On Terminal Nodes and the Degree Profile of Preferential Dynamic Attachment Circuits

Panpan Zhang*

Abstract

We investigate terminal nodes and the degree profile in preferential dynamic attachment circuits. We study the distribution of the number of terminal nodes, which are the nodes that have not recruited other nodes, as the circuit ages. The expectation and variance of the number of terminal nodes are both linear with respect to the age of the circuit. We show via martingale that the number of terminal nodes asymptotically follows a Gaussian law. We also study the exact distribution of the degree of a specific node as the circuit grows. The exact expectation and variance of the degree of a node are determined via a series of Pólya–Eggenberger urn models with "hiccups" in between and recurrence methods. Phase transitions of these degrees are discussed briefly.

1 Introduction

Networks are proliferating all around us. They appear in many forms, such as hardwired, amorphous cyber and virtual constructs, routes on navigation and trading maps, etc. There is need for models and analysis of networks. In this manuscript, we take up a type of network that has recently received attention, the *preferential attachment circuit*. These are circuits (networks) that grow with new comers favoring to attach themselves to nodes of higher degrees in the network, a manifestation of the economic principles "the rich get richer" and "success breeds success."

In this research, we discuss two properties in preferential attachment circuits: terminal nodes and the degree profile of a node as the circuit ages. Terminal nodes are the nodes that have not been chosen as parents for the newly inserted nodes in the circuit (e.g., leaves of a random recursive tree). Terminal nodes are the outputs of random circuits and they have minimal outdegree (i.e., 0). The study of the distribution of degrees (degree profile) in random networks has been a popular topic. Knowing the degree of a node can tell, for instance, how popular the node is in a social network, or how much demand there is on it in a routing network, which can help allocate the appropriate resources. In the context of circuits, the outdegree of a node determines the amount of electric current that flows through the node.

Related circuit models are in [14] (uniform choice of parents), in [7] (a uniform positional model), and in [11] (a model without the consideration of the dynamical aspect in the present paper). We got inspiration to study a dynamic model after reading [12], and [13]. The source [9] discusses several variants of these circuits. The plane-oriented recursive tree (PORT), introduced in [6], is a special case of our circuit model. The dynamic flavor in our model makes it more adaptable to the real world, and this flavor will be introduced in the next section.

2 Preferential Attachment Circuits

The circuit (network) model that we discuss in this paper refers to a structure (graph) connecting discrete objects (nodes). We consider a circuit model that grows in the following way. At the beginning of discrete time (time 0), there is an originator, which is a single isolated node labeled with 0. We shall refer to the nodes by their labels. So, the originator is node 0. At time $n \ge 1$, a new node n appears and attaches itself to $C_{n-1}^{(m)}$, the circuit of a fixed index m ($m \ge 1$) existing at time $n-1$. The network $C_n^{(m)}$ grows from its shape $C_{n-1}^{(m)}$ in the previous time step $n-1$ by choosing $m \ge 1$ nodes from $C_{n-1}^{(m)}$ as *parents*. We

*Department of Statistics, The George Washington University, Washington, D.C. 20052, USA

call the sequence of nodes chosen as parents of node n, the nth *sample*. The sample is taken with replacement. Each of the m parents in the nth sample is joined by an edge to node n.

As the sampling is with replacement, a node j can be chosen multiple times in the sample at step n. For example, m can be equal to five, and j may appear three times in the sample of parents for node n. Hence, three new edges will appear joining the nodes j and n. The dynamic probabilities are quite sensitive to the order (sequence) of nodes in the sample. For example, the probability of the event that the three choices of node j in the nth sample appear as first, second and third may not be the same as its being first, second and fifth. This is a critical element in our probability calculation, and the notion will be made precise later, as we shall derive these probabilities for all possible sequences of parent choices. As the model allocates higher probability to nodes of higher degrees, the nodes that recruited children are more likely to attract more children, and the model rightly deserves to be described with attributes like "the rich get richer" and "success breeds success."

The choice of the parents is not uniform and the probabilities are dynamic during the construction. The choice rather depends on the degrees of the nodes present in $C_{n-1}^{(m)}$. A parent node is chosen with probability associated with its degree in $C_{n-1}^{(m)}$ (more precisely, proportional to its outdegree plus one). For example, once the first of the m new edges adjoins an existing node j of outdegree d, to node n, it changes the outdegree of node j to $d+1$, so node j now has outdegree $d+1$, and its chance to be chosen again to construct the second edge as parent of n is now proportional to $d+2$. The second parent, be it j or some other node, is joined to n. The process continues in this fashion, adding new edges and accounting for the dynamic change in their degrees, which is reflected in the probabilities of choosing nodes as parents, till the mth member of the sample has been collected and all the members in the set of parent nodes are joined to n. At this point we consider that node n has completed its parent selection. Once settled, n will not change its parents.

Figure 1 illustrates the growth of a preferential attachment circuit with index $m = 3$ in the first two insertions of node 1 and 2, showing the dynamical changes within each sampling step (edge addition). From left to right, the probabilities of this particular development (conditioning on the previous evolution) are $1, 1, 1, \frac{1}{5}, \frac{4}{6}, \frac{2}{7}$.

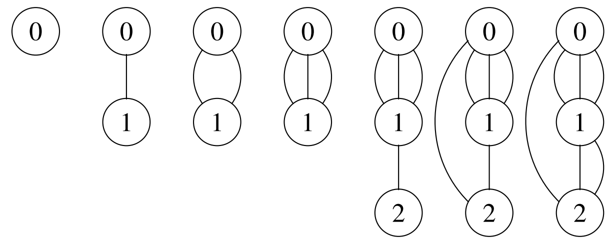

Figure 1: The evolution of a preferential attachment circuit of index 3 in two steps.

When $m = 1$, the circuit is a tree known in the literature as the plane-oriented recursive tree (PORT), introduced in [6], and its derivatives have received quite a bit of attention in recent years [1, 2, 5, 10].

It aids the analysis to consider extended circuits, where each node is supplemented with external nodes that appear as children in the "gaps" between the edges emanating out of a node and connect them to the node. We think of the insertion position to the left (right) of all the edges out of a node as a virtual gap, too. Figure 2 shows the circuit of Figure 1 after it has been extended. In Figure 2 the external nodes are shown as squares, some empty and some blackened. That is a color code that will be explained in the sequel.

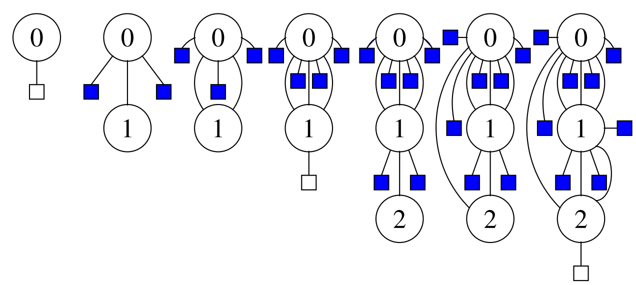

Figure 2: The evolution of an extended preferential attachment circuit of index 3 in two steps.

After the completion of each node insertion, a total of $(m+1)$ external nodes added to the extended circuit, of which m are added to the existing nodes, and one is added to the newly inserted node (see the 4th, and 7th graph in Figure 2). Therefore, after n node insertions, we have a total of

$$\tau_n = (m+1)n + 1 \tag{1}$$

external nodes. This notation will be used throughout the paper.

