FASCICULI MATHEMATICI

Nr 28 1998

DEDICATED TO PROFESSOR DOBIESŁAW BOBROWSKI
ON THE OCCASION OF HIS 70^{th} BIRTHDAY

MARIAN DONDAJEWSKI* AND JERZY SZYMAŃSKI**

VERTICES OF SMALL DEGREES IN RANDOM RECURSIVE DAGS

ABSTRACT. In this paper random recursive dags (directed acyclic graphs) are considered. For such combinatorial structures the expected number of vertices of small outdegrees as well as the degree of a given vertex are studied.

KEY WORDS. directed acyclic graph, random graphs, recursive tree, vertex degree.

1. Introduction

A *uniform random recursive directed acyclic graph (dag)* on n vertices is a graph recursively constructed by picking by the i-th vertex r other vertices (*parents*) uniformly among the first $i-1$ vertices, for $i = m+1,...,n$ (compare [4]). It is assumed that $r \leq m \leq n$. So, the first m vertices are *roots* of the dag. See [1] or [3] for definitions not given here.

Note that for $m = r = 1$ a random recursive dag is a random recursive tree. So results for recursive dags generalize corresponding results for random recursive trees, which can be found in [2] and [5–9].

The *outdegree* of a vertex is a number of its children. The number of parents (*indegree*) is r for non-root vertices, and 0 for the roots.

Let X_{nk} denote the number of vertices of outdegree k in a random recursive dag with n vertices. A *leaf* of a dag is a vertex with outdegree 0. Let $L_n = X_{n,0}$ denote the number of leaves in a recursive dag with n vertices. Moreover, let d_{ni} denote the outdegree of a vertex i in a uniform random recursive dag. The object of this note is to study the distribution of random variables L_n, X_{nk} and d_{ni}.

The following notation will be used in this paper:

- $\zeta_n(z) = \sum_{k=1}^{n} k^{-z}$ for the incomplete Riemann zeta function,
- $H_n = \zeta_n(1) = 1 + \frac{1}{2} + ... \frac{1}{n}$ for the harmonic number, and
- $(x)_n = x(x-1)...(x-n+1)$ for the falling factorial.

Moreover, for a sequence of random variables we will write $\xrightarrow{D}$ and $\xrightarrow{P}$ for the convergence in distribution and in probability, respectively.

2. Number of leaves

Let L_n denote the number of vertices of outdegree zero in a random recursive dag on n vertices. The way in which a random recursive dag is constructed implies that a random recursive dag on m vertices has m leaves, while a random recursive dag on $m+1$ vertices has $m-r+1$ leaves.

Theorem 1. The expected number of leaves L_n in a random recursive dag on n vertices satisfies the following recurrence relation

$$E[L_{n+1}] = \frac{n-r}{n} E[L_n] + 1$$

for $r \le m$, with the initial condition

$$E[L_m] = m.$$

Proof. After adding to the recursive dag a vertex with label $n+1$, one more leaf (namely vertex $n+1$) in the dag occurs. Moreover, the probability that k leaves increase their outdegrees ($1 \le k \le r$) is equal to

$$\frac{\binom{L_n}{k}\binom{n-L_n}{r-k}}{\binom{n}{r}}.$$

Therefore by a well-known property of binomial coefficients

$$\sum_{k=0}^{r} k \frac{\binom{m}{k}\binom{n-m}{r-k}}{\binom{n}{r}} = \frac{mr}{n}$$

we obtain

$$E[L_{n+1}] = E[L_n] + 1 - \sum_{k=0}^{r} k \frac{\binom{E[L_n]}{k}\binom{n-E[L_n]}{r-k}}{\binom{n}{r}} =$$

$$= E[L_n] + 1 - \frac{rE[L_n]}{n} =$$

$$= \frac{n-r}{n} E[L_n] + 1,$$

which completes the proof. ■

Theorem 2. The expected number of leaves L_n in a random recursive dag on n vertices equals

$$E[L_n]=\frac{m}{r+1}+\frac{mr}{r+1}\prod_{j=m}^{n-1}\frac{j-r}{j}.$$

Proof. It is enough to prove that the recurrence relation from Theorem 1 as well as the initial condition $E[L_m]=m$ are fulfilled. As a matter of fact

