The root cluster after percolation on preferential attachment trees

(Extended abstract)

Colin Desmarais[*]   Cecilia Holmgren[*]   Stephan Wagner[*,†]

Abstract

The class of linear preferential attachment trees includes recursive trees, plane-oriented recursive trees, binary search trees, and increasing d-ary trees. Bond percolation with parameter p is performed by considering every edge in a graph independently, and either keeping the edge with probability p or removing it otherwise. The resulting connected components are called clusters. In this extended abstract, we demonstrate how to use methods from analytic combinatorics to compute limiting distributions, after rescaling, for the size of the cluster containing the root. These results are part of a larger work on broadcasting induced colorings of preferential attachment trees.

1 Introduction

For a real number α, a sequence of random (linear) preferential attachment trees $(\mathcal{T}_n)_{n=1}^{\infty}$ is constructed in the following manner. The tree $\mathcal{T}_1$ consists of a single vertex labelled 1. For $n \geq 1$, a vertex v is chosen from $\mathcal{T}_n$ with probability

$$\frac{\alpha \deg^+(v) + 1}{\sum_{u \in V(\mathcal{T}_n)}(\alpha \deg^+(u) + 1)} = \frac{\alpha \deg^+(v) + 1}{\alpha(n-1) + n}, \tag{1}$$

---
[*]Department of Mathematics, Uppsala University, Uppsala, Sweden. E-mail: colindesmarais@gmail.com, {cecilia.holmgren, stephan.wagner} @math.uu.se

[†]Department of Mathematical Sciences, Stellenbosch University, Stellenbosch, South Africa

where $\deg^+(u)$ (called the *outdegree* of u) is the number of children of u. A vertex labelled $n+1$ is then added as a child of v to construct $\mathcal{T}_{n+1}$. The equality above holds since $\sum_{u \in V(\mathcal{T}_n)} \deg^+(u) = n-1$, the number of edges in the tree $\mathcal{T}_n$. The parameter α is restricted to $\{\ldots, -\frac{1}{4}, -\frac{1}{3}, -\frac{1}{2}\} \bigcup [0, \infty)$ to avoid the degenerate case $\alpha = -1$ that produces a path, and to avoid the cases where (1) might be negative.

We then perform bond percolation with parameter p on preferential attachment trees. Each edge is considered independently and either kept (or remains *open*) with probability p or removed (*closed*) with probability $1 - p$. We then ask how many vertices remain in the component (or *cluster*) $\mathcal{C}_n$ that contains the root. For $\alpha \geq 0$, Baur proved that $|\mathcal{C}_n|/n^{(p+\alpha)/(1+\alpha)}$ converges in distribution to a random variable $\mathcal{C}$ and provided the first two moments [2]. In the special case $\alpha = 0$, it is already known that $\mathcal{C}$ is a Mittag-Leffler random variable [3, 4]. We extend these results by showing that $|\mathcal{C}_n|/n^{(p+\alpha)(1+\alpha)}$ converges in distribution for $p > -\alpha$ and provide a recursion for the moments. This recursion uses the partial Bell polynomials (see [5, Chapter 3.3])

$$B_{k,j}(x_1, \ldots, x_{k-j+1}) = \sum_{\substack{m_1 + \cdots + (k-j+1)m_{k-j+1}=k \\ m_1 + \cdots + m_{k-j+1}=j}} k! \prod_{i=1}^{k-j+1} \frac{x_i^{m_i}}{m_i! i!^{m_i}}.$$

Theorem 1.1. *Let $\alpha > 0$. Then $|\mathcal{C}_n|/n^{(p+\alpha)/(1+\alpha)} \xrightarrow{d} \mathcal{C}$, where $\mathcal{C}$ has integer moments*

$$\mathbb{E}[\mathcal{C}^k] = \frac{C_k(1+\alpha)\Gamma(1/(1+\alpha))}{\alpha\Gamma((kp + \alpha(k-1))/(\alpha+1))},$$

where C_k satisfies the recursion $C_1 = \alpha/(p+\alpha)$ and

$$(k-1)(p/\alpha+1)C_k = \sum_{j=2}^{k} \frac{p^j \Gamma(1/\alpha + j)}{\Gamma(1/\alpha)} B_{k,j}(C_1, \ldots, C_{k-j+1}). \tag{2}$$

