Probabilistic Relation between In-Degree and PageRank

Nelly Litvak, Werner R.W. Scheinhardt, and Yana Volkovich

University of Twente, Dept. of Applied Mathematics,
P.O. Box 217, 7500AE Enschede, The Netherlands
{n.litvak,w.r.w.scheinhardt,y.volkovich}@ewi.utwente.nl

Abstract. This paper presents a novel stochastic model that explains the relation between power laws of In-Degree and PageRank. PageRank is a popularity measure designed by Google to rank Web pages. We model the relation between PageRank and In-Degree through a stochastic equation, which is inspired by the original definition of PageRank. Using the theory of regular variation and Tauberian theorems, we prove 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 Web data.

Categories and Subject Descriptors
H.3.3:[**Information Storage and Retrieval**]: Information Search and Retrieval– *Retrieval models*; G.3:[**Mathematics of Computing**]: Probability and statistics – *Stochastic processes, Distribution functions*

General Terms
Theory, Verification, Experimentation, Algorithms

Keywords: PageRank, In-Degree, Power law, Regular variation, Stochastic equation, Web measurement.

1 Introduction

We study the relation between the probability distributions of the PageRank and the In-Degree of a randomly selected Web page. In this paper we present the mathematical model and main results while more detailed discussion and proofs can be found in the extended version [1]. The notion of *PageRank* was introduced by Google in order to numerically characterize the popularity of Web pages. The original description of PageRank presented in [2] is as follows:

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

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 simply the number of incoming hyperlinks to that page. From

W. Aiello et al. (Eds.): WAW 2006, LNCS 4936, pp. 72–83, 2008.

equation (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 simply mean the fraction of pages $\mathbb{P}(PR > x)$ having PageRank greater than x, where x is large. 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 equivalent, and thus, we can approximate $\mathbb{P}(PR > x)$ by $p(x)$ for large enough x.

Pandurangan et al. [3] observed that the tails of PageRank and In-Degree distributions for Web data seem to follow power laws with the same exponent. Recent extensive experiments by Donato et al. [4] and Fortunato et al. [5] confirmed this phenomenon. Becchetti and Castillo [6] investigated the influence of the damping factor c on the power law behavior of PageRank. 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.

Obviously, equation (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 [7,8] and proved very successful for designing efficient computational methods, seem to be insufficient for modelling 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. Our approach is inspired by techniques from applied probability and stochastic operations research. The relation between PageRank and In-Degree is modelled through a distributional identity, which is analogous to the equation for the busy period in the M/G/1 queue (see e.g. [9]). Further, we analyze our model using the approach employed in [10] for studying the tail behavior of the busy period in case the service times are regularly varying random variables. This fits in our research because regular variation is in fact a formalization of the power law, and it has been widely used in queueing theory to model self-similarity, long-range dependence and heavy tails [11]. Thus, we use the notion of regular variation to model the power law distribution of In-Degree.

To obtain the tail behavior of PageRank in our model, we use Laplace-Stieltjes transforms and apply Tauberian theorems presented in the paper by Bingham and Doney [12], see also Theorem 8.1.6 in [13]. 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 to explicitly derive the multiplicative constant that quantifies the difference between PageRank and In-Degree tail behavior. Our analytical results show a good agreement with Web data.

2 Preliminaries

This section describes important properties of regularly varying random variables. We follow definitions and notations by Bingham and Doney [12], Meyer and Teugels [10], and Zwart [11]. More comprehensive details can be found in [13].

Definition 1. *A function L is said to be* slowly varying *if for every $t > 0$,*

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

Definition 2. *A random variable X is said to be* regularly varying *with index α if its distribution is such that*

$$\mathbb{P}(X > x) \sim x^{-\alpha} L(x) \quad as \quad x \to \infty,$$

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 $\xi_n = \mathbb{E}X^n$ be the nth moment of X, where $n \in \mathbb{N}$. The successive moments of X can be obtained by expanding f in a series at $s = 0$. More precisely, we have the following.

