DEPTH-FIRST SEARCH PERFORMANCE IN RANDOM DIGRAPHS

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 an arbitrary distribution with finite variance. The results include asymptotics for the distribution of the stack index and depths of the search. The search yields a series of trees of finite size before and after the exploration of a giant tree. Our analysis mainly concerns the giant tree. Most results are first order. 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

This paper is a continuation of our earlier paper [5], which studies a special case using a simpler method. The motivation of both paper is a new section in Donald Knuth's *The Art of Computer Programming* [6], which is dedicated to Depth-First Search (DFS) in a digraph. The DFS is an important computing tool dedicated to the exploration of large unstructured dataset, mostly organised as large directed graphs, and the DFS is the fundation of the daily crawling process of the graph of the Web performed by Google [9]. 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 it 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 [6] for details as well as for historical notes. (See also S1–S2 in Section 2.) Note that the digraphs in [6] and here are multi-digraphs, where loops and multiple arcs are allowed. (Although in our random model they 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 stack index and the depth of vertices in the depth-first forest.

The random digraph model that we consider has n vertices and a given outdegree distribution $\mathbf{P}$. 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

---

Date: 13 March, 2024.

Supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council (VR) .

arcs. (Therefore, an arc can loop back to the starting vertex, and multiple arcs can occur.) We consider asymptotics as $n \to \infty$ for a fixed outdegree distribution. In [5], 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 1.1. Related results are proved by Enriquez, Faraud and Ménard [3] 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 Diskin and Krivelevich [2]. DFS in the random digraph $D(n, p)$ has also been considered previously, for example in the proof of [7, 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 $\mathrm{Po}(\lambda)$. Hence, the main result of [3], which shows convergence of the depth profile in the depth-first forest to a certain deterministic limit, is essentially the special case $\mathbf{P} = \mathrm{Po}(\lambda)$ of our result for the depths (Theorem 3.2). The proofs are quite different. See also Enriquez, Faraud, Ménard and Noiry [4], 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 [5], $d(t)$ 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 $o_{\mathrm{p}}(n)$ 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 that the method in Section 3, with deviation within $O_{\mathrm{p}}(n^{\beta})$ with β 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 η 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 η_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 λ. 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 $n \to \infty$.

For random variables X_n and positive numbers a_n, we write $X_n = o_p(a_n)$ if, as $n \to \infty$, $X_n/a_n \xrightarrow{p} 0$, *i.e.*, if for every $\varepsilon > 0$, we have $\mathbb{P}(|X_n| > \varepsilon a_n) \to 0$. Furthermore, $X_n = O_p(a_n)$ means that the family $\{X_n/a_n\}$ is bounded in probability, *i.e.*, if for every $\varepsilon > 0$ there exists C such that $\mathbb{P}(|X_n| > Ca_n) < \varepsilon$ for all n. $X_n = o_{L^2}(a_n)$ means $\mathbb{E}\big[|X_n/a_n|^2\big] \to 0$, and $X_n = O_{L^2}(a_n)$ means $\mathbb{E}\big[|X_n/a_n|^2\big] = O(1)$.

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

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

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

2. Stack index analysis

As said above, unlike the special case $\mathbf{P} = \mathrm{Ge}(1-p)$ studied in [5], in general the depth $d(t)$ does not follow a Markov chain. Therefore, we instead first consider the stack index $I(t)$ defined below, which does evolve as a Markov chain; many arguments for $I(t)$ below are similar to arguments for $d(t)$ in [5, Section 2].

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 $\eta(v)$ of v and add to the stack $\eta(v)$ new arcs from v, with unspecified endpoints. GOTO S1.

