NOTE ON THE EXPONENTIAL RECURSIVE $k$-ARY TREES

MINA GHASEMI¹, MEHRI JAVANIAN²,* and RAMIN IMANY NABIYYI¹

Abstract. In the present paper, we consider exponential recursive trees with no node of outdegree greater than $k$, called exponential recursive $k$-ary trees ($k \geq 2$). At each step of growing of these trees, every external node (insertion position) is changed into a leaf with probability $p$, or fails to do so with probability $1 - p$. We investigate limiting behavior of fundamental parameters such as size, leaves and distances in exponential recursive $k$-ary trees.

Mathematics Subject Classification. 05C05, 60C05, 60F05, 60G42.

Received April 6, 2022. Accepted February 17, 2022.

1. Introduction

A social network site is a web-based platform which people use to build social relationships with family, friends, colleagues, customers, or clients. Social networking can have a social purpose, a business purpose, or both, through sites like Instagram, WhatsApp, Facebook, Telegram, YouTube, and Twitter. These sites allow people and corporations to connect with one another so they can develop relationships very quickly. Certain aspects of fast-growing social networks can be modeled by exponential recursive trees, see [1, 4, 7].

By a recursive tree, we mean a labeled rooted tree where the nodes on a path from the root to any other node form an increasing sequence. Recursive trees grow dynamically, by attaching every new node to one of the already present nodes. The recursive trees are slow-growing where one node is added at each step (see Fig. 1). So, the recursive tree models cannot be suitable for fast-growing phenomena (e.g., the Corona virus spreads very quickly from person to person). See a survey of results of recursive trees in [10].

A recursive $k$-ary tree is a slow version of recursive trees in which every node has at most $k$ children. The nodes that already exist in the tree are called internal. In order to identify the locations of possible insertions of a new node (internal node), a recursive $k$-ary tree is extended by having $k - d$ external nodes attached as children of a node (internal node) of outdegree $d$ (e.g., for $k = 3$, in Fig. 2, see growing a recursive 3-ary tree of size 3). In an extended recursive $k$-ary tree of $n$ nodes, the number of all external nodes is $(k - 1)n + 1$, see [5].

An alternative fast-growing model, the exponential recursive $k$-ary tree ($k \geq 2$), grows by having every external node either converted into a leaf (an internal node of outdegree 0) with probability $p$ or stay unchanged with probability $q := 1 - p$.

We define the random exponential recursive $k$-ary tree of index $p$ and age $n$ grown as follows. Initially (at age 0), there is an external node. For $n \geq 1$, at the $n$th step (at age $n$), every external node independently is

Keywords and phrases: Recursive $k$-ary trees, contraction method, almost sure convergence, martingale.

¹ Department of Statistics, Faculty of Mathematical Sciences, University of Tabriz, Tabriz, Iran.
² Department of Statistics, Faculty of Sciences, University of Zanjan, Zanjan, Iran.
* Corresponding author: javanian@znu.ac.ir
changed into a leaf with probability \( p \in (0, 1) \), or not to be changed with probability \( q = 1 - p \). Figure 3 shows all exponential recursive ternary trees of ages 0, 1 and 2 and their probabilities. In the figure, external nodes and internal nodes are represented by squares and circles, respectively. The probability of getting each tree from the previous one appears above the arrow leading to it. For example, the leftmost tree in the third row occurs with probability \( p^2q^2 \).

2. Size

Let \( S_n \) be the size of a random exponential recursive \( k \)-ary tree of index \( p \) and age \( n \) \((k \geq 2)\), and let \( Y_n \) be the number of external nodes in the extension. Thus, we have

\[
Y_n = (k - 1)S_n + 1.
\]  

(2.1)

Let the notation \( \text{Bin}(m, p) \) stands for a binomial random variable that counts the number of successes in \( m \) independent identically distributed trials with success probability \( p \) in each trial. Each external node that succeeds to be converted into a leaf adds \( k - 1 \) external nodes; we have

\[
Y_n = Y_{n-1} + (k - 1)\text{Bin}(Y_{n-1}, p).
\]

Let \( \mathcal{F}_n \) be the sigma field generated by the first \( n \) steps of evolution. Then,

\[
\mathbb{E}[Y_n | \mathcal{F}_{n-1}] = Y_{n-1} + (k - 1)pY_{n-1} = ((k - 1)p + 1)Y_{n-1}.
\]

(2.2)

\[
\mathbb{E}[Y_n^2 | \mathcal{F}_{n-1}] = ((k - 1)p + 1)^2Y_{n-1}^2 + (k - 1)^2p(1 - p)Y_{n-1}.
\]
It follows that \( Y_n/((k-1)p+1)^n \) is a martingale with bounded moments

\[
\mathbb{E}\left[ \frac{Y_n}{((k-1)p+1)^n} \right] = \mathbb{E}[Y_0] = 1, \quad \mathbb{E}\left[ \frac{Y_n^2}{((k-1)p+1)^{2n}} \right] = \frac{(k-1)p((k-1)p+1)+k}{((k-1)p+1)^2} < \infty.
\]