3 Organization

The rest of this manuscript is organized as follows. In Section 4, we discuss the number of terminal nodes. The first paragraph of the section offers a color code of external nodes into white and blue. The section is divided into four subsections. Subsection 4.1 introduces the distribution of blue (white) external nodes in each sample, proved by double mathematical induction; Subsection 4.2 establishes a stochastic recurrence (pointwise in the sample space) for terminal nodes. The recurrence is utilized to find the exact expectation and variance of the number of white external nodes. In Subsection 4.3, a martingale underlying terminal nodes is uncovered and used to establish a central limit theorem for terminal nodes. In the last section, Section 5, we look at the degree of a specified node as the circuit ages. We find the exact distribution of that degree via a series of Pólya–Eggenberger urns with "hiccups" in between. That distribution is unwieldy for moment calculations, so we resort to the stochastic recurrence again to find the exact mean and variance of the degree. Upon inspecting the average degree of a specified node, we realize that there arise phases as the circuit ages. These phases are discussed at the end of the manuscript.

4 Terminal nodes

Let us look at *terminal* nodes, which are the nodes that joined the circuit, but were not chosen as parents of any other nodes. A terminal node has outdegree 0, and thus has only one external node attached to it. To study terminal nodes, we distinguish the external nodes by colors: Those that are attached to terminal nodes have one color (say white (W)), and those coming out of other nonterminal nodes are colored with a different color (say blue (B)); see Figure 2 for an illustration. The number of white external nodes is the number of terminal nodes in the circuit.

4.1 Distribution of the number of white (blue) external nodes in the sample

Let $W_n(B_n)$ be the number of white (blue) external nodes, after n nodes have been added to the circuit. Note that W_n is also the number of terminal nodes. We have a stochastic recurrence governing the evolution: The number of terminal nodes after n node insertions is what it was after $n-1$ node insertions, plus an adjustment occurring during the selection of m parents for the nth node. The adjustment can be an increase by 1, no change, or a decrease by $1, 2, \dots, m-1$, depending on the number of blue external nodes selected in the nth sample drawing—if k blue external nodes are selected in the sample (of size m), the adjustment will be a decrease of $m-k-1$.

Let $\mathbb{F}_n$ be the σ-field generated by the first n node insertions. Denote by K_n the number of blue external balls in the nth sample. When $K_n = k$ for $0 \le k \le m$, there are $\binom{m}{k}$ possibilities (associated with different positions of the blue external nodes in the sample), and each has a possibly different probability (conditioning on $\mathbb{F}_{n-1}$). We shall cast the results in terms of Pochhammer's symbols for the rising and falling factorials. Pochhammer's symbol for the rising factorial is

$$\langle x \rangle_s = x(x+1)(x+2)\cdots(x+s-1)$$

for any $x \in \mathbb{R}$, and any integer $s \ge 0$, with the interpretation that $\langle x \rangle_0 = 1$. Pochhammer's symbol for the falling factorial is

$$(x)_s = x(x-1)(x-2)\cdots(x-s+1)$$

for any $x \in \mathbb{R}$, and any integer $s \ge 0$, with the interpretation that $(x)_0 = 1$.

Proposition 1. *Let K_n be the number of blue external nodes drawn in the sample in a preferential attachment circuit with index m at time n. For any integer $0 \le k \le m$, we have*

$$\mathbb{P}(K_n = k \,|\, \mathbb{F}_{n-1}) = \binom{m}{k} \frac{(W_{n-1})_{m-k}(B_{n-1}+m-1)_k}{\langle \tau_{n-1} \rangle_m},$$

Proof. We prove this proposition by a double induction on $0 \leq k \leq m$ and $m \geq 1$. The proof progresses in the style of dynamic programming to fill an infinite lower triangle table. We index the rows by m and the columns by k. We first initialize the first column (i.e., $k = 0$ for all m) and the diagonal (i.e., $k = m$ for all m) as two bases of the double induction, then proceed to fill the other entries by induction.

For $m \geq 1$ and $k = 0$, we have

$$\mathbb{P}(K_n = 0 \,|\, \mathbb{F}_{n-1}) = \frac{(W_{n-1})_m}{\langle \tau_{n-1} \rangle_m};$$

and for $k = m$, we have

$$\mathbb{P}(K_n = m \,|\, \mathbb{F}_{n-1}) = \frac{\langle B_{n-1} \rangle_m}{\langle \tau_{n-1} \rangle_m} = \frac{(B_{n-1} + m - 1)_m}{\langle \tau_{n-1} \rangle_m}.$$

Then, we assume as an induction hypothesis that the formula holds for all rows (in the lower triangle table) up to $m - 1$. Given $\mathbb{F}_{n-1}$, the probability of $K_n = k$, with the k choices of a blue ball occurring at positions $1 \leq i_1 < i_2 < \cdots < i_k \leq m$ is

$$\frac{(W_{n-1})_{m-k}}{\langle \tau_{n-1} \rangle_m} (B_{n-1} + 2i_1 - 2)(B_{n-1} + 2i_2 - 3) \times \cdots \times \big(B_{n-1} + 2i_k - (k+1)\big).$$

We next get the conditional probability, regardless of the position of the k blue nodes in the sample:

$$\begin{aligned}
&\mathbb{P}(K_n^{(m)} = k \,|\, \mathbb{F}_{n-1}) \\
&= \sum_{1 \leq i_1 < \cdots < i_k \leq m} \frac{(W_{n-1})_{m-k}}{\langle \tau_{n-1} \rangle_m} (B_{n-1} + 2i_1 - 2) \\
&\qquad \times (B_{n-1} + 2i_2 - 3) \cdots \big(B_{n-1} + 2i_k - (k+1)\big) \\
&= \sum_{1 \leq i_1 < \cdots < i_k \leq m-1} \frac{(W_{n-1})_{m-k}}{\langle \tau_{n-1} \rangle_m} (B_{n-1} + 2i_1 - 2) \\
&\qquad \times (B_{n-1} + 2i_2 - 3) \cdots \big(B_{n-1} + 2i_k - (k+1)\big) \\
&\quad + \sum_{1 \leq i_1 < \cdots < i_k = m} \frac{(W_{n-1})_{m-k}}{\langle \tau_{n-1} \rangle_m} (B_{n-1} + 2i_1 - 2) \\
&\qquad \times (B_{n-1} + 2i_2 - 3) \cdots \big(B_{n-1} + 2i_k - (k+1)\big)
\end{aligned}$$

Note that the right-hand side of the equation has the interpretation

$$\begin{aligned}
\mathbb{P}(K_n^{(m)} = k \,|\, \mathbb{F}_{n-1}) &= \mathbb{P}(K_n^{(m-1)} = k \,|\, \mathbb{F}_{n-1}) \\
&\quad + \mathbb{P}(K_n^{(m-1)} = k - 1 \,|\, \mathbb{F}_{n-1}).
\end{aligned}$$

We take into account the fact that numbers like W_{n-1} on the right-hand side depend on m. We only need their values in the Pochammar's symbols, which can be mimicked by developments of their counterparts in a circuit of index $m-1$. The two conditional probabilities on the right-hand side are in our induction hypotheses, leading to

$$\begin{aligned}
&\mathbb{P}(K_n^{(m)} = k \,|\, \mathbb{F}_{n-1}) \\
&= \frac{(W_{n-1})_{m-k}}{\langle \tau_{n-1} \rangle_m} \Big[\binom{m-1}{k} (B_{n-1} + m - 2)_k \\
&\qquad + \binom{m-1}{k-1} (B_{n-1} + m - 2)_{k-1} \\
&\qquad \times \big(B_{n-1} + 2m - (k+1)\big) \Big] \\
&= \binom{m}{k} \frac{(W_{n-1})_{m-k} (B_{n-1} + m - 1)_k}{\langle \tau_{n-1} \rangle_m}.
\end{aligned}$$

