Google's PageRank

The Math Behind the Search Engine

REBECCA S. WILLS

Approximately 94 million American adults use the Internet on a typical day [24]. The number-one Internet activity is reading and writing e-mail. Search engine use is next in line and continues to increase in popularity. In fact, survey findings indicate that nearly 60 million American adults use search engines on a given day. Even though there are many Internet search engines, Google, Yahoo!, and MSN receive over 81% of all search requests [27]. Despite claims that the quality of search provided by Yahoo! and MSN now equals that of Google [11], Google continues to thrive as the search engine of choice, receiving over 46% of all search requests, nearly double the volume of Yahoo! and over four times that of MSN.

I use Google's search engine on a daily basis and rarely request information from other search engines. One day, I decided to visit the homepages of Google, Yahoo!, and MSN to compare the quality of search results. Coffee was on my mind that day, so I entered the simple query "coffee" in the search box at each homepage. Table 1 shows the top ten (unsponsored) results returned by each search engine. Although ordered differently, two webpages, *www.peets.com* and *www.coffeegeek.com*, appear in all three top ten lists. In addition, each pairing of top ten lists has two additional results in common.

Depending on the information I hoped to obtain about coffee by using the search engines, I could argue that any one of the three returned better results; however, I was not looking for a particular webpage, so all three listings of search results seemed of equal quality. Thus, I plan to continue using Google. My decision is indicative of the problem Yahoo!, MSN, and other search engine companies face in the quest to obtain a larger percentage of Internet search volume. Search engine users are loyal to one or a few search engines and are generally happy with search results [14, 28]. Thus, as long as Google continues to provide results deemed high in quality, Google likely will remain the top search engine. But what set Google apart from its competitors in the first place? The answer is PageRank. In this article I explain this simple mathematical algorithm that revolutionized Web search.

Google's Search Engine

Google founders Sergey Brin and Larry Page met in 1995 when Page visited the computer science department of Stanford University during a recruitment weekend [2, 9]. Brin, a second-year graduate student at the time, served as a guide for potential recruits, and Page was part of his group. They discussed many topics during their first meeting and disagreed on nearly every issue. Soon after he began graduate study at Stanford, Page began working on a Web project, initially called BackRub, that exploited the link structure of the Web. Brin found Page's work on BackRub interesting, so the two started working together on a project that would permanently change Web search. Brin and Page realized that they were creating a search engine that adapted to the ever-increasing size of the Web, so they replaced the name BackRub with Google (a common misspelling of *googol*, the number 10^{100}). Unable to convince existing search engine companies to adopt the technology they had developed but certain their technology was superior to any being used, Brin and Page decided to start their own company. With the financial assistance of a small group of initial investors, Brin and Page founded the Web search engine company Google, Inc. in September 1998.

Almost immediately, the general public noticed what Brin, Page, and others in the academic Web search community already knew—the Google search engine produced much higher-quality results than those produced by other Web search engines. Other search engines relied entirely on webpage content to determine ranking of results, and

Table 1. Top Ten Results for Search Query "coffee" at *www.google.com*, *www.yahoo.com*, and *www.msn.com*, April 10, 2006

Order	Google	Yahoo!	MSN
1	www.starbucks.com (◇)	www.gevalia.com (◇)	www.peets.com (*)
2	www.coffeereview.com (†)	en.wikipedia.org/wiki/Coffee (△)	en.wikipedia.org/wiki/Coffee (△)
3	www.peets.com (*)	www.nationalgeographic.com/coffee	www.coffeegeek.com (*)
4	www.coffeegeek.com (*)	www.peets.com (*)	coffeetea.about.com (△)
5	www.coffeeuniverse.com (†)	www.starbucks.com (◇)	coffeebean.com
6	www.coffeescience.org	www.coffeegeek.com (*)	www.coffeereview.com (†)
7	www.gevalia.com (◇)	coffeetea.about.com (△)	www.coffeeuniverse.com (†)
8	www.coffeebreakarcade.com	kaffee.netfirms.com/Coffee	www.tmcm.com
9	https://www.dunkindonuts.com	www.strong-enough.net/coffee	www.coffeeforums.com
10	www.cariboucoffee.com	www.cl.cam.ac.uk/coffee/coffee.html	www.communitycoffee.com
Approximate Number of Results:			
	447,000,000	151,000,000	46,850,246

Shared results for Google, Yahoo!, and MSN (*); Google and Yahoo! (◇); Google and MSN (†); and Yahoo! and MSN (△)

Brin and Page realized that webpage developers could easily manipulate the ordering of search results by placing concealed information on webpages.[1] Brin and Page developed a ranking algorithm, named PageRank after Larry Page, that uses the link structure of the Web to determine the importance of webpages. During the processing of a query, Google's search algorithm combines precomputed PageRank scores with text-matching scores to obtain an overall ranking score for each webpage.