When $\alpha = 1$ the above recursion is used to express the moments of $\mathcal{C}$ in a closed form

$$\mathbb{E}[\mathcal{C}^k] = \frac{2p^{k-1}\Gamma(kp+k-1)\sqrt{\pi}}{(p+1)^{2k-1}\Gamma(kp)\Gamma((kp+k-1)/2)}.$$

Theorem 1.2. *Let $\alpha = -1/d$, where $d \geq 2$ is a positive integer, and let $p > -\alpha$. Then $|\mathcal{C}_n|/n^{(pd-1)/(d-1)} \xrightarrow{d} \mathcal{C}$, where $\mathcal{C}$ has integer moments*

$$\mathbb{E}[\mathcal{C}^k] = \frac{D_k \Gamma(1/(d-1))}{\Gamma((kpd-k+1)/(d-1))},$$

where D_k satisfies the recursion $D_1 = 1/(pd-1)$ and

$$(k-1)(pd-1)D_k = \sum_{j=2}^{\min\{k,d\}} \frac{p^j d!}{(d-j)!} B_{k,j}(D_1, \ldots, D_{k-j+1}).$$

In the special case $\alpha = -1/2$ (so $\mathcal{T}_n$ is a random binary search tree) the recursion above produces the closed form

$$\mathbb{E}[\mathcal{C}^k] = \frac{k! p^{2(k-1)}}{(2p-1)^{2k-1} \Gamma(k(2p-1)+1)}.$$

When $p \leq -\alpha$, we are also able to prove that $|\mathcal{C}_n|$ is almost surely finite and converges to a Galton-Watson tree with binomial $\mathrm{Bin}(d,p)$ offspring distribution.

The results presented in this extended abstract are part of a larger work on broadcasting-induced colorings on preferential attachment trees [6]. A broadcasting process on a tree is performed by assigning the root either the bit 1 or 0 with equal probability. Starting with the children of the root, every other vertex takes the same bit as its parent with probability p and the other bit with probability $1-p$. The reconstruction problem is then to reconstruct the bit value of the root ρ from the bit values of some subset of vertices in T after broadcasting. Addario-Berry, Devroye, Lugosi, and Velona studied the reconstruction problem in random recursive trees and preferential attachment trees [1].

We then colour a vertex red if its bit value is 0 and blue if its bit value is 1 to obtain a broadcasting induced coloring. If we remove edges between vertices with different colours we are left with a forest of trees corresponding to clusters after performing bond percolation with parameter p. Along with results on the size of the root cluster, we also prove in [6] limiting distributions after rescaling for the number of vertices, clusters, and leaves of each colour, as well as the number of fringe subtrees with two-colorings.

2 Proof outline

Consider the function

$$\phi(\delta) = \begin{cases} 1 & \alpha = 0, \\ \frac{\Gamma(\delta+1/\alpha)}{\Gamma(1/\alpha)} & \alpha > 0, \\ \frac{d!}{(d-\delta)!} & \alpha = -\frac{1}{d}, d \in \mathbb{Z}^+. \end{cases} \qquad (3)$$

For a particular rooted labelled tree T on n vertices with increasing labels, define the weight of T to be $w(T) = \prod_{v \in V(T)} \phi(\deg^+(v))$. Then the probability that the recursive process described above produces T is given by $\mathbb{P}(\mathcal{T}_n = T) = w(T)/\sum_{T'} w(T')$, where the sum is taken over all rooted labelled trees T on n vertices with increasing labels. Letting b_n be the denominator $\sum_{T'} w(T')$, the exponential generating function for b_n is given by $B(x) = \sum_{n \geq 1} \frac{b_n}{n!} x^n = -\ln(1-x)$ when $\alpha = 0$, $B(x) = 1 - (1 - (1+1/\alpha)x)^{\frac{\alpha}{1+\alpha}}$ when $\alpha > 0$, and $B(x) = (1 - (d-1)x)^{-\frac{1}{d-1}} - 1$ when $\alpha = -\frac{1}{d}, d = 2, 3, 4, \ldots$. See [7] for detailed derivations of $B(x)$.

