MaxRank: Discovering and Leveraging the Most Valuable Links for Ranking

Hengshuai Yao
University of Alberta
hengshua@cs.ualberta.ca

Abstract
On the Web, visits of a page are often introduced by one or more valuable linking sources. Indeed, good back links are valuable resources for Web pages and sites. We propose to discovering and leveraging the best backlinks of pages for ranking. Similar to PageRank, MaxRank scores are updated recursively. In particular, with probability λ, the MaxRank of a document is updated from the backlink source with the maximum score; with probability 1 − λ, the MaxRank of a document is updated from a random backlink source. MaxRank has an interesting relation to PageRank. When λ = 0, MaxRank reduces to PageRank; when λ = 1, MaxRank only looks at the best backlink it thinks. Empirical results on Wikipedia shows that the global authorities are very influential; Overall large λs (but smaller than 1) perform best: the convergence is dramatically faster than PageRank, but the performance is still comparable. We study the influence of these sources and propose a few measures such as the times of being the best backlink for others, and related properties of the proposed algorithm. The introduction of best backlink sources provides new insights for link analysis. Besides ranking, our method can be used to discover the most valuable linking sources for a page or Website, which is useful for both search engines and site owners.

1 Introduction
The gigantic size and diverse content of modern databases have made ranking algorithms fundamental components of search systems [1]. The link analysis approach to ranking has been shown to be very effective in evaluating the qualities of Webpages [21, 22], with widely practice from industry and intensive studies from academics. The success has proven that the hyperlinks on the Web are useful in finding high quality sources, which is hard based only on the content of pages. PageRank and HITS are two seminal algorithms in literature. PageRank finds authorities which are the pages frequently visited by a random surfer. HITS finds both authorities and hubs, which are defined recursively—the authorities are frequently linked by the hubs which turn out to be the pages frequently linked by authorities. In this paper, we will be focused on finding authorities in the spirit of PageRank, though our techniques may also apply to HITS and other link analysis algorithms.

PageRank. In PageRank formulation, with probability c, a random surfer model follows the links on a page uniformly at random, and with probability 1 − c, the surfer model jumps to a new page selected uniformly at random from the database. The PageRank value of a page is defined as the probability of visiting the page in the long run of the random walk, e.g., see [23, 24].

Suppose there are N documents in the database. All vectors are column vectors. The transpose of a matrix X is denoted by XT. We need the following notations.

L be an adjacency matrix of the database. That is, L(i, j) = 1 if there is a link from document i to document j, otherwise L(i, j) = 0, i, j = 1, 2, . . . , N;
L be a row normalized matrix of L;
e be a vector of all 1s, and v be a vector of probabilities that sum to one; and
S be a stochastic matrix such that S = L + (aeT/N), where ai = 1 if document i is dangling (i.e., document i has no forward link) and 0 otherwise.

The transition probability matrix used by PageRank is

\[ G = cS + (1 - c)e v^T, \]

where v (often called the teleportation vector) is a probability vector that sums to one. Matrix G is sometimes called the Google matrix in literature [23]. One merit of the Google matrix is that it is stochastic and primitive and thus its steady state distribution (also called the stationary distribution) exists. In fact, PageRank (denoted by $\pi$) is exactly the steady state distribution vector of G, satisfying

\[ \pi = G^T \pi. \]

The other merit of G is that it does not have to be stored, and the power iteration of computing $\pi$ can take advantage of the rank-1 matrix evT, manipulating $S$, c, e and v directly, e.g., see [12].

Considerable efforts have been devoted to the computation problem of PageRank due to its large scale applications. This is especially important when one wants to compute multiple PageRank vectors depending on queries and users. For a detailed discussion, please refer to Section 5. In this paper, we present a method utilizing the best backlinks, which
have a much faster convergence than PageRank but the perfor-
mance is still comparable. We are interested in under-
standing the roles of these influential links and their implica-
tion for link analysis algorithms especially PageRank.

2 Research Questions

Link analysis takes advantage of the linking information in
calculating document importances. For example, Pagerank
uses the back links of a document in updating its score.
Intuitively, there are influential links which contribute a large
portion to the score, and there are unimportant links which
only contribute a negligible portion. We would like to ask
the following questions.

- Where are the influential links from? What types of
documents are the influential sources? Are they
authorities, hubs, or anything else? What relations are
they to the nodes that are influenced by them?

- How many such influential sources are there?

- How influential is a backlink to the score of a docu-
ment? Most importantly, how influential are those in-
fluential back links?

These questions are interesting for all link analysis
algorithms. In this paper, we will be dealing with PageRank.
By answering these questions, we wish to gain insights
into the connectivity of large, real-world graphs and the
quality of documents, and provide a better ranking. A result
of this study is a ranking method that takes advantage of
the most influential back links to discover authorities and
communities.

3 The Best Back Links and MaxRank

In the case of PageRank-style authority discovery, a natural
definition of the best back link of a page is the one with the
largest score.

We discover the best back links in the same process of
authority score update, giving a so-called MaxRank method.
The basic idea of this algorithm is, with probability λ, the
contributing score comes from the best backlink of the page;
with probability 1 − λ, the contributing scores come from a
random backlink of the page.