$$E[L_n]=\frac{m}{r+1}+\frac{mr}{r+1}\prod_{j=m}^{m-1}\frac{j-r}{j}=$$

$$=\frac{m}{r+1}+\frac{mr}{r+1}=m,$$

because an empty product is equal to 1. Furthermore, one can also verify that

$$E[L_m]=\frac{m+1}{r+1}+\frac{mr}{r+1}\prod_{j=m}^{m}\frac{j-r}{j}=\frac{m+1}{r+1}+\frac{mr}{r+1}\frac{m-r}{r+1}=$$

$$=\frac{m+1+r(m-r)}{r+1}=m-r+1.$$

Now, let us show that the recurrence relation from Theorem 1 holds as well. We have

$$E[L_{n+1}]-\frac{n-r}{n}E[L_n]=$$

$$=\frac{n+1}{r+1}+\frac{mr}{r+1}\prod_{j=m}^{n}\frac{j-r}{j}-\frac{n-r}{n}\left(\frac{n}{r+1}+\frac{mr}{r+1}\prod_{j=m}^{n-1}\frac{j-r}{j}\right)=$$

$$=\frac{n+1}{r+1}+\frac{mr}{r+1}\frac{n-r}{n}\prod_{j=m}^{n-1}\frac{j-r}{j}-\frac{n-r}{n+1}-\frac{mr}{r+1}\frac{n-r}{n}\prod_{j=m}^{n-1}\frac{j-r}{j}=1,$$

which completes the proof. ■

In a case when $m=r=1$ (i.e. in a case of a *uniform random recursive tree*) Theorem 2 provides already known result (compare [6] or [7]).

Corollary 1. In a random recursive tree on n vertices the expected number of leaves L_n equals $E[L_n]=\frac{n}{2}$. ■

The case $m=r$ leads to

Corollary 2. In a random recursive dag with $m = r$ the expected number of leaves L_n equals $E[L_n] = \frac{n}{r+1}$. ■

An asymptotic behavior of $E[L_n]$ if characterized by the following result.

Corollary 3. In a random recursive dag with fixed m and r, the expected number of leaves L_n has the following asymptotic behavior

$$E[L_n] = \frac{n}{r+1} + o(1)$$

as $n \to \infty$.

Proof. Theorem 2 implies that $E[L_n] = \frac{n}{r+1} + x_n$ where

$$x_n = \frac{mr}{r+1} \prod_{j=m}^{n-1} \frac{j-r}{j}.$$

So for large enough n

$$x_n = \frac{mr}{r+1} \frac{(m-1)_r}{(n-1)_r} = O(n^{-r}) = o(1). \quad ■$$

Now, let us consider *random recursive forests* with exactly m trees. We begin with a case when $r = 1$.

Corollary 4. In a random recursive forest with m trees (components) the expected number of leaves L_n is

$$E[L_n] = \frac{n}{2} + \frac{m(m-1)}{2(n-1)}.$$

Proof. A random recursive forest on m trees is a random recursive dag with $r = 1$. Consequently by Theorem 2

$$E[L_n] = \frac{n}{2} + \frac{m}{2} \prod_{j=m}^{n-1} \frac{j-1}{j} =$$

$$= \frac{n}{2} + \frac{m}{2} \frac{m-1}{n-1},$$

which completes the proof. ■

Now, let us consider a case when $r = 2$.

Corollary 5. In a case when $r = 2$ the expected number of leaves L_n equals

$$E[L_n] = \frac{n}{3} + \frac{2m(m-1)(m-2)}{3(n-1)(n-2)}.$$

Proof. Substituting $r = 2$ into Theorem 2 we get

$$E[L_n] = \frac{n}{3} + \frac{2m}{3} \prod_{j=m}^{n-1} \frac{j-2}{j} =$$
$$= \frac{n}{3} + \frac{2m}{3} \frac{(m-1)(m-2)}{(n-1)(n-2)}.$$

3. NUMBER OF VERTICES OF A FIXED DEGREE

Let X_{nk} denote the number of vertices of outdegree k in a random recursive dag. So, the number of leaves $L_n = X_{n,0}$. From the way in which random recursive dag is constructed, one can easily see that a random recursive dag with $m + k - 1$ vertices has 0 vertices of outdegree k.