For a particular tree T with bond percolation performed, let $\mathcal{C}(T)$ be the cluster that contains the root and let $|\mathcal{C}(T)|$ be the number of vertices in $\mathcal{C}(T)$. Define $r_{n,k} = \sum_{T:|T|=n} w(T)\mathbb{P}(|\mathcal{C}(T)| = k)$. Then $\sum_{k=1}^{n} r_{n,k} = b_n$, and $\mathbb{P}(|\mathcal{C}_n| = k) = r_{n,k}/b_n$.

We develop a recursion for $r_{n,k}$. Take any tree T on n vertices with increasing labels whose root cluster after percolation has size k. Let δ be the outdegree of the root and

let $T_1, \cdots, T_\delta$ be the subtrees of T rooted at the children of the root such that the edges from the root to $T_1, \ldots, T_s$ are open and the edges from the root to $T_{s+1}, \ldots, T_\delta$ are closed. Then

$$w(T)\mathbb{P}(|\mathcal{C}(T)| = k) = \phi(\delta) \prod_{i=1}^{s} pw(T_i)\mathbb{P}(|\mathcal{C}(T_i)| = k_i) \prod_{j=s+1}^{\delta} (1-p)w(T_j),$$

where $k_1 + \cdots + k_s = k - 1$. If the trees $T_1, \ldots, T_\delta$ are of size $n_1, \ldots, n_\delta$ then $n_1 + \cdots + n_\delta = n - 1$. Summing over all such trees T on n vertices with root cluster of size k, we get the recursion

$$r_{n,k} = \sum_{\delta}^{n-1} \sum_{s=0}^{\delta} \binom{\delta}{s} \frac{\phi(\delta)}{\delta!} \sum_{n_1,\ldots,n_\delta} \binom{n-1}{n_1, \ldots, n_\delta} \sum_{k_1,\ldots,k_s} \prod_{i=1}^{s} pr_{n_i,k_i} \prod_{j=s+1}^{\delta} (1-p)b_{n_j}.$$

Let $R(x, u)$ be the bivariate (exponential) generating function for $r_{n,k}$, so $R(x, u) = \sum_{n,k\geq 1} = \frac{r_{n,k}}{n!} x^n u^k$. To use the moment of methods to prove that $|\mathcal{C}_n|$ converges in distribution (after rescaling) to a random variable $\mathcal{C}$, we first extract the factorial moments of $|\mathcal{C}_n|$ from the generating function $R(x, u)$. The next steps are to show that after proper rescaling, the factorial moments and integer moments conincide asymptotically and converge to the moments of a random variable $\mathcal{C}$, and to prove that $\mathcal{C}$ is uniquely determined by its moments.

Let $R_k(x) := \frac{\partial^k}{\partial u^k} R(x, u)\big|_{u=1}$. Then from standard methods (see for example [8, Proposition III.2]), the k'th factorial moment of $|\mathcal{C}_n|$ is given by $[x^n]R_k(x)/[x^n]R(x, 1)$. To study $R_k(x)$, we first use our recursion to establish the differential equation

$$\frac{\partial}{\partial x} R(x, u) = u \sum_{\delta=0}^{\infty} \frac{\phi(\delta)}{\delta!} (pR(x, u) + (1-p)B(x))^\delta. \tag{4}$$

When $\alpha = 0$ the above differential equation simplifies to

$$\frac{\partial}{\partial x} R(x, u) = u \exp (pR(x, u) - (1-p)\ln(1-x)),$$

and with the initial condition $R(0, u) = 0$, this linear differential equation has the solution $R(x, u) = \frac{-1}{p} \ln (1 - u + u(1-x)^p)$. Then $R_k(x) = \frac{1}{p}(k-1)!((1-x)^{-p} - 1)^k$, and so the factorial moments of $|\mathcal{C}_n|$ are given by $(k-1)!n^{pk}/(p\Gamma(pk))$. Once divided by n^{pk}, the factorial moments and integer moments conincide asymptotically and

$$\mathbb{E}\left[\frac{|\mathcal{C}_n|^k}{n^{pk}}\right] \sim \frac{(k-1)!n^{pk}}{n^{pk}p\Gamma(pk)} = \frac{k!}{\Gamma(pk+1)},$$

which are the moments of the Mittag-Leffler distribution with paramter p, a distribution uniquely determined by its moments. Therefore, we have that $|\mathcal{C}_n|/n^p$ converges in distribution to a random variable with Mittag-Leffler distribution.