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

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

If $\xi_n < \infty$ then we introduce the notation

$$f_n(s) = (-1)^{n+1} \left(f(s) - \sum_{i=0}^{n} \frac{\xi_i}{i!}(-s)^i \right). \tag{2}$$

Note 1. It follows from Lemma 1 that $\mathbb{E}X^n < \infty$ 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 1. *(Tauberian Theorem) If $n \in \mathbb{N}$, $\xi_n < \infty$, $\alpha = n + \beta$, $\beta \in (0, 1)$, 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) $\mathbb{P}(X > x) \sim x^{-\alpha} L(x)$ *as* $x \to \infty$.

Here and in the remainder of the paper we use the letter α to denote the index of the tail probability $\mathbb{P}(X > x)$.

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 (1), and is analytically tractable, thus enabling us to obtain the asymptotic behavior of PageRank. As will become clear, we make several rather strong 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) almost unchanged but we transform it into a stochastic equation. Let R be the PageRank of a randomly chosen page. We treat R simply as a random variable whose distribution we want to determine. Further, we view the In-Degree of a random page as a random variable N, which follows a power law. The model for N will be specified in Section 3.2 below. In this work, we assume that the number of outgoing links (*Out-Degree*) $d \geq 1$ is the same for each page. This assumption is obviously not realistic; in particular it ignores the presence of 'hubs' (pages with extremely high Out-Degree) and 'dangling nodes' (pages with Out-Degree zero). The idea behind this rigid simplification is that we want to focus on the influence of the In-Degree, without considering other factors. Besides, it is a common belief that Out-Degrees do not affect the PageRank distribution, and it is also well-known (see e.g. [14]) that dangling nodes alter the PageRank vector only by a multiplicative constant. We note however that the proposed stochastic model allows for extensions. For instance, in the upcoming paper [15], we account for dangling nodes and allow for an arbitrary Out-Degree distribution.

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

$$R \stackrel{d}{=} c \sum_{j=1}^{N} \frac{1}{d} R_j + (1 - c). \tag{3}$$

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 not true in general. However, it is also not the case 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 PageRank as a random variable which solves a certain stochastic equation. We believe, 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. In fact, this point of view is in line with Pandurangan et al. [3] and other authors who consistently present log-log *histograms* of PageRank.

One of the nice features of the stochastic equation (3) is that it has the same form as the original formula (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).

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 to employ the theory of regular variation. We model the In-Degree of a randomly chosen page as a nonnegative, integer, regularly varying random variable, which is distributed as $N(X)$, where X is regularly varying with index α:

$$\mathbb{P}(X > x) \sim x^{-\alpha} L(x) \;\text{ as } x \to \infty,$$

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 at the same time ensure that the In-Degree is integer. It is intuitively clear that $N(X)$ is asymptotically equivalent to X, that is, $N(X)$ and X follow the same power law. Specifically, we have

$$\mathbb{P}(N(X) > x) \sim \mathbb{P}(X > x) \;\text{ as } x \to \infty. \tag{4}$$

For the proof of (4) using the Tauberian theorem (Theorem 1) see e.g. [1].

3.3 The Main Stochastic Equation

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

$$R \stackrel{d}{=} c \sum_{j=1}^{N(X)} \frac{1}{d} R_j + (1 - c), \tag{5}$$

where $c \in (0, 1)$ is the damping factor, $d \geq 1$ 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 (5) adequately captures several important aspects of the PageRank distribution and its relation to the In-Degree distribution. Moreover, our model is completely formalized, and thus we can apply analytical methods in order to derive the tail behavior of the random variable R representing PageRank.

Linear stochastic equations like (5) have a long history. In particular, (5) is similar to the famous equation that arises in the theory of branching processes and describes many real-life phenomena, for instance, 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 [9] and other books on queueing theory for more details. Also, see Zwart [11] for an excellent detailed treatment of queues with regular variation, and specifically the busy period problem. We would like to add that our equation (5) is a special case in a rich class of stochastic recursive equations that were discussed in detail in the recent survey by Aldous and Bandyopadhyay [16].

