Three results on the PageRank vector: eigenstructure, sensitivity, and the derivative

David Gleich[1], Peter Glynn[2], Gene Golub[3], Chen Greif[4]

[1] Stanford University, Institute for Computational and Mathematical Engineering
Stanford, CA 94305 USA
dgleich@stanford.edu
[2] Stanford University, Department of Computer Science
Stanford, CA 94305 USA
golub@stanford.edu
[3] Stanford University, Department of Management Science and Engineering
Stanford, CA 94305 USA
glynn@stanford.edu
[4] University of British Columbia, Department of Computer Science
Vancouver, BC Canada
greif@cs.ubc.ca

Abstract. The three results on the PageRank vector are preliminary but shed light on the eigenstructure of a PageRank modified Markov chain and what happens when changing the teleportation parameter in the PageRank model. Computations with the derivative of the PageRank vector with respect to the teleportation parameter show predictive ability and identify an interesting set of pages from Wikipedia.

Keywords. PageRank, PageRank derivative, PageRank sensitivity, PageRank eigenstructure

1 Introduction

We present three results that touch on various aspects of PageRank computation [1–3]. Section 3 presents an analysis of the random walk in the PageRank modification of a Markov chain. This analysis yields the eigenvalues and eigenvectors of the PageRank Markov chain. While the eigenstructure of the PageRank Markov matrix was previously known [4–7], our techniques are based on looking at the n-step transition matrix.

Second, we consider the PageRank vector as a function of the teleportation parameter α and examine a first order Taylor approximation in section 4. For positive step sizes until $\alpha = 1$, the Taylor approximation is a PageRank vector for a different teleportation distribution vector.

⋆ The first author is supported by NSF grant CCF-0430617; the third author is partially supported by NSF grant CCF-0430617; and the fourth author is partially supported by Natural Sciences and Engineering Research Council of Canada.

Dagstuhl Seminar Proceedings 07071
Web Information Retrieval and Linear Algebra Algorithms
http://drops.dagstuhl.de/opus/volltexte/2007/1061

The next section presents an algorithm to compute the derivative of the PageRank vector with respect to the teleportation parameter by solving two PageRank problems. Algorithms to compute the derivative exist [8], but ours only requires the solution of PageRank problems. The motivation for studying the derivative is to understand how the ordering of the PageRank vector changes with α.

Each of our theoretical contributions is illustrated with a computational example in section 7. Two additional experiments examine the derivative. The first explores the relationship between the numerical value of the derivative and the change in rank of a page when increasing the teleportation parameter (section 8.1), and the second exhibits a few pages with the largest derivative (section 8.2).

An extended version of this report can be at found at `http://sccm.stanford.edu/wrap/pub-tech.html`.

2 Notation

The vector $\mathbf{e}$ is the vector of all ones. Let $\mathbf{P}$ represent a column stochastic matrix, that is $\mathbf{e}^T\mathbf{P} = \mathbf{e}^T$ and $p_{i,j} \geq 0$. We use the scalar variable $\alpha,\ 0 < \alpha < 1$ to denote the teleportation parameter in the PageRank problem and the vector $\mathbf{v}$ to denote the teleportation distribution vector where $\mathbf{e}^T\mathbf{v} = 1$ and $v_i > 0$.[1] The PageRank vector is denoted $\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha)$ and is the unique positive eigenvector with $\mathbf{e}^T\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha) = 1$ of the eigenvalue problem

$$\left[\alpha\mathbf{P} + (1-\alpha)\mathbf{v}\mathbf{e}^T\right]\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha) = \mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha). \tag{1}$$

This equation can be interpreted as a Markov chain where we denote the transposed transition matrix by

$$\mathbf{M} = \alpha\mathbf{P} + (1-\alpha)\mathbf{v}\mathbf{e}^T \tag{2}$$

and interpret $\mathbf{M}$ as a modified Markov matrix for the PageRank problem. We also exploit the normalization $\mathbf{e}^T\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha) = 1$ and write the PageRank vector as the solution of the linear system

$$(\mathbf{I} - \alpha\mathbf{P})\,\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha) = (1-\alpha)\mathbf{v}. \tag{3}$$

Where we believe the parameters are clear from context, we will shorten the PageRank vector to either $\mathbf{x}$ or $\mathbf{x}(\alpha)$. Langville and Meyer summarize numerous additional properties of these two formulations of the PageRank problem in their book [3].