That is, \( \mathbb{E}[Y_n] = ((k-1)p+1)^n \) and, by (2.1),

\[
\mathbb{E}[S_n] = \frac{((k-1)p+1)^n - 1}{k-1}.
\]

According to the martingale convergence theorem (Thm. 6.6.9 in [2]), we have

\[
\frac{Y_n}{((k-1)p+1)^n} \xrightarrow{L^2 \text{ & a.s.}} Y_*,
\]

where \( Y_* \) is an integrable limit random variable. Since \( \frac{1}{((k-1)p+1)^n} \to 0 \), we have

\[
\frac{S_n}{((k-1)p+1)^n} \xrightarrow{\text{a.s.}} S_* := \frac{Y_*}{k-1}.
\]
for a limiting random variable $S_\ast$.
We now characterize the limiting distribution of $S_n/((k-1)p+1)^n$ by inductively-constructed moments. The calculation techniques were taken from [1].

**Theorem 2.1.** Let $S_n$ be the number of nodes in an exponential recursive $k$-ary tree of age $n$ and index $p$ ($k \geq 2$). As $n \to \infty$, we have almost sure convergence

$$\frac{S_n}{((k-1)p+1)^n} \xrightarrow{a.s.} S_\ast,$$

where the limiting random variable $S_\ast$ has moments $a_m := \mathbb{E}[S_m^m]$ defined inductively by

$$a_m = \frac{p}{((k-1)p+1)^m - ((k-1)p+1)} \sum_{i_1, \ldots, i_k \geq 0 \atop i_1 \neq m, \ldots, i_k \neq m} \binom{m}{i_1, \ldots, i_k} a_{i_1} \cdots a_{i_k},$$

for $m \geq 2$, with $a_1 = 1/(k-1)$.

**Proof.** After one step of evolution, the initial external node either succeeds to recruit a leaf (event $R$) or not. Let $\mathbb{I}$ be the indicator of event $R$, that assumes the value 1, if $R$ occurs, and is 0 otherwise. If $R$ occurs, the extended exponential recursive $k$-ary tree has an internal node with $k$ external nodes. By time $n$, these $k$ external nodes produce $k$ independent exponential recursive $k$-ary subtrees of sizes $S_{n-1}^{(1)}, S_{n-1}^{(2)}, \ldots, S_{n-1}^{(k)}$. These $k$ random variables are independent and each is distributed like $S_{n-1}$. Alternatively, if $R$ does not occur, the size of the exponential recursive $k$-ary tree at step $n$ is the same as $S_{n-1}$. Hence, the size satisfies the distributional equation

$$S_n \overset{d}{=} (1 - \mathbb{I})S_{n-1} + \mathbb{I}(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1), \tag{2.6}$$

where the symbol $\overset{d}{=}$ denotes the equality in distribution.

Raise both sides of the equation (2.6), to the $m$th power. Since $(1 - \mathbb{I})^i \mathbb{I}^{m-i} = 0$, for $i > 0$ and $i < m$, we have

$$S_n^m = \sum_{i=0}^{m} \binom{m}{i} (1 - \mathbb{I})^i S_{n-1}^i \times \mathbb{I}^{m-i}(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1)^{m-i}$$

$$= (1 - \mathbb{I})^m S_{n-1}^m + \mathbb{I}^m(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1)^m$$

$$\overset{d}{=} (1 - \mathbb{I})S_{n-1}^m + \mathbb{I}(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1)^m. \tag{2.7}$$

Take expectations both sides of the equation (2.7). By the independence relations and identical distribution in the $k$ subtrees, we get

$$\mathbb{E}[S_n^m] = \mathbb{E}[1 - \mathbb{I}]\mathbb{E}[S_{n-1}^m] + \mathbb{E}[\mathbb{I}] \sum_{i_1, \ldots, i_k, j \geq 0 \atop i_1 + \cdots + i_k + j = m} \binom{m}{i_1, \ldots, i_k, j} \mathbb{E}[(S_{n-1}^{(1)})^{i_1}] \cdots \mathbb{E}[(S_{n-1}^{(k)})^{i_k}]$$

$$= (1 - p)\mathbb{E}[S_{n-1}^m] + p \sum_{i_1, \ldots, i_k, j \geq 0 \atop i_1 + \cdots + i_k + j = m} \binom{m}{i_1, \ldots, i_k, j} \mathbb{E}[(S_{n-1}^{(1)})^{i_1}] \cdots \mathbb{E}[(S_{n-1}^{(k)})^{i_k}].$$
Scale the above recurrence by \((k - 1)p + 1)^m\) to get

\[
\mathbb{E}\left[\left(\frac{S_n}{((k - 1)p + 1)^n}\right)^m\right] = \frac{1 - p}{((k - 1)p + 1)^m} \mathbb{E}\left[\left(\frac{S_{n-1}}{((k - 1)p + 1)^{n-1}}\right)^m\right] + \frac{p}{((k - 1)p + 1)^m} \sum_{i_1, \ldots, i_k, j \geq 0} \binom{m}{i_1, \ldots, i_k, j = m} \frac{1}{((k - 1)p + 1)^{(n-1)}}
\]

