PageRank of Scale Free Growing Networks

Konstantin Avrachenkov, Dmitri Lebedev

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

PageRank of Scale Free Growing Networks

Konstantin Avrachenkov — Dmitri Lebedev

N° 5858

March 2006

Thème COM

ISSN 0249-6399 ISRN INRIA/RR--5858--FR+ENG

INRIA
SOPHIA ANTIPOLIS

PageRank of Scale Free Growing Networks

Konstantin Avrachenkov* , Dmitri Lebedev †

Thème COM — Systèmes communicants
Projet Maestro

Rapport de recherche n° 5858 — March 2006 — 24 pages

Abstract: PageRank is one of the principle criteria according to which Google ranks Web pages. PageRank can be interpreted as a frequency of Web page visits by a random surfer and thus it reflects the popularity of a Web page. 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 shows that PageRank follows closely a power law. The exponent of the theoretical power law matches very well the value found from measurements of the Web. Finally, we provide a mathematical insight for the choice of the damping factor in PageRank definition.

Key-words: PageRank, Web Graph, Growing scale free networks, Pólya-Eggenberger urn models, Power law, scale-free

The present work is partially supported by EGIDE ECO-NET grant no. 10191XC and the European Research Grant BIONETS.

* INRIA Sophia Antipolis, France, e-mail: K.Avrachenkov@sophia.inria.fr
† Ecole Polytechnique, France, e-mail: Dmitri.Lebedev@polytechnique.edu

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

PageRank dans les Modèles Scale Free de Réseaux Croissants

Résumé : PageRank est un des principaux critères de classement des pages Web par Google. PageRank peut être interpreté comme la fréquence de visites d'une page Web par un utilisateur aléatoire, on peut donc aussi l'appeler la popularité de cette page Web. Dans ce travail, nous donnons une expression analitique pour le PageRank moyen dans les modéles scale-free de réseaux croissants. Cette expression est obtenue comme une fonction de l'âge du modèle et de l'âge d'un nœud. En plus, on obtient les asymptotiques qui démontrent que la distribution approche une loi en puissance. L'exposant théorique de cette loi est trés proche des valeurs trouvées dans les mesures expérimentales du Web. L'expression ainsi trouvée fournit une base de raisonnement mathématique au choix du facteur d'abandon par Google.

Mots-clés : PageRank, World Wide Web, Graphes aléatoires, Modeèles d'urnes de Pólya-Eggenberger, loi en puissance, scale-free

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 non-trivial task. The original idea of Google presented in [10] is to list pages according to their PageRank which reflects popularity of a page. The PageRank is defined in the following way. 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 has $k > 0$ ougtoing 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 Web page with the uniform distribution. Thus, the PageRank is defined as a stationary distribution of a Markov chain whose state space is the set of all Web pages, and the transition matrix is

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

where E is a matrix whose all entries are 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

$$\pi\tilde{P} = \pi, \quad \pi\underline{1} = 1, \tag{2}$$

where $\underline{1}$ is a column vector of ones. The row vector π satisfying (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 [1] have 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. Then, Bollobás et al. [6] have refined their model and proved rigorously that in and out degree distributions satisfy power laws. Pandurangan et al. [20] applied the "mean-field" heuristics from [1, 2, 3] to show that the PageRank distribution in the scale free growing network model satisfies the power law with exponent 2. They have also proposed a model where 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 [20] 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 shows that PageRank follows closely a power law with exponent 2.08. 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 formulae for Pagerank for 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 Appendices.

2 Scale Free Network Models

Inspired by the power law in and out degree distributions of the World Wide Web, Barabási and Albert [1] have proposed growing network model with preferential attachment. In their model a new node is attached to some old nodes with probability proportional to the in degree of the old nodes. The authors of [1, 2, 3] have developed the "mean-field" heuristics, which allowed them to derive approximations to the power law degree distributions. Then, Bollobás et al. [6] have added some missing parts to the Barabási-Albert model and have shown rigorously that the degree distributions of the scale free growing network model indeed satisfy power laws. The model of [6] 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, Google when computes the PageRank disregards the hyperlinks within the same Web page. Taking into account the above 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 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 *in* degrees plus m. Namely, the probability that node n connects to node v, $v < n$, is given by

$$\mathbb{P}[n \to v | G_{n-1}^m] = \frac{d_v(n-1)+m}{\sum_{k=0}^{n-1}(d_k(n-1)+m)} = \frac{d_v(n-1)+m}{2m(n-1)+m}. \quad (3)$$

  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_n^1 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 [12] is the closest model to ours. An interested reader can find a detail overview of growing network models in the surveys [8, 11, 19].

3 PageRank of Growing Networks

Let us study the PageRank for growing networks with fixed outdegree 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 in a value of the common factor for all nodes $v \geq 1$ [15]. Since it turns out that this factor is much simpler in the case of the initial node with self loop, we choose the first option in the present work.

We denote by $\pi_v(n)$ the PageRank of node v after the n-th step of the growing network 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 $L_v(n)$ the set of all paths from nodes $v+1, ..., n$ to v and by $|l|$ the length of a path l. Then, the PageRank vector of a growing network realization can be calculated by an explicit formula given in the next theorem.