□

4.2 A stochastic recurrence for the terminal nodes

Recall the relation between blue external nodes in the sample and adjustment of white external nodes in the circuit indicated in the first paragraph of Section 4, we are able to construct recurrences for the first and second moments of W_n, using the conditional distribution function of K_n given $\mathbb{F}_{n-1}$. We need the following three combinatorial identities in the proofs of the forthcoming proposition and theorem. We refer the reader to [3] for more details of the three identities.

$$\sum_{k=0}^{n} \binom{n}{k} (x)_k (y)_{n-k} = (x + y)_n. \tag{2}$$

$$\sum_{k=0}^{n} k \binom{n}{k} (x)_k (y)_{n-k} = nx(x + y - 1)_{n-1}. \tag{3}$$

$$\sum_{k=0}^{n} k^2 \binom{n}{k} (x)_k (y)_{n-k} = nx(nx + y - n) \times (x + y - 2)_{n-2}. \tag{4}$$

Proposition 2. *Let W_n be the number of white external nodes in a preferential attachment circuit with index m (i.e., the number of terminal nodes) after n insertions. We* *have the conditional expectation*

$$\mathbb{E}[W_n \,|\, \mathbb{F}_{n-1}] = \frac{(m+1)(n-1)}{(m+1)(n-1)+m} W_{n-1} + 1. \tag{5}$$

Proof. We have a pointwise relation for the history of W_n. Namely, it is

$$W_n = W_{n-1} - (m - K_n - 1), \tag{6}$$

where K_n is the number of blue external nodes in the sample at time n as defined in the previous subsection. Accordingly, we have

$$\begin{aligned}
&\mathbb{E}[W_n \,|\, \mathbb{F}_{n-1}]\\
&= W_{n-1} - \sum_{k=0}^{m} (m-k-1) \mathbb{P}(K_n = k \,|\, \mathbb{F}_{n-1})\\
&= W_{n-1} + 1\\
&\qquad - m \sum_{k=0}^{m} \frac{\binom{m}{k} (W_{n-1})_{m-k} (B_{n-1}+m-1)_k}{\langle \tau_{n-1} \rangle_m}\\
&\qquad + \sum_{k=0}^{m} k \frac{\binom{m}{k} (W_{n-1})_{m-k} (B_{n-1}+m-1)_k}{\langle \tau_{n-1} \rangle_m}.
\end{aligned}$$

We obtain the conditional probability of W_n by applying (2) and (3) to the two summations of the equation in the last two lines. □

For $n = 0$, we only have one node in the circuit, and it is the terminal node carrying one white external position.

Theorem 1. *Let W_n be the number of white external nodes in a preferential attachment circuit with index m (i.e., the number of terminal nodes) after n insertions. For $n \geq 1$, we have*

$$\begin{aligned}
\mathbb{E}[W_n] &= \frac{m+1}{2m+1} n + \frac{m}{2m+1},\\
\mathbb{E}[W_n^2] &= \frac{1}{(2m+1)^2 (3m+1)(mn+n-1)}\\
&\qquad \times \big((3m+1)(m+1)^3 n^3\\
&\qquad\qquad + (m+1)^2 (8m^2 - m - 1) n^2\\
&\qquad\qquad + m(m+1)(3m^2 - 6m - 1) n\\
&\qquad\qquad - m(2m^3 + 6m^2 + 3m + 1) \big).
\end{aligned}$$

Proof. Let us take the expectation of both sides of (5), obtaining a recurrence for $\mathbb{E}[W_n]$, namely

$$\mathbb{E}[W_n] = \frac{(m+1)(n-1)}{(m+1)(n-1)+m} \mathbb{E}[W_{n-1}] + 1.$$

We complete the proof of the first moment of W_n by solving the recurrence under the initial condition of $\mathbb{E}[W_0] = 1$.

Next, we formulate a recurrence for the second moment of W_n (i.e., $\mathbb{E}[W_n^2 \,|\, \mathbb{F}_{n-1}]$) by squaring the relation (6). Consequently, we have

$$\begin{aligned}
\mathbb{E}[W_n^2 \,|\, \mathbb{F}_{n-1}] &= W_{n-1}^2 - 2(m-1) W_{n-1} + (m-1)^2\\
&\qquad + \sum_{k=0}^{m} k^2 \, \mathbb{P}(K_n = k \,|\, \mathbb{F}_{n-1})\\
&\qquad + 2(W_{n-1} - m + 1)\\
&\qquad\qquad \times \sum_{k=0}^{m} k \, \mathbb{P}(K_n = k \,|\, \mathbb{F}_{n-1}).
\end{aligned}$$

We apply (4) and (3) to the summations in the equation and have

$$\begin{aligned}
&\mathbb{E}[W_n^2 \,|\, \mathbb{F}_{n-1}]\\
&= W_{n-1}^2 - 2(m-1) W_{n-1} + (m-1)^2\\
&\qquad - \frac{2m(W_{n-1} - m + 1)(W_{n-1} - mn - n + 1)}{mn + n - 1}\\
&\qquad + m(W_{n-1} - mn - n + 1)\\
&\qquad\quad \times \frac{\big((m-1) W_{n-1} - m^2 n - mn + 2m\big)}{(mn+n-1)(mn+n-2)}.
\end{aligned}$$

Take another expectation both sides, and plug in the result of $E[W_{n-1}]$, and we obtain a recurrence for $\mathbb{E}[W_n^2]$.

$$\begin{aligned}
\mathbb{E}[W_n^2] &= \frac{(m+1)(n-1)(mn+n-m-2)}{(mn+n-1)(mn+n-2)} \mathbb{E}[W_{n-1}^2]\\
&\qquad + \frac{(m+1)(n-1)(2mn+m+2n-4)}{(mn+n-1)(mn+n-2)}\\
&\qquad\quad \times \frac{(m+1)(n-1)+m}{2m+1} + 1.
\end{aligned}$$

In addition to the initial condition $\mathbb{E}[W_0^2] = 1$, we solve the recurrence and get the solution as stated. □

We want to point out that when $n = 1$, the recurrence for $\mathbb{E}[W_1]$ is reduced to $\mathbb{E}[W_1] = 1$, which also can be used as the initial condition in our recurrence calculation.

Corollary 1. *For $n \geq 1$, the variance of random variable W_n is given by*

$$\mathbb{V}\mathrm{ar}[W_n] = \frac{(m+1)m}{\big((m+1)n-1\big)(2m+1)^2(3m+1)} \times \big(2m(m+1)n^2 - (m-1)n - (2m^2+m+1)\big).$$

When $n = 1$, the variance of W_n is equal to 0 for any m, since at time 1 there is one white external node added to node 1, and this is deterministic. We observe that the variance of W_n has the highest order n. The relatively small variance is helpful to derive the limit distribution of W_n.

