Communications in Combinatorics and Optimization
Vol. xx, No. x (xxxx), pp. 1-15
https://doi.org/10.22049/cco.2026.31556.2887

CCO
Commun. Comb. Optim.

Research Article

Path length and Sackin index of random m-oriented recursive trees

Ramin Kazemi[*], Sedigheh Zamani Mehreyan[†]

Department of Statistics, Imam Khomeini International University, Qazvin, Iran
[*]r.kazemi@sci.ikiu.ac.ir
[†]s.zamani@sci.ikiu.ac.ir

Received: 31 May 2026; Accepted: 4 September 2026
Published Online: 8 September 2026

Abstract: The main purpose of this article is to study two distance-based quantities, internal path length and Sackin index, in random m-oriented recursive trees of size n. Unlike the traditional method, the mean and variance of the internal path length I_n are *directly* calculated through a simple recurrence. The complexity of the calculations is due to the dependence of the probability of attracting a new node on the outdegree of the node. We show that there exists a random variable I such that $\frac{I_n - \frac{n}{m}\log n}{mn} \to I$ almost surely and in L^2, as $n \to \infty$. Then, under two assumptions, some results related to the Sackin index (S_n) of these tree models are given. Specifically, based on Chebychev's inequality, we show that $\frac{(m+1)S_n}{n \log n} \to 1$ in probability.

Keywords: random m-oriented recursive tree, path length, Sackin index, limiting rule.

AMS Subject Classification: 05C05, 60F05

1. Introduction

A tree of size n with nodes labelled by $1, 2, \ldots, n$ is called a *recursive tree* if the node labelled 1 is the root and, for all $2 \le k \le n$, the sequence of node labels in the path from the root to k is increasing [9]. If different orientations are taken to represent different trees, we arrive at a definition of an oriented recursive tree. A *random m-oriented recursive tree* (RMORT) of size n is constructed as follows. The tree starts from a root labelled 1. At stage $2 \le i \le n$ a new node i is attached to any previous node j of outdegree $d(j)$ of the already grown tree T_{i-1} of size $i-1$ with probability $\frac{(m-1)d(j)+1}{m(i-2)+1}$ (when $i = 2$, i.e., the tree has just the root node, then the node $i = 2$

* *Corresponding Author*

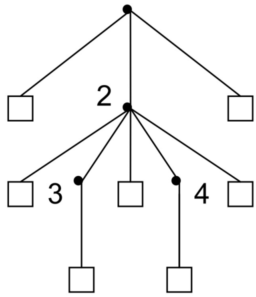

Figure 1. An extended 2-oriented recursive tree. The internal nodes are shown as circles and external nodes are shown as squares [4].

will attach to the root ($d(1) = 0$) with probability 1). Thus the higher outdegree nodes possess a higher attraction for n [4]. There is a different type of node called *external* at each possible insertion position. If a node v has outdegree $d(v)$; there are $d(v)$ children under it, with $(m-1)d(v)+1$ external nodes (see Figure 1). One of the applications of this model is in the discussion of investment in stock markets. For example, investors prefer to invest in cases where the demand for stock exchange has been higher. This generalization of the plane-oriented recursive trees at first were defined in [11].

Let D_i be the depth of i-th node in a random m-oriented recursive tree of size n (the number of edges from the root to the ith node). The internal path length is defined as $I_n = \sum_{i=1}^{n} D_i$. For external nodes labelled $1, 2, \ldots, m(n-1)+1$ when x_j is the depth of the jth external node, $X_n = \sum_{j=1}^{m(n-1)+1} x_j$ is defined as the external path length [9]. Also, the Sackin index S_n of a tree of size n is defined as the sum of the depths of its leaves (see [1, 8, 10, 12] for some related quantities). Thus, in m-oriented recursive trees, the external path length is the Sackin index of external nodes.

The traditional method for examining the internal path length is that the node depth variable is examined first. If the average depth of the node is calculated, the expected internal path length can be determined using the linear property. To calculate the variance I_n, it should use an indirect method because the direct method requires the calculation of the covariance between the depth of the nodes [6]. Suppose $q_{j,m} = m(j-1)+1$ for $j \geq 1$. Dobrow and Smythe [2] proved for a random m-oriented recursive tree of size n:

$$\mathbb{E}(D_n) = \sum_{j=1}^{n-1} \frac{1}{q_{j,m}} \sim \frac{1}{m} \log n, \tag{1.1}$$

$$\mathbb{V}(D_n) = \sum_{j=1}^{n-1} \frac{m(j-1)}{q_{j,m}^2} \sim \frac{1}{m} \log n \qquad n, m \geq 2. \tag{1.2}$$

In random m-oriented recursive trees, according to equality $X_n = mI_n + n$, it is easier to check variable X_n first, because the probability of attracting the nth label is equal to $\frac{1}{q_{n-1,m}}$. With this approach, Javanian and Vahidi-asl [4] showed that:

$$\mathbb{E}(I_n) = \frac{q_{n,m}}{m}\sum_{j=1}^{n}\frac{1}{q_{j,m}} - \frac{n}{m} \sim \frac{n}{m}\log n, \tag{1.3}$$

$$\mathbb{E}(X_n) = q_{n,m}\sum_{j=1}^{n}\frac{1}{q_{j,m}} \sim n\log n, \qquad n,m \geq 2. \tag{1.4}$$

