

THE NUMBER OF DESCENDANTS IN A RANDOM DIRECTED ACYCLIC GRAPH

SVANTE JANSON

ABSTRACT. We consider a well known model of random directed acyclic graphs of order n , obtained by recursively adding vertices, where each new vertex has a fixed outdegree $d \geq 2$ and the endpoints of the d edges from it are chosen uniformly at random among previously existing vertices.

Our main results concern the number $X^{(n)}$ of vertices that are descendants of n . We show that $X^{(n)}/n^{(d-1)/d}$ converges in distribution; the limit distribution is, up to a constant factor, given by the d th root of a Gamma distributed variable. $\Gamma(d/(d-1))$. When $d = 2$, the limit distribution can also be described as a chi distribution $\chi(4)$. We also show convergence of moments, and find thus the asymptotics of the mean and higher moments.

1. INTRODUCTION

A *dag* is a directed acyclic (multi)graph, and a *d-dag* is a dag where one or several vertices are *roots* with outdegree 0, and all other vertices have outdegrees d . (Here, d is a positive integer; we assume below $d \geq 2$.)

We consider, as many before us, the random *d-dag* D_n on n vertices constructed recursively by starting with a single root 1, and then adding vertices $2, 3, \dots, n$ one by one, giving each new vertex, k say, d outgoing edges with endpoints uniformly and independently chosen at random among the already existing vertices $\{1, \dots, k-1\}$. (We thus allow multiple edges, so D_n is a directed multigraph.) Two minor variations that will be discussed in Section 10 are that we may start with any number $m \geq 1$ of roots, and that we may select the d parents of a new node without replacement, thus not allowing multiple edges. (In the latter case, we have to start with $\geq d$ roots.)

Note that for $d = 1$, the model becomes the well known *random recursive tree*; the properties in this case are quite different from the case $d \geq 2$, and we assume throughout the paper $d \geq 2$. In fact, to concentrate on the essential features, in the bulk of the paper we consider the most important case $d = 2$; the minor differences in the case $d > 2$ are briefly treated in Section 8.

The random *d-dag* has been studied as a model for a random circuit where each gate has d inputs chosen at random [10; 19; 1; 18; 6; 16]. (In this case it seems more natural to reverse all edges, and regard a *d-dag* as a graph with indegrees 0 or d . In the present paper, we direct the edges towards the root(s) as above.) The model has also been studied in connection with constraint satisfaction [13, Exercise 7.2.2.3–371]. Among results shown earlier for random *d-dags*, we mention results on vertex degrees and leaves [9; 18; 15; 16; 14], and on lengths of paths and depth [10; 19; 1; 8; 6].

In the present paper, we study the following problem, as far as we know first considered by Knuth [13, Exercises 7.2.2.3–371 and 372]: How many descendants does

Date: 27 February, 2023.

Supported by the Knut and Alice Wallenberg Foundation.

vertex n have? In other words, how many vertices can be reached by a directed path from vertex n ? In the random circuit interpretation, this is the number of gates (and inputs) that are used in the calculation of an output.

We state our main results in the next subsection, and prove them in Sections 2–8. Along the way, we prove some results on the structure of the subgraph of descendants which may be of independent interest. Some further results are given in Section 9. As said above, we discuss two variations of the model in Section 10.

Remark 1.1. We emphasise that we in this paper exclusively consider random dags constructed by *uniform* attachment. Another popular model that has been studied by many authors (often as an undirected graph) is preferential attachment, see e.g. [3] and [5]. A different model of non-uniform attachment is studied in [6]. $\triangle$

Problem 1.2. Find results for preferential attachment random dags corresponding to the results above!

Do the same for the model in [6]!

1.1. Main result. We introduce some notation; for further (mainly standard) notation, see Section 1.2. We let $d \geq 2$ be fixed and consider asymptotics as $n \rightarrow \infty$.

Let D_n be the random d -dag defined above, let $\hat{D}_n$ be the subdigraph of D_n consisting of all vertices and edges that can be reached by a directed path from vertex n (including vertex n itself), and let $X^{(n)} := |\hat{D}_n|$, the number of descendants of n . We thus want to find the asymptotic behaviour of the random variable $X^{(n)}$ and its expectation $\mathbb{E} X^{(n)}$ as $n \rightarrow \infty$. Note that $\hat{D}_n$ also is a d -dag, and has 1 root; thus the number of edges in $\hat{D}_n$ is $d(X^{(n)} - 1)$, and hence our results also yield the asymptotics of the number of edges.

Our main result in the case $d = 2$ is the following theorem, proved in two parts in Sections 6 and 7.

Let χ_4 denote a random variable with the $\chi(4)$ distribution. Recall that this means that χ_4 has the distribution of $|\eta|$ where η is a standard normal random vector in $\mathbb{R}^4$, and that thus (or by (1.7) and a change of variables) χ_4 has density function

$$f_{\chi_4}(x) = \frac{1}{2}x^3e^{-x^2/2}, \quad x > 0. \quad (1.1)$$

Theorem 1.3. *Let $d = 2$. Then, as $n \rightarrow \infty$,*

$$X^{(n)}/\sqrt{n} \xrightarrow{d} \frac{\pi}{2\sqrt{2}}\chi_4 \quad (1.2)$$

with convergence of all moments. Hence, for every fixed $r > 0$,

$$\mathbb{E}(X^{(n)})^r \sim \left(\frac{\pi}{2}\right)^r \Gamma\left(\frac{r}{2} + 2\right)n^{r/2} \quad (1.3)$$

and, in particular,

$$\mathbb{E} X^{(n)} \sim \frac{3\pi^{3/2}}{8}\sqrt{n}. \quad (1.4)$$

More generally, for any fixed $d \geq 2$, we prove in Section 8 the following:

Theorem 1.4. *Let $d \geq 2$. Then, as $n \rightarrow \infty$,*

$$X^{(n)}/n^{(d-1)/d} \xrightarrow{d} \frac{\pi(d-1)^{1/d}}{d \sin(\pi/d)}\gamma^{1/d}, \quad (1.5)$$

with convergence of all moments, where $\gamma \in \Gamma(\frac{d}{d-1})$. Hence, for every fixed $r > 0$,

$$\mathbb{E}(X^{(n)})^r \sim \left(\frac{(d-1)^{1/d} \pi}{d \sin(\pi/d)} \right)^r \frac{\Gamma(\frac{d}{d-1} + \frac{r}{d})}{\Gamma(\frac{d}{d-1})} n^{r(d-1)/d} \quad (1.6)$$

We note that the convergence in (1.2) and (1.5) does *not* hold a.s.; see Remark 9.6.

We will see in Section 10 that the same results hold for the variations with $m \geq 1$ roots (as long as m is fixed or does not grow too fast) and without multiple edges (i.e., drawing without replacement).

Example 1.5. Knuth [13, Answer 7.2.2.3–371(b)] considers the version with $d = 2$, $m \geq 2$ roots, and drawing without replacement (i.e., no multiple edges); for this version he provides recursion formulas that yield the exact value of $\mathbb{E} X^{(n)}$ (there denoted $C_{m,n}$). For example, for $m = 2$ and $n = 100$, his formulas yield $\mathbb{E} X^{(n)} \doteq 20.79$ while the asymptotic value (1.4) is $\doteq 20.88$, with an error of less than 0.5%. $\triangle$

1.2. Notation. The random d -dag D_n , its subdigraph $\hat{D}_n$, and the number $X^{(n)}$ of descendants of n are defined above. The outdegree d is fixed and not shown in the notation. As said above we usually assume $d = 2$; in particular this is the case in the proof in Sections 2–7, while we consider general $d \geq 2$ in Section 8.

We say that the vertices and edges of $\hat{D}_n$ are *red*. Thus $X^{(n)} := |\hat{D}_n|$ is the number of red vertices in D_n . (For any digraph D , we let $|D|$ denote its number of vertices.)

Essentially all random variables below depend on n . We may denote the dependency on n by a superscript $^{(n)}$ for clarity (in particular in limit statements), but we often omit this. We sometimes in the proofs tacitly assume that n is large enough. Unspecified limits are as $n \rightarrow \infty$.

We will in the proofs consider three different phases of the dag D_n , see Sections 3–5. We will then use fixed integers $n_1 = n_1^{(n)}$ and $n_2 = n_2^{(n)}$; these can be chosen rather arbitrarily with $n_1/n \rightarrow 0$ slowly and $n_2/\sqrt{n} \rightarrow \infty$ slowly, see the beginnings of Sections 3 and 4.

We use $\xrightarrow{p}, \xrightarrow{d}, \xrightarrow{L^1}$, for convergence in probability, distribution and L^1 , respectively, and $\stackrel{d}{=}$ for equality in distribution.

As usual, a.s. (almost surely) means with probability 1, while w.h.p. (with high probability) means with probability tending to 1 as $n \rightarrow \infty$.

We recall some classical probability distributions. The Gamma distribution $\Gamma(s, a)$, where $s > 0$ and $a > 0$, has density

$$\Gamma(s)^{-1} a^{-s} x^{s-1} e^{-x/a}, \quad x > 0. \quad (1.7)$$

We write $\Gamma(s) = \Gamma(s, 1)$. (There should be no risk of confusion with the Gamma function.) In particular, $\Gamma(1) = \text{Exp}(1)$, the standard exponential distribution. If $\gamma \in \Gamma(s, a)$, then (1.7) implies

$$\mathbb{E} \gamma^r = a^r \frac{\Gamma(s+r)}{\Gamma(s)}, \quad r \geq 0. \quad (1.8)$$

The chi-square distribution $\chi^2(r) = \Gamma(r/2, 2)$, and the chi-distribution $\chi(r)$ is the distribution of $\sqrt{\xi}$ where $\xi \in \chi^2(r)$. (This shows that when $d = 2$, the limits in (1.2) and (1.5) agree.)

We use ‘increasing’ and ‘positive’ in the weak sense.

Remark 1.6. For simplicity, and to avoid unnecessary distraction, we often state results with convergence in probability, also when the proof yields the stronger convergence in L^1 . (For example, this applies to all three results in Section 4.) Actually, in many (all?) cases, convergence in probability can be improved to convergence in L^p for any $p < \infty$, as a consequence of the estimates in Section 7. $\triangle$

Remark 1.7. The construction of the random dag D_n naturally constructs D_n for all $n \geq 1$ together. In other words, it yields a coupling of D_n for all $n \geq 1$. However, in the proofs below we will *not* use this coupling; instead we regard D_n as constructed separately for each n , which allows us to use a different coupling in the proof. $\triangle$

2. BASIC ANALYSIS

For simplicity, we assume $d = 2$ from now on until the proof of Theorem 1.3 is completed at the end of Section 7. The modifications for general d are discussed in Section 8.

2.1. A stochastic recursion. We consider in the sequel only the red subgraph $\hat{D}_n$ of D_n , which we recall consists of the descendants of n and all edges between them.

In the definition in Section 1 of the dag D_n , we start with vertex 1 and add vertices in increasing order. In our analysis, we will instead start at vertex n and go backwards to 1. The red dag $\hat{D}_n$ then may be generated by the following procedure.

- (1) Start by declaring vertex n to be *red*, and all others *black*. Let $k := n$.
- (2) If vertex k is red, then create two new edges from that vertex, with endpoints that are randomly drawn from $1, \dots, k-1$, and declare these endpoints red. If k is black, delete k (and do nothing else).
- (3) If $k = 2$ then STOP; otherwise let $k := k-1$ and REPEAT from (2).

Let Y_k be the number of edges in $\hat{D}_n$ that start in $\{k+1, \dots, n\}$ and end in $\{1, \dots, k\}$. In other words, Y_k is the number of edges that cross the gap between $k+1$ and k . Furthermore, let Z_k be the number of these edges that end in k . We here consider integers k with $0 \leq k \leq n-1$, and have the boundary conditions $Y_{n-1} = 2$ and $Y_0 = 0$; also $Z_1 = Y_1$ and $Z_0 = 0$.

Let also, for $1 \leq k \leq n-1$,

$$J_k := \mathbf{1}\{Z_k \geq 1\}, \quad (2.1)$$

the indicator that at least one edge ends at k , which equals the indicator that k is red, and thus can be reached from n .

We will study the random dag $\hat{D}_n$ by travelling from vertex n backwards to the root; we thus consider the sequence $Y_{n-1}, \dots, Y_1, Y_0$ in reverse order. In the procedure above, there are Z_k edges that end at k , and $2J_k$ edges that start there; hence, for $2 \leq k \leq n-1$,

$$Y_{k-1} = Y_k - Z_k + 2J_k = Y_k - Z_k + 2 \cdot \mathbf{1}\{Z_k \geq 1\}. \quad (2.2)$$

In our analysis, we modify the procedure above by not revealing the endpoint of the edges until needed. This means that when coming to a vertex $k \in \{1, \dots, n-1\}$, we have a list of Y_k edges where we know only the start but not the end (except that the end should be in $\{1, \dots, k\}$). We then randomly select a subset by throwing a coin with success probability $1/k$ for each of the Y_k edges; these edges end at k and are removed from the list, and thus Z_k is the number of them. This determines also J_k by (2.1), and if $J_k = 1$, we add two new edges starting at k to our list. It is evident that this

gives the same distribution of random edges as the original algorithm above. (It is here important that the two edges from a given vertex are chosen with replacement, so that we can treat the Y_k edges passing over the gap between $k+1$ and k as independent. Note that the endpoints of these edges are uniformly distributed on $\{1, \dots, k\}$.)

It follows from the modified procedure that $Y_{n-1}, \dots, Y_1$ is a Markov chain. More precisely, let $\mathcal{F}_k$ be the σ -field generated by our coin tosses at vertices $n-1, \dots, k+1$, and note that these coin tosses determine Y_k (and also $Y_{n-1}, \dots, Y_{k+1}$). Then, for $1 \leq k \leq n-1$, conditioned on $\mathcal{F}_k$, Z_k has a binomial distribution

$$Z_k \in \text{Bin}(Y_k, 1/k). \quad (2.3)$$

Thus (2.2) and (2.3) give a stochastic recursion of Markov type for Y_k .