Theorem 2. *Let W_n be the number of external white nodes in a preferential attachment circuit with index m after n insertions*[1]*. Then, we have*

$$W_n = \frac{m+1}{2m+1}n + O_{L_1}\big(n^{\frac{1}{2}}\big)$$

Proof. By Theorem 1 and Corollary 1, we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big(W_n - \frac{m+1}{2m+1}n\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\big(W_n - \mathbb{E}[W_n]\big) + \Big(\mathbb{E}[W_n] - \frac{m+1}{2m+1}n\Big)\Big)^2\Big] \\
&= \mathbb{V}\mathrm{ar}[W_n] + \Big(\mathbb{E}[W_n] - \frac{m+1}{2m+1}n\Big)^2 \\
&= O(n).
\end{aligned}$$

So, by the Cauchy–Schwarz inequality, we have

$$\mathbb{E}\Big[\Big|W_n - \frac{m+1}{2m+1}n\Big|\Big] \leq \sqrt{\mathbb{E}\Big[\Big(W_n - \frac{m+1}{2m+1}n\Big)^2\Big]} = O(n^{\frac{1}{2}}),$$

which completes the proof. □

[1] By saying a sequence of random variables W_n is $O_{L_1}(g(n))$, we mean there exists a positive constant C and a positive integer n_0, such that $\mathbb{E}[W_n] \leq C|g(n)|$, for all $n \geq n_0$.

Corollary 2. *We have*

$$\begin{aligned}
\frac{W_n}{n} &\xrightarrow{L_1} \frac{m+1}{2m+1} \\
\frac{W_n^2}{n^2} &\xrightarrow{L_1} \Big(\frac{m+1}{2m+1}\Big)^2.
\end{aligned}$$

So, both convergences take place in probability, as well.

In fact, according to the proof of Theorem 2, we can also conclude that $\frac{W_n}{n}$ converges in L_2 space.

4.3 The Martingale Structure

According to (5), W_n is not a martingale, but we can transform W_n into a martingale by introducing two deterministic functions α_n and β_n. In the following lemma, we prove that $M_n = \alpha_n W_n + \beta_n$ is a martingale for suitable choice of α_n and β_n.

Lemma 1.

$$M_n = \frac{\Gamma\big(n + \frac{m}{m+1}\big)}{\Gamma(n)\,\Gamma\big(1 + \frac{m}{m+1}\big)} W_n - \frac{(m+1)\Gamma\big(n + 1 + \frac{m}{m+1}\big)}{(2m+1)\Gamma(n)\,\Gamma\big(1 + \frac{m}{m+1}\big)}$$

is a martingale for $n \geq 1$.

Proof. Set

$$M_n = \alpha_n W_n + \beta_n$$

for yet-to-be computed deterministic functions that will render M_n a martingale. Recall that we also denote the σ-field generated by the first n evolutionary steps by $\mathbb{F}_n$. According to the fundamental property of martingales, we want to have

$$\begin{aligned}
\mathbb{E}[M_n \,|\, \mathbb{F}_{n-1}] &= \alpha_n \mathbb{E}[W_n \,|\, \mathbb{F}_{n-1}] + \beta_n \\
&= \alpha_n \Big(\frac{(m+1)(n-1)}{(m+1)(n-1)+m} W_{n-1} + 1\Big) + \beta_n \\
&= M_{n-1} \\
&= \alpha_{n-1} W_{n-1} + \beta_{n-1}.
\end{aligned}$$

This is possible, if we equate the coefficients of W_{n-1} and also the constant terms. Equating the coefficients of

W_{n-1} gives the recurrence for α_n. For $n \geq 2$, we obtain

$$\alpha_n = \frac{(m+1)(n-1)+m}{(m+1)(n-1)}\,\alpha_{n-1},$$

which has a solution

$$\alpha_n = \prod_{j=2}^{n} \frac{(m+1)(j-1)+m}{(m+1)(j-1)}\,\alpha_1,$$

for any arbitrary value α_1. For simplicity, let us take $\alpha_1 = 1$, and simplify the solution to a ratio of product of Gamma functions as follows. For $n \geq 1$, we have

$$\alpha_n = \frac{\Gamma\big(n + \frac{m}{m+1}\big)}{\Gamma(n)\,\Gamma\big(1+\frac{m}{m+1}\big)}, \tag{7}$$

Equating the constant terms gives

$$\beta_n = \beta_{n-1} - \alpha_n.$$

Solving this recurrence, we have

$$\beta_n = \beta_0 - \sum_{j=1}^{n} \alpha_j,$$

for any arbitrary value β_0. Setting $\beta_0 = 0$ and using the determined values of α_n in (7), we have

$$\beta_n = -\frac{(m+1)\Gamma\big(n+1+\frac{m}{m+1}\big)}{(2m+1)\Gamma(n)\,\Gamma\big(1+\frac{m}{m+1}\big)}. \tag{8}$$

□

Corollary 3. *As $n \to \infty$, the scaling (α_n) and shifting (β_n) coefficients in the martingale structure have the following asymptotic values:*

$$\alpha_n \sim \frac{1}{\Gamma\big(1+\frac{m}{m+1}\big)} n^{\frac{m}{m+1}}, \tag{9}$$

$$\beta_n \sim -\frac{m+1}{(2m+1)\Gamma\big(1+\frac{m}{m+1}\big)} n^{\frac{2m+1}{m+1}}. \tag{10}$$

Proof. We obtain the asymptotic values of α_n and β_n by applying Stirling's approximation to (7) and (8), respectively. □

To get a Gaussian law for $M_n(W_n)$, we appeal to the martingale central limit theorem. The conditions that we plan to check are a set of *conditional Lindeberg's condition* and *conditional variance condition*, see [4] for more details. To verify this set of conditions, we introduce an operation, ∇, which denotes the backward difference operator. For instance, we define $\nabla h_n = h_n - h_{n-1}$, for $n \in \mathbb{N}$.

The conditional Lindeberg's condition holds if, for some positive increasing sequence $\{\xi_n\}$, and for all $\varepsilon > 0$, we have

$$U_n := \sum_{j=1}^{n} \mathbb{E}\Big[\Big(\frac{\nabla M_j}{\xi_n}\Big)^2 \mathbb{I}\Big(\Big|\frac{\nabla M_j}{\xi_n}\Big| > \varepsilon\Big)\,\Big|\,\mathbb{F}_{j-1}\Big] \xrightarrow{P} 0;$$

the conditional variance condition requires that, for some square integrable random variable $Y \neq 0$, we have

$$V_n := \sum_{j=1}^{n} \mathbb{E}\Big[\Big(\frac{\nabla M_j}{\xi_n}\Big)^2\,\Big|\,\mathbb{F}_{j-1}\Big] \xrightarrow{P} Y.$$

When these conditions are satisfied, we get

$$\frac{M_n}{\xi_n} \xrightarrow{D} \mathcal{N}(0, Y),$$

where the right-hand side is a mixture of normally distributed random variables, with Y being the mixer. It will turn out that the correct scale factor ξ_n is $n^{(3m+1)/[2(m+1)]}$, and Y is deterministic, and thus there is only one normal distribution in the mixture.

We construct a uniform bound for ∇M_j after proper scaling for $1 \leq j \leq n$, shown in the following lemma, to verify the conditional Lindeberg's condition.