3.1 The Algorithm In particular, MaxRank of a page \( j \)
\((j = 1, 2, \ldots, N)\) is defined by

\[
R(j) = \begin{bmatrix}
\lambda P(i^*, j) R(i^*) + (1 - \lambda) \
\sum_{i \in B(j)} P(i, j) R(i) \\
(1 - c) v(j)
\end{bmatrix}
\]

where

\[
i^* = \arg \max_{i \in B(j)} R(i),
\]

\(\lambda \in [0, 1], B(j)\) is the set of backlink pages of page \( j \), and
\(P(i, j)\) is the probability of going from page \( i \) to page \( j \).

3.2 Convergence of MaxRank In this section, we first
give a theorem showing that both variants of MaxRank are
well defined. A straightforward application of this theorem
is that power iteration of computing MaxRank is guaranteed
to converge for \( \lambda \in [0, 1] \).

Theorem 1. For \( c \in (0, 1) \) and \( \lambda \in [0, 1] \), MaxRank is
well defined.

Proof. For notational convenience, we define

\[
T(R, j) = \left[ \lambda P(i^*, j) \max_{i \in B(j)} R(i) + (1 - \lambda) \sum_{i \in B(j)} P(i, j) R(i) \right].
\]

Accordingly, we have

\[
R(j) = cT(R, j) + (1 - c)v(j)
\]

\(j = 1, 2, \ldots, N\). In matrix form, we have

\[
R = cT(R) + (1 - c)v,
\]

where \(T(R)\) is a vector, with \(T(R) = T(R, j), j = 1, 2, \ldots, N\).

Next we are to prove that \(T(R)\) is a non-expansion
operator with respect to the 1-norm, which means that

\[
||T(R)||_1 \leq ||R||_1.
\]

According to the definition of \(T(R)\), \(T(R)(j)\) and \(T(R, j)\),
we have

\[
T(R) = T \cdot R,
\]

where \(T\) is a \(N \times N\) matrix, with \(T(j, i) = P(i, j)\), if
page \( i \) is the best backlink of page \( j \); otherwise \(T(j, i) = (1 - \lambda) P(i, j)\).

Then the inequality (3.3) can be proven in the following
steps:

\[
||T(R)||_1 \leq ||T||_1 ||R||_1
\]

\[
= \max_{i=1,2,\ldots,N} \sum_{j=1}^{N} T(j, i) ||R||_1
\]

\[
\leq \max_{i=1,2,\ldots,N} \sum_{j=1}^{N} P(i, j) ||R||_1
\]

\[
= ||R||_1.
\]

For the third equation, the equality holds when \( i \) is the best
backlink for all pages.

Thus \( T \) is a non-expansion mapping in 1-norm. Accord-
ing to equation (3.2), \( R \) is defined by a contraction mapping
composed of \( T \) and \( c \). Hence \( R \) is finite.
The definition of MaxRank enables straightforward estimation using power iteration starting from any initial guess. The convergence of power iteration is guaranteed following an argument similar to Theorem 1.

**Theorem 2. (Convergence)** For $c \in (0, 1)$ and $\lambda \in [0, 1]$, power iteration of solving MaxRank converges to the true vector defined in (3.7), irrespective of any initial vector.

We will consider the random surfer in the remainder of this paper. That is, the probability of going from a (non-dangling) page $i$ to a page $j$ is $1/n_i$, where $n_i$ is the number of (forward) links on page $i$. In this case, it is noticeable that when $\lambda = 0$, the algorithm reduces to PageRank. When $\lambda = 1$, the algorithm only considers the “best” backlink it finds and ignores the contribution from the others. However, in our experience, this usually gives poor ranking results because the selected best backlink pages are usually not good in quality.

**4 Empirical Results**

In this section, we study the proposed algorithm and questions on the Wikipedia English article dump, which contains about 6 million pages (articles or categories). For all algorithms, $c = 0.85$ was used. The teleportation probabilities were uniformly set to $1/N$. All algorithms are updated by the standard power iteration. No sophisticated update is used for any algorithm.

Recall that we would like to study the following questions. What are the sources of the best back links? How many are they? How influential are they? For space limitation we show only the case of $\lambda = 0.1$ in this paper.

**4.1 Sources of the Best Back links** Table 3 shows the sources of the best backlinks for the top-50 pages on Wikipedia, using algorithm MaxRanked with $\lambda = 0.1$. Note that this choice produces a similar scoring to PageRank, as will be shown later. The sources of the best backlinks are mostly global authorities. The very top pages are seen to support many top pages. For example, “United States” influences many other concepts which further influence the remaining of the site. The effect is that this classifies the site into clusters of nodes, in each of which there are only a small number of dominant nodes.

There are only 775, 438 unique backlink sources with MaxRank. They support the whole site and form a core. The size of this core is only about 0.7% of the total number of links (117, 864, 053), and about 13.5% of the total number of the pages (5, 743, 047). On average a core page “supports” about 3, 620, 343/775, 438 ≈ 4.7 pages. This is also an estimate of the average size of the clusters. The size of the best backlink core for various $\lambda$ is shown in Figure 1. $\lambda = 0.3$ leads to the smallest core for this example. For $\lambda$ larger than 0.7, the core size is much larger and increases much quicker with respect to $\lambda$.

**4.2 Influence of the Best Backlink Sources** We measure the influence of the best backlink sources in three distinct aspects. The first measure is the collective influence of the best backlink sources in the graph, which is defined as the ratio of the sum of the scores of all the best backlink sources over the sum of the scores of all the pages. The collective influence of the core is 53.1%. Note that the number of core pages is only about 13.5% of the whole graph. Thus the influence of the core is significant.