[1] In the most general setting, we need not require $v_i > 0$ but this requirement simplifies many technical matters associated with the results.

3 The Markovian Eigenstructure of the Modified Markov Matrix

Our first result is about the eigenstructure of the modified Markov matrix in the PageRank problem. One possible (and common) interpretation of equation (1) is that the PageRank vector is the stationary distribution of a random walk where with probability α, the walk follows an existing random walk given by $\mathbf{P}$, and with probability $1-\alpha$, the walk resets to a prior distribution over states given by $\mathbf{v}$.

In the following derivation, we set the rank-1 matrix $\mathbf{v}\mathbf{e}^T = \mathbf{R}$, so $\mathbf{M} = \alpha\mathbf{P} + (1-\alpha)\mathbf{R}$. Our first lemma is about the n-step transition probabilities of the PageRank random walk.

Lemma 1. *The column-oriented n-step probability transition matrix for the modified random walk in the PageRank problem is*

$$\mathbf{M}^n = \alpha^n\mathbf{P}^n + (1-\alpha)\left[\sum_{k=1}^{n} \alpha^{k-1}\mathbf{P}^{k-1}\right]\mathbf{R}. \tag{4}$$

We now use the lemma and a few carefully chosen expressions to rewrite the n-step transition matrix and expose the eigenstructure of $\mathbf{M}$. Using the Neumann series of $(\mathbf{I}-\alpha\mathbf{P})^{-1} = \sum_{i=0}^{\infty}\alpha^i\mathbf{P}^i$, the term $\sum_{k=1}^{n}\alpha^{k-1}\mathbf{P}^{k-1} = (\mathbf{I}-\alpha^n\mathbf{P}^n)(\mathbf{I}-\alpha\mathbf{P})^{-1}$. We then introduce the Césaro limit matrix of $\mathbf{P}$, $\mathbf{\Pi} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathbf{P}^i$ (which always exists for stochastic matrices), so that

$$\begin{aligned}\mathbf{M}^n = {} & (1-\alpha)(\mathbf{I}-\alpha\mathbf{P})^{-1}\mathbf{R} \\ & + \alpha^n\mathbf{\Pi}[\mathbf{I}-(1-\alpha)(\mathbf{I}-\alpha\mathbf{P})^{-1}\mathbf{R}] \\ & + \alpha^n(\mathbf{P}^n-\mathbf{\Pi})[\mathbf{I}-(1-\alpha)(\mathbf{I}-\alpha\mathbf{P})^{-1}\mathbf{R}].\end{aligned} \tag{5}$$

Theorem 1. *The matrix $\mathbf{M}^n$ has the following properties.*

1. *It has a column eigenvector $\mathbf{x}(\mathbf{P},\mathbf{v},\alpha)$ and row eigenvector $\mathbf{e}^T$ with eigenvalue 1.*
2. *If $\mathbf{y}^T$ is a row eigenvector of $\mathbf{P}$ with eigenvalue 1 orthogonal to $\mathbf{v}$, $\mathbf{y}^T\mathbf{P} = \mathbf{y}^T$,[2] then $\mathbf{y}^T$ is a row eigenvector of $\mathbf{M}^n$ with eigenvalue α^n.*
3. *If $(\mathbf{f}^T,\gamma)$ is an eigenpair of $\mathbf{P}$, $\mathbf{f}^T\mathbf{P} = \gamma\mathbf{f}^T$ with $\gamma \neq 1$, then $\mathbf{f}^T[\mathbf{I}-(1-\alpha)(\mathbf{I}-\alpha\mathbf{P})^{-1}\mathbf{R}]$ is a row eigenvector of M^n with eigenvalue $\alpha^n\gamma^n$.*

From the eigenvectors and eigenvalues of $\mathbf{M}^n$ we can write the eigenvalues and eigenvectors of the matrix $\mathbf{M}$ itself. Let $\mathbf{Y}^T$ be the matrix of row eigenvectors of $\mathbf{P}$ with eigenvalue 1 orthogonal to $\mathbf{v}$, that is $\mathbf{Y}^T\mathbf{P} = \mathbf{P}$ and $\mathbf{Y}^T\mathbf{v} = 0$. Also, let $\mathbf{F}^T$ be the matrix of row eigenvectors of $\mathbf{P}$ with non-dominant eigenvalues

[2] Such a vector need not exist.