Theorem 1. *The PageRank of a growing network realization of node v, $v > 0$, at time step n is given by equation*

$$\pi_v(n) = \frac{1-c}{n+1}\left(1 + \sum_{l \in L_v(n)} \left(\frac{c}{m}\right)^{|l|}\right), \quad (4)$$

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

$$\pi_0(n) = \frac{1}{n+1}\left(1 + \sum_{l \in L_0(n)} \left(\frac{c}{m}\right)^{|l|}\right). \tag{5}$$

Proof. The PageRank vector of any network can be expressed by the formula [17, 4, 15]

$$\pi = \frac{1-c}{n+1}\underline{1}^T[I - cP]^{-1}, \tag{6}$$

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

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

Next we note that the (i, j) element of matrix $[I - cP]^{-1}$ corresponds to the sum of $(c/m)^{|l|}$ 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. In the case $v > 0$, there are no loops and hence we obtain formula (4). In the case $v = 0$, each path to the initial node ends with a self loop. Because of this self loop each term $(c/m)^{|l|}$ is multiplied by the series $1 + c + c^2 + ...$ The sum of the later series is equal to $1/(1-c)$, which cancels the factor $1 - c$ in (6) and results in the particular expression (5) 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 the formulae (4) and (5). 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, $n > v$. Also we denote by $X_n(v, w)$ the distance between v and w at step n (of course, we should have $w \in T_v(n)$). We shall also call $X_n(v, w)$ *height* of w in $T_v(n)$. Let us denote the number of nodes in $T_v(n)$ by $Y_v(n)$. Then, we have the following corollary from Theorem 1.

Corollary 1. *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:*

$$\pi_v(n) = \frac{1-c}{n+1}\left(1 + \sum_{\alpha \in T_v(n)} c^{X_n(v,\alpha)}\right), \tag{7}$$

or in its alternative local time form with respect to the subtree $T_v(n)$,

$$\pi_v(n) = \frac{1-c}{n+1}(1 + \sum_{k=1}^{Y_v(n)} c^{X_n(v,k)}), \tag{8}$$

and the PageRank of the initial node 0 is given by

$$\pi_0(n) = \frac{1}{n+1}\left(1 + \sum_{k=1}^{n} c^{X_n(0,k)}\right). \tag{9}$$

4 The case $m > 1$ can be reduced to the case $m = 1$

It follows from Corollary 1 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 the following Theorem 2 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 PageRank of node v at time step n for our growing network model G_n^m.

Theorem 2. *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*

$$\mathbb{E}\pi_v^m(n) = \mathbb{E}\pi_v^1(n), \quad v < n. \tag{10}$$

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:

$$\mathbb{E}\pi_v^m(v+1) = \sum_{j=0}^{m} \frac{jc}{m}\frac{1-c}{v+2}\mathbb{P}\left[v+1 \text{ has } j \text{ links to } v\right] + \frac{1-c}{v+2}, \tag{11}$$

it is equal to

$$\mathbb{E}\pi_v^m(v+1) = \frac{c}{m}\frac{1-c}{v+2}\sum_{j=0}^{m} j\mathbb{P}\left[v+1 \text{ has } j \text{ links to } v\right] + \frac{1-c}{v+2}. \tag{12}$$

The probability that a link will be created from $v + 1$ to v is equal to $\frac{m}{2mv+m} = \frac{1}{2v+1}$. Therefore, the sum in (12) is the average number of the links from $v + 1$ to v or, in other words, the average number of the successes in m Bernoulli trials with the probability of the success equal to $\frac{1}{2v+1}$. Therefore, we can write

$$\mathbb{E}\pi_v^m(v+1) = \frac{c}{m}\frac{1-c}{v+2}\frac{m}{2v+1} + \frac{1-c}{v+2} = \frac{c(1-c)}{(v+2)(2v+1)} + \frac{1-c}{v+2}. \tag{13}$$

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 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 which 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

$$ {}^{\lambda}\pi_v^m(v+t) = \sum_{k=v+1}^{v+t} \frac{c}{m} {}^{\lambda}\pi_k^m(v+t)\mathcal{M}\{k \to v,\ \lambda\} + \frac{1-c}{v+t+1}, \tag{14} $$

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 (14) over all possible realizations λ:

$$ \mathbb{E}\pi_v^m(v+t) = \sum_{k=v+1}^{v+t} \frac{c}{m} \mathbb{E}\left({}^{\lambda}\pi_k^m(v+t)\mathcal{M}\{k \to v,\ \lambda\}\right) + \frac{1-c}{v+t+1}. \tag{15} $$

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 in time later than node 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 hence, we can write

$$ \begin{aligned} \mathbb{E}\pi_v^m(v+t) &= \sum_{k=v+1}^{v+t} \frac{c}{m}\mathbb{E}\pi_k^m\mathbb{E}\mathcal{M}\{k \to v\} + \frac{1-c}{v+t+1} \qquad (16)\\ &= \sum_{k=v+1}^{v+t-1} \frac{c}{m}\mathbb{E}\pi_k^m\mathbb{E}\mathcal{M}\{k \to v\} + \frac{c}{m}\mathbb{E}\mathcal{M}\{v+t \to v\}\frac{1-c}{v+t+1} + \frac{1-c}{v+t+1} \end{aligned} $$

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

$$ \mathbb{E}(d_k(n) + m | d_k(n-1)) = d_k(n-1) + m + m\frac{m + d_k(n-1)}{2m(n-1)+m}, \tag{17} $$

taking the average over all possible network realizations, we get

$$ \mathbb{E}(d_k(n) + m) = \mathbb{E}(d_k(n-1) + m) + \frac{\mathbb{E}(d_k(n-1)+m)}{2n-1}. \tag{18} $$

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. This marks the end of the proof.