Note that $\mathcal{F}_k \subset \mathcal{F}_{k-1}$, so $\mathcal{F}_1, \dots, \mathcal{F}_{n-1}$ form a *decreasing* sequence of σ -fields, i.e., a reverse filtration. We therefore may change sign of the indices and consider, for example, Y_{-j} and $\mathcal{F}_{-j}$ for $j \in \{-(n-1), \dots, -1\}$ so that we have a filtration of the standard type.

The recursion (2.2)–(2.3) yields, for $2 \leq k \leq n-1$,

$$\begin{aligned} \mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) &= Y_k - \mathbb{E}(Z_k \mid \mathcal{F}_k) + 2\mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k) \\ &= Y_k - \frac{1}{k}Y_k + 2\left(1 - \left(1 - \frac{1}{k}\right)^{Y_k}\right). \end{aligned} \quad (2.4)$$

We obtain also, by Markov's inequality,

$$\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) \leq Y_k - \mathbb{E}(Z_k \mid \mathcal{F}_k) + 2\mathbb{E}(Z_k \mid \mathcal{F}_k) = Y_k + \frac{1}{k}Y_k = \frac{k+1}{k}Y_k. \quad (2.5)$$

2.2. A reverse supermartingale and some estimates. We define, for $0 \leq k \leq n-1$,

$$W_k := (k+1)Y_k, \quad (2.6)$$

and find from (2.5)

$$\mathbb{E}(W_{k-1} \mid \mathcal{F}_k) = k\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) \leq (k+1)Y_k = W_k. \quad (2.7)$$

This shows that W_{-j} , $-(n-1) \leq j \leq 0$, is a supermartingale for the filtration $(\mathcal{F}_{-j})$; in other words, $W_0, \dots, W_{n-1}$ is a reverse supermartingale. We have the initial value

$$W_{n-1} = nY_{n-1} = 2n. \quad (2.8)$$

We thus have the Doob decomposition

$$W_k = M_k - A_k, \quad 0 \leq k \leq n-1, \quad (2.9)$$

where

$$M_k := 2n + \sum_{i=k+1}^{n-1} (W_{i-1} - \mathbb{E}(W_{i-1} \mid \mathcal{F}_i)) \quad (2.10)$$

is a reverse martingale: $\mathbb{E}(M_{k-1} \mid \mathcal{F}_k) = M_k$, and

$$A_k := \sum_{i=k+1}^{n-1} (W_i - \mathbb{E}(W_{i-1} \mid \mathcal{F}_i)) \quad (2.11)$$

is positive and reverse increasing: (2.7) yields

$$0 = A_{n-1} \leq \dots \leq A_1 \leq A_0. \quad (2.12)$$

In particular, $W_k \leq M_k$ and

$$\mathbb{E}W_k \leq \mathbb{E}M_k = M_{n-1} = 2n, \quad 0 \leq k \leq n-1. \quad (2.13)$$

We note also from (2.4) the exact formula

$$\mathbb{E}(W_{k-1} \mid \mathcal{F}_k) = k \mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) = (k-1)Y_k + 2k(1 - (1 - \frac{1}{k})^{Y_k}) \quad (2.14)$$

and thus

$$\begin{aligned} A_{k-1} - A_k &= W_k - \mathbb{E}(W_{k-1} \mid \mathcal{F}_k) = 2Y_k - 2k(1 - (1 - \frac{1}{k})^{Y_k}) \\ &= 2k[(1 - \frac{1}{k})^{Y_k} - 1 + \frac{Y_k}{k}]. \end{aligned} \quad (2.15)$$

Furthermore, (2.2) and (2.3) also yield (rather crudely, but we do not need the exact formula, nor optimal constants), for $1 \leq k \leq n-1$,

$$\begin{aligned} \text{Var}(Y_{k-1} \mid \mathcal{F}_k) &= \text{Var}(Z_k - 2 \cdot \mathbf{1}\{Z_k \geq 1\} \mid \mathcal{F}_k) \\ &\leq 2 \text{Var}(Z_k \mid \mathcal{F}_k) + 2 \text{Var}(2 \cdot \mathbf{1}\{Z_k \geq 1\} \mid \mathcal{F}_k) \\ &= 2Y_k \frac{1}{k} (1 - \frac{1}{k}) + 8 \mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k) (1 - \mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k)) \\ &\leq 2Y_k \frac{1}{k} + 8 \mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k) \leq 2Y_k \frac{1}{k} + 8 \mathbb{E}(Z_k \mid \mathcal{F}_k) \\ &\leq \frac{10}{k} Y_k \end{aligned} \quad (2.16)$$

and thus

$$\text{Var}(W_{k-1} \mid \mathcal{F}_k) = k^2 \text{Var}(Y_{k-1} \mid \mathcal{F}_k) \leq 10kY_k \leq 10W_k. \quad (2.17)$$

Hence, (2.10) yields, using the (reverse) martingale property, (2.17), and (2.13), for $0 \leq k \leq n-1$,

$$\begin{aligned} \text{Var } M_k &= \mathbb{E}(M_k - 2n)^2 = \sum_{i=k+1}^{n-1} \mathbb{E} \text{Var}(W_{i-1} \mid \mathcal{F}_i) \leq 10 \sum_{i=k+1}^{n-1} \mathbb{E} W_i \\ &\leq 10(n-1-k) \cdot 2n \leq 20n^2. \end{aligned} \quad (2.18)$$

Consequently, since $W_k \leq M_k$ by (2.9),

$$\mathbb{E} W_k^2 \leq \mathbb{E} M_k^2 = \text{Var } M_k + (\mathbb{E} M_k)^2 \leq 20n^2 + (2n)^2 = 24n^2. \quad (2.19)$$

We extend this to a maximal inequality.

Lemma 2.1. *We have*

$$\mathbb{E} \max_{n-1 \geq k \geq 0} W_k^2 \leq \mathbb{E} \max_{n-1 \geq k \geq 0} M_k^2 \leq 96n^2. \quad (2.20)$$

Proof. By Doob's inequality [11, Theorem 10.9.4] for the reverse martingale M_k and (2.19),

$$\mathbb{E} \max_{n-1 \geq k \geq 0} M_k^2 \leq 4 \mathbb{E} M_0^2 \leq 96n^2. \quad (2.21)$$

The result follows, recalling again $W_k \leq M_k$. $\square$

We show some further estimates used later.

Lemma 2.2. *For $1 \leq k \leq n-1$,*

$$\mathbb{P}(Z_k \geq 1) \leq \frac{2n}{k^2}, \quad (2.22)$$

$$\mathbb{P}(Z_k \geq 2) \leq \frac{24n^2}{k^4}. \quad (2.23)$$

Proof. By Markov's inequality, (2.3), and (2.6), we have

$$\mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k) \leq \mathbb{E}(Z_k \mid \mathcal{F}_k) = \frac{Y_k}{k} \leq \frac{W_k}{k^2} \quad (2.24)$$

and

$$\mathbb{P}(Z_k \geq 2 \mid \mathcal{F}_k) \leq \mathbb{E}\left(\binom{Z_k}{2} \mid \mathcal{F}_k\right) = \binom{Y_k}{2} \frac{1}{k^2} \leq \frac{Y_k^2}{k^2} \leq \frac{W_k^2}{k^4}. \quad (2.25)$$

The results (2.22)–(2.23) follow by taking expectations, using (2.13) and (2.19). $\square$

Lemma 2.3. *For $1 \leq k \leq n-1$,*

$$A_{k-1} - A_k \leq \frac{W_k^2}{k^3}, \quad (2.26)$$

$$\mathbb{E} A_k \leq 12 \frac{n^2}{k^2}. \quad (2.27)$$

Proof. By (2.15) and Taylor's formula (or the Bonferroni inequalities),

$$A_{k-1} - A_k = 2k \left(\left(1 - \frac{1}{k}\right)^{Y_k} - 1 + \frac{Y_k}{k} \right) \leq 2k \binom{Y_k}{2} \frac{1}{k^2} \leq \frac{Y_k^2}{k} \leq \frac{W_k^2}{k^3}, \quad (2.28)$$

which is (2.26). As a consequence,

$$A_k \leq \sum_{i=k+1}^{n-1} \frac{W_i^2}{i^3} \quad (2.29)$$

and thus, by (2.19),

$$\mathbb{E} A_k \leq \sum_{i=k+1}^{n-1} \frac{\mathbb{E} W_i^2}{i^3} \leq 24n^2 \sum_{i=k+1}^{\infty} \frac{1}{i^3} \leq 12 \frac{n^2}{k^2}. \quad (2.30)$$

$\square$

3. PHASE I: A YULE PROCESS

In this section we consider the first part of the evolution of the red dag $\hat{D}_n$, and consider the variables $Y_{n-1}, \dots, Y_{n_1}$, where (for definiteness) we let $n_1 := n_1^{(n)} := \lfloor n/\log n \rfloor$. (We might choose $n_1 = n_1^{(n)}$ as any (deterministic) sequence of integers such that $n_1/n \rightarrow 0$ slowly; in particular, any such sequence with $n_1 \geq n/\log n$ will also do. We leave it to the reader to see precisely how small n_1 can be.) We will show that the variables $Y_{n-1}, \dots, Y_{n_1}$ can be approximated (as $n \rightarrow \infty$) by a time-changed *Yule process*.

Recall that the Yule process is a continuous-time branching process, where each particle lives a lifetime that has an exponential $\text{Exp}(1)$ distribution, and then the particle splits into two new particles. (All lifetimes are independent.) Let $\mathcal{Y}_t$ be the number of particles at time t . The standard version, which we denote by $\mathcal{Y}'_t$, starts with one particle at time 0, but we start with $\mathcal{Y}_0 = 2$; thus the process $\mathcal{Y}_t$ can be seen as the sum of two independent copies of the standard Yule process $\mathcal{Y}'_t$.

It is well known, see e.g. [2, Section III.5], that for the standard Yule process, the number of particles at time t has the geometric distribution $\text{Ge}(e^{-t})$ with mean e^t and

$$\mathbb{P}(|\mathcal{Y}'_t| = k) = e^{-t}(1 - e^{-t})^{k-1}, \quad k \geq 1. \quad (3.1)$$

Moreover, $\mathcal{Y}'_t/e^t \xrightarrow{\text{a.s.}} \hat{\xi}$ as $t \rightarrow \infty$, where (e.g. as a consequence of (3.1)) $\hat{\xi} \in \text{Exp}(1)$. Hence, $\mathcal{Y}_t$ has a shifted negative binomial distribution $\text{NegBin}(2, e^{-t}) + 2$ with

$$\mathbb{P}(|\mathcal{Y}_t| = k) = (k-1)e^{-2t}(1-e^{-t})^{k-2}, \quad k \geq 2. \quad (3.2)$$

In particular, for all $t \geq 0$ we have

$$\mathbb{E} \mathcal{Y}_t = 2 \mathbb{E} \mathcal{Y}'_t = 2e^t, \quad (3.3)$$

and, as $t \rightarrow \infty$,

$$e^{-t} \mathcal{Y}_t \xrightarrow{\text{a.s.}} \xi := \hat{\xi}_1 + \hat{\xi}_2 \in \Gamma(2), \quad (3.4)$$

with $\hat{\xi}_1, \hat{\xi}_2 \in \text{Exp}(1)$ independent, so that their sum has a Gamma distribution.

We may also regard the Yule process $\mathcal{Y}$ as an infinite tree (the *Yule tree*), with one vertex $\gamma_0 := 0$ (the root), and one vertex γ_i at each time a particle splits (a.s. these times are distinct, and we may number them in increasing order); each particle is then represented by an edge from its time of birth to its time of death. Note that $\mathcal{Y}_t$, the number of living particles, equals the number of edges alive at time t , and that the number of particles that have died before (or at) t is $\mathcal{Y}_t - 2$.

We now change time by the mapping $t \mapsto e^{-t}$; thus the vertices in the Yule tree are mapped to the points $e^{-\gamma_i} \in (0, 1]$. The root is now at 1, and edges go from a larger label to a smaller. If a particle is born at one of these times $x = e^{-\gamma_i}$, and its lifetime in the original Yule process is $\tau \in \text{Exp}(1)$, then it lives there from γ_i to $\gamma_i + \tau$, and after the time change it is represented by an edge from $x = e^{-\gamma_i}$ to $e^{-(\gamma_i + \tau)} = xe^{-\tau} = xU$, where $U := e^{-\tau} \in \text{U}(0, 1)$ has a uniform distribution. Going backwards in time, we thus begin with two particles (edges) starting at 1. Each edge starting at a point x has endpoints xU'_x and xU''_x , where $U'_x, U''_x \in \text{U}(0, 1)$, and all these uniform random variables are independent. As before, we start two new edges at each endpoint. We let $\hat{\mathcal{Y}}$ denote this (infinite) random tree with vertices in $(0, 1]$, and let $\hat{\mathcal{Y}}_x$ be the number of particles (edges) alive at time x .

We may now compare the time-changed Yule tree to the red dag $\hat{D}_n$ constructed above, scaled to $[0, 1]$. An edge from a vertex k ends at a vertex uniformly distributed on $\{1, \dots, k-1\}$, which we may construct as $\lfloor (k-1)U \rfloor + 1$, where $U \in \text{U}(0, 1)$. We thus start with one point at n , and add again two edges from it and from the endpoint of every edge (except at 1), where now an edge started at $j+1$ goes to $\lfloor jU \rfloor + 1$ with $U \in \text{U}(0, 1)$. However, if two or more edges have the same endpoint, we still only start two new edges there.

