Probability in the Engineering and Informational Sciences, **35**, 2021, 839–857.
doi:10.1017/S0269964820000261

ON SEVERAL PROPERTIES OF A CLASS OF PREFERENTIAL ATTACHMENT TREES—PLANE-ORIENTED RECURSIVE TREES

PANPAN ZHANG

Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

E-mail: panpan.zhang@pennmedicine.upenn.edu

In this paper, several properties of a class of trees presenting preferential attachment phenomenon—plane-oriented recursive trees (PORTs) are uncovered. Specifically, we investigate the degree profile of a PORT by determining the exact probability mass function of the degree of a node with a fixed label. We compute the expectation and the variance of degree variable via a Pólya urn approach. In addition, we study a topological index, Zagreb index, of this class of trees. We calculate the exact first two moments of the Zagreb index (of PORTs) by using recurrence methods. Lastly, we determine the limiting degree distribution in PORTs that grow in continuous time, where the embedding is done in a Poissonization framework. We show that it is exponential after proper scaling.

Keywords: degree distribution, plane-oriented recursive trees, Poissonization, Pólya urn, preferential attachment, Zagreb index

1. INTRODUCTION

In graph theory, a tree refers to a connected structure with no cycles [6, p. 24]. A random recursive tree is an unordered labeled tree with label set such that there exists an increasing unique path from the root (the most primitive node labeled with 1) to the node labeled with j for all $2 \le j \le n$. This class of uniform recursive trees was proposed in the late 1960s and has found applications in a plethora of areas, such as spread of epidemics [29], genealogy [30], and the pyramid scheme [18].

In this paper, we consider a class of nonuniform random recursive trees—plane-oriented recursive trees (PORTs). A plane-ordered recursive tree is a tree in which descendants of each node are ordered. At time $n \ge 1$, we denote the structure of a PORT as T_n, that is, a PORT consisting of n nodes. The tree T_n is obtained by starting with a single node labeled with 1 (i.e., root). Upon each insertion point $n \ge 2$, a node labeled with n joins into the tree, connected by a directed edge emanating out of an existing node, which results in an increase of the outdegree of the selected node by 1 as well as an increase of the indegree of the newcomer (the node labeled with n) by 1; the probability of the newcomer adjoint to the node labeled with i, for $1 \le i \le n-1$, in T_{n-1} is proportional to the degree of the recruiter (the node labeled with i). The key feature of this class of trees is that a parent node (recruiter) with higher degrees is more attractive to newcomers, which coincides with a manifestation of the economic principles—"the rich get richer" and "success breeds success."

Precursory research on PORTs traced back to the late 1980s and the early 1990s. The exact and asymptotic moments of two degree profile random variables, the number of nodes of a given degree and the degree of a fixed node, were investigated in [35]. The distribution of the depth of nodes was determined by Mahmoud [25]. The exact and asymptotic distribution of leaves (terminal vertices) in PORTs and subtrees (branches) were studied by Mahmoud *et al.* [28]. The asymptotic average of internal path length was characterized by Chen and Ni [10]. Several concentration results were developed by Lu and Feng [24]. More recently, PORTs again caught researchers' attention since its evolutionary characteristic coincides with a network property of great interest in the community—preferential attachment [5]. Hence, PORTs are also known as preferential attachment trees in the literature; e.g., [20,23]. The joint asymptotic distribution of the numbers of nodes of different outdegrees in PORTs was shown to be normal by Drmota *et al.* [11] and Janson [22]. Several other limiting results for PORTs were presented in Hwang [21]. A slightly different random graph model accounting for self loops was considered in Bollobás *et al.* [8], where the degree distribution was determined and a power-law was developed. In Avrachenkov and Lebedev [3], the authors studied a generalized network model with applications to webpage ranking called PageRank; specifically, the distribution of subtree sizes, the heights of nodes in subtrees, and several distance-based properties were investigated.

The rest of this paper is organized as follows. We begin with introducing some notations and preliminaries in Section 2. In Section 3, we determine the exact distribution of the degree of a node with a fixed label. More specifically, we develop the probability mass function via an elementary approach—two-dimensional induction—in Section 3.1 and calculate its moments by exploiting a Pólya urn model in Section 3.2. In Section 4, we look into the Zagreb index for this class of trees. This section is split into two parts. In Section 4.1, we compute the mean and variance of the Zagreb index of PORTs via recurrence methods, while in Section 4.2, we study the convergence of the Zagreb index. In Section 5, we investigate the degree profile of PORTs embedded into continuous time. We find that the asymptotic distribution of the degree variable under the Poissonization framework is exponential. Lastly, we address some concluding remarks and propose some future work in Section 6.

2. NOTATIONS AND PRELIMINARIES

Let $D_{n,j}$ be the degree of the node with label j in T_n, for $1 \le j \le n$. Let $\mathbb{F}_n$ denote the σ-field generated by the first n stages of T_n. Many results in this paper are given in terms of gamma functions, $\Gamma(\cdot)$; see a classic text [12, p. 47] for its definition and fundamental properties. For a nonnegative integer z, the double factorial of z is $z!! = \prod_{i=0}^{\lceil z/2 \rceil - 1}(z-2i)$, with the interpretation of $0!! = 1$. The Pochhammer symbol for the rising factorial is defined as

$$\langle x \rangle_k = x(x+1)\cdots(x+k-1)$$

for any real x and nonnegative integer k, with the interpretation of $\langle x \rangle_0 = 1$. The Kronecker delta function of two variables s and t, denoted by $\delta_{s,t}$, equals 1 for $s = t$; 0, otherwise. The little o and big O notations define relations between two real-valued functions $f(x)$ and $g(x)$. We have $f(x) = o(g(x))$ equivalent to $\lim_{x\to\infty}(f(x)/g(x)) = 0$ provided that $g(x) \neq 0$; on the other hand, $f(x) = O(g(x))$ if there exists $M > 0$ and $x_0 \in \mathbb{R}$ such that $|f(x)| \le M|g(x)|$ for all $x \ge x_0$. Generalized hypergeometric functions are defined in terms of Pochhammer symbols of rising factorials; that is,

$$ {}_pF_q(a_1 \ldots, a_p; b_1 \ldots, b_q; z) = \sum_{s=0}^{\infty} \frac{\langle a_1 \rangle_s \cdots \langle a_p \rangle_s}{\langle b_1 \rangle_s \cdots \langle b_q \rangle_s} \frac{z^s}{s!}. $$

Much of the study in this paper relies on an extensively studied probabilistic model—Pólya urn model. We give quick words about Pólya urns. A two-color Pólya urn scheme is an urn containing balls of two different colors (say white and blue). At each point of discrete time, we draw a ball from the urn at random, observe its color and put it back in the urn, then execute some ball additions (or removals) according to predesignated rules: If the ball withdrawn is white, we add a white balls and b blue balls; otherwise, the ball withdrawn is blue, in which case we add c white balls and d blue balls. The dynamics of the urn can, thus, be represented by the following replacement matrix:

$$ \begin{pmatrix} a & b \\ c & d \end{pmatrix}, $$

in which the rows from top to bottom are indexed by white and blue, and the columns from left to right are also indexed by white and blue. We refer the interested readers to [26] for a text-style elaboration of Pólya urns.

3. DEGREE DISTRIBUTION

In this section, we investigate the degree profile of PORTs, that is, the distribution of the degree variable $D_{n,j}$ for a fixed $1 \le j \le n$. First, we determine the distribution of $D_{n,j}$ by developing the exact expression of its probability mass function. Next, we characterize its behavior by looking into the first two moments.