\[
\times \mathbb{E}\left[\left(\frac{S_{n-1}^{(1)}}{((k - 1)p + 1)^{n-1}}\right)^{i_1} \ldots \mathbb{E}\left[\left(\frac{S_{n-1}^{(k)}}{((k - 1)p + 1)^{n-1}}\right)^{i_k}\right]\right],
\]

for \(m \geq 2\). According to (2.3), \(a_1 := \lim_{n \to \infty} \mathbb{E}[S_n]/((k - 1)p + 1)^n = 1/(k - 1)\) exists. So, by induction on \(m\), \(a_m := \lim_{n \to \infty} \mathbb{E}[S_n^m]/((k - 1)p + 1)^{mn}\) exists. Taking limits, as \(n \to \infty\), we get

\[
((k - 1)p + 1)^m a_m = (1 - p)a_m + p \sum_{i_1, \ldots, i_k, j \geq 0} \binom{m}{i_1, \ldots, i_k} a_{i_1} \cdots a_{i_k}.
\]

Isolating the cases \(i_j = m\), for \(j = 1, 2, \ldots, k\), and rearranging, we get the claimed recurrence.

Now, by induction on \(m\), we show that \(a_m/m! \leq 1/(k - 1)\). This assertion is true for \(m = 1\). For \(m \geq 2\) and \(1 \leq i \leq m - 1\), we assume \(a_i/i! \leq 1/(k - 1)\). From this assumption, we have

\[
\frac{a_m}{m!} = \frac{p}{((k - 1)p + 1)^m - ((k - 1)p + 1)} \sum_{i_1, \ldots, i_k, j \geq 0} \frac{a_{i_1}}{i_1!} \cdots \frac{a_{i_k}}{i_k!} \leq \frac{p}{\sum_{j=0}^{m} \binom{m}{j} (k - 1)^{j+1}} \sum_{i_1, \ldots, i_k, j \geq 0} \frac{1}{i_1! \cdots i_k!} \frac{1}{k - 1}
\]

\[
= \frac{(m + k - 1)p - kp}{(m - 1)(k - 1)^m + \sum_{j=2}^{m} \binom{m}{j} (k - 1)^{j+1}} \cdot \frac{1}{(k - 1)^m}
\]

\[
= \frac{(m + k - 1)(m + k - 2) \cdots (m + k - (m - 1))k - km!}{m!(m - 1)(k - 1)^m} \cdot \frac{1}{k - 1}
\]

\[
\leq \frac{1}{k - 1},
\]

which completes the induction. Subsequently, we have

\[
\sum_{m=0}^{\infty} \frac{a_m}{m!} z^m \leq \sum_{m=0}^{\infty} \frac{a_m}{m!} |z|^m \leq \frac{1}{k - 1} \sum_{m=0}^{\infty} |z|^m \leq \frac{1}{1 - |z|} < \infty,
\]

for all complex \(z\), with \(|z| < 1\). Therefore, by Theorem 30.1 in [3], we conclude the uniqueness of the distribution with the moments \(a_m, m \geq 1\). This verifies the convergence in distribution \(S_n/((k - 1)p + 1)^n \xrightarrow{D} S\), for some random variable \(S\) with the moments \(a_m = \mathbb{E}[S^m], m \geq 1\). Since \(S_n/((k - 1)p + 1)^n \xrightarrow{a.s.} S\), so it converges in distribution to \(S\). That is \(S \overset{d}{=} S\). Thus, \(S\) has the moments given by the stated recurrence. \(\square\)
3. Leaves

Let $L_n$ be the number of leaves in an exponential $k$-ary tree of age $n$ and index $p$. After one step of evolution, if the initial external node succeeds to recruit a leaf, then the extended exponential recursive $k$-ary tree has an internal node with $k$ external nodes. In this case, for $i = 1, \ldots, k$, let $J_{n-1}^{(i)}$ be an indicator of the event that the $i$th external node does not progress into a tree of age $n - 1$. By the $n$th step, these $k$ external nodes produce $k$ independent exponential recursive $k$-ary subtrees of age $n - 1$. Let the number of leaves arising from these $k$ subtrees be respectively $L_{n-1}^{(1)}$, $L_{n-1}^{(2)}$, \ldots, $L_{n-1}^{(k)}$. These $k$ random variables are independent and each is distributed like $L_{n-1}$. After $n$ steps of evolution, the root of the tree is a leaf if and only if $J_{n-1}^{(1)} \cdots J_{n-1}^{(k)} = 1$. Alternatively, if the initial external node does not succeed to recruit a leaf, then the number of leaves in the exponential recursive $k$-ary tree at step $n$ is the same as $L_{n-1}$. Hence, the number of leaves satisfies the distributional equation

$$L_n \overset{d}{=} (1 - p) L_{n-1} + 1 \left( L_{n-1}^{(1)} + L_{n-1}^{(2)} + \ldots + L_{n-1}^{(k)} + J_{n-1}^{(1)} \cdots J_{n-1}^{(k)} \right).$$

(3.1)