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

$$\Big(\frac{t}{n}\Big)^{k-1}\Big(1 - \frac{t}{n}\Big), \tag{2.1}$$

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

$$I(t+1) = \big(I(t) + \eta_t - 1 - \xi_t\big)^+, \qquad 1 \leqslant t < n, \tag{2.2}$$

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

We define

$$\zeta_t := \eta_t - \xi_t - 1. \tag{2.3}$$

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

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

Note that the random variables ζ_t are independent, but have different distributions. We define also

$$\widetilde{I}(t) := \sum_{i=1}^{t-1}(\eta_i - 1 - \xi_i) = \sum_{i=1}^{t-1}\zeta_i \tag{2.5}$$

and

$$I_*(t) := \min_{1\leqslant j\leqslant t} \widetilde{I}(j). \tag{2.6}$$

It follows from the recursion (2.2) and induction that

$$I(t) = \widetilde{I}(t) - I_*(t). \tag{2.7}$$

By (2.5) and (2.7),

$$\widetilde{I}(t+1) = \widetilde{I}(t) + \zeta_t = I_*(t) + I(t) + \zeta_t \tag{2.8}$$

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

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

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

We have

$$\mathbb{E}\,\zeta_t = \mathbb{E}\,\eta_t - \mathbb{E}\,\xi_t - 1 = \lambda - \frac{t/n}{1-t/n} - 1 = \lambda - \frac{1}{1-t/n}. \tag{2.10}$$

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

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

where

$$\widetilde{\iota}(\theta) := \int_0^\theta \Bigl(\lambda - \frac{1}{1-x}\Bigr)\,\mathrm{d}x = \lambda\theta + \log(1-\theta). \tag{2.12}$$

Let, as in [5],

$$\theta_0 := \bigl(1-\lambda^{-1}\bigr)^+ = \begin{cases} 1-\lambda^{-1}, & \lambda > 1,\\ 0, & \lambda \leqslant 1,\end{cases} \tag{2.13}$$

and let θ_1 is the largest root in $[0,1)$ of $\widetilde{\iota}(\theta_1) = 0$; thus

$$\log(1-\theta_1) = -\lambda\theta_1 \tag{2.14}$$

and θ_1 equals the survival probability of a Bienayme–Galton–Watson (BGW) process with $\mathrm{Po}(\lambda)$ offspring distribution. Define

$$\widetilde{\iota}^+(\theta) := [\widetilde{\iota}(\theta)]^+ = \begin{cases} \lambda\theta + \log(1-\theta), & 0 \leqslant \theta \leqslant \theta_1,\\ 0, & \theta_1 \leqslant \theta \leqslant 1,\end{cases} \tag{2.15}$$

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

We now can argue similarly to [5], using $I(t)$ instead of $d(t)$. The proof of [5, Theorem 2.4] applies with very minor differences, and yields:

Theorem 2.1. *Suppose that the outdegree distribution has finite variance. Then*

$$\max_{1\leqslant t\leqslant n}\bigl|I(t)-n\widetilde{\iota}^{+}(t/n)\bigr| = O_{L^2}(n^{1/2}). \tag{2.16}$$

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

$$\max_{1\leqslant t\leqslant (1-\varepsilon)n}\bigl|\widetilde{I}(t)-n\widetilde{\iota}(t/n)\bigr| = O_{L^2}(n^{1/2}). \tag{2.17}$$

We can use Theorem 2.1 to show the following results, extending [5, Theorems 4.1 and 4.3] to general outdegree distributions. The proofs are similar to the ones in [5]; details will be given in the full paper.

Theorem 2.2. *Suppose that the outdegree distribution has finite variance. Let N be the number of trees in the depth-first forest. Then*

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

where

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

Figure 1 shows the parameter ψ as a function of the average degree λ.

