In-Degree and PageRank: Why Do They Follow Similar Power Laws?

N. Litvak, W. R. W. Scheinhardt, and Y. Volkovich

Abstract. PageRank is a popularity measure designed by Google to rank Web pages. Experiments confirm that PageRank values obey a power law with the same exponent as In-Degree values. This paper presents a novel mathematical model that explains this phenomenon. The relation between PageRank and In-Degree is modeled through a stochastic equation, which is inspired by the original definition of PageRank, and is analogous to the well-known distributional identity for the busy period in the $M/G/1$ queue. Further, we employ the theory of regular variation and Tauberian theorems to prove analytically that the tail distributions of PageRank and In-Degree differ only by a multiplicative constant, for which we derive a closed-form expression. Our analytical results are in good agreement with experimental data.

1. Introduction

In this paper we study the relation between the probability distributions of the PageRank and the In-Degree of a randomly selected Web page. The notion of *PageRank* was introduced by Google in order to characterize the popularity of Web pages numerically. The original description of PageRank presented in [Brin and Page 98] is as follows:

$$\mathrm{PR}(i) = c \sum_{j \to i} \frac{1}{d_j} \mathrm{PR}(j) + (1 - c), \tag{1.1}$$

where $PR(i)$ is the PageRank of page i, d_j is the number of outgoing links of page j, the sum is taken over all pages j that link to page i, and c is the "damping factor," which is some constant between 0 and 1. The *In-Degree* of a Web page denotes the number of incoming hyperlinks to that page. From Equation (1.1) it is clear that the PageRank of a page depends on its In-Degree and the importance (i.e., PageRanks) of the pages that link to it.

We focus in particular on the tail asymptotics for PageRank and its connection to In-Degree. By *tail* of the PageRank distribution we mean the fraction of pages $\mathbb{P}(PR > x)$ having PageRank greater than x, where x is large. Thus, we are concerned only with pages of high ranking. A common way to analyze tail behavior is to find an asymptotic expression $p(x)$ such that $\mathbb{P}(PR > x)/p(x) \to 1$ as $x \to \infty$. In this case, $p(x)$ and $\mathbb{P}(PR > x)$ are asymptotically similar, and thus we can approximate $\mathbb{P}(PR > x)$ by $p(x)$ for x large enough.

Pandurangan et al. [Pandurangan et al. 02] observed that the tails of PageRank and In-Degree distributions for Web data seem to follow power laws with the same exponent. Loosely speaking, a power law with exponent α means that the probability that the random variable takes values greater than some large number x is approximately proportional to $x^{-\alpha}$. Formally, this can be modeled as asymptotic similarity of PageRank and In-Degree tail distributions to some power law functions. It turns out that, for both PageRank and In-Degree distributions, the power law exponent α is about 1.1.

Recent extensive experiments [Donato et al. 04] and [Fortunato et al. 08] confirmed the similarity in tail behavior observed in [Pandurangan et al. 02]. Becchetti and Castillo extensively investigated the influence of the damping factor c on the power law behavior of PageRank [Becchetti and Castillo 06]. They have shown that the PageRank of the top 10% of the nodes always follows a power law with the same exponent independent of the value of the damping factor. Our own experiments based on Web data from [Kamvar 06] are also in agreement with [Pandurangan et al. 02] (see Figure 1 in Section 5.2 and the discussion there).

Obviously, Equation (1.1) suggests that PageRank and In-Degree are intimately related, but this formula by itself does not explain the observed similarity in tail behavior. Furthermore, the linear algebra methods that have been commonly used in the PageRank literature [Berkhin 05, Langville and Meyer 03] and proved to be very successful for designing efficient computational methods, seem to be insufficient for modeling and analyzing the asymptotic properties of the PageRank distribution.

The goal of our paper is to provide mathematical evidence for the power-law behavior of PageRank and its relation to the In-Degree distribution. We

propose a stochastic model to explain this relation. Our approach is inspired by techniques from applied probability and stochastic operations research. The relation between PageRank and In-Degree is modeled through a distributional identity analogous to the equation for the busy period in the $M/G/1$ queue (see, e.g., [Robert 03]). Further, we analyze our model using the approach employed in [De Meyer and Teugels 80] for studying the tail behavior of the busy period in the case where the service times are regularly varying random variables. This fits in our research because regular variation is, in fact, a generalization of the power law, and it has been widely used in queueing theory to model self-similarity, long-range dependence and heavy tails [Zwart 01]. Thus, we use the notion of regular variation to model the power law distribution of the In-Degree. For the sake of completeness, in Section 2 we introduce regularly varying random variables and describe their basic properties.

To obtain the tail behavior of PageRank in our model, we use Laplace-Stieltjes transforms and apply Tauberian theorems presented in [Bingham and Doney 74]; see also Theorem 8.1.6 in [Bingham et al. 89]. Even though our model contains some rather rigid simplifying assumptions—the most notable being independence between pages that link to the same page and a constant Out-Degree for all pages—these techniques allow us to prove the similarity in tail behavior for PageRank and In-Degree, thus suggesting that our assumptions do not touch upon the underlying reasons for this similarity. Moreover, our analysis allows us to derive explicitly the multiplicative constant that quantifies the difference between PageRank and In-Degree tail behavior.

We believe that our approach is extremely promising for analyzing the PageRank distribution and solving other problems related to the structural properties of the Web. At the end of this paper, we will briefly mention other possibilities for probabilistic analysis of the PageRank distribution. In particular, we will provide experimental results for Growing Networks [Barabási and Albert 99], and we will draw a parallel between the recent studies [Avrachenkov and Lebedev 06, Fortunato and Flammini 06] on PageRank behavior in this class of graph models and our present work.

2. Preliminaries

This section describes important properties of regularly varying random variables. We follow definitions and notations from [Bingham and Doney 74], [De Meyer and Teugels 80], and [Zwart 01]. More comprehensive details can be found in [Bingham et al. 89].

Definition 2.1. A function $V(x)$ is *regularly varying* of index $\alpha \in \mathbb{R}$ if for every $t > 0$,

$$\frac{V(tx)}{V(x)} \to t^\alpha \quad \text{as} \quad x \to \infty.$$

If $\alpha = 0$, then V is called *slowly varying*.

Definition 2.2. A function L is *slowly varying* if for every $t > 0$,

$$\frac{L(tx)}{L(x)} \to 1 \quad \text{as} \quad x \to \infty.$$

A function $V(x)$ is regularly varying if and only if it can be written in the form

$$V(x) = x^\alpha L(x),$$

for some slowly varying $L(x)$.

The following lemma provides a useful bound for slowly varying functions.

Lemma 2.3. (Potter Bounds.)   *Let L be a slowly varying function. Then, for any fixed $A > 1, \delta > 0$ there exists a finite constant $K > 1$ such that for all $x_1, x_2 > K$,*

$$\frac{L(x_1)}{L(x_2)} \le A \max \left\{ \left(\frac{x_1}{x_2} \right)^\delta, \left(\frac{x_1}{x_2} \right)^{-\delta} \right\}.$$

Definition 2.4. In probability theory, a random variable X is said to be *regularly varying* with index α if its distribution F is such that

$$1 - F(x) \sim x^{-\alpha} L(x) \quad \text{as} \quad x \to \infty,$$

for some positive slowly varying function $L(x)$. Here, as in the remainder of this paper, the notation $a(x) \sim b(x)$ means that $a(x)/b(x) \to 1$.