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 down an equation for the Laplace-Stieltjes transforms of X and R and then make use of the Tauberian theorem to prove that R is regularly varying with the same index as X. Since X and $N(X)$ are asymptotically equivalent, this will give us the desired similarity in tail behavior of the PageRank R and the In-Degree $N(X)$.

Let r be the the Laplace-Stieltjes transform of R. As a result of the assumptions from Section 3, we can use (5) to express r in terms of f, the Laplace-Stieltjes transform of X, as follows:

$$\begin{aligned}
r(s) := \mathbb{E}e^{-sR} &= e^{-s(1-c)}\mathbb{E}\left[\mathbb{E}\left[\exp\left(-s\frac{c}{d}\sum_{i=1}^{N(X)} R_i\right)\middle| N(X)\right]\right] \\
&= e^{-s(1-c)}\mathbb{E}\left[\left(\mathbb{E}\left[\exp\left(-s\frac{c}{d}R_i\right)\right]\right)^{N(X)}\right] \\
&= e^{-s(1-c)}\mathbb{E}\left[\mathbb{E}\left[\left(r\left(s\,\frac{c}{d}\right)\right)^{N(X)}\middle| X\right]\right] \\
&= e^{-s(1-c)}\mathbb{E}\exp\left(-\left(1-r\left(s\,\frac{c}{d}\right)\right)X\right) = e^{-s(1-c)}f\left(1-r\left(\frac{c}{d}s\right)\right).
\end{aligned}$$

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

We start the analysis with providing the correspondence between existence of the n-th moments of X and R. We remind 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 define

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

as in (2). Note that taking expectations on both sides of (5) 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.

Lemma 2. *The following are equivalent*

(i) $\xi_n < \infty$,
(ii) $\rho_n < \infty$.

Note 2. Similar as in Note 1, we can reformulate Lemma 2 as

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

Note 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 R is at least as heavy as the tail of the In-Degree $N(X)$.

The proof of Lemma 2 is quite lengthy and is therefore omitted. The interested reader is referred to the full version of this paper, see [1]. Same applies to the proof of Corollary 1 below.

Corollary 1. *The following holds:*

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

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

Now we are ready to explain the similarity between the In-Degree and PageRank distributions. Specifically, we show that the tail probabilities $\mathbb{P}(R > x)$ and $\mathbb{P}(N(X) > x)$ for PageRank and In-Degree, respectively, approximately differ by a multiplicative constant as x grows large. The next theorem formalizes this statement.

Theorem 2. *The following are equivalent*

(i) $\mathbb{P}(N(X) > x) \sim x^{-\alpha}L(x) \quad \text{as} \quad x \to \infty$,
(ii) $\mathbb{P}(R > x) \sim \dfrac{c^\alpha}{d^\alpha - c^\alpha d} x^{-\alpha}L(x) \quad \text{as} \quad x \to \infty$.

Proof.
$(i) \to (ii)$ From (i) and (4) it follows that

$$\mathbb{P}(X > x) \sim x^{-\alpha}L(x) \quad \text{as} \quad x \to \infty. \tag{6}$$

Theorem 1 also implies that (6) is equivalent to $f_n(t) \sim (-1)^n \Gamma(1-\alpha) t^\alpha L\left(\frac{1}{t}\right)$, where $t(s) = 1 - r\left(\frac{c}{d}s\right) \sim (c/d)s$ as $s \to 0$. Hence, by Corollary 1 we obtain

$$r_n(s) - dr_n\left(\frac{c}{d}s\right) \sim (-1)^n \Gamma(1-\alpha) \left(\frac{c}{d}\right)^\alpha s^\alpha L\left(\frac{1}{s}\right) \quad \text{as } s \to 0. \tag{7}$$

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

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

Next, we write $r_n(s)$ in the form of an infinite sum as follows:

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

From the above representation we obtain

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

Now we again invoke Theorem 1, which leads to *(ii)*.
(ii) $\to$ *(i)* The proof follows easily from *(ii)* and Corollary 1.

Thus, we have shown that the asymptotic behaviors of PageRank and In-Degree differ 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

We verified our findings by computing PageRank on the public data of the Stanford Web from [17]. To identify the power law behavior, we used cumulative log-log plots, which are much less noisy than histograms.