The paper is organized as follows. In Section 2, the internal path length (through a recurrence) is studied by direct method. The exact mean and variance of this variable are presented. Like the external path length, the normalized internal path length is shown to converge almost surely and in L^2 to a limiting random variable via martingales. In Section 3, the Sackin index is examined. Again, the mean and variance of this variable are directly calculated through a recurrence. To calculate the variance, we need to calculate the covariance between S_n and I_n as well as S_n and S_{n-1}. To avoid lengthy calculations, we make two logical assumptions to arrive at asymptotic results. Convergence in probability of this variable will be proved by Chebyshev's inequality.

2. The path length

In this section, unlike the procedure presented in [4], the internal path length is studied directly. Due to the occurrence of outdegree in the probability of attracting a new node, this study needs to use the following lemma.

Lemma 1. *[7] For each rooted tree T_n of size n, we have*

$$\sum_{v\in V(T_n)} d(v)\, D(v) = I_n - (n-1),$$

where $D(v)$ is the depth of node v and $V(T_n)$ is the node set of T_n.

Theorem 1. *The mean of I_n in a random m-oriented recursive tree of size n is given by*

$$\mathbb{E}(I_n) = \frac{q_{n,m}}{m}\sum_{j=1}^{n}\frac{1}{q_{j,m}} - \frac{n}{m}, \qquad n \geq 3.$$

Proof. Let U_{n-1} be a randomly chosen node belonging to a RMORT of size $n-1$. If n is attached to U_{n-1} in T_{n-1}, then

$$I_n = I_{n-1} + D(U_{n-1}) + 1.$$

Let $\mathcal{F}_n$ be the σ-field generated by the first n stages of these models [3, 5]. Then

$$\begin{aligned}
\mathbb{E}(I_n|\mathcal{F}_{n-1}) &= I_{n-1} + \mathbb{E}(D(U_{n-1})|\mathcal{F}_{n-1}) + 1 \\
&= I_{n-1} + \sum_{v \in V(T_{n-1})} \frac{1+(m-1)d(v)}{q_{n-1,m}} D(v) + 1 \\
&= I_{n-1} + \frac{1}{q_{n-1,m}} \left[\sum_{v \in V(T_{n-1})} D(v) + \sum_{v \in V(T_{n-1})} (m-1)d(v)D(v) \right] + 1.
\end{aligned}$$

From Lemma 1,

$$\begin{aligned}
\mathbb{E}(I_n|\mathcal{F}_{n-1}) &= I_{n-1} + \frac{1}{q_{n-1,m}} \Big[(I_{n-1} + (m-1)(I_{n-1} - (n-2) \Big] + 1 \\
&= I_{n-1} + \frac{1}{q_{n-1,m}} \Big[mI_{n-1} - (n-2)(m-1) \Big] + 1 \\
&= \left(1 + \frac{m}{q_{n-1,m}}\right) I_{n-1} + \frac{n-1}{q_{n-1,m}}.
\end{aligned}$$

Hence

$$\mathbb{E}(I_n|\mathcal{F}_{n-1}) = \frac{q_{n,m}}{q_{n-1,m}} I_{n-1} + \frac{n-1}{q_{n-1,m}}. \tag{2.1}$$

Taking expectations of the relation (2.1),

$$\mathbb{E}(I_n) = \frac{q_{n,m}}{q_{n-1,m}} \mathbb{E}(I_{n-1}) + \beta_{n-1,m}.$$

where for $j \geq 1$, $\beta_{j,m} := \frac{j}{q_{j,m}}$. By iteration,

$$\mathbb{E}(I_n) = q_{n,m} \sum_{j=1}^{n-1} \frac{\beta_{j,m}}{q_{j+1,m}}.$$

since $\mathbb{E}(I_1) = 0$. By definition of $\beta_{j,m}$,

$$\begin{aligned}
\mathbb{E}(I_n) &= q_{n,m} \sum_{j=1}^{n-1} \frac{j}{q_{j,m} q_{j+1,m}} \\
&= q_{n,m} \sum_{j=1}^{n-1} \frac{j}{m} \left(\frac{1}{q_{j,m}} - \frac{1}{q_{j+1,m}} \right) \\
&= \frac{q_{n,m}}{m} \left(\sum_{j=1}^{n-1} \frac{j}{q_{j,m}} - \sum_{j=1}^{n-1} \frac{j}{q_{j+1,m}} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{q_{n,m}}{m}\left(\sum_{j=1}^{n-1}\frac{j}{q_{j,m}} - \sum_{j=2}^{n}\frac{j-1}{q_{j,m}}\right)\\
&= \frac{q_{n,m}}{m}\left(1 - \frac{n}{q_{n,m}} + \sum_{j=2}^{n}\frac{1}{q_{j,m}}\right)\\
&= \frac{q_{n,m}}{m} - \frac{n}{m} + \frac{q_{n,m}}{m}\sum_{j=2}^{n}\frac{1}{q_{j,m}}\\
&= \frac{q_{n,m}}{m}\sum_{j=1}^{n}\frac{1}{q_{j,m}} - \frac{n}{m},
\end{aligned}
$$

and proof is completed. □

In the proof of Theorem 2, we will use the equality established in the following proposition.