□

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

Let us provide clarifications to 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 1 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 preferential attachment mechanism the evolution of the size of $T_v(n)$ is a Markov chain: at every step n the size $|T_v(n)|$ 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 are connected to each other, therefore, the overall attractiveness of the tree $T_v(n)$ can be calculated directly and it is equal to $2|T_v(n)| - 1$. 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 tree $T_v(n)$ are given in Section 5.

- Let us consider the tree $T_v(\cdot)$ at the moment, when it has n' 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 to be at some height h depends on the number of the existing nodes with the height $h-1$ and their "popularity" (the number of the nodes at the height h). Actually, we can express this probability as a number of the nodes of the heights $h-1$ and h divided by the number of the nodes n'. 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 (8) we have a sum of a random number of random variables. To find the expectation of this sum we need some generalization of the Wald's equation. Such a generalization is provided by Kolmogorov-Prokhorov equation (Appendix A), which allows to obtain the final result in Section 7 by combining the expressions for the tree size distribution and the moment generating function of the nodes' heights.

5 Distribution of the subtree size $Y_v(n)$

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

Lemma 1. *The probability that at time step n the subtree rooted in node v has k nodes is given by*

$$\mathbb{P}[Y_v(n) = k] = \frac{\Gamma(n-v+1)\Gamma(k+1/2)\Gamma(n-i)\Gamma(v+1/2)}{\Gamma(n-v-k+1)\Gamma(k+1)\Gamma(1/2)\Gamma(v)\Gamma(n+1/2)} \tag{19}$$

Proof. We show that the evolution of the subtree size $Y_v(n) = |T_v(n)|$ can be described by the Pólya-Eggenberger urn model (see Appendix B).

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 back together with $s = 2$ balls of the same color.

The balls correspond to the in and out degrees of the nodes. The number of the 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 the 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 is connects itself to the subtree of v. Otherwise, the new node 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 Appendix B) for our problem

$$\begin{aligned}\mathbb{P}[Y_v(n) = k] &= \binom{n-v}{k} 1(1+2)\dots(1+2n) \times \\ &\quad \times \frac{2v(2v+2)\dots(2(n-k))}{(1+2v)(1+2v+2)\dots(1+2n)},\end{aligned} \tag{20}$$

or, equivalently, in its Gamma function form it gives the expression (19). □

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 outgoing degree of the nodes marked with their own numbers and the second row corresponds to the incoming degrees of the nodes. At time step 3 we have an urn with 7 balls: 6 white and 1 black. Node 0 has the in degree $d_0 = 2$, therefore, there are 3 balls bearing the mark 0. If we draw from the urn a white ball, like on Fig. 1 (b), no matter which number it has (here it is 2) we fall out of T_3. Therefore, two white balls are added. On contrary, if we choose a black ball, then the new node falls inside the tree T_3, and, therefore, we add two black

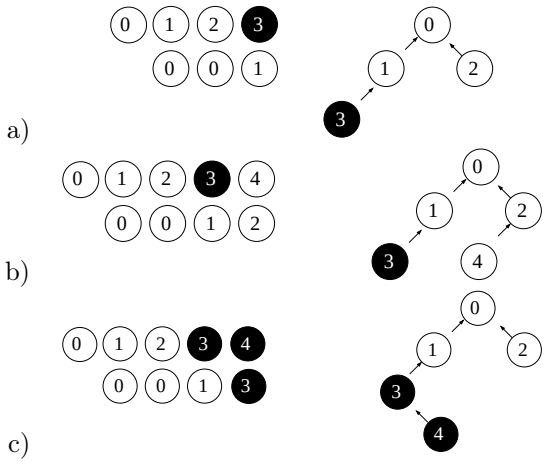

Figure 1: Illustration to the urn model: (a)Growing network after 3 steps. We choose to follow the subtree T_3. (b) Node 4 is not linked to T_3. (c) Node 4 is linked to T_3.

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 of G_n^1

In this section we obtain the moment generating function of the nodes' heights inside the subtree $G_{n'}^1 = T_v(n)$, where $n' = |T_v(n)|$. As we already remarked in the previous section, n' can be seen as a local time inside $T_v(n)$. Since $G_{n'}^1$ and G_n^1 have the same distribution, we can consider without loss of generality only the distribution of the heights in G_n^1. The later explains why we have taken a particular care for the choice of the initial weight for the initial node.