A point in $\hat{D}_n$ that is m generations away from the root, thus has label

$$X = \lfloor \dots \lfloor (n-1)U_{\nu_1} \rfloor \dots U_{\nu_m} \rfloor + 1, \quad (3.5)$$

for the some $U_{\nu_1}, \dots, U_{\nu_m} \in [0, 1]$ (from the construction of the edges), and then

$$nU_{\nu_1} \dots U_{\nu_m} + 1 \geq X \geq nU_{\nu_1} \dots U_{\nu_m} - m. \quad (3.6)$$

Let $\hat{D}'_n$ denote the random red dag $\hat{D}_n$ with all labels divided by n ; thus the vertices are now points in $(0, 1]$. We then see that $\hat{D}'_n$ coincides with the time-changed Yule tree up to small errors. More precisely, we couple the two by first constructing the Yule tree $\mathcal{Y}$, and its time-changed version $\hat{\mathcal{Y}}$, and then making a perturbation of $\hat{\mathcal{Y}}$ by replacing each label $U_{\nu_1} \dots U_{\nu_m}$ by X/n with X as in (3.5). This gives a dag that coincides (in distribution) with $\hat{D}'_n$ until the first time that two edges in $\hat{D}'_n$ have the same endpoint.

Theorem 3.1. *We may w.h.p. couple the random dag $\hat{D}'_n$ and the time-changed Yule tree $\hat{\mathcal{Y}}$, such that considering only vertices with labels in $[n_1/n, 1]$, and edges with the starting point in this set, there is a bijection between these sets of vertices in the two models which displaces each label by at most $\log^2 n/n$, and a corresponding bijection between the edges (preserving the incidence relations).*

Proof. We have $\hat{\mathcal{Y}}_x = \mathcal{Y}_{-\log x}$ for every $x \in (0, 1]$, and thus by (3.3)

$$\mathbb{E} \hat{\mathcal{Y}}_x = \mathbb{E} \mathcal{Y}_{-\log x} = 2e^{-\log x} = 2/x. \quad (3.7)$$

The number of vertices with labels in $[x, 1]$ is $\hat{\mathcal{Y}}_x - 1$, and taking $x = n_1/n \sim 1/\log n$, we thus have $O_p(\log n)$ vertices; in particular w.h.p. less than $\log^2 n$ vertices. Consequently, w.h.p., the number of generations from the root to any point in $[n_1/n, 1]$ is at most $\log^2 n$, and then the bound (3.6) shows that all vertex displacements are at most $(\log n)^2/n$.

Furthermore, it follows from (3.7) that the expected number of vertices in $\hat{\mathcal{Y}}$ that are within $(\log n)^2/n$ from n_1/n is

$$\begin{aligned} \mathbb{E} (\hat{\mathcal{Y}}_{n_1/n - (\log n)^2/n} - \hat{\mathcal{Y}}_{n_1/n + (\log n)^2/n}) &= \frac{2}{n_1/n - (\log n)^2/n} - \frac{2}{n_1/n + (\log n)^2/n} \\ &\sim \frac{4(\log n)^2/n}{(n_1/n)^2} = O\left(\frac{\log^4 n}{n}\right) = o(1), \end{aligned} \quad (3.8)$$

and thus w.h.p. no vertex is pushed across the boundary n_1/n by the displacements in the coupling.

Finally, it follows from Lemma 2.2 that the probability that two edges in the dag $\hat{D}_n$ have the same endpoint k for some $k \geq n_1$ is at most

$$\sum_{k=n_1}^{n-1} \mathbb{P}(Z_k \geq 2) \leq 24n^2 \sum_{k=n_1}^{\infty} k^{-4} = O(n^2/n_1^3) = o(1). \quad (3.9)$$

Consequently, w.h.p. the coupling above between $\hat{\mathcal{Y}}$ and $\hat{D}_n$ yields a bijection for vertices in $[n_1/n, n]$ and their edges. $\square$

We define a random variable that will play an important role later: let

$$\Xi = \Xi^{(n)} := \frac{W_{n_1}}{n}. \quad (3.10)$$

Lemma 3.2. *As $n \rightarrow \infty$,*

$$\Xi^{(n)} = \frac{W_{n_1}^{(n)}}{n} \xrightarrow{d} \xi \in \Gamma(2). \quad (3.11)$$

Proof. We use the coupling in Theorem 3.1 for each n , recalling Remark 1.7. Then, w.h.p.,

$$\hat{\mathcal{Y}}_{n_1/n + (\log n)^2/n} \leq Y_{n_1}^{(n)} \leq \hat{\mathcal{Y}}_{n_1/n - (\log n)^2/n} \quad (3.12)$$

and thus

$$|Y_{n_1}^{(n)} - \hat{\mathcal{Y}}_{n_1/n}| \leq \hat{\mathcal{Y}}_{n_1/n - (\log n)^2/n} - \hat{\mathcal{Y}}_{n_1/n + (\log n)^2/n}. \quad (3.13)$$

In particular, (3.8) implies

$$|Y_{n_1}^{(n)} - \hat{\mathcal{Y}}_{n_1/n}| \xrightarrow{P} 0. \quad (3.14)$$

Moreover, (3.4) implies

$$x\hat{\mathcal{Y}}_x = x\mathcal{Y}_{-\log x} \xrightarrow{\text{a.s.}} \xi \quad \text{as } x \rightarrow 0, \quad (3.15)$$

with $\xi \in \Gamma(2)$. Consequently, by (3.14) and (3.15),

$$\frac{n_1}{n} Y_{n_1}^{(n)} = \frac{n_1}{n} (Y_{n_1}^{(n)} - \hat{\mathcal{Y}}_{n_1/n}) + \frac{n_1}{n} \hat{\mathcal{Y}}_{n_1/n} \xrightarrow{\text{p}} \xi. \quad (3.16)$$

Hence, recalling (2.6),

$$\frac{W_{n_1}^{(n)}}{n} = \frac{n_1 + 1}{n_1} \cdot \frac{n_1}{n} Y_{n_1}^{(n)} \xrightarrow{\text{p}} \xi. \quad (3.17)$$

The convergence in probability in (3.16)–(3.17) depends on the coupling used above, but it follows that convergence in distribution holds also without it, which completes the proof. $\square$

4. PHASE II: A BORING FLAT PART

Let $n_2 = n_2^{(n)}$ be any sequence of integers with $\sqrt{n} \ll n_2 \leq n_1$. We will show that in the range $n_1 \geq k \geq n_2$, the variable W_k essentially does not change, so it is equal to a random constant. We begin with two lemmas valid for larger ranges.

Lemma 4.1. *As $n \rightarrow \infty$,*

$$\max_{n-1 \geq k \geq n_2} \left| \frac{A_k}{n} \right| = \frac{A_{n_2}}{n} \xrightarrow{\text{p}} 0. \quad (4.1)$$

Proof. By Lemma 2.3,

$$\mathbb{E} \frac{A_{n_2}}{n} \leq 12 \frac{n}{n_2^2} = o(1), \quad (4.2)$$

which shows (4.1), recalling (2.12). $\square$

Lemma 4.2. *As $n \rightarrow \infty$,*

$$\max_{n_1 \geq k \geq 0} \left| \frac{M_k}{n} - \Xi^{(n)} \right| \xrightarrow{\text{p}} 0. \quad (4.3)$$

Proof. By Doob's inequality for the reverse martingale M_k and using (2.10), (2.17) and (2.13) as in (2.18) (cf. the proof of Lemma 2.1),

$$\begin{aligned} \mathbb{E} \max_{n_1 \geq k \geq 0} |M_k - M_{n_1}|^2 &\leq 4 \mathbb{E} |M_0 - M_{n_1}|^2 = 4 \sum_{i=1}^{n_1} \mathbb{E} \text{Var}(W_{i-1} | \mathcal{F}_i) \\ &\leq 40 \sum_{i=1}^{n_1} \mathbb{E} W_i \leq 80nn_1 = o(n^2). \end{aligned} \quad (4.4)$$

We have, using (3.10) and $W_{n_1} = M_{n_1} - A_{n_1}$,

$$\max_{n_1 \geq k \geq 0} \left| \frac{M_k}{n} - \Xi \right| \leq \max_{n_1 \geq k \geq 0} \left| \frac{M_k}{n} - \frac{M_{n_1}}{n} \right| + \left| \frac{A_{n_1}}{n} \right| \xrightarrow{\text{p}} 0, \quad (4.5)$$

where the convergence follows by (4.4) and Lemma 4.1. $\square$

Theorem 4.3. *As $n \rightarrow \infty$,*

$$\max_{n_1 \geq k \geq n_2} \left| \frac{W_k}{n} - \Xi^{(n)} \right| \xrightarrow{\text{p}} 0. \quad (4.6)$$

Proof. We have, for any k ,

$$\left| \frac{W_k}{n} - \Xi \right| \leq \left| \frac{M_k}{n} - \Xi \right| + \left| \frac{A_k}{n} \right| \quad (4.7)$$

and thus the result follows from Lemmas 4.1 and 4.2. $\square$

5. PHASE III: DETERMINISTIC DECAY FROM A RANDOM LEVEL

We extend the processes W_k , M_k and A_k to real arguments $t \in [0, n-1]$ by linear interpolation. Since the extended version A_t is piecewise linear, it is differentiable everywhere except at integer points, where we (arbitrarily) take the left derivative.

Lemma 5.1. *Let $\delta > 0$. Then*

$$\mathbb{E} \left[\sup_{\delta\sqrt{n} \leq t \leq n-1} \left| \frac{d}{dt} A_t \right| \right] \leq \frac{96}{\delta^3} n^{1/2}. \quad (5.1)$$

Proof. Let $k := [t]$, so $k-1 < t \leq k$. Then, by (2.26),

$$0 \leq -\frac{d}{dt} A_t = A_{k-1} - A_k \leq \frac{W_k^2}{k^3} \leq \frac{W_k^2}{t^3}. \quad (5.2)$$

The result (5.1) follows by Lemma 2.1. $\square$

We rescale and define

$$\hat{A}_t^{(n)} := n^{-1} A_{t\sqrt{n}}^{(n)}, \quad t \geq 0. \quad (5.3)$$

Recall also that $C[a, b]$ is the (Banach) space of continuous functions on $[a, b]$.

Lemma 5.2. *Let $0 < \delta < b < \infty$. Then the stochastic processes $\hat{A}_t^{(n)}$, $n \geq 1$, are tight in $C[\delta, b]$.*

Proof. We have, temporarily writing $A(t) := A_t^{(n)}$,

$$\frac{d}{dt} \hat{A}_t^{(n)} = n^{-1/2} A'(t\sqrt{n}). \quad (5.4)$$

Hence, Lemma 5.1 yields

$$\mathbb{E} \left[\sup_{\delta \leq t \leq b} \left| \frac{d}{dt} \hat{A}_t^{(n)} \right| \right] = \mathbb{E} \left[n^{-1/2} \sup_{\delta\sqrt{n} \leq t \leq b\sqrt{n}} \left| \frac{d}{dt} A_t \right| \right] \leq \frac{96}{\delta^3}, \quad (5.5)$$

and thus the supremum in the left-hand side forms a tight family of random variables as n varies.

Moreover, for a fixed $t \in [\delta, b]$ we have by Lemma 2.3

$$\mathbb{E} \hat{A}_t^{(n)} = n^{-1} \mathbb{E} A_{t\sqrt{n}}^{(n)} \leq n^{-1} \cdot 12 \frac{n^2}{[t\sqrt{n}]^2} = 12(1 + o(1))t^{-2} = O(1), \quad (5.6)$$

and thus also the family $\hat{A}_t^{(n)}$ is tight. The result follows, see [4, Theorem 8.2]. $\square$

Theorem 5.3. *We have*

$$\sup_{0 \leq t \leq b} \left| n^{-1} W_{t\sqrt{n}}^{(n)} - t^2 \log(1 + \Xi^{(n)}/t^2) \right| \xrightarrow{\mathbb{P}} 0, \quad (5.7)$$

for every fixed $b > 0$.

Remark 5.4. We may note that (5.7) means convergence, in probability, in the space $C[0, \infty)$ with its standard topology (uniform convergence on compact sets). Equivalently, we may consider the step functions $n^{-1}W_{[t\sqrt{n}]}^{(n)}$ and convergence in $D[0, \infty)$. $\triangle$

Proof. We divide the proof into several steps.

Step 1: A subsequence. By Lemma 5.2 and Prohorov's theorem [4, Theorem 6.1], for every compact interval $[\delta, b] \subset (0, \infty)$ we can find a subsequence (n_ν) such that, along the subsequence,

$$\hat{A}_t^{(n)} \xrightarrow{d} \mathcal{A}_{\delta,b}(t) \quad \text{in } C[\delta, b] \quad (5.8)$$

for some continuous random function $\mathcal{A}_{\delta,b}(t)$ on $[\delta, b]$. Furthermore, it suffices to consider a countable set of such intervals, for example $\mathcal{I} := \{[m^{-1}, m], m \geq 2\}$, and by considering convergence in the product space $\prod_{[\delta,b] \in \mathcal{I}} C[\delta, b]$ we can find a subsequence such that (5.8) holds jointly for all compact intervals $[\delta, b] \in \mathcal{I}$; by adding a factor $\mathbb{R}^2$, we may also assume that this holds jointly with (3.11) and (4.3). We consider until the last step of the proof only this subsequence.

Step 2: A coupling. By the Skorohod coupling theorem [12, Theorem 4.30], we may couple D_n for different n such that the convergence in (5.8) holds a.s. for every $[\delta, b] \in \mathcal{I}$, and also (3.11) and (4.3) hold a.s. Since convergence in $C[\delta, b]$ means uniform convergence, this means that a.s. $\hat{A}_t^{(n)} \rightarrow \mathcal{A}_{\delta,b}(t)$ uniformly on $[\delta, b]$ for each $[\delta, b] \in \mathcal{I}$. It is evident that a.s. the different limits $\mathcal{A}_{\delta,b}(t)$ have to agree whenever intervals overlap, and thus there exists a continuous random function $\mathcal{A}(t)$ defined on $(0, \infty)$ such that a.s.

$$\hat{A}_t^{(n)} \rightarrow \mathcal{A}(t) \quad (5.9)$$

uniformly on each compact interval $[\delta, b] \subset (0, \infty)$. (In other words, $\hat{A}_t^{(n)} \xrightarrow{\text{a.s.}} \mathcal{A}(t)$ in the space $C(0, \infty)$.) Clearly, $\mathcal{A}(t) \geq 0$. Furthermore, we now a.s. have

$$\Xi^{(n)} \rightarrow \xi \in \Gamma(2), \quad (5.10)$$

$$\sup_{0 \leq t \leq n_1} |n^{-1}M_t^{(n)} - \Xi^{(n)}| \rightarrow 0. \quad (5.11)$$

It follows from (5.9)–(5.11) that a.s.,

$$n^{-1}W_{t\sqrt{n}}^{(n)} = n^{-1}M_{t\sqrt{n}}^{(n)} - n^{-1}A_{t\sqrt{n}}^{(n)} = n^{-1}M_{t\sqrt{n}}^{(n)} - \hat{A}_t^{(n)} \rightarrow \xi - \mathcal{A}(t), \quad (5.12)$$

uniformly on each compact interval in $(0, \infty)$.