Note that, for each $n \geq 1$, the vectors $(L_{n-1}^{(i)}, J_{n-1}^{(i)})$, for $i = 1, \ldots, k$, are independent. So, for $m \geq 1$, by raising both sides of the distributional equation (3.1) to the $m$th power, taking expectations, and observing independence and identical distribution in the subtrees, we have

$$\mathbb{E}[L_n^m] = (1 - p) \mathbb{E}[L_{n-1}^m] + p \sum_{i_1, \ldots, i_k \geq 0 \atop i_1 + \ldots + i_k = m} \binom{m}{i_1, \ldots, i_k} \mathbb{E}\left[ (L_{n-1}^{(1)})^{i_1} (J_{n-1}^{(1)})^j \right] \cdots \mathbb{E}\left[ (L_{n-1}^{(k)})^{i_k} (J_{n-1}^{(k)})^j \right]$$

(3.2)

$$= (1 - p) \mathbb{E}[L_{n-1}^m] + p \sum_{i_1, \ldots, i_k \geq 0 \atop i_1 + \ldots + i_k = m} \binom{m}{i_1, \ldots, i_k} \mathbb{E}\left[ (L_{n-1}^{(1)})^{i_1} \right] \cdots \mathbb{E}\left[ (L_{n-1}^{(k)})^{i_k} \right] + p(1 - p)^k(n - 1).$$

From (3.2), in particular, we can calculate the exact mean, and find its asymptotic equivalent.

**Proposition 3.1.** Let $L_n$ be the number of leaves in an exponential $k$-ary tree of age $n$ and index $p$. Then we have

$$\mathbb{E}[L_n] = \frac{p ((k - 1)p + 1)^n - (1 - p)^{kn}}{(k - 1)p + 1 - (1 - p)^k} \sim \frac{p((k - 1)p + 1)^n}{(k - 1)p + 1 - (1 - p)^k}, \quad \text{as } n \to \infty.$$

**Proof.** Take the expectation of (3.1), to write

$$\mathbb{E}[L_n] = (1 - p) \mathbb{E}[L_{n-1}] + p \mathbb{E}\left[ L_{n-1}^{(1)} \right] + \ldots + p \mathbb{E}\left[ L_{n-1}^{(k)} \right] + p(1 - p)^k(n - 1)$$

$$= ((k - 1)p + 1) \mathbb{E}[L_{n-1}] + p(1 - p)^k(n - 1).$$

The solution to the recurrence of the mean is

$$\mathbb{E}[L_n] = p \sum_{j=0}^{n-1} (1 - p)^j ((k - 1)p + 1)^{n-j-1}$$

$$= p((k - 1)p + 1)^n \sum_{j=0}^{n-1} \left( \frac{(1 - p)^k}{(k - 1)p + 1} \right)^j.$$
The sum is a geometric series, yielding

$$E[L_n] = p((k-1)p+1)^{n-1} \cdot \frac{1 - \left(\frac{(1-p)^k}{(k-1)p+1}\right)^n}{1 - \left(\frac{(1-p)^k}{(k-1)p+1}\right)}.$$ 

This implies the assertion.

The calculation techniques were taken from [1].

**Theorem 3.2.** Let $L_n$ be the number of leaves in an exponential recursive $k$-ary tree of age $n$ and index $p$. We have the convergence in distribution

$$\frac{L_n}{((k-1)p+1)^n} \xrightarrow{d} L_\ast,$$

where the limiting random variable $L_\ast$ has moments $b_m := E[L_\ast^m]$ defined inductively by

$$b_m = \frac{p}{((k-1)p+1)^m - ((k-1)p+1)} \sum_{i_1, \ldots, i_k \geq 0 \atop i_1 \neq m, \ldots, i_k \neq m} \binom{m}{i_1, \ldots, i_k} b_{i_1} \cdots b_{i_k},$$

for $m \geq 2$ with $b_1 = p/((k-1)p+1 - (1-p)^k)$.

**Proof.** Scale the recurrence (3.2) by $((k-1)p+1)^{nm}$ to get

$$E\left[\left(\frac{L_n}{((k-1)p+1)^n}\right)^m\right] = \frac{1-p}{((k-1)p+1)^m} E\left[\left(\frac{L_{n-1}}{((k-1)p+1)^{n-1}}\right)^m\right]$$

$$+ \frac{p}{((k-1)p+1)^m} \sum_{i_1, \ldots, i_k \geq 0 \atop i_1 \neq m, \ldots, i_k \neq m} \binom{m}{i_1, \ldots, i_k} E\left[\left(\frac{L_{n-1}^{(1)}}{((k-1)p+1)^{n-1}}\right)^i_1\right]$$

$$\times \cdots \times E\left[\left(\frac{L_{n-1}^{(k)}}{((k-1)p+1)^{n-1}}\right)^i_k\right] + \frac{p(1-p)^{k(n-1)}}{((k-1)p+1)^{nm}}.$$

Thus, inductively the sequence of the limit $b_m$ satisfies the recurrence

$$((k-1)p+1)^m b_m = (1-p)b_m + pkb_m + p \sum_{i_1, \ldots, i_k \geq 0 \atop i_1 \neq m, \ldots, i_k \neq m} \binom{m}{i_1, \ldots, i_k} b_{i_1} \cdots b_{i_k},$$

for $m \geq 2$.