Lemma 2. *If X_n is a distance between the initial node and node n, then*

$$\mathbb{E}c^{X_n} = \frac{\Gamma(n+\frac{c}{2})\sqrt{\pi}}{\Gamma(n+\frac{1}{2})\Gamma(\frac{c}{2})}. \tag{21}$$

Proof. The evolution of X_n can be described without reference to any particular network realization, sub-lying graph or tree structure. If node n has the height k in G_n^1, then it means that it is connected to a node with the height $k-1$. The conditional probability of

such event is the the number of nodes located at the height $k-1$ plus the number of nodes located at the height k, normalized by $2n-1$, that is,

$$\mathbb{P}[X_n = k | X_{n-1}, \ldots, X_0] = \frac{\sum_{i=0}^{n-1} \mathbb{1}(X_i = k) + \sum_{i=0}^{n-1} \mathbb{1}(X_i = k-1)}{2n-1}, \tag{22}$$

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

Using (22), we can calculate the conditional moment generating function of the nodes' heights as follows:

$$\begin{aligned}
\mathbb{E}(c^{X_n} | X_{n-1}, X_{n-2} \ldots, X_0) &= \sum_{k=0}^{n} c^k \mathbb{P}[X_n = k | X_{n-1}, \ldots, X_0] && (23)\\
&= \frac{\sum_{k=0}^{n} c^k \mathbb{1}(X_{n-1} = k) + \sum_{k=1}^{n} c^k \mathbb{1}(X_{n-1} = k-1)}{2n-1} + \\
&\quad + \frac{2n-3}{2n-1} \mathbb{E}(c^X_{n-1} | X_{n-2} \ldots, X_0) && (24)\\
&= \frac{\sum_{k=0}^{n} c^k \mathbb{1}(X_{n-1} = k) + c \sum_{k=0}^{n-1} c^k \mathbb{1}(X_{n-1} = k)}{2n-1} + \\
&\quad + \frac{2n-3}{2n-1} \mathbb{E}(c^X_{n-1} | X_{n-2} \ldots, X_0), && (25)
\end{aligned}$$

where $\mathbb{1}(X_i = n) = 0$ for all $i < n$. Next, applying the double expectation value rule, we obtain the following recurrent equation

$$\mathbb{E}c^{X_n} = (1 - \frac{1-c}{2n-1}) \mathbb{E}c^{X_{n-1}}. \tag{26}$$

The above recurrent equation gives

$$\mathbb{E}c^{X_n} = \Pi_{k=1}^{n} \left[1 - \frac{1-c}{2k-1} \right] = \frac{\Gamma(n + \frac{c}{2}) \sqrt{\pi}}{\Gamma(n + \frac{1}{2}) \Gamma(\frac{c}{2})}, \tag{27}$$

which completes the proof. □

Using the derivations in the proof of Lemma 2, we can also estimate the average tree height. Namely, we have

$$\mathbb{E}X_n = \mathbb{E}X_{n-1} + \frac{1}{2n-1}$$

and, consequently,

$$\mathbb{E}X_n = \sum_{k=1}^{n} \frac{1}{2k-1}. \tag{28}$$

The equation (28) can be interpreted as follows:

Lemma 3. *The average height of the scale free growing network model G_n^1 after n time steps is of order* $\log(n)$.

This result is in line with the results of [7].

Now we can already calculate the expected PageRank value of the initial node. After taking the expectation in (9), we substitute in (9) the expression for $\mathbb{E}c^{X_n(0,k)}$ given in (21). Then, simplifying the sum, we obtain

$$\mathbb{E}\pi_0(n) = \frac{1}{1+n}\left(\frac{1}{c+1} + \frac{2\sqrt{\pi}\Gamma(n+\frac{c}{2}+1)}{(c+1)\Gamma(\frac{c}{2})\Gamma(n+1/2)}\right). \tag{29}$$

7 Final Result Statement and Asymptotics

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

Theorem 3. *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*

$$\mathbb{E}\pi_v(n) = \frac{1-c}{1+n}\left(\frac{1}{1+c} + \frac{c\Gamma(v+\frac{1}{2})\Gamma(n+\frac{c}{2}+1)}{(1+c)\Gamma(v+\frac{c}{2}+1)\Gamma(n+\frac{1}{2})}\right), \tag{30}$$

for $v > 0$, and

$$\mathbb{E}\pi_0(n) = \frac{1}{1+n}\left(\frac{1}{c+1} + \frac{2\sqrt{\pi}\Gamma(n+\frac{c}{2}+1)}{(c+1)\Gamma(\frac{c}{2})\Gamma(n+1/2)}\right), \tag{31}$$

for the particular case of $v = 0$.

Proof. First, we reduce the case $m > 1$ to the case $m = 1$ by Theorem 2.

Then, we apply Kolmogorov-Prokhorov Theorem (see Appendix A) to equation (8). Namely, we have

$$\mathbb{E}\pi_v(n) = \frac{1-c}{n+1}\left(1 + \sum_{i=1}^{\infty}\mathbb{P}[Y_v(n) \geq i]\mathbb{E}c^{X_i}\right). \tag{32}$$

Using the property that the height of a node will not be greater than the size of the tree, i.e., $\mathbb{P}[Y_v(n) > n - v] = 0$, we transform equation (32) to

$$\begin{aligned}\mathbb{E}\pi_v(n) &= \frac{1-c}{n+1}\left(1 + \sum_{i=1}^{n-v}\left(\sum_{k=i}^{n-v}\mathbb{P}[Y_v(n)=k]\right)\mathbb{E}c^{X_i}\right)\\ &= \frac{1-c}{n+1}\left(1 + \sum_{i=1}^{n-v}\mathbb{P}[Y_v(n)=i]\left(\sum_{k=1}^{i}\mathbb{E}c^{X_k}\right)\right).\end{aligned} \tag{33}$$

