Internet Mathematics Vol. 3, No. 2: 207-231

PageRank of Scale-Free Growing Networks

Konstantin Avrachenkov and Dmitri Lebedev

Abstract. PageRank is one of the principle criteria according to which Google ranks webpages. PageRank can be interpreted as a frequency of webpage visits by a random surfer, and thus it reflects the popularity of a webpage. In the present work we find an analytical expression for the expected PageRank value in a scale-free growing network model as a function of the age of the growing network and the age of a particular node. Then, we derive asymptotics that show that PageRank follows closely a power law in the middle range of its values. The exponent of the theoretical power law matches very well the value found from measurements of the World Wide Web. Finally, we provide a mathematical insight for the choice of the damping factor in PageRank definition.

1. Introduction

Surfers on the Internet frequently use search engines to find pages satisfying their query. However, there are typically hundreds or thousands of relevant pages available on the web. Thus, listing them in a proper order is a crucial and nontrivial task. One can use several criteria to sort relevant answers. It turns out that the link-based criteria which represent well the popularity of webpages provide rankings that appear to be very satisfactory to Internet users. Examples of link-based criteria are PageRank [Brin and Page 98] used by search engine Google, HITS [Kleinberg 99] used by search engines Teoma and Ask, and SALSA [Lempel and Moran 00]. In the present work we restrict ourselves to the analysis of the PageRank criterion and use the following definition of PageRank from [Langville and Meyer 06]. Denote by n the total number of pages on the web and define the $n \times n$ hyperlink matrix P as follows. Suppose that page i

1542-7951/06 $0.50 per page

has $k > 0$ outgoing links. Then, $p_{ij} = 1/k$ if j is one of the outgoing links and $p_{ij} = 0$ otherwise. If a page does not have outgoing links, the probability is spread among all pages of the web, namely, $p_{ij} = 1/n$. In order to make the hyperlink graph connected, it is assumed that a random surfer goes with some probability to an arbitrary webpage with uniform distribution. Thus, the PageRank is defined as a stationary distribution of a Markov chain whose state space is the set of all webpages, and the transition matrix is

$$\tilde{P} = cP + (1-c)(1/n)E, \tag{1.1}$$

where E is a matrix whose entries are all equal to one and $c \in (0,1)$ is the probability of following a link on the page and not jumping to a random page (it is chosen by Google to be 0.85). The constant c is often referred to as a *damping factor*. The Google matrix $\tilde{P}$ is stochastic, aperiodic, and irreducible, so there exists a unique row vector π such that

$$\pi\tilde{P} = \pi, \quad \pi\underline{1} = 1, \tag{1.2}$$

where $\underline{1}$ is a column vector of ones. The row vector π satisfying (1.2) is called a PageRank vector, or simply PageRank. If a surfer follows a hyperlink with probability c and jumps to a random page with probability $1-c$, then π_i can be interpreted as a stationary probability that the surfer is at page i.

Barabási and Albert [Barabási and Albert 99] proposed a scale-free growing network model to understand the evolution of the World Wide Web and in particular to explain the power law for in- and out-degree distributions. Bollobás et al. [Bollobás et al. 01] refined this model and proved rigorously that in- and out-degree distributions satisfy power laws. Pandurangan et al. [Pandurangan et al. 02] applied the “mean-field” heuristics from [Barabási and Albert 99, Barabási et al. 99, Barabási et al. 00] to show that the PageRank distribution in the scale-free growing network model satisfies the power law with exponent 2. They also proposed a model in which new nodes attach with weighted probability that takes into account the in-degree as well as PageRank. By studying two large samples of the web, the authors of [Pandurangan et al. 02] found that PageRank closely follows a power law with exponent 2.1.

In the present work we find an analytical expression for the expected PageRank value in a scale-free growing network model as a function of the age of the growing network and the age of a particular node. We prove that the average PageRank value does not depend on the number of outgoing links. This fact helps us significantly, since we can deal with tree graphs instead of directed acyclic graphs. Then, we derive asymptotics that show that PageRank follows closely a power law with exponent 2.08 in the middle range of its values. Finally, our expressions give a mathematical insight for the choice of the damping factor c.

The structure of the paper is as follows: In Section 2 we describe the scale-free growing network model, which is used in the present work, and its relation to the other scale-free growing network models. In Section 3 we derive an explicit formula for the PageRank of directed acyclic graphs and tree graphs. In Section 4 we prove that in our model the average PageRank does not depend on the number of outgoing links. Sections 5 and 6 provide auxiliary results on the moment-generating function of the nodes' heights in subtrees and on the subtree size distribution, which lead to the final results and asymptotics given in Section 7. The paper is concluded by Section 8, where we discuss the results and compare them with the related results from the literature. Some techniques that we use in the present work are explained in more detail in the appendices (Sections 9 and 10).

2. Scale-Free Network Models

Inspired by the power-law in- and out-degree distributions of the World Wide Web, Barabási and Albert [Barabási and Albert 99] proposed a growing network model with a preferential attachment mechanism. In their model a new node is attached to some old nodes with probability proportional to the in-degree of the old nodes. Barabási, Albert, and Jeong developed the "mean-field" heuristics, which allowed them to derive approximations to the power-law degree distributions [Barabási and Albert 99, Barabási et al. 99, Barabási et al. 00]. Then, Bollobás et al. [Bollobás et al. 01] added some missing parts to the Barabási-Albert model and showed rigorously that the degree distributions of the scale-free growing network model indeed satisfy power laws. The model in that paper allows self-loops and multiple links.

It turns out that there is an explicit analytic expression (it is given in the next section) for the PageRank of directed acyclic graphs. Furthermore, when Google computes the PageRank, it disregards the hyperlinks within the same webpage. Taking into account these two reasons, we have decided to work with the following scale-free growing network model: The time is discrete. The network grows at the speed of one node per time step. We fix a parameter m, the number of outgoing links from each node. At each time step a new node creates m links to the existing nodes. Let us denote the growing network at arbitrary time step n by G_n^m. At this point we need to define the way the links of a new node connect to the existing nodes. We denote by $d_v(n)$ the *in-degree* of node v at time step n.

- At time step 0, the initial node 0 is created and it has no links. The initial node has weight m by definition.

- Then, at the next time step 1, a new node has no other choice but to connect its m links to the initial node. Node 1 receives the weight m and the weight of node 0 becomes $2m$.

- A new node that appears after time step 1 connects each of its m edges independently with probability proportional to the existing nodes' weights equal to the in-degrees plus m. Namely, the probability that node n connects to node v, $v < n$, is given by

$$\mathbb{P}[n \to v | G^m_{n-1}] = \frac{d_v(n-1) + m}{\sum_{k=0}^{n-1}(d_k(n-1)+m)} = \frac{d_v(n-1)+m}{2m(n-1)+m}. \tag{2.1}$$

  For instance, node 2 connects with probability 2/3 to the initial node 0 and with probability 1/3 to node 1.

It is easy to see that, in the case of $m = 1$, the growing network G^1_n is a tree. This fact will be used extensively later in the paper. We would like to note that the scale-free growing network model of Dorogovtsev at al. [Dorogovtsev et al. 00] is the closest model to ours. An interested reader can find a detailed overview of growing network models in the surveys [Bollobás and Riordan 02, Dorogovtsev and Mendes 02, Newman 03].