Lemma 2. *The terms $\Big|\frac{\nabla M_j}{n^{\frac{m}{m+1}}}\Big|$ are uniformly bounded in both j and n, for $1 \leq j \leq n$.*

Proof. By the definition of operation ∇ and the construction of martingale, we write the absolute difference,

$|\nabla M_j|$ for $j = 1, 2, \ldots, n$, as

$$\begin{aligned}|\nabla M_j| &= |M_j - M_{j-1}| \\ &= |\alpha_j W_j + \beta_j - (\alpha_{j-1} W_{j-1} + \beta_{j-1})| \\ &\le |\alpha_j W_j - \alpha_{j-1} W_{j-1}| + |\beta_j - \beta_{j-1}| \\ &= \alpha_j \Big| W_j - \frac{(m+1)(j-1)}{(m+1)(j-1)+m} W_{j-1} \Big| + \alpha_j \\ &\le \alpha_j (|W_j - W_{j-1}| + m + 1) \\ &\le \alpha_j (m - 1 + m + 1) \\ &= 2m\alpha_j .\end{aligned}$$

According to (9), we conclude that $\alpha_j \le c_m j^{\frac{m}{m+1}}$, for some positive c_m only depending on m, and this completes the proof. $\square$

Lemma 3.

$$U_n = \sum_{j=1}^{n} \mathbb{E}\Big[\Big(\frac{\nabla M_j}{n^{\frac{3m+1}{2(m+1)}}} \Big)^2 \mathbb{I}\Big(\Big| \frac{\nabla M_j}{n^{\frac{3m+1}{2(m+1)}}} \Big| > \varepsilon \Big) \,\Big|\, \mathbb{F}_{j-1} \Big] \xrightarrow{P} 0.$$

Proof. For every $\varepsilon > 0$, there exists $n_0(\varepsilon) > 0$, such that for all $n > n_0(\varepsilon)$, the sets $\big\{ \big| \frac{\nabla M_j}{n^{(3m+1)/[2(m+1)]}} \big| > \varepsilon \big\}$ are empty. Therefore, the sum in U_n can be truncated at $n_0(\varepsilon)$, and by the uniform bound established in Lemma 2, we can write

$$\begin{aligned}U_n &= \sum_{j=1}^{n_0(\varepsilon)} \mathbb{E}\Big[\Big(\frac{\nabla M_j}{n^{\frac{3m+1}{2(m+1)}}} \Big)^2 \mathbb{I}\Big(\frac{\nabla M_j}{n^{\frac{3m+1}{2(m+1)}}} > \varepsilon \Big) \,\Big|\, \mathbb{F}_{j-1} \Big] \\ &\le \sum_{j=1}^{n_0(\varepsilon)} \mathbb{E}\Big[\Big(\frac{\nabla M_j}{n^{\frac{3m+1}{2(m+1)}}} \Big)^2 \,\Big|\, \mathbb{F}_{j-1} \Big] \\ &\le \frac{n_0(\varepsilon)(2mc_m)^2}{n} \\ &\to 0.\end{aligned}$$

This convergence is stronger than the required in-probability convergence, which completes the verification of Lindeberg's condition. $\square$

In the next lemma, we verify the conditional variance condition.

Lemma 4.

$$\begin{aligned}V_n = \sum_{j=1}^{n} \mathbb{E}\Big[\Big(\frac{\nabla M_j}{n^{\frac{3m+1}{2(m+1)}}} \Big)^2 \,\Big|\, \mathbb{F}_{j-1} \Big] \\ \xrightarrow{P} \frac{2m^2(m+1)}{(2m+1)^2(3m+1)\Gamma\big(1 + \frac{m}{m+1}\big)}.\end{aligned}$$

Proof. Write

$$V_n = \frac{1}{n^{\frac{3m+1}{m+1}}} \sum_{j=1}^{n} \mathbb{E}\big[\big(\nabla(\alpha_j W_j) + \nabla\beta_j \big)^2 \,\big|\, \mathbb{F}_{j-1} \big].$$

We expand the quadratic form in the conditional expectation,

$$\sum_{j=1}^{n} \mathbb{E}\big[\big(\nabla(\alpha_j W_j)\big)^2 + 2\nabla(\alpha_j W_j)\nabla\beta_j + (\nabla\beta_j)^2 \,\big|\, \mathbb{F}_{j-1} \big],$$

and consider the summand in three parts.

1. The first part is

$$\begin{aligned}&\mathbb{E}\big[\big(\nabla(\alpha_j W_j)\big)^2 \,\big|\, F_{j-1} \big] \\ &= \mathbb{E}\big[(\alpha_j W_j - \alpha_{j-1} W_{j-1})^2 \,|\, F_{j-1} \big] \\ &= \alpha_j^2 \mathbb{E}[W_j^2 \,|\, \mathbb{F}_{j-1}] + \alpha_{j-1}^2 W_{j-1}^2 \\ &\qquad - 2\alpha_j \alpha_{j-1} W_{j-1} \mathbb{E}[W_j \,|\, \mathbb{F}_{j-1}] \\ &= \Big(\alpha_j^2 \frac{(m+1)(j-1)(mj+j-m-2)}{(mj+j-1)(mj+j-2)} + \alpha_{j-1}^2 \\ &\qquad - 2\alpha_j \alpha_{j-1} \frac{(m+1)(j-1)}{(m+1)(j-1)+m} \Big) W_{j-1}^2 \\ &\quad + \Big(\alpha_j^2 \frac{(m+1)(j-1)(2mj+m+2j-4)}{(mj+j-1)(mj+j-2)} \\ &\qquad - 2\alpha_j \alpha_{j-1} \Big) W_{j-1} + \alpha_j^2 .\end{aligned}$$

2. The second part is

$$\begin{aligned}&\mathbb{E}\big[2\nabla(\alpha_j W_j)\nabla\beta_j \,\big|\, F_{j-1} \big] \\ &= 2\mathbb{E}[(\alpha_j W_j - \alpha_{j-1} W_{j-1})(\beta_j - \beta_{j-1}) \,|\, \mathbb{F}_{j-1}] \\ &= -2\big(\alpha_j^2 \mathbb{E}[W_j \,|\, \mathbb{F}_{j-1}] - \alpha_j \alpha_{j-1} W_{j-1} \big) \\ &= 2\Big(\alpha_j \alpha_{j-1} - \alpha_j^2 \frac{(m+1)(j-1)}{(m+1)(j-1)+m} \Big) W_{j-1} \\ &\qquad - 2\alpha_j^2 .\end{aligned}$$

3. The third part is
$$\mathbb{E}[(\nabla\beta_j)^2 \,|\, \mathbb{F}_{j-1}] = \mathbb{E}[(\beta_j - \beta_{j-1})^2 \,|\, F_{j-1}] = \alpha_j^2.$$

Apply the asymptotic value of W_n in Theorem 2 and the asymptotic value of α_n in Corollary 3, and we find that the summand in V_n is
$$\frac{2m^2}{(2m+1)^2\Gamma^2\big(1+\frac{m}{m+1}\big)} j^{\frac{2m}{m+1}} + O_{L_1}\big(j^{\frac{3m-1}{2m+2}}\big).$$
Summing these terms up for $j = 1, 2, \dots, n$ and letting $n \to \infty$, we obtain
$$V_n = \frac{2m^2}{(2m+1)^2\Gamma^2\big(1+\frac{m}{m+1}\big)n^{\frac{3m+1}{m+1}}}\Big(\frac{m+1}{3m+1}n^{\frac{3m+1}{m+1}} + O_{L_1}\big(n^{\frac{5m+1}{2m+2}}\big)\Big)$$
$$\xrightarrow{L_1} \frac{2m^2(m+1)}{(2m+1)^2(3m+1)\Gamma^2\big(1+\frac{m}{m+1}\big)}.$$
This L_1 convergence is stronger than the in-probability convergence required in conditional variance condition. □