Denote by $f(s) = \mathbb{E}e^{-sX}$, $s > 0$, the Laplace-Stieltjes transform of X, and let $\xi_n = \int_0^\infty x^n dF(x)$ be the nth moment of X. The successive moments of F can be obtained by expanding $f(s)$ in a series at $s = 0$. More precisely, we have the following.

Lemma 2.5. *The nth moment of X is finite if and only if there exist numbers $\xi_0 = 1$ and $\xi_1, ..., \xi_n$, such that*

$$f(s) - \sum_{i=0}^n \frac{\xi_i}{i!}(-s)^i = o(s^n) \text{ as } s \to 0.$$

In that case, ξ_i is the ith moment of X.

If $\xi_n < \infty$ then we introduce the notation $(n \in \mathbb{N})$

$$f_n(s) = (-1)^{n+1}\left(f(s) - \sum_{i=0}^{n}\frac{\xi_i}{i!}(-s^i)\right).\tag{2.1}$$

Remark 2.6. It follows from Lemma 2.5 that the nth moment of X is finite if and only if there exist numbers $\xi_0 = 1$ and $\xi_1, ..., \xi_n$ such that $f_n(s) = o(s^n)$ as $s \to 0$.

The following theorem establishes the relation between asymptotic behavior of a regularly varying distribution and its Laplace-Stieltjes transform. This result plays an essential role in our analysis.

Theorem 2.7. (Tauberian Theorem.)   *If $n \in \mathbb{N}$, $\xi_n < \infty$, $\alpha \in (n, n+1)$, then the following are equivalent:*

(i) $f_n(s) \sim (-1)^n\Gamma(1 - \alpha)s^\alpha L(\frac{1}{s})$ *as* $s \to 0$,

(ii) $1 - F(x) \sim x^{-\alpha}L(x)$ *as* $x \to \infty$.

Here and in the remainder of the paper we use the letter α to denote the index of the complementary distribution function $1 - F(x)$ rather than that of the density.

3.   Model

In this section we introduce a model that describes the relation between PageRank and In-Degree in the form of a stochastic equation. This model naturally follows from the definition of PageRank in Equation (1.1), and is analytically tractable, thus enabling us to obtain the asymptotic behavior of PageRank. As will become clear, we make several rather severe simplifying assumptions. Nevertheless, the theoretical results of this model show a good match with observed Web graph behavior.

3.1.   Relation between In-Degree and PageRank

Our goal now is to describe the relation between PageRank and In-Degree. To this end, we keep Equation (1.1) almost unchanged but we make several assumptions. First, let R be the PageRank of a randomly chosen page. We treat R simply as a random variable whose distribution we want to determine. Second, we assume that the number of outgoing links $d \geq 1$ is the same for each page.

Although this assumption is not realistic, it helps us to focus on the influence of
In-Degree without considering other factors. We note, however, that the present
model allows for various generalizations. For instance, we can take into account
pages without outgoing links (*dangling nodes*) as will be discussed below.

Under the assumptions above, the random variable R satisfies a distributional
identity

$$R \stackrel{d}{=} c \sum_{j=1}^{N} \frac{1}{d} R_j + (1 - c), \tag{3.1}$$

where N is the In-Degree of the considered random page.

We now make the assumption that N and the R_j's are independent, and that
the R_j's have the same distribution as R itself. We note that the independence
assumption is obviously not true in general. It is also not the case, however, that
the PageRank values of the pages linking to the same page i are directly related,
so we may assume independence in this study.

The novelty of our approach is that we treat the PageRank as a random
variable that solves a certain stochastic equation. We believe that this approach
is quite natural if our goal is to explain the power law behavior of PageRank,
because the power law is merely a description of a certain class of probability
distributions.

One of the nice features of the stochastic equation (3.1) is that it has the same
form as the original formula (1.1). Thus, we may hope that our model correctly
describes the relation between In-Degree and PageRank. This is easy to verify
in the extreme (unrealistic) case when all pages have the same In-Degree d. In
this situation, the PageRanks of all pages are equal, and it is easy to see that
$R \equiv 1$ constitutes the unique solution of (3.1).

Note that in the present model, we can easily account for pages without out-
going links. There are different ways to deal with such nodes when defining the
ranking. We consider a classic scenario, where a dangling node is equivalent to a
node that has an outgoing link to every page in the Web. Then Equation (1.1)
becomes

$$\mathrm{PR}(i) = c \sum_{j \to i} \frac{1}{d_j} \mathrm{PR}(j) + \frac{c}{n} \sum_{j \in \mathcal{D}} \mathrm{PR}(j) + (1 - c), \tag{3.2}$$

where $\mathcal{D}$ is the set of dangling nodes, and n is the number of pages in the Web.
Assume further that the PageRank of a random page does not depend on whether
the page is dangling. Indeed, it can be shown that the PageRank of a page cannot
be altered significantly by modifying its outgoing links [Avrachenkov and Litvak
06]. Moreover, experiments (e.g., in [Eiron et al. 04]) show that dangling nodes
are often just regular pages whose links have not been crawled. Besides, even
authentically dangling pages such as `.pdf` or `.ps` files often contain important

information and gain high ranking independently of the fact that they do not have outgoing links. Such independence implies, in particular, that the average PageRank of dangling nodes is 1, and thus the fraction of the total PageRank mass concentrated in dangling nodes, approximately equals the fraction of dangling nodes p_0:

$$p_0 = \frac{|\mathcal{D}|}{n} \approx \frac{1}{n} \sum_{j \in \mathcal{D}} \mathrm{PR}(j).$$

Note that in the presence of dangling nodes, the average Out-Degree of non-dangling nodes is $d(1 - p_0)^{-1}$.

Now, exactly as our main stochastic equation (3.1) is analogous to (1.1), we can also provide a stochastic equation analogous to (3.2) as follows:

$$R \stackrel{d}{=} c \sum_{j=1}^{N} \frac{1 - p_0}{d} R_j + [1 - c(1 - p_0)].$$

Observe that this is the same stochastic equation as (3.1), only with $c(1 - p_0)$ instead of c. Thus, the results of our current work can be applied directly after the corresponding straightforward adjustments. To simplify notation, we will just use c throughout the paper.

3.2. In-Degree Distribution

It is well known that the In-Degree of Web pages follows a power law. For our analysis, however, we need a more formal description of this random variable; thus, we suggest employing the theory of regular variation. We model the In-Degree of a randomly chosen page as a non-negative, integer, regularly varying random variable, distributed as $N(X)$, where X is regularly varying with index α and $N(x)$ is the number of Poisson arrivals on the time interval $[0, x]$. Without loss of generality, we assume that the rate of the Poisson process is equal to 1.

The advantage of this construction is that we do not need to impose any restrictions on X and can at the same time ensure that the In-Degree is an integer. We claim that the random variable $N(X)$ is regularly varying with the same index as X, or, more informally, $N(X)$ follows a power law with the same exponent. Thus, we can think of $N(X)$ as the In-Degree of a random Web page. For the sake of completeness we present the formal statement and its proof in the remainder of this section.