Proposition 1. *[4] Suppose that $q_{j,m} = m(j-1)+1$ for $j \geq 1$. Then*

$$\sum_{j=3}^{n-1}\frac{\sum_{k=1}^{j-1}\frac{m(k-1)}{q_{k,m}^2}}{q_{j+1,m}q_{j+2,m}} = \frac{1}{q_{n+1,m}}\sum_{j=2}^{n-1}\frac{m(n-j)(j-1)}{q_{j,m}^2 q_{j+1,m}}.$$

Our approach to prove the theorem is the same as the approach used for X_n in [9] and [4].

Theorem 2. *There exists a limiting random variable I such that*

$$\frac{I_n - \frac{n}{m}\log n}{mn} \longrightarrow I,$$

in L^2 and almost surely.

Proof. We prove the theorem by showing that

$$W_n = \frac{I_n - \mathbb{E}(I_n)}{q_{n,m}}$$

is a martingale over the sequence of fields $\{\mathcal{F}_n\}_{n\geq 1}$ with uniformly bounded second moments. Absolute integrability of the sequence W_n for $n \geq 1$ is guaranteed by Theorem 1. But

$$
\begin{aligned}
\mathbb{E}(W_n|\mathcal{F}_{n-1}) &= \frac{\mathbb{E}(I_n|\mathcal{F}_{n-1}) - \mathbb{E}(I_n)}{q_{n,m}}\\
&= \frac{1}{q_{n,m}}\left(\frac{q_{n,m}}{q_{n-1,m}}I_{n-1} + \frac{n-1}{q_{n-1,m}} - \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I_{n-1}) - \frac{n-1}{q_{n-1,m}}\right)\\
&= \frac{I_{n-1} - \mathbb{E}(I_{n-1})}{q_{n-1,m}}\\
&= W_{n-1}.
\end{aligned}
$$

Thus the sequence W_n is a zero-mean martingale. We have $I_n = I_{n-1} + D_n$. Then

$$\begin{aligned} W_n &= \frac{I_n - \mathbb{E}(I_n)}{q_{n,m}} \\ &= \frac{I_{n-1} + D_n - \mathbb{E}(I_{n-1} + D_n)}{q_{n,m}} \\ &= \frac{q_{n-1,m}}{q_{n,m}} W_{n-1} + \frac{D_n - \mathbb{E}(D_n)}{q_{n,m}}, \end{aligned} \tag{2.2}$$

and by (2.1),

$$\begin{aligned} \mathbb{E}(D_n|\mathcal{F}_{n-1}) &= \mathbb{E}(I_n|\mathcal{F}_{n-1}) - I_{n-1} \\ &= \frac{q_{n,m}}{q_{n-1,m}} I_{n-1} + \frac{n-1}{q_{n-1,m}} - I_{n-1} \\ &= \frac{m}{q_{n-1,m}} I_{n-1} + \frac{n-1}{q_{n-1,m}} \\ &= \frac{m}{q_{n-1,m}} I_{n-1} + \frac{n-1}{q_{n-1,m}} \pm m \frac{\mathbb{E}(I_{n-1})}{q_{n-1,m}} \\ &= m W_{n-1} + \frac{m}{q_{n-1,m}} \mathbb{E}(I_{n-1}) + \frac{n-1}{q_{n-1,m}}. \end{aligned} \tag{2.3}$$

From (2.2), to compute the second moment of W_n, we have

$$\begin{aligned} \mathbb{E}(W_n^2) &= \left(\frac{q_{n-1,m}}{q_{n,m}}\right)^2 \mathbb{E}(W_{n-1}^2) + \frac{1}{q_{n,m}^2} \mathbb{V}(D_n) \\ &+ 2 \frac{q_{n-1,m}}{q_{n,m}^2} \mathbb{E}(W_{n-1}(D_n - \mathbb{E}(D_n))). \end{aligned}$$

From (2.3),

$$\begin{aligned} \mathbb{E}(W_{n-1}D_n) &= \mathbb{E}(W_{n-1}\mathbb{E}(D_n|\mathcal{F}_{n-1})) \\ &= \mathbb{E}\Bigg(W_{n-1}\Bigg(m W_{n-1} + \frac{m}{q_{n-1,m}} \mathbb{E}(I_{n-1}) + \frac{n-1}{q_{n-1,m}}\Bigg)\Bigg) \\ &= m \mathbb{E}(W_{n-1}^2). \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}(W_n^2) &= \left(\frac{q_{n-1,m}}{q_{n,m}}\right)^2 \mathbb{E}(W_{n-1}^2) + \frac{1}{q_{n,m}^2} \mathbb{V}(D_n) \\ &+ 2 \frac{q_{n-1,m}}{q_{n,m}^2} m \mathbb{E}(W_{n-1}^2) \\ &= \frac{q_{n-1,m} q_{n+1,m}}{q_{n,m}^2} \mathbb{E}(W_{n-1}^2) + \frac{1}{q_{n,m}^2} \mathbb{V}(D_n). \end{aligned} \tag{2.4}$$