Although many factors determine Google's overall ranking of search engine results, Google maintains that the heart of its search engine software is PageRank [3]. A few quick searches on the Internet reveal that both the business and academic communities hold PageRank in high regard. The business community is mindful that Google remains the search engine of choice and that PageRank plays a substantial role in the order in which webpages are displayed. Maximizing the PageRank score of a webpage, therefore, has become an important component of company marketing strategies. The academic community recognizes that PageRank has connections to numerous areas of mathematics and computer science such as matrix theory, numerical analysis, information retrieval, and graph theory. As a result, much research continues to be devoted to explaining and improving PageRank.

The Mathematics of PageRank

The PageRank algorithm assigns a PageRank score to each of more than 25 billion webpages [7]. The algorithm models the behavior of an idealized *random Web surfer* [12, 23]. This Internet user randomly chooses a webpage to view from the listing of available webpages. Then, the surfer randomly selects a link from that webpage to another webpage. The surfer continues the process of selecting links at random from successive webpages until deciding to move to another webpage by some means other than selecting a link. The choice of which webpage to visit next does not depend on the previously visited webpages, and the idealized Web surfer never grows tired of visiting webpages. Thus, the PageRank score of a webpage represents the probability that a random Web surfer chooses to view that webpage.

Directed Web Graph

To model the activity of the random Web surfer, the PageRank algorithm represents the link structure of the Web as a directed graph. Webpages are nodes of the graph, and links from webpages to other webpages are edges that show direction of movement. Although the directed Web graph is very large, the PageRank algorithm can be applied to a directed graph of any size. To faciliate our discussion of PageRank, we apply the PageRank algorithm to the directed graph with 4 nodes shown in Figure 1.

AUTHOR

REBECCA S. WILLS received her undergraduate degree from High Point University. Since August 2002 she has been a graduate student at North Carolina State University, working on matrix theory and numerical linear algebra. She enjoys a good cup of coffee in the morning and a sitcom or two in the evening.

Rebecca S. Wills
Department of Mathematics
North Carolina State University
Raleigh, NC 27695
USA
e-mail: rmwills@ncsu.edu

[1] That is, a developer could add text in the same color as the background of the page, invisible to the user but detected by automated search engines. If the terms of a search query occurred many times in the hidden text, that webpage could appear higher in rank than webpages that were really more informative.

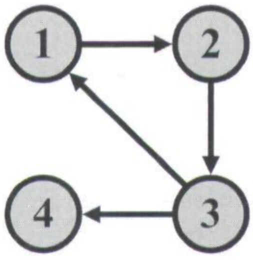

Figure 1. Directed graph with 4 nodes.

Web Hyperlink Matrix

The process for determining PageRank begins by expressing the directed Web graph as the $n \times n$ "hyperlink matrix" H, where n is the number of webpages. If webpage i has $l_i \geq 1$ links to other webpages and webpage i links to webpage j, then the element in row i and column j of H is $H_{ij} = \frac{1}{l_i}$. Otherwise, $H_{ij} = 0$. Thus, H_{ij} represents the likelihood that a random surfer will select a link from webpage i to webpage j. For the directed graph in Figure 1,

$$H = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Node 4 is a *dangling node* because it does not link to other nodes. As a result, all entries in row 4 of the example matrix are zero. This means the probability is zero that a random surfer moves from node 4 to any other node in the directed graph. The majority of webpages are dangling nodes (e.g., postscript files and image files), so there are many rows with all zero entries in the Web hyperlink matrix. When a Web surfer lands on dangling node webpages, the surfer can either stop surfing or move to another webpage, perhaps by entering the Uniform Resource Locator (URL) of a different webpage in the address line of a Web browser. Since H does not model the possibility of moving from dangling node webpages to other webpages, the long-term behavior of Web surfers cannot be determined from H alone.