3.1. Probability Mass Function

To determine the probability mass function of $D_{n,j}$, we separate the cases of $\{j \ge 2\}$ and $\{j = 1\}$ for clarity. When $j = 1$, the random variable $D_{n,j} = D_{n,1}$ refers to the degree of the root of T_n. The root is the originator of the tree, so it has no parent. The root is the only node in the tree that has indegree 0.

Proposition 1: *For a fixed* $2 \le j \le n$, *we have*

$$ \mathbb{P}(D_{n,j} = d) = \frac{\Gamma(d)\Gamma\left(j - \frac{1}{2}\right) \sum_{i=0}^{d-1} \frac{(-1)^i \Gamma\left(n-1-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d-i)\Gamma\left(j-1-\frac{i}{2}\right)}}{\Gamma\left(n - \frac{1}{2}\right)}, \qquad (1) $$

for $d = 1, 2, \ldots, n - j + 1$.

Proof: We prove the proposition by a two-dimensional induction on $n \ge j$ and $d \ge 1$. The proof progresses in the style of filling an infinite lower triangular table, in which the rows are indexed by n and the columns are indexed by d. A (similar) graphic interpretation of the method can be found in [37, p. 69]. We initialize the first column and the diagonal of the table to be the basis of the induction. The event of $\{D_{n,j} = 1\}$ for all $n \ge j$ is that the node labeled with j is never chosen as a parent for any newcomer since its first appearance in the tree till time n. Thus, we have

$$\mathbb{P}(D_{n,j}=1)=\frac{2j-2}{2j-1}\times\frac{2j}{2j+1}\times\cdots\times\frac{2n-4}{2n-3}=\frac{\Gamma(n-1)\Gamma\left(j-\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)\Gamma(j-1)}.$$

On the other hand, the event of $\{D_{n,j}=n-j+1\}$ for all $n\ge j$ is that the node labeled with j is selected as parents for newcomers at all points from $j+1$ to n. It follows that

$$\begin{aligned}\mathbb{P}(D_{n,j}=n-j+1)&=\frac{1}{2j-1}\times\frac{2}{2j+1}\times\cdots\times\frac{n-j}{2n-3}\\&=\frac{\Gamma(n-j+1)\Gamma\left(j-\frac{1}{2}\right)}{2^{n-j}\Gamma\left(n-\frac{1}{2}\right)}.\end{aligned}$$

We assume that Eq. (1) holds for all d up to row $(n-1)$ in the table. Noticing that the degree of the node labeled with j increases at most by one at each point, we have

$$\begin{aligned}\mathbb{P}(D_{n,j}=d)&=\frac{d-1}{2n-3}\mathbb{P}(D_{n-1,j}=d-1)+\frac{2n-3-d}{2n-3}\mathbb{P}(D_{n-1,j}=d)\\&=\frac{d-1}{2n-3}\,\frac{\Gamma(d-1)\Gamma\left(j-\frac{1}{2}\right)\sum_{i=0}^{d-2}\frac{(-1)^i\Gamma\left(n-2-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d-1-i)\Gamma\left(j-1-\frac{i}{2}\right)}}{\Gamma\left(n-\frac{3}{2}\right)}\\&\quad+\frac{2n-3-d}{2n-3}\,\frac{\Gamma(d)\Gamma\left(j-\frac{1}{2}\right)\sum_{i=0}^{d-1}\frac{(-1)^i\Gamma\left(n-2-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d-i)\Gamma\left(j-1-\frac{i}{2}\right)}}{\Gamma\left(n-\frac{3}{2}\right)}\\&=\frac{\Gamma(d)\Gamma\left(j-\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)}\left[\frac{1}{2}\sum_{i=0}^{d-2}\frac{(-1)^i\Gamma\left(n-2-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d-1-i)\Gamma\left(j-1-\frac{i}{2}\right)}\right.\\&\quad\left.+\left(n-\frac{d}{2}-\frac{3}{2}\right)\sum_{i=0}^{d-1}\frac{(-1)^i\Gamma\left(n-2-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d-i)\Gamma\left(j-1-\frac{i}{2}\right)}\right]\\&=\frac{\Gamma(d)\Gamma\left(j-\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)}\left[\sum_{i=0}^{d-2}\left(n-2-\frac{i}{2}\right)\frac{(-1)^i\Gamma\left(n-2-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d-i)\Gamma\left(j-1-\frac{i}{2}\right)}\right.\\&\quad\left.+\left(n-\frac{d}{2}-\frac{3}{2}\right)\frac{(-1)^{d-1}\Gamma\left(n-\frac{d}{2}-\frac{3}{2}\right)}{\Gamma(d)\Gamma\left(j-\frac{d}{2}-\frac{3}{2}\right)}\right].\end{aligned}$$

This is equivalent to Eq. (1) stated in the proposition. ■

The probability mass function of $D_{n,j}$ is given by the sum of an alternating sequence. We split the total sum into two parts: a partial sum of odd indices and a partial sum of even indices. We then, respectively, evaluate the two partial sums to obtain an alternative expression of the probability mass function of $D_{n,j}$. The result is given in terms of generalized hypergeometric functions, presented in the next corollary.

COROLLARY 2: *For a fixed $2 \le j \le n$, we have*

$$\mathbb{P}(D_{n,j} = d) = \frac{\Gamma(d)\Gamma\left(j-\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)} \left(\frac{\Gamma(n-1)_3F_2 \frac{2-d}{2}, \frac{1-d}{2}, 2-j\frac{1}{2}, 2-n1}{\Gamma(d)\Gamma(j-1)} - \frac{\Gamma(n-\frac{3}{2})_3F_2 \frac{3-d}{2}, \frac{2-d}{2}, \frac{5}{2}-j\frac{3}{2}, \frac{5}{2}-n1}{\Gamma(d-1)\Gamma\left(j-\frac{3}{2}\right)} \right).$$

The first generalized hypergeometric function (on the top row) in the result stated in Corollary 1 can be further simplified for small choices of j. We present the probability mass functions of $D_{n,j}$ for $j = 2,3$ as examples:

$$\mathbb{P}(D_{n,2} = d) = \frac{1}{(2n-3)\Gamma\left(n-\frac{1}{2}\right)} \left[\sqrt{\pi}\left(n-\frac{3}{2}\right)\Gamma(n-1) - (d-1)\Gamma\left(n-\frac{1}{2}\right) {}_3F_2 \frac{3-d}{2}, \frac{2-d}{2}, \frac{1}{2}\frac{3}{2}, \frac{5}{2}-n1 \right];$$

$$\mathbb{P}(D_{n,3} = d) = \frac{3}{(2n-3)\Gamma\left(n-\frac{1}{2}\right)} \left[\sqrt{\pi}\left(n-\frac{3}{2}\right)\frac{d^2-3d+2n-2}{4}\Gamma(n-2) - (d-1)\Gamma\left(n-\frac{1}{2}\right) {}_3F_2 \frac{3-d}{2}, \frac{2-d}{2}, -\frac{1}{2}\frac{3}{2}, \frac{5}{2}-n1 \right].$$

Simplifications for the probability mass function of $D_{n,j}$ for higher values of j are also available, done in a similar manner.

Next, we look at the degree distribution of the root of a PORT. For $j = 1$, the probability mass function of $D_{n,j}$ (i.e., $D_{n,1}$) cannot be directly derived from Eq. (1). Notice that the main difference between the root and other nodes is that the root has indegree 0, while each of the other nodes has indegree 1. Thus, we can tweak Eq. (1) by substituting d by $d+1$, and then letting $j = 1$. Under such setting, we find that the probability mass function of $D_{n,1}$ can be substantially simplified to the following neat and closed form.