Next, we substitute the expressions obtained for X_n and $Y_v(n)$, equations (21) and (19), respectively, into (33) to obtain

$$\begin{aligned}\mathbb{E}\pi_v(n) &= \frac{1-c}{n+1}\left(1+\sum_{i=1}^{n-v}\frac{\Gamma(n-v+1)}{\Gamma(n-v-i+1)\Gamma(i+1)}\times\right.\\ &\left.\times\frac{\Gamma(i+1/2)\Gamma(n-i)\Gamma(v+1/2)}{\Gamma(v)\Gamma(n+1/2)}\sum_{k=1}^{i}\frac{\Gamma(k+c/2)}{\Gamma(k+1/2)\Gamma(c/2)}\right). \qquad (34)\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}\pi_v(n) &= \frac{1-c}{n+1}\left(1+\frac{\Gamma(n-v+1)\Gamma(v+1/2)}{\Gamma(v)\Gamma(n+1/2)}\times\right.\\ &\left.\sum_{i=1}^{n-v}\frac{\Gamma(n-i)\Gamma(i+1/2)}{\Gamma(n-v-i+1)\Gamma(i+1)}\left(\frac{2\sqrt{\pi}\Gamma(i+1+c/2)}{(1+c)\Gamma(i+1/2)\Gamma(c/2)}-\frac{c}{c+1}\right)\right)=\\ &= \frac{1-c}{n+1}\left(1+\frac{2\sqrt{\pi}\Gamma(n-v+1)\Gamma(v+1/2)}{(1+c)\Gamma(c/2)\Gamma(v)\Gamma(n+1/2)}\sum_{i=1}^{n-v}\frac{\Gamma(n-i)\Gamma(i+1+c/2)}{\Gamma(n-v-i+1)\Gamma(i+1)}\right.\\ &\left.-\frac{c\Gamma(n-v+1)\Gamma(v+1/2)}{(1+c)\Gamma(v)\Gamma(n+1/2)}\sum_{i=1}^{n-v}\frac{\Gamma(n-i)\Gamma(i+1/2)}{\Gamma(n-v-i+1)\Gamma(i+1)}\right). \qquad (35)\end{aligned}$$

By using the Zeilberger's algorithm and his package EKHAD for Maple, we prove (see Lemma 4 in Appendix C) the following hypergeometric identity

$$\sum_{i=1}^{n-v}\frac{\Gamma(n-i)\Gamma(i+1+c/2)}{\Gamma(n-v-i+1)\Gamma(i+1)}=\frac{\Gamma(v)\Gamma(n+c/2+1)\Gamma(1+c/2)}{\Gamma(v+c/2+1)\Gamma(n-v+1)}-\frac{\Gamma(n)\Gamma(1+c/2)}{\Gamma(n-v+1)}. \qquad (36)$$

We can apply this identity to the both sums in (35), 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 (30). □

The expression (30) 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

$$\mathbb{E}\pi_v(n)\approx\frac{1-c}{1+n}\left(\frac{1}{1+c}+\frac{c}{1+c}(v+\frac{1}{2})^{-\frac{1+c}{2}}(n+\frac{1}{2})^{\frac{1+c}{2}}\right). \qquad (37)$$

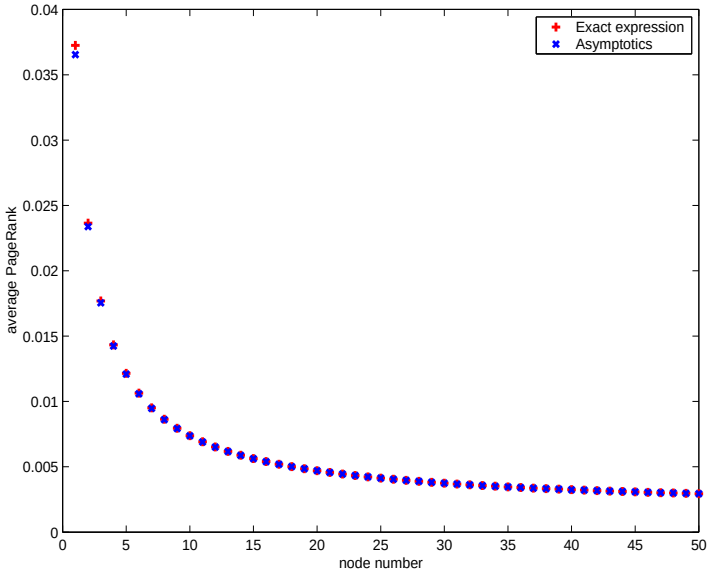

Figure 2: Comparison between the asymptotics (37) and the exact expression (30).

or, neglecting the first term,

$$\mathbb{E}\pi_v(n) \approx \frac{1-c}{1+c} c v^{-\frac{1+c}{2}} n^{-\frac{1-c}{2}}. \tag{38}$$

In particular, for the zero node, we have

$$\mathbb{E}\pi_0(n) \approx \frac{2\sqrt{\pi}}{(1+c)\Gamma(\frac{c}{2})} n^{-\frac{1-c}{2}}. \tag{39}$$

As one can see from Figure 2, the asymptotics (37) indeed closely follows the exact expression (30).