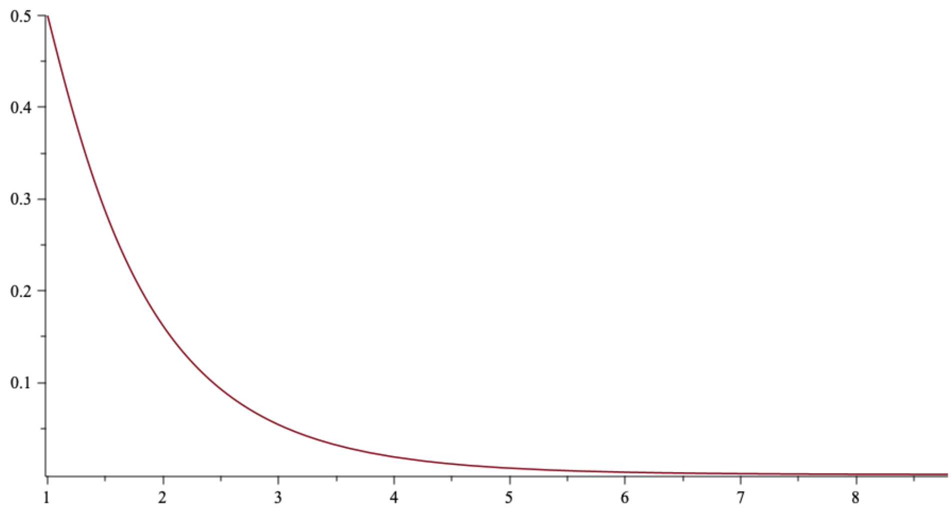

FIGURE 1. ψ, as function of λ.

Theorem 2.3. *Suppose that the outdegree distribution has finite variance. Let $\mathbf{T}_1$ be the largest tree in the depth-first forest.*

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

More precise results will be given in the full paper.

3. Stack size and depth

We can recover the depth of the vertices from the stack size process, or more precisely, from the process $\widetilde{I}(t)$ defined above; this uses a method that has been used by several authors in the study of other random trees, see for example [8, (1.1)] and [1, Proposition 4].

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

Lemma 3.1. (i) *For any $s, t \in [n]$, v_s is a descendant of v_t if and only if $s > t$ and*

$$\min_{t \leqslant i \leqslant s} \widetilde{I}(i) = \widetilde{I}(t). \tag{3.1}$$

(ii) *For any $t \geqslant 1$, the ancestors of v_t are the vertices v_s corresponding to the weak right-to-left minima of $(\widetilde{I}(s))_{1 \leqslant s \leqslant t}$, excluding $s = t$. Hence the depth $d(t)$ of v_t is the number of such right-to-left minima.*

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

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

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

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

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

Define

$$\widetilde{\ell}(\theta) := \int_0^\theta \rho(x)\,\mathrm{d}x, \qquad 0 \leqslant \theta \leqslant \theta_0, \tag{3.4}$$

and

$$\widetilde{\ell}^+(\theta) := \begin{cases} \widetilde{\ell}(\theta) = \int_0^\theta \rho(x)\,\mathrm{d}x, & 0 \leqslant \theta \leqslant \theta_0, \\ \widetilde{\ell}(\breve{\theta}), \text{ where } \breve{\theta} \in (0, \theta_0) \text{ and } \widetilde{\iota}(\breve{\theta}) = \widetilde{\iota}(\theta), & \theta_0 < \theta < \theta_1, \\ 0, & \theta_1 \leqslant \theta \leqslant 1. \end{cases} \tag{3.5}$$

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

Using Lemma 3.1 and Theorem 2.1 we can obtain the following result; the detailed argument (given in the full paper) is unfortunately rather long.

Theorem 3.2. *We have*

$$\max_{1\leqslant t\leqslant n}\left|d(t) - n\widetilde{\ell}^{+}(t/n)\right| = o_{L^2}(n). \tag{3.6}$$

4. Depth analysis

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

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

4.1. Splitting the giant tree.

Recall the definitions of θ_0 and θ_1 from (2.13)–(2.14). It is easily seen that these are, roughly, the proportions of discovered nodes when we find the top of the giant tree, and when we leave the giant tree, respectively. Let $\varepsilon \in (0,1)$ be a fixed number which will be specified later. We divide the interval $[0, n\theta_1]$ into three intervals: $\mathcal{U}_n = [0, \theta_0 n - n^{1-\varepsilon}]$ called the *climbing zone*, $\mathcal{D}_n = [\theta_0 n + n^{1-\varepsilon}, \theta_1 n]$, called the *descending zone*, and the band $\mathcal{B}_n = [\theta_0 n - n^{1-\varepsilon}, \theta_0 n + n^{1-\varepsilon}]$, called the *neutral zone*.