Different core sources have different strength of influence. Some contribute many best backlinks, while others only contribute a few. Thus this suggests a measure for influential sources, in particular, by the times of being the best back link (TBB) to other nodes. Note that the TBB of an influential page is equal to the number of pages that the page supports. Table 1 shows the ordering of the sources according to the TBB measure. In addition, we also show in this table the ratio of TBB to the out-degree of the sources, which measures the percentage of competitive links cast by the sources. In this top list we see many hubs and authorities, and the number of hubs is more than the number of authorities. Thus on Wikipedia the more links an article has the more likely it is influential to others.

A log-log plot of the distributions of the out-degree and the TBB is shown in the left plot of Figure 3. Some key observations are as follows. First, the number of pages that have been the best backlink only a few times is very large, while the number of pages that have been the best backlink many times is very small. Second, the log-log curve of TBB distribution is more straight, which means the TBB distribution follows an exponential distribution in a more strict way. Third, it can be seen that for $x > 10$, the two curves follow a similar exponential distribution with a close
exponent, with a large shift in the $x$ direction which indicates that the TBB of a page is much smaller than the out-degree of the same page.

The (sorted) ratio between the TBB and the out-degree for each core page is shown in the middle plot of Figure 3. First, the sources whose value of the ratio is smaller than 0.2 are about 66% of the total core. This means the majority of the core has only 20% of their links being the best backlinks. Second, the number of those pages whose ratio is equal to 1.0 is about 87,193 (11% of the total core). Astonishingly, 86,763 (99.5%) of them have only one link. This sheds lights on the structure of Wikipedia. Most of them are due to the existence of “redirect pages” in Wikipedia, which contains no content but a “link” to another article. Third, the remaining sources have a ratio larger than 0.2. Together with the nontrivial sources whose ratio is 1.0, they form the most competitive link sources of the core. They take about 23% of the total core. Of them, only 4,632 sources have a ratio larger than 0.5, and 360 sources have a ratio larger than 0.8. In short, the number of nontrivial, competitive backlink sources is very small.

The third is from the perspective of an ordinary page (either in the core or not in the core), a measure of being influenced by the best back link, by the ratio of the score contributed by the best back link over the overall score of the page. We expect this measure can distinguish authoritative sources. This ratio for all the pages is shown in the right plot of Figure 3. For authorities with high scores, this ratio is very small. Thus they are not easily influenced even by the best backlink source. As pages become less authoritative (along the negative direction of the $x$-axis), the values of this ratio become more diverse. For example, we can observe the values of this ratio cover almost the whole range of $(0, 1)$ for pages with a score equal to $10^{-5}$.

Figure 2 shows the TBB versus the out-degree for the top authorities. For the very top-100 authorities, the curve is almost linear, and very close to $y = x$. (Note that all points are below $y = x$.) Thus their links are very influential. Further down the ordering of the authorities, we can observe that there are more and more less influential pages.

4.3 Convergence Studies Figure 4 (Left) compares the convergence rates of MaxRank and PageRank, measured in terms of the (1-norm) errors between successive iterations. MaxRank is faster than PageRank. The advantage is very significant for large $\lambda$. MaxRank with $\lambda = 0.1$ needs about 20 iterations to reach the accuracy by PageRank at the 30th iteration, while with $\lambda = 0.9$ MaxRank only needs 3 or 4 iterations.

4.4 Performance of MaxRank We compared the top list for the three algorithms, since it is usually the most important in practice. Table 2 shows the top 50 pages by PageRank (MaxRank with $\lambda = 0$). Table 3 and Table 4 show the top results of MaxRank with $\lambda = 0.1, 0.5, 0.9$. We also tested $\lambda = 1$ for MaxRank, but the results were very poor. The intuition is that the found “best backlinks” are not good without considering the wisdom of the majority. Note that in the tables, “ISBN is short for International Standard Book Number”, and “Inter-Air-Trans-code” is short for “International Air Transport Association airport code”.

The top lists of these algorithms have some similarities, and also some differences. In order to measure the similarity between the algorithms, we performed comparisons using two measurements. One is the percentage of common pages in the top-$k$ lists by two algorithms,

$$c_k = \frac{\# \text{Common pages in top}-k}{k} \in [0, 1].$$

The other is Kendall’s tau coefficient which measures the correlation in two rankings [19]. Here we care about whether MaxRank ranks the top-$k$ pages of PageRank in a consistent manner to PageRank, so the measure used is

$$\tau_k = \frac{n_k}{C_k} \in [0, 1],$$

where $n_k$ is the number of concordant orderings for every two pages from the top-$k$ pages of PageRank. The results of $c_k$ are summarized in the middle plot of Figure 4 for $k = 5, 10, 30, 50, 80, 100, 300, 500, 800, 1000$. Notice
that MaxRank performs remarkably similarly to PageRank for \( \lambda = 0.1 \), due to that the effect of the best backlinks is made small. In general, the smaller \( \lambda \) is, the more similar ranking of MaxRank to that of PageRank.