Let us define

$$H_n = \frac{q_{n,m}}{q_{n+1,m}}\mathbb{E}(W_n^2), \qquad n \geq 1.$$

Then the recurrence (2.4) translate to the simple recurrence

$$H_n = H_{n-1} + \frac{1}{q_{n,m}q_{n+1,m}}\mathbb{V}(D_n).$$

By iteration and Proposition 1,

$$\begin{aligned} H_n &= \sum_{j=3}^{n-1} \frac{\sum_{k=1}^{j-1} \frac{m(k-1)}{q_{k,m}^2}}{q_{j+1,m}q_{j+2,m}} \\ &= \frac{1}{q_{n+1,m}} \sum_{j=2}^{n-1} \frac{m(n-j)(j-1)}{q_{j,m}^2 q_{j+1,m}}. \end{aligned}$$

Hence

$$\mathbb{E}(W_n^2) = \frac{1}{q_{n,m}} \sum_{j=2}^{n-1} \frac{m(n-j)(j-1)}{q_{j,m}^2 q_{j+1,m}}.$$

The sum in $\mathbb{E}(W_n^2)$ clearly converges. Convergence $a.s.$ and in L^2 follows from the martingale convergence theorem [3]. □

Corollary 1. *We have,*

$$\mathbb{V}(I_n) = q_{n,m}^2\mathbb{E}(W_n^2) = q_{n,m} \sum_{j=2}^{n-1} \frac{m(n-j)(j-1)}{q_{j,m}^2 q_{j+1,m}}$$

and

$$\mathbb{V}(X_n) = m^2\mathbb{V}(I_n) = q_{n,m} \sum_{j=2}^{n-1} \frac{m^3(n-j)(j-1)}{q_{j,m}^2 q_{j+1,m}}.$$

A standard argument based on Chebychev's inequality shows that $\frac{mI_n}{n\log n} \to 1$ *and* $\frac{X_n}{n\log n} \to 1$ *in probability. Also*

$$\mathbb{C}ov(I_n, X_n) = \mathbb{C}ov(I_n, mI_n + n) = m\mathbb{V}(I_n) = q_{n,m} \sum_{j=2}^{n-1} \frac{m^2(n-j)(j-1)}{q_{j,m}^2 q_{j+1,m}}.$$

From (2.1), $(q_{n,m}^{-1}(I_n - \mathbb{E}(I_n)))_{n\geq 1}$ *is a martingale over the sequence of fields* $\{\mathcal{F}_n\}_{n\geq 1}$. *Then*

$$\begin{aligned} \mathbb{C}ov(I_n, I_{n-1}) &= \mathbb{E}(\mathbb{E}((I_n - \mathbb{E}(I_n))(I_{n-1} - \mathbb{E}(I_{n-1})))|\mathcal{F}_{n-1})) \\ &= \frac{q_{n-1,m}}{q_{n,m}}\mathbb{V}(I_n) \\ &= q_{n-1,m} \sum_{j=2}^{n-1} \frac{m(n-j)(j-1)}{q_{j,m}^2 q_{j+1,m}}. \end{aligned}$$

3. The Sackin index

Since the Sackin index is only dependent on a part of the nodes (leaves), its study is more complex than the internal path length. Given tree T_{n-1} of size $n-1$, pick a node at random. If leaf v is picked, then tree T_n of size n is formed by making node n as a child of v and $S_n = S_{n-1} + 1$. If non-leaf v is picked, then $S_n = S_{n-1} + D_n$. Then, for each $v \in V(T_n)$ we define the indicator $I(v)$ as below:

$$I(v) = \begin{cases} 0, & d(v) = 0 \\ D(v), & d(v) \geq 1. \end{cases}$$

Now, if n is attached to a randomly chosen node U_{n-1} in T_{n-1}, then

$$S_n = S_{n-1} + I(U_{n-1}) + 1.$$

First, we determine the mean of this index according to the gamma function. It is not difficult to show that for each rooted tree of size n [7],

$$\sum_{v \in V(T_n)} D(v)^2 = 2I_n - (n-1) + \sum_{v \in V(T_n), d(v) \geq 1} d(v) D(v)^2.$$

Therefore, with very long calculations, it is possible to obtain the exact value of the variance of the Sackin index. To avoid long calculations, we consider two logical assumptions and based on them, we obtain a value of the variance of this index:
Assumption 1. There exists $k_n > 0$ such that $\mathbb{E}(I(U_{n-1})D(U_{n-1})) = k_n \mathbb{E}(I(U_{n-1}))$.
Assumption 2. There exists $r_n > 0$ such that $\mathbb{V}(I(U_{n-1})) = r_n \mathbb{V}(D_{n-1})$.
Using the gamma function define

$$\gamma_k[n,m] = \frac{\Gamma\left(n - 1 + \frac{1}{m}\right)}{\Gamma\left(n - 1 + \frac{1-k}{m}\right)}, \qquad k \geq 1, \quad n, m \geq 2 \ (n > m).$$

First, we prove the following lemma, which is vital for performing calculations.