3. PageRank of Growing Networks

Let us study the PageRank for growing networks with fixed out-degree m. We would like to emphasize that in this section we do not assume any preferential attachment of new nodes. It is only assumed that at each time step a new node is added to the network and makes m links to previously created nodes. Thus, if the initial node does not have any outgoing links, a growing network realization is a *directed acyclic graph* (DAG) at each time step. To calculate the PageRank one needs to attribute some outgoing links to the initial node. There are two natural options: either to make a self-loop in the initial node or to connect the initial node to all nodes in the network. The difference between these two cases is a value of the common factor for all nodes $v \geq 1$ [Langville and Meyer 05]. Since it turns out that this factor is much simpler in the case of the initial node with a self-loop, we choose the first option in the present work.

We denote by $\pi_v(n)$ the PageRank of node v after the nth step of the growing network's evolution. Of course, $n \geq v$. We note that at time step n the PageRank value of a newly created node n is minimal and is given by $\pi_n(n) = \frac{1-c}{n+1}$.

Let us denote by $P_v(n)$ the set of all paths from nodes $v+1,...,n$ to v and by $l(p)$ the length of a path p. Then, the PageRank vector of a growing network realization can be calculated by the explicit formula given in the following lemma.

Lemma 3.1. *The PageRank of a growing network realization of node v, $v > 0$, at time step n is given by*

$$\pi_v(n) = \frac{1-c}{n+1}\left(1 + \sum_{p \in P_v(n)} \left(\frac{c}{m}\right)^{l(p)}\right), \tag{3.1}$$

and the PageRank of the initial node $v = 0$ is given by

$$\pi_0(n) = \frac{1}{n+1}\left(1 + \sum_{p \in P_0(n)} \left(\frac{c}{m}\right)^{l(p)}\right). \tag{3.2}$$

Proof. The PageRank vector of any network can be expressed by the formula [Moler 04, Bianchini et al. 05, Langville and Meyer 05]

$$\pi = \frac{1-c}{n+1}\underline{1}^T[I - cP]^{-1}, \tag{3.3}$$

where $\underline{1}^T$ is the row vector of ones and P is the hyperlink matrix as in (1.1). We can rewrite the inverse matrix as a power series:

$$[I - cP]^{-1} = I + cP + c^2P^2 + \ldots .$$

Next, we note that the (i, j) element of matrix $[I - cP]^{-1}$ corresponds to the sum of the terms $(c/m)^{l(p)}$ over all possible paths from node i to node j. The premultiplication of $[I - cP]^{-1}$ by vector $\underline{1}^T$ gives the sum of all paths to node j. We note that all paths to node v originate from nodes which appeared after node v. In the case $v > 0$, there are no loops and hence we obtain formula (3.1). In the case $v = 0$, each path to the initial node ends with a self-loop. Because of this self-loop, each term $(c/m)^{l(p)}$ is multiplied by the series $1 + c + c^2 + ...$ The sum of the latter series is equal to $1/(1 - c)$, which cancels the factor $1 - c$ in (3.3) and results in the particular expression (3.2) for the PageRank of the initial node. □

Next, we note that if $m = 1$ every realization of the growing network becomes a tree. This simplifies further Equations (3.1) and (3.2). In the case $m = 1$, let us denote by $T_v(n)$ the subtree of the growing network with the root in node

v at time step n, $n > v$. We denote by $Y_v(n)$ the number of nodes in $T_v(n)$, not counting the root node v. Let us fix v and denote by $\{\tau_s\}$ the sequence of nodes that are attached to the subtree $T_v(n)$. In particular, $\tau_0 = v$, τ_1 is the first node that attaches to the subtree $T_v(n)$ (inevitably, it attaches to node v), τ_2 is the second node that attaches to the subtree $T_v(n)$, and so on. Thus, the subscript index s can be considered as the local time for subtree $T_v(n)$. Next, we denote by $X(v, s)$ the distance between node v and the node attached to $T_v(n)$ at time step τ_s. We shall also call this distance the *height* of node τ_s in $T_v(n)$. In particular, we have $X(v, 0) = 0$ and $X(v, 1) = 1$. The following corollary follows from Lemma 3.1.

Corollary 3.2. *If all the distances between the root node v and all nodes in $T_v(n)$ are known, then $\pi_v(n)$, the PageRank of node v, $v > 0$, can be expressed explicitly as follows:*

$$\pi_v(n) = \frac{1-c}{n+1}\left(1 + \sum_{\alpha \in T_v(n)} c^{l(v,\alpha)}\right), \tag{3.4}$$

where $l(v, \alpha)$ is the distance between nodes v and α, or in its alternative local time form with respect to the subtree $T_v(n)$,

$$\pi_v(n) = \frac{1-c}{n+1}\left(1 + \sum_{s=1}^{Y_v(n)} c^{X(v,s)}\right). \tag{3.5}$$

The PageRank of the initial node 0 is given by

$$\pi_0(n) = \frac{1}{n+1}\left(1 + \sum_{s=1}^{n} c^{X(0,s)}\right). \tag{3.6}$$

4. The Case $m > 1$ Reduced to the Case $m = 1$

It follows from Corollary 3.2 that the calculation of PageRank is much simpler in the case of tree graphs than in the case of directed acyclic graphs. In particular, in the case of tree graphs, there is a one-to-one correspondence between the paths and the nodes. Fortunately, as Theorem 4.1 demonstrates, the expected values of PageRank in the cases $m > 1$ and $m = 1$ are equal for the corresponding nodes of the same age. Denote by $\mathbb{E}\pi_v^m(n)$ the expected value of the PageRank of node v at time step n for our growing network model G_n^m.

Theorem 4.1. *In the present scale-free growing network model G_n^m, the average PageRank of node v does not depend on m. Namely, we have*

$$\mathbb{E}\pi_v^m(n) = \mathbb{E}\pi_v^1(n), \quad v < n. \tag{4.1}$$

Proof. The proof is done by induction on the node age. Thus, we fix v and consider time steps $n = v+1, v+2, \dots$.

As the induction base, consider node v at time step $v+1$. There is a new node $v+1$ that is being added to the network and this new node has m links, with j links to node v, $0 \le j \le m$, and $m-j$ links to the rest of the nodes. Let us find the expected value of PageRank for node v:

$$\mathbb{E}\pi_v^m(v+1) = \sum_{j=0}^{m} \frac{jc}{m}\frac{1-c}{v+2}\mathbb{P}\left[v+1 \text{ has } j \text{ links to } v\right] + \frac{1-c}{v+2}. \tag{4.2}$$

This is equal to

$$\mathbb{E}\pi_v^m(v+1) = \frac{c}{m}\frac{1-c}{v+2}\sum_{j=0}^{m} j\mathbb{P}\left[v+1 \text{ has } j \text{ links to } v\right] + \frac{1-c}{v+2}. \tag{4.3}$$

The probability that a link will be created from node $v+1$ to node v is equal to

$$\frac{m}{2mv+m} = \frac{1}{2v+1}.$$

Therefore, the sum in (4.3) is the average number of the links from node $v+1$ to node v or, in other words, the average number of the successes in m Bernoulli trials with the probability of success equal to $\frac{1}{2v+1}$. Therefore, we can write

$$\mathbb{E}\pi_v^m(v+1) = \frac{c}{m}\frac{1-c}{v+2}\frac{m}{2v+1} + \frac{1-c}{v+2} = \frac{c(1-c)}{(v+2)(2v+1)} + \frac{1-c}{v+2}. \tag{4.4}$$

Thus, $\mathbb{E}\pi_v^m(v+1)$ does not depend on m, and the induction base is proven.

Next, we consider node v at its age of t, or equivalently at time step $n = v+t$, and we suppose that all the average PageRanks $\mathbb{E}\pi_k^m(v+t)$ of the nodes k, $v < k \le v+t$, do not depend on m. The nodes k, $v < k \le v+t$, are the nodes that are "younger" than node v. We shall prove that $\mathbb{E}\pi_v^m(v+t)$, the expected value of the PageRank of node v at time step $v+t$, also does not depend on m.