Step 3: Identifying the limit. Since the limit in (5.12) is continuous, (5.12) and (2.6) yield, again a.s. uniformly on each compact interval in $(0, \infty)$,

$$\frac{Y_{[t\sqrt{n}]}^{(n)}}{[t\sqrt{n}]} = \frac{W_{[t\sqrt{n}]}^{(n)}}{(t^2 + o(1))n} \rightarrow \mathcal{B}(t) := t^{-2}(\xi - \mathcal{A}(t)). \quad (5.13)$$

By (5.4) and (2.15), with $k := [t\sqrt{n}]$,

$$\frac{d}{dt} \hat{A}_t^{(n)} = n^{-1/2}(A_k - A_{k-1}) = -n^{-1/2} \cdot 2k \left(\left(1 - \frac{1}{k}\right)^{Y_k} - 1 + \frac{Y_k}{k} \right), \quad (5.14)$$

and thus (5.13) implies that a.s., uniformly on each compact interval in $(0, \infty)$,

$$\frac{d}{dt} \hat{A}_t^{(n)} \rightarrow -2t(e^{-\mathcal{B}(t)} - 1 + \mathcal{B}(t)). \quad (5.15)$$

It follows from (5.9) and (5.15) that a.s., if $0 < t_1 < t_2 < \infty$,

$$\mathcal{A}(t_2) - \mathcal{A}(t_1) = \lim_{n \rightarrow \infty} (\hat{A}_{t_2}^{(n)} - \hat{A}_{t_1}^{(n)}) = \int_{t_1}^{t_2} \left(-2t(e^{-\mathcal{B}(t)} - 1 + \mathcal{B}(t)) \right) dt. \quad (5.16)$$

Consequently, a.s. the random function $\mathcal{A}(t)$ is continuously differentiable on $(0, \infty)$, with derivative

$$\mathcal{A}'(t) = -2t(e^{-\mathcal{B}(t)} - 1 + \mathcal{B}(t)), \quad 0 < t < \infty. \quad (5.17)$$

This and the definition of $\mathcal{B}(t)$ in (5.13) yield a differential equation for $\mathcal{A}(t)$, which we solve as follows. First, let

$$\mathcal{C}(t) := t^2 \mathcal{B}(t) = \xi - \mathcal{A}(t). \quad (5.18)$$

Then

$$2t\mathcal{B}(t) + t^2\mathcal{B}'(t) = \mathcal{C}'(t) = -\mathcal{A}'(t) = 2t(e^{-\mathcal{B}(t)} - 1 + \mathcal{B}(t)) \quad (5.19)$$

and thus

$$\mathcal{B}'(t) = \frac{2}{t}(e^{-\mathcal{B}(t)} - 1) \quad (5.20)$$

which can be written as

$$\frac{e^{\mathcal{B}(t)} d\mathcal{B}(t)}{e^{\mathcal{B}(t)} - 1} = -\frac{2 dt}{t} \quad (5.21)$$

with the solution, for some $c \in \mathbb{R}$,

$$\log(e^{\mathcal{B}(t)} - 1) = c - 2 \log t \quad (5.22)$$

and thus, with $C := e^c > 0$,

$$\mathcal{B}(t) = \log(1 + C/t^2), \quad t > 0. \quad (5.23)$$

Note that the constants c and C may be random.

We have shown that (5.23) holds a.s., for some random C , and (5.13) then yields

$$\mathcal{A}(t) = \xi - t^2 \mathcal{B}(t) = \xi - t^2 \log(1 + C/t^2), \quad t > 0. \quad (5.24)$$

It follows that a.s.

$$\mathcal{A}(t) \rightarrow \xi - C \quad \text{as } t \rightarrow \infty. \quad (5.25)$$

On the other hand, for every fixed $t > 0$, as in (5.6),

$$\mathbb{E} \hat{A}_t^{(n)} \leq 12 \frac{n}{[t\sqrt{n}]^2} \rightarrow \frac{12}{t^2}, \quad (5.26)$$

which by (5.9) and Fatou's lemma implies

$$\mathbb{E} \mathcal{A}(t) \leq \frac{12}{t^2}, \quad t > 0. \quad (5.27)$$

In particular, $\mathcal{A}(t) \xrightarrow{\mathbb{P}} 0$ as $t \rightarrow \infty$, which together with (5.25) yields $C = \xi$.

We thus have shown that (5.9), (5.12) and (5.13) a.s. hold uniformly on each compact interval in $(0, \infty)$, with

$$\mathcal{A}(t) = \xi - t^2 \log(1 + \xi/t^2). \quad (5.28)$$

Step 4: Convergence on $[0, \infty)$. We extend the results just shown from $(0, \infty)$ to $[0, \infty)$ as follows. First, note that $\mathcal{A}(t)$ in (5.28) extends to a continuous function on $[0, \infty)$,

with $\mathcal{A}(0) = \xi$. We have $A_0^{(n)} \leq M_0^{(n)}$ by (2.9), and $n^{-1}M_0^{(n)} \rightarrow \xi$ a.s. by (5.10) and (5.11). Hence, a.s.,

$$\limsup_{n \rightarrow \infty} n^{-1}A_0^{(n)} \leq \limsup_{n \rightarrow \infty} n^{-1}M_0^{(n)} = \xi = \mathcal{A}(0). \quad (5.29)$$

On the other hand, for every $t > 0$, $A_{t\sqrt{n}}^{(n)} \leq A_0^{(n)}$ and thus by (5.9), a.s.

$$\liminf_{n \rightarrow \infty} n^{-1}A_0^{(n)} \geq \liminf_{n \rightarrow \infty} n^{-1}A_{t\sqrt{n}}^{(n)} = \mathcal{A}(t). \quad (5.30)$$

Letting $t \searrow 0$ yields $\mathcal{A}(t) \nearrow \mathcal{A}(0)$ and thus

$$\liminf_{n \rightarrow \infty} n^{-1}A_0^{(n)} \geq \mathcal{A}(0). \quad (5.31)$$

Consequently, a.s., $\hat{A}_t^{(n)} \rightarrow \mathcal{A}(t)$ for $t = 0$ too. We thus have (5.9) a.s. for each fixed $t \geq 0$. Since $\hat{A}_t^{(n)}$ and $\mathcal{A}(t)$ are decreasing in t , and $\mathcal{A}(t)$ is continuous, this implies uniform convergence on each compact interval $[0, b] \subset [0, \infty)$. It follows from (5.10)–(5.11) that (5.12) also holds a.s. on each compact interval in $[0, \infty)$, i.e., in $C[0, \infty)$. This means, by (5.28), a.s. uniformly on each compact interval,

$$n^{-1}W_{t\sqrt{n}}^{(n)} \rightarrow t^2 \log(1 + \xi/t^2). \quad (5.32)$$

By (5.10) and the fact that $\frac{d}{dx} \log(1+x) \leq 1$, (5.32) yields also

$$n^{-1}W_{t\sqrt{n}}^{(n)} - t^2 \log(1 + \Xi^{(n)}/t^2) \rightarrow 0, \quad (5.33)$$

a.s. uniformly on each compact interval in $[0, \infty)$.

Step 5: Uncoupling. The a.s. convergence in (5.33) depends on the chosen coupling of D_n for different n , but this yields (5.33) with convergence in probability in general, i.e., (5.7).

Step 6: Conclusion. We have so far proved (5.7) only for a subsequence, but the same proof shows that every subsequence has a subsubsequence such that (5.7) holds, which as is well known implies that (5.7) holds for the full sequence, see e.g. [11, Section 5.7]. $\square$

6. THE NUMBER OF DESCENDANTS

Recall that the random variable $X = X^{(n)}$ is the number of descendants of n , i.e. red vertices, and thus, counting the root n separately,

$$X = 1 + \sum_{k=1}^{n-1} J_k. \quad (6.1)$$

We make a Doob decomposition similar to (2.9); in this case it takes the form, since J_k is $\mathcal{F}_{k-1}$ -measurable,

$$X = 1 + L_0 + B_0, \quad (6.2)$$

where

$$L_k := \sum_{i=k+1}^{n-1} (J_i - \mathbb{E}(J_i \mid \mathcal{F}_i)) \quad (6.3)$$

so that $(L_k)_0^{n-1}$ is a reverse martingale with $L_{n-1} = 0$: $\mathbb{E}(L_{k-1} \mid \mathcal{F}_k) = L_k$, and, using (2.1) and (2.3),

$$B_k := \sum_{i=k+1}^{n-1} \mathbb{E}(J_i \mid \mathcal{F}_i) = \sum_{i=k+1}^{n-1} \mathbb{P}(Z_i \geq 1 \mid \mathcal{F}_i) = \sum_{i=k+1}^{n-1} \left(1 - \left(1 - \frac{1}{i}\right)^{Y_i}\right) \quad (6.4)$$

is positive and increasing backwards:

$$0 = B_{n-1} \leq \dots \leq B_1 \leq B_0. \quad (6.5)$$

By (6.4) and Lemma 2.2, for every $k \leq n-1$,

$$\mathbb{E} B_k = \sum_{i=k+1}^{n-1} \mathbb{E} J_i = \sum_{i=k+1}^{n-1} \mathbb{P}(Z_i \geq 1) \leq \sum_{i=k+1}^{\infty} \frac{2n}{i^2} \leq \frac{2n}{k}. \quad (6.6)$$

This is too coarse for small k ; however, since $0 \leq J_i \leq 1$ for every i , we also have $B_0 - B_\ell \leq \ell$ for every $\ell \leq n-1$. Hence, (6.6) implies

$$\mathbb{E} B_0 \leq \mathbb{E} B_{\lfloor \sqrt{n} \rfloor} + \lfloor \sqrt{n} \rfloor \leq 4\sqrt{n}. \quad (6.7)$$

Since $(J_i \mid \mathcal{F}_i)$ is a Bernoulli variable, $\text{Var}(J_i \mid \mathcal{F}_i) \leq \mathbb{E}(J_i \mid \mathcal{F}_i)$, and thus the (reverse) martingale property of (L_k) yields

$$\mathbb{E} L_0^2 = \sum_{i=1}^{n-1} \mathbb{E} [\text{Var}(J_i \mid \mathcal{F}_i)] \leq \sum_{i=1}^{n-1} \mathbb{E} [\mathbb{E}(J_i \mid \mathcal{F}_i)] = \sum_{i=1}^{n-1} \mathbb{E} J_i = \mathbb{E} B_0 \leq 4\sqrt{n}. \quad (6.8)$$

In particular, $L_0/\sqrt{n} \xrightarrow{P} 0$, which will show that L_0 is negligible in (6.2).

Lemma 6.1. *As $n \rightarrow \infty$,*

$$\left| \frac{X^{(n)}}{\sqrt{n}} - \frac{\pi}{2} \sqrt{\Xi^{(n)}} \right| \xrightarrow{P} 0. \quad (6.9)$$

Thus,

$$\frac{X^{(n)}}{\sqrt{n}} \xrightarrow{d} \frac{\pi}{2} \sqrt{\xi}, \quad (6.10)$$

with $\xi \in \Gamma(2)$.

Proof. For convenience, we use the Skorohod coupling theorem as in the proof of Theorem 5.3; we may thus assume that all a.s. convergence results in the proof of Theorem 5.3 hold. (We may for simplicity consider the same subsequence as in the proof of Theorem 5.3, and then draw the conclusion for the full sequence as there; alternatively, we may argue that now when Theorem 5.3 is proved, we may consider the full sequence when we apply the Skorohod coupling theorem.) In particular, (5.13) and (5.28) (or (5.32)) yield

$$\frac{Y_{\lfloor t\sqrt{n} \rfloor}^{(n)}}{\lfloor t\sqrt{n} \rfloor} = \frac{W_{\lfloor t\sqrt{n} \rfloor}^{(n)}}{(t^2 + o(1))n} \rightarrow \mathcal{B}(t) = \log(1 + \xi/t^2) \quad (6.11)$$

a.s. uniformly on each compact interval in $(0, \infty)$.