Lemma 2. *For each integer $k \geq 1$,*
1)

$$\frac{\gamma_k[n-1,m]}{\gamma_k[n,m]} = \frac{q_{n-1,m} - k}{q_{n-1,m}}.$$

2)

$$\frac{\gamma_{2k}[n-1,m]}{\gamma_{2k}[n,m]} = 2\frac{\gamma_k[n-1,m]}{\gamma_k[n,m]} - 1.$$

3) The function $\gamma_k[n,m]$ is decreasing in n and

$$\gamma_k[n,m] = n^{\frac{k}{m}}(1 + \mathcal{O}(n^{-1})).$$

Proof. 1) The proof of (1) is obvious and straightforward, since $\Gamma(x+1) = x\Gamma(x)$. For part (2),

$$2\frac{\gamma_k[n-1,m]}{\gamma_k[n,m]} - 1 = \frac{q_{n-1,m} - 2k}{q_{n-1,m}} = \frac{\gamma_{2k}[n-1,m]}{\gamma_{2k}[n,m]}.$$

By Stirling formula,

$$n! = \sqrt{2\pi n}(n/e)^n(1 + O(n^{-1})).$$

Thus, part (3) is an immediate consequence of Stirling formula. □

Theorem 3. *The mean of S_n in a random m-oriented recursive tree of size n is given by*

$$\mathbb{E}(S_n) = \frac{1}{\gamma_1[n,m]} \sum_{j=1}^{n-1} \gamma_1[j+1,m]\mathbb{E}(D_j), \qquad n \geq 4.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}(S_n|\mathcal{F}_{n-1}) &= S_{n-1} + \mathbb{E}(I(U_{n-1})|\mathcal{F}_{n-1}) + 1 \\
&= S_{n-1} + \sum_{v \in V(T_{n-1})} \frac{1 + (m-1)d(v)}{q_{n-1,m}} I(v) + 1 \\
&= S_{n-1} + \frac{1}{q_{n-1,m}} \left[\sum_{v \in V(T_{n-1})} I(v) + \sum_{v \in V(T_{n-1})} (m-1)d(v)D(v) \right] + 1.
\end{aligned}$$

From Lemma 1,

$$\begin{aligned}
\mathbb{E}(S_n|\mathcal{F}_{n-1}) &= S_{n-1} + \frac{1}{q_{n-1,m}} \Big[(I_{n-1} - S_{n-1}) + (m-1)(I_{n-1} - (n-2) \Big] + 1 \\
&= S_{n-1} + \frac{1}{q_{n-1,m}} \Big[mI_{n-1} - S_{n-1} + (n-1) - q_{n-1,m} \Big] + 1.
\end{aligned}$$

Now, by $X_n = mI_n + n$ and (1.4),

$$\begin{aligned}
\mathbb{E}(S_n|\mathcal{F}_{n-1}) &= \frac{q_{n-1,m} - 1}{q_{n-1,m}} S_{n-1} + \frac{1}{q_{n-1,m}} (X_{n-1} - q_{n-1,m}) + 1 \\
&= \frac{q_{n-1,m} - 1}{q_{n-1,m}} S_{n-1} + \frac{1}{q_{n-1,m}} X_{n-1}. \qquad (3.1)
\end{aligned}$$

From Lemma 2, part (1),

$$\mathbb{E}(S_n|\mathcal{F}_{n-1}) = \frac{\gamma_1[n-1,m]}{\gamma_1[n,m]} S_{n-1} + \frac{1}{q_{n-1,m}} X_{n-1}. \qquad (3.2)$$

Let us define $\alpha_{n,m} = \sum_{j=1}^{n} \frac{1}{q_{j,m}} = \mathbb{E}(D_{n+1})$. Taking expectations of the relation (3.2) and (1.4), we have

$$\mathbb{E}(S_n) = \frac{\gamma_1[n-1,m]}{\gamma_1[n,m]}\mathbb{E}(S_{n-1}) + \alpha_{n-1,m}.$$

By iteration,

$$\mathbb{E}(S_n) = \frac{1}{\gamma_1[n,m]}\sum_{j=1}^{n-1} \gamma_1[j+1,m]\alpha_{j,m},$$

since $\mathbb{E}(S_1) = 0$ and proof is completed.