Dangling Node Fix

Several options exist for modeling the behavior of a random Web surfer after landing on a dangling node, and Google does not reveal which option it employs. One option replaces each dangling node row of H by the same *probability distribution vector*, w, a vector with nonnegative elements that sum to 1. The resulting matrix is $S = H + dw$, where d is a column vector that identifies dangling nodes, meaning $d_i = 1$ if $l_i = 0$ and $d_i = 0$ otherwise; and $w = (w_1\ w_2\ \ldots\ w_n)$ is a row vector with $w_j \geq 0$ for all $1 \leq j \leq n$ and $\sum_{j=1}^{n} w_j = 1$. The most popular choice for w is the uniform row vector, $w = \left(\frac{1}{n}\ \frac{1}{n}\ \ldots\ \frac{1}{n}\right)$. This amounts to adding artificial links from dangling nodes to all webpages. With $w = \left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right)$, the directed graph in Figure 1 changes (see Figure 2).

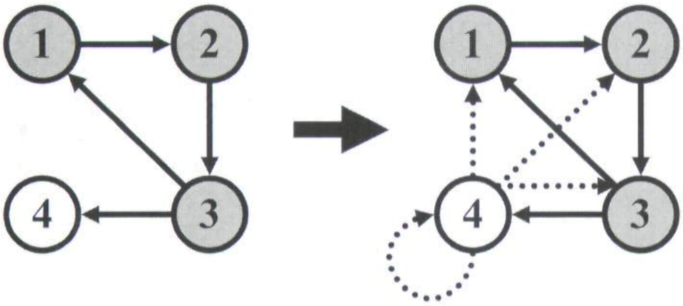

Figure 2. Dangling node fix to Figure 1.

The new matrix $S = H + dw$ is,

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \left(\frac{1}{4}\ \ \frac{1}{4}\ \ \frac{1}{4}\ \ \frac{1}{4}\right)$$

$$= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Regardless of the option chosen to deal with dangling nodes, Google creates a new matrix S that models the tendency of random Web surfers to leave a dangling node; however, the model is not yet complete. Even when webpages have links to other webpages, a random Web surfer might grow tired of continually selecting links and decide to move to a different webpage some other way. For the graph in Figure 2, there is no directed edge from node 2 to node 1. On the Web, though, a surfer can move directly from node 2 to node 1 by entering the URL for node 1 in the address line of a Web browser. The matrix S does not consider this possibility.

Google Matrix

To model the overall behavior of a random Web surfer, Google forms the matrix $G = \alpha S + (1 - \alpha)\mathbb{1}v$, where $0 \leq \alpha < 1$ is a scalar, $\mathbb{1}$ is the column vector of ones, and v is a row probability distribution vector called the *personalization vector*. The *damping factor*, α, in the Google matrix indicates that random Web surfers move to a different webpage by some means other than selecting a link with probability $1 - \alpha$. The majority of experiments performed by Brin and Page during the development of the PageRank algorithm used $\alpha = 0.85$ and $v = \left(\frac{1}{n}\ \frac{1}{n}\ \ldots\ \frac{1}{n}\right)$ [12, 23]. Values of α ranging from 0.85 to 0.99 appear in most research papers on the PageRank algorithm.

Assigning the uniform vector for v suggests Web surfers randomly choose new webpages to view when not selecting links. The uniform vector makes PageRank highly susceptible to *link spamming*, so Google does not use it to determine actual PageRank scores. Link spamming is the practice by some search engine optimization experts of adding more links to their clients' webpages for the sole purpose of increasing the PageRank score of those webpages. This attempt to manipulate PageRank scores is one reason Google does not reveal the current damping factor or personalization vector for the Google matrix. In 2004, however, Gyöngyi, Garcia-Molina, and Pederson developed the TrustRank algorithm to create a personalization vector that decreases the harmful effect of link spamming [17], and Google registered the trademark for TrustRank on March 16, 2005 [6].

Because each element G_{ij} of G lies between 0 and 1 ($0 \leq G_{ij} \leq 1$) and the sum of elements in each row of G is 1, the Google matrix is called a *row-stochastic* matrix. It is known that $\lambda = 1$ is not a repeated eigenvalue of G and is greater in magnitude than any other eigenvalue of G [18, 26]. Hence the eigensystem $\pi G = \pi$ has a unique solution, where π is a row probability distribution vector.[2] We say that $\lambda = 1$ is the *dominant eigenvalue* of G, and π is the corresponding *dominant left eigenvector* of G. The ith entry of π is the PageRank score for webpage i, and π is called the PageRank vector.