Using this recurrence and $b_1 < 1/(k-1)$, by (2.8), we can conclude that $\frac{b_m}{m!} < 1/(k-1)$ for $m \geq 1$. So, for $|z| < 1$, the series $\sum_{m=0}^{\infty} \frac{b_m}{m!} z^m$ converges. Therefore, by Theorem 30.1 in [3], $L_n/((k-1)p+1)^n$ converges in distribution to a unique limit $L_\ast$. 

4. **EXTERNAL PATH LENGTH**

Let $W_n$ be the external path length, which is the sum of the root-to-external node distances (the sum of the depth of external nodes). We use $T_n$ for an exponential recursive $k$-ary tree of age $n$ and index $p$. Assume, we
number the external nodes of $T_n$ arbitrarily with the numbers $1, \ldots, Y_n$. Let $D_i$ be the depth of external node $i$, for $i = 1, \ldots, Y_n$. Then the external path length of $T_n$ is

$$W_n = \sum_{i=1}^{Y_n} D_i.$$ 

Similar to the equation (2.6), for $n \geq 1$, $W_n$ and $Y_n$ satisfy the distributional equations

$$W_n \overset{d}{=} (1 - \mathbb{1})W_{n-1}^{\prime} + \mathbb{1}(W_{n-1}^{(1)} + \cdots + W_{n-1}^{(k)} + Y_{n-1}^{(1)} + \cdots + Y_{n-1}^{(k)}),$$

(4.1) 

$$Y_n \overset{d}{=} (1 - \mathbb{1})Y_{n-1}^{\prime} + \mathbb{1}(Y_{n-1}^{(1)} + \cdots + Y_{n-1}^{(k)}), \quad (W_0 = 0, Y_0 = 1).$$

(4.2)

Here $W_{n-1}^{(i)}$, $i = 0, 1, \ldots, k$, and $W_{n-1}^{\prime}$ are independent copies of $W_{n-1}$ and independent of $\mathbb{1}$; $Y_{n-1}^{(i)}$, $i = 0, 1, \ldots, k$, and $Y_{n-1}^{\prime}$ are independent copies of $Y_{n-1}$ and independent of $\mathbb{1}$.

4.1. The expectation and variance

The pair of equations (4.1) and (4.2) is sufficient to determine the means and the quadratic order moments exactly.

**Theorem 4.1.** Let $W_n$ be the external path length in an exponential $k$-ary tree of age $n$ and index $p$. We have

$$\mathbb{E}[W_n] = kp(n - 1)p + 1)^{n-1},$$

(4.3) 

$$\text{Var}[W_n] \sim k^2(k - 1)p^2(1 - p)n^2 ((k - 1)p + 1)^{2n-3}, \quad \text{as } n \to \infty.$$ 

**Proof.** Raise both sides of (4.2) to the second power. So we get

$$Y_n^2 \overset{d}{=} (1 - \mathbb{1})Y_{n-1}^2 + \mathbb{1}\left(\sum_{j=1}^{k} (Y_{n-1}^{(j)})^2 + \sum_{1 \leq j_1, j_2 \leq k \atop j_1 \neq j_2} Y_{n-1}^{(j_1)} Y_{n-1}^{(j_2)}\right).$$

(4.4)

Take the expectation of (4.4) and observe the independence $Y_{n-1}^{(i)}$ and $Y_{n-1}^{(j)}$ for $i \neq j$, to write

$$\mathbb{E}[Y_n^2] = (1 - p)\mathbb{E}[Y_{n-1}^2] + p\left(k\mathbb{E}[Y_{n-1}^2] + k(k - 1)(\mathbb{E}[Y_{n-1}])^2\right)$$

$$= ((k - 1)p + 1)\mathbb{E}[Y_{n-1}^2] + k(k - 1)p((k - 1)p + 1)^{2n-2},$$

with initial condition $\mathbb{E}[Y_0^2] = 1$. Iterating this formula, we find

$$\mathbb{E}[Y_n^2] = ((k - 1)p + 1)^n + k(k - 1)p \sum_{j=0}^{n-1} ((k - 1)p + 1)^{2n-j-2}$$

$$= \left(k((k - 1)p + 1)^{2n-1}(1 + o(1)), \quad \text{as } n \to \infty. \right.$$

(4.5)

Taking the expectation of (4.1), we get

$$\mathbb{E}[W_n] = (1 - p)\mathbb{E}[W_{n-1}] + kp\mathbb{E}[W_{n-1}] + kp\mathbb{E}[Y_{n-1}]$$

$$= ((k - 1)p + 1)\mathbb{E}[W_{n-1}] + kp((k - 1)p + 1)^{n-1}$$

$$= \left(k((k - 1)p + 1)^{n}(1 + o(1)), \quad \text{as } n \to \infty. \right.$$
with initial condition $\mathbb{E}[W_0] = 0$. By iterating, the mean value follows.