$\mathbf{\Lambda}$, $\mathbf{F}^T\mathbf{P} = \mathbf{\Lambda}\mathbf{F}^T$, then

$$\begin{bmatrix} \mathbf{e}^T \\ \mathbf{Y}^T \\ \mathbf{F}^T[\mathbf{I} - (1-\alpha)(\mathbf{I} - \alpha\mathbf{P})^{-1}\mathbf{R}] \end{bmatrix} \mathbf{M} = \begin{bmatrix} 1 & & \\ & \alpha\mathbf{I} & \\ & & \alpha\mathbf{\Lambda} \end{bmatrix} \begin{bmatrix} \mathbf{e}^T \\ \mathbf{Y}^T \\ \mathbf{F}^T[\mathbf{I} - (1-\alpha)(\mathbf{I} - \alpha\mathbf{P})^{-1}\mathbf{R}] \end{bmatrix}. \tag{6}$$

Another outcome of the previous theorem is that the eigenspace of $\mathbf{M}$ for eigenvalue α is insensitive to changes in the teleportation parameter α.

4 A Result on a Taylor Step

We discovered the result in this section while investigating the sensitivity of the PageRank vector $\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha)$ with respect to the teleportation parameter α. As in Golub and Greif's earlier work [9], we look at sensitivity in the context of the derivative of $\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha)$ with respect to α. In the remainder of this section, we write the PageRank vector as $\mathbf{x}(\alpha)$ and drop the dependence on $\mathbf{P}$ and $\mathbf{v}$. We write $\mathbf{x}'(\alpha)$ for $\frac{\partial}{\partial\alpha}\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha)$.

We derive an analytic form for the derivative from the linear system formulation of PageRank. Rewriting equation (3) slightly,

$$\mathbf{x}(\alpha) = \alpha\mathbf{P}\mathbf{x}(\alpha) + (1-\alpha)\mathbf{v} \tag{7}$$

and if we differentiate both sides with respect to α

$$\mathbf{x}'(\alpha) = \mathbf{P}\mathbf{x}(\alpha) + \alpha\mathbf{P}\mathbf{x}'(\alpha) - \mathbf{v}. \tag{8}$$

Theorem 2. *Fix $\mathbf{P}, \mathbf{v}$, and α and let $\mathbf{x}(\alpha)$ and $\mathbf{x}'(\alpha)$ be the PageRank vector and derivative with respect to α, then $\mathbf{y}(\gamma) = \mathbf{x}(\alpha) + \gamma\mathbf{x}'(\alpha)$ is a PageRank vector for $0 \le \gamma < 1-\alpha$ with teleportation distribution vector $\mathbf{w}(\gamma) = \frac{1}{1-\alpha}((1-\alpha-\gamma)\mathbf{v} + \gamma\mathbf{P}\mathbf{x}(\alpha))$.*

An immediate implication of the previous theorem is that $x'(\alpha)_i < \frac{1}{1-\alpha}$ (otherwise $y(1-\alpha-\varepsilon)_i > 1$ for some small but positive ε). Using different techniques, Langville and Meyer prove a slightly stronger version of the previous remark that $|x'(\alpha)_i| < \frac{1}{1-\alpha}$ [3, p.66].

5 Computing the Derivative by Solving PageRank

In this section we address computing the PageRank derivative vector with respect to α. One property of the derivative that we exploit is $\mathbf{e}^T\mathbf{x}'(\alpha) = 0$, which follows directly from the fact that $\mathbf{e}^T\mathbf{x}(\alpha) = 1$ for any α. For additional theoretical properties of the derivative, see Langville and Meyer [3, section 6.5]. In this section, we adopt the notation of the previous section to refer to a PageRank vector $\mathbf{x}(\alpha)$ and its derivative $\mathbf{x}'(\alpha) = \frac{\partial}{\partial\alpha}\mathbf{x}(\mathbf{P}, \mathbf{v}, \alpha)$.

In contrast with other methods for computing the derivative [8, 2, 9], our method involves solving only PageRank problems. This feature allows our algorithm to benefit from any algorithmic advances in computing PageRank.