Theorem 3. The expected number of vertices of outdegree k in a random recursive dag with n vertices satisfies the following recurrence relation

$$E[X_{n+1,k}] = \frac{n-r}{n} E[X_{nk}] + \frac{r}{n} E[X_{n,k-1}].$$

Proof. After adding to the recursive dag a vertex with label $n+1$ the probability that j new vertices of degree k occur is $\binom{X_{n,k-1}}{j}\binom{n-X_{n,j-1}}{r-j}\Big/\binom{n}{r}$. Moreover, the probability that j vertices of outdegree k increase their outdegrees is $\binom{X_{nk}}{j}\binom{n-X_{nk}}{r-j}\Big/\binom{n}{r}$. So, by the property of binomial coefficients (see proof of Theorem 1)

$$E[X_{n+1,k}] = E[X_{nk}] + \sum_{j=0}^{r} j \frac{\binom{X_{n,k-1}}{j}\binom{n-X_{n,j-1}}{r-j}}{\binom{n}{r}} - \sum_{j=0}^{r} j \frac{\binom{X_{nk}}{j}\binom{n-X_{nk}}{r-j}}{\binom{n}{r}} =$$
$$= E[X_{nk}] + \frac{r\,E[X_{n,k-1}]}{n} - \frac{r\,E[X_{nk}]}{n} =$$
$$= \frac{n-r}{n} E[X_{nk}] + \frac{r}{n} E[X_{n,k-1}],$$

and the proof is completed. ■

We find asymptotic solution of the recurrence equation from Theorem 3 for fixed values of m and r.

Theorem 4. If m and r are fixed and $n \to \infty$ then

$$E[X_{nk}] = \left(\frac{r}{r+1}\right)^k \frac{n}{r+1} + o(1).$$

Proof. Assume that $E[X_{nk}] = \alpha_k n + o(1)$, where α_k is fixed. Then from Theorem 3 one can obtain

$$\alpha_k(n+1) + o(1) = \frac{n-r}{n}(\alpha_k n + o(1)) + \frac{r}{n}(\alpha_{k-1} n + o(1))$$

or equivalently

$$\alpha_k(r+1) - \alpha_{k-1} r = o(1).$$

The left-hand side of the last relation does not depend on n, so the right-hand side has to be zero. This leads to a recurrence relation

$$(r+1)\alpha_k = r\alpha_{k-1}$$

with the initial condition $\alpha_0 = \frac{1}{r+1}$ (compare with Corollary 3). It has solution $\alpha_k = \frac{r^k}{(r+1)^{k+1}}$. From the uniqueness of the solution of the recurrence equation we get our result. ■

Notice that in a case when $r = m = 1$ we obtain a known result for the random recursive tree, namely $E[X_k] = \frac{n}{2^{k+1}}$.

4. DEGREES OF ROOTS

From the construction of a random recursive dag, one can easily see that random variables $d_{n,1}, d_{n,2}, \ldots, d_{n,m}$ (the degrees of the roots) are identically distributed. So let ρ_n denote the degree of a root of random recursive dag. Then the following result holds.

Theorem 5. In a random recursive dag with n vertices the degree ρ_n of the root satisfies

$$E[\rho_n] = r(H_{n-1} - H_{m-1})$$

and

$$Var[\rho_n] = r(H_{n-1} - H_{m-1}) - r^2(\zeta_{n-1}(2) - \zeta_{m-1}(2)).$$

Moreover, if $n \to \infty$ for fixed m and r then

$$E[\rho_n] \sim r \log n$$

and

$$Var[\rho_n] \sim r \log n.$$

The limiting distribution of ρ_n satisfies

$$\frac{\rho_n}{E[\rho_n]} \xrightarrow{P} 1$$

and

$$\frac{\rho_n - E[\rho_n]}{\sqrt{Var[\rho_n]}} \xrightarrow{D} N(0,1)$$

To prove the above theorem we need the following result.