Let F_X and $F_{N(X)}$, f and ϕ be the distribution functions and the Laplace-Stieltjes transforms of X and $N(X)$, respectively. Since the random variable X is regularly varying, we have by definition

$$1 - F_X(x) \sim x^{-\alpha} L(x) \quad \text{as } x \to \infty, \tag{3.3}$$

where $L(x)$ is some slowly varying function. Then we claim that for $N(X)$ the following also holds:

$$1 - F_{N(X)}(x) \sim x^{-\alpha} L(x) \quad \text{as } x \to \infty. \tag{3.4}$$

For completeness, we prove this statement using the Tauberian theorem (Theorem 2.7). To this end, we first have to show that the corresponding moments of X and $N(X)$ always exist together. Assuming that $\mathbb{E}X = d$ we immediately get $\mathbb{E}N(X) = d$. Next, we consider the generating function of $N(X)$,

$$\mathbb{G}_{N(X)}(s) = \mathbb{E}s^{N(X)} = \int_0^\infty \mathbb{E}s^{N(t)} dF_X(t) = \int_0^\infty e^{-t(1-s)} dF_X(t) = f(1-s),$$
$$\tag{3.5}$$

from which we derive the Laplace-Stieltjes transform of $N(X)$ in terms of the Laplace-Stieltjes transform of X:

$$\phi(s) = \mathbb{E}e^{-sN(X)} = f(1 - e^{-s}).$$

Now, denote by $\xi_1 = d, \xi_2, \ldots, \xi_n$ and $\nu_1 = d, \nu_2, \ldots, \nu_n$ the first n moments of X and $N(X)$, respectively, and put $\xi_0 = \nu_0 = 1$. Then, provided that ξ_n and ν_n are finite, we define, correspondingly,

$$f_n(s) = (-1)^{n+1} \left(f(s) - \sum_{i=0}^n \frac{\xi_i}{i!}(-s)^i \right) \quad \text{and}$$

$$\phi_n(s) = (-1)^{n+1} \left(\phi(s) - \sum_{i=0}^n \frac{\nu_i}{i!}(-s)^i \right)$$

as in (2.1).

We will establish an auxiliary result formulated in the next lemma (see Section 7 for the proof).

Lemma 3.1. *For $n \geq 1$, the following are equivalent:*

(i) $\xi_n < \infty$,

(ii) $\nu_n < \infty$.

Remark 3.2. We can also reformulate Lemma 3.1 as follows:

$$f_n(s) = o(s^n) \quad \text{if and only if} \quad \phi_n(s) = o(s^n).$$

Remark 3.3. It follows from the proof of Lemma 3.1 that if $\xi_n < \infty$, then

$$f_n\left(1 - e^{-s}\right) = \phi_n(s) + O(s^{n+1}).$$

Now, we can use Theorem 2.7 to prove that (3.3) implies (3.4). In fact the reverse also holds, as stated in the following theorem.

Theorem 3.4. *The following are equivalent:*

(i) $1 - F_X(x) \sim x^{-\alpha}L(x)$ *as* $x \to \infty$,

(ii) $1 - F_{N(X)}(x) \sim x^{-\alpha}L(x)$ *as* $x \to \infty$.

Proof. $(i) \to (ii)$. From Theorem 2.7 for X, we know that

$$1 - F_X(x) \sim x^{-\alpha}L(x) \quad \text{as} \quad x \to \infty$$

implies

$$f_n(s) \sim (-1)^n\Gamma(1 - \alpha)s^\alpha L\left(\frac{1}{s}\right) \quad \text{as} \quad s \to 0, \tag{3.6}$$

where $\alpha > 1$ is not integer and n is the greatest integer less than α.

Since $1 - e^{-s} \sim s$ as $s \to 0$, it follows from (3.6) and Lemma 2.3 that $f_n(s) \sim f_n(1 - e^{-s})$ as $s \to 0$. Then we use (3.6) and Remark 3.3 to obtain

$$\phi_n(s) \sim (-1)^n\Gamma(1 - \alpha)s^\alpha L\left(\frac{1}{s}\right) \quad \text{as} \quad s \to 0.$$

Now we again apply Theorem 2.7 to conclude that

$$1 - F_{N(X)} \sim x^{-\alpha}L(x) \quad \text{as} \quad x \to \infty.$$

$(ii) \to (i)$. Similar to the first part of the proof. $\square$

Thus, our model for the number of incoming links properly describes an In-Degree distribution that follows a power law with finite expectation and a non-integer exponent.

3.3. Main Stochastic Equation

Combining the ideas from Sections 3.1 and 3.2, we arrive at the following equation:

$$R \overset{d}{=} c \sum_{j=1}^{N(X)} \frac{1}{d}R_j + (1 - c), \tag{3.7}$$

where $c \in (0,1)$ is the damping factor, $d \in \{1, 2, \ldots\}$ is the fixed Out-Degree of each page, and $N(X)$ describes the In-Degree of a randomly chosen page as the number of Poisson arrivals on a regularly varying time interval X. As we discussed above, stochastic equation (3.7) adequately captures several important aspects of the PageRank distribution and its relation to the In-Degree distribution. Moreover, our model is completely formalized, so we can apply analytical methods in order to derive the tail behavior of the random variable R representing the PageRank.

Linear stochastic equations like (3.7) have a long history. In particular, (3.7) is similar to the famous equation that arises in the theory of branching processes and describes many real-life phenomena, such as the distribution of the busy period in the $M/G/1$ queue:

$$B \stackrel{d}{=} \sum_{i=1}^{N(S_1)} B_i + S_1,$$

where B is the distribution of the busy period (the time interval during which the queue is non-empty), S_1 is the service time of the customer that initiated the busy period, $N(S_1)$ is the number of Poisson arrivals during this service time and the B_i's are independent and distributed as B. We refer to [Robert 03] and other books on queueing theory for more details. Also, see [Zwart 01] for an excellent detailed treatment of queues with regular variation, and specifically the busy period problem. We note also that our Equation (3.7) is a special case in a rich class of stochastic recursive equations that were discussed in detail in the recent survey [Aldous and Bandyopadhyay 05].

This concludes the model description. The next step will be to use our model for providing a rigorous explanation of the indicated connection between the distributions of In-Degree and PageRank.

4. Analysis

The idea of our analysis is to write the equation for the Laplace-Stieltjes transforms of X and R and then make use of the Tauberian theorems to prove that R is regularly varying with the same index as X. According to Theorem 3.4, this will give us the desired similarity in tail behavior of the PageRank R and the In-Degree $N(X)$.