The key to this result is an observation by Golub and Greif [9], $\mathbf{P}\mathbf{x}(\alpha) - \mathbf{v} = \frac{1}{\alpha}(\mathbf{x}(\alpha) - \mathbf{v})$. We use this result to simplify equation (8) and express the PageRank derivative vector as $\mathbf{x}'(\alpha) = \beta\mathbf{z}(\alpha) - \beta\mathbf{x}(\alpha)$ where $(\mathbf{I} - \alpha\mathbf{P})\mathbf{z}(\alpha) = (1-\alpha)\mathbf{x}(\alpha)$ and $\beta = \frac{1}{\alpha(1-\alpha)}$. This idea yields an algorithm for computing the PageRank derivative as the solution of two PageRank systems with different teleportation distribution vectors with the same value of α. Berkhin also made this observation [2].

One concern with the previous approach is that it requires computing PageRank for a column stochastic matrix $\mathbf{P}$. For computational reasons, many codes for PageRank often choose to work with an almost stochastic matrix $\overline{\mathbf{P}}$ where $(\mathbf{e}^T\mathbf{P})_i = \{0, 1\}$. Often the matrix $\mathbf{P} = \overline{\mathbf{P}} + \mathbf{v}\mathbf{d}^T$ where $\mathbf{d}^T = \mathbf{e}^T - \mathbf{e}^T\overline{\mathbf{P}}$ and $\mathbf{v}$ is the same teleportation distribution vector [10]. Boldi et al. call the ensuing formulation of PageRank *strongly preferential PageRank* in contrast with *weakly preferential PageRank* where $\mathbf{P} = \overline{\mathbf{P}} + \mathbf{u}\mathbf{d}^T$ for a distinct distribution vector $\mathbf{u}$ [11]. To maximize our computational advantage, we want to solve only strongly preferential PageRank problems. A simple application of the previous idea no longer works because the strongly preferential PageRank system for $\mathbf{z}(\alpha)$ is not linear with respect to the derivative linear system.

To address the non-linearity issue, we can rewrite the derivative as the solution of the linear system $(\mathbf{I} - \alpha\overline{\mathbf{P}})\mathbf{x}'(\alpha) = \frac{1}{\alpha}\mathbf{x}(\alpha) + \eta\mathbf{v}$, where η is an unknown scalar. The solution of a strongly preferential PageRank problem is also a solution vector for the system $\mathbf{I} - \alpha\overline{\mathbf{P}}$ with a rescaled right hand side. To compute η, we exploit the fact that $\mathbf{e}^T\mathbf{x}'(\alpha) = 0$. Combining these ideas yields the following algorithm.

1. Compute $\mathbf{x}(\alpha)$ as the solution to the original strongly preferential PageRank problem.
2. Compute $\mathbf{z}(\alpha)$ as the solution to the strongly preferential PageRank problem with teleportation distribution $\mathbf{x}(\alpha)$.
3. Set $\tilde{\mathbf{z}} = \frac{1}{\alpha(1-\alpha+\alpha\mathbf{d}^T\mathbf{z}(\alpha))}\mathbf{z}(\alpha)$.
4. Compute $\eta = \frac{-\mathbf{e}^T\tilde{\mathbf{z}}}{\mathbf{e}^T\mathbf{x}(\alpha)}$.
5. Return $\mathbf{x}'(\alpha) = \tilde{\mathbf{z}} + \eta\mathbf{x}(\alpha)$.

6 Datasets

Table 1 shows a series of properties about the datasets used in the forthcoming experiments. The graphs `aa-stan`, `ee-stan`, and `cs-stan` correspond to the web graphs for the hosts `aa.stanford.edu`, `ee.stanford.edu`, and `cs.stanford.edu`, respectively. These graphs were formed as a subset of the Webbase 2001 crawl [12] compressed with the Webgraph framework [13]. The graph cnr-2000 is the result of an Ubicrawler crawl [14].

Wikipedia provides access to semi-regular copies of its English page database. We downloaded the databases from late September 2006 [15] and early November 2006 [16]. From each database, we formed an article-article link graph, where an article is a page in the main Wikipedia namespace, for example `http://en.wikipedia.org/wiki/PageRank`; a category page, for example `http://en.wikipedia.org/wiki/Category:Matrix_theory`; or a portal page, for example `http://en.wikipedia.org/wiki/Portal:Mathematics`. We removed all other pages and links.