We extend B_k to real arguments by linear interpolation and define also, similarly to (5.3) but with a different scaling,

$$\hat{B}_t^{(n)} := n^{-1/2} B_{t\sqrt{n}}^{(n)}. \quad (6.12)$$

Then, with $k := \lceil t\sqrt{n} \rceil$,

$$\frac{d}{dt} \hat{B}_t^{(n)} = -\mathbb{E}(J_k | \mathcal{F}_k) = \left(1 - \frac{1}{k}\right)^{Y_k} - 1 \quad (6.13)$$

and thus it follows from (6.11) that, uniformly on each compact interval in $(0, \infty)$,

$$\frac{d}{dt} \hat{B}_t^{(n)} \rightarrow e^{-\mathcal{B}(t)} - 1 = \frac{1}{1 + \xi/t^2} - 1 = -\frac{\xi}{\xi + t^2}. \quad (6.14)$$

Consequently, if $0 < t_1 < t_2 < \infty$, a.s.

$$\hat{B}_{t_1}^{(n)} - \hat{B}_{t_2}^{(n)} = -\int_{t_1}^{t_2} \frac{d}{dt} \hat{B}_t^{(n)} dt \rightarrow \int_{t_1}^{t_2} \frac{\xi}{\xi + t^2} dt = \sqrt{\xi} \left(\arctan \frac{t_2}{\sqrt{\xi}} - \arctan \frac{t_1}{\sqrt{\xi}} \right). \quad (6.15)$$

Since (6.13) implies $|\frac{d}{dt} \hat{B}_t^{(n)}| \leq 1$, we have

$$|\hat{B}_0^{(n)} - \hat{B}_{t_1}^{(n)}| \leq t_1. \quad (6.16)$$

Furthermore, (6.6) implies, for $t_2 \geq 1$,

$$\mathbb{E} \hat{B}_{t_2}^{(n)} = n^{-1/2} \mathbb{E} B_{t_2 \sqrt{n}} \leq \frac{2\sqrt{n}}{\lfloor t_2 \sqrt{n} \rfloor} \leq \frac{4}{t_2}. \quad (6.17)$$

Thus, letting $t_1 \rightarrow 0$ and $t_2 \rightarrow \infty$, we have $\hat{B}_0^{(n)} - (\hat{B}_{t_1}^{(n)} - \hat{B}_{t_2}^{(n)}) \xrightarrow{p} 0$, uniformly in n , and it follows from (6.15) by standard arguments that

$$\hat{B}_0^{(n)} \xrightarrow{p} \int_0^\infty \frac{\xi}{\xi + t^2} dt = \frac{\pi}{2} \sqrt{\xi}. \quad (6.18)$$

Recall from (6.12) that $B_0 = \sqrt{n} \hat{B}_0^{(n)}$. The results (6.9)–(6.10) now follow from (6.18) by (6.2), (6.8), and (5.10). $\square$

Proof of Theorem 1.3, first part. The limit in distribution (1.2) follows immediately from (6.10), using the well known facts that $\chi_4^2 \in \chi^2(4)$ and thus $\frac{1}{2}\chi_4^2 \in \Gamma(2)$, see Section 1.2, and consequently

$$\sqrt{\xi} \stackrel{d}{=} 2^{-1/2} \chi_4. \quad (6.19)$$

$\square$

7. HIGHER MOMENTS

In this section we prove some inequalities for higher moments. We do not care about exact constants, and we use the convention that c_p stands for constants that may (and will) depend on the parameter p , but not on n ; the value of c_p may change from one occurrence to another.

We consider first the reverse martingale M_k . We define the maximal function

$$M^* := \max_{n-1 \geq k \geq 0} M_k, \quad (7.1)$$

the martingale differences, for $n-1 \geq k \geq 1$, recalling (2.10), (2.6) and (2.2),

$$\begin{aligned} \Delta M_k &:= M_{k-1} - M_k = W_{k-1} - \mathbb{E}(W_{k-1} | \mathcal{F}_k) \\ &= k(Y_{k-1} - \mathbb{E}(Y_{k-1} | \mathcal{F}_k)) \\ &= -k(Z_k - \mathbb{E}(Z_k | \mathcal{F}_k)) + 2k(J_k - \mathbb{E}(J_k | \mathcal{F}_k)), \end{aligned} \quad (7.2)$$

and the conditional square function

$$s(M) := \left(\sum_{i=1}^{n-1} \mathbb{E}((\Delta M_i)^2 \mid \mathcal{F}_i) \right)^{1/2}. \quad (7.3)$$

We use one of Burkholder's martingale inequalities [7, Theorem 21.1], [11, Corollary 10.9.1] on the martingale $M_k - M_{n-1} = M_k - 2n$, which yields

$$\begin{aligned} \mathbb{E}(M^*)^p &\leq c_p(2n)^p + c_p \mathbb{E}(\max_k |M_k - 2n|)^p \\ &\leq c_p n^p + c_p \mathbb{E} s(M)^p + c_p \mathbb{E}(\max_k |\Delta M_k|)^p \\ &\leq c_p n^p + c_p \mathbb{E} s(M)^p + c_p \sum_{k=1}^{n-1} \mathbb{E} |\Delta M_k|^p. \end{aligned} \quad (7.4)$$

(This is valid for any $p > 0$, although we only use $p \geq 2$.)

Lemma 7.1. *For every $p > 0$,*

$$\mathbb{E}(M^*)^p \leq c_p n^p. \quad (7.5)$$

Proof. By Lyapunov's inequality, it suffices to prove (7.5) for $p = 2^j$, $j \geq 1$ integer. We use induction on j . The base case $p = 2$ is proved in Lemma 2.1. In the rest of the proof, we thus assume $p \geq 4$ and that (7.5) holds for the exponent $p/2$ (or smaller). We use (7.4), and it remains to estimate the two last terms on its right-hand side.

First, by (7.2) and (2.17),

$$\mathbb{E}((\Delta M_i)^2 \mid \mathcal{F}_i) = \text{Var}(W_{k-1} \mid \mathcal{F}_k) \leq 10W_k \leq 10M_k \leq 10M^*. \quad (7.6)$$

Hence, (7.3) yields

$$s(M) \leq \sqrt{10nM^*} \quad (7.7)$$

and the induction hypothesis yields

$$\mathbb{E} s(M)^p \leq c_p n^{p/2} \mathbb{E}(M^*)^{p/2} \leq c_p n^p. \quad (7.8)$$

Next, we recall the well known moment estimate for the binomial distribution

$$\mathbb{E} |\zeta - \mathbb{E} \zeta|^p \leq c_p (Nq)^{p/2} + c_p Nq, \quad \zeta \in \text{Bin}(N, q). \quad (7.9)$$

(Coincidentally, this can be shown by the Burkholder inequality used in (7.4), writing the binomial variable $\zeta - \mathbb{E} \zeta$ as a sum of N independent centred Bernoulli variables.) Hence, recalling the conditional distribution (2.3), we have

$$\begin{aligned} \mathbb{E}(|Z_k - \mathbb{E}(Z_k \mid \mathcal{F}_k)|^p \mid \mathcal{F}_k) &\leq c_p (Y_k/k)^{p/2} + c_p (Y_k/k) \leq c_p (W_k/k^2)^{p/2} + c_p (W_k/k^2) \\ &\leq c_p (M^*/k^2)^{p/2} + c_p M^*/k^2 \end{aligned} \quad (7.10)$$

and thus

$$\mathbb{E} |Z_k - \mathbb{E}(Z_k \mid \mathcal{F}_k)|^p \leq c_p k^{-p} \mathbb{E}(M^*)^{p/2} + c_p k^{-2} \mathbb{E} M^*. \quad (7.11)$$

Consequently, by the induction hypothesis,

$$k^p \mathbb{E} |Z_k - \mathbb{E}(Z_k \mid \mathcal{F}_k)|^p \leq c_p n^{p/2} + c_p k^{p-2} n \leq c_p n^{p-1}. \quad (7.12)$$

Similarly, since J_k has a conditional Bernoulli distribution, and using (2.24),

$$\begin{aligned} \mathbb{E}(|J_k - \mathbb{E}(J_k \mid \mathcal{F}_k)|^p \mid \mathcal{F}_k) &\leq c_p \mathbb{E}(|J_k|^p \mid \mathcal{F}_k) = c_p \mathbb{E}(J_k \mid \mathcal{F}_k) \\ &\leq c_p W_k/k^2 \leq c_p M_k/k^2 \leq c_p M^*/k^2 \end{aligned} \quad (7.13)$$

and thus, using again (2.20),

$$k^p \mathbb{E} |J_k - \mathbb{E}(J_k | \mathcal{F}_k)|^p \leq c_p k^{p-2} \mathbb{E} M^* \leq c_p k^{p-2} n \leq c_p n^{p-1}. \quad (7.14)$$

Hence, (7.2), (7.12) and (7.14) yield

$$\mathbb{E} |\Delta M_k|^p \leq c_p k^p \mathbb{E} |Z_k - \mathbb{E}(Z_k | \mathcal{F}_k)|^p + c_p k^p \mathbb{E} |J_k - \mathbb{E}(J_k | \mathcal{F}_k)|^p \leq c_p n^{p-1}. \quad (7.15)$$

The induction step is shown by (7.4), (7.8) and (7.15), which completes the proof. $\square$

We proceed to our main objective, the number X of vertices in $\hat{D}_n$.

Lemma 7.2. *For every $p > 0$,*

$$\mathbb{E} (X^{(n)})^p \leq c_p n^{p/2}. \quad (7.16)$$

Proof. We use the decomposition (6.2) and argue similarly as in the proof of Lemma 7.1. First, by (6.4), cf. (6.6)–(6.7),

$$B_k \leq \sum_{i=k+1}^{n-1} \frac{Y_i}{i} \leq \sum_{i=k+1}^{n-1} \frac{W_i}{i^2} \leq M^* \sum_{i=k+1}^{\infty} \frac{1}{i^2} \leq \frac{M^*}{k} \quad (7.17)$$

and thus

$$B_0 \leq B_{\lfloor \sqrt{n} \rfloor} + \lceil \sqrt{n} \rceil \leq n^{-1/2} M^* + 2n^{1/2}. \quad (7.18)$$

Hence, by Lemma 7.1, for every $p > 0$,

$$\mathbb{E} B_0^p \leq c_p n^{-p/2} \mathbb{E} (M^*)^p + c_p n^{p/2} \leq c_p n^{p/2}. \quad (7.19)$$

Next, the conditional square function of the reverse martingale L_k is given by, see (6.3) and (6.4) and recall again that $(J_i | \mathcal{F}_i)$ is a Bernoulli variable,

$$s(L)^2 = \sum_{i=1}^{n-1} \text{Var}(J_i | \mathcal{F}_i) \leq \sum_{i=1}^{n-1} \mathbb{E}(J_i | \mathcal{F}_i) = B_0. \quad (7.20)$$

Consequently, using (7.19), for every $p > 0$,

$$\mathbb{E} s(L)^p \leq \mathbb{E} B_0^{p/2} \leq c_p n^{p/4}. \quad (7.21)$$

Furthermore,

$$\Delta L_k := L_{k-1} - L_k = J_k - \mathbb{E}(J_k | \mathcal{F}_k) \quad (7.22)$$

and thus $|\Delta L_k| \leq 1$. Consequently, the conditional Burkholder inequality in (7.4) yields

$$\mathbb{E} |L_0|^p \leq c_p \mathbb{E} s(L)^p + c_p \mathbb{E} \left(\max_k |\Delta L_k| \right)^p \leq c_p n^{p/4}. \quad (7.23)$$

The result (7.16) now follows from (6.2), (7.19) and (7.23). $\square$

Proof of Theorem 1.3, conclusion. Lemma 7.2 shows that $\mathbb{E} |X^{(n)}/\sqrt{n}|^p = O(1)$ for every fixed $p > 0$. By a standard argument, see e.g. [11, Theorems 5.4.2 and 5.5.9], this implies uniform integrability of the sequence $|X^{(n)}/\sqrt{n}|^p$ for every $p > 0$ and thus convergence of all moments in (6.10). (Recall that convergence in distribution was proved in Section 6.)

Finally, (1.3)–(1.4) now follow from the formula

$$\mathbb{E} \chi_4^r = 2^{r/2} \Gamma\left(\frac{r}{2} + 2\right), \quad (7.24)$$

which is a simple consequence of (1.1), or of (6.19) and (1.8). This completes the proof. $\square$

8. HIGHER DEGREE d

We have so far considered the random 2-dag, with outdegree $d = 2$. The arguments and results above extend to any constant $d \geq 2$ with minor modifications which we sketch here, omitting straightforward details. We let $d \geq 2$ be fixed, and let c and c_p denote constants that may depend on d (and p); these may change value from one occurrence to the next. Note that the case $d = 2$ treated above is included as a special case below.

We define Y_k , Z_k , J_k , and $\mathcal{F}_k$ as in Section 2; thus $Y_{n-1} = d$, (2.1) and (2.3) still hold, but (2.2) is replaced by

$$Y_{k-1} = Y_k - Z_k + dJ_k. \quad (8.1)$$

Then, instead of (2.4)–(2.5),

$$\begin{aligned} \mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) &= Y_k - \mathbb{E}(Z_k \mid \mathcal{F}_k) + d\mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k) \\ &= Y_k - \frac{1}{k}Y_k + d(1 - (1 - \frac{1}{k})^{Y_k}) \end{aligned} \quad (8.2)$$

and

$$\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) \leq Y_k - \frac{1}{k}Y_k + \frac{d}{k}Y_k = \frac{k+d-1}{k}Y_k. \quad (8.3)$$

We now define, letting $m^{\bar{\ell}} := m(m+1) \cdots (m+\ell-1)$ denote the rising factorial,

$$W_k := (k+1)^{\overline{d-1}}Y_k = (k+1) \cdots (k+d-1)Y_k. \quad (8.4)$$

Then, (8.3) yields

$$\mathbb{E}(W_{k-1} \mid \mathcal{F}_k) \leq W_k, \quad (8.5)$$

and thus again W_k is a reverse supermartingale, with a Doob decomposition (2.9) where now

$$M_{n-1} = W_{n-1} = dn^{\overline{d-1}} = O(n^{d-1}). \quad (8.6)$$

We still have (2.16), up to the numerical constants (which depend on d), while we now have

$$\text{Var}(W_{k-1} \mid \mathcal{F}_k) = O(k^{2d-3}Y_k) = O(k^{d-2}W_k) \quad (8.7)$$

and

$$\mathbb{E}W_k^2 \leq \mathbb{E}M_k^2 \leq \mathbb{E}(M^*)^2 = O(n^{2d-2}). \quad (8.8)$$