For $t \in \mathcal{U}_n$, we say that v_t is a *ladder vertex* if for all $t' \in \mathcal{U}_n$: $t' > t \Rightarrow d(t') \geqslant d(t)$. In other words, t is a right minimum of the function $d(t)$ in $\mathcal{U}$. For an integer i, we define $d_1^{-1}(i)$ 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 ladder vertex. Also, we say that t is an *up-time* if v_t is a ladder vertex and t is the first time it is visited.

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

4.2. Average interval between two up-times.

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

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

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

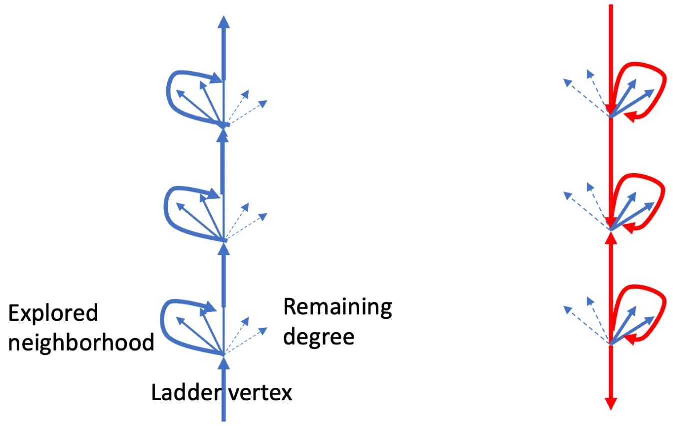

FIGURE 2. The DFS in the giant tree: the climbing phase (left) and the descending phase (right) through ladder vertices.

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

Theorem 4.1. *We have the identity*

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

and thus the expectated number of visited nodes in the BGW tree is

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

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

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

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

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

$$\begin{aligned} z\frac{\rho(\theta)}{1-g_\theta(z)}\left(G_{\eta_\theta}(1) - G_{\eta_\theta}(g_\theta(z))\right) &= \frac{\rho(\theta)}{1-g_\theta(z)}(z-g_\theta(z)) \\ &= \rho(\theta)\left(\frac{z-1}{1-g_\theta(z)}+1\right). \end{aligned} \tag{4.5}$$

To get the p.g.f. $H(z,\theta)$ of the number of discovered nodes conditioned on $\mathcal{T}_\theta$ having at least one infinite subtree, one must divide by $\rho(\theta)$, which is the probability that the BGW tree is infinite. The error term comes from the fact that the fraction of discovered vertices varies during the interval in the proportion of the size of the interval divided by n. If we look at the generating function $H(z,\theta)$ it turns out (see the appendix) that the interval multiplied by $(\theta_0-\theta)^2$ is bounded in probability, thus a variation of θ of order $O\big((\theta_0-\theta)^{-2}/n\big)$. □

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

$$\begin{cases} \mathbb{P}(\Delta_i+\Delta_{i+1}+\cdots+\Delta_{i+k}>t'-t) & = & \mathbb{P}(u(t,t')<k) \\ \mathbb{P}(\Delta_i+\Delta_{i+1}+\cdots+\Delta_{i+k}\leqslant t'-t) & = & \mathbb{P}(u(t,t')\geqslant k). \end{cases} \tag{4.6}$$