Let us denote a realization of the network G_n^m at time step $v+t-1$ as λ. At time step $v+t$ a new node $v+t$ is born that connects itself with m links to the older nodes according to the preferential attachment rule. The PageRank of node v at time step $v+t$, knowing that the configuration at time step $v+t-1$ was λ, is given by

$$^{\lambda}\pi_v^m(v+t) = \sum_{k=v+1}^{v+t} \frac{c}{m} {}^{\lambda}\pi_k^m(v+t)\mathcal{M}\{k \to v,\ \lambda\} + \frac{1-c}{v+t+1}, \tag{4.5}$$

where $\mathcal{M}\{k \to v,\ \lambda\}$ is the number of edges from node k to node v. In particular, we note that the PageRank of an arbitrary node depends only on those nodes that appear later in time. Now we consider the expectation of (4.5) over all possible realizations λ:

$$\mathbb{E}\pi_v^m(v+t) = \sum_{k=v+1}^{v+t} \frac{c}{m}\mathbb{E}\left({}^{\lambda}\pi_k^m(v+t)\mathcal{M}\{k \to v,\ \lambda\}\right) + \frac{1-c}{v+t+1}. \tag{4.6}$$

We claim that ${}^{\lambda}\pi_k^m$ and $\mathcal{M}\{k \to v,\ \lambda\}$ are independent.

In fact, as mentioned above, the PageRank ${}^{\lambda}\pi_k^m$ of node k depends on the nodes that appear later than time step k, whereas the number of the links between k and v depends only on the nodes that appeared before node k due to the preferential attachment rule.

Therefore,

$$\mathbb{E}\left({}^{\lambda}\pi_k^m(v+t)\mathcal{M}\{k \to v,\ \lambda\}\right) = \mathbb{E}\left({}^{\lambda}\pi_k^m(v+t)\right)\mathbb{E}\left(\mathcal{M}\{k \to v,\ \lambda\}\right),$$

and we can write

$$\begin{aligned}\mathbb{E}\pi_v^m(v+t) &= \sum_{k=v+1}^{v+t} \frac{c}{m}\mathbb{E}\pi_k^m(v+t)\mathbb{E}\mathcal{M}\{k \to v\} + \frac{1-c}{v+t+1} \\ &= \sum_{k=v+1}^{v+t-1} \frac{c}{m}\mathbb{E}\pi_k^m(v+t)\mathbb{E}\mathcal{M}\{k \to v\} \\ &\quad + \frac{c}{m}\mathbb{E}\mathcal{M}\{v+t \to v\}\frac{1-c}{v+t+1} + \frac{1-c}{v+t+1}.\end{aligned} \tag{4.7}$$

Since each outgoing link from node k is created independently, we have

$$\mathbb{E}\mathcal{M}\{k \to v\} = m\mathbb{P}[\text{one link from } k \text{ to } v].$$

Due to the preferential attachment rule (see (2.1)), the probability $\mathbb{P}$[one link from k to v] does not depend on m if the expected weight of node v is proportional to m. Let us show this:

$$\mathbb{E}(d_k(n) + m | d_k(n-1)) = d_k(n-1) + m + m\frac{m + d_k(n-1)}{2m(n-1)+m}, \tag{4.8}$$

and taking the average over all possible network realizations, we get

$$\mathbb{E}(d_k(n)+m) = \mathbb{E}(d_k(n-1)+m) + \frac{\mathbb{E}(d_k(n-1)+m)}{2n-1}. \tag{4.9}$$

Knowing that $\mathbb{E}(d_k(k)+m) = m$, we conclude, even without calculating the final expression for $\mathbb{E}d_k(n)$, that it is proportional to m.

Since $\mathbb{P}$[one link from k to v] does not depend on m and $\mathbb{E}\pi_k^m(v+t)$ for $k = v+1, ..., v+t-1$ also does not depend on m by the induction hypothesis, the induction step is proven. $\square$

Theorem 4.1 allows us to concentrate on the case $m = 1$, when each realization of the growing network is a tree, in our study of the PageRank of the growing network model G_n^m.

Let us clarify the claim that the case $m = 1$ is much simpler than the case $m > 1$, thus outlining the steps of the ensuing analysis presented in the next sections. It follows from Corollary 3.2 that the PageRank of a given node v depends on the *number of nodes* in the subtree $T_v(n)$ and on the *distances* from these nodes to node v. Both these values can be described using the Markov-type random processes:

- The size of the tree $T_v(n)$ is a random variable, and it is easy to see that, in a growing network model with a preferential attachment mechanism, the evolution of the size of $T_v(n)$ is a Markov chain: at every step n the size $Y_v(n)$ of the tree $T_v(n)$ depends only on the size of the tree at the previous step $n-1$. Inside the tree $T_v(n)$ all nodes (except node v) are only connected to nodes from $T_v(n)$; therefore, the overall attractiveness of the tree $T_v(n)$ can be calculated directly, and it is equal to $2Y_v(n)+1$. The term "+1" is explained by the fact that we consider the node v to be inside the tree $T_v(n)$, but its "participation" in the attractiveness of $T_v(n)$ is just its out-degree 1. Further details on the evolution of the subtree $T_v(n)$ are given in Section 5.

- Let us consider the tree $T_v(\cdot)$ at the moment when it has k nodes. By the above arguments about the tree formation, we can limit our consideration only to the nodes that belong to the tree and ignore the rest of the network. In particular, node v becomes the initial node, and the moments of attachment of new nodes to the tree can be considered as the local time of $T_v(\cdot)$. When a new node is connected to some already existing node in the tree $T_v(\cdot)$, its distance to the root (or its height) depends only on the height of that node. Therefore, in the model with the preferential attachment mechanism, the probability of a new node being at some height h

depends on the number of existing nodes with the height $h-1$ and their popularity (the number of the nodes at height h). Actually, we can express this probability as the number of nodes of heights $h-1$ and h divided by the number of the nodes k. It does not depend on other details, for example, how the nodes are exactly connected inside the tree. Using this fact, we calculate the moment-generating function of the nodes' heights in Section 6.

In (3.5) we have a sum of a random number of random variables. We calculate this sum in Section 7 by combining the expressions for the distribution of subtree sizes and the moment-generating function of the nodes' heights.

5. Distribution of Subtree Sizes

We start with a lemma that gives an explicit expression for the distribution of the subtree size.

Lemma 5.1. *The probability that at time step n the subtree rooted in node v has k nodes is given by*

$$\mathbb{P}[Y_v(n)=k]=\frac{\Gamma(n-v+1)\Gamma(k+1/2)\Gamma(n-k)\Gamma(v+1/2)}{\Gamma(n-v-k+1)\Gamma(k+1)\Gamma(1/2)\Gamma(v)\Gamma(n+1/2)}. \tag{5.1}$$

Proof. We show that the evolution of the subtree size can be described by the Pólya-Eggenberger urn model (see Section 9).

There are balls of two colors, black and white, in one urn. Initially the urn contains $b=1$ black balls and $w=2v$ white ones. At every step one ball is drawn at random from the urn, then it is returned together with $s=2$ balls of the same color.

The balls correspond to the in- and out-degrees of the nodes. The number of balls is the sum of the degrees. The black balls correspond to the nodes from the subtree $T_v(n)$. The white balls, therefore, correspond to the nodes outside the subtree. Every existing edge (k,l) in G_n^1 corresponds to two balls in the urn model. Namely, one ball corresponds to the out-degree of node k, and the other ball corresponds to the in-degree of node l. Therefore, the Pólya-Eggenberger distribution can be used to estimate the number of black (or white) balls in the urn at time step n.