The results of \( r_k \) are summarized in the right plot of Figure 4. Similarly, the smaller the parameter \( \lambda \) is, the more similar the ranking is to PageRank. In particular, MaxRank with \( \lambda = 0.1 \) has a very similar \( r_k \) to PageRank for all \( k \). For large \( \lambda \) like 0.9 and 0.99, MaxRank still has about 80% similarities on average, and 65% similarities at worst to PageRank. The difference between the orderings of 0.9 and 0.99 for MaxRank is relatively small for all \( k \). This suggests that increasing \( \lambda \) to large values close to 1 produces stable rankings.

5 Discussion

The size of the Web creates a large computation burden for PageRank. Currently most large commercial search engines index 10 to 100 billion pages. However, the Web is actually much larger, e.g., there were already 1 trillion unique URLs in 2008 according to Google. Computing a single, global PageRank for the Web is already very demanding. Page et. al. used power iteration to take advantage of the sparse nature of the link structure of the Web\(^1\). Kamvar et al. proposed an adaptive method which monitors the change in the PageRank update for each page, and removes those pages whose update no longer changes\(^2\). Methods proposed in\(^3\),\(^4\),\(^5\),\(^6\),\(^7\),\(^8\) take advantage of the structure of link matrices and compute PageRank block-wise. Other linear system solvers were also used to update PageRank. For example, Kamvar et. al. used extrapolation methods\(^9\); Gleich et. al. proposed an inner-outer iteration procedure\(^10\), which essentially applies preconditioning incrementally; and Langville and Meyer, and Ipsen and Kirkland studied aggregation/disaggregation methods\(^11\),\(^12\).

It becomes more severe when one wants to computes many score vectors, such as many personalized PageRank\(^13\),\(^14\), context-sensitive or query-dependent scores\(^15\),\(^16\),\(^17\),\(^18\) which has numerous applications in search systems. Jeh and Widom proposed a scalable method by pre-computing some components of PageRank and saving them for efficient future computation\(^19\). Fogaras et. al. simulated a number of random walks and used Monte-Carlo methods to estimate personalized PageRank vectors\(^20\).

The computation of PageRank is very demanding. Thus distributed, parallel computation becomes necessary for large graphs. The methods in\(^21\),\(^22\),\(^23\) partition the whole graph into disjoint subgraphs, and then compute local PageRank for each subgraph. The local PageRanks are then merged, considering the links between the subgraphs. The methods in\(^24\),\(^25\) feature in the use of advanced linear system solvers. Some researchers also considered efficient hardware structures, such as specially optimized circuits\(^26\). For excellent surveys of PageRank, please refer to\(^27\),\(^28\),\(^29\).

Our method is very different from these efforts in literature, though it should be noted that these techniques also apply to our algorithms in a straightforward way. Our work takes advantage of the influential links in updating PageRank-style scores. We hope by doing so one can gain speedup in convergence and the performance is similar or comparable to PageRank.

6 Conclusion

The observation leading to this paper is that there exists one or more valuable backlinks for a page with a nonzero number of backlinks. We show that by leveraging the best backlinks a recursive update can have a much faster convergence than PageRank. The algorithm has a parameter \( \lambda \in [0, 1] \), which controls the effects of the best backlinks discovered. When \( \lambda = 0 \), the algorithm reduces to PageRank. Empirical

---

\(^1\)http://googleblog.blogspot.com/2008/07/we-knew-the-web-was-big.html

\(^2\)PageRank-style scores. We hope by doing so one can gain speedup in convergence and the performance is similar or comparable to PageRank.
Table 1: The ordering of the best backlink sources according to the “Times of being the Best Backlinks” (TBB). \( \lambda = 0.1 \).