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

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

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

Lemma 4.2. *When the quantity θ is smaller than but close to θ_0, we have $\rho(\theta)=1-g_\theta(1)=2\frac{\lambda^3}{\lambda_2}(\theta_0-\theta)+O((\theta_0-\theta)^2)$ where $\lambda_2=G''_\eta(1)$. Furthermore the generating function $g_\theta(z)$ 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 $|z-1|\ll(\theta_0-\theta)^2$ we have*

$$g_\theta(z)=1-2\frac{\lambda^3}{\lambda_2}(\theta_0-\theta)+\frac{z-1}{\lambda(\theta_0-\theta)}+O\Big(\frac{(z-1)^2}{(\theta_0-\theta)^3}+(\theta_0-\theta)^2\Big). \tag{4.8}$$

$$H(z,\theta)=1+\frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)}+O\Big(\frac{(z-1)^2}{(\theta_0-\theta)^3}+(\theta_0-\theta)^2\Big) \tag{4.9}$$

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

Proof. In Appendix A. □

Lemma 4.3. *Let $\varepsilon>0$ and $\frac{1-\varepsilon}{2}<\alpha<1-\epsilon$. For $a,A>0$ there exist $B>0$ and $C>0$ such that for $t<t'\in\mathcal{U}_n$ with $an^{1-\varepsilon}\leqslant t'-t\leqslant An^{1-\varepsilon}$ we have*

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

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

Proof. Now t and t' can be within $n^{1-\varepsilon}$ from $n\theta_0$ which has an impact on the estimate of $\log H(e^x,\theta)$. For the super-critical BGW tree, we know from Lemma 4.2 and its proof that $H'(1,\theta)=O((\theta_0-\theta)^{-1})$ and that, for $|x|\ll(\theta_0-\theta)^2$,

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

for some $D>0$.

For (4.11), let $t'-t=kH'(1,\theta)-Bn^\alpha$ for some $B>0$ and $0<\alpha<1-\varepsilon$. Thus $k=\frac{1}{H'(1,\theta)}(t'-t-Bn^\alpha)$ which makes k of the order $n^{1-\varepsilon}(\theta_0-\theta)\gtrsim n^{1-2\varepsilon}$.

We use the estimates (4.7). The quantity $H(e^x,\theta)^k e^{-(t'-t)x}$ in (4.7) is smaller than $\exp(\frac{Dkx^2}{(\theta_0-\theta)^3}-Bxn^\alpha)$, since the terms in $kxH'(1,\theta)$ cancel. The minimal value of this quantity is reached for $x=\frac{B}{2Dk}n^\alpha(\theta_0-\theta)^3$ which is of order $\frac{B}{2D(t'-t)}n^\alpha(\theta_0-\theta)^3H'(1,\theta)\lesssim n^{\alpha-1+\varepsilon}(\theta_0-\theta)^2 \ll (\theta_0-\theta)^2$. This minimal value is $\exp(-\frac{B^2}{4kD}(\theta_0-\theta)^3H'(1,\theta)n^{2\alpha})\leqslant \exp(-Cn^{2\alpha-1-\varepsilon})$ for some $C>0$.

The proof of (4.12) is essentially the same. □

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

Theorem 4.4. *For all $\beta>\frac{4}{5}$ for all $t\in\mathcal{U}_n$ we have $d(t)=n\int_0^{t/n}\rho(\theta)d\theta+o_{\mathrm{p}}(n^\beta)$.*

Proof. Assume first that t is an up-time, and that there is an increasing sequence of up-times (not necessarily consecutive) $t_0,t_1,\dots,t_\ell$ with $t_0=0$ and $t_\ell=t$ with the constraint that for all $i<\ell$: $t_{i+1}-t_i\leqslant An^{1-\varepsilon}$. Since $u(0,t)=\sum_i u(t_i,i+1)$, it turns out that 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 smaller than $n\exp(-Cn^{2\alpha-1-\varepsilon})$ as shown in Lemma 4.3. Therefore with high probability $\sum_i u(t_i,t_{i+1})\leqslant\sum_i\frac{t_{i+1}-t_i}{H'(1,t_i/n)}+O(n^{\alpha+\varepsilon})$, provided that $2\alpha-1-\varepsilon>0$.