The choice of a black ball from the urn corresponds to the event that a new $(n+1)$th node connects itself to the subtree of v. Otherwise, the new node

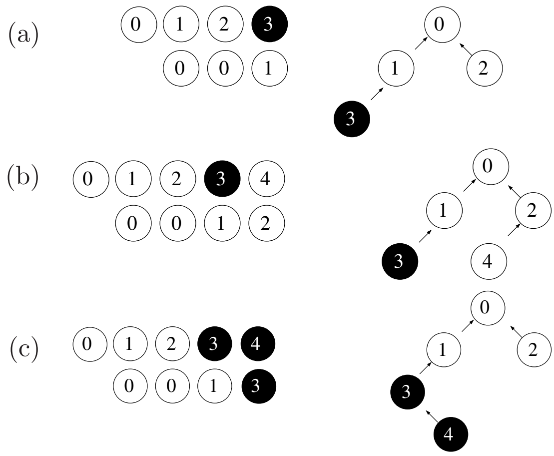

Figure 1. Illustration of the urn model: (a) Growing network after three steps. We choose to follow the subtree T_3. (b) Node 4 is not linked to T_3. (c) Node 4 is linked to T_3.

connects itself to some node outside of the subtree $T_v(\cdot)$, and, therefore, neither this node nor its subtree nodes will ever connect themselves to v with a path lying in the subtree of v.

We specify the expression for the Pólya-Eggenberger distribution (see Section 9) for our problem. The probability that after $n-v$ steps a black ball is drawn from the urn k times is

$$\begin{aligned}\mathbb{P}[Y_v(n)=k] &= \binom{n-v}{k}1(1+2)\dots(2k-1)\\ &\quad\times\frac{2v(2v+2)\dots(2(n-k-1))}{(1+2v)(1+2v+2)\dots(2n-1))},\end{aligned} \tag{5.2}$$

or, equivalently, in its Gamma function form it gives the expression (5.1). □

Let us illustrate the application of the urn model to the growing network formation by a simple example (see Figure 1). The upper row of the balls corresponds to the out-degrees of the nodes marked with their own numbers, and the second row corresponds to the in-degrees of the nodes. At time step 3 we have an urn with seven balls: six white and one black. Node 0 has in-degree $d_0 = 2$; therefore, there are three balls bearing the mark 0. If we draw from the urn a white ball, as in Figure 1(b), no matter which number it has (here it is 2), we fall out of T_3. Therefore, two white balls are added. In contrast, if we choose a black ball, then the new node falls inside the tree T_3, and, therefore, we add two black balls. Now it is easy to see that if we erase the number marks from the balls but leaving the ball colors, then we will not change anything in the formation of the number of balls of each color. Thus, the evolution of $Y_3(n)$ does not depend on the topology of T_3, but it depends only on the number of nodes inside and outside the subtree T_3.

6. Moment Generating Function of the Nodes' Heights in Subtrees

In this section we obtain the moment generating function of the nodes' heights inside the subtree $T_v(n)$. Namely, we have the following result.

Lemma 6.1. *If $X(v,s)$ is the height of the sth node inside the subtree $T_v(n)$, then*

$$\mathbb{E}[c^{X(v,s)}|Y_v(n) \geq s] = \frac{\Gamma(s+\frac{c}{2})\sqrt{\pi}}{\Gamma(s+\frac{1}{2})\Gamma(\frac{c}{2})}, \tag{6.1}$$

where $Y_v(n)$ is the number of nodes inside the subtree $T_v(n)$ not counting the root node v.

Proof. Let us fix v and define $X_s = X(v,s)$ for the sake of short notation. The evolution of X_s can be described without reference to any particular network realization or any particular tree structure. If the sth node of subtree $T_v(n)$ has height k in $T_v(n)$, it means that it is connected to a node with the height $k-1$ in $T_v(n)$. The conditional probability of such an event is the number of nodes located at height $k-1$ plus the number of nodes located at height k, normalized by $2s-1$; that is,

$$\begin{aligned}\mathbb{P}[X_s = k|X_{s-1},\ldots,X_0; Y_v(n)\geq s] = \\ \frac{\sum_{i=0}^{s-1}\mathbb{1}(X_i = k) + \sum_{i=0}^{s-1}\mathbb{1}(X_i = k-1)}{2s-1},\end{aligned} \tag{6.2}$$

where $\mathbb{1}(\cdot)$ is an indicator function.

Using (6.2), we can calculate the conditional moment-generating function of the nodes' heights as follows:

$$\mathbb{E}[c^{X_s}|X_{s-1},\ldots,X_0; Y_v(n)\geq s] = \sum_{k=0}^{s} c^k \mathbb{P}[X_s = k|X_{s-1},\ldots,X_0; Y_v(n)\geq s]$$

$$
\begin{aligned}
&= \frac{\sum_{k=0}^{s} c^k \mathbb{I}(X_{s-1}=k) + \sum_{k=1}^{s} c^k \mathbb{I}(X_{s-1}=k-1)}{2s-1} \\
&\quad + \frac{2s-3}{2s-1}\mathbb{E}[c^{X_{s-1}}|X_{s-2}\ldots,X_0;Y_v(n)\geq s] \\
&= \frac{\sum_{k=0}^{s} c^k \mathbb{I}(X_{s-1}=k) + c\sum_{k=0}^{s-1} c^k \mathbb{I}(X_{s-1}=k)}{2s-1} \\
&\quad + \frac{2s-3}{2s-1}\mathbb{E}[c^{X_{s-1}}|X_{s-2}\ldots,X_0;Y_v(n)\geq s],
\end{aligned}
$$

where $\mathbb{I}(X_i = s) = 0$ for all $i < s$. Next, applying the double expectation value rule with respect to $\mathbb{E}[\cdot|Y_v(n) \geq s]$, we obtain the following recurrent equation:

$$
\mathbb{E}[c^{X_s}|Y_v(n)\geq s] = \left(1 - \frac{1-c}{2s-1}\right)\mathbb{E}[c^{X_{s-1}}|Y_v(n)\geq s]. \tag{6.3}
$$

The above recurrent equation gives

$$
\mathbb{E}[c^{X_s}|Y_v(n)\geq s] = \prod_{k=1}^{s}\left[1-\frac{1-c}{2s-1}\right] = \frac{\Gamma(s+\frac{c}{2})\sqrt{\pi}}{\Gamma(s+\frac{1}{2})\Gamma(\frac{c}{2})}, \tag{6.4}
$$

which completes the proof. □

Using the derivations in the proof of Lemma 6.1, we can also estimate the average subtree height. Namely, we have

$$
\mathbb{E}[X_s|Y_v(n)\geq s] = \mathbb{E}[X_{s-1}|Y_v(n)\geq s] + \frac{1}{2s-1}
$$

and, consequently,

$$
\mathbb{E}[X_s|Y_v(n)\geq s] = \sum_{k=1}^{s}\frac{1}{2s-1}. \tag{6.5}
$$

Equation (6.5) can be interpreted as follows.

Lemma 6.2. *The average height of the subtree $T_v(n)$ after s steps in local time is of order* $\log(s)$.

This result is in line with the results of [Bollobás and Riordan 04].