Then multiplying (4.1) and (4.2), we deduce that,

$$W_n Y_n = (1 - I) W_{n-1} Y_{n-1} + I \left( \sum_{j=1}^{k} (Y_n^{(j)})^2 + \sum_{1 \leq j_1, j_2 \leq k, j_1 \neq j_2} Y_n^{(j_1)} Y_n^{(j_2)} \right)$$

$$+ I \left( \sum_{j=1}^{k} W_{n-1}^{(j)} Y_{n-1} + \sum_{1 \leq j_1, j_2 \leq k, j_1 \neq j_2} W_{n-1}^{(j_1)} Y_{n-1}^{(j_2)} \right).$$

We next take the expectation and get the following recurrence for $\mathbb{E}[W_n Y_n]$,

$$\mathbb{E}[W_n Y_n] = ((k - 1)p + 1)\mathbb{E}[W_{n-1} Y_{n-1}] + kp\mathbb{E}[Y_{n-1}^2] + k(k - 1)p \left( \mathbb{E}[W_{n-1}] \mathbb{E}[Y_{n-1}] + (\mathbb{E}[Y_{n-1}])^2 \right)$$

$$= ((k - 1)p + 1)\mathbb{E}[W_{n-1} Y_{n-1}] + kp((k - 1)p + 1)^{n-1} + k(k - 1)p((k - 1)p + 1)^{2n-2} + k^2(k - 1)p^2 (n - 1)((k - 1)p + 1)^{2n-3} + k^2(k - 1)p^2 \sum_{i=0}^{n-2} ((k - 1)p + 1)^{2n-i-4},$$

with initial condition $\mathbb{E}[W_0 Y_0] = 0$. This gives the solution

$$\mathbb{E}[W_n Y_n] = kp(n((k - 1)p + 1)^{n-1} + k^2(k - 1)p^2 \sum_{j=0}^{n-1} (n - j - 1)((k - 1)p + 1)^{2n-j-3}$$

$$+ k(k - 1)p \sum_{j=0}^{n-1} ((k - 1)p + 1)^{2n-j-2} + k^2(k - 1)p^2 \sum_{j=0}^{n-1} \sum_{i=0}^{n-j-2} ((k - 1)p + 1)^{2n-j-i-4}$$

$$= \left( k^2pn((k - 1)p + 1)^{2n-2} \right) (1 + o(1)), \quad \text{as } n \to \infty. \quad (4.6)$$

Now, we take the square of the equation (4.1). Then, by taking expectation, we have

$$\mathbb{E}[W_n^2] = ((k - 1)p + 1)\mathbb{E}[W_{n-1}^2] + kp\mathbb{E}[Y_{n-1}^2] + k(k - 1)p(\mathbb{E}[Y_{n-1}])^2$$

$$+ k(k - 1)p\mathbb{E}[W_{n-1}]\mathbb{E}[Y_{n-1}] + k(k - 1)p(\mathbb{E}[W_{n-1}])^2 + kp\mathbb{E}[W_{n-1} Y_{n-1}].$$

Again, by iterating, we obtain

$$\mathbb{E}[W_n^2] = kp(n((k - 1)p + 1)^{n-1} + k^2(k - 1)p^2 \sum_{l=0}^{n-1} \sum_{j=0}^{n-l-2} ((k - 1)p + 1)^{2n-l-j-4}$$

$$+ k(k - 1)p \sum_{l=0}^{n-1} ((k - 1)p + 1)^{2n-l-2} + k^2(k - 1)p^2 \sum_{l=0}^{n-1} (n - l - 1)((k - 1)p + 1)^{2n-l-3}$$

$$+ k^3(k - 1)p^3 \sum_{l=0}^{n-1} (n - l - 1)^2((k - 1)p + 1)^{2n-l-4} + k^2p^2 \sum_{l=0}^{n-1} (n - l - 1)((k - 1)p + 1)^{n-l-2}$$

$$+ k^2(k - 1)p^2 \sum_{l=0}^{n-1} \sum_{j=0}^{n-l-2} ((k - 1)p + 1)^{2n-l-j-4} + kp \sum_{i=0}^{n-l-j-3} ((k - 1)p + 1)^{2n-l-j-i-6}.$$
Theorem 4.2. The scaled external path length

\[ W_n/kpn((k-1)p+1)^n \]

converges in distribution to some random variable \( W \) with distribution, the unique solution of the following equation

\[ \hat{W} = \frac{1}{(k-1)p+1} \hat{W}(0) + \frac{1}{(k-1)p+1} (\hat{W}(1) + \ldots + \hat{W}(k)) \]

(4.8)

where \( \hat{W}(i), i = 0, 1, \ldots, k, \) are independent copies of \( \hat{W} \) and independent of \( I \).

Proof. The proof is based on the Lemma 3.1 page 502 of the paper [8]. Let \( \mathcal{M}_2 \) be the space of all distributions with finite absolute second moment.