Having checked the set of conditions for the martingale central limit theorem, we conclude that

(11)
$$\frac{M_n}{n^{\frac{3m+1}{2(m+1)}}} \xrightarrow{D} \mathcal{N}\Big(0, \frac{2m^2(m+1)}{(2m+1)^2(3m+1)\Gamma^2\big(1+\frac{m}{m+1}\big)}\Big).$$

Theorem 3. *Let W_n be the number of white external nodes in a preferential attachment circuit with index m at time n. Then, we have*
$$\frac{W_n - \frac{m+1}{2m+1}n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}\Big(0, \frac{2m^2(m+1)}{(2m+1)^2(3m+1)}\Big).$$

Proof. Rewrite (11) as
$$\frac{\alpha_n W_n + \beta_n}{n^{\frac{3m+1}{2(m+1)}}} \xrightarrow{D} \mathcal{N}\Big(0, \frac{2m^2(m+1)}{(2m+1)^2(3m+1)\Gamma^2\big(1+\frac{m}{m+1}\big)}\Big).$$
Expressing β_n in terms of α_n and using the asymptotic value of α_n in (9), we arrive at
$$\frac{\frac{n^{\frac{m}{m+1}}}{\Gamma\left(1+\frac{m}{m+1}\right)}W_n - \frac{(m+1)n^{\frac{2m+1}{m+1}}}{(2m+1)\Gamma\left(1+\frac{m}{m+1}\right)}}{n^{\frac{3m+1}{2(m+1)}}} \xrightarrow{D} \mathcal{N}\Big(0, \frac{2m^2(m+1)}{(2m+1)^2(3m+1)\Gamma^2\big(1+\frac{m}{m+1}\big)}\Big),$$
which is equivalent to the stated convergence in distribution. □

5 Degree Profile

In this section, we study the evolution of the degree of a node as the circuit ages. Let $D_{j,n}^{(m)}$, for $j = 0, \dots n$, be the degree of node j in a preferential attachment circuit of index m and of age n. The random variables $D_{j,n}^{(m)}$, for $j = 0, \dots, n$, describe a profile of degrees in the random circuit.

Upon inserting the nth node (i.e., at time $n > j$), we allocate node n and choose m parent nodes (dynamically, one by one as discussed) for it. Each time node j appears in the dynamic sampling, its degree goes up by one; otherwise, its degree remains the same. The transition from the $(n-1)$st to the nth insertion coincides with the working of a two-color Pólya-Eggenbegrer urn scheme, as we shall shortly discuss. We would like to alert the reader to that we are not employing one continual urn underlying the circuit growth process. Rather, for each transition to add a new node there is an underlying urn. Right after the insertion of node n in the circuit, we have used n different urns, and the circuit experiences "hiccups" in between, jolting the contents of an urn at an insertion to become the starting conditions for the next urn used in the next insertion. The notion of a hiccup will be made precise in the following paragraphs.

Let us color external nodes attached to node j with white, and color those that are attached to other nodes with blue. We shall reuse the notations W_n and B_n, but here they mean the number of white external nodes dangling out of node j after n insertions, and the number of blue balls dangling out of all the other nodes after n insertions. As the originator does not have parents, while each other node has m parents, we have the relation
$$D_{j,n}^{(m)} = \begin{cases} W_n - 1, & j = 0; \\ W_n - 1 + m, & j > 0. \end{cases}$$

For compactness, we write this double-decker expression in one line with the aid of Kronecker's delta $\delta_{\cdot,\cdot}$, so it takes the form

(12)
$$D_{j,n}^{(m)} = W_n - 1 + m(1 - \delta_{j,0}).$$

Consider the sampling process at time $n > j$ to choose m parents for the node n. At this point, the circuit has W_{n-1} white external nodes and B_{n-1} blue external nodes. When a white external node is selected, one white external node is added to the circuit; whereas when a blue external node is selected, one blue external node is adjoined to the circuit. This goes on for m steps. These additions are just like the dynamics of a standard Pólya–Eggenberger urn starting with W_{n-1} white balls and B_{n-1} blue balls and evolving in m draws (where one ball is sampled in each drawing), see [**?**]. The number of white (blue) external nodes added to the circuit is the same as the number of times a white (blue) ball is picked from such an urn during m ball draws.

And now comes the hiccup we alluded to. The urn after m ball additions (of either color) is not the right urn to model the next m parent selections, without an adjustment. At the end of the sampling process (i.e., after all m parents are selected), the newly added node (i.e., node n) acquires one additional blue external node. So, the next urn that models the transition from the nth insertion step to the $(n+1)$st step needs one extra blue ball than those at the end of adding m balls to the urn used in modeling step n.

Pólya–Eggenberger urns enjoy an exchangeability property that is helpful to our analysis—the order of choosing white and blue balls in all m-long sequences that have the same number of white balls in them have the same probability. For example, the sequence $WWBWBW$ has the same probability as the sequence $BBWWWW$, where W's and B's stand for white and blue choices. This is a fundamental component in building the well-known Pólya distribution; see [8]. If U is the number of blue balls added in the nth step, we have

$$\mathbb{P}(U = k \,|\, \mathbb{F}_{n-1}) = \binom{m}{k} \frac{\langle W_{n-1}\rangle_{m-k} \langle B_{n-1}\rangle_k}{\langle \tau_{n-1}\rangle_m},$$

where $\tau_{n-1} = W_{n-1} + B_{n-1}$.

Theorem 4. *Let $D_{n,j}^{(m)}$ be the degree of node j of a preferential dynamic attachment circuit with index m at time n. For $n \ge j$, we have*

$$\begin{aligned}
&\mathbb{P}\big(D_{n,j}^{(m)} = d\big)\\
&= (d - m(1-\delta_{j,0}))!\\
&\qquad \times \sum_{\substack{b_1+\cdots+b_{n-j}=m(n-j+1-\delta_{j,0})-d \\ 0\le b_1\le \min\{m,(m+1)j\}}} \left(\prod_{r=1}^{n-j} \binom{m}{b_r}\right)\\
&\qquad\qquad \times \frac{\prod_{r=1}^{n-j} \langle (m+1)j + \sum_{\ell=1}^{r-1} b_\ell + r - 1\rangle_{b_r}}{\prod_{r=0}^{n-j-1} \langle (m+1)(j+r)+1\rangle_m}.
\end{aligned}$$

Proof. We wait until node j appears in the circuit, then monitor the growth thereafter. Let X_n be the number of blue external nodes in the circuit (balls in the urn) that appear in the sample of inserting node n. This variable has the distribution of the Pólya–Eggenberger urn used in the nth node insertion. We note again, the urns in the different node insertions are different. For the insertion of node $j+1$, we have an urn starting with one white external node and $(m+1)j$ blue external nodes. Thus, for $0 \le b_1 \le \min\{m, (m+1)j\}^2$, we have

$$\mathbb{P}(X_{j+1} = b_1) = \binom{m}{b_1} \frac{(m-b_1)!\,\langle (m+1)j\rangle_{b_1}}{\langle \tau_j\rangle_m}.$$