8 Discussion and comparison with related work

First, let us compare our results with the results of Pandurangan et al. [20]. In the present work we have obtained 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. [20] the authors have used "mean-field" approach [1, 2, 3] 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 [20]. Specifically, suppose that

n is fixed, PageRank depends continuously on v and the node age is uniformly distributed. Then, using our asymptotic expression (37), we obtain

$$\begin{aligned} P(x) &= \mathbb{P}[\pi_v < x] \\ &\approx \mathbb{P}\left[v > \left(\left(\frac{1+n}{1-c}x - \frac{1}{1+c}\right)\frac{1+c}{c}(n+\frac{1}{2})^{-\frac{1+c}{2}}\right)^{-\frac{2}{1+c}} - 1/2\right] \\ &= 1 - \left(\left(\left(\frac{1+n}{1-c}x - \frac{1}{1+c}\right)\frac{1+c}{c}(n+\frac{1}{2})^{-\frac{1+c}{2}}\right)^{-\frac{2}{1+c}} - 1/2\right)/n \\ &= 1 + \frac{1}{2n} - \left(1+\frac{1}{2n}\right)c^{\frac{2}{1+c}}\left(\frac{1+c}{1-c}(n+1)x - 1\right)^{-\frac{2}{1+c}}. \end{aligned} \tag{40}$$

In particular, we note that $P(\frac{1-c}{n+1}) = 0$, as $x = \frac{1-c}{n+1}$ is the minimal value of PageRank. Taking the derivative of (40), we obtain the density distribution function of the PageRank value

$$p(x) = \frac{2}{1-c}(n+1)\left(1+\frac{1}{2n}\right)c^{\frac{2}{1+c}}\left(\frac{1+c}{1-c}(n+1)x - 1\right)^{-\frac{3+c}{1+c}}. \tag{41}$$

For large values of n and for values of x, which are not too small and not too close to one, the expression (41) 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 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. [20] gives the exponent 2 and the experiments with the real Web data in Pandurangan et al. [20] give the exponent 2.1.

To test the mean-field estimation (40), we have run simulations of our growing network model. The network was grown up to $n = 1000$ for 100000 simulation runs with $m = 10$. In Figure 3 the mean-field estimation (40) 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 [18], when dealing with power laws, it is preferable to work with the cumulative complimentary distribution function rather then 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 the exponent $x^{-\alpha+1}$. When calculating the PageRank, we have used $c = 0.85$. We note that plot is indeed close to a stright line for the middle segment of the PageRank range. In [20] the authors also noticed that PageRank follows a power law except those pages with very small PageRank. This phenomenon can easily be explained with the

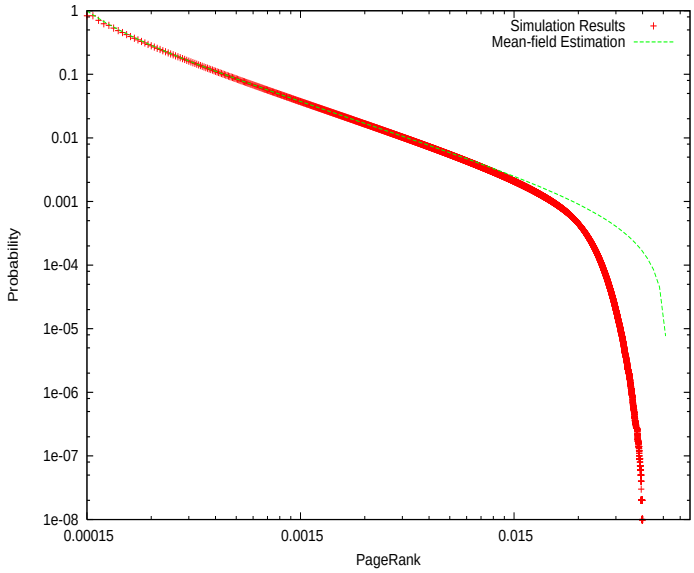

Figure 3: Cumulative complimentary distribution function: Simulation results compared to the mean-field estimation.

help of (41). The term $\frac{1+c}{1-c}(n+1)x$ becomes comparable with 1 in (41) for values of x too close to the minimal PageRank $\frac{1-c}{n+1}$ and the distribution density function can not 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 is simply not enough such nodes to use the "averaging" argument.

As it can be observed from (31) and (30), 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

$$\tilde{\pi}_v = \lim_{n\to\infty} \frac{\mathbb{E}\pi_v(n)}{\mathbb{E}\pi_0(n)} = \frac{(1-c)c\Gamma(\frac{c}{2})}{2\sqrt{\pi}} \frac{\Gamma(v+\frac{1}{2})}{\Gamma(v+\frac{c}{2}+1)}. \tag{42}$$

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 some initial nodes.

Recall that Google divides the whole range of PageRank in 10 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:

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

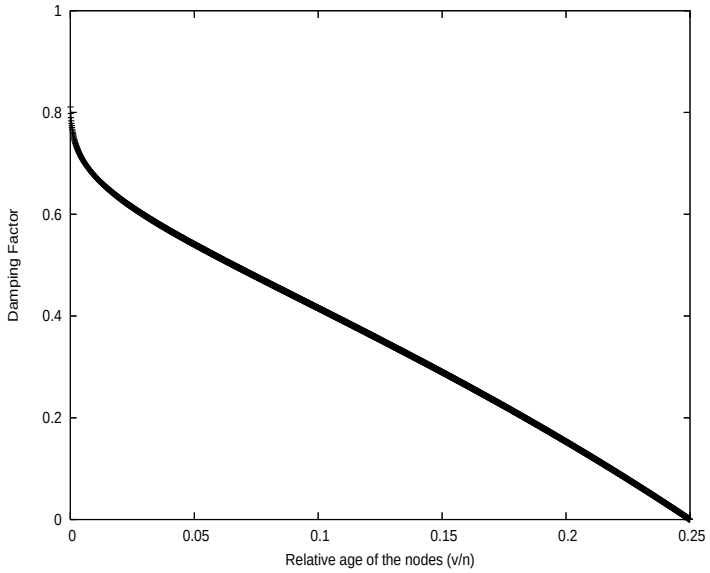

Figure 4: The optimal value of c as a function of v/n.