As a result of the assumptions from Section 3, we can express the Laplace-Stieltjes transform $r(s)$ of the PageRank distribution R in terms of the probability generating function of $N(X)$ using (3.7):

$$r(s) = \mathbb{E}e^{-sR} = e^{-s(1-c)}\mathbb{E}\exp\left(-s\frac{c}{d}\sum_{i=1}^{N(X)}R_i\right)$$

$$= e^{-s(1-c)}\sum_{k=0}^{\infty}\mathbb{E}\exp\left(-s\frac{c}{d}\sum_{i=1}^{k}R_i\right)\mathbb{P}(N(X)=k)$$

$$= e^{-s(1-c)}\sum_{k=0}^{\infty}\left(r\left(s\frac{c}{d}\right)\right)^k\mathbb{P}(N(X)=k)$$

$$= e^{-s(1-c)}\mathbb{G}_{N(X)}\left(r\left(s\frac{c}{d}\right)\right).$$

Since, by (3.5), $\mathbb{G}_{N(X)}(s) = f(1-s)$, we arrive at

$$r(s) = f\left(1 - r\left(\frac{c}{d}s\right)\right)e^{-s(1-c)}. \tag{4.1}$$

It can be shown that, for the typical values $d > 1$ and $0 < c < 1$, Equation (4.1) has a unique solution $r(s)$ which is completely monotone and has $r(0) = 1$.

As in Section 3.2, we will start the analysis by providing the correspondence between existence of the nth moments of X and R. Recall that $\xi_1, \ldots, \xi_n$ denote the first n moments of X. Further, denote the first n moments of R by $\rho_1, \ldots, \rho_n$, and, if $\rho_n < \infty$, we define

$$r_n(s) = (-1)^{n+1}\left(r(s) - \sum_{k=0}^{n}\frac{\rho_k}{k!}(-s)^k\right),$$

as in (2.1). Note that taking expectations on both sides of (3.7) we easily obtain $\mathbb{E}R = \rho_1 = 1$. This follows from the independence of $N(X)$ and the R_j's and the fact that $\mathbb{E}N(X) = \mathbb{E}X = \xi_1 = d$.

The next lemma holds (the proof is provided in Section 7).

Lemma 4.1. *For $n \geq 1$, the following are equivalent:*

(i) $\xi_n < \infty$,

(ii) $\rho_n < \infty$.

Remark 4.2. Similar to Remark 2.6, we can reformulate Lemma 4.1 as

$$f_n(s) = o(s^n) \quad \text{if and only if} \quad r_n(s) = o(s^n).$$

Remark 4.3. Note that the stochastic inequality $R \overset{d}{>} (1-c)\left(\frac{c}{d}N(X)+1\right)$ implies that the tail of the PageRank is at least as heavy as the tail of the In-Degree.

To establish the main result we only need to make one technical observation, which is proved in Section 7.

Corollary 4.4. *If $\xi_n < \infty$, then the following holds:*

$$r_n(s) - dr_n\left(\frac{c}{d}s\right) = f_n(t) + O\left(t^{n+1}\right),$$

where

$$t = 1 - r\left(\frac{c}{d}s\right).$$

Now we are ready to explain the similarity between In-Degree and PageRank distributions. The next theorem formalizes this main statement.

Theorem 4.5. *The following are equivalent:*

(i) $1 - F_{N(X)}(x) \sim x^{-\alpha}L(x) \quad as \quad x \to \infty,$

(ii) $1 - F_R(x) \sim \dfrac{c^\alpha}{d^\alpha - c^\alpha d}x^{-\alpha}L(x) \quad as \quad x \to \infty.$

Proof. $(i) \to (ii)$. From (i) and Theorem 3.4 it follows that

$$1 - F_X(x) \sim x^{-\alpha}L(x) \quad as \quad x \to \infty. \tag{4.2}$$

Theorem 2.7 also implies that (4.2) is equivalent to $f_n(t) \sim (-1)^n\Gamma(1 - \alpha)t^\alpha L\left(\frac{1}{t}\right)$, where n is the greatest integer less than α, and $t(s) \sim (c/d)s$ as $s \to 0$, since $r(s) = 1 - s + o(s)$. Then, by Corollary 4.4 and Lemma 2.3, we obtain

$$r_n(s) - dr_n\left(\frac{c}{d}s\right) \sim (-1)^n\Gamma(1-\alpha)\left(\frac{c}{d}\right)^\alpha s^\alpha L\left(\frac{1}{s}\right) \quad as\ s \to 0.$$

Then for every $k \geq 0$, as $s \to 0$, we have

$$r_n\left(\left(\frac{c}{d}\right)^k s\right) - dr_n\left(\left(\frac{c}{d}\right)^{k+1} s\right) \sim (-1)^n\Gamma(1-\alpha)\left(\frac{c}{d}\right)^\alpha\left(\frac{c}{d}\right)^{\alpha k} s^\alpha L\left(\frac{1}{\left(\frac{c}{d}\right)^k s}\right)$$

$$\sim (-1)^n\Gamma(1-\alpha)\left(\frac{c}{d}\right)^\alpha\left(\frac{c}{d}\right)^{\alpha k} s^\alpha L\left(\frac{1}{s}\right).$$

Using the infinite-sum representation for $r_n(s)$ (see Equation (7.5) in the proof of Lemma 4.1 in Section 7),

$$r_n(s) = \sum_{k=0}^{\infty} d^k\left(r_n\left(\left(\frac{c}{d}\right)^k s\right) - dr_n\left(\left(\frac{c}{d}\right)^{k+1} s\right)\right),$$

we directly obtain

$$r_n(s) \sim (-1)^n \Gamma(1-\alpha) \frac{d^\alpha}{d^\alpha - c^\alpha d} \left(\frac{c}{d}\right)^\alpha s^\alpha L\left(\frac{1}{s}\right) \quad \text{as } s \to 0.$$

We again apply Theorem 2.7, which leads to (ii).

$(ii) \to (i)$. The proof follows easily from (ii) and Corollary 4.4. $\qquad\square$

Thus, we have shown that the asymptotic behaviors of PageRank and In-Degree differ only by the multiplicative constant $\frac{c^\alpha}{d^\alpha - c^\alpha d}$, while the power law exponent remains the same. In the next section we will experimentally verify this result.

5. Numerical Results

5.1. Power Law Identification

The identification and measuring of power law behavior is not always simple. In this section we provide a brief overview of techniques that we used to plot and identify numerically power law distributions.

The standard strategy is to plot a histogram of a quantity on logarithmic scales to obtain a straight line, which is a typical feature of the power law. This technique is often not efficient, however. In [Newman 05], the author clearly illustrated that, even for generated random numbers with a known distribution, the noise in the tail region has a strong influence on the estimation of the power law parameters. Newman suggested plotting the fraction of measurements that are not less than a given value, i.e., the complementary cumulative distribution function $1 - F(x) = \mathbb{P}(X > x)$, rather than the histogram. The advantage is obtaining a less noisy plot. In addition, this idea is consistent with the analysis in Section 4, which was based on complementary cumulative distribution functions. We note that if the distribution of X follows a power law with exponent α so that $1 - F(x) \sim C x^{-\alpha}$, $x \to \infty$, where C is some constant, then the corresponding histogram has an exponent $\alpha + 1$. Thus, the plot of $1 - F(x)$ on logarithmic scales has a lesser slope than the plot of the histogram.

Computing the correct slope from the observed data is also not trivial. Goldstein et al., and later Newman, have proposed using a maximum likelihood estimator, which provides a more robust estimation of the power law exponent than the standard least-squares fit method [Goldstein et al. 04, Newman 05]. Thus, we compute the exponent α using the next formula from [Newman 05]:

$$\alpha = 1 + N \left(\sum_{i=1}^{N} \ln \frac{x_i}{x_{\min}}\right)^{-1}. \tag{5.1}$$

Here the quantities x_i, $i = 1, \ldots, N$, are the measured values of X, and $x_{\min}$ usually corresponds to the least value of X for which the power law behavior is assumed to hold.