| Rank | Page | TBB | #Links | TBB/#Links | Score |
|------|------|-----|--------|------------|-------|
| 1    | List_of_endangered_species | 3850 | 5097 | 0.755346 | 0.000003 |
| 2    | Arvados_in_Germany | 3044 | 4032 | 0.754960 | 0.000084 |
| 3    | Village_development_committee | 2764 | 2982 | 0.936302 | 0.000024 |
| 4    | Index_of_India_related_articles | 2581 | 4772 | 0.540863 | 0.000006 |
| 5    | List_of_Tachirthdaya_genera_and_species | 2547 | 2635 | 0.966603 | 0.000003 |
| 6    | List_of_gears | 2325 | 3885 | 0.598456 | 0.000105 |
| 7    | List_of_auxiliaries_of_the_United_States_Navy | 1751 | 1866 | 0.928420 | 0.000017 |
| 8    | List_of_municipalities_in_Switzerland | 1749 | 1967 | 0.889171 | 0.000049 |
| 9    | List_of_fish | 1717 | 1742 | 0.985649 | 0.000001 |
| 10   | List_of_municipalities_and_towns_in_Slovakia | 1692 | 2302 | 0.739013 | 0.000008 |
| 11   | 2007 | 1632 | 1856 | 0.879310 | 0.001658 |
| 12   | List_of_state_recreational_fish | 1604 | 2026 | 0.791708 | 0.000007 |
| 13   | United_States | 1448 | 1448 | 1.000000 | 0.000930 |
| 14   | List_of_Roman_Catholic_dioceses_(alphabetical) | 1434 | 2326 | 0.616509 | 0.000002 |
| 15   | List_of_uranium_conditions | 1306 | 1829 | 0.714051 | 0.00013 |
| 16   | United_Kingdom | 1140 | 1505 | 0.757475 | 0.000913 |
| 17   | List_of_Olympic_medalists_in_athletics_(men) | 1127 | 2069 | 0.544708 | 0.00013 |
| 18   | List_of_gosho_Lords_in_Germany | 1092 | 1304 | 0.837423 | 0.000084 |
| 19   | List_of_Vanity_Fair_articatures | 1090 | 1815 | 0.600551 | 0.000001 |
| 20   | Russia | 1087 | 1597 | 0.680651 | 0.001464 |
| 21   | List_of_mantis_genera_and_species | 1001 | 1057 | 0.947020 | 0.000025 |
| 22   | List_of_United_States_representatives_from_New_York | 996 | 1342 | 0.742179 | 0.000025 |
| 23   | List_of_benchmarks | 992 | 1168 | 0.849315 | 0.000033 |
| 24   | Catholic_University | 990 | 1166 | 0.849057 | 0.000013 |
| 25   | Index_of_statistics_articles | 984 | 2025 | 0.485868 | 0.000016 |
| 26   | 2006_music | 953 | 2162 | 0.440796 | 0.000017 |
| 27   | List_of_paleontological_fish | 946 | 1049 | 0.901811 | 0.000007 |
| 28   | England | 927 | 1551 | 0.597679 | 0.000024 |
| 29   | List_of numberOfRows (C) | 920 | 1141 | 0.806310 | 0.000006 |
| 30   | List_of_marine_aquarium_fish_species | 898 | 1103 | 0.814143 | 0.000002 |
| 31   | List_of_school_districts_in_Texas | 891 | 943 | 0.944857 | 0.000007 |
| 32   | Peerage_of_the_United_Kingdom | 856 | 1132 | 0.756184 | 0.000052 |
| 33   | List_of_rivers_of_New_Zealand | 843 | 919 | 0.917301 | 0.000019 |
| 34   | List_of_baronetcies | 841 | 1577 | 0.533291 | 0.000003 |
| 35   | London | 838 | 1323 | 0.633409 | 0.001363 |
| 36   | List_of_subjects_in_Gray's_Anatomy_XI_Neurology | 825 | 1426 | 0.578541 | 0.000037 |
| 37   | 2006_music | 824 | 1780 | 0.462921 | 0.000018 |
| 38   | Pronunciation_of_easter_nam | 823 | 910 | 0.904396 | 0.000030 |
| 39   | List_of_destroyers_in_U.S._Navy | 814 | 1030 | 0.790291 | 0.000012 |
| 40   | California | 807 | 1192 | 0.677013 | 0.000155 |
| 41   | List_of_number_of_C (1) | 799 | 1063 | 0.751646 | 0.000013 |
| 42   | Cocaine | 789 | 1231 | 0.640942 | 0.000005 |
| 43   | Sibley-Morse_checklist_18 | 789 | 1844 | 0.427874 | 0.000000 |
| 44   | List_of_United_States_representatives_from_Pennsylvania | 789 | 1085 | 0.727189 | 0.000021 |
| 45   | List_of_Digimon | 788 | 788 | 1.000000 | 0.000186 |
| 46   | List_of_third_genera | 788 | 1929 | 0.408502 | 0.000005 |
| 47   | List_of_subjects_in_Gray's_Anatomy_XII_Splanchnology | 786 | 1116 | 0.704301 | 0.000046 |
| 48   | List_of_state_roads_in_New_York | 772 | 985 | 0.783756 | 0.000025 |
| 49   | Italy | 762 | 1310 | 0.581679 | 0.001633 |
| 50   | List_of_JCT_Canoe_Sprint_World_Championships_medallists_in_men_kayak | 761 | 865 | 0.879769 | 0.000007 |

Results show that with large \( \lambda \)s (but smaller than 1) the new algorithm converges dramatically faster, but the results still have 80% similarities to PageRank on average (measured with Kendall’s tau). Thus our algorithm is advantageous for ranking in large search systems, where the computation of many personalized, query-dependent or context-sensitive score vectors is demanding.

Results on Wikipedia show that the number of unique best backlink sources (the so-called “core” in the paper) is only about 13.5% of the total number of pages. However, the sum of their scores is more than a half (about 53.1%) of the total scores. We propose to measure a source in the core by the times of being the best backlinks (TBB) and the ratio between TBB and the out-degree. Results show that TBB follows an exponential distribution with a similar exponent to the distribution of the out-degrees. With these two measures, the number of competitive backlink sources is very small. Results also show that a top authority is not easily influenced by the best backlink source.

References

[1] Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., and Raghavan, S. (2001). Searching the Web. ACM Transactions on Internet Technology, 1(1):2-43.
[2] Berkhin, P. (2005). A survey on PageRank computing. Internet Mathematics, 2(3):73-120.
[3] Bianchini, M., Gori, M., and Scarselli, F. (2005). Inside pagerank. ACM Transactions on Internet Technologies, 1(1):2-43.
[4] Brinkmeier, M. (2006). PageRank revisited. ACM Transactions on Internet Technology, 6(3):282-301.
Figure 4: Left: convergence rate comparisons of MaxRank and PageRank. Middle and Right: Percentage of the common pages and Kendall’s tau for the top-$k$ lists of MaxRank and PageRank.

[5] Broder, A. Z., Lempel, R., Maghoul, F., and Pedersen, J. (2004). Efficient PageRank approximation via graph aggregation. WWW.