PROPOSITION 3: *The probability mass function of the root of a PORT is given by*

$$\mathbb{P}(D_{n,1} = d) = \frac{d(2n-d-3)!}{2^{n-d-1}(n-d-1)!(2n-3)!!}, \tag{2}$$

for $d = 1, 2, \ldots, n-1$.

PROOF: Recall Eq. (1), and set $j = 1$. Replacing d with $d+1$ in the equation, we have

$$\mathbb{P}(D_{n,1} = d) = \frac{\Gamma(d+1)\Gamma\left(\frac{1}{2}\right)\sum_{i=0}^{d} \frac{(-1)^i\Gamma\left(n-1-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d+1-i)\Gamma\left(j-1-\frac{i}{2}\right)}}{\Gamma\left(n-\frac{1}{2}\right)}.$$

Reimplementing the strategy of writing the total sum into partial sums with respect to odd indicies and even indicies, we apply the Euler's reflection formula to gamma functions

https://doi.org/10.1017/S0269964820000261 Published online by Cambridge University Press

and obtain

$$\begin{aligned}\mathbb{P}(D_{n,1}=d) &= \frac{\Gamma(d+1)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)}\left(\sum_{\substack{i \text{ is even}\\ 0\le i\le d}}\frac{\Gamma\left(n-1-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d+1-i)\Gamma\left(j-1-\frac{i}{2}\right)}\right.\\ &\quad \left. -\sum_{\substack{i \text{ is odd}\\ 0\le i\le d}}\frac{\Gamma\left(n-1-\frac{i}{2}\right)}{\Gamma(i+1)\Gamma(d+1-i)\Gamma\left(j-1-\frac{i}{2}\right)}\right)\\ &= \frac{\Gamma(d+1)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)}\frac{2^{d+1}\Gamma(d+1-n)}{4^n\Gamma(d)\Gamma(d+3-2n)\cos(n\pi)}\\ &= \frac{d(2n-d-3)\Gamma(2n-d-3)2^n}{2^{2n-d-1}(n-d-1)\Gamma(n-d-1)(2n-3)!!}\\ &= \frac{d(2n-d-3)!}{2^{n-d-1}(n-d-1)!(2n-3)!!}.\end{aligned}$$

■

The probability mass function of $D_{n,1}$ in Proposition 2 agrees with that derived in [38]. The proof in [38] requires massive algebraic computations and simplifications, whereas the proof given in this paper appears more concise and succinct.

3.2. Moments

In general, the probability mass function (c.f. Eq. (1)) is unwieldy for moment computations. Alternatively, we appeal to a two-color Pólya urn model [26] to calculate the mean and variance of $D_{j,n}$. Imagine that there is an urn containing balls of two colors (white and blue). Let W_n be the degree of the node labeled with j (white balls) at time $n \ge j$, and B_n be the total degree of all the other nodes (blue balls). At time $(n+1)$, if the node labeled with j is selected, W_n increases by one, and B_n also increases by one, which is contributed by the edge incident to the node labeled with $(n+1)$; if any other node is selected, B_n increases by two. This dynamic can be represented by the following replacement matrix:

$$\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}. \tag{3}$$

This Pólya urn scheme appropriately interprets the mechanism of preferential attachment, as, upon the insertion at time point $n+1$, the probability of the node labeled with j being selected is exactly $W_n/(W_n+B_n)$. Another equivalent approach to modeling the dynamics of degree change is to employ an extended PORT. The basic idea is to fill all the gaps in the original tree with external nodes, which represent insertion positions. We omit the details in this section but will revisit this strategy in the sequel.

The replacement matrix (c.f. Matrix (3)) is triangular, so the Pólya urn associated with this kind of replacement matrix is called triangular Pólya urn. Triangular urns are well studied, and the moments of white balls are explicitly characterized in [42, Thm. 3.1]. We exploit those results to get the following proposition.

Proposition 4: *For a fixed* $1 \le j \le n$ *and* $n \ge 2$*, we have*

$$\mathbb{E}[D_{n,j}] = \frac{\Gamma(n)\Gamma\left(j-\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)\Gamma(j)} - \delta_{j,1},$$

$$\mathbb{V}\mathrm{ar}[D_{n,j}] = -\frac{\Gamma^2(n)\Gamma^2\left(j-\frac{1}{2}\right)}{\Gamma^2\left(n-\frac{1}{2}\right)\Gamma^2(j)} - \frac{\Gamma(n)\Gamma\left(j-\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)\Gamma(j)} + \frac{4n-2}{2j-1}.$$

We discover that when n is large, both $\mathbb{E}[D_{n,j}]$ and $\mathbb{V}\mathrm{ar}[D_{n,j}]$ experience phase transitions. To compute the asymptotic expectation and variance, we apply the Stirling's approximation to the expectation and variance of $D_{n,j}$ in Proposition 3. As $n \to \infty$, we have

$$\mathbb{E}[D_{n,j}] \sim \frac{\Gamma\left(j-\frac{1}{2}\right)}{\Gamma(j)} n^{1/2}, \tag{4}$$

$$\mathbb{V}\mathrm{ar}[D_{n,j}] \sim \left(\frac{4}{2j-1} - \frac{\Gamma^2\left(j-\frac{1}{2}\right)}{\Gamma^2(j)}\right) n - \frac{\Gamma\left(j-\frac{1}{2}\right)}{\Gamma(j)} n^{1/2}, \tag{5}$$

where "$\sim$" is a standard notation standing for "asymptotic equivalence"; that is, given well-defined real-valued functions $f(n)$ and $g(n)$, $f(n) \sim g(n)$ is equivalent to $\lim_{n\to\infty}(f(n)/g(n)) = 1$. We keep the second highest order term (i.e., the term that involves $n^{1/2}$) in the asymptotic variance of $D_{n,j}$ because it makes a contribution when j grows in the linear phase (with respect to n). We reapply the Stirling's approximation to Eqs. (4) and (5), respectively, and obtain the next corollary.

Corollary 5: *As* $n \to \infty$*, we have*

$$\mathbb{E}[D_{n,j}] \sim \begin{cases} (\Gamma(j-1/2)/\Gamma(j))n^{1/2}, & \text{for fixed } j, \\ (n/j)^{1/2}, & \text{for } j \to \infty, \end{cases}$$

and

$$\mathbb{V}\mathrm{ar}[D_{n,j}] \sim \begin{cases} \left(\frac{4}{2j-1} - \frac{\Gamma^2(j-1/2)}{\Gamma^2(j)}\right) n, & \text{for fixed } j, \\ n/j, & \text{for } j \to \infty,\ j = o(n), \\ 1/\theta - 1/\sqrt{\theta}, & \text{for } j/n = \theta,\ 0 < \theta < 1. \end{cases}$$

The formulation of $\mathbb{E}[D_{n,j}]$ coincides with that developed in [35]. In addition, $\mathbb{V}\mathrm{ar}[D_{n,j}]$ is also reported in [35], where it is presented in terms of a sum of binomial coefficients. In this paper, we provide an alternative approach to determining $\mathbb{E}[D_{n,j}]$ and $\mathbb{V}\mathrm{ar}[D_{n,j}]$, and both of them are in neat and closed forms.