In the next sections we will present our experiments on real Web Data and on a graph that represents a well-known mathematical model of the Web (Growing Networks). In both cases, for each value x, we plot in log-log scale the fraction of measurements that are not less than x.

5.2. Web Data

To confirm our results on asymptotic similarity between PageRank and In-Degree distributions, we performed experiments on the public data of the Stanford Web from [Kamvar 06]. We calculated all PageRank values for a Web graph with $281,903$ nodes (pages) and ~ 2.3 million edges (links) using the standard power method (see, e.g., [Langville and Meyer 03]). For this data set, the average Out-Degree, and hence average In-Degree, is 8.2. Since the number of dangling nodes is negligibly small in this data set, we do not take them into account.

In Figure 1, we show the log-log plots for In-Degree and PageRank of the Stanford Web Data for different values of the damping factor ($c = 0.1$, 0.5 and 0.9). Clearly, these empirical values of In-Degree and PageRank lead to parallel lines for all values of the damping factor, provided that the PageRank values are reasonably large. It was observed in [Becchetti and Castillo 06] that, in general, PageRank depends on the damping factor but the PageRank of the top 10% of pages obeys a power law with the same exponent as the In-Degree, independent of the damping factor. This is in perfect agreement with our experimental results and the mathematical model, which is focused on the tail behavior of the PageRank distribution.

The calculations based on the maximum likelihood method (5.1) yield a slope -1.1, which verifies that In-Degree and PageRank have power laws with the same exponent $\alpha = 1.1$ (this corresponds to the well-known value 2.1 for the histogram). More precisely, in Figure 1 we fitted the lines $y = -1.1x + 0.08$, $y = -1.1x - 0.87$, $y = -1.1x - 1.27$, and $y = -1.1x - 2.07$ to the plots of In-Degree and PageRank (with $c = 0.9$, $c = 0.5$, and $c = 0.1$, respectively).

We also investigated whether Theorem 4.5 correctly predicts the multiplicative constant

$$y(c) = \frac{c^\alpha}{d^\alpha - c^\alpha d}.$$

In Figure 2, we plot $\log_{10}(y(c))$ and compare it to the observed differences between the logarithms of the complementary cumulative distribution functions of
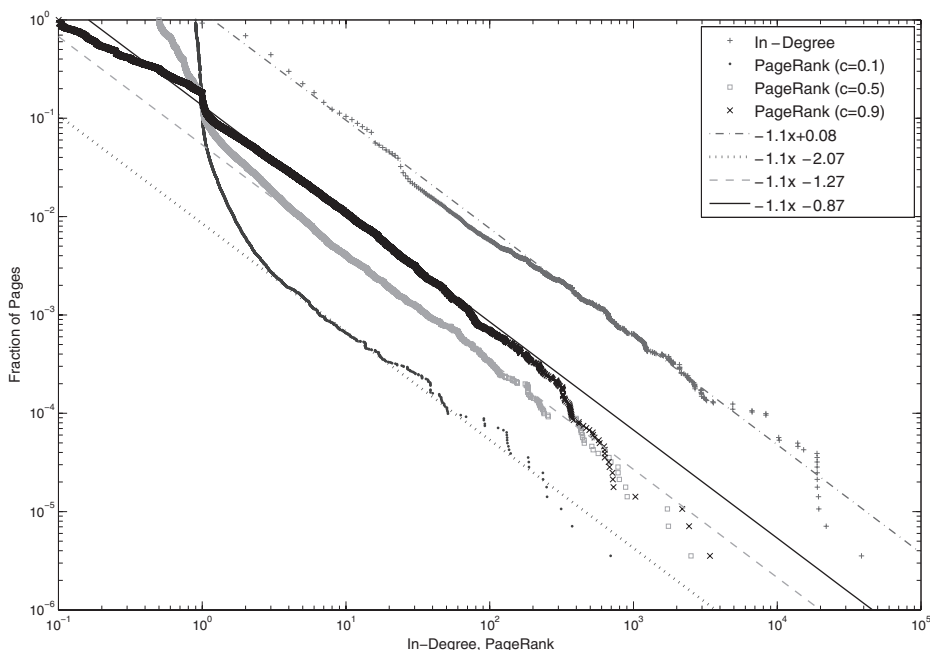
Figure 1. Plots for the Web data: fraction of pages with In-Degree/PageRank not less than x versus x in log-log scale, and the fitted lines.
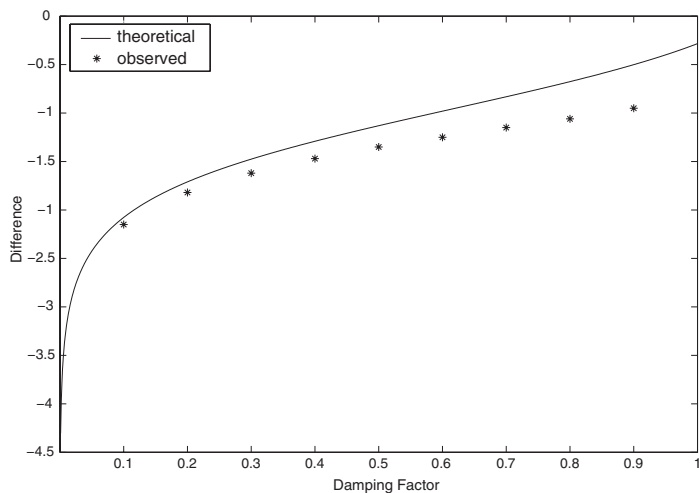

Figure 2. The theoretical and observed differences between asymptotics of In-Degree and PageRank.

PageRank and In-Degree, for different values of the damping factor. Again, $d = 8.2$ because that is the average In/Out-Degree in the data set. As can be seen, the theoretical and observed values are quite close. For typical values of c between 0.8 and 0.9, for example, the difference is 0.41, resulting in a factor $y(c)$ that is only a factor of 2.57 greater than in the observed data. Thus, our model not only allows us to prove the similarity in the power law behavior but also gives a good approximation for the difference between the two distributions.

The discrepancy between the predicted and observed values of the multiplicative factor suggests that our model does not capture PageRank behavior to the full extent. For instance, the assumption of the independence of PageRank values of pages that have a common neighbor may be too strong. We believe, however, that the achieved precision, especially for lesser values of c, is quite good for our relatively simple stochastic model.

5.3.   Growing Networks

Growing Networks, introduced in [Barabási and Albert 99], now represent a large class of models that are commonly accepted as a possible scenario of Web growth. In particular, these models provide a mathematical explanation for the power law behavior of In-Degree [Bollobas et al. 01]. The recent studies [Avrachenkov and Lebedev 06] and [Fortunato and Flammini 06] have addressed for the first time the PageRank distribution in Growing Networks.

Growing Network models are characterized by preferential attachment. This entails that a newly created node connects to the existing nodes with probabilities that are proportional to the current In-Degrees of the existing nodes. We simulated a slightly modified version of this model, where a new link points to a randomly chosen page with probability q, and with probability $(1 - q)$ the preferential attachment selection rule is used. This allows us to fine-tune the exponent of the resulting power law [Newman 05].