Now, we can already calculate the expected PageRank value of the initial node. Using (3.6) and the fact that $Y_0(n) = n$, we obtain

$$
\mathbb{E}\pi_0(n) = \frac{1}{1+n}\left(\frac{1}{c+1} + \frac{2\sqrt{\pi}\Gamma(n+\frac{c}{2}+1)}{(c+1)\Gamma(\frac{c}{2})\Gamma(n+1/2)}\right). \tag{6.6}
$$

7. Final Result Statement and Asymptotics

The expected value of PageRank is provided by the following theorem.

Theorem 7.1. *The expected value of PageRank $\pi_v(n)$ of node v at time step n in the present growing network model G_n^m is given by*

$$\mathbb{E}\pi_v(n) = \frac{1-c}{1+n}\left(\frac{1}{1+c} + \frac{c\Gamma(v+\frac{1}{2})\Gamma(n+\frac{c}{2}+1)}{(1+c)\Gamma(v+\frac{c}{2}+1)\Gamma(n+\frac{1}{2})}\right), \tag{7.1}$$

for $v > 0$, and

$$\mathbb{E}\pi_0(n) = \frac{1}{1+n}\left(\frac{1}{c+1} + \frac{2\sqrt{\pi}\Gamma(n+\frac{c}{2}+1)}{(c+1)\Gamma(\frac{c}{2})\Gamma(n+1/2)}\right), \tag{7.2}$$

for the particular case of $v = 0$.

Proof. First, we reduce the case $m > 1$ to the case $m = 1$ by Theorem 4.1.

Then, we calculate $\mathbb{E}\pi_v(n)$ using Equation (3.5). We note that in (3.5) we sum a random number of random variables. The calculation of this sum is similar in spirit to the proof of Wald's identity or Kolmogorov-Prokhorov identity (see, e.g., [Ross 92] or [Borovkov 98, Chapter 4.4]). Namely, we have

$$\begin{aligned}
\mathbb{E}\pi_v(n) &= \frac{1-c}{n+1}\left(1 + \mathbb{E}\sum_{s=1}^{Y_v(n)} c^{X(v,s)}\right) \\
&= \frac{1-c}{n+1}\left(1 + \mathbb{E}\sum_{s=1}^{n-v} c^{X(v,s)} 1\{Y_v(n) \geq s\}\right) \\
&= \frac{1-c}{n+1}\left(1 + \sum_{s=1}^{n-v} \mathbb{E}[c^{X(v,s)} 1\{Y_v(n) \geq s\}]\right) \\
&= \frac{1-c}{n+1}\left(1 + \sum_{s=1}^{n-v} \mathbb{E}[c^{X(v,s)} | Y_v(n) \geq s] P[Y_v(n) \geq s]\right) \\
&= \frac{1-c}{n+1}\left(1 + \sum_{s=1}^{n-v} \mathbb{E}[c^{X(v,s)} | Y_v(n) \geq s] \sum_{i=s}^{n-v} P[Y_v(n) = i]\right).
\end{aligned}$$

Changing the order of summation, we get

$$\mathbb{E}\pi_v(n) = \frac{1-c}{n+1}\left(1 + \sum_{i=1}^{n-v} P[Y_v(n) = i] \sum_{s=1}^{i} \mathbb{E}[c^{X(v,s)} | Y_v(n) \geq s]\right).$$

Next, we substitute into the above expression Equations (5.1) and (6.1) for $\mathbb{P}[Y_v(n) = i]$ and $\mathbb{E}[c^{X(v,s)}|Y_v(n) \geq s]$, respectively, to obtain

$$\begin{aligned}\mathbb{E}\pi_v(n) &= \frac{1-c}{n+1}\left(1 + \sum_{i=1}^{n-v} \frac{\Gamma(n-v+1)}{\Gamma(n-v-i+1)\Gamma(i+1)}\right. \\ &\quad \left. \times \frac{\Gamma(i+1/2)\Gamma(n-i)\Gamma(v+1/2)}{\Gamma(v)\Gamma(n+1/2)} \sum_{k=1}^{i} \frac{\Gamma(k+c/2)}{\Gamma(k+1/2)\Gamma(c/2)}\right). \end{aligned} \tag{7.3}$$

Simplifying the internal sum in the above equation, we obtain the following expression

$$\begin{aligned}\mathbb{E}\pi_v(n) &= \frac{1-c}{n+1}\left(1 + \frac{\Gamma(n-v+1)\Gamma(v+1/2)}{\Gamma(v)\Gamma(n+1/2)}\right. \\ &\quad \times \sum_{i=1}^{n-v} \frac{\Gamma(n-i)\Gamma(i+1/2)}{\Gamma(n-v-i+1)\Gamma(i+1)} \\ &\quad \left. \times \left(\frac{2\sqrt{\pi}\Gamma(i+1+c/2)}{(1+c)\Gamma(i+1/2)\Gamma(c/2)} - \frac{c}{c+1}\right)\right) \\ &= \frac{1-c}{n+1}\left(1 + \frac{2\sqrt{\pi}\Gamma(n-v+1)\Gamma(v+1/2)}{(1+c)\Gamma(c/2)\Gamma(v)\Gamma(n+1/2)}\right. \\ &\quad \times \sum_{i=1}^{n-v} \frac{\Gamma(n-i)\Gamma(i+1+c/2)}{\Gamma(n-v-i+1)\Gamma(i+1)} \\ &\quad - \frac{c\Gamma(n-v+1)\Gamma(v+1/2)}{(1+c)\Gamma(v)\Gamma(n+1/2)} \\ &\quad \left. \times \sum_{i=1}^{n-v} \frac{\Gamma(n-i)\Gamma(i+1/2)}{\Gamma(n-v-i+1)\Gamma(i+1)}\right). \end{aligned} \tag{7.4}$$

By using Zeilberger's algorithm and his package EKHAD for Maple, we prove (see Lemma 10.1) the following hypergeometric identity:

$$\sum_{i=1}^{n-v} \frac{\Gamma(n-i)\Gamma(i+1+c/2)}{\Gamma(n-v-i+1)\Gamma(i+1)} = \frac{\Gamma(v)\Gamma(n+c/2+1)\Gamma(1+c/2)}{\Gamma(v+c/2+1)\Gamma(n-v+1)} - \frac{\Gamma(n)\Gamma(1+c/2)}{\Gamma(n-v+1)}. \tag{7.5}$$

We can apply this identity to both sums in (7.4), since we can think of the second sum as a particular case of the first one, with $c = -1/2$. After some simplifications we obtain the final result (7.1). $\square$

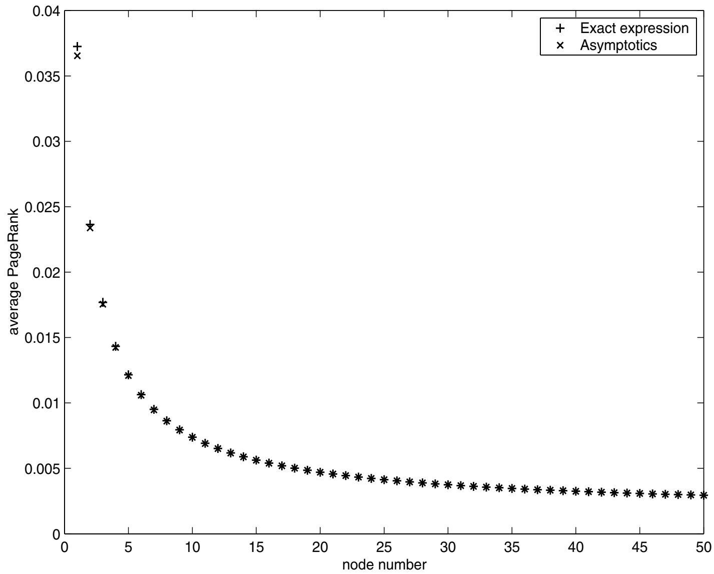

Figure 2. Comparison between the asymptotics (7.6) and the exact expression (7.1).