4. ZAGREB INDEX

A topological index of a graph quantifies it by turning its structure into a number. Capturing structures in numbers allows researchers to compare graphs according to certain criteria. There are many possible indices that can be constructed for static and random graphs. Each index tends to capture certain features of the graphs, such as sparseness, regularity, and centrality. Examples of indices that have been introduced for random graphs include the Zagreb index [13], the Randíc index [16], the Wiener index [17,31], the Gini index [4,40], and a topological index measuring graph weight [41].

In this section, we investigate the Zagreb index for the class of PORTs. Zagreb index was first introduced by Gutman and Trinajstić [19] in 1972. It has been a popular topological index to study molecules and complexity of selected classes of molecules in mathematical chemistry [32] and to model quantitative structure–property relationship (QSPR) and quantitative structure–activity relationship (QSAR) in chemoinformatics [36]. We refer the readers to [33] and the references therein for a concise review of the Zagreb index and its applications.

In the field of (random) graph theory, a series of research papers were recently produced to look into the Zagreb index of various classes of random trees, including random recursive trees (RRTs) [13], scale-free trees [14], and (generalized) b-ary recursive trees [15]. Motivated by these sources, we compute the Zagreb index of PORTs in this section and compare the result with that of RRTs.

4.1. Mean and Variance

The Zagreb index of a graph is defined as the sum of the squared degrees of all the nodes therein. Given a PORT at time n, T_n, its Zagreb index is thus given by

$$Z_n = \mathbf{Zagreb}(T_n) = \sum_{j=1}^{n} D_{n,j}^2,$$

where $D_{n,j}$, again, is the degree of the node labeled with j in T_n. Let $\mathbb{I}(n,j)$ indicate the event that the node labeled with j is selected at time n. In the next proposition, we present the exact expectation of Z_n as well as a weak law.

Proposition 6: *The mean of the Zagreb index of a PORT at time $n \geq 1$ is given by*

$$\mathbb{E}[Z_n] = 2(n-1)(\Psi(n) + \gamma),$$

where $\Psi(\cdot)$ is the digamma function, and γ is the Euler's constant. As $n \to \infty$, we have

$$\frac{Z_n}{n \log n} \xrightarrow{P} 2.$$

Proof: Upon the insertion taking place at time point n, we have the following recurrence of Z_n conditional on $\mathbb{F}_{n-1}$ and $\mathbb{I}(n,j)$:

$$Z_n = Z_{n-1} + (D_{n-1,j} + 1)^2 - D_{n-1,j}^2 + 1, \tag{6}$$

where the terms $((D_{n-1,j} + 1)^2 - D_{n-1,j}^2)$ altogether indicate the contribution by the node labeled with j (to the Zagreb index) by the degree increase, and the last term 1 comes from the contribution by the newcomer (the node labeled with n). We simplify Eq. (6) and take

https://doi.org/10.1017/S0269964820000261 Published online by Cambridge University Press

the expectation with respect to $\mathbb{I}(n, j)$ to get

$$\begin{aligned}\mathbb{E}[Z_n \,|\, \mathbb{F}_{n-1}] &= Z_{n-1} + 2\sum_{j=1}^{n-1} D_{n-1,j} \times \mathbb{P}(\mathbb{I}(n,j)) + 2 \\ &= Z_{n-1} + 2\sum_{j=1}^{n-1} D_{n-1,j} \times \frac{D_{n-1,j}}{2(n-2)} + 2 \\ &= Z_{n-1} + \frac{\sum_{j=1}^{n-1} D_{n-1,j}^2}{n-2} + 2 \\ &= \left(1 + \frac{1}{n-2}\right) Z_{n-1} + 2.\end{aligned}$$

Taking another expectation on both sides, we receive a recurrence on the mean of Z_n, namely

$$\mathbb{E}[Z_n] = \frac{n-1}{n-2}\mathbb{E}[Z_{n-1}] + 2.$$

This recurrence is well defined for $n \geq 3$, so we can set the initial condition at $\mathbb{E}[Z_2] = Z_2 = 2$. Solving the recurrence, we obtain the result stated in the proposition. Notice that the result is well defined for all $n \geq 1$, albeit the developed recurrence is undefined for $n = 2$.

As $n \to \infty$, we have $\Psi(n) \sim \log n$. Hence, we obtain the following convergence in L_1-space:

$$\frac{Z_n}{n \log n} \xrightarrow{L_1} 2.$$

This convergence takes place in probability as well. ■

Toward the computation of the second moment of Z_n, we consider a new topological index that is the sum of cubic degrees of nodes in a graph. Let $Y_n = \sum_{j=1}^{n} D_{n,j}^3$ be such an index of T_n. In the next lemma, we derive the mean of Y_n, and a weak law as well.

LEMMA 7: *The mean of Y_n of a PORT at time $n \geq 2$ is given by*

$$\mathbb{E}[Y_n] = \frac{32\Gamma(n+1/2)}{\sqrt{\pi}\Gamma(n-1)} - 6(n-1)\left(\Psi(n) + \gamma + \frac{8}{3}\right).$$

As $n \to \infty$, we have

$$\frac{Y_n}{n^{3/2}} \xrightarrow{P} \frac{32}{\sqrt{\pi}}.$$

PROOF: We consider a recurrence for Y_n conditional on $\mathbb{F}_{n-1}$ and $\mathbb{I}(n, j)$, mimicking that for Z_n in Eq. (6) as follows:

$$\begin{aligned}Y_n &= Y_{n-1} + (D_{n-1,j} + 1)^3 - D_{n-1,j}^3 + 1 \\ &= Y_{n-1} + 3D_{n-1,j}^2 + 3D_{n-1,j} + 2.\end{aligned}$$

https://doi.org/10.1017/S0269964820000261 Published online by Cambridge University Press

Taking the expectation with respect to $\mathbb{I}(n, j)$, we get

$$\begin{aligned}\mathbb{E}[Y_n \,|\, \mathbb{F}_{n-1}] &= Y_{n-1} + 3\sum_{j=1}^{n-1} D_{n-1,j}^2 \times \frac{D_{n-1,j}}{2(n-2)} + 3\sum_{j=1}^{n-1} D_{n-1,j} \times \frac{D_{n-1,j}}{2(n-2)} + 2 \\ &= Y_{n-1} + \frac{3}{2(n-2)} Y_{n-1} + \frac{3}{2(n-2)} Z_{n-1} + 2.\end{aligned}$$

Take another expectation on both sides and plug in the result of $\mathbb{E}[Z_{n-1}]$ to receive a recurrence on $\mathbb{E}[Y_n]$:

$$\begin{aligned}\mathbb{E}[Y_n] &= \frac{2n-1}{2(n-2)}\mathbb{E}[Y_{n-1}] + 3(\Psi(n-1)+\gamma) + 2 \\ &= \frac{2n-1}{2(n-2)}\left(\frac{2n-3}{2(n-3)}\mathbb{E}[Y_{n-2}] + 3(\Psi(n-2)+\gamma) + 2\right) \\ &\quad + 3(\Psi(n-1)+\gamma) + 2 \\ &= \frac{\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)}{(n-2)(n-3)}\mathbb{E}[Y_{n-2}] + 3\left(\frac{n-\frac{1}{2}}{n-2}\Psi(n-2) + \Psi(n-1)\right) \\ &\quad + \left(\frac{n-\frac{1}{2}}{n-2} + 1\right)(3\gamma+2) \\ &= \frac{\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(n-1)}\left(\sum_{k=2}^{n-1}\frac{\sqrt{\pi}\Gamma(k)(3\Psi(k)+3\gamma+2)}{\Gamma\left(k+\frac{3}{2}\right)} + \frac{8}{3}\right),\end{aligned}$$

after plugging in the initial value $\mathbb{E}[Y_2] = Y_2 = 2$.