□

Remark 1. From Lemma 1, part (3), and (1.1),

$$\mathbb{E}(S_n) \sim n^{-\frac{1}{m}} \int_0^n \frac{1}{m} \log nx^{\frac{1}{m}}\, dx = \frac{n}{m+1}\log n.$$

In fact, $\frac{n}{m+1}\log n$ is an upper bound for $\mathbb{E}(S_n)$ as $n \to \infty$. Suppose S'_n is the sum of the depths of non-leaves of a random m-oriented recursive tree of size n. Then

$$\begin{aligned}\mathbb{E}(S'_n) &= \mathbb{E}(I_n) - \mathbb{E}(S_n)\\ &= \frac{q_{n,m}}{m}\sum_{j=1}^{n}\frac{1}{q_{j,m}} - \frac{n}{m} - \frac{1}{\gamma_1[n,m]}\sum_{j=1}^{n-1}\gamma_1[j+1,m]\mathbb{E}(D_j).\end{aligned}$$

From (1.3),

$$\mathbb{E}(S'_n) \sim \frac{n}{m(m+1)}\log n.$$

In fact, $\frac{n}{m(m+1)}\log n$ is a lower bound for $\mathbb{E}(S'_n)$ as $n \to \infty$. Thus, as $n \to \infty$,

$$\mathbb{E}(S_n) \sim \frac{m}{m+1}\mathbb{E}(I_n).$$

Theorem 4. *Under assumption 1, for $n \geq 3$,*

$$\mathbb{E}(I(U_{n-1})D_n) = (k_n+1)\left(\frac{q_{n,m}}{q^2_{n-1,m}}\mathbb{E}(X_{n-1}) - \frac{q_{n,m}}{q^2_{n-1,m}}\mathbb{E}(S_{n-1}) - \frac{q_{n,m}}{q_{n-1,m}}\right).$$

Proof. We have

$$\begin{aligned}
\mathbb{E}(I(U_{n-1})I_n) &= \mathbb{E}(I(U_{n-1})\mathbb{E}(I_n|\mathcal{F}_{n-1})) \\
&= \mathbb{E}\bigg(I(U_{n-1})\bigg(\frac{q_{n,m}}{q_{n-1,m}}I_{n-1} + \frac{n-1}{q_{n-1,m}}\bigg)\bigg) \\
&= \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})I_{n-1}) + \frac{n-1}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})) \\
&= \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})(I_n - D(U_{n-1}) - 1)) + \frac{n-1}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})) \\
&= \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})I_n) - \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})D(U_{n-1})) \\
&- \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})) + \frac{n-1}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})).
\end{aligned}$$

Then

$$\begin{aligned}
\frac{m}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})I_n) &= \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})D(U_{n-1})) \\
&+ \frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})) - \frac{n-1}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})). \qquad (3.3)
\end{aligned}$$

But

$$\begin{aligned}
\mathbb{E}(I(U_{n-1})D_n) &= \mathbb{E}(I(U_{n-1})\mathbb{E}(D_n|\mathcal{F}_{n-1})) \\
&= \mathbb{E}\bigg(I(U_{n-1})\bigg(\frac{m}{q_{n-1,m}}I_{n-1} + \frac{n-1}{q_{n-1,m}}\bigg)\bigg) \\
&= \frac{m}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})(I_n - D_n)) + \frac{n-1}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})) \\
&= \frac{m}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})I_n) - \frac{m}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})D_n) \\
&+ \frac{n-1}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})).
\end{aligned}$$

Then

$$\frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})D_n) = \frac{m}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})I_n) + \frac{n-1}{q_{n-1,m}}\mathbb{E}(I(U_{n-1})),$$

and thus

$$\mathbb{E}(I(U_{n-1})D_n) = \frac{m}{q_{n,m}}\mathbb{E}(I(U_{n-1})I_n) + \frac{n-1}{q_{n,m}}\mathbb{E}(I(U_{n-1})).$$

Now, from (3.3),

$$\begin{aligned}\mathbb{E}(I(U_{n-1})D_n) &= (k_n+1)\frac{q_{n,m}}{q_{n-1,m}}\mathbb{E}(I(U_{n-1}))\\ &= (k_n+1)\left(\frac{q_{n,m}}{q_{n-1,m}^2}\mathbb{E}(X_{n-1}) - \frac{q_{n,m}}{q_{n-1,m}^2}\mathbb{E}(S_{n-1}) - \frac{q_{n,m}}{q_{n-1,m}}\right).\end{aligned}$$

□

Set

$$\eta_{j,m} = \frac{\mathbb{E}(I_jX_j)}{q_{j,m}} + \frac{j}{q_{j,m}}\mathbb{E}(S_j) + \mathbb{E}(I(U_j)D_{j+1}) + \mathbb{E}(D_{j+1}), \quad j \geq 1.$$

Since $\mathbb{E}(I_jX_j) = \mathbb{C}ov\,(I_j, X_j) + \mathbb{E}(I_j)\mathbb{E}(X_j)$, all parts of $\eta_{j,m}$ are calculated.