We simulate our Growing Network using MATLAB. We start with d nodes and at each step add a new node that links to d existing nodes. To ensure the same number of outgoing links for all pages, at the end of the simulation we link the first d nodes to randomly chosen pages. In the example presented below we set $q = 0.2$ and obtain a network of $50,000$ nodes with Out-Degree $d = 8$.

In Figure 3 we present the numerical data for In-Degree and PageRank in the Growing Network. Clearly, the Web data from Section 5.2 shows a much better agreement with our model than the data generated by the preferential attachment algorithm. In the next section we briefly compare recent results on PageRank in Growing Networks to our present study, and we indicate possible directions for further research.
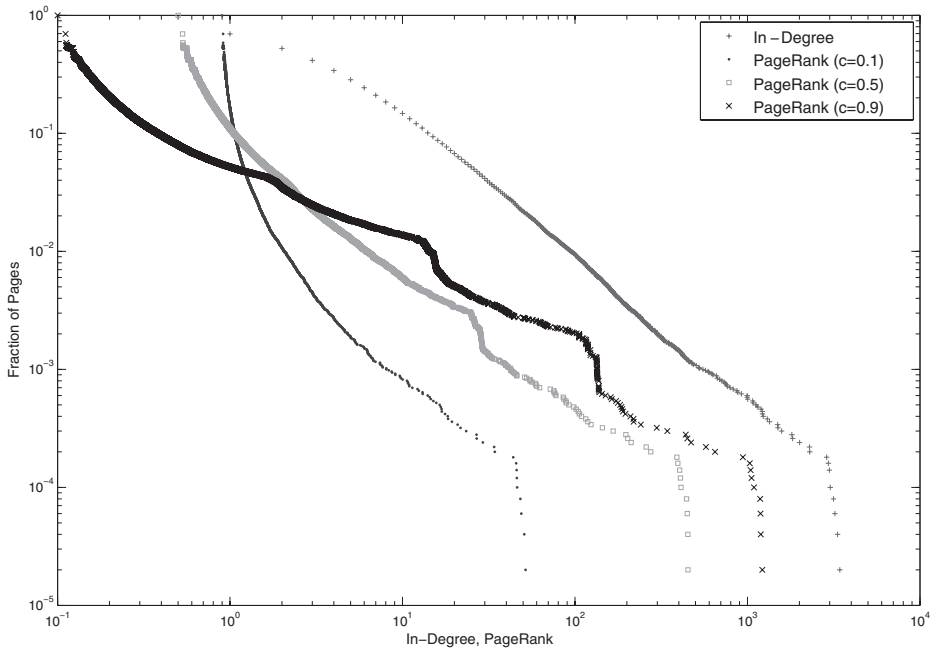
Figure 3. Plots for the Growing Network model: fraction of pages with In-Degree/PageRank not less than x versus x in log-log scale.

6. Discussion

Our model and analysis resulted in the conclusion that PageRank and In-Degree should follow power laws with the same exponent. Growing Network models may provide an alternative explanation [Avrachenkov and Lebedev 06, Fortunato and Flammini 06]. For instance, in the recent paper [Avrachenkov and Lebedev 06], it was shown that the expected PageRank in Growing Networks follows a power law with an exponent that does depend on the damping factor but equals ≈ 1.08 for $c = 0.85$. Thus, the model in [Avrachenkov and Lebedev 06] can also be used to explain the tail behavior of PageRank, but it leads to a slightly different result than our model, because in our case, the power law exponent of PageRank does not depend on the damping factor. The reason for this could be that we focus only on asymptotics, whereas [Avrachenkov and Lebedev 06] employs a mean-field approximation. Indeed, experiments show that the shape of the PageRank distribution depends on the damping factor and thus may affect the average values, whereas the tail behavior remains the same for all values of c.

We emphasize that, compared to [Avrachenkov and Lebedev 06, Fortunato and Flammini 06], our model provides a completely different approach for modeling the relation between In-Degree and PageRank. Specifically, we do not make any assumption on the underlying Web graph, whereas [Avrachenkov and Lebedev 06, Fortunato and Flammini 06] choose the preferential attachment structure, thus exploiting the fact that the graph model correctly captures the In-Degree distribution. We believe that both approaches are useful for research on the PageRank distribution.

One of the important innovations in the present work is the analogy between the PageRank equation and the equation for the busy period that enables us to apply the techniques from [De Meyer and Teugels 80]. In fact, queueing systems with heavy tails and, in particular, the busy period problem allow for a more sophisticated probabilistic analysis (see, e.g., [Zwart 01]). It is interesting to apply these advanced methods to the problems related to the Web and PageRank.

Our model definitely lacks the dependencies between the PageRank values of pages sharing a common neighbor. Such dependencies must be present in the Web due to the high clustering of the Web graph [Newman 05] (roughly speaking, *clustering* means that, with high probability, two neighbors of the same page are connected to each other). Thus, in our further research we will try to include some sort of dependencies along with random Out-Degrees, as well as considering personalization or topic sensitivity [Haveliwala 03]. The impact of these factors on the PageRank distribution could be determined by extending and generalizing the proposed analytical model.

7. Appendix

Proof of Lemma 3.1. $(i) \rightarrow (ii)$. By Lemma 2.5, we know that $\xi_n < \infty$ if and only if $f_n(y) = o\left(y^n\right)$. Let us consider

$$y(s) = 1 - e^{-s} = \sum_{k=1}^{n+1} (-1)^{k+1} \frac{s^k}{k!} + o(s^{n+1}).$$

Then we can actually find $y^i(s)$:

$$y^i(s) = \sum_{j=i}^{n+i} \mu_{i,j} s^j + o\left(s^{n+i}\right)$$

for $i \geq 1$ and appropriate constants $\mu_{i,j}, j = i, \ldots, n + i$. Thus, we easily obtain

$$
\begin{aligned}
f_n(y(s)) &= (-1)^{n+1} \left(f(y(s)) - \sum_{i=0}^{n} \frac{\xi_i}{i!} (-y(s))^i \right) \\
&= (-1)^{n+1} \left(\phi(s) - 1 - \sum_{i=1}^{n} \frac{\xi_i}{i!} (-1)^i \left(\sum_{j=i}^{n+i} \mu_{i,j} s^j + o\left(s^{n+i}\right) \right) \right) \\
&= (-1)^{n+1} \left(\phi(s) - \sum_{i=0}^{n} \frac{\hat{\nu}_i}{i!} (-s)^i + O\left(s^{n+1}\right) \right)
\end{aligned}
$$

for some finite constants $\hat{\nu}_0 = 1$ and $\hat{\nu}_1, \ldots \hat{\nu}_n$, which can be expressed in terms of $\xi_1, \ldots, \xi_n$. Thus, we find

$$
\phi(s) = \sum_{i=0}^{n} \frac{\hat{\nu}_i}{i!} (-s)^i + (-1)^{n+1} f_n(y(s)) + O(s^{n+1}) = \sum_{i=0}^{n} \frac{\hat{\nu}_i}{i!} (-s)^i + o(s^n),
$$

since $y(s) = s + o(s)$. By uniqueness of the power series expansion we have $\nu_i = \hat{\nu}_i$, $i = 0 \ldots n - 1$, and by Lemma 2.5, we have $\nu_n = \hat{\nu}_n < \infty$.