[6] Fogaras, D. and Rácz, B. (2004). Towards scaling fully personalized PageRank. WWW.

[7] Gleich, D. and Zhukov, L. (2005). Scalable computing for power law graphs: Experience with parallel pagerank. Technical report, Yahoo! Research Labs Technical Report.

[8] Gleich, D., Zhukov, L., and Berkhin, P. (2004). Fast parallel PageRank: A linear system approach. Technical report, Yahoo! Research Labs Technical Report, YRL-2004-038.

[9] Gleich, D. F. (2009). Models and Algorithms for PageRank sensitivity. PhD thesis, Stanford University.

[10] Gleich, D. F., Gray, A. P., Greif, C., and Lau, T. (2010). An inner-outer iteration for PageRank. SIAM Journal of Scientific Computing, 32(1):349–371.

[11] Haveliwala, T. (1999). Efficient computation of PageRank. Technical Report 1999–31, Database Group, Computer Science Department, Stanford University.

[12] Haveliwala, T. (2005). Context-Sensitive Web Search. PhD thesis, Stanford University.

[13] Haveliwala, T. H. (2002). Topic-sensitive PageRank. WWW.

[14] Ipsen, C. F. and Kirklad, S. (2004). Convergence analysis of an improved pagerank algorithm. Technical report, NCSU CRSC Technical Report.

[15] Jeh, G. and Widom, J. (2003). Scaling personalized web search. WWW.

[16] Kamvar, S., Haveliwala, T., and Golub, G. (2003a). Adaptive methods for the computation of PageRank. Technical report, Stanford University.

[17] Kamvar, S., Haveliwala, T., Manning, C., and Golub, G. (2003b). Exploiting the block structure of the web for computing PageRank. Technical report, Stanford University.

[18] Kamvar, S. D., Haveliwala, T. H., Manning, C. D., and Golub, G. H. (2003c). Extrapolation methods for accelerating PageRank computations. WWW.

[19] Kendall, S. M. G. (1975). Rank Correlation Methods. Charles Griffin & Company Limited.

[20] Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. SODA.

[21] Langville, A. and Meyer, C. (2004a). Updating PageRank with iterative aggregation. WWW.

[22] Langville, A. N. and Meyer, C. D. (2004b). Deeper inside PageRank. Internet Mathematics, 1.

[23] Langville, A. N. and Meyer, C. D. (2006). Google’s PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press.

[24] Liu, B. (2007). Web Data Mining: Exploring Hyperlinks, Contents and Usage data. Springer.

[25] Page, L., Brin, S., Motwani, R., and Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web. Technical report, Stanford University.

[26] Rafiei, D. and Mendelson, A. O. (2000). What is this page known for? computing web page reputations. WWW.

[27] Richardson, M. and Domingos, P. (2002). The intelligent surfer: Probabilistic combination of link and content information in PageRank. NIPS.

[28] Sargolzaei, P. and Soleymani, F. (2010). Pagerank problem, survey and future research directions. International Mathematical Forum, 5(19):937–956.

[29] Seamas, M., Dermot, G., and McElroy, C. (2008). Towards an FPGA solver for the pagerank eigenvector problem. Advances in Parallel Computing, 15.

[30] Zhu, Y., Ye, S., and Li, X. (2005). Distributed PageRank computation based on iterative aggregation-disaggregation methods. CIKM.
| Rank | Page                  | Score  | Best backlink       |
|------|----------------------|--------|---------------------|
| 1    | United States        | 0.013911 | ISBN                |
| 2    | ISBN                 | 0.007283 | United States       |
| 3    | United Kingdom       | 0.006135 | United States       |
| 4    | Wikimedia Commons     | 0.005986 | Wiktionary          |
| 5    | Wiktionary           | 0.004081 | United States       |
| 6    | France               | 0.004049 | United States       |
| 7    | Canada               | 0.003964 | Wiki                |
| 8    | Biography            | 0.003860 | United States       |
| 9    | England              | 0.003766 | United States       |
| 10   | Biological classification | 0.003562 | Arthropod           |
| 11   | English language     | 0.003522 | United States       |
| 12   | Australia            | 0.003426 | United States       |
| 13   | World War II         | 0.003336 | United States       |
| 14   | Binomial nomenclature | 0.003176 | Biological classification |
| 15   | Japan                | 0.003091 | United States       |
| 16   | India                | 0.003026 | United States       |
| 17   | Internet Movie Database | 0.002907 | Alexa, Internet     |
| 18   | Abbreviation         | 0.002882 | USA                 |
| 19   | Music genre          | 0.002844 | Poland              |
| 20   | Association football | 0.002766 | United States       |
| 21   | Europe               | 0.002753 | United States       |
| 22   | Record label         | 0.002734 | Music genre         |
| 23   | Italy                | 0.002612 | United States       |
| 24   | 2007                 | 0.002603 | Australia            |
| 25   | 2008                 | 0.002505 | Australia            |
| 26   | Russia               | 0.002399 | United States       |
| 27   | 2006                 | 0.002339 | United States       |
| 28   | 2005                 | 0.002130 | Poland              |
| 29   | 2006                 | 0.002054 | Music genre         |
| 30   | 2008                 | 0.002030 | Germany             |
| 31   | 2008                 | 0.001989 | Given_name          |
| 32   | 2008                 | 0.001919 | Germany             |
| 33   | 2008                 | 0.001850 | United States       |
| 34   | 2008                 | 0.001841 | United States       |
| 35   | 2008                 | 0.001829 | United States       |
| 36   | 2008                 | 0.001825 | United States       |
| 37   | 2008                 | 0.001752 | Biological classification |
| 38   | 2008                 | 0.001752 | Wikimedia Commons    |
| 39   | 2008                 | 0.001663 | United States       |
| 40   | 2008                 | 0.001663 | United States       |
| 41   | 2008                 | 0.001658 | World War II        |
| 42   | 2008                 | 0.001651 | United States       |
| 43   | 2008                 | 0.001643 | United States       |
| 44   | 2008                 | 0.001623 | France              |
| 45   | 2008                 | 0.001620 | United States       |
| 46   | 2008                 | 0.001592 | United States       |
| 47   | 2008                 | 0.001569 | United States       |
| 48   | 2008                 | 0.001559 | France              |
| 49   | 2008                 | 0.001554 | United States       |
| 50   | 2008                 | 0.001515 | United States       |