Theorem 5. *Let I_n and S_n be the internal path length and Sackin index of a random m-oriented recursive tree of size n, respectively. For $n \geq 4$,*

$$\mathbb{C}ov\,(S_n, I_n) = \sum_{i=1}^{n-1}\eta_{i,m} - \mathbb{E}(S_n)\mathbb{E}(I_n).$$

Proof. For $n = 1, 2, 3$, $\mathbb{C}ov\,(S_n, I_n) = 0$. Assume $n \geq 4$. By definition,

$$\mathbb{C}ov\,(S_n, I_n) = \mathbb{E}(I_nS_n) - \mathbb{E}(S_n)\mathbb{E}(I_n),$$

where

$$\mathbb{E}(I_nS_n) = \mathbb{E}(S_n(I_{n-1} + D_n)) = \mathbb{E}(S_nI_{n-1}) + \mathbb{E}(S_nD_n).$$

From (3.1),

$$\begin{aligned}\mathbb{E}(S_nI_{n-1}) &= \mathbb{E}(I_{n-1}\mathbb{E}(S_n|\mathcal{F}_{n-1}))\\ &= \frac{q_{n-1,m}-1}{q_{n,m}}\mathbb{E}(I_{n-1}S_{n-1}) + \frac{1}{q_{n-1,m}}\mathbb{E}(I_{n-1}X_{n-1}).\end{aligned} \tag{3.4}$$

Also

$$\begin{aligned}\mathbb{E}(S_nD_n) &= \mathbb{E}((S_{n-1} + I(U_{n-1}) + 1)D_n)\\ &= \mathbb{E}(S_{n-1}D_n) + \mathbb{E}(I(U_{n-1})D_n) + \mathbb{E}(D_n).\end{aligned} \tag{3.5}$$

But

$$\begin{aligned}\mathbb{E}(S_{n-1}D_n) &= \mathbb{E}(S_{n-1}\mathbb{E}(D_n|\mathcal{F}_{n-1}))\\ &= \frac{m}{q_{n-1,m}}\mathbb{E}(S_{n-1}I_{n-1}) + \frac{n-1}{q_{n-1,m}}\mathbb{E}(S_{n-1}).\end{aligned} \tag{3.6}$$

Now, from (3.4), (3.5) and (3.6),

$$\begin{aligned}\mathbb{E}(S_nI_n) &= \frac{q_{n,m}}{q_{n,m}}\mathbb{E}(S_{n-1}I_{n-1}) + \frac{1}{q_{n-1,m}}\mathbb{E}(I_{n-1}X_{n-1})\\ &+ \frac{n-1}{q_{n-1,m}}\mathbb{E}(S_{n-1}) + \mathbb{E}(I(U_{n-1})D_n) + \mathbb{E}(D_n)\\ &= \vdots\\ &= \sum_{i=1}^{n-1}\left(\frac{\mathbb{E}(I_iX_i) + i\mathbb{E}(S_i)}{q_{i,m}} + \mathbb{E}(I(U_i)D_{i+1}) + \mathbb{E}(D_{i+1})\right)\\ &= \sum_{i=1}^{n-1}\eta_{i,m},\end{aligned}$$

since $\mathbb{E}(S_1I_1) = 0$ and proof is completed. □

Lemma 3. *For $n \geq 4$,*

$$\mathbb{C}ov(S_n, S_{n-1}) = \frac{\gamma_1[n-1,m]}{\gamma_1[n,m]}\mathbb{V}(S_{n-1}) + \frac{m}{q_{n-1,m}}\mathbb{C}ov(I_{n-1}, S_{n-1}).$$

Proof. For $n = 1, 2, 3$, $\mathbb{C}ov\,(S_n, S_{n-1}) = 0$. We have

$$\begin{aligned}\mathbb{E}(S_n - \mathbb{E}(S_n)|\mathcal{F}_{n-1}) &= \frac{\gamma_1[n-1,m]}{\gamma_1[n,m]}(S_{n-1} - \mathbb{E}(S_{n-1})) + \frac{X_{n-1} - \mathbb{E}(X_{n-1})}{q_{n-1,m}}\\ &= \frac{\gamma_1[n-1,m]}{\gamma_1[n,m]}(S_{n-1} - \mathbb{E}(S_{n-1})) + m\frac{I_{n-1} - \mathbb{E}(I_{n-1})}{q_{n-1,m}}.\end{aligned}$$

Then

$$\begin{aligned}\mathbb{C}ov(S_n, S_{n-1}) &= \mathbb{E}(\mathbb{E}((S_n - \mathbb{E}(S_n))(S_{n-1} - \mathbb{E}(S_{n-1})))|\mathcal{F}_{n-1}))\\ &= \mathbb{E}((S_{n-1} - \mathbb{E}(S_{n-1}))\mathbb{E}(S_n - \mathbb{E}(S_n))|\mathcal{F}_{n-1}))\\ &= \frac{\gamma_1[n-1,m]}{\gamma_1[n,m]}\mathbb{V}(S_{n-1}) + \frac{m}{q_{n-1,m}}\mathbb{C}ov(I_{n-1}, S_{n-1}).\end{aligned}$$

□

Theorem 6. *Under assumption 2, for $n \geq 4$,*

$$\mathbb{V}(S_n) = \frac{1}{\gamma_2[n,m]}\sum_{j=3}^{n-1}\gamma_2[j+1,m]\mu_{j,m}$$

where

$$\mu_{j,m} = \frac{2m}{q_{j,m}}\ \mathbb{C}ov(S_j, I_j) + r_n\mathbb{V}(D_j), \qquad j \geq 1.$$