Right before inserting node $j+2$, one additional blue external node is added to node $j+1$ (the hiccup mentioned above). Therefore, conditioning on the event $X_{j+1} = b_1$, before inserting node $j+2$, we have $1+m-b_1$ white external nodes and $(m+1)j + b_1 + 1$ blue external nodes. According to the urn associated with this insertion, we compute

$$\begin{aligned}
\mathbb{P}(X_{j+2} = b_2 \,|\, X_{j+1} = b_1) = \binom{m}{b_2} &\frac{\langle m - b_1 + 1\rangle_{m-b_2}}{\langle \tau_{j+1}\rangle_m}\\
&\times \langle (m+1)j + b_1 + 1\rangle_{b_2}.
\end{aligned}$$

So, we have the joint distribution of the number of exter-

[2] This condition is only needed when $j = 0$, in which cases b_1 only can be taken 0.

nal nodes that appear in the first two insertions:

$$
\begin{aligned}
&\mathbb{P}(X_{j+1} = b_1, X_{j+2} = b_2) \\
&= \mathbb{P}(X_{j+2} = b_2 \,|\, X_{j+1} = b_1)\, \mathbb{P}(X_{j+1} = b_1) \\
&= \binom{m}{b_1} \frac{(m-b_1)!\, \langle (m+1)j \rangle_{b_1}}{\langle \tau_j \rangle_m} \\
&\qquad \times \binom{m}{b_2} \frac{\langle m - b_1 + 1 \rangle_{m-b_2} \langle (m+1)j + b_1 + 1 \rangle_{b_2}}{\langle \tau_{j+1} \rangle_m} \\
&= \binom{m}{b_1}\binom{m}{b_2} \frac{(2m - b_1 - b_2)!\, \langle (m+1)j \rangle_{b_1}}{\langle \tau_j \rangle_m} \\
&\qquad \times \frac{\langle (m+1)j + b_1 + 1 \rangle_{b_2}}{\langle \tau_{j+1} \rangle_m}.
\end{aligned}
$$

We can get the joint distribution of $X_{j+1}, X_{j+2}, X_{j+3}$, by conditioning on $X_{j+1} = b_1$ and $X_{j+2} = b_2$, then uncondition via the joint distribution of X_{j+1}, X_{j+2}. We can continue in this fashion to get the joint distribution of $X_{j+1}, X_{j+2}, \dots, X_n$. We establish

$$
\begin{aligned}
&\mathbb{P}(X_{j+1} = b_1, \dots, X_n = b_{n-j}) \\
&= \Big((n-j)m - \sum_{r=1}^{n-j} b_r \Big)! \left(\prod_{r=1}^{n-j} \binom{m}{b_r} \right) \\
&\qquad \times \frac{\prod_{r=1}^{n-j} \langle (m+1)j + \sum_{\ell=1}^{r-1} b_\ell + r - 1 \rangle_{b_r}}{\prod_{s=j}^{n-1} \langle \tau_s \rangle_m}.
\end{aligned}
$$

The exact probability distribution of the number of blue external nodes (blue balls) at time n follows by summing the joint probabilities over every feasible tuple $(b_1, \dots, b_{n-j})$ with nonnegative components adding up to $b - mj - n$, giving us:

$$
\begin{aligned}
&\mathbb{P}(B_n = b) \\
&= ((m+1)n - b)! \sum_{b_1 + \cdots + b_{n-j} = b - mj - n} \left(\prod_{r=1}^{n-j} \binom{m}{b_r} \right) \\
&\qquad \times \frac{\prod_{r=1}^{n-j} \langle (m+1)j + \sum_{\ell=1}^{r-1} b_\ell + r - 1 \rangle_{b_r}}{\prod_{r=0}^{n-j-1} \langle (m+1)(j+r) + 1 \rangle_m}.
\end{aligned}
$$

The theorem follows by calculating the exact probability distribution when $b = (m+1)n - d + m(1 - \delta_{j,0})$, which is equivalent to the probability of $W_n = d - m(1 - \delta_{j,0}) + 1$, and W_n translates into $D_{n,j}^{(m)}$ by (12). □

The probability mass function of $D_{n,j}^{(m)}$ is unwieldy. In general, it does not give good insight into the asymptotic distribution associated with short and long term insertions in the circuit. However, we are able to investigate the exact distribution of the degree profile of some particular nodes from simple networks. For instance, when $m = 1$ and $j = 0$, the random variable $D_{n,0}^{(1)}$ represents the degree of the root of a PORT at time n. According to Theorem 4, we simplify the distribution function and it arrive at

$$
\mathbb{P}(D_{n,0}^{(1)} = d) = \frac{(2n - d - 1)!\, d}{(n-d)!(2n-1)!!\, 2^{n-d}}.
$$

As it is not easy to use the exact distribution to do a calculation of the exact or asymptotic average or variance for bigger node labels j or more complicated circuits (i.e., big m), we resort to transparent stochastic recurrence techniques.

Consider W_n, the number of white external nodes at time n, according to the known results of *Pólya–Eggenberger Urn* schemes (see [8]), we have

$$
\mathbb{E}[W_n \,|\, \mathbb{F}_{n-1}] = \frac{W_{n-1}}{\tau_{n-1}} m + W_{n-1}, \tag{13}
$$

$$
\mathbb{V}\mathrm{ar}[W_n \,|\, \mathbb{F}_{n-1}] = \frac{W_{n-1} B_{n-1} m (m + \tau_{n-1})}{\tau_{n-1}^2 (\tau_{n-1} + 1)}. \tag{14}
$$

We want to point out that the total number of external nodes (i.e., τ_n) in the recurrences fixes the problem of one additional blue external node added to the circuit at the end of every sampling process (i.e., hiccup). Solving the recurrences, we obtain the exact expectation and variance of W_n as stated in the following theorem.

Theorem 5. *Let $D_{n,j}^{(m)}$ be the degree of node j from a preferential dynamic attachment circuit with index m at* *time n. For $n \geq j$, we have*

$$\mathbb{E}\big[D_{n,j}^{(m)}\big] = \frac{\Gamma(n+1)\,\Gamma\big(j+\frac{1}{m+1}\big)}{\Gamma\big(n+\frac{1}{m+1}\big)\,\Gamma(j+1)} - 1 + m(1-\delta_{j,0}),$$

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\big[D_{n,j}^{(m)}\big]\\ &= \frac{\Gamma(n+1)}{((m+1)j+1)\,\Gamma^2\big(n+\frac{1}{m+1}\big)\,\Gamma^2(j+1)\,\Gamma\big(n+\frac{2}{m+1}\big)}\\ &\quad\times\Big\{2((m+1)n+1)\,\Gamma(j+1)\,\Gamma^2\Big(n+\frac{1}{m+1}\Big)\\ &\quad\times\Gamma\Big(j+\frac{2}{m+1}\Big) - ((m+1)j+1)\Big[\Gamma(n+1)\\ &\quad\times\Gamma\Big(n+\frac{2}{m+1}\Big)\,\Gamma^2\Big(j+\frac{1}{m+1}\Big)\\ &\quad+\Gamma\Big(n+\frac{2}{m+1}\Big)\,\Gamma\Big(n+\frac{1}{m+1}\Big)\,\Gamma(j+1)\\ &\quad\times\Gamma\Big(j+\frac{1}{m+1}\Big)\Big]\Big\}.\end{aligned}$$