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

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

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

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 $2\alpha-1-\varepsilon>0$. The smallest error order is when $1-\varepsilon=\alpha+\varepsilon$, thus when $1-\varepsilon=\alpha+\varepsilon>4/5$.

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

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

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

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

Theorem 4.5. *The p.g.f. of the remaining degree is asymptotically*

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

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

$$\begin{aligned}\tilde{\eta}_{\theta,k}(z) &= \sum_{\ell<k}(1-a(u(\theta))^\ell\, a(u(\theta))z^{k-1-\ell} && (4.17)\\ &= a(u(\theta))\frac{z^k-(1-a(u(\theta)))^k}{z-1+a(u(\theta))} && (4.18)\end{aligned}$$

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

Summing over all outdegrees weighted by their probabilities we get the unconditional 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 $a(u(\theta))\frac{G_\eta(z)-1+\rho(u(\theta))}{z-1+a(u(\theta))}$ since (3.3) yields $G_\eta(1-a(u(\theta))) = G_{\eta_{u(\theta)}}\big(1-\rho(u(\theta))\big) = 1-\rho(u(\theta))$. The above expression has value $\rho(u(\theta))$ at $z=1$ and therefore must be divided by this factor in order to get the p.g.f. $\tilde{\eta}_\theta(z)$. □

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

Theorem 4.6. *The p.g.f. of the number of visited vertices between two consecutive down-times is asymptotically*

$$\overline{H}_\theta(z) = (1-u(\theta))\frac{\frac{g_\theta(z)}{z}-1+\rho\,(u(\theta))}{(1-\theta)(g_\theta(z)-1)+(1-u(\theta))\rho\,(u(\theta))} \tag{4.19}$$

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

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

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

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

and

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

We have a theorem whose proof is absolutely similar to the proof of Theorem 4.4:

Theorem 4.7. *For all $\beta > \frac{4}{5}$ with high probability for all $t \in \mathcal{D}_n$ we have $d(t) = n\int_{t/n}^{\theta_1}\frac{\mathrm{d}\theta}{\overline{H}'_\theta(1)} + O(n^\beta)$.*

For $\theta \in [\theta_0, \theta_1]$, define $\widehat{\ell}(\theta) = \int_\theta^{\theta_1} \frac{\mathrm{d}\tau}{\overline{H}'_\theta(1)}$. Since $d(t) = d(t')$, where asymptotically $t/n = \theta \in [\theta_0, \theta_1]$ and $t'/n = u(\theta) \in [0, \theta_0]$, we obtain from Theorems 4.4 and 4.7, recalling (3.4),

$$\widehat{\ell}(\theta) = \int_0^{u(\theta)} \rho(x)\,\mathrm{d}x = \widetilde{\ell}(u(\theta)). \tag{4.22}$$

Theorem 4.8. *We have the identity* $u(\theta) = \breve{\theta}$ *in* (3.5).

Proof. Theorem 4.6 yields

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

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

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

Similarly, the identity $\widetilde{\iota}(\breve{\theta}) = \widetilde{\iota}(\theta)$ in (3.5) yields, recalling (2.12),

$$\frac{\mathrm{d}\breve{\theta}}{\mathrm{d}\theta} = \frac{\widetilde{\iota}'(\theta)}{\widetilde{\iota}'(\breve{\theta})} = \frac{\frac{1}{1-\theta} - \lambda}{\frac{1}{1-\breve{\theta}} - \lambda}. \tag{4.25}$$

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

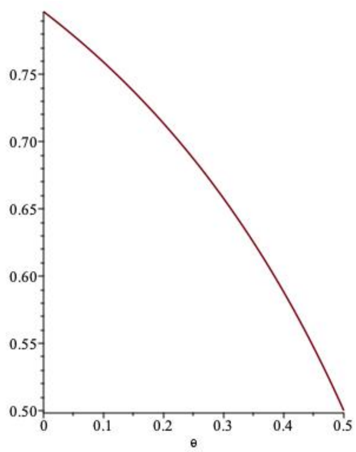

FIGURE 3. The mirror function for $\lambda = 2$.