In order to compute the slope α, we used the following maximum likelihood estimator proposed by Newman [18]:

$$\alpha = 1 + n\left(\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right)^{-1}. \tag{8}$$

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

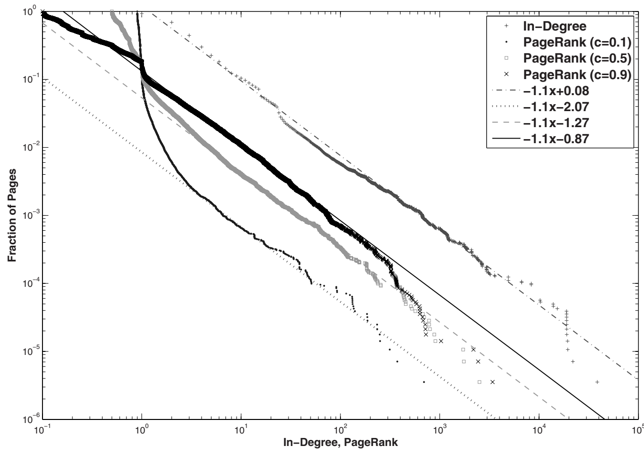

Fig. 1. Plots for the Web data. Fraction of pages with In-Degree/PageRank greater than x versus x in log-log scale, and the fitted straight lines.

There are several papers, see [3,4,5], and [6] that describe similar experiments for different domains and different number of pages, and they all confirm that PageRank and In-Degree follow power laws with the same exponent, around 1.1 for the cumulative distribution function.

We calculated all PageRank values for the Web graph with 281903 nodes (pages) and ~ 2.3 million edges (links) using the standard power method (see e.g. [8]). On this dataset, the average Out-Degree, and hence average In-Degree is 8.2. 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 constitute parallel straight lines for all values of the damping factor, provided that the PageRank values are reasonably large. It was observed in [6] 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 right tail behavior of the PageRank distribution.

The calculations based on the maximum likelihood method 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 2 correctly predicts the multiplicative constant

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

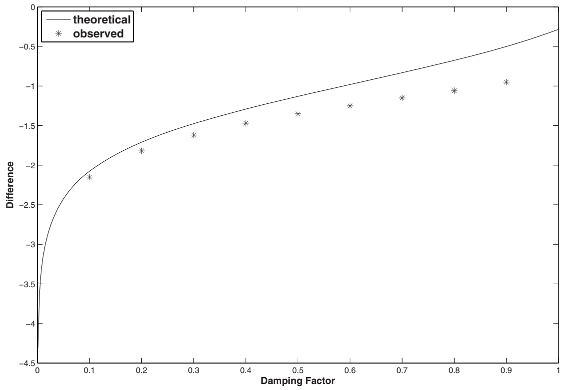

Fig. 2. The theoretical and observed differences between logarithmic asymptotics of In-Degree and PageRank

In Figure 2 we plotted $\log_{10}(y(c))$ and we compared it to the observed differences between the logarithms of the complementary cumulative distribution functions of PageRank and In-Degree, for different values of the damping factor. Obviously, in the data set, the assumption that all Out-Degrees are equal to some constant d is not satisfied. Therefore, we take $d = 8.2$, which is equal to the average In/Out-Degree in the Web data. As can be seen, the theoretical and observed values are quite close. E.g., for typical values of c between 0.8 and 0.9, the difference is 0.41, resulting in a factor $y(c)$ that is only a factor 2.57 larger than in the observed data. Thus, our model not only allows 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 constant 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 small values of c, is quite good for our relatively simple stochastic model.

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 [19,20]. For instance, Avrachenkov and Lebedev [19] showed that in Growing Networks, introduced by Barabási and Albert [21], the *expected* PageRank follows a power law with an exponent, which does depend on the damping factor but equals ≈ 1.08 for $c = 0.85$. Note that our present model suggests that the power law exponent of PageRank does *not* depend on the damping factor. We emphasize that compared to [19,20], our model provides a completely different approach for modelling the relation between In-Degree and PageRank because we do not make any assumption on the structure or growth of the underlying Web graph.