Consider the transformation \( T: \mathcal{M}_2 \rightarrow \mathcal{M}_2 \) defined by

\[ T(\mu) \overset{d}{=} \frac{1 - I}{(k-1)p+1} \hat{W}(0) + \frac{1}{(k-1)p+1} (\hat{W}(1) + \ldots + \hat{W}(k)) \]

(4.9)

where \( \hat{W}(i), i = 0, 1, \ldots, k \) and \( I \) are independent and \( \hat{W}(i) \) have \( \mu \) as distribution.

At a first step we have to prove that with respect to the metric \( L^2_2 \) defined on \( \mathcal{M}_2 \) by

\[ L^2_2(\mu, \nu) := \min \{ \mathbb{E}[|X-Y|^2] \}, \text{ where } X, Y \text{ have } \mu \text{ and } \nu \text{ as distribution, respectively} \]

the transformation have a unique fixed point. In fact let \( \mu \) and \( \nu \) be two measures of \( \mathcal{M}_2 ")
\( \mathcal{L}_2^2(T(\mu), T(\nu)) \leq \mathbb{E} \left[ (T(\mu) - T(\nu))^2 \right] \)

\[
= \frac{k \mathbb{E}[1^2]}{((k-1)p + 1)^2} \mathbb{E}[\hat{W}_\mu - \hat{W}_\nu]^2 + \frac{\mathbb{E}[(1-1)^2]}{((k-1)p + 1)^2} \mathbb{E}[\hat{W}_\mu - \hat{W}_\nu]^2 \\
= \frac{1}{(k-1)p + 1} \mathbb{E}[\hat{W}_\mu - \hat{W}_\nu]^2 = \frac{1}{(k-1)p + 1} \mathcal{L}_2^2(\mu, \nu).
\]

Since \(1/((k-1)p + 1) < 1\), then \(T\) is a contracted transform and it has only one fixed point.

Now, if \(\hat{W}_n\) converges in distribution, the limit will be the unique fixed point of the transformation \(T\). Namely, we have to prove that

\[
\lim_{n \to \infty} \mathcal{L}_2^2(\hat{W}, \hat{W}_n) = 0,
\]

to conclude that \(\hat{W}_n\) converges in distribution and in \(L_2\) to \(\hat{W}\). We have

\[
\mathcal{L}_2^2(\hat{W}, \hat{W}_n) \leq \frac{p}{((k-1)p + 1)^2} \left( |\hat{W}^{(1)} - \hat{W}^{(1)}_{n-1}| \frac{n-1}{n} \right)^2 + \ldots + \frac{1 - p}{((k-1)p + 1)^2} \left( |\hat{W}^{(k)} - \hat{W}^{(k)}_{n-1}| \frac{n-1}{n} \right)^2 \\
+ \frac{2}{k((k-1)p + 1)n} \sum_{j=1}^{k} \sum_{i=1}^{k} \mathbb{E} \left[ \frac{Y^{(i)}_{n-1}}{((k-1)p + 1)^{n-1}} \left| \hat{W}^{(j)} - \hat{W}^{(j)}_{n-1} \frac{n-1}{n} \right| \right].
\]

Then we deduce

\[
\lim_{n \to \infty} \mathcal{L}_2^2(\hat{W}, \hat{W}_n) \leq \frac{1}{(k-1)p + 1} \lim_{n \to \infty} \mathcal{L}_2^2(\hat{W}, \hat{W}_{n-1}) < \lim_{n \to \infty} \mathcal{L}_2^2(\hat{W}, \hat{W}_n).
\]

This implies \(\lim_{n \to \infty} \mathcal{L}_2^2(\hat{W}, \hat{W}_n) = 0\).

\[\square\]

4.3. Almost sure convergence

In this subsection, the calculation techniques were taken from [1].

**Theorem 4.3.** Let \(W_n\) be the external path length in an exponential k-ary tree of age \(n\) and index \(p\). As \(n \to \infty\), we have the almost sure and \(L^2\) convergence,

\[
\frac{W_n}{n((k-1)p + 1)^n} \overset{L^2 \; \text{a.s.}}{\longrightarrow} Y_\ast.
\]

**Proof.** Let \(B_k\) be a sequence of independent Bernoulli \((p)\) random variables, defined on the same probability space as the trees. Consider \(\{B_{k}\}_{k=1}^\infty\) to be independent of the structure of the tree at any age, too. If the \(i\)th external node recruits a leaf, it is converted into an internal node (a leaf node) at its own level and it adds \(k\) external nodes at the next level (as the children of the new leaf). The contribution of the \(i\)th external node of
\[ T_{n-1} \text{ into } W_n, \text{ is } D_i + (k(D_i + 1) - D_i)\mathbb{B}_i, \text{ and we write} \]
\[ W_n = \sum_{i=1}^{Y_{n-1}} \left( D_i + (k(D_i + 1) - D_i)\mathbb{B}_i \right). \]

Taking an expectation conditioned of \( \mathcal{F}_{n-1} \), we obtain
\[ \mathbb{E}[W_n | \mathcal{F}_{n-1}] = \sum_{i=1}^{Y_{n-1}} D_i + p \sum_{i=1}^{Y_{n-1}} (k(D_i + 1) - D_i) = ((k-1)p+1)W_{n-1} + kpY_{n-1}. \] (4.10)