Remark 4.9. Theorems 3.2 and 4.7 show that $\widehat{\ell}(\theta) = \widetilde{\ell}^+(\theta)$ for $\theta \in [\theta_0, \theta_1]$. Furthermore, the identity $u(\theta) = \breve{\theta}$ follows also directly from Theorem 3.2 and (3.5) together with $d(t) = d(t')$.

5. An example

We take the example of a constant outdegree. In the case of outdegree 2, $G_\eta(z) = z^2$, $\lambda = 2$, and $\theta_0 = \frac{1}{2}$. Consider for simplicity only $\theta \in [0, \theta_0]$. Then, (3.3) has the explicit solution $\rho(\theta) = 1 - \left(\frac{\theta}{1-\theta}\right)^2$, and Theorems 2.1 and 3.2 show that if $t/n \to \theta$, then

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

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

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

6. Conclusion

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

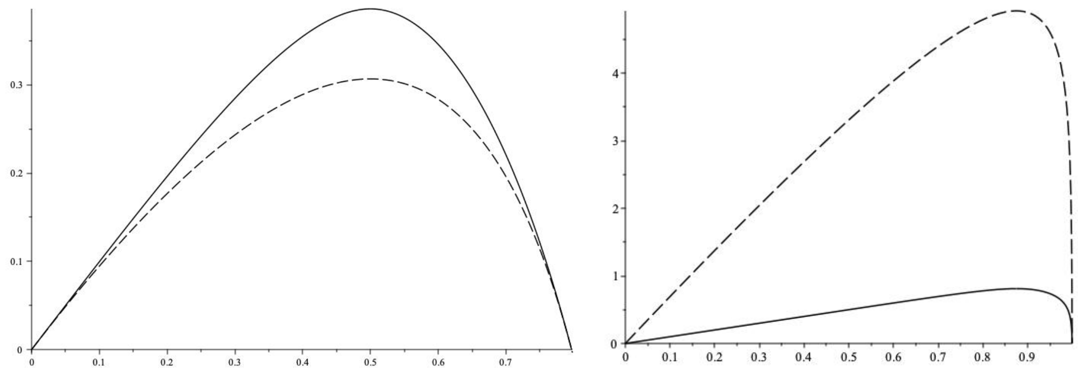

Figure 4. The limit of $\frac{1}{n}I(t)$ (dashed) and $\frac{1}{n}d(t)$ (solid) as functions of $\theta = t/n$ for $\theta \in [0, \theta_0]$, for the example $\eta = 2$ (left), $\eta = 8$ (right).

References

[1] Jürgen Bennies & Götz Kersting. A random walk approach to Galton–Watson trees. *J. Theoret. Probab.* **13** (2000), no. 3, 777–803.

[2] Sahar Diskin & Michael Krivelevich. On the performance of the Depth First Search algorithm in supercritical random graphs. *Electron. J. Combin.* **29** (2022), no. 3, Paper No. 3.64, 7 pp.
[3] Nathanaël Enriquez, Gabriel Faraud & Laurent Ménard. Limiting shape of the depth first search tree in an Erdős-Rényi graph. *Random Structures Algorithms* **56** (2020), no. 2, 501–516.
[4] Nathanaël Enriquez, Gabriel Faraud, Laurent Ménard & Nathan Noiry. Depth first exploration of a configuration model. *Electron. J. Probab.* **27** (2022), Paper No. 53, 27 pp.
[5] Philippe Jacquet & Svante Janson. Depth-First Search performance in a random digraph with geometric outdegree distribution. *La Matematica* (2024).
Extended abstract: Depth-First Search performance in a random digraph with geometric degree distribution. *33rd International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2022)*, Leibniz International Proceedings in Informatics (LIPIcs) **225** (2022), pp. 11:1-11:15.
[6] 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`
[7] Michael Krivelevich & Benny Sudakov. The phase transition in random graphs: a simple proof. *Random Structures Algorithms* **43** (2013), no. 2, 131–138.
[8] Jean-François Le Gall & Yves Le Jan. Branching processes in Lévy processes: the exploration process. *Ann. Probab.* **26** (1998), no. 1, 213–252.
[9] Jayant Madhavan, David Ko, Łucja Kot, Vignesh Ganapathy, Alex Rasmussen & Alon Halevy. Google's Deep Web crawl. *Proceedings of the VLDB Endowment* **1**, no. 2 (2008), 1241–1252.