Proof. We consider the external nodes in preferential attachment circuits. We obtain the recurrence for $\mathbb{E}[W_n]$ by taking another expectation of (13). In addition to the initial condition $W_j \equiv 1$, we get

$$\mathbb{E}[W_n] = \frac{\Gamma(n+1)\,\Gamma\big(j+\frac{1}{m+1}\big)}{\Gamma\big(n+\frac{1}{m+1}\big)\,\Gamma(j+1)}.$$

According to the law of total variance and the result of $\mathbb{E}[W_n]$, we have

$$\begin{aligned}&\mathbb{V}\mathrm{ar}[W_n]\\ &= \mathbb{E}\big[\mathbb{V}\mathrm{ar}[W_n \,|\, \mathbb{F}_{n-1}]\big] + \mathbb{V}\mathrm{ar}\big[\mathbb{E}[W_n \,|\, \mathbb{F}_{n-1}]\big]\\ &= \frac{((m+1)n+1)n(m+1)}{((m+1)n-m+1)((m+1)n-m)}\mathbb{V}\mathrm{ar}[W_{n-1}]\\ &\quad-\frac{mn(m+1)\,\Gamma\Big(j+\frac{1}{m+1}\Big)\,\Gamma(n)}{((m+1)n-m)^2((m+1)n-m+1)}\\ &\quad\times\frac{1}{\Gamma^2\big(n-\frac{m}{m+1}\big)\Gamma(j+1)^2}\Big(\Gamma\Big(j+\frac{1}{m+1}\Big)\,\Gamma(n)\\ &\quad-\Gamma(j+1)\,\Gamma\Big(n-\frac{m}{m+1}\Big)((m+1)n-m)\Big).\end{aligned}$$

The exact variance of W_n is obtained by solving the variance recurrence above with the initial condition $\mathbb{V}\mathrm{ar}[W_j] = 0$.

Finally, both $\mathbb{E}[W_n]$ and $\mathbb{V}\mathrm{ar}[W_n]$ translate to $\mathbb{E}\big[D_{n,j}^{(m)}\big]$ and $\mathbb{V}\mathrm{ar}\big[D_{n,j}^{(m)}\big]$ via relation (12). □

Corollary 4. *As $n \to \infty$, the asymptotic expectation and variance of $D_{n,j}^{(m)}$ are*

$$\mathbb{E}\big[D_{n,j}^{(m)}\big] \sim \frac{\Gamma\big(j+\frac{1}{m+1}\big)}{\Gamma(j+1)}\, n^{\frac{m}{m+1}},$$

$$\mathbb{V}\mathrm{ar}\big[D_{n,j}^{(m)}\big] \sim \Big(\frac{2(m+1)\,\Gamma\big(j+\frac{2}{m+1}\big)}{((m+1)j+1)\,\Gamma(j+1)} - \frac{\Gamma^2\big(j+\frac{1}{m+1}\big)}{\Gamma^2(j+1)}\Big)\, n^{\frac{2m}{m+1}}.$$

Proof. The asymptotic expectation and variance of $D_{n,j}^{(m)}$ are obtained by applying Stirling's approximation to the exact expectation and variance of $D_{n,j}^{(m)}$ in Theorem 5. □

We have an *exact* expression for the mean, and we can analyze it asymptotically for different regimes of j, whereupon we discover "phase changes." If $j = O(1)$, we have

$$\mathbb{E}\big[D_{n,j}^{(m)}\big] \sim \frac{\Gamma\big(j+\frac{1}{m+1}\big)}{\Gamma(j+1)}\, n^{\frac{m}{m+1}}$$

as stated in Corollary 4. For $j = j_n$, $n \to \infty$, the degree is not asymptotically much different from the case of fixed j, though we can simplify the gamma functions involving j_n via Stirling's approximation. For such slowly growing nodes j_n (e.g., $j_n = \lfloor\sqrt{n}+3\rfloor$), we have

$$\mathbb{E}\big[D_{n,j}^{(m)}\big] \sim \Big(\frac{n}{j_n}\Big)^{\frac{m}{m+1}}.$$

If j_n grows with n in such a way that $j_n/n \to \theta$, with $0 < \theta \leq 1$ (e.g., $j_n = \lceil\frac{1}{6}n + 3\ln n\rceil$), we have

$$\mathbb{E}\big[D_{n,j}^{(m)}\big] \sim \frac{1}{\theta^{\frac{m}{m+1}}} - 1 + m.$$

Note something special about the case $\theta = 1$, (e.g., $j_n = \lfloor\sqrt{n} - 2\ln\ln n\rfloor$), where we get $\mathbb{E}\big[D_{n,j}^{(m)}\big] \sim m$. These very late arrivals join, and on average they do not recruit (outdegree 0). The only contribution to their degree is m (indegree), coming from the connections to the parents.

Acknowledgment

I would like to thank Professor Hosam Mahmoud for insightful advice on this paper. I also want to thank three anonymous reviewers for helpful suggestions and critical comments.

References

[1] Drmota, M. , Gittenberger, B. and Panholzer, A. (2008). The degree distribution of thickened trees. *Discrete Mathematics and Theoretical Computer Science*; *Proceedings of the Fifth Colloquium on Mathematics and Computer Science*, **Proceedings AI**, 149–152.

[2] Fuchs, M. Hwang, H.-K. and Neininger, R. (2006) Profile of Random Trees: Limit Theorems for Random Recursive Trees and Binary Search Trees. *Algorithmica*, **46**, 367–407.

[3] Graham, R. , Knuth, D. and Patashnik, O. (1994). *Concrete Mathematics*. Addison-Wesley, Reading, Massachusetts.

[4] Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, Inc., New York.

[5] Hwang, H.-K. (2007). Profiles of random trees: plane-oriented recursive trees. *Random Structures and Algorithms*, **30**, 380–413.

[6] Mahmoud, H. Smythe, R. and Szymanski, J. (1993). On the structure of plane-oriented recursive trees and their branches. *Random Structures and Algorithms*, **4**, 151–176.

[7] Mahmoud, H. and Tsukiji, T. (2004). Limit laws for terminal nodes in random circuits with restricted fan-out: a family of graphs generalizing binary search trees. Acta Informatica, 42, 1432-0525.

[8] Mahmoud, H. (2008). *Pólya Urn Models*. Chapman-Hall, Orlando, Florida.

[9] Mahmoud, H. (2014). The degree profile in some classes of random graphs that generalize recursive trees. *Methodology and Computing in Applied Probability*, **16**, 643–673.

[10] Gopaladesikan, M. and Mahmoud, H. and Ward, M. (2014) Asymptotic Joint Normality of Counts of Uncorrelated Motifs in Recursive Trees *Methodology and Computing in Applied Probability*, **16**, 863–884.

[11] Moler, J., Plo, F., and Urmeneta, H. (2013). A generalized Pólya urn and limit laws for the number of outputs in a family of random circuits. *Test*, **22**, 46–61.

[12] Peköz, E. , Röllin, A. and Ross, N. (2013). Degree Asymptotics with rates for preferential attachment random graphs. *The Annals of Applied Probability*, **23**, 1188–1218.

[13] Ross, N. (2013) Power laws in preferential attachment graphs and Stein's method for the negative binomial distribution. *Advances in Applied Probability*, **45**, 876–893.

[14] Tsukiji, T. and Mahmoud, H. (2001). A limit law for outputs in random circuits. *Algorithmica*, **31**, 403–412.