Lemma 1. If H_k is the k-th harmonic number and $\zeta_n(z)$ is the incomplete Riemann zeta function then for $n \geq 1$

$$\sum_{k=1}^{n} \frac{H_k}{k} = \frac{1}{2}(H_n^2 + \zeta_n(2)).$$

Proof. Let

$$a_n = \sum_{k=1}^{n} \frac{H_k}{k}$$

and

$$b_n = \frac{1}{2}(H_n^2 + \zeta_n(2)).$$

Then a_n can be defined as a solution of the following recurrence problem

$$a_1 = 1,$$

$$\Delta a_n = \frac{H_{n+1}}{n+1}.$$

It is easy to check that $b_1 = 1$. Now let us compute Δb_n:

$$\begin{aligned}\Delta b_n &= b_{n+1} - b_n = \\ &= \tfrac{1}{2}(H_{n+1}^2 - H_n^2 + \zeta_{n+1}(2) - \zeta_n(2)) = \\ &= \frac{1}{2}\left(H_{n+1}^2 - \left(H_{n+1} - \frac{1}{n+1}\right)^2 + \frac{1}{(n+1)^2}\right) = \\ &= \frac{1}{2}\left(\frac{2H_{n+1}}{n+1}\right) = \frac{H_{n+1}}{n+1}.\end{aligned}$$

This implies that, b_n fulfills the recurrence equation for a_n and by the uniqueness of the solution we obtain that $a_n = b_n$ for all natural n. ■

Proof of the Theorem 5. Let

$$R_i = \begin{cases} 1 & \text{if } i \to 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $i \to j$ means that vertex i chooses the vertex j as one of its parents. From the construction of a random dag we obtain

$$E[\rho_n] = \sum_{i=m+1}^{n} P(R_i = 1),$$

on the other hand

$$P(R_i = 1) = E[R_i] = \frac{\binom{i-2}{r-1}}{\binom{i-1}{r}} = \frac{r}{i-1}.$$

Therefore,

$$E[\rho_n] = \sum_{i=m+1}^{n} \frac{r}{i-1} = r(H_{n-1} - H_{m-1}).$$

Now, because

$$Var[R_i] = \frac{r}{i-1} - \left(\frac{r}{i-1}\right)^2$$

the variance of ρ_n is equal to

$$Var[\rho_n] = \sum_{i=m+1}^{n} Var[R_i] =$$

$$= E[\rho_n] - r^2 \sum_{i=m+1}^{n} \frac{1}{(i-1)^2} =$$

$$= r(H_{n-1} - H_{m-1}) - r^2(\zeta_{n-1}(2) - \zeta_{m-1}(2)).$$

Consequently the limiting distribution of ρ_n follows from the Chebyshev's inequality and the Lindeberg-Feller's theorem. ■

In a case when $m = r = 1$ (i.e. in a case of a uniform random recursive tree) Theorem 5 provides a well-known result, summarized in the following corollary.

Corollary 6. In a random recursive tree on n vertices, the root degree d_{n1} satisfies

$$E[d_{n1}] = H_{n-1}$$

and

$$Var[d_{n1}] = H_{n-1} - \zeta_{n-1}(2).$$

Moreover, if $n \to \infty$ then

$$E[d_{n1}] \sim \log n$$

and

$$Var[d_{n1}] \sim \log n.$$

Furthermore, the limiting distribution of d_{n1} satisfies

$$\frac{d_{n1}}{E[d_{n1}]} \xrightarrow{P} 1$$

and

$$\frac{d_{n1} - E[d_{n1}]}{\sqrt{Var[d_{n1}]}} \xrightarrow{D} N(0,1). \quad \blacksquare$$

5. Degree of non-root vertices

In this section a degrees of non-root vertices are considered.

Theorem 6. In a random recursive dag on n vertices, the outdegree d_{n1} of i-th vertex satisfies $(m < i \le n)$

$$E[d_{n1}] = r(H_{n-1} - H_{i-1})$$

and

$$Var[d_{n1}] = r(H_{n-1} - H_{i-1}) - r^2(\zeta_{n-1}(2) - \zeta_{i-1}(2)).$$

Moreover, for fixed *m, r* and $i = o(n)$

$$E[d_{ni}] \sim r \log n$$

and

$$Var[d_{ni}] \sim r \log n.$$

Furthermore, the limiting distribution of d_{ni} satisfies

$$\frac{d_{ni}}{E[d_{ni}]} \xrightarrow{P} 1$$

and

$$\frac{d_{ni} - E[d_{ni}]}{\sqrt{Var[d_{ni}]}} \xrightarrow{D} N(0,1)$$

Proof. Let

$$R_{ji} = \begin{cases} 1 & \text{if } j \to i, \\ 0 & \text{otherwise,} \end{cases}$$