We utilize mathematical induction to show

$$\sum_{k=2}^{n-1}\frac{\sqrt{\pi}\Gamma(k)(3\Psi(k)+3\gamma+2)}{\Gamma\left(k+\frac{3}{2}\right)} = \frac{88}{3} - \frac{2\sqrt{\pi}\Gamma(n)(3\Psi(n)+3\gamma+8)}{\Gamma\left(n+\frac{1}{2}\right)}. \tag{7}$$

For better readability, we present the details in Appendix A. The result stated in the lemma follows after simple algebra.

It is obvious that the right-hand side of Eq. (7) converges to $88/3$ as n tends to infinity since the asymptotic of the latter term is 0 according to the Stirling's approximation. To compute the limit of Y_n, we again apply the Stirling's approximation to the gamma functions in the expression of $\mathbb{E}[Y_n]$ to get

$$\mathbb{E}[Y_n] = \frac{32}{\sqrt{\pi}}n^{3/2} + O(n\log n).$$

Thus, we obtain an L_1 convergence for $Y_n/n^{3/2}$ as well as a weak law. ■

Note that in Lemma 1, the expression of $\mathbb{E}[Y_n]$ is well defined for $n \geq 2$. As $n \to 1$, $\Gamma(n-1)$ in the denominator of the first term approaches infinity, and $(n-1)$ in the second term approaches 0, rendering $\mathbb{E}[Y_n] \to 0$. This is consistent with the fact of $\mathbb{E}[Y_1] = Y_1 = 0$, as there is an isolated node in the tree.

We are now ready to calculate the second moment of Z_n as well as the variance of Z_n.

PROPOSITION 8: *The second moment of the Zagreb index of a PORT at time* $n \geq 1$ *is given by*

$$\mathbb{E}[Z_n^2] = 4(n\log n)^2 + 8\gamma(n^2\log n) + \left(16 + 4\gamma^2 - \frac{2\pi^2}{3}\right)n^2 + O(n^{3/2}),$$

and the variance of Z_n *is*

$$\mathbb{V}\mathrm{ar}[Z_n] = \left(16 - \frac{2\pi^2}{3}\right)n^2 + O(n^{3/2}).$$

PROOF: We revisit the almost-sure recurrence for Z_n in Eq. (6) and square both sides to get

$$Z_n^2 = Z_{n-1}^2 + 4D_{n-1,j}^2 + 4 + 4Z_{n-1}D_{n-1,j} + 4Z_{n-1} + 8D_{n-1,j}.$$

Averaging it out with respect to $\mathbb{I}(n,j)$, we have

$$\begin{aligned}\mathbb{E}\left[Z_n^2 \,|\, \mathbb{F}_{n-1}\right] &= Z_{n-1}^2 + 4\sum_{j=1}^{n} D_{n-1,j}^2 \times \frac{D_{n-1,j}}{2(n-2)} + 4\\ &\quad + 4Z_{n-1}\sum_{j=1}^{n} D_{n-1,j} \times \frac{D_{n-1,j}}{2(n-2)} + 4Z_{n-1}\\ &\quad + 8\sum_{j=1}^{n} D_{n-1,j} \times \frac{D_{n-1,j}}{2(n-2)}\\ &= Z_{n-1}^2 + \frac{2}{n-2}Y_{n-1} + 4 + \frac{2}{n-2}Z_{n-1}^2 + 4Z_{n-1} + \frac{4}{n-2}Z_{n-1}\\ &= \frac{n}{n-2}Z_{n-1}^2 + \frac{2}{n-2}Y_{n-1} + \frac{4(n-1)}{n-2}Z_{n-1} + 4.\end{aligned}$$

The recurrence for $\mathbb{E}[Z_n^2]$ is thus obtained by taking the expectation of the formula above on both sides and by plugging in the results of $\mathbb{E}[Y_n]$ and $\mathbb{E}[Z_n]$; that is,

$$\begin{aligned}\mathbb{E}[Z_n^2] &= \frac{n}{n-2}\mathbb{E}[Z_{n-1}^2] + \frac{64\Gamma\left(n-\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(n-1)} + 4(2n-5)(\Psi(n-1)+\gamma) - 28\\ &= \frac{n}{n-2}\mathbb{E}[Z_{n-1}^2] + q_n,\end{aligned}$$

where q_n denotes the latter terms in the above recurrence relation for brevity. This is a simple linear recurrence relation, the solution of which is

$$\begin{aligned}\mathbb{E}[Z_n^2] &= \frac{n}{n-2}\left(\frac{n-1}{n-3}\mathbb{E}[Z_{n-2}^2] + q_{n-1}\right) + q_n\\ &= \frac{n(n-1)}{(n-2)(n-3)}\mathbb{E}[Z_{n-2}^2] + \frac{n}{n-2}q_{n-1} + q_n\\ &= n(n-1)\left(\sum_{k=2}^{n-1}\frac{q_{k+1}}{k(k+1)} + 2\right)\end{aligned}$$

after plugging in the initial value $\mathbb{E}[Z_2^2] = Z_2^2 = 4$. Simplify the summand $q_k/(k(k+1))$ to get

$$\frac{64\Gamma\left(k+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(k+2)} + \frac{4(2k-3)\Psi(k)}{k(k+1)} + \frac{4(2k-3)\gamma - 28}{k(k+1)}. \tag{8}$$

Our next task is to evaluate the sum of the three parts over k from 2 to $(n-1)$ one after another. Details of the computations are given in Appendix B. The stated result presents the exact expressions of a few highest order terms and include the remainder in a big O notation. The orders of these leading terms and their corresponding coefficients are determined by applying the Stirling's approximation to gamma functions and the fact that

$$\Psi(n) = \log n - \frac{1}{2n} + O\left(\frac{1}{n^2}\right),$$
$$\Psi(1, n) = \frac{1}{n} + O\left(\frac{1}{n^2}\right).$$

In what follows, we obtain the variance of Z_n by computing $\mathbb{E}[Z_2^2] - \mathbb{E}^2[Z_n]$. ■

Notice that Z_n^2 converges to $4(n\log n)^2$ in L_1-space as well as in probability, both directly from the continuous mapping theorem. Besides, we would like to point out that we derive the exact solution of $\mathbb{E}[Z_2^2]$ but do not present it in the manuscript for better readability. However, the exact expression of the second moment of Z_n is available upon request.

4.2. Investigation of Asymptotic Behavior

In this section, we exploit a martingale formulation to investigate the asymptotic behavior of Z_n. Based on the recurrence developed in the Proof of Proposition 4, we assert that $\{Z_n\}_n$ is not a martingale array. Consider the following transformation such that the transformed random variables, M_n, form a martingale.

LEMMA 9: *For $n > 1$, the sequence consisting of*

$$M_n = \frac{2}{n-1} Z_n - 4(\Psi(n) + \gamma)$$

is a martingale (with respect to $\mathbb{F}_n$).