The expression (7.1) is already simple enough. However, it can be made even more transparent by using the following asymptotics:

$$\Gamma(x+a)/\Gamma(x) \approx x^a,$$

when $0 < a < 1$ and $x \to +\infty$. Thus, we have

$$\mathbb{E}\pi_v(n) \approx \frac{1-c}{1+n}\left(\frac{1}{1+c} + \frac{c}{1+c}(v+\frac{1}{2})^{-\frac{1+c}{2}}(n+\frac{1}{2})^{\frac{1+c}{2}}\right) \tag{7.6}$$

or, neglecting the first term,

$$\mathbb{E}\pi_v(n) \approx \frac{1-c}{1+c} c v^{-\frac{1+c}{2}} n^{-\frac{1-c}{2}}. \tag{7.7}$$

In particular, for the zero node we have

$$\mathbb{E}\pi_0(n) \approx \frac{2\sqrt{\pi}}{(1+c)\Gamma(\frac{c}{2})} n^{-\frac{1-c}{2}}. \tag{7.8}$$

As one can see from Figure 2, the asymptotics (7.6) indeed closely follows the exact expression (7.1).

8. Discussion and Comparison with Related Work

First, let us compare our results with the results of Pandurangan et al. [Pandurangan et al. 02]. In the present work we have obtained an exact analytical expression and asymptotics for the expected value of the PageRank as a function of the age of the growing network and the age of a particular node. In Pandurangan et al. [Pandurangan et al. 02] the authors have used the "mean-field" approach [Barabási and Albert 99, Barabási et al. 99, Barabási et al. 00] to obtain an approximation for the PageRank distribution. Let us use our results on the expected value of PageRank for the "mean-field" calculations of [Pandurangan et al. 02]. Specifically, suppose that n is fixed, PageRank depends continuously on v, and the node age is uniformly distributed. Then, using our asymptotic expression (7.6), we obtain

$$\begin{aligned} P(x) &= \mathbb{P}[\pi_v < x] \\ &\approx \mathbb{P}\left[v > \left(\left(\frac{1+n}{1-c}x - \frac{1}{1+c} \right) \frac{1+c}{c} (n+\frac{1}{2})^{-\frac{1+c}{2}} \right)^{-\frac{2}{1+c}} - 1/2 \right] \\ &= 1 - \left(\left(\left(\frac{1+n}{1-c}x - \frac{1}{1+c} \right) \frac{1+c}{c} (n+\frac{1}{2})^{-\frac{1+c}{2}} \right)^{-\frac{2}{1+c}} - 1/2 \right) /n \\ &= 1 + \frac{1}{2n} - \left(1+\frac{1}{2n}\right) c^{\frac{2}{1+c}} \left(\frac{1+c}{1-c}(n+1)x - 1 \right)^{-\frac{2}{1+c}}. \end{aligned} \tag{8.1}$$

In particular, we note that $P(\frac{1-c}{n+1}) = 0$, since $x = \frac{1-c}{n+1}$ is the minimal value of PageRank. Taking the derivative of (8.1), we obtain the density distribution function of the PageRank value:

$$p(x) = \frac{2}{1-c}(n+1)\left(1+\frac{1}{2n}\right) c^{\frac{2}{1+c}} \left(\frac{1+c}{1-c}(n+1)x - 1 \right)^{-\frac{3+c}{1+c}}. \tag{8.2}$$

For large values of n and for values of x that are not too small and not too close to one, the expression (8.2) is close to the power law

$$p(x) \asymp \frac{1}{x^{\frac{3+c}{1+c}}}.$$

For instance, for the dumping factor $c = 0.85$, we can conclude that the density distribution of PageRank for the nodes whose numbers are not too small and not too close to n can be approximated by a power law with the exponent 2.08. Note that the "mean-field" approximation of [Pandurangan et al. 02] gives the exponent 2 and the experiments with the real web data in [Pandurangan et al. 02] give the exponent 2.1.

To test the mean-field estimation (8.1), we ran simulations of our growing network model. The network grew up to $n = 1000$ for 100,000 simulation runs

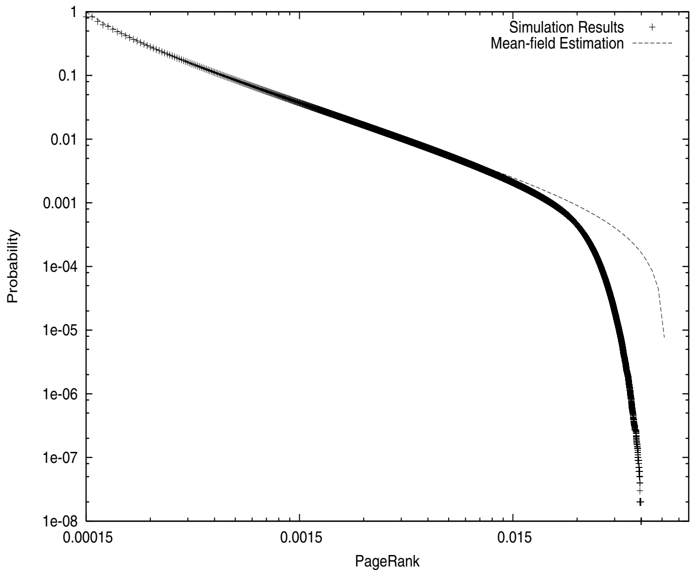

Figure 3. Cumulative complimentary distribution function: Simulation results compared to the mean-field estimation.

with $m = 10$. In Figure 3 the mean-field estimation (8.1) is compared with the cumulative complimentary distribution function $\mathbb{P}[\pi_v > x] = 1 - \mathbb{P}[\pi_v < x]$ obtained from the simulations. As pointed out in [Newman 05], when dealing with power laws, it is preferable to work with the cumulative complimentary distribution function rather than with the density distribution function or the histogram. The cumulative distribution function of a power law $x^{-\alpha}$ also follows the power law, but with exponent $x^{-\alpha+1}$. When calculating the PageRank, we used $c = 0.85$. We note that the plot is indeed close to a straight line for the middle segment of the PageRank range. In [Pandurangan et al. 02] the authors also noticed that PageRank follows a power law except for those pages with very small PageRank. This phenomenon can easily be explained with the help of (8.2). The term $\frac{1+c}{1-c}(n+1)x$ becomes comparable with 1 in (8.2) for values of x too close to the minimal PageRank $\frac{1-c}{n+1}$, and the distribution density function cannot be, in this case, approximated by $O(x^{-\alpha})$. The mismatch for large values of PageRank can be explained as follows: the “mean-field” approach cannot be applied to the nodes with large PageRank because there are simply not enough such nodes to use the “averaging” argument.

As it can be observed from (7.2) and (7.1), the zero node is special. As n grows, its PageRank converges to 0, but, nevertheless, its value is bigger than the PageRank of other nodes. We can normalize the expected value of PageRank of all nodes by $\mathbb{E}\pi_0$. In fact, we have

$$\tilde{\pi}_v = \lim_{n\to\infty} \frac{\mathbb{E}\pi_v(n)}{\mathbb{E}\pi_0(n)} = \frac{(1-c)c\Gamma(\frac{c}{2})}{2\sqrt{\pi}} \frac{\Gamma(v+\frac{1}{2})}{\Gamma(v+\frac{c}{2}+1)}. \tag{8.3}$$

Let us call $\tilde{\pi}$ the relative PageRank. We would like to emphasize that the *relative PageRank* does not depend on time. The relative PageRank closely follows a power law except for some initial nodes.