where, as before $j \to i$ means that a vertex j chooses vertex i as one of its parents. A construction of a random dag implies that

$$E[d_{ni}] = \sum_{j=i+1}^{n} P(R_{ji} = 1).$$

Similarly as in the proof of a Theorem 5 one can obtain that

$$P(R_{ji} = 1) = E[R_{ji}] = \frac{\binom{j-2}{r-1}}{\binom{j-1}{r}} = \frac{r}{j-1}.$$

Consequently,

$$E[d_{ni}] = \sum_{j=i+1}^{n} \frac{r}{j-1} = r(H_{n-1} - H_{i-1}).$$

Now, because

$$Var[R_{ji}] = \frac{r}{j-1} - \left(\frac{r}{j-1}\right)^2$$

the variance of d_{ni} is equal to

$$Var[d_{ni}] = \sum_{j=i+1}^{n} Var[R_{ji}] =$$

$$= E[d_{ni}] - r^2 \sum_{j=i+1}^{n} \frac{1}{(j-1)^2} =$$

$$= r(H_{n-1} - H_{i-1}) - r^2(\zeta_{n-1}(2) - \zeta_{i-1}(2)).$$

The limiting distribution of d_{ni} follows from the Chebyshev's inequality and the Lindeberg-Feller's theorem. ■

In a case when $m = r = 1$ Theorem 6 provides a well-known result, summarized in the following corollary.

Corollary 7. In a random recursive tree on n vertices the degree d_{ni} of i-th vertex satisfies

$$E[d_{ni}] = H_{n-1} - H_{i-1}$$

and

$$Var[d_{ni}] = H_{n-1} - H_{i-1} - \zeta_{n-1}(2) + \zeta_{i-1}(2).$$

Moreover, if i is fixed and $n \to \infty$ then

$$E[d_{ni}] \sim \log n$$

and

$$Var[d_{ni}] \sim \log n.$$

The limiting distribution of d_{ni} satisfies

$$\frac{d_{ni}}{E[d_{ni}]} \xrightarrow{P} 1$$

and

$$\frac{d_{ni} - E[d_{ni}]}{\sqrt{Var[d_{ni}]}} \xrightarrow{D} N(0,1).$$

REFERENCES

[1] C. Berge, *Graphs and Hypergraphs*, North-Holland, 1973.

[2] F. Bergeron, P. Flajolet, B. Salvy, *Varieties of increasing trees*, INRIA Research Report (1992) .

[3] B. Bollobás, *Random Graphs*, Academic Press, New York, 1985.

[4] L. Devroye, J. Lu, *The strong convergence of maximal degrees in uniform random recursive trees and dags*, Random Structures & Algorithms (to appear).

[5] Meir, J.W. Moon, Recursive trees with no nodes of out-degree one, *Congressus Numerantium*, 66(1988), 49-62.

[6] H. Na, A. Rapoport, Distribution of nodes of a tree by degree, *Math. Biosciences*, 6(1970) 313-329.

[7] D. Najock, C. Heyde, On the number of terminal vertices in certain random trees with an applications to stemma construction in philology, *J. Applied Prob.*, 19(1982), 675-680.

[8] J. Szymański, On the maximum degree and the height of a random recursive tree, *Random Graphs*'87 (M. Karoński, J. Jaworski and A. Ruciński, eds.), John Wiley & Sons, 1990, 313-324.

[9] J. Szymański, *On the complexity of algorithms on recursive* trees, *Theoretical Computer Science* 74(1990), 355-361.

(* Institute of Mathematics, Poznan University of Thechnology, 60-965 Poznań, Poland, *E-mail address*: dondajew@math.put.poznan.pl

**Department of Discrete Mathematics, Adam Mickiewicz University, 60-769 Poznań, Poland, *E-mail address*: jszymans@math.amu.edu.pl)

Received on 05.12.1997 and, in revised form, on 12.05.1998.

Research supported by grant KBN 2 P03A 023 09