Graph	$\lvert\mathbf{V}\rvert$	$\lvert\mathbf{E}\rvert$	$\lvert\mathcal{C}\rvert$	**max** $\lvert\mathcal{C}_\mathbf{i}\rvert$	**max** $\mathbf{d}_{\text{out}}$	**max** $\mathbf{d}_{\text{in}}$	$\lvert\mathbf{d}_{\text{out}} = \mathbf{0}\rvert$	$\lvert\mathbf{d}_{\text{in}} = \mathbf{0}\rvert$
aa-stan	114	229	3	112	109	110	2	0
ee-stan	1,615	7,046	531	653	210	221	465	38
cs-stan	9,914	36,854	4,391	2,759	277	340	2,861	699
cnr-2000	325,557	3,216,152	100,977	112,023	2,716	18,235	78,056	0
wiki-2006-09	2,983,494	37,269,096	975,731	2,003,668	5,852	159,378	88,970	873,634
wiki-2006-11	3,148,440	39,383,235	1,040,035	2,104,115	6,576	168,685	91,462	932,906

Table 1. The table presents properties of the datasets we use for experiments. Each row of the table lists the number of vertices ($|V|$), number of edges ($|E|$), number of strongly connected components ($|\mathcal{C}|$), size of the largest strongly connected component ($\max |\mathcal{C}_i|$), maximum out-degree of a vertex ($\max d_{\text{out}}$), maximum in-degree of a vertex ($\max d_{\text{in}}$), count of nodes with no out-links ($|d_{\text{out}} = 0|$), and count of nodes with no in-links ($|d_{\text{in}} = 0|$).

7 Theoretical Examples

Unless otherwise noted, the experiments use the strongly preferential PageRank model with $\alpha = 0.85$ and a uniform teleportation distribution vector. For a directed adjacency matrix $\mathbf{A}$, $\overline{\mathbf{P}} = \mathbf{A}\mathbf{D}^{+}$ where $\mathbf{D}^{+}$ is the pseudo-inverse of the diagonal matrix of out-degrees for each node.

7.1 Eigenstructure

Tables 2-3 show experiments on the three smallest graphs to check the results of the eigenstructure of the matrix $\mathbf{M}$. The results confirm the analysis with good accuracy. We believe the eigenvalue counts in table 2 are not exact between $\mathbf{M}$ and $\mathbf{P}$ due to large numbers of eigenvalues quite close to the real and imaginary axes which made counting them exactly difficult. On cs-stan, ARPACK found an additional eigenvalue at α in $\mathbf{M}$ not predicted by our theory, but the third column in table 3 shows ARPACK missed this eigenvalue in $\mathbf{P}$.

	aa-stan		ee-stan		cs-stan	
	P	**M**	**P**	**M**	**P**	**M**
$\#\lambda_i \in \mathbb{R}$	114	114	1391	1496	198	198
$\#\lambda_i \in \mathbb{C}$	0	0	224	238	2	2
$\#\lvert\lambda_i\rvert < \varepsilon$	108	108	762	764	0	5
$\#\lvert\lambda_i - 1\rvert < \varepsilon$	1	1	13	1	111	1
$\#\lvert\lambda_i - (-1)\rvert < \varepsilon$	1	1	0	0	16	0
$\#\lvert\lambda_i - \alpha\rvert < \varepsilon$	0	0	0	12	0	112
$\#\lvert\lambda_i - (-\alpha)\rvert < \varepsilon$	0	0	0	0	0	16

Table 2. An evaluation of the spectrum for three of the examples. The entries show the number of real and complex eigenvalues for each matrix as well as the number of eigenvalues close to $0, 1, -1, \alpha$, and $-\alpha$. For these experiments, we took $\varepsilon = 100000\varepsilon_{\text{mach}} \approx 2 \times 10^{-11}$. The graph cs-stan was too large to compute all the eigenvalues and we used ARPACK [17] from Matlab's `eigs` function to compute the top 200 eigenvalues and eigenvectors by largest magnitude.