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

The authors of [5] investigated the effect of damping factor c on PageRank. In their numerical example they have noticed that 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 which maximizes the PageRank expression (37) 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 value of PageRank for node 1 and node 2 when $n = 10000$. If the World Wide Web has 8 billion pages, then the present model suggests that the pages that mostly benefit 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 Web pages 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 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 different preferential attachment process. The expression (19) would change slightly, but there is now guarantee, that one could find a simple closed form of the final expression (30).

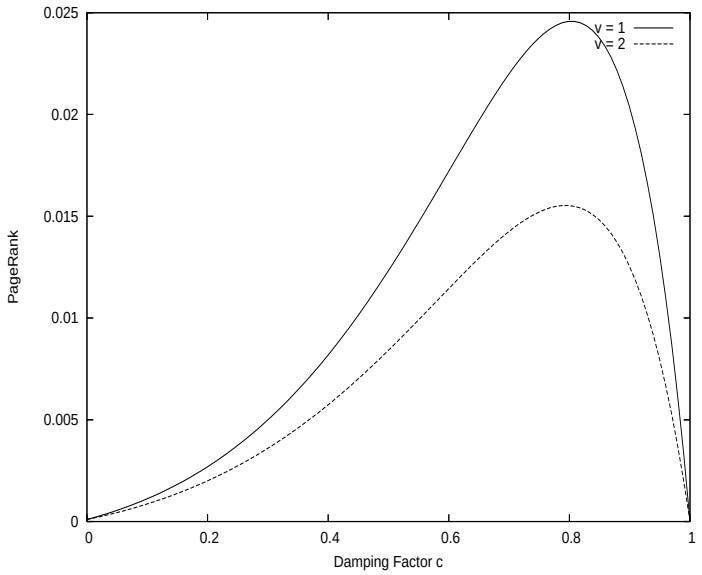

Figure 5: The PageRank as a function of c for $v = 1, 2$ and $n = 10000$

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 is an interesting perspective research direction.

Acknowledgments

We would like to thank Nelly Litvak, Vladimir Dobrynin, Bruno Salvy and Son Kim Pham for discussions and helpful remarks during the manuscript preparation.

A Kolmogorov-Prokhorov equation

We present the Kolmogorov-Prokhorov equation following the book of A. Borovkov ([9], chapter 4.4). The Kolmogorov-Prokhorov equation can be seen as a generalization of Wald's equation.

Theorem 4. *If an integer non-negative random variable ν does not depend on the future with the respect to the sequence of random variables $\{\xi_n\}$ and*

$$\sum_{k=1}^{\infty} \mathbb{P}(\nu \geq k)\mathbb{E}|\xi_k| < \infty,$$

then

$$\mathbb{E}\sum_{k=1}^{\nu}\xi_k = \sum_{k=1}^{\infty}\mathbb{P}(\nu \geq k)\mathbb{E}\xi_k.$$

In our case $\nu = Y_v(n)$ and $\xi_k = c^{X_k}$. The random variable $Y_v(n)$ is independent of $\{X_k\}$, and c^{X_k} are positive, therefore, $\mathbb{E}|c^{X_k}| = \mathbb{E}c^{X_k}$ and the Kolmogorov-Prokhorov equation can be applied to the problem we consider in this paper.

B Pólya-Eggenberger urn model

We follow the book [14] in our description of the urn models. The Pólya-Eggenberger urn model starts with one urn where one can find $b+w$ 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, then it is returned back 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 have k black balls in the urn at time step n can then be expressed as

$$\begin{aligned}\mathcal{P}_{n,k}(w,b,s) &= \binom{n}{k} b(b+s)\dots(b+(k-1)s)\times \\ &\quad \times \frac{w(w+s)\dots(w+(n-k-1)s)}{(b+w)(b+w+s)\dots(b+w+(n-1)s)},\end{aligned} \tag{43}$$

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

$$\begin{aligned}\mathcal{P}_{n,k}(w,b,s) &= \binom{n}{k}\frac{\Gamma(\frac{b+w}{s})\Gamma(\frac{b}{s}+k)\Gamma(\frac{w}{s}+n-k)}{\Gamma(b/s)\Gamma(w/s)\Gamma(\frac{b+w}{s}+n)} \\ &= \binom{n}{k}\frac{B(\frac{b}{s}+k,\frac{w}{s}+n-k)}{B(\frac{b}{s},\frac{w}{s})}\end{aligned} \tag{44}$$

It is worthy to note here that the problem of the tree height, which we study in the section 6, can also be described in terms of the urn model with a node height value as a mark (or color).