Recall that Google divides the whole range of PageRank into ten intervals using logarithmic scale. Curiously enough, if PageRank exactly followed a power law, then this division would be independent of c and the exponent of the power law but would depend only on n, the age of the network. Specifically, in such a case, the following formula holds for the boundaries of the ranking intervals:

$$v_k^* = (n)^{\frac{k}{10}}, k = 1, ..., 10.$$

The above observation justifies further the scale-free term for the growing network model.

The authors of [Boldi et al. 05] investigated the effect of the damping factor c on PageRank. In their numerical example they have noticed that the PageRank for some nodes attains a maximal value for some value of c. Let us investigate the dependence of PageRank on c in our growing network model. In our case the value of c that maximizes the PageRank expression (7.6) depends on the ratio v/n. In Figure 4 we plot the optimal value of c as a function of the ratio v/n. As an example, in Figure 5 we plot the expected values of PageRank for node 1 and node 2 when $n = 10000$. If the World Wide Web has eight billion pages, then the present model suggests that the pages which benefit the most from the value of $c = 0.85$ are around the node $v = 46212$. Thus, it looks like the damping factor $c = 0.85$ benefits only a small fraction of old pages. Thus, to give a better ranking to less established webpages and to distribute PageRank more fairly, it is necessary to decrease the value of c. Of course, this will also have a positive effect on the convergence of the numerical methods for PageRank computation. The question of by how much the damping factor can be reduced merits a careful special investigation.

Finally, we would like to note that the choice of the initial weight for the zero node was a crucial factor for the derivation of simple explicit expressions. This choice affects only the preferential attachment process. In fact, all the methods in the present work can be applied to the growing network models with a different preferential attachment process. The expression (5.1) would change slightly, but there is no guarantee that one could find a simple closed form of the final expression (7.1).

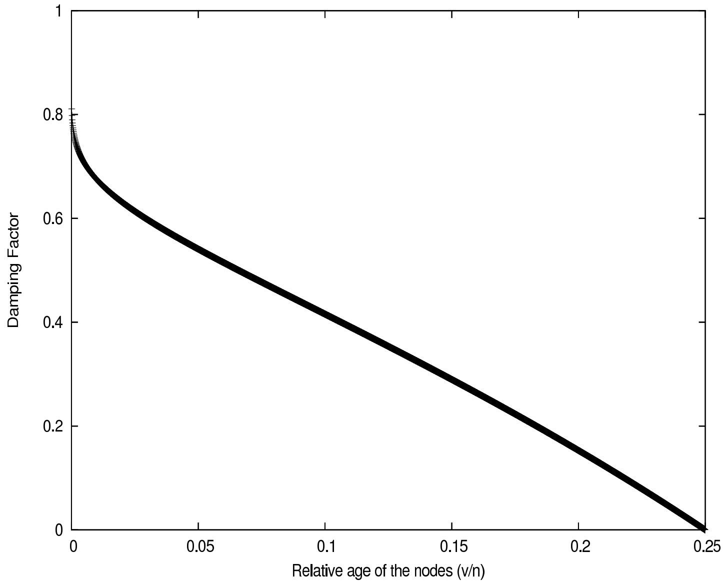

Figure 4. The optimal value of c as a function of v/n.

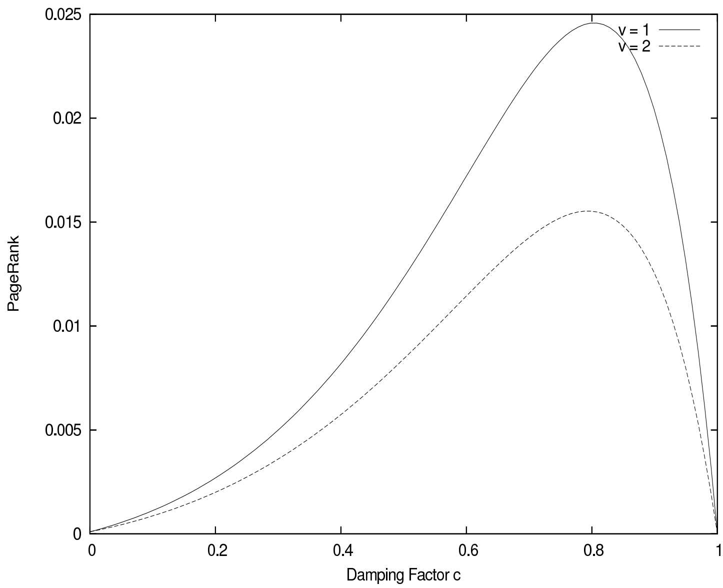

Figure 5. The expected value of PageRank as a function of c for $v = 1, 2$ and $n = 10000$.

Curiously enough, we have tested several scale-free growing network models, and in all our experiments the results were very close. Thus, the analysis of PageRank for different growing network models and further generalization of the results are interesting perspective research directions.

9. Appendix: Polya-Eggenberger Urn Model

We follow the book [Johnson and Kotz 77] in our description of the urn models. The Pólya-Eggenberger urn model starts with one urn containing $b+w$ balls of two colors: black and white. Let b be the number of black balls, and let w be the number of white balls. At every time step one ball is drawn at random from the urn, and then it is returned together with s balls of the same color. The Pólya-Eggenberger distribution is used to estimate the number of black (or white) balls at time step n. The probability to draw a black ball k times from the urn after n steps can then be expressed as

$$\begin{aligned}\mathcal{P}_{n,k}(w,b,s) &= \binom{n}{k} b(b+s)\ldots(b+(k-1)s)\\ &\quad\times\frac{w(w+s)\ldots(w+(n-k-1)s)}{(b+w)(b+w+s)\ldots(b+w+(n-1)s)},\end{aligned} \tag{9.1}$$

for $k=0,1,\ldots,n$. Using the gamma function, the above formula can be rewritten as follows:

$$\begin{aligned}\mathcal{P}_{n,k}(w,b,s) &= \binom{n}{k}\frac{\Gamma(\frac{b+w}{s})\Gamma(\frac{b}{s}+k)\Gamma(\frac{w}{s}+n-k)}{\Gamma(b/s)\Gamma(w/s)\Gamma(\frac{b+w}{s}+n)}\\ &= \binom{n}{k}\frac{B(\frac{b}{s}+k,\frac{w}{s}+n-k)}{B(\frac{b}{s},\frac{w}{s})}.\end{aligned} \tag{9.2}$$

It is worthwhile to note here that the problem of the nodes' heights in subtrees, which we study in Section 6, can also be described in terms of the urn model with a node height value as a mark (or color).

10. Appendix: Zeilberger's Algorithm

We follow the book [Petkovsek et al. 96] in our description of Zeilberger's Algorithm. Let us consider a sum

$$f(n)=\sum_k F(n,k). \tag{10.1}$$

The goal of Zeilberger's Algorithm is to find function $G(\cdot,\cdot)$ and coefficients $a_j(n)$ such that

$$\sum_{j=0}^{J} a_j(n)F(n+j,k) = G(n,k+1) - G(n,k). \tag{10.2}$$

This method is also called the method of *creative telescoping*. When such a representation is obtained, we can sum equation (10.2) by k, and, if we are lucky with the values F and G, the right part of the sum might collapse to 0, leaving us with an equation of the type

$$\sum_{j=0}^{J} a_j(n)f(n+j) = 0. \tag{10.3}$$

For example, if $J = 1$, we find the recurrence $a_0(n)f(n) + a_1(n)f(n+1) = 0$, and then $f(n)$ is easy to find.