Proof. Since the empty sum is zero, the result holds for $n \in \{1, 2, 3\}$ and we can assume that $n \geq 4$. We have

$$\mathbb{E}(S_n - S_{n-1} - 1)^2 \;=\; \mathbb{E}(I^2(U_{n-1})). \tag{3.7}$$

From Lemma 3,

$$\begin{aligned}
& \mathbb{E}(S_n - S_{n-1} - 1)^2 \\
= \; & \mathbb{E}(S_n - \mathbb{E}(S_n) - S_{n-1} + \mathbb{E}(S_{n-1}) + \mathbb{E}(S_n) - \mathbb{E}(S_{n-1}) - 1)^2 \\
= \; & \mathbb{E}(S_n - \mathbb{E}(S_n) - S_{n-1} + \mathbb{E}(S_{n-1}))^2 \\
+ \; & \mathbb{E}(\mathbb{E}(S_n) - \mathbb{E}(S_{n-1}) - 1)^2 \\
+ \; & 2\mathbb{E}\Big((S_n - \mathbb{E}(S_n) - S_{n-1} + \mathbb{E}(S_{n-1})) \;\; (\mathbb{E}(S_n) - \mathbb{E}(S_{n-1}) - 1)\Big) \\
= \; & \mathbb{V}ar(S_n) + \Big(1 - 2\frac{\gamma_1[n-1, m]}{\gamma_1[n, m]}\Big)\mathbb{V}ar(S_{n-1}) \\
- \; & \frac{2m}{q_{n-1,m}} \, \mathbb{C}ov(S_{n-1}, I_{n-1}) + \mathbb{E}^2(I(U_{n-1})).
\end{aligned} \tag{3.8}$$

Now, from (3.7) and (3.8) we see that

$$\begin{aligned}
\mathbb{V}ar(S_n) \;&=\; \Big(2\frac{\gamma_1[n-1, m]}{\gamma_1[n, m]} - 1\Big)\mathbb{V}ar(S_{n-1}) + \mu_{n-1,m} \\
&=\; \frac{\gamma_2[n-1, m]}{\gamma_2[n, m]}\mathbb{V}ar(S_{n-1}) + \mu_{n-1,m}.
\end{aligned}$$

By iteration, proof is completed because $\mathbb{V}ar(S_3) = 0$. □

Corollary 2. *Under assumptions 1 and 2, and Corollary 1, $\mathbb{V}(S_n) = O(n^2)$. Then, based on Chebychev's inequality,*

$$\frac{(m+1)S_n}{n \log n} \to 1,$$

in probability.

Conflicts of Interest: The authors have no competing interests to declare that are relevant to the content of this article.

Consent for Publication: Authors declare the consent for manuscript publication.

Availability of data and materials: No data were used to support this study.

Funding: Not applicable. No organization has supported this research.

Authors' Contributions: All authors read and approved the final manuscript.

References

[1] T.M. Coronado, A. Mir, F. Rossello, and L. Rotger, *On Sackin's original proposal: the variance of the leaves' depths as a phylogenetic balance index*, BMC Bioinformatics **21** (2020), no. 1, 154.
https://doi.org/10.1186/s12859-020-3405-1.

[2] R.P. Dobrow and R.T. Smythe, *Poisson approximations for functionals of random trees*, Random Structures and Algorithms **9** (1996), no. 1-2, 79–92.

[3] P. Hall and C.C. Heyde, *Martingale Limit Theory and its Application*, Academic Press, New York, 1980.

[4] M. Javanian and M.Q. Vahidi-Asl, *External path length of random m-oriented recursive trees*, J. Natural Sci. Math. **50** (2010), no. 1-2, 11–18.

[5] R. Kazemi, *On the multiplicative Zagreb indices of bucket recursive trees*, Iranian J. Math. Chem. **8** (2017), no. 1, 37–45.
https://doi.org/10.22052/ijmc.2017.15385.

[6] ———, *Total path length and Sackin index of random recursive trees*, 14th Iranian Statistics Conference, 2018, pp. 420–425.

[7] R. Kazemi and A. Behtoei, *The moments of the Sackin index of random d-ary increasing trees*, Mathematicki Vesnik **73** (2021), no. 1, 55–62.

[8] M.C. King and N.A. Rosenberg, *A simple derivation of the mean of the Sackin index of tree balance under the uniform model on rooted binary labeled trees*, Math. Biosci. **342** (2021), 108688.
https://doi.org/10.1016/j.mbs.2021.108688.

[9] H.M. Mahmoud, *Distances in random plane-oriented recursive trees*, J. Comput. Appl. Math. **41** (1992), no. 1-2, 237–245.
https://doi.org/10.1016/0377-0427(92)90252-S.

[10] A. Mir, F. Rossello, and L. Rotger, *A new balance index for phylogenetic trees*, Math. Biosci. **241** (2013), no. 1, 125–136.
https://doi.org/10.1016/j.mbs.2012.10.005.

[11] B. Pittel, *Note on the heights of random recursive trees and random m-ary search trees*, Random Structures and Algorithms **5** (1994), no. 2, 337–347.
https://doi.org/10.1002/rsa.3240050207.

[12] M.J. Sackin, *Good and bad phenograms*, Systematic Biology **21** (1972), no. 2, 225–226.
https://doi.org/10.1093/sysbio/21.2.225.