Appendix A

Proof of Lemma 4.2. We want first to determine the first fixed point $g_\theta(1)$ of G_{η_θ}. We have the expansion

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

where λ_2 is $G''_\eta(1)$. Thus

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

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

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

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 claimed result since $\frac{1}{1-\theta} = \lambda + O(\theta_0 - \theta)$.

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

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

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

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

$$y=1-2\frac{\lambda^3}{\lambda_2}(\theta_0-\theta)+\frac{z-1}{\lambda(\theta_0-\theta)}+O\Big(\frac{(z-1)^2}{(\theta_0-\theta)^3}+(\theta_0-\theta)^2\Big). \tag{A.5}$$

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

$$\begin{aligned}
H(z,\theta) &= 1+\frac{z-1}{1-g_\theta(z)} && \text{(A.6)}\\
&= 1+\frac{z-1}{2\frac{\lambda^3}{\lambda_2}(\theta_0-\theta)-\frac{z-1}{\lambda(\theta_0-\theta)}+O\big(\frac{(z-1)^2}{(\theta_0-\theta)^3}+(\theta_0-\theta)^2\big)} && \text{(A.7)}\\
&= 1+\frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)}\left(\frac{1}{1-\frac{\lambda_2(z-1)}{2\lambda^4(\theta_0-\theta)^2}+O\big(\frac{(z-1)^2}{(\theta_0-\theta)^4}+(\theta_0-\theta)\big)}\right) && \text{(A.8)}\\
&= 1+\frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)}+\frac{\lambda_2^2(z-1)^2}{4\lambda^7(\theta_0-\theta)^3}+O\Big(\frac{|z-1|^3}{(\theta_0-\theta)^5}+|z-1|\Big) && \text{(A.9)}\\
&= 1+\frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)}+O\Big(\frac{(z-1)^2}{(\theta_0-\theta)^3}+|z-1|\Big) && \text{(A.10)}
\end{aligned}$$

In particular,

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

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

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

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

$$\begin{aligned}
H(z,\theta)-1-(z-1)H'(1,\theta) &= \frac{z-1}{1-g_\theta(z)}-\frac{z-1}{1-g_\theta(1)}\\
&= (z-1)\frac{g_\theta(z)-g_\theta(1)}{\big(1-g_\theta(z)\big)\big(1-g_\theta(1)\big)} && \text{(A.13)}
\end{aligned}$$

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

$$\begin{aligned}
g_\theta'(z) &= \frac{G_{\eta\theta}(g_\theta(z))}{1-zG_{\eta\theta}'(g_\theta(z))}=\frac{g_\theta(z)/z}{-\lambda(\theta_0-\theta)-(1-\theta)^2\lambda_2(g_\theta(z)-1)+o(|\theta_0-\theta|^2)}\\
&= \frac{1}{\lambda(\theta_0-\theta)+o(|\theta_0-\theta|^2)}=O\big((\theta_0-\theta)^{-1}\big). && \text{(A.14)}
\end{aligned}$$

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

$$H(z,\theta)-1-(z-1)H'(1,\theta)=O\big(|z-1|^2/(\theta_o-\theta)^3\big). \tag{A.15}$$

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

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

□

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/