PROOF: We consider two constant sequences $\{\alpha_n\}_n$ and $\{\beta_n\}_n$ such that the following martingale property holds for all $n \geq 3$.

$$\begin{aligned}\mathbb{E}[\alpha_n Z_n + \beta_n \,|\, \mathbb{F}_{n-1}] &= \alpha_n \mathbb{E}[Z_n \,|\, \mathbb{F}_{n-1}] + \beta_n \\ &= \frac{\alpha_n(n-1)}{n-2} Z_{n-1} + 2\alpha_n + \beta_n \\ &= \alpha_{n-1} Z_{n-1} + \beta_{n-1}.\end{aligned}$$

This produces two recurrences on α_n and β_n, respectively,

$$\alpha_n = \frac{n-2}{n-1}\alpha_{n-1} \quad \text{and} \quad \beta_n = \beta_{n-1} - 2\alpha_n,$$

with arbitrary choices of initial conditions. We thus obtain the following solutions

$$\alpha_n = \frac{2}{n-1} \quad \text{and} \quad \beta_n = -4(\Psi(n) + \gamma),$$

by choosing initial conditions $\alpha_3 = 1$ for the former and $\beta_1 = 0$ for the latter, respectively. As a matter of fact, the result stated in the lemma is well defined for all $n > 1$. ■

Noting that the martingale M_n is equivalent to

$$M_n = \frac{Z_n - \mathbb{E}[Z_n]}{(n-1)/2},$$

we have

$$\mathbb{E}[M_n] = 0 \quad \text{and} \quad \mathbb{E}[M_n^2] = \frac{\mathbb{V}\text{ar}[Z_n]}{(n-1)^2/4} \sim 64 - \frac{8\pi^2}{3} < +\infty,$$

leading to the fact that $\{M_n\}_n$ is a mean-zero and square-integrable martingale. According to the Doob's martingale convergence theorem, there exists an L_2-measurable random variable, to which M_n converges almost surely, leading to an almost-sure convergence of Z_n after proper scaling.

REMARK 10: *The authors of [13] proved that the limit distribution of the Zagreb index of RRTs is normal with mean* $6n$ *and variance* $8n$*. However, asymptotic normality does not exist for PORTs. It is evident that the limit distribution of the Zagreb index of PORTs is right skewed, as the Pearson's moment coefficient of skewness is always positive [39, Thm. 2].*

5. POISSONIZED PLANE-ORIENTED RECURSIVE TREES

Many real structures do not grow in discrete time, but in continuous time. In this section, we study PORTs embedded into continuous time. The embedding is done by changing the interarrival times between node additions from equispaced discrete units to more general renewal intervals. The author of [2] suggests to use exponential random variables as interarrival time. Under this choice, a count of the arrival points constitutes a Poisson process [34]. Hence, such embedding is commonly called Poissonization [1]. The advantage of Poissonization is that the underlying exponential random variables (interarrival times) share an appealing property—the memoryless property.

We elaborate the growth of a Poissonized PORT by employing an extended graph analogous to that for RRTs [27], that is, extended PORT, as mentioned in Section 3.2. Under the Poissonization framework, each external node is endowed with an independent clock that rings in Exp(1). When the clock of an external node rings, a newcomer joins in the tree, and is connected with the node (in the original tree) that carries that external node by an edge. Then, all the new gaps are filled by new external nodes instantaneously. Upon each renewal, the clocks of existing external nodes are reset owing to the memoryless property, and the new external nodes come endowed with their own independent clocks. We do not consider the time loss of the execution of node additions. Thus, this growth process is Markovian.

To investigate the degree distribution of the node labeled with j, we assume that t_0, the time of its first appearance in the tree, is finite. At this point, there is 1 external node carried by the node labeled with j, and we paint it white; Meanwhile, there is a total of $(2j-3)$ external nodes carried by all the other nodes, and we paint them blue. In the two-color Pólya urn framework, the dynamic of ball addition (at each renewal point) is analogous to that for the discrete-time counterpart, so it also can be represented by Matrix (3). The feature of preferential attachment is reflected in the number of external nodes adjacent to the nodes from the original tree. Let $W(t)$ and $B(t)$ be the numbers of white and blue balls (external nodes) at time $t \geq t_0$, respectively. Noting that $W(t)$ is exactly equal to the degree of the node labeled with j at time t, we thus place our focus on the distribution of $W(t)$.

Recall that Matrix (3) is triangular, so the associated (Poissonized) Pólya urn process is called the triangular Pólya process. This class of urn models was recently investigated by [9]. In this source, the moment generating function of $W(t)$ in a more general framework was developed. Under our specific setting, we present the moment generating function of $W(t)$ in the next proposition.

Proposition 11: *At time $t \geq t_0$, the moment generating function of $W(t)$ is given by*

$$\phi_{W(t)}(u) = \frac{e^{u-(t-t_0)}}{1-(1-e^{-(t-t_0)})e^u}.$$

This result is obtained directly from [9, Lemma 4.3] by plugging in appropriate parameters. Then, the rth moment of $W(t)$ can be derived from $\phi_{W(t)}(u)$ for all $r \geq 1$. The expression of the rth moment of $W(t)$ is available but not in a closed form, rather in a partial sum of an alternating sequence involving Stirling numbers of the second kind and gamma functions. Thus, we do not present all the moments of $W(t)$ in this paper, but only the first two moments (after simplifications) and accordingly the variance in the next corollary.

Corollary 12: *At time t, the first moment, second moment, and variance of $W(t)$, respectively, are*

$$\mathbb{E}[W(t)] = e^{t-t_0},$$
$$\mathbb{E}[W^2(t)] = 2e^{2(t-t_0)} - e^{t-t_0},$$
$$\mathbb{V}\text{ar}[W(t)] = e^{2(t-t_0)} - e^{t-t_0}.$$

Noticing that the probability distribution of a random variable is uniquely determined by its moment generating function provided that it exists, we give the asymptotic distribution of $W(t)$ after proper scaling in the next theorem.

Theorem 13: *As $t \to \infty$, we have*

$$\frac{W(t)}{e^t} \xrightarrow{d} \text{Exp}\left(\frac{1}{e^{t_0}}\right).$$

Proof: Using the moment generating function of $W(t)$, we derive the moment generating function of $\tilde{W}(t) = W(t)/e^{t-t_0}$ as follows:

$$\phi_{\tilde{W}(t)}(u) = \mathbb{E}[e^{(u/e^t)W(t)}] = \frac{e^{u/e^{t-t_0}-(t-t_0)}}{1-(1-e^{-(t-t_0)})e^{u/e^{t-t_0}}},$$

which converges to $1/(1-u)$ as $t \to \infty$. Noticing that $1/(1-u)$ is the moment generating function of Exp(1), we thus, by Billingsley [7, Thm. 30.1, pp. 342], have

$$\frac{W(t)}{e^{t-t_0}} \xrightarrow{d} \text{Exp}(1).$$

The result stated in the theorem follows from the scaling property of exponential random variables. ■

6. CONCLUDING REMARKS

In this section, we add some concluding remarks and propose some future work. In this paper, we investigate three properties of PORTs. First, we determine the degree distribution of a node with a fixed degree by developing its probability mass function. Additionally, we compute the first two moments by exploiting a two-color triangular urn.

Second, we look into the Zagreb index of the class of PORTs. We calculate the exact mean and variance via recurrence methods. We formulate a martingale to characterize the asymptotic behavior of the index. We show that it converges almost surely to a finite random variable after proper scaling, but there does not exist a Gaussian law. We plan, in the future, to investigate many other topological indices, such as the Gini index and the Randić index, for PORTs.

Last, we study the degree profile of PORTs embedded into continuous time, so-called Poissonized PORTs. We interpret the growth of Poissonized PORTs by introducing extended trees. The exact moment generating function of the degree variable is determined. We show that the asymptotic distribution of the degree variable scaled by e^t is exponential.

Acknowledgments

The author thanks two anonymous referees for their insightful comments that help improve the quality of the paper. Besides, the author is especially grateful to Professor Hosam Mahmoud from the George Washington University for his encouragement and advice throughout this research.