Let \( M_n := \alpha_n W_n + \beta_n Y_n \), for specified factors \( \alpha_n \) and \( \beta_n \) that transformed \( M_n \) a martingale. So, by (2.2), (4.10) and a martingalization procedure, we have
\[ \mathbb{E}[M_n | \mathcal{F}_{n-1}] = \alpha_n((k-1)p+1)W_{n-1} + kp\alpha_nY_{n-1} + \beta_n((k-1)p+1)Y_{n-1} \]
\[ = M_{n-1} = \alpha_{n-1}W_{n-1} + \beta_{n-1}Y_{n-1}. \]

This is possible, if we choose
\[ \alpha_{n-1} = \alpha_n((k-1)p+1), \quad \beta_{n-1} = kp\alpha_n + \beta_n((k-1)p+1). \]

From these solvable recurrences, we obtain
\[ \alpha_n = \frac{\alpha_0}{((k-1)p+1)^n}, \quad \beta_n = \frac{\beta_0}{((k-1)p+1)^n} - \frac{kpn}{((k-1)p+1)^{n+1}}. \]

For simplicity, we take \( \alpha_0 = 1 \) and \( \beta_0 = 0 \). Thus, by (4.3), (4.5), (4.6) and (4.7),
\[ M_n = \frac{1}{((k-1)p+1)^n} W_n - \frac{kpn}{((k-1)p+1)^{n+1}} Y_n, \]
is a martingale with bounded moments
\[ \mathbb{E}[M_n] = \frac{1}{((k-1)p+1)^n} \mathbb{E}[W_n] - \frac{kpn}{((k-1)p+1)^{n+1}} \mathbb{E}[Y_n] \]
\[ = \frac{kpn ((k-1)p+1)^{n-1}}{((k-1)p+1)^n} - \frac{kpn ((k-1)p+1)^n}{((k-1)p+1)^{n+1}} = \frac{kpn}{(k-1)p+1} - \frac{kpn}{(k-1)p+1} = 0 < \infty, \]
\[ \mathbb{E}[M_n^2] = \frac{1}{((k-1)p+1)^{2n}} \mathbb{E}[W_n^2] + \frac{k^2 p^2 n^2}{((k-1)p+1)^{2n+2}} \mathbb{E}[Y_n^2] - 2 \frac{kpn}{((k-1)p+1)^{2n+1}} \mathbb{E}[W_n Y_n] \]
\[ = k^3 p^2 n^2 \left( \frac{((k-1)p+1)^{2n-3}}{((k-1)p+1)^{2n}} + \frac{((k-1)p+1)^{2n-1}}{((k-1)p+1)^{2n+2}} - 2 \frac{((k-1)p+1)^{2n-2}}{((k-1)p+1)^{2n+1}} \right) + o(1) \]
\[ = \left( \frac{k^3 p^2 n^2}{((k-1)p+1)^3} + \frac{k^3 p^2 n^2}{((k-1)p+1)^3} - 2 \frac{k^3 p^2 n^2}{((k-1)p+1)^3} \right) + o(1) = 0 + o(1) < \infty. \]
So, by the martingale convergence theorem (Thm. 6.6.9 in [2]), $M_n$ converges to some $M_\ast$ almost surely and in $L^2$. Consequently,

$$
\frac{M_n}{n} = \frac{1}{n ((k-1)p+1)^n} W_n - \frac{kp}{((k-1)p+1)^{n+1}} Y_n \xrightarrow{L^2 \text{ a.s.}} 0.
$$

Hence, by (2.4), the result follows. \hfill \Box

**Corollary 4.4.** Let $D_n^\ast$ be the depth of a randomly chosen external nodes in an exponential $k$-ary tree of age $n$ and index $p$. Then we have

$$
\frac{\mathbb{E}[D_n^\ast]}{n} \xrightarrow{\text{a.s.}} \frac{kp}{(k-1)p+1}.
$$

**Proof.** In $T_n$, we have

$$
\mathbb{E}[D_n^\ast | T_n] = \frac{D_1 + D_2 + \cdots + D_{Y_n}}{Y_n} = \frac{W_n}{Y_n}.
$$

Therefore, by Theorem (4.3), we have

$$
\frac{\mathbb{E}[D_n^\ast | T_n]}{n} = \frac{W_n / (n((k-1)p+1)^n)}{Y_n / (((k-1)p+1)^{n+1})} \xrightarrow{\text{a.s.}} \frac{kp}{(k-1)p+1}.
$$

Since $W_n \leq nY_n$, then $\mathbb{E}[D_n^\ast | T_n]/n \leq 1$. Consequently,

$$
\lim_{n \to \infty} \frac{\mathbb{E}[D_n^\ast]}{n} = \lim_{n \to \infty} \mathbb{E} \left[ \frac{\mathbb{E}[D_n^\ast | T_n]}{n} \right] = \mathbb{E} \left[ \lim_{n \to \infty} \frac{\mathbb{E}[D_n^\ast | T_n]}{n} \right]
$$

yields the result. \hfill \Box

**References**