Table 3: Top 50 Wikipedia pages by MaxRank, λ = 0.1.

| Rank | Page                  | Score  | Best backlink       |
|------|----------------------|--------|---------------------|
| 1    | United States        | 0.009093 | ISBN                |
| 2    | ISBN                 | 0.004404 | United States       |
| 3    | United Kingdom       | 0.003863 | United States       |
| 4    | Wikimedia Commons     | 0.003614 | Wiktionary          |
| 5    | Biogaphy             | 0.003035 | Wiki                |
| 6    | Biological classification | 0.002773 | Arthropod           |
| 7    | Canada               | 0.002626 | United States       |
| 8    | France               | 0.002546 | United States       |
| 9    | Wiktionary           | 0.002474 | Wikimedia Commons    |
| 10   | England              | 0.002462 | United States       |
| 11   | Germany              | 0.002452 | United States       |
| 12   | Biological classification | 0.002273 | Biological classification |
| 13   | Australia            | 0.002063 | United States       |
| 14   | England              | 0.002017 | United States       |
| 15   | Germany              | 0.002012 | United States       |
| 16   | Russia               | 0.001968 | United States       |
| 17   | Poland               | 0.001968 | United States       |
| 18   | 2007                 | 0.001658 | Australia            |
| 19   | 2008                 | 0.001633 | Australia            |
| 20   | 2006                 | 0.001599 | Given_name          |
| 21   | Russia               | 0.001464 | United States       |
| 22   | 2008                 | 0.001363 | United Kingdom       |
| 23   | 2008                 | 0.001322 | Germany             |
| 24   | 2008                 | 0.001274 | Germany             |
| 25   | 2008                 | 0.001236 | Biological classification |
| 26   | 2008                 | 0.001183 | United States       |
| 27   | 2008                 | 0.001155 | United States       |
| 28   | 2008                 | 0.001140 | United States       |
| 29   | 2008                 | 0.001136 | United States       |
| 30   | 2008                 | 0.001120 | Wikimedia Commons    |
| 31   | 2008                 | 0.001077 | United States       |
| 32   | 2008                 | 0.001076 | United States       |
| 33   | 2008                 | 0.001055 | United States       |
| 34   | 2008                 | 0.001024 | Music genre         |
| 35   | 2008                 | 0.001023 | Japan               |
| 36   | 2008                 | 0.001007 | United States       |
| 37   | 2008                 | 0.001007 | United States       |
| 38   | 2008                 | 0.001004 | United States       |
| 39   | 2008                 | 0.001004 | United States       |
| 40   | 2008                 | 0.001004 | United States       |
| 41   | 2008                 | 0.001004 | United States       |
| 42   | 2008                 | 0.001004 | United States       |
| 43   | 2008                 | 0.001004 | United States       |
| 44   | 2008                 | 0.001004 | United States       |
| 45   | 2008                 | 0.001004 | United States       |
| 46   | 2008                 | 0.001004 | United States       |
| 47   | 2008                 | 0.001004 | United States       |
| 48   | 2008                 | 0.001004 | United States       |
| 49   | 2008                 | 0.001004 | United States       |
| 50   | 2008                 | 0.001004 | Video game          |
### Table 4: Top 50 Wikipedia pages by MaxRank, $\lambda = 0.5$.  
| Rank | Page                  | Score   | Best backlink |
|------|-----------------------|---------|---------------|
| 1    | United_States         | 0.002444 | ISBN          |
| 2    | Biography             | 0.001198 | Genre         |
| 3    | Biological_classification | 0.001103 | Arthropod     |
| 4    | “ISBN”                | 0.000949 | United_States |
| 5    | United_Kingdom        | 0.000929 | United_States |
| 6    | Music_genre           | 0.000753 | Record_producer |
| 7    | Record_label          | 0.000726 | Music_genre   |
| 8    | Wikimedia_Commons      | 0.000712 | Association_football |
| 9    | Canada                | 0.000685 | United_States |
| 10   | England               | 0.000671 | United_States |
| 11   | Binomial_nomenclature | 0.000656 | Biological_classification |
| 12   | Personal_name         | 0.000640 | Given_name    |
| 13   | Internet_Movie_Database | 0.000630 | Royal_Navy   |
| 14   | Germany               | 0.000607 | United_States |
| 15   | France                | 0.000603 | United_States |
| 16   | Australia             | 0.000558 | United_States |
| 17   | India                 | 0.000543 | United_States |
| 18   | Association_football  | 0.000533 | United_States |
| 19   | Japan                 | 0.000523 | United_States |
| 20   | Wiktionary            | 0.000510 | United_States |
| 21   | English_language      | 0.000502 | United_States |
| 22   | 2007                  | 0.000454 | Australia     |
| 23   | Abbreviation          | 0.000436 | USA           |
| 24   | Arthropod             | 0.000423 | Lepidoptera   |
| 25   | World_War_II          | 0.000415 | United_States |
| 26   | Italy                 | 0.000392 | United_States |
| 27   | Europe                | 0.000373 | United_States |
| 28   | Studio_album          | 0.000370 | Music_genre   |
| 29   | Politician            | 0.000364 | Video_game    |
| 30   | Record_producer       | 0.000363 | Music_genre   |
| 31   | 2008                  | 0.000349 | Germany       |
| 32   | Russia                | 0.000343 | United_States |
| 33   | 2006                  | 0.000340 | Germany       |
| 34   | Romania               | 0.000310 | United_States |
| 35   | Scientific_name       | 0.000326 | Biological_classification |
| 36   | Poland                | 0.000318 | United_States |
| 37   | Spain                 | 0.000313 | United_States |
| 38   | Romania               | 0.000310 | United_States |
| 39   | New_York_City         | 0.000294 | United_States |
| 40   | Public_domain         | 0.000293 | Wikimedia_Commons |
| 41   | Time_zone             | 0.000288 | United_States |
| 42   | Brazil                | 0.000282 | Russia        |
| 43   | Sweden                | 0.000280 | United_States |
| 44   | California            | 0.000278 | United_States |
| 45   | Television            | 0.000270 | United_Kingdom |
| 46   | Drainage_basin        | 0.000267 | New_York_City |
| 47   | Netherlands           | 0.000259 | United_States |
| 48   | 2005                  | 0.000255 | France        |
| 49   | New_York              | 0.000249 | United_States |
| 50   | New_Zealand           | 0.000249 | United_States |