$(ii) \rightarrow (i)$. Similar to the first part of the proof, using $s(y) = -\ln(1 - y)$. $\square$

Proof of Lemma 4.1. $(i) \rightarrow (ii)$. We use induction, starting from $n = 1$, for which both (i) and (ii) are valid. Assume that for $k = 1, 2, \ldots, n - 1$, it has been shown that (i)$\rightarrow$(ii). Denote

$$
g(s) = e^{-s(1-c)},
$$

and

$$
t(s) = 1 - r\left(\frac{c}{d} s\right).
$$

Then we can write (4.1) as

$$
r(s) = f(t)g(s). \tag{7.1}
$$

We know from (i) that

$$
\begin{aligned}
f(t) &= 1 - dt + \sum_{k=2}^{n} \frac{\xi_k (-t)^k}{k!} + o(t^n) \\
&= 1 - d\left(1 - r\left(\frac{c}{d} s\right)\right) + \sum_{k=2}^{n} \frac{\xi_k (-t)^k}{k!} + o(t^n).
\end{aligned}
$$

Thus, from (7.1) we obtain

$$r(s) - dg(s)r\left(\frac{c}{d}s\right) = \left(1 - d + \sum_{k=2}^{n} \frac{\xi_k(-t)^k}{k!} + o(t^n)\right)g(s). \qquad (7.2)$$

It follows from the induction hypothesis for $n-1$, however, that

$$r(s) = 1 - s + \sum_{k=2}^{n-1} \frac{\rho_k}{k!}(-s)^k + o(s^{n-1}),$$

so we can present $t(s)$ as a sum

$$t(s) = -\sum_{k=1}^{n-1} \frac{\rho_k}{k!}\left(\frac{c}{d}\right)^k (-s)^k + o(s^{n-1}).$$

Using this, we can actually find $t^k(s)$:

$$t^k(s) = \sum_{i=k}^{n+k-2} \beta_{k,i}s^i + o(s^{n+k-2}), \qquad (7.3)$$

for $k \geq 1$ and appropriate constants $\beta_{k,i}, i = k, \ldots, n+k-2$. We obtain by (7.2) and (7.3)

$$r(s) - dg(s)r\left(\frac{c}{d}s\right) = \left(\sum_{i=0}^{n} \gamma_i s^i + o(s^n)\right)g(s)$$

for appropriate constants $\gamma_0, \ldots, \gamma_n$. Using the expansion of $g(s)$, it is not difficult to show that, for appropriate constants $\eta_0, \ldots, \eta_n$, we also have

$$r(s) - dr\left(\frac{c}{d}s\right) = \sum_{i=0}^{n} \eta_i s^i + o(s^n).$$

Because of the uniqueness of the series expansion, we can rewrite the last equation as

$$r_{n-1}(s) - dr_{n-1}\left(\frac{c}{d}s\right) - \eta_n s^n = o(s^n). \qquad (7.4)$$

We will now show that this implies (ii), to which end we consider the partial sums

$$\begin{aligned}
r_{n-1}^m(s) &= \sum_{k=0}^{m} d^k \left(r_{n-1}\left(\left(\frac{c}{d}\right)^k s\right) - dr_{n-1}\left(\left(\frac{c}{d}\right)^{k+1} s\right)\right) \\
&= r_{n-1}(s) - d^{m+1} r_{n-1}\left(\left(\frac{c}{d}\right)^{m+1} s\right).
\end{aligned}$$

Taking the limit as $m \to \infty$, we have for the last term that

$$\lim_{m \to \infty} d^{m+1} r_{n-1} \left(\left(\frac{c}{d}\right)^{m+1} s \right)$$

$$= \lim_{m \to \infty} \frac{r_{n-1} \left(\left(\frac{c}{d}\right)^{m+1} s \right)}{\left(\left(\frac{c}{d}\right)^{m+1} s \right)^{n-1}} \lim_{m \to \infty} \left(\frac{c}{d}\right)^{(m+1)(n-2)} s^{n-1} c^{m+1} = 0,$$

where we used the induction hypothesis $r_{n-1}(s) = o(s^{n-1})$ together with $n \geq 2$, $0 < c < 1$ and $d > 1$. It follows that we can express $r_{n-1}(s)$ as an infinite sum,

$$r_{n-1}(s) = \sum_{k=0}^{\infty} d^k \left(r_{n-1} \left(\left(\frac{c}{d}\right)^k s \right) - d r_{n-1} \left(\left(\frac{c}{d}\right)^{k+1} s \right) \right), \qquad (7.5)$$

where we can apply (7.4) to each of the terms. Further, by definition of $o(s^n)$, for all $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon)$ such that $\left| r_{n-1}(s) - d r_{n-1} \left(\frac{c}{d} s \right) - \eta_n s^n \right| < \varepsilon s^n$ whenever $0 < s \leq \delta$. Moreover, for this ε and δ, and $0 < s \leq \delta$, we also have

$$\left| r_{n-1}(s) - \frac{d^{n-1}}{d^{n-1} - c^n} \eta_n s^n \right|$$

$$= \left| \sum_{k=0}^{\infty} d^k \left(r_{n-1} \left(\left(\frac{c}{d}\right)^k s \right) - d r_{n-1} \left(\left(\frac{c}{d}\right)^{k+1} s \right) - \left(\frac{c}{d}\right)^{kn} \eta_n s^n \right) \right|$$

$$\leq \sum_{k=0}^{\infty} \left| d^k \left(r_{n-1} \left(\left(\frac{c}{d}\right)^k s \right) - d r_{n-1} \left(\left(\frac{c}{d}\right)^{k+1} s \right) - \left(\frac{c}{d}\right)^{kn} \eta_n s^n \right) \right|$$

$$< \sum_{k=0}^{\infty} \varepsilon d^k \left(\frac{c}{d}\right)^{kn} s^n = \frac{d^{n-1}}{d^{n-1} - c^n} \varepsilon s^n. \qquad (7.6)$$

Here the second inequality holds because $0 < \left(\frac{c}{d}\right)^k s \leq \delta$ for every $k \geq 0$. Since, for every $\varepsilon_0 > 0$ there exists δ_0 such that

$$\left| r_{n-1}(s) - d r_{n-1} \left(\frac{c}{d} s\right) - \eta_n s^n \right| < \frac{d^{n-1} - c^n}{d^{n-1}} \varepsilon_0 s^n$$

for $0 < s \leq \delta_0$, then, according to (7.6), we have $\left| r_{n-1}(s) - \frac{d^{n-1}}{d^{n-1} - c^n} \eta_n s^n \right| < \varepsilon_0 s^n$ whenever $0 < s \leq \delta_0$, by which we have shown that

$$r_{n-1}(s) - \frac{d^{n-1}}{d^{n-1} - c^n} \eta_n s^n = o(s^n).$$

Taking $\rho_n = -\frac{d^{n-1}}{d^{n-1} - c^n} \eta_n n!$, from Lemma 2.5 and the last equation, we conclude that ρ_n is the nth moment of R and is finite.