References

1. Aldous, D. (1989). *Probability approximations via the Poisson clumping heuristic*. Applied Mathematical Sciences No. 77. New York, NY: Springer-Verlag, xvi+269 pp.
2. Athreya, K., & Karlin, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *The Annals of Mathematical Statistics* 39: 1801–1817.
3. Avrachenkov, K., & Lebedev, D. (2006). PageRank of scale-free growing networks. *Internet Mathematics* 3: 207–231.
4. Balaji, H., & Mahmoud, H. (2017). The Gini index of random trees with an application to caterpillars. *Journal of Applied Probability* 54: 701–709.
5. Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science* 286: 509–512.
6. Berge, C. (1973). *Graphs and hypergraphs*. Translated from the French by Edward Minieka. *North-Holland Mathematical Library*, vol. 6. Amsterdam-London/New York: North-Holland Publishing Co./American Elsevier Publishing Co., Inc., xiv+528 pp.
7. Billingsley, P. (1995). *Probability and measure*, 3rd ed. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York, NY: John Wiley & Sons, Inc., xiv+593 pp.
8. Bollobás, B., Riordan, O., Spencer, J., & Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms* 18: 279–290.
9. Chen, C., & Mahmoud, H. (2018). The continuous-time triangular Pólya process. *Annals of the Institute of Statistical Mathematics* 70: 303–321.
10. Chen, W.-C., & Ni, W.-C. (1994). Internal path length of the binary representation of heap-ordered trees. *Information Processing Letters* 51: 129–132.
11. Drmota, M., Gittenberger, B., & Panholzer, A. (2008). *The degree distribution of thickened trees* (English summary). In *Fifth Colloquium on Mathematics and Computer Science, Discrete Math. Theor. Comput. Sci. Proc.*, vol. AI. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, pp. 149–161.
12. Feller, W. (1971). *An introduction to probability theory and its applications*, 2nd ed., vol. II, New York-London-Sydney: John Wiley & Sons, Inc., xxiv+669 pp.
13. Feng, Q., & Hu, Z. (2011). On the Zagreb index of random recursive trees. *Journal of Applied Probability* 48: 1189–1196.
14. Feng, Q., & Hu, Z. (2013). Phase changes in the topological indices of scale-free trees. *Journal of Applied Probability* 50: 516–532.

15. Feng, Q., & Hu, Z. (2015). Asymptotic normality of the Zagreb index of random b-ary recursive trees. *Dal'nevostochnyi Matematicheskii Zhurnal* 15: 91–101.
16. Feng, Q., Mahmoud, H., & Panholzer, A. (2010). Limit laws for the Randić index of random binary trees. *Annals of the Institute of Statistical Mathematics* 60: 319–343.
17. Fuchs, M., & Lee, C.-K. (2015). The Wiener index of random digital trees. *SIAM Journal on Discrete Mathematics* 29: 586–614.
18. Gastwirth, J. (1977). A probability model of a pyramid scheme. *The American Statistician* 31: 79–82.
19. Gutman, I., & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total ϕ-electron energy of alternant hydrocarbons. *Chemical Physics Letters* 17: 535–538.
20. Holmgren, C., & Janson, S. (2017). Fringe trees, Crump-Mode-Jagers branching processes and m-ary search trees. *Probability Surveys* 14: 53–154.
21. Hwang, H.-K. (2007). Profiles of random trees: plane-oriented recursive trees. *Random Structures & Algorithms* 30: 380–413.
22. Janson, S. (2005). Asymptotic degree distribution in random recursive trees. *Random Structures & Algorithms* 25: 69–83.
23. Janson, S. (2019). Random recursive trees and preferential attachment trees are random split trees. *Combinatorics, Probability and Computing* 28: 81–99.
24. Lu, J., & Feng, Q. (1998). Strong consistency of the number of vertices of given degrees in nonuniform random recursive trees. *Yokohama Mathematical Journal* 45: 61–69.
25. Mahmoud, H. (1992). Distances in random plane-oriented recursive trees. *Journal of Computational and Applied Mathematics* 41: 237–245.
26. Mahmoud, H. (2009). *Pólya urn models.* Texts in Statistical Science Series. Boca Raton, FL: CRC Press, xii+290 pp.
27. Mahmoud, H., & Smythe, R. (1992). Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures & Algorithms* 3: 255–266.
28. Mahmoud, H., Smythe, R., & Szymański, J. (1993). On the structure of random plane-oriented recursive trees and their branches. *Random Structures & Algorithms* 4: 151–176.
29. Moon, J. (1974). The distance between nodes in recursive trees. In *Combinatorics (Proc. British Combinatorial Conf., Univ. Coll. Wales, Aberystwyth, 1973).* London Math. Soc. Lecture Note Ser., vol. 13. London: Cambridge Univ. Press, pp. 125–132.
30. Najock, D., & Heyde, C.C. (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *Journal of Applied Probability* 19: 675–680.
31. Neininger, R. (2002). The Wiener index of random trees. *Combinatorics, Probability and Computing* 11: 587–597.
32. Nikolić, S., Tolić, I., Trinajstić, N., & Baučić, I. (2000). On the Zagreb indices as complexity indices. *Croatica Chemica Acta* 73: 909–921.
33. Nikolić, S., Kovačević, G., Miličević, A., & Trinajstić, N. (2003). The Zagreb indices 30 years after. *Croatica Chemica Acta* 76: 113–124.
34. Ross, S. (1996). *Stochastic processes*, 2nd ed. Wiley Series in Probability and Statistics: Probability and Statistics. New York, NY: John Wiley & Sons, Inc., xvi+510 pp.
35. Szymański, J. (1987). *On a nonuniform random recursive tree.* In *Random graphs '85 (Poznań, 1985). North-Holland Math. Stud.*, vol. 144. Amsterdam: North-Holland, pp. 297–306.
36. Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics.* Hoboken, NJ: Wiley, 1257 pp.
37. Zhang, P. (2016). On properties of several random networks. (Ph.D. Thesis), The George Washington Univeristy, 131 pp.
38. Zhang, P. (2016). On terminal nodes and the degree profile of preferential dynamic attachment circuits. In *Thirteenth workshop on analytic algorithmics and combinatorics.* Arlington: SIAM, pp. 80–92.
39. Zhang, P. (2019). The Zagreb index of several random models. ArXiv:1901.04657.
40. Zhang, P., & Dey, D. (2019). The degree profile and Gini index of random caterpillar trees. *Probability in the Engineering and Informational Sciences* 33: 511–527.
41. Zhang, P., & Mahmoud, H. (2016). The degree profile and weight in Apollonian networks and k-trees. *Advances in Applied Probability* 48: 163–175.
42. Zhang, P., Chen, C., & Mahmoud, H. (2015). Explicit characterization of moments of balanced triangular Pólya urns by an elementary approach. *Statistics & Probability Letters* 96: 149–154.