C Zeilberger's Algorithm

We follow the book [21] in our description of the Zeilberger's algorithm. Let us consider a sum

$$f(n) = \sum_k F(n,k) \tag{45}$$

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

$$\sum_{j=0}^{J} a_j(n)F(n+j,k) = G(n,k+1) - G(n,k) \tag{46}$$

This method is also called the method of *creative telescoping*. When such representation is obtained, we can sum the equation (46) 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:

$$\sum_{j=0}^{J} a_j(n)f(n+j) = 0 \tag{47}$$

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 has written the package EKHAD [13] for Maple [16], which implements his algorithm and finds $a_0, \ldots, a_J$ and $R(\cdot,\cdot)$ such that

$$G(n,k) = R(n,k)F(n,k) \tag{48}$$

Fortunately, the Zeilberger's Algorithm gives a satisfying result for the sum in (34). Let us prove the following lemma.

Lemma 4.

$$\sum_{i=1}^{n-v} \frac{\Gamma(n-i)\Gamma(i+1+c/2)}{\Gamma(n-v-i+1)\Gamma(i+1)} = \frac{\Gamma(v)\Gamma(n+c/2+1)\Gamma(1+c/2)}{\Gamma(v+c/2+1)\Gamma(n-v+1)} - \frac{\Gamma(n)\Gamma(1+c/2)}{\Gamma(n-v+1)} \tag{49}$$

Proof. Consider the internal sum (34). We introduce the following notation:

$$F(n,i) = \frac{\Gamma(n+v-i)\Gamma(i+1+c/2)\Gamma(v+c/2+1)\Gamma(n+v+1)}{\Gamma(n-i+1)\Gamma(i+1)\Gamma(v)\Gamma(n+v+c/2+1)\Gamma(1+c/2)} \tag{50}$$

It is the summand from (49) divided by the result we want to prove (it was guessed from the values of $\mathbb{E}\pi_v$ for $v = 1, 2, 3$), plus we change the variable from $n \to n - v$ and we add the 0th summand. Now, we want to prove that

$$f(n) = \sum_{i=0}^{n} F(n,i) \equiv 1 \tag{51}$$

The function *zeil* from EKHAD [13] package in Maple [16] finds the following identities: $a_0 = -1$, $a_1 = 1$ and

$$R(n,i) = -\frac{(n+v-i)i}{(n+v+\frac{c}{2}+1)(n-i+1)} \tag{52}$$

Therefore, in our case (46) takes the following form

$$F(n,i) - F(n+1,i) = R(n,i+1)F(n,i+1) - R(n,i)F(n,i) \tag{53}$$

Then we sum the equation (53) for the values $i = 0, \ldots, n-1$ 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 (50) and (52). Therefore, $f(n) = f(n+1)$. As $f(0) = 1$, it completes the prove.

Note, that it is indeed a proof, because Maple [16] and EKHAD [13] provide the identities (i.e. the values of a_i, $i = 0, 1$ and (52)) that can be easily checked. □

Contents

1 Introduction **3**

2 Scale Free Network Models **4**

3 PageRank of Growing Networks **5**

4 The case $m > 1$ can be reduced to the case $m = 1$ **7**

5 Distribution of the subtree size $Y_v(n)$ **10**

6 Moment Generating function of the nodes' heights in subtrees of G_n^1 **11**

7 Final Result Statement and Asymptotics **13**

8 Discussion and comparison with related work **15**

A Kolmogorov-Prokhorov equation **19**

B Pólya-Eggenberger urn model **20**

C Zeilberger's Algorithm **20**

References

[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[2] A.-L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272:173–187, 1999.

[3] A.-L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: The topology of the world wide web. *Physica A*, 281:69–77, 2000.

[4] M. Bianchini, M. Gori, and F. Scarselli. Inside pagerank. *ACM Trans. Internet Technology*, 5(1):92–128, 2005.

[5] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Pagerank as a function of the damping factor. In *Proc. WWW 2005 conference*. ACM Press, 2005.

[6] 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):279–290, 2001.

[7] Béla Bollobás and Oliver Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[8] Béla Bollobás and Oliver M. Riordan. Mathematical results in scale-free random graphs. *in Handbook of Graphs and Networks, eds. S. Bornholdt and H.G. Schuster, Wiley-VCH*, pages 1–34, 2002.

[9] A.A. Borovkov. *Probability Theory.* Gordon and Breach Science Publishers, 1998.

[10] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1-7):107–117, 1998.

[11] Sergei N. Dorogovtsev and Jose F.F. Mendes. Evolution of networks. *Advances in Physics*, 51:1079–1187, 2002.

[12] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Structure of growing networks with preferential linking. *Phys. Rev. Lett.*, 85:4633–4636, 2000.

[13] Ekhad software package. http://www.math.temple.edu/~zeilberg/programs.html.

[14] Norman L. Johnson and Samuel Kotz. *Urn Models and Their Application.* John Wiley and Sons, 1977.

[15] A.N. Langville and C.D.Meyer. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2005.

[16] Maple software, http://www.maplesoft.com.

[17] C.B. Moler. *Numerical Computing with MATLAB.* SIAM, 2004.

[18] M E J Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323, 2005.

[19] Mark E.J. Newman. The structure and function of complex networks. *SIAM Reviews*, 45(2):167–256, 2003.

[20] Gopal Pandurangan, Prabhakara Raghavan, and Eli Upfal. Using pagerank to characterize web structure. In *8th Annual International Computing and Combinatorics Conference (COCOON)*, 2002.

[21] Marko Petkovsek, Herbert S. Wilf, and Doron Zeilberger. *A=B.* A. K. Peters, 1996.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399