Lemmas 2.2–2.3 take the form

$$\mathbb{P}(Z_k \geq 1) \leq c \frac{n^{d-1}}{k^d}, \quad (8.9)$$

$$\mathbb{P}(Z_k \geq 2) \leq c \frac{n^{2d-2}}{k^{2d}}, \quad (8.10)$$

$$A_{k-1} - A_k \leq c \frac{W_k^2}{k^{d+1}}, \quad (8.11)$$

$$\mathbb{E}A_k \leq c \frac{n^{2d-2}}{k^d}. \quad (8.12)$$

The moment estimates in Section 7 extend too. We find $s(M) \leq c\sqrt{n^{d-1}M^*}$ and obtain by induction, for every $p > 0$,

$$\mathbb{E}(M^*)^p \leq c_p n^{p(d-1)}. \quad (8.13)$$

Instead of (7.17)–(7.19) and (7.23) we obtain

$$B_k \leq M^*/k^{d-1}, \quad (8.14)$$

$$B_0 \leq B_{\lceil n^{(d-1)/d} \rceil} + \lceil n^{(d-1)/d} \rceil \leq n^{-(d-1)^2/d} M^* + 2n^{(d-1)/d}, \quad (8.15)$$

$$\mathbb{E} B_0^p \leq c_p n^{p(d-1)/d}, \quad (8.16)$$

$$\mathbb{E} |L_0|^p \leq c_p n^{p(d-1)/(2d)}, \quad (8.17)$$

and thus (6.2) yields

$$\mathbb{E} (X^{(n)})^p \leq c_p n^{p(d-1)/d}. \quad (8.18)$$

We may couple the initial phase of the dag with a branching process as in Section 3; however, now each particle splits into d new particles. The corresponding standard process $\mathcal{Y}'_t$ starting with one particle has probability generating function, see e.g. [2, Remark III.5.1],

$$\mathbb{E} s^{\mathcal{Y}'_t} = s e^{-t} [1 - (1 - e^{-(d-1)t}) s^{d-1}]^{-1/(d-1)}, \quad (8.19)$$

which means that $(\mathcal{Y}'_t - 1)/(d-1)$ has a negative binomial distribution $\text{NegBin}(\frac{1}{d-1}, e^{-(d-1)t})$. Since our version $\mathcal{Y}_t$ starts with d particles as $t = 0$, and thus $\mathcal{Y}_t$ is the sum of d independent copies of $\mathcal{Y}'_t$, it follows that $(\mathcal{Y}_t - d)/(d-1) \in \text{NegBin}(\frac{d}{d-1}, e^{-(d-1)t})$. It also follows that

$$\begin{aligned} \mathbb{E} \mathcal{Y}_t &= d \mathbb{E} \mathcal{Y}'_t = d + (d-1) \mathbb{E} \text{NegBin}\left(\frac{d}{d-1}, e^{-(d-1)t}\right) \\ &= d + (d-1) \frac{d}{d-1} (e^{(d-1)t} - 1) = d e^{(d-1)t} \end{aligned} \quad (8.20)$$

and thus, after the same time change as before,

$$\mathbb{E} \hat{\mathcal{Y}}_x = \mathbb{E} \mathcal{Y}_{-\log x} = d/x^{d-1}, \quad 0 < x \leq 1. \quad (8.21)$$

Moreover, also from (8.19), as $t \rightarrow \infty$,

$$e^{-(d-1)t} \mathcal{Y}'_t \xrightarrow{\text{a.s.}} \Gamma\left(\frac{1}{d-1}, d-1\right) \quad (8.22)$$

and thus, with $x = e^{-t} \rightarrow 0$,

$$e^{-(d-1)t} \mathcal{Y}_t = x^{d-1} \hat{\mathcal{Y}}_x \xrightarrow{\text{a.s.}} \Gamma\left(\frac{d}{d-1}, d-1\right). \quad (8.23)$$

We may choose $n_1 := \lfloor n/\log n \rfloor$ as in Section 3, and then Theorem 3.1 holds, except that $\log^2 n/n$ is replaced by $\log^d n/n$. Furthermore, we now have

$$\Xi^{(n)} := \frac{W_{n_1}^{(n)}}{n^{d-1}} \xrightarrow{d} \xi \in \Gamma\left(\frac{d}{d-1}, d-1\right). \quad (8.24)$$

In Section 4, we now choose $n_2 \gg n^{(d-1)/d}$, and we have

$$\max_{n_1 \geq k \geq 0} \left| \frac{M_k}{n^{d-1}} - \Xi^{(n)} \right| \xrightarrow{p} 0, \quad (8.25)$$

$$\max_{n_1 \geq k \geq n_2} \left| \frac{W_k}{n^{d-1}} - \Xi^{(n)} \right| \xrightarrow{p} 0. \quad (8.26)$$

In Section 5, we define

$$\hat{A}_t^{(n)} := n^{-(d-1)} A_{tn^{(d-1)/d}}^{(n)}, \quad t \geq 0. \quad (8.27)$$

Then tightness holds as in Lemma 5.2, and we can argue as in the proof of Theorem 5.3 using a suitable subsequence and a suitable coupling. Then (5.9) holds a.s. uniformly on compact intervals, and (5.13) becomes

$$\frac{Y_{[tn^{(d-1)/d}]}^{(n)}}{[tn^{(d-1)/d}]} = \frac{W_{[tn^{(d-1)/d}]}^{(n)}}{(t^d + o(1))n^{d-1}} \rightarrow \mathcal{B}(t) := t^{-d}(\xi - \mathcal{A}(t)). \quad (8.28)$$

This leads by the arguments above to

$$\frac{d}{dt} \hat{A}_t^{(n)} \rightarrow -dt^{d-1}(e^{-\mathcal{B}(t)} - 1 + \mathcal{B}(t)), \quad (8.29)$$

uniformly on compact intervals in $(0, \infty)$, and then to the differential equation (instead of (5.17))

$$\mathcal{A}'(t) = -dt^{d-1}(e^{-\mathcal{B}(t)} - 1 + \mathcal{B}(t)), \quad 0 < t < \infty, \quad (8.30)$$

with the solution

$$\mathcal{B}(t) = \log(1 + C/t^d), \quad t > 0, \quad (8.31)$$

where again we find $C = \xi$ a.s., and consequently

$$\mathcal{A}(t) = \xi - t^d \log(1 + \xi/t^d). \quad (8.32)$$

Finally, we extend the convergence to $[0, \infty)$ as above, and reach the conclusion that (generalizing Theorem 5.3)

$$\sup_{0 \leq t \leq b} \left| n^{-d} W_{tn^{(d-1)/d}}^{(n)} - t^d \log(1 + \Xi^{(n)}/t^d) \right| \xrightarrow{p} 0, \quad (8.33)$$

for every fixed $b > 0$.

In Section 6, we replace (6.11) by

$$\frac{Y_{[tn^{(d-1)/d}]}^{(n)}}{[tn^{(d-1)/d}]} = \frac{W_{[tn^{(d-1)/d}]}^{(n)}}{(t^d + o(1))n^{d-1}} \rightarrow \mathcal{B}(t) = \log(1 + \xi/t^d) \quad (8.34)$$

and define

$$\hat{B}_t^{(n)} := n^{-(d-1)/d} B_{tn^{(d-1)/d}}^{(n)}. \quad (8.35)$$

This leads to

$$n^{-(d-1)/d} B_0^{(n)} \xrightarrow{p} \int_0^\infty \frac{\xi}{\xi + t^d} dt = \xi^{1/d} \int_0^\infty \frac{1}{1 + x^d} dx = \frac{\pi}{d \sin(\pi/d)} \xi^{1/d}, \quad (8.36)$$

where the integral is evaluated by a substitution yielding a Beta integral [17, 5.12.3, together with 5.12.1 and 5.5.3]:

$$\begin{aligned} \int_0^\infty \frac{1}{1 + x^d} dx &= \frac{1}{d} \int_0^\infty \frac{y^{\frac{1}{d}-1}}{1 + y} dy = \frac{1}{d} B\left(\frac{1}{d}, \frac{d-1}{d}\right) = \frac{1}{d} \Gamma\left(\frac{1}{d}\right) \Gamma\left(\frac{d-1}{d}\right) \\ &= \frac{\pi}{d \sin(\pi/d)}. \end{aligned} \quad (8.37)$$

Proof of Theorem 1.4. The limit in distribution (1.5) follows from (6.2), (8.17) and (8.36), recalling (8.24) and writing $\xi = (d-1)\gamma$. Moment convergence then follows from the bounds (8.16) and (8.17) as in the case $d = 2$. Finally, the moment convergence and (1.8) yield (1.6). $\square$

9. FURTHER RESULTS

We give here some further results on the structure of the random dag $\hat{D}_n$. Again, we consider for simplicity only the case $d = 2$, and leave the straightforward extensions to larger d to the reader.

9.1. Density of descendants. The proof of Theorem 1.3 shows that most vertices in $\hat{D}_n$ are in the range $O(\sqrt{n})$. More precisely, let $0 \leq a \leq b \leq \infty$, and let

$$X_{a,b}^{(n)} := |\hat{D}_n \cap (a\sqrt{n}, b\sqrt{n}]|, \quad (9.1)$$

the number of descendants of n (red vertices) in the interval $(a\sqrt{n}, b\sqrt{n}]$. (Thus, $X^{(n)} = X_{0,\infty}^{(n)}$.) Then, Lemma 6.1 can be extended:

Lemma 9.1. *If $0 \leq a \leq b \leq \infty$ are fixed, then as $n \rightarrow \infty$,*

$$\frac{X_{a,b}^{(n)}}{\sqrt{n}} - \int_a^b \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} dt \xrightarrow{L^1} 0 \quad (9.2)$$

and thus

$$\mathbb{E} \left(\frac{X_{a,b}^{(n)}}{\sqrt{n}} \mid \Xi^{(n)} \right) - \int_a^b \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} dt \xrightarrow{L^1} 0 \quad (9.3)$$

and, unconditionally,

$$\frac{\mathbb{E} X_{a,b}^{(n)}}{\sqrt{n}} \rightarrow \int_a^b p(t) dt, \quad (9.4)$$

where

$$p(t) := \mathbb{E} \frac{\xi}{\xi + t^2} = \int_0^\infty \frac{x^2}{x + t^2} e^{-x} dx. \quad (9.5)$$

Proof. If $0 < a \leq b < \infty$, let $k_a := \lfloor a\sqrt{n} \rfloor$ and $k_b := \lfloor b\sqrt{n} \rfloor$. Then (6.3)–(6.4) show that, provided n is so large that $b\sqrt{n} < n$,

$$X_{a,b}^{(n)} = \sum_{k=k_a+1}^{k_b} J_k = B_{k_a} - B_{k_b} + L_{k_a} - L_{k_b}. \quad (9.6)$$

Convergence in probability in (9.2) then follows from (6.12) and (6.15) together with (6.8) (and, for example, Doob's inequality), and as always (5.10). If $a = 0$ or $b = \infty$, this result follows similarly using also (6.16)–(6.17) as in the proof of Lemma 6.1.

Thus, (9.2) holds in probability. This implies convergence also in L^1 , since uniform integrability holds because $X_{a,b}^{(n)}/\sqrt{n} \leq X^{(n)}/\sqrt{n}$ and, recalling (3.10),

$$\int_a^b \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} dt \leq \int_0^\infty \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} dt = \frac{\pi}{2} \sqrt{\Xi^{(n)}} \leq \frac{\pi}{2} \sqrt{\frac{M^*}{n}} \quad (9.7)$$

and these are uniformly integrable by Lemmas 7.2 and Lemma 7.1.

Next, (9.3) follows from (9.2) by taking the conditional expectation, and (9.4) follows by taking the unconditional expectation, using (3.11) and Fubini's theorem, and again the uniform integrability of (9.7). The final equality in (9.4) follows since $\xi \in \Gamma(2)$ has density function xe^{-x} by (1.7). $\square$

Remark 9.2. The function $p(t)$ can be expressed using the exponential integral $E_1(x)$, see [17, 6.2.1-2 and 6.7.1]:

$$p(t) = \int_0^\infty \left(x - t^2 + \frac{t^4}{x + t^2} \right) e^{-x} dx = 1 - t^2 + t^4 e^{t^2} E_1(t^2). \quad (9.8)$$

△

Informally, Lemma 9.1 says that, asymptotically, the density of descendants of n around any $k < n$ is $\Xi/(\Xi + k^2/n)$ conditioned on Ξ , and $p(k/\sqrt{n})$ unconditionally. Another aspect of this is the following theorem, where we consider a single vertex k .

Theorem 9.3. *Conditioned on $\Xi^{(n)}$, the probability that vertex k is a descendant of n (i.e., belongs to $\hat{D}_n$) is*

$$\mathbb{P}(J_k^{(n)} = 1 \mid \Xi^{(n)}) = \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} + o_p(1), \quad (9.9)$$

uniformly in $k \leq n_1$. Hence, the unconditional probability is, with $p(t)$ given by (9.5),

$$\mathbb{P}(J_k^{(n)} = 1) = \mathbb{E} \frac{\xi}{\xi + k^2/n} + o(1) = p(k/\sqrt{n}) + o(1), \quad (9.10)$$

uniformly in $k \leq n_1$.

Proof. Recall that $J_k = \mathbf{1}\{k \in \hat{D}_n\}$ is a Bernoulli variable; hence $\mathbb{P}(J_k = 1) = \mathbb{E} J_k$, and this holds also conditionally.