We can further exploit the analogy between the PageRank equation and the equation for the busy period in $M/G/1$ queue, since sophisticated probabilistic techniques have been developed for analyzing queueing systems with heavy tails and in particular the busy period problem (see e.g. [11]). It is interesting to apply these advanced methods to the problems related to the Web and PageRank.

Our current model lacks the dependencies between PageRank values of pages sharing a common neighbor. Such dependencies must be present in the Web in particular due to the high clustering of the Web graph [18] (roughly speaking, clustering means that with high probability, two neighbors of the same page are connected to each other). In our further research we will try to include some sort of dependencies along with dangling nodes and random Out-Degrees [15]. Besides, we could also consider personalization or topic sensitivity [22]. The impact of these factors on the PageRank distribution could be determined by extending and generalizing the proposed analytical model.

Acknowledgment

Nelly Litvak gratefully acknowledges the financial support of the Netherlands Organization for Scientific Research (NWO) under the Meervoud grant 632.002.401.

References

1. Litvak, N., Scheinhardt, W.R.W., Volkovich, Y.: In-Degree and PageRank: Why do they follow similar power laws? (to appear in Internet Math.)
2. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 33, 107–117 (1998)
3. Pandurangan, G., Raghavan, P., Upfal, E.: Using PageRank to characterize Web structure. In: H. Ibarra, O., Zhang, L. (eds.) COCOON 2002. LNCS, vol. 2387, Springer, Heidelberg (2002)
4. Donato, D., Laura, L., Leonardi, S., Millozi, S.: Large scale properties of the Webgraph. Eur. Phys. J. 38, 239–243 (2004)
5. Fortunato, S., Flammini, A., Menczer, F., Vespignani, A.: The egalitarian effect of search engines (2005), `arxiv.org/cs/0511005`
6. Becchetti, L., Castillo, C.: 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. ACM Press, New York (2006)
7. Berkhin, P.: A survey on PageRank computing. Internet Math. 2, 73–120 (2005)
8. Langville, A.N., Meyer, C.D.: Deeper inside PageRank. Internet Math. 1, 335–380 (2003)
9. Robert, P.: Stochastic networks and queues. Springer, New York (2003)
10. Meyer, A.D., Teugels, J.L.: On the asymptotic behaviour of the distributions of the busy period and service time in M/G/1. J. App. Probab. 17, 802–813 (1980)
11. Zwart, A.P.: Queueing Systems with Heavy Tails. PhD thesis, Eindhoven University of Technology (2001)

12. Bingham, N.H., Doney, R.A.: Asymptotic properties of supercritical branching processes. I. The Galton-Watson process. Advances in Appl. Probability 6, 711–731 (1974)
13. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation. Cambridge University Press, Cambridge (1989)
14. Avrachenkov, K., Litvak, N., Nemirovsky, D., Osipova, N.: Monte Carlo methods in PageRank computation: When one iteration is sufficient (electronic). SIAM Journal on Numerical Analysis 45(2), 890–904 (2007)
15. Volkovich, Y., Litvak, N., Donato, D.: Determining factors behind the PageRank log-log plot. In: Bonato, A., Chung, F.R.K. (eds.) WAW 2007. LNCS, vol. 4863, Springer, Heidelberg (2007)
16. Aldous, D., Bandyopadhyay, A.: A survey of max-type recursive distributional equations. Ann. Appl. Probab. 15, 1047–1110 (2005)
17. Stanford dataset: (Accessed in March 2006), http://www.stanford.edu/simsdkamvar/research.html
18. Newman, M.E.J.: Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46, 323–351 (2005)
19. Avrachenkov, K., Lebedev, D.: PageRank of scale free growing networks. Internet Mathematics 3(2), 207–231 (2006)
20. Fortunato, S., Flammini, A.: Random walks on directed networks: The case of PageRank (2006), arxiv.org/physics/0604203
21. Albert, R., Barabási, A.L.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
22. Haveliwala, T.H.: Topic-sensitive PageRank. In: Proceedings of the Eleventh International World Wide Web Conference, Honolulu, Hawaii (2002)