### Table 5: Top 50 Wikipedia pages by MaxRank, $\lambda = 0.9$.  
| Rank | Page                  | Score   | Best backlink |
|------|-----------------------|---------|---------------|
| 1    | United_States         | 0.000311 | United_Kingdom |
| 2    | Biography             | 0.000194 | Autobiography |
| 3    | Biological_classification | 0.000175 | Arthropod     |
| 4    | Music_genre           | 0.000111 | Record_producer |
| 5    | United_Kingdom        | 0.000109 | United_States |
| 6    | Record_label          | 0.000106 | Music_genre   |
| 7    | Personal_name         | 0.000105 | Given_name    |
| 8    | “ISBN”                | 0.000099 | United_States |
| 9    | England               | 0.000088 | United_States |
| 10   | United_States         | 0.000087 | Royal_Navy   |
| 11   | Canada                | 0.000085 | United_States |
| 12   | Binomial_nomenclature | 0.000079 | Biological_classification |
| 13   | Arthropod             | 0.000075 | Lepidoptera   |
| 14   | India                 | 0.000073 | United_States |
| 15   | Germany               | 0.000072 | United_States |
| 16   | France                | 0.000067 | United_States |
| 17   | Australia             | 0.000062 | United_States |
| 18   | Japan                 | 0.000061 | United_States |
| 19   | Wiktionary            | 0.000060 | United_States |
| 20   | English_language      | 0.000059 | United_States |
| 21   | Politician            | 0.000058 | Music_genre   |
| 22   | Abbreviation          | 0.000058 | USA           |
| 23   | Arthropod             | 0.000057 | United_States |
| 24   | 2007                  | 0.000057 | Australia     |
| 25   | 2008                  | 0.000056 | Germany       |
| 26   | Record_producer       | 0.000054 | United_States |
| 27   | Wiktionary            | 0.000052 | Wikimedia_Commons |
| 28   | Studio_album          | 0.000050 | United_States |
| 29   | Geocode               | 0.000048 | UN/LOCODE     |
| 30   | UN/LOCODE             | 0.000047 | “Inter-Air-Trans-code” |
| 31   | Italy                 | 0.000046 | United_States |
| 32   | Russia                | 0.000044 | Germany       |
| 33   | Poland                | 0.000044 | United_States |
| 34   | Lepidoptera           | 0.000042 | Moth          |
| 35   | New_York_City         | 0.000042 | United_States |
| 36   | Drainage_basin        | 0.000040 | United_States |
| 37   | London                | 0.000039 | United_Kingdom |
| 38   | Television            | 0.000039 | United_States |
| 39   | Europe                | 0.000039 | United_States |
| 40   | Russia                | 0.000038 | United_States |
| 41   | Germany               | 0.000037 | Biological_classification |
| 42   | New_York_City         | 0.000036 | IUCN_Red_List |
| 43   | Poland                | 0.000035 | Wikimedia_Commons |
| 44   | Brazil                | 0.000034 | United_States |
| 45   | California            | 0.000033 | United_States |
| 46   | New_York_City         | 0.000032 | United_States |
| 47   | France                | 0.000031 | United_States |
| 48   | New_York_City         | 0.000030 | United_States |
| 49   | California            | 0.000029 | United_States |
| 50   | IUCN_Red_List         | 0.000028 | Conservation_status |