Consider first $k \in [\delta\sqrt{n}, b\sqrt{n}]$ for some fixed $0 < \delta < b < \infty$, and let $t := k/\sqrt{n} \in [\delta, b]$. Using again the simplifying assumptions in the proof of Theorem 5.3, we see from (6.13)–(6.14) and (5.10) that

$$\max_{k \in [\delta\sqrt{n}, b\sqrt{n}]} \left| \mathbb{E}(J_k \mid \mathcal{F}_k) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \xrightarrow{P} 0. \quad (9.11)$$

Hence, by dominated convergence (the max is bounded by 1),

$$\mathbb{E} \max_{k \in [\delta\sqrt{n}, b\sqrt{n}]} \left| \mathbb{E}(J_k \mid \mathcal{F}_k) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \rightarrow 0. \quad (9.12)$$

Thus, by taking the conditional expectation with respect to $\Xi^{(n)}$, assuming that n is so large that $b\sqrt{n} \leq n_1$ and thus $\mathbb{E}(J_k \mid \Xi^{(n)}) = \mathbb{E}[\mathbb{E}(J_k \mid \mathcal{F}_k) \mid \Xi^{(n)}]$,

$$\begin{aligned} & \max_{k \in [\delta\sqrt{n}, b\sqrt{n}]} \left| \mathbb{E}(J_k \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \\ & \leq \max_{k \in [\delta\sqrt{n}, b\sqrt{n}]} \mathbb{E} \left(\left| \mathbb{E}(J_k \mid \mathcal{F}_k) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \mid \Xi^{(n)} \right) \\ & \leq \mathbb{E} \left(\max_{k \in [\delta\sqrt{n}, b\sqrt{n}]} \left| \mathbb{E}(J_k \mid \mathcal{F}_k) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \mid \Xi^{(n)} \right) \xrightarrow{L^1} 0, \end{aligned} \quad (9.13)$$

Furthermore, if $n_1 \geq k > \ell \geq 1$, then when the evolution comes to k , we have Y_k red edges, and each of them ends at k with probability $1/k$. We have the same probability for each of these edges to end at ℓ instead, and since the endpoints are independent, we see that conditioned on $\mathcal{F}_k$, Z_ℓ is stochastically larger than Z_k . (Larger, since there may also be red edges ending at ℓ that start at k or later.) Hence,

$$\mathbb{E}(J_\ell \mid \mathcal{F}_k) = \mathbb{P}(Z_\ell \geq 1 \mid \mathcal{F}_k) \geq \mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k) = \mathbb{E}(J_k \mid \mathcal{F}_k) \quad (9.14)$$

and thus $\mathbb{E}(J_\ell \mid \Xi^{(n)}) \geq \mathbb{E}(J_k \mid \Xi^{(n)})$. In other words, $\mathbb{E}(J_k \mid \Xi^{(n)})$ is decreasing in $k \in [1, n_1]$. The same obviously holds for $\Xi^{(n)}/(\Xi^{(n)} + k^2/n)$. Consequently, with $k_b := \lfloor b\sqrt{n} \rfloor$,

$$\begin{aligned} & \max_{k \in [b\sqrt{n}, n_1]} \left| \mathbb{E}(J_k \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \\ & \leq \max_{k \in [b\sqrt{n}, n_1]} \left(\mathbb{E}(J_k \mid \Xi^{(n)}) + \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right) \\ & \leq \left| \mathbb{E}(J_{k_b} \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k_b^2/n} \right| + 2 \frac{\Xi^{(n)}}{\Xi^{(n)} + k_b^2/n}. \end{aligned} \quad (9.15)$$

Hence, using (9.13),

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \mathbb{E} \max_{k \in [b\sqrt{n}, n_1]} \left| \mathbb{E}(J_k \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \\ & \leq 2 \limsup_{n \rightarrow \infty} \mathbb{E} \frac{\Xi^{(n)}}{\Xi^{(n)} + k_b^2/n} = 2 \mathbb{E} \frac{\xi}{\xi + b^2}, \end{aligned} \quad (9.16)$$

which can be made arbitrarily small by choosing b large (by dominated convergence). Similarly, with $k_\delta := \lfloor \delta\sqrt{n} \rfloor$,

$$\begin{aligned} & \max_{k \in [1, \delta\sqrt{n}]} \left| \mathbb{E}(J_k \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \\ & \leq \max_{k \in [1, \delta\sqrt{n}]} \left(\left(1 - \mathbb{E}(J_k \mid \Xi^{(n)})\right) + \left(1 - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n}\right) \right) \\ & \leq \left| \mathbb{E}(J_{k_\delta} \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k_\delta^2/n} \right| + 2 \left(1 - \frac{\Xi^{(n)}}{\Xi^{(n)} + k_\delta^2/n}\right) \end{aligned} \quad (9.17)$$

and thus

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \mathbb{E} \max_{k \in [1, \delta\sqrt{n}]} \left| \mathbb{E}(J_k \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \\ & \leq 2 \limsup_{n \rightarrow \infty} \mathbb{E} \frac{k_\delta^2/n}{\Xi^{(n)} + k_\delta^2/n} = 2 \mathbb{E} \frac{\delta^2}{\xi + \delta^2}, \end{aligned} \quad (9.18)$$

which can be made arbitrarily small by choosing δ small.

It follows from (9.13), (9.16) and (9.18) that

$$\max_{k \in [1, n_1]} \left| \mathbb{E}(J_k \mid \Xi^{(n)}) - \frac{\Xi^{(n)}}{\Xi^{(n)} + k^2/n} \right| \xrightarrow{L^1} 0, \quad (9.19)$$

which is a more precise version of (9.9)

Finally, (5.10) implies

$$\sup_{t>0} \left| \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} - \frac{\xi}{\xi + t^2} \right| = \sup_{t>0} \frac{t^2 |\Xi^{(n)} - \xi|}{(\Xi^{(n)} + t^2)(\xi + t^2)} \leq \frac{|\Xi^{(n)} - \xi|}{\xi} \xrightarrow{\text{a.s.}} 0, \quad (9.20)$$

and thus, by dominated convergence,

$$\sup_{t>0} \left| \mathbb{E} \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} - \mathbb{E} \frac{\xi}{\xi + t^2} \right| \leq \mathbb{E} \sup_{t>0} \left| \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} - \frac{\xi}{\xi + t^2} \right| \rightarrow 0, \quad (9.21)$$

Hence, taking the expectation in (9.19) yields (9.10). $\square$

9.2. Different n yield asymptotically independent results. As noted in Remark 1.7, the construction naturally constructs the dags D_n for all n together. Using this coupling, we may consider the joint distribution of, for example, $X^{(n)}$ and $X^{(n+1)}$. Somewhat surprisingly, $X^{(n)}$ and $X^{(n+1)}$ are asymptotically independent:

Theorem 9.4. *As $n \rightarrow \infty$,*

$$(X^{(n)}/\sqrt{n}, X^{(n+1)}/\sqrt{n+1}) \xrightarrow{d} (\zeta, \zeta'), \quad (9.22)$$

where ζ and ζ' are independent copies of the limit $(\pi/\sqrt{8})\chi_4$ in (1.2).

Proof. Consider the evolutions of the red dags $\hat{D}_n$ and $\hat{D}_{n+1}$ together, starting at n and $n+1$ and going down, as always; these evolutions are independent until they first have a common vertex. The probability that k is the first common vertex is thus at most the probability that two independent versions of $\hat{D}_n$ and $\hat{D}_{n+1}$ both contain k , which by (2.22) is

$$\begin{cases} \mathbb{P}(Z_k^{(n)} \geq 1) \mathbb{P}(Z_k^{(n+1)} \geq 1) \leq \frac{4n(n+1)}{k^4} \leq c \frac{n^2}{k^4}, & k < n, \\ \mathbb{P}(Z_n^{(n+1)} \geq 1) \leq \frac{2(n+1)}{n^2}, & k = n. \end{cases} \quad (9.23)$$

Consequently, the probability that $\hat{D}_n$ and $\hat{D}_{n+1}$ meet before $n_1^{(n)}$ is

$$\leq \sum_{k=n_1}^{n-1} c \frac{n^2}{k^4} + \frac{c}{n} \leq c \frac{n^2}{n_1^3} + \frac{c}{n} = o(1). \quad (9.24)$$

Consequently, w.h.p. $\hat{D}_n$ and $\hat{D}_{n+1}$ are independent until n_1 ; more formally, we may couple the pair $(\hat{D}_n, \hat{D}_{n+1})$ with a pair $(\hat{D}'_n, \hat{D}'_{n+1})$ of independent copies of them such that the two pairs w.h.p. coincide until $n_1^{(n)}$. In particular, this and the definition (3.10) show that the pair $(\Xi^{(n)}, \Xi^{(n+1)})$ can be coupled with a pair of independent copies of them (defined in the same way from $\hat{D}'_n$ and $\hat{D}'_{n+1}$) such that the two pairs coincide w.h.p. Consequently, Lemma 3.2 implies that

$$(\Xi^{(n)}, \Xi^{(n+1)}) \xrightarrow{d} (\xi, \xi'), \quad (9.25)$$

where $\xi, \xi' \in \Gamma(2)$ are independent. The result then follows by (6.9). $\square$

This result may seem surprising, since we have seen that most vertices k in $\hat{D}_n$ and $\hat{D}_{n+1}$ have k of the order $\sqrt{n}$, and that in this range, the density of vertices is high, which means that $\hat{D}_n$ and $\hat{D}_{n+1}$ necessarily have a large number of common vertices. Since the $\hat{D}_n$ and $\hat{D}_{n+1}$ have the same descendants of any common vertex, it follows that the graphs $\hat{D}_n$ and $\hat{D}_{n+1}$ are strongly dependent. Nevertheless, the proof above shows that $\hat{D}_n$ and $\hat{D}_{n+1}$ are essentially independent in the first phase, which determines $\Xi^{(n)}$ and $\Xi^{(n+1)}$. Almost all vertices that contribute to $X^{(n)}$ and $X^{(n+1)}$ are in the later dense phase, where there is strong dependence, but this does not prevent the asymptotic independence of $X^{(n)}$ and $X^{(n+1)}$ because in this phase, there are so many vertices and edges that the evolution is governed by a law of large numbers and is essentially deterministic; hence the strong dependence here does not matter.

Remark 9.5. We considered above $X^{(n)}$ and $X^{(n+1)}$ only to be concrete. The result extends to $X^{(n'_\nu)}$ and $X^{(n''_\nu)}$ for any two sequences n'_ν and n''_ν that tend to infinity, with $n'_\nu < n''_\nu$. (This follows by the same proof, where we treat the cases $(n''_\nu)_1 \leq n'_\nu$ and $(n''_\nu)_1 > n'_\nu$ separately in the first part.)

Furthermore, the theorem extends to any finite number of such sequences. $\triangle$

Remark 9.6. Theorem 9.4 shows that the sequence $X^{(n)}/\sqrt{n}$ does *not* converge a.s.; in contrast, the theorem (with Remark 9.5) implies that this sequence a.s. oscillates wildly. $\triangle$

It seems interesting to understand the relations between $\hat{D}^{(n)}$ and $\hat{D}^{(n+1)}$ further. For example, consider the number of common vertices

$$\Upsilon^{(n)} := |\hat{D}^{(n)} \cap \hat{D}^{(n+1)}|. \quad (9.26)$$

We have $\Upsilon^{(n)} \leq \min(X^{(n)}, X^{(n+1)})$, and thus Theorem 9.4 implies (since uniform integrability of $\min(X^{(n)}, X^{(n+1)})/\sqrt{n}$ follows from Lemma 7.2, or indeed from Theorem 1.3)

$$\limsup_{n \rightarrow \infty} \mathbb{E}(\Upsilon^{(n)}/\sqrt{n}) \leq \mathbb{E}[\min(\zeta, \zeta')] = \int_0^\infty \mathbb{P}(\zeta > x)^2 dx = \frac{27\pi^{3/2}}{64\sqrt{2}}. \quad (9.27)$$

(We omit the final, straightforward calculation.) On the other hand, it follows easily from the results in Section 9.1 that

$$\liminf_{n \rightarrow \infty} \mathbb{E}(\Upsilon^{(n)}/\sqrt{n}) > 0. \quad (9.28)$$

Problem 9.7. What is the asymptotics of the number of common vertices $\Upsilon^{(n)} := |\hat{D}^{(n)} \cap \hat{D}^{(n+1)}|$? (I.e., the vertices that are descendants of both n and $n+1$.)

We conjecture that $\mathbb{E}(\Upsilon^{(n)}/\sqrt{n}) \rightarrow v$ for some constant $v > 0$. Show this! What is v ? What is the asymptotic distribution of $\Upsilon^{(n)}/\sqrt{n}$? (Assuming that it exists.)

10. SOME VARIATIONS

We consider here the two variations mentioned in the introduction, and show that the same results hold for them too.

10.1. Several roots. We may start with any given number $m \geq 1$ roots, and then add $n - m$ vertices with outdegree d recursively as above. (We assume $1 \leq m \leq n$.) Denote the resulting random d -dag by $D_{n,m}$, and let $\hat{D}_{n,m}$ be the subgraph consisting of all vertices and edges that can be reached from n .

Note that $D_{n,m}$ can be obtained from D_n by simply removing all edges between the roots, i.e., all edges within $[1, m]$. Consequently, $D_{n,m}$ and D_n have the same descendants in the interval $(m, n]$, and it follows that

$$|\hat{D}_n| - m < |\hat{D}_{n,m}| \leq |\hat{D}_n|. \quad (10.1)$$

Theorem 10.1. *If the process starts with $m = o(n^{(d-1)/d})$ roots, and we thus define $X^{(n)} := |\hat{D}_{n,m}|$, then the results in Theorems 1.3 and 1.4 still hold.*

Proof. An immediate consequence of (10.1). $\square$

We may also obtain results for larger m . For simplicity we consider only the case $d = 2$. Define, for $\mu > 0$ and $x > 0$,

$$\begin{aligned} \psi_\mu(x) &:= \int_0^\infty \frac{x}{x + \max\{t, \mu\}^2} dt = \int_\mu^\infty \frac{x}{x + t^2} dt + \frac{\mu x}{x + \mu^2} \\ &= \sqrt{x} \arctan \frac{\sqrt{x}}{\mu} + \frac{\mu x}{x + \mu^2}. \end{aligned} \quad (10.2)$$

Theorem 10.2. *Let $d = 2$. Suppose that $m = m_n \rightarrow \infty$ such that $m/\sqrt{n} \rightarrow \mu \in (0, \infty)$. Then*

$$\frac{|\hat{D}_{n,m}|}{\sqrt{n}} \xrightarrow{d} \psi_\mu(\xi), \quad (10.3)$$

with convergence of all moments, where $\xi \in \Gamma(2)$. Moreover,

$$\frac{|\hat{D}_{n,m}|}{\sqrt{n}} - \psi_\mu(\Xi^{(n)}) \xrightarrow{p} 0. \quad (10.4)$$

Proof. The number of non-roots in $\hat{D}_{n,m}$ is, using the notation (9.1),

$$|\hat{D}_{n,m} \cap (m, n]| = |\hat{D}_n \cap (m, n]| = X_{m/\sqrt{n}, \infty}^{(n)}, \quad (10.5)$$

and thus it follows from Lemma 9.1 that

$$\frac{|\hat{D}_{n,m} \cap (m, n]|}{\sqrt{n}} - \int_\mu^\infty \frac{\Xi^{(n)}}{\Xi^{(n)} + t^2} dt \xrightarrow{L^1} 0. \quad (10.6)$$

Let $R_{n,m} := |\hat{D}_n \cap [1, m]|$ be the number of roots that are descendants of n . When the procedure in Section 2 reaches m , there are Y_m edges left. Each of these selects an endpoint in $\{1, \dots, m\}$ at random, uniformly and independently, and $R_{n,m}$ is the number of vertices in $\{1, \dots, m\}$ that are selected at least once. (This is a classical occupancy problem, often described as throwing Y_m balls into m cells.)