Graph	$\lVert\mathbf{Y^T M} - \alpha\mathbf{Y^T}\rVert_\mathbf{F}$	$\lVert\mathbf{\tilde{F}^T M} - \alpha\mathbf{\Lambda\tilde{F}^T}\rVert_\mathbf{F}$	$\lVert\mathbf{Z^T P} - \mathbf{Z^T}\rVert_\mathbf{F}$
aa-stan	-	1.7×10^{-15}	-
ee-stan	1.1×10^{-13}	3.4×10^{-13}	8.4×10^{-13}
cs-stan	1.1×10^{-12}	5.2×10^{-13}	1.0×10^{-12}

Table 3. The matrix aa-stan has only a single dominant eigenvalue so the set of eigenvectors in $\mathbf{Y}$ is empty. The notation for the matrices is from section 3 but with $\tilde{\mathbf{F}}^T = \mathbf{F}^T[\mathbf{I} - (1-\alpha)(\mathbf{I} - \alpha\mathbf{P})^{-1}\mathbf{R}]$. The matrix $\mathbf{Z}$ is the set of eigenvectors corresponding to eigenvalue α computed for the matrix $\mathbf{M}$.

7.2 Sensitivity

To confirm theorem 2, we examine the difference $\mathbf{y}(\gamma) - \mathbf{x}(\mathbf{P}, \mathbf{w}(\gamma), \alpha)$ in table 4. The norms of the difference are quite small, demonstrating experimental evidence for the theorem.

Graph	$\gamma = \mathbf{0.001}$	$\gamma = \mathbf{0.01}$	$\gamma = \mathbf{0.1}$
aa-stan	1.72×10^{-10}	1.72×10^{-9}	4.30×10^{-8}
ee-stan	5.62×10^{-11}	5.62×10^{-10}	5.62×10^{-9}
cs-stan	5.31×10^{-11}	5.31×10^{-10}	2.90×10^{-10}
cnr-2000	1.79×10^{-10}	1.79×10^{-9}	5.35×10^{-9}

Table 4. The table entries show the value of $\|\mathbf{y}(\gamma) - \mathbf{x}(\mathbf{P}, \mathbf{w}(\gamma), \alpha)\|_2$ using the notation from section 4.

8 Experiments on the PageRank vector and its derivative

8.1 Does a negative derivative justify a change in ranking?

One of the most promising uses of the derivative vector is to evaluate what happens in the PageRank vector at different values of α. Table 5 shows some preliminary results on this idea where we look at the fraction of pages with negative derivative that actually decrease in rank when increasing α by a value γ. The fraction predicted by the derivative is higher than the average fraction predicted by a random vector. Currently, we do not consider the magnitude of the derivative with these predictions.

8.2 What are the pages with largest derivative?

For the largest strongly connected component of wiki-2006-11, table 6 lists the top 20 pages with largest derivative for a few values of α. Most of the pages that appear in the top 20 list are also highly ranked according to the PageRank value. Additionally, pages in the category namespace in Wikipedia are highly ranked by both PageRank and its derivative for the two largest values of α evaluated.

References

1. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford University (1999)
2. Berkhin, P.: A survey on PageRank computing. Internet Mathematics **2** (2005) 73–120

Graph	$\gamma = \mathbf{0.001}$		$\gamma = \mathbf{0.01}$		$\gamma = \mathbf{0.1}$	
	$\mathbf{x}'$	$\bar{\mathbf{r}}$	$\mathbf{x}'$	$\bar{\mathbf{r}}$	$\mathbf{x}'$	$\bar{\mathbf{r}}$
aa-stan	0.000	0.000	0.000	0.000	0.000	0.000
ee-stan	0.079	0.078	0.286	0.266	0.478	0.453
cs-stan	0.257	0.237	0.441	0.372	0.505	0.432
cnr-2000	0.557	0.477	0.621	0.527	0.641	0.553
wiki-2006-09	0.385	0.362	0.385	0.362	0.361	0.342
wiki-2006-11	0.385	0.361	0.383	0.360	0.360	0.341

Table 5. The $\mathbf{x}'$ entries show the fraction of pages with negative derivative that decreased in rank when increasing α by the value of γ in the table heading. These values are compared with the $\bar{\mathbf{r}}$ entries, which show the average fraction over 50 trials where the derivative is replaced by a random vector generated with `randn`.