For other cases of α, we were unable to find a closed form for $R(x, u)$. However, we only need to estimate $R_k(x)$ to extract approximations for the factorial moments of $|\mathcal{C}_n|$. When $\alpha > 0$, the differential equation 4 becomes

$$\frac{\partial}{\partial x}R(x, u) = u\left(1 - \left(pR(x, u) + (1 - p)\left(1 - (1 - (1 + 1/\alpha)x)^{\alpha/(1+\alpha)}\right)\right)\right)^{-1/\alpha}.$$

By differentiating both sides k times with respect to u and evaluating at $u = 1$, we achieve a differential equation for $R_k(x)$. With the help of induction we prove the following:

Lemma 2.1. *Let $\alpha > 0$. Then $R_k(x)$ is analytic on the cut plane $\mathbb{C} \setminus [1/(1 + 1/\alpha), \infty)$ and*

$$R_k(x) = C_k(1 - (1 + 1/\alpha)x)^{\frac{-kp - \alpha(k-1)}{1+\alpha}} + O\left((1 - (1 + 1/\alpha)x)^{\frac{-kp - \alpha(k-1)}{1+\alpha} + \varepsilon}\right)$$

for some $\varepsilon > 0$, where C_k satisfies the recursion given in (2).

Using a transfer theorem (see [8, Corollary VI.1]), we can approximate the coefficients of $R_k(x)$. Using [8, Proposition III.2] again, we extract the k'th factorial moments of $|\mathcal{C}_n|$. After rescaling by $n^{(p+\alpha)/(1+\alpha)}$, the factorial and integer moments coincide asymptotically and

$$\mathbb{E}\left[\frac{|\mathcal{C}_n|^k}{n^{(p+\alpha)/(1+\alpha)}}\right] \to \frac{C_k(1 + \alpha)\Gamma(1/(1 + \alpha))}{\alpha\Gamma((kp + \alpha(k - 1))/(\alpha + 1))}. \tag{5}$$

All that is left is to prove that the limiting distribution is determined by its moments. We can prove that the exponential generating function for the coefficients C_k above exists for a positive radius around 0. The exponential generating function generated by the limiting integer moments in (5) has a larger radius of convergence. Therefore the distribution $\mathcal{C}$ with integer moments given by (5) has a moment generating function that exists for a positive radius, and so $\mathcal{C}$ is uniquely determined by its moments. The moment of methods can therefore be applied to prove Theorem 1.1. Theorem 1.2 is proved in a similar manner.

References

[1] L. Addario-Berry, L. Devroye, G. Lugosi, and V. Velona, Broadcasting on random recursive trees, *Ann. Appl. Probab.* **32** (2022), 497–528.

[2] E. Baur, On a class of random walks with reinforced memory, *J. Stat. Phys.* **181** (2020), 772–802.

[3] E. Baur and J. Bertoin, The fragmentation process of an infinite recursive tree and Ornstein-Uhlenbeck type processes, *Electron. J. Probab.* **20** (2015), 1–20.

[4] S. Businger, The shark random swim, *J. Stat. Phys.* **172** (2018), 701–717.

[5] L. Comtet, *Advanced combinatorics. The art of finite and infinite expansions*, D. Reidel Publishing Co., 1974.

[6] C. Desmarais, C. Holmgren, and S. Wagner, Broadcasting-induce colorings of preferential attachment trees, *Random Structures & Algorithms* (2023), available online `https://doi.org/10.1002/rsa.21142`, 42pp.

[7] M. Drmota, *Random trees: an interplay between combinatorics and probability*, Springer Science and Business Media, 2009.

[8] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009.