Conditioned on $\mathcal{F}_m$, each vertex $k \leq m$ thus has the same probability $\mathbb{E} J_k = 1 - (1 - \frac{1}{m})^{Y_m}$ of becoming red. The covariances can easily be calculated, but we note instead that if we also condition on $J_k = 0$, this increases the probability that $J_\ell = 1$ for every $\ell \neq k$; thus $\text{Cov}(J_k, J_\ell | \mathcal{F}_m) \leq 0$, and

$$\begin{aligned} \text{Var}(R_{n,m} | \mathcal{F}_m) &= \sum_{k, \ell=1}^m \text{Cov}(J_k, J_\ell | \mathcal{F}_m) \leq \sum_{k=1}^m \text{Var}(J_k | \mathcal{F}_m) = m \text{Var}(J_m | \mathcal{F}_m) \\ &\leq m \mathbb{E}(J_m | \mathcal{F}_m) \leq m \mathbb{E}(Z_m | \mathcal{F}_m) = m \frac{Y_m}{m} = Y_m. \end{aligned} \quad (10.7)$$

Hence, recalling (2.6) and (2.13),

$$\mathbb{E} \left(\frac{R_{n,m} - \mathbb{E}(R_{n,m} | \mathcal{F}_m)}{\sqrt{n}} \right)^2 = \frac{\mathbb{E}[\text{Var}(R_{n,m} | \mathcal{F}_m)]}{n} \leq \frac{\mathbb{E} Y_m}{n} \leq \frac{\mathbb{E} W_m}{mn} \leq \frac{2}{m} \rightarrow 0. \quad (10.8)$$

Consequently,

$$\frac{R_{n,m} - \mathbb{E}(R_{n,m} | \mathcal{F}_m)}{\sqrt{n}} \xrightarrow{p} 0. \quad (10.9)$$

Furthermore, by symmetry and (9.11),

$$\mathbb{E}(R_{n,m} | \mathcal{F}_m) = m \mathbb{E}(J_m | \mathcal{F}_m) = m \frac{\Xi^{(n)}}{\Xi^{(n)} + m^2/n} + o_p(m), \quad (10.10)$$

where $o_p(m)$ is a (random) quantity such that $o_p(m)/m \xrightarrow{p} 0$. It follows from (10.9) and (10.10) that

$$\frac{R_{n,m}}{\sqrt{n}} - \mu \frac{\Xi^{(n)}}{\Xi^{(n)} + \mu^2} \xrightarrow{p} 0. \quad (10.11)$$

We obtain (10.4) by summing (10.6) and (10.11), and this implies (10.3) by (3.11). Finally, moment convergence follows, since every power is uniformly integrable by $|\hat{D}_{n,m}| \leq X^{(n)}$ and Lemma 7.2. $\square$

It is possible to obtain results also for the case $m/\sqrt{n} \rightarrow \infty$ by our methods, but we leave this case to the reader.

10.2. Drawing without replacement. Consider now the case when the endpoints of the d edges from a vertex k are selected by drawing without replacement; in other words, the endpoints form a uniformly random subset of $\{1, \dots, k-1\}$ with d elements. (We start with $m \geq d$ roots.) Thus there are no multiple edges and D_n is a simple multigraph.

The analysis in Section 2 is based on the independence of the endpoints of different edges; this is no longer true since edges from the same vertex now are dependent. However, a minor variation of the arguments allows us to reach the same conclusions. For simplicity, we consider again the case $d = 2$, and leave the straightforward generalization to higher d to the reader. We use the same notations as above, with the additions below.

Say that the two edges starting together from a red vertex are *twins*. We thus now do not allow two twins to have the same endpoint.

Consider the Y_k red edges that cross the gap between $k+1$ and k . Some of these come in pairs of twins, while others are single (because their twin has already found an endpoint). Let $Y_{k,1}$ be the number of single edges, and $Y_{k,2}$ the number of pairs of twins among these edges. Thus

$$Y_k = Y_{k,1} + 2Y_{k,2}. \quad (10.12)$$

Similarly, let $Z_{k,1}$ be the number of single edges that end at k , and let $Z_{k,2}$ be the number of edges that end in k and still having a living twin (that will later find an endpoint $\ell < k$). Thus

$$Z_k = Z_{k,1} + Z_{k,2}. \quad (10.13)$$

We still have (2.2), but also the more detailed recursion

$$Y_{k-1,1} = Y_{k,1} - Z_{k,1} + Z_{k,2}, \quad (10.14)$$

$$Y_{k-1,2} = Y_{k,2} - Z_{k,2} + J_k = Y_{k,2} - Z_{k,2} + \mathbf{1}\{Z_{k,1} + Z_{k,2} \geq 1\}. \quad (10.15)$$

Each of the $Y_{k,1}$ single edges ends at k with probability $1/k$, and each of the $Y_{k,2}$ pairs of twins has one edge ending at k (and thus leaving one single edge) with probability $2/k$. Hence, conditioned on $\mathcal{F}_k$, we now have

$$Z_{k,1} \in \text{Bin}(Y_{k,1}, 1/k), \quad (10.16)$$

$$Z_{k,2} \in \text{Bin}(Y_{k,2}, 2/k), \quad (10.17)$$

with $Z_{k,1}$ and $Z_{k,2}$ (conditionally) independent.

Taking conditional expectations yields, instead of (2.4),

$$\mathbb{E}(Y_{k-1,1} \mid \mathcal{F}_k) = Y_{k,1} - \frac{1}{k}Y_{k,1} + \frac{2}{k}Y_{k,2}, \quad (10.18)$$

$$\mathbb{E}(Y_{k-1,2} \mid \mathcal{F}_k) = Y_{k,2} - \frac{2}{k}Y_{k,2} + \mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k). \quad (10.19)$$

Thus, using (10.12),

$$\begin{aligned}\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) &= Y_k - \frac{1}{k}Y_k + 2\mathbb{P}(Z_k \geq 1 \mid \mathcal{F}_k) \\ &= Y_k - \frac{1}{k}Y_k + 2\left(1 - \left(1 - \frac{1}{k}\right)^{Y_{k,1}} \left(1 - \frac{2}{k}\right)^{Y_{k,2}}\right)\end{aligned}\quad (10.20)$$

and also, exactly as in (2.5),

$$\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) = Y_k - \frac{1}{k}Y_k + 2\mathbb{P}(Z_k \geq 1) \leq \frac{k-1}{k}Y_k + 2\mathbb{E}Z_k = \frac{k+1}{k}Y_k. \quad (10.21)$$

Thus W_k is still a reverse supermartingale, M_k is a reverse martingale, and A_k is reverse increasing; (2.6)–(2.13) hold without any changes. The exact formulas in (2.14) and (2.15) are replaced by

$$\mathbb{E}(W_{k-1} \mid \mathcal{F}_k) = (k-1)Y_k + 2k\left(1 - \left(1 - \frac{1}{k}\right)^{Y_{k,1}} \left(1 - \frac{2}{k}\right)^{Y_{k,2}}\right) \quad (10.22)$$

and thus

$$A_{k-1} - A_k = W_k - \mathbb{E}(W_{k-1} \mid \mathcal{F}_k) = 2k\left(\left(1 - \frac{1}{k}\right)^{Y_{k,1}} \left(1 - \frac{2}{k}\right)^{Y_{k,2}} - 1 + \frac{Y_k}{k}\right). \quad (10.23)$$

The rest of Section 2 holds with minor changes: the numerical constants in inequalities may change (perhaps including cases where we had constant 1), we estimate (conditional) variances of $Z_{k,1}$ and $Z_{k,2}$ separately in (2.16), the exact formula in (2.28) is modified as above, and the equality in (2.25) is modified; we omit the details.

In Section 3, we note that for the version studied in the previous sections, the probability that two twins starting at k have the same endpoint is $1/(k-1)$. Hence, the expected number of such collisions among twins starting at $k \geq n_1$ is (with $J_n := 1$), using (2.22),

$$\sum_{k=n_1}^n \mathbb{P}(J_k = 1) \frac{1}{k-1} \leq \frac{1}{n-1} + \sum_{k=n_1}^{n-1} \frac{2n}{k^2(k-1)} \leq \frac{1}{n-1} + \frac{3n}{n_1^2} = o(1). \quad (10.24)$$

Thus, w.h.p. there are no such collisions, which means that we may couple the versions using drawing with and without replacement such that they w.h.p. coincide on the interval $[n_1, n]$. Consequently, Theorem 3.1 giving a coupling with the Yule process holds also for drawing without replacement.

The results in Sections 3–7 now hold as before, with some numerical constants changed and a few minor changes. The most important is that (5.14) now, by (10.23), becomes

$$\frac{d}{dt}\hat{A}_t^{(n)} = -n^{-1/2} \cdot 2k\left(\left(1 - \frac{1}{k}\right)^{Y_{k,1}} \left(1 - \frac{2}{k}\right)^{Y_{k,2}} - 1 + \frac{Y_k}{k}\right), \quad (10.25)$$

but it is easily seen that this together with (5.13) still yields (5.15), since

$$\begin{aligned}\log\left(\left(1 - \frac{1}{k}\right)^{Y_{k,1}} \left(1 - \frac{2}{k}\right)^{Y_{k,2}}\right) &= -Y_{k,1} \cdot \frac{1}{k} - Y_{k,2} \cdot \frac{2}{k} + O\left(Y_k \cdot \frac{1}{k^2}\right) \\ &= -\frac{Y_k}{k} + o(1).\end{aligned}\quad (10.26)$$

There is a similar modification in (6.13), but again the conclusion (6.14) holds by (10.26). In Section 7, we argue as in (7.10) for $Z_{k,1}$ and $Z_{k,2}$ separately.

Hence, Theorem 1.3 holds also for drawing without replacement. (And so does Theorem 1.4, by similar arguments.)

ACKNOWLEDGEMENT

I thank Donald Knuth and Philippe Jacquet for drawing my attention to this problem.

REFERENCES

- [1] Sunil Arya, Mordecai J Golin & Kurt Mehlhorn. On the expected depth of random circuits. *Combin. Probab. Comput.* **8** (1999), no. 3, 209–227.
- [2] Krishna B. Athreya & Peter E. Ney. *Branching Processes*. Springer-Verlag, Berlin, 1972.
- [3] Albert-László Barabási & Réka Albert. Emergence of scaling in random networks. *Science* **286** (1999), no. 5439, 509–512.
- [4] Patrick Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [5] Béla Bollobás, Oliver Riordan, Joel Spencer and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** (2001), no. 3, 279–290.
- [6] Nicolas Broutin & Omar Fawzi. Longest path distance in random circuits. *Combin. Probab. Comput.* **21** (2012), no. 6, 856–881.
- [7] Donald L. Burkholder. Distribution function inequalities for martingales. *Ann. Probability* **1** (1973), 19–42.
- [8] Luc Devroye & Svante Janson. Long and short paths in uniform random recursive dags. *Arkiv för Matematik* **49** (2011), 61–77.
- [9] Luc Devroye & Jiang Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures Algorithms* **7** (1995), no. 1, 1–14.
- [10] Josep Díaz, Maria J. Serna, Paul Spirakis & Jacobo Torán: On the expected depth of Boolean circuits. Technical Report LSI-94-7-R, Univ. Politéc. de Catalunya, Dept. LSI (1994).
- [11] Allan Gut. *Probability: A Graduate Course*, 2nd edn. Springer, New York, 2013.
- [12] Olav Kallenberg. *Foundations of Modern Probability*, 2nd edn. Springer, New York, 2002.
- [13] Donald E. Knuth. *The Art of Computer Programming*, Section 7.2.2.3 (Preliminary draft, 29 January 2023).
<http://cs.stanford.edu/~knuth/fasc7a.ps.gz>
- [14] Markus Kuba & Henning Sulzbach. On martingale tail sums in affine two-color urn models with multiple drawings. *J. Appl. Probab.* **54** (2017), no. 1, 96–117.
- [15] Hosam Mahmoud & Tatsuie Tsukiji. On the internal structure of random recursive circuits. *J. Comput. Appl. Math.* **142** (2002), no. 1, 155–171.
- [16] J. A. Moler, F. Plo, M. San Miguel & H. Urmeneta. Asymptotics in random recursive circuits. *Journal of Mathematical Sciences* **196** (2014), 70–74.
- [17] *NIST Handbook of Mathematical Functions*. Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert and Charles W. Clark. Cambridge Univ. Press, 2010.
Also available as *NIST Digital Library of Mathematical Functions*,
<http://dlmf.nist.gov/>
- [18] T. Tsukiji & H. Mahmoud. A limit law for outputs in random recursive circuits. *Algorithmica* **31** (2001), 403–412. MR1855257

- [19] Tatsuie Tsukiji & Fatos Xhafa. On the depth of randomly generated circuits.
In *Algorithms – ESA ’96*, 208–220, Lecture Notes in Computer Science **1136**,
Springer, Berlin, 1996.

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/`