$\alpha = \mathbf{0.5}$		$\alpha = \mathbf{0.85}$		$\alpha = \mathbf{0.95}$	
Page	**Rank**	**Page**	**Rank**	**Page**	**Rank**
United States	1	Portal:List	6	Portal:List	2
Race (US Census)	6	C:Main topic classif.	4	C:Main topic classif.	3
C:Categories by country	23	C:Society	3	C:Society	4
United Kingdom	4	C:Political geography	10	Wikipedia	8
2006	3	Wikipedia	24	C:Fundamental	9
England	5	C:Fundamental	22	C:Science	21
Canada	7	C:Geography	25	C:Social sciences	12
2005	8	C:Social sciences	29	C:Geography	10
France	10	C:Politics	27	Portal:Browse	83
C:Society	108	C:Science	49	C:Portals	79
C:People	63	C:Human geography	41	C:Academic disciplines	36
C:Living people	2	C:Countries	28	C:Political geography	5
Germany	12	Human	36	C:Politics	16
Australia	11	C:Business	45	C:Nature	40
2004	9	C:People	16	List of acad. disciplines	63
World War II	18	C:Academic disciplines	98	Human	25
C:Political geography	188	C:Nature	94	C:Humans	41
Japan	14	C:Categories by country	5	Academia	75
Europe	32	C:Geography by place	38	Philosophy	64

Table 6. The columns show the top 20 pages with largest derivative for three values of α computed on the largest strongly connected component in wiki-2006-11. The pages are presented in order of the derivative so “United States” has the largest derivative at $\alpha = 0.50$ and its rank according to PageRank is 1. Pages in the category namespace are abbreviated “C:” instead of the full “Category:”. The page “Category:Main topic classification” is abbreviated “C:Main topic classif.” Likewise ”List of academic disciplines” is abbreviated “List of acad. disciplines.”

3. Langville, A.N., Meyer, C.D.: Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press (2006)
4. Haveliwala, T.H., Kamvar, S.D.: The second eigenvalue of the Google matrix. Technical Report 2003-20, Stanford University (2003) Accessed online, 8 April 2007.
5. Eldén, L.: A note on the eigenvalues of the Google matrix. arXiv e-print math/0401177 (2004)
6. Serra-Capizzano, S.: Jordan canonical form of the Google matrix: A potential contribution to the PageRank computation. SIAM Journal on Matrix Analysis and Applications **27** (2005) 305–312
7. Brauer, A.: Limits for the characteristic roots of a matrix. IV. Duke Mathematics Journal **19** (1952) 75–91
8. Boldi, P., Santini, M., Vigna, S.: PageRank as a function of the damping factor. In: Proceedings of the 14th international conference on the World Wide Web, Chiba, Japan, ACM Press (2005)
9. Golub, G., Greif, C.: An arnoldi-type algorithm for computing PageRank. BIT Numerical Mathematics **46** (2006) 759–771
10. Kamvar, S.D., Haveliwala, T.H., Manning, C.D., Golub, G.H.: Extrapolation methods for accelerating PageRank computations. In: Proceedings of the 12th international conference on the World Wide Web, New York, NY, USA, ACM Press (2003) 261–270
11. Boldi, P., Posenato, R., Santini, M., Vigna, S.: Traps and pitfalls of topic-biased PageRank. In: WAW 2006. Fourth Workshop on Algorithms and Models for the Web-Graph. LNCS, Springer–Verlag (2007)
12. Hirai, J., Raghavan, S., Garcia-Molina, H., Paepcke, A.: Webbase: a repository of web pages. Computer Networks **33** (2000) 277–293
13. Boldi, P., Vigna, S.: Codes for the world wide web. Internet Mathematics **2** (2005) 407–429
14. Boldi, P., Codenotti, B., Santini, M., Vigna, S.: UbiCrawler: A scalable fully distributed web crawler. Software: Practice & Experience **34** (2004) 711–726
15. : Wikipedia XML database dump from september 25, 2006. Accessed from `http://en.wikipedia.org/wiki/Wikipedia:Database_download` (2006)
16. : Wikipedia XML database dump from november 4, 2006. Accessed from `http://en.wikipedia.org/wiki/Wikipedia:Database_download` (2006)
17. Lehoucq, R.B., Sorensen, D.C., Yang, C.: ARPACK User's Guide: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods. (1997)