}(y-1)^2 + O((y-1)^3) \quad (\text{A.2})$$

438 We look at the fixed point equation  $G_{\eta_\theta}(y) - y = 0$  which is equivalent to

$$439 \quad \lambda(\theta_0 - \theta)(y-1) + \frac{\lambda_2(1-\theta)^2}{2}(y-1)^2 + O((y-1)^3) = 0 \quad (\text{A.3})$$

440 The natural root is 1 but the other root is  $1 - 2 \frac{\lambda(\theta_0 - \theta)}{\lambda_2(1-\theta)^2} + O((\theta_0 - \theta)^2)$  which yields the  
441 claimed result since  $\frac{1}{1-\theta} = \lambda + O(\theta_0 - \theta)$ .

442 To determine the function  $g_\theta(z)$  one must solve the equation  $zG_{\eta_\theta}(y) = y$  which writes:

$$443 \quad 1 - \frac{1}{z} + (1 + \lambda(\theta_0 - \theta) - \frac{1}{z})(y-1) + \frac{\lambda_2(1-\theta)^2}{2}(y-1)^2 + O((y-1)^3) = 0. \quad (\text{A.4})$$

444 Retaining only the second order (the extra order will be in  $O((\theta_0 - \theta)^3)$ ) we get an equation  
445 which is degenerate when  $(1 + \lambda(\theta_0 - \theta) - \frac{1}{z})^2 - 2\lambda_2(1-\theta)^2(1 - \frac{1}{z}) = 0$ , *i.e.* when  $1 - \frac{1}{z} =$   
446  $\frac{\lambda^2(\theta_0 - \theta)^2}{2\lambda_2(1-\theta)^2} + O((\theta_0 - \theta)^3)$ . In the above expression we can substitute  $(1 - \theta)$  with  $\frac{1}{\lambda}$  to a  
447  $O((\theta_0 - \theta)^3)$  error term.

448 Assuming  $|z - 1| \ll (\theta_0 - \theta)^2$  we have the expression of the root  $y = g_\theta(z)$ :

$$449 \quad y = 1 - 2 \frac{\lambda^3}{\lambda_2}(\theta_0 - \theta) + \frac{z-1}{\lambda(\theta_0 - \theta)} + O\left(\frac{(z-1)^2}{(\theta_0 - \theta)^3} + (\theta_0 - \theta)^2\right). \quad (\text{A.5})$$

450 Now the expression of  $H(z, \theta)$  satisfies

$$451 \quad H(z, \theta) = 1 + \frac{z-1}{1-g_\theta(z)} \quad (\text{A.6})$$

$$452 \quad = 1 + \frac{z-1}{2\frac{\lambda_2^3}{\lambda_2}(\theta_0-\theta) - \frac{z-1}{\lambda(\theta_0-\theta)} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^3} + (\theta_0-\theta)^2\right)} \quad (\text{A.7})$$

$$453 \quad = 1 + \frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)} \left( \frac{1}{1 - \frac{\lambda_2(z-1)}{2\lambda^4(\theta_0-\theta)^2} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^4} + (\theta_0-\theta)\right)} \right) \quad (\text{A.8})$$

$$454 \quad = 1 + \frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)} + \frac{\lambda_2^2(z-1)^2}{4\lambda^7(\theta_0-\theta)^3} + O\left(\frac{|z-1|^3}{(\theta_0-\theta)^5} + |z-1|\right) \quad (\text{A.9})$$

$$455 \quad = 1 + \frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^3} + |z-1|\right) \quad (\text{A.10})$$

456 In particular,

$$457 \quad H'(1, \theta) = \frac{\lambda_2}{2\lambda^3(\theta_0-\theta)} + O(1). \quad (\text{A.11})$$

458 Since the terms added after the unity are small compared to 1 we can equivalently state that

$$459 \quad H(z, \theta) = \exp\left(\frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^3} + |z-1|\right)\right). \quad (\text{A.12})$$

460 We also have by (4.2), still assuming  $|z-1| \ll (\theta_0-\theta)^2$ ,

$$461 \quad H(z, \theta) - 1 - (z-1)H'(1, \theta) = \frac{z-1}{1-g_\theta(z)} - \frac{z-1}{1-g_\theta(1)} \\ 462 \quad = (z-1) \frac{g_\theta(z) - g_\theta(1)}{(1-g_\theta(z))(1-g_\theta(1))} \quad (\text{A.13})$$

463 Furthermore, by taking the derivative of the fixed point equation  $zG_{\eta_\theta}(g_\theta(z)) = g_\theta(z)$ , we  
464 find, using (A.2) and (A.5),

$$465 \quad g'_\theta(z) = \frac{G_{\eta_\theta}(g_\theta(z))}{1 - zG'_{\eta_\theta}(g_\theta(z))} = \frac{g_\theta(z)/z}{-\lambda(\theta_0-\theta) - (1-\theta)^2\lambda_2(g_\theta(z)-1) + o(|\theta_0-\theta|^2)} \\ 466 \quad = \frac{1}{\lambda(\theta_0-\theta) + o(|\theta_0-\theta|^2)} = O((\theta_0-\theta)^{-1}). \quad (\text{A.14})$$

467 Thus (A.13) and the mean-value theorem together with (A.5) yield

$$468 \quad H(z, \theta) - 1 - (z-1)H'(1, \theta) = O(|z-1|^2/(\theta_0-\theta)^3). \quad (\text{A.15})$$

469 We can rewrite this as, for  $|x| \ll (\theta_0-\theta)^2$ , using the estimate (A.11),

$$470 \quad H(e^x, \theta) = \exp\left(xH'(1, \theta) + O\left(\frac{x^2}{(\theta_0-\theta)^3}\right)\right). \quad (\text{A.16})$$

471

◀