APPENDIX A

LEMMA 14: *For $n \geq 3$, we have*

$$\sum_{k=2}^{n-1} \frac{\sqrt{\pi}\Gamma(k)(3\Psi(k)+3\gamma+2)}{\Gamma\left(k+\frac{3}{2}\right)} = \frac{88}{3} - \frac{2\sqrt{\pi}\Gamma(n)(3\Psi(n)+3\gamma+8)}{\Gamma\left(n+\frac{1}{2}\right)}. \tag{A.1}$$

PROOF: We show the stated result by induction. For the base, that is, $n = 3$, the left-hand side of and the right-hand side of Eq. (A.1) are equal; that is,

$$\frac{\sqrt{\pi}\Gamma(2)(3\Psi(2)+3\gamma+2)}{\Gamma\left(2+\frac{3}{2}\right)} = \frac{8}{3} = \frac{88}{3} - \frac{2\sqrt{\pi}\Gamma(3)(3\Psi(3)+3\gamma+8)}{\Gamma\left(3+\frac{1}{2}\right)}.$$

We assume that Eq. (A.1) holds for n. In the inductive step, we get

$$\begin{aligned}\sum_{k=2}^{n} \frac{\sqrt{\pi}\Gamma(k)(3\Psi(k)+3\gamma+2)}{\Gamma\left(k+\frac{3}{2}\right)} &= \frac{88}{3} - \frac{2\sqrt{\pi}\Gamma(n)(3\Psi(n)+3\gamma+8)}{\Gamma\left(n+\frac{1}{2}\right)} \\ &\quad + \frac{\sqrt{\pi}\Gamma(n)(3\Psi(n)+3\gamma+2)}{\Gamma\left(n+\frac{3}{2}\right)} \\ &= \frac{88}{3} + \frac{\sqrt{\pi}\Gamma(n)(-6n\Psi(n)-6-6\gamma n-16n)}{\Gamma\left(n+\frac{3}{2}\right)}\end{aligned}$$

The proof is completed by applying its well-known recurrence relation for the digamma function: $\Psi(n+1) = \Psi(n) + 1/n$. ■

APPENDIX B

LEMMA 15: *For the three parts in Eq. (8), we take the sum over k from 2 to $(n-1)$, and respectively get*

$$\sum_{k=2}^{n-1} \frac{64\,\Gamma\left(k+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(k+2)} = 48 - \frac{128\,\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(n+1)}$$

$$\begin{aligned}\sum_{k=2}^{n-1} \frac{4(2k-3)\Psi(k)}{k(k+1)} &= 4(\Psi^2(n+1)+3\Psi(n+1)+\Psi(1,n+1)+2) \\ &\quad - \frac{12(n-1)\Psi(n)}{n} - \frac{2\pi^2+12\gamma^2-54\gamma+60}{3},\end{aligned}$$

$$\sum_{k=2}^{n-1} \frac{4\,(2k-3)\,\gamma-28}{k(k+1)} = 8\gamma\Psi(n) + 8\gamma^2 - 18\gamma - 14 + \frac{4(5\gamma+7)}{n},$$

where $\Psi(1,x) := (d/dx)\Psi(x)$ is the first derivative of digamma function $\Psi(x)$, known as the first-order polygamma function.

PROOF: To begin with, we introduce two notations. Let $H_k := \sum_{j=1}^{k}(1/j)$ be the kth order harmonic number and $H_{2,k} := \sum_{j=1}^{k}(1/j^2)$be the generalized kth order harmonic number, respectively.

https://doi.org/10.1017/S0269964820000261 Published online by Cambridge University Press

The computation of the first sum is based on the following well-known identity for gamma functions:

$$\sum_{k=1}^{n}\frac{\Gamma(k+a)}{\Gamma(k+b)}=\frac{(n+b)\Gamma(n+a+1)}{(a-b+1)\Gamma(n+b+1)}-\frac{b\Gamma(a+1)}{(a-b+1)\Gamma(b+1)},$$

for $a, b \in \mathbb{R}, a \neq b-1$. Set $a = 1/2$ and $b = 2$ to get

$$\begin{aligned}\sum_{k=2}^{n-1}\frac{64\Gamma\left(k+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(k+2)}&=\frac{64}{\sqrt{\pi}}\left(\sum_{k=1}^{\infty}\frac{\Gamma\left(k+\frac{1}{2}\right)}{\Gamma(k+2)}-\frac{\sqrt{\pi}}{4}-\frac{\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+2)}\right)\\&=\frac{64}{\sqrt{\pi}}\left(-\frac{2(n+2)\Gamma\left(n+\frac{3}{2}\right)}{\Gamma(n+3)}+\frac{3\sqrt{\pi}}{4}-\frac{\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+2)}\right)\\&=\frac{64}{\sqrt{\pi}}\left(-\frac{2\,\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+1)}+\frac{3\sqrt{\pi}}{4}\right).\end{aligned}$$

We then look at the third sum and then the second sum. We can divide the third sum into two parts:

$$\sum_{k=2}^{n-1}\frac{4(2k-3)\gamma-28}{k(k+1)}=\sum_{k=2}^{n-1}\frac{8\gamma}{k+1}-\sum_{k=2}^{n-1}\frac{12\gamma+28}{k(k+1)}. \tag{B.1}$$

It is obvious that

$$\sum_{k=2}^{n-1}\frac{1}{k+1}=\sum_{k=1}^{n}\frac{1}{k}-1-\frac{1}{2}=H_n-\frac{3}{2}$$

and

$$\sum_{k=2}^{n-1}\frac{1}{k(k+1)}=\sum_{k=2}^{n-1}\left(\frac{1}{k}-\frac{1}{k+1}\right)=\frac{1}{2}-\frac{1}{n}.$$

Eq. (B.1) can be simplified to

$$\sum_{k=2}^{n-1}\frac{4\,(2k-3)\,\gamma-28}{k(k+1)}=8\gamma\left(H_n-\frac{3}{2}\right)-4(3\gamma+7)\left(\frac{1}{2}-\frac{1}{n}\right),$$

which is equivalent to the stated results by considering the well-known relation between digamma functions and harmonic numbers: $H_{n-1}=\Psi(n)+\gamma$.

Lastly, we compute the sum of the second part. At first, we present two useful identities for harmonic numbers in the next lemma.

Lemma 16: *Let $H_k := \sum_{j=1}^{k}(1/j)$ be the kth order harmonic number and $H_{2,k} := \sum_{j=1}^{k}(1/j^2)$ be the generalized kth order harmonic number, respectively. We have*

$$\sum_{k=1}^{n}\frac{H_{k-1}}{k+1}=\frac{H_{n+1}^2-H_{2,n+1}}{2}+\frac{1}{n+1}-1,$$

$$\sum_{k=1}^{n}\frac{H_{k-1}}{k(k+1)}=-\frac{nH_{n-1}-n^2+1}{n(n+1)}.$$

The proofs for both are straightforward by via mathematical induction and the definition of harmonic numbers. Hence, we omit the tedious algebra.

https://doi.org/10.1017/S0269964820000261 Published online by Cambridge University Press

Again, we consider the relation between digamma functions and harmonic numbers, that is, $H_{k-1} = \Psi(k) + \gamma$, and then further divide the second sum into four parts to get

$$\sum_{k=2}^{n-1} \frac{4(2k-3)\Psi(k)}{k(k+1)} = \sum_{k=2}^{n-1} \frac{4(2k-3)(H_{k-1}-\gamma)}{k(k+1)}$$
$$= \sum_{k=2}^{n-1} \frac{8H_{k-1}}{k+1} - \sum_{k=2}^{n-1} \frac{12H_{k-1}}{k(k+1)} - \sum_{k=2}^{n-1} \frac{8\gamma}{k+1} + \sum_{k=2}^{n-1} \frac{12\gamma}{k(k+1)}.$$

The computations of the former two can be done by applying the identities in Lemma B.2, whereas the computations of the latter two can be handled in a similar manner as the third sum. Putting everything together, we obtain the stated result by converting harmonic numbers to digamma and polygamma functions according to the following relation: $\Psi(n) = H_{n-1} - \gamma$, $\Psi(1, n+1) = \pi^2/6 - H_{2,n}$. ■

https://doi.org/10.1017/S0269964820000261 Published online by Cambridge University Press