$(ii) \rightarrow (i)$. Assume that there exists a nonnegative random variable R satisfying (3.7). Then, obviously, $R \geq 1 - c$. Moreover, (3.7) also implies that R is stochastically greater than $(1 - c)\left(\frac{c}{d}N(X) + 1\right)$. Hence, the existence of the nth moment of R ensures the existence of the nth moment of $N(X)$, which in turn, by Lemma 3.1, ensures the existence of the nth moment of X. $\qquad \square$

Proof of Corollary 4.4. The proof follows from the first part of the proof of Lemma 4.1. By definitions of $r_n(s)$, $f_n(t)$, $t(s)$ and Lemma 4.1, it follows from (7.1) that, for fixed n,

$$(-1)^{n+1} r_n(s) + \sum_{k=0}^{n} \frac{\rho_k}{k!}(-s)^k$$

$$= \left((-1)^{n+1} f_n(t) + 1 - dt + \sum_{k=2}^{n} \frac{\xi_k(-t)^k}{k!}\right) g(s)$$

$$= \left((-1)^{n+1} f_n(t) + 1 - d + d\left((-1)^{n+1} r_n\left(\frac{c}{d}s\right) + \sum_{k=0}^{n} \frac{\rho_k}{k!}\left(\frac{c}{d}\right)^k (-s)^k\right)\right.$$

$$\left. + \sum_{k=2}^{n} \frac{\xi_k(-t)^k}{k!}\right)\left(\sum_{k=0}^{n+1} \frac{(1-c)^k}{k!}(-s)^k + o\left(s^{n+1}\right)\right). \tag{7.7}$$

Because $r_n(s) = o(s^n)$, we can extend (7.3) for $k \geq 1$ and appropriate constants $\beta_{k,i}, i = k, ..., k + n - 1$,

$$t^k(s) = \sum_{i=k}^{n+k-1} \beta_{k,i} s^i + o(s^{n+k-1}),$$

and rewrite (7.7) as

$$(-1)^{n+1} r_n(s) + \sum_{k=0}^{n} \frac{\rho_k}{k!}(-s)^k$$

$$= (-1)^{n+1} f_n(t) + d(-1)^{n+1} r_n\left(\frac{c}{d}s\right) + \sum_{k=0}^{n+1} \tau_k s^k + o(s^{n+1}),$$

where $\tau_0, \ldots, \tau_{n+1}$ are corresponding constants. Now, due to the uniqueness of the series expansion, we can reduce the above formula to

$$r_n(s) = f_n(t) + dr_n\left(\frac{c}{d}s\right) + (-1)^{n+1}\tau_{n+1}s^{n+1} + o(s^{n+1}).$$

Then, since $t(s) \sim (c/d)s$ as $s \rightarrow 0$, we get

$$r_n(s) - dr_n\left(\frac{c}{d}s\right) = f_n(t) + O(t^{n+1}). \qquad \square$$

References

[Aldous and Bandyopadhyay 05] D. J. Aldous and A. Bandyopadhyay. "A Survey of Max-Type Recursive Distributional Equations." *Ann. Appl. Probab.* 15 (2005), 1047–1110.

[Avrachenkov and Lebedev 06] K. Avrachenkov and D. Lebedev. "PageRank of Scale-Free Growing Networks." *Internet Mathematics* 3:2 (2006), 207–231.

[Avrachenkov and Litvak 06] K. Avrachenkov and N. Litvak. "The Effect of New Links on Google PageRank." *Stoch. Models* 22:2 (2006), 319–331.

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

[Becchetti and Castillo 06] L. Becchetti and C. Castillo. "The Distribution of PageRank Follows a Power-Law Only for Particular Values of the Damping Factor." In *Proceedings of the 15th International Conference on World Wide Web*, pp. 941–942. New York: ACM Press, 2006.

[Berkhin 05] P. Berkhin. "A Survey on PageRank Computing." *Internet Mathematics* 2 (2005), 73–120.

[Bingham and Doney 74] N. H. Bingham and R. A. Doney. "Asymptotic Properties of Supercritical Branching Processes. I. The Galton-Watson Process." *Advances in Appl. Probability* 6 (1974), 711–731.

[Bingham et al. 89] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation.* Cambridge, UK: Cambridge University Press, 1989.

[Bollobas 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 (2001), 279–290.

[Brin and Page 98] S. Brin and L. Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Comput. Networks ISDN Systems* 33 (1998), 107–117.

[De Meyer and Teugels 80] A. De Meyer and J. L. Teugels. "On the Asymptotic Behaviour of the Distributions of the Busy Period and Service Time in M/G/1." *J. App. Probab.* 17 (1980), 802–813.

[Donato et al. 04] D. Donato, L. Laura, S. Leonardi, and S. Millozi." "Large Scale Properties of the Webgraph. *Eur. Phys. J.* 38 (2004), 239–243.

[Eiron et al. 04] N. Eiron, K. S. McCurley, and J. A. Tomlin. "Ranking the Web frontier." In *Proceedings of the 13th International Conference on World Wide Web*, pp. 309–318. New York: ACM Press, 2004.

[Fortunato and Flammini 06] S. Fortunato and A. Flammini. "Random Walks on Directed Networks: The Case of PageRank." Technical Report 0604203, arXiv/physics, 2006.

[Fortunato et al. 08] S. Fortunato, M. Boguna, A. Flammini, and F.Menczer. "Approximating PageRank from In-Degree." In *Algorithms and Models for the Web Graph: Fourth International Workshop, WAW 2006, Banff, Canada, November 30–December 1, 2006, Revised Papers*, Lecture Notes in Computer Science 4936, pp. 59–71. Berlin: Springer, 2008.

[Goldstein et al. 04] M. L. Goldstein, S. A. Morris, and G. G. Yen." "Problems with Fitting to the Power-Law Distribution. *Eur. Phys. J.* 41 (2004), 255–258.

[Haveliwala 03] T. H. Haveliwala. "Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search." *IEEE TKDE*, 15:4 (2003), 784–796.

[Kamvar 06] S. Kamvar. "Sep Kamvar: Research and Portfolio." http://www.stanford.edu/~sdkamvar/research.html, accessed in March 2006.

[Langville and Meyer 03] A. N. Langville and C. D. Meyer. "Deeper Inside PageRank." *Internet Mathematics* 1 (2003), 335–380.

[Newman 05] M. E. J. Newman. "Power Laws, Pareto Distributions and Zipf's law." *Contemporary Physics* 46 (2005), 323–351.

[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*, Lecture Notes in Computer Science 2387, 330–339. Berlin: Springer, 2002.

[Robert 03] P. Robert. *Stochastic Networks and Queues*. New York: Springer, 2003.

[Zwart 01] A. P. Zwart. *Queueing Systems with Heavy Tails*. PhD thesis, Eindhoven University of Technology, 2001.

N. Litvak, University of Twente, Department of Applied Mathematics, P.O. Box 217, Enschede 7500 AE, Netherlands (n.litvak@ewi.utwente.nl)

W. R. W. Scheinhardt, University of Twente, Department of Applied Mathematics, P.O. Box 217, Enschede 7500 AE, Netherlands (w.r.w.scheinhardt@ewi.utwente.nl)

Y. Volkovich, University of Twente, Department of Applied Mathematics, P.O. Box 217, Enschede 7500 AE, Netherlands (y.volkovich@ewi.utwente.nl)