D. Zeilberger wrote the package EKHAD [Zeilberger 01] for Maple, which implements his algorithm and finds $a_0, \ldots, a_J$ and $R(\cdot,\cdot)$ such that

$$G(n,k) = R(n,k)F(n,k). \tag{10.4}$$

Fortunately, Zeilberger's Algorithm gives a satisfying result for the sum in (7.3). Let us prove the following lemma.

Lemma 10.1.

$$\sum_{i=1}^{n-v} \frac{\Gamma(n-i)\Gamma(i+1+c/2)}{\Gamma(n-v-i+1)\Gamma(i+1)} = \frac{\Gamma(v)\Gamma(n+c/2+1)\Gamma(1+c/2)}{\Gamma(v+c/2+1)\Gamma(n-v+1)} - \frac{\Gamma(n)\Gamma(1+c/2)}{\Gamma(n-v+1)}. \tag{10.5}$$

Proof. Consider the internal sum (7.3). We introduce the following notation:

$$F(n,i) = \frac{\Gamma(n+v-i)\Gamma(i+1+c/2)\Gamma(v+c/2+1)\Gamma(n+v+1)}{\Gamma(n-i+1)\Gamma(i+1)\Gamma(v)\Gamma(n+v+c/2+1)\Gamma(1+c/2)}. \tag{10.6}$$

This is the summand from (10.5) divided by the result that we want to prove (it was guessed from the values of $\mathbb{E}\pi_v$ for $v = 1,2,3$); also, we change the variable from n to $n-v$, and we add the zeroeth summand. Now, we want to prove that

$$f(n) = \sum_{i=0}^{n} F(n,i) \equiv 1. \tag{10.7}$$

The function *zeil* from the EKHAD package [Zeilberger 01] finds the following identities: $a_0 = -1$, $a_1 = 1$, and

$$R(n,i) = -\frac{(n+v-i)i}{(n+v+\frac{c}{2}+1)(n-i+1)}. \tag{10.8}$$

Therefore, in our case (10.2) takes the following form:

$$F(n,i) - F(n+1,i) = R(n,i+1)F(n,i+1) - R(n,i)F(n,i). \tag{10.9}$$

Then, we sum Equation (10.9) for the values $i = 0, \dots, n-1$, and we find that

$$R(n,n)F(n,n) - R(n,0)F(n,0) - F(n,n) + F(n+1,n) + F(n+1,n+1) = 0$$

for the values F and R in (10.6) and (10.8). Therefore, $f(n) = f(n+1)$. As $f(0) = 1$, it completes the prove.

Note that it is indeed a proof, because Maple and EKHAD provide the identities (i.e., the values of a_i, $i = 0, 1$ and (10.8)), which can be easily checked. □

Acknowledgments. We would like to thank Nelly Litvak, Fan Chung Graham, Vladimir Dobrynin, Bruno Salvy, and Son Kim Pham for discussions and helpful remarks during the manuscript preparation. The present work is partially supported by EGIDE ECO-NET grant no. 10191XC and the European Research Grant BIONETS.

References

[Barabási and Albert 99] A.-L. Barabási and R. Albert. "Emergence of Scaling in Random Networks." *Science* 286 (1999), 509–512.

[Barabási et al. 99] A.-L. Barabási, R. Albert, and H. Jeong. "Mean-Field Theory for Scale-Free Random Networks." *Physica A* 272 (1999), 173–187.

[Barabási et al. 00] A.-L. Barabási, R. Albert, and H. Jeong. "Scale-Free Characteristics of Random Networks: The Topology of the World Wide Web." *Physica A* 281 (2000), 69–77.

[Bianchini et al. 05] M. Bianchini, M. Gori, and F. Scarselli. "Inside PageRank." *ACM Trans. Internet Technology* 5:1 (2005), 92–128.

[Boldi et al. 05] P. Boldi, M. Santini, and S. Vigna. "PageRank as a Function of the Damping Factor." In *Proceedings of the 14th International Conference on World Wide Web*, pp. 557–566. New York: ACM Press, 2005.

[Bollobás and Riordan 02] B. Bollobás and O. Riordan. "Mathematical Results in Scale-Free Random Graphs." In *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster, pp. 1–34. Weinheim: Wiley-VCH, 2002.

[Bollobás and Riordan 04] B. Bollobás and O. Riordan. "The Diameter of a Scale-Free Random Graph." *Combinatorica* 24:1 (2004), 5–34.

[Bollobás et al. 01] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. "The Degree Sequence of a Scale-Free Random Graph Process." *Random Structures and Algorithms* 18:3 (2001), 279–290.

[Borovkov 98] A. A. Borovkov. *Probability Theory.* New York: Gordon and Breach Science Publishers, 1998.

[Brin and Page 98] S. Brin and L. Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks* 30:1–7 (1998), 107–117.

[Dorogovtsev and Mendes 02] S. N. Dorogovtsev and J. F. F. Mendes. "Evolution of Networks." *Advances in Physics* 51 (2002), 1079–1187.

[Dorogovtsev et al. 00] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. "Structure of Growing Networks with Preferential Linking." *Phys. Rev. Lett.* 85 (2000), 4633–4636.

[Johnson and Kotz 77] N. L. Johnson and S. Kotz. *Urn Models and Their Application.* New York: John Wiley and Sons, 1977.

[Kleinberg 99] J. Kleinberg. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM* 46:5 (1999), 604–632.

[Langville and Meyer 05] A. N. Langville and C. D. Meyer. "Deeper Inside PageRank." *Internet Mathematics* 1:3 (2005), 335–380.

[Langville and Meyer 06] A. N. Langville and C. D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton, NJ: Princeton University Press, 2006.

[Lempel and Moran 00] R. Lempel and S. Moran. "The Stochastic Approach for Link-Structure Analysis (SALSA) and the TKC Effect." In *Proceedings of the Ninth International World Wide Web Conference.* Available at http://www9.org/w9cdrom/175/175.html, 2000.

[Moler 04] C. B. Moler. *Numerical Computing with MATLAB.* Philadelphia: SIAM, 2004.

[Newman 03] M. E. J. Newman. "The Structure and Function of Complex Networks." *SIAM Reviews* 45:2 (2003), 167–256.

[Newman 05] M. E. J. Newman. "Power Laws, Pareto Distributions and Zipf's Law." *Contemporary Physics* 46 (2005), 323.

[Pandurangan et al. 02] G. Pandurangan, P. Raghavan, and E. Upfal. "Using PageRank to Characterize Web Structure." In *Computing and Combinatorics: 8th Annual International Conference, COCOON 2002, Singapore, August 15–17, 2002, Proceedings*, pp. 330–339, Lecture Notes in Computer Science 2387. Berlin: Springer, 2002.

[Petkovsek et al. 96] M. Petkovsek, H. S. Wilf, and D. Zeilberger. *A=B.* Wellesley, MA: A K Peters, 1996.

[Ross 92] S. M. Ross. *Applied Probability Models With Optimization Applications.* New York: Courier Dover Publications, 1992.

[Zeilberger 01] D. Zeilberger. "Doron Zeilberger's Maple Packages and Programs." Available at http://www.math.temple.edu/~zeilberg/programs.html, 2001.

Konstantin Avrachenkov, INRIA Sophia Antipolis, 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France (K.Avrachenkov@sophia.inria.fr)

Dmitri Lebedev, Laboratoire d'Informatique, Ecole Polytechnique, 91128 Palaiseaux CEDEX, France (Dmitri.Lebedev@polytechnique.edu)

Received March 23, 2006; accepted November 4, 2006.