Table 2 shows four different Google matrices and their corresponding PageRank vectors (approximated to two decimal places) for the directed graph in Figure 2. The table indicates that the personalization vector has more influence on the PageRank scores for smaller damping factors. For instance, when $\alpha = 0.85$, as is the case for the first and second models, the PageRank scores and the ordering of the scores differ significantly. The first model assigns the uniform vector to v, and node 1 is one of the nodes with the lowest PageRank score. The second model uses $v = (1\ 0\ 0\ 0)$, and node 1 receives the highest PageRank score. This personalization vector suggests that when Web surfers grow tired of following the link structure of the Web, they always move to node 1. For the third and fourth models, $\alpha = 0.95$. The difference in PageRank scores and ordering of scores for these models is less significant. Even though $v = (1\ 0\ 0\ 0)$ in the fourth model, the higher damping factor decreases the influence of v.

Computing PageRank Scores

For small Google matrices like the ones in Table 2, we can quickly find exact solutions to the eigensystem, $\pi G = \pi$. The Google matrix for the entire Web has more than 25 billion rows and columns, so computing the exact solution requires extensive time and computing resources. The oldest and easiest technique for approximating a dominant eigenvector of a matrix is the power method. The power method converges for most starting vectors when the dominant eigenvalue is not a repeated eigenvalue [13, §9.4]. Since $\lambda = 1$ is the dominant eigenvalue of G and π is the dominant left eigenvector, the power method applied to G converges to the PageRank vector. This method was the original choice for computing the PageRank vector.

Given a starting vector $\pi^{(0)}$, e.g. $\pi^{(0)} = v$, the power method calculates successive iterates

$$\pi^{(k)} = \pi^{(k-1)}G, \text{ where } k = 1, 2, \ldots,$$

until some convergence criterion is satisfied. Notice that $\pi^{(k)} = \pi^{(k-1)}G$ can also be stated $\pi^{(k)} = \pi^{(0)}G^k$. As the number of nonzero elements of the personalization vector increases, the number of nonzero elements of G increases. Thus, the multiplication of $\pi^{(k-1)}$ with G is expensive; however, since $S = H + dw$ and $G = \alpha S + (1-\alpha)\mathbb{1}v$, we can express the multiplication as follows:

$$\begin{aligned}\pi^{(k)} &= \pi^{(k-1)}G \\ &= \pi^{(k-1)}\,[\alpha(H + dw) + (1-\alpha)\,\mathbb{1}v] \\ &= \alpha\pi^{(k-1)}\,H + \alpha(\pi^{(k-1)}\,d)\,w + (1-\alpha)\,(\pi^{(k-1)}\,\mathbb{1})\,v \\ &= \alpha\pi^{(k-1)}\,H + \alpha(\pi^{(k-1)}\,d)\,w + (1-\alpha)\,v,\end{aligned}$$

because $\pi^{(k-1)}\mathbb{1} = 1$, $\pi^{(k-1)}$ is a probability vector. This is a sum of three vectors: a multiple of $\pi^{(k-1)}H$, a multiple of w, and a multiple of v. (Notice that $\pi^{(k-1)}d$ is a scalar.) The only matrix-vector multiplication required is with the hyperlink matrix H. A 2004 investigation of Web documents estimates that the average number of outlinks for a webpage is 52 [22]. This means that for a typical row of the hyperlink matrix only 52 of the 25 billion elements are nonzero, so the majority of elements in H are 0 (H is very sparse). Since all computations involve the sparse matrix H and vectors w and v, an iteration of the power method is cheap (the operation count is proportional to the matrix dimension n).

Writing a subroutine to approximate the PageRank vector using the power method is quick and easy. For a simple program (in MATLAB), see Langville and Meyer [20, §4.6].

The ratio of the two eigenvalues largest in magnitude for a given matrix determines how quickly the power method converges [16]. Haveliwala and Kamvar were the first to prove that the second-largest eigenvalue in magnitude of G is less than or equal to the damping factor α [18]. This means that the ratio is less than or equal to α for the Google matrix. Thus, the power method converges quickly when α is less than 1. This might explain why Brin and Page originally used $\alpha = 0.85$. No more than 29 iterations are required for the maximal element of the difference in successive iterates, $\pi^{(k+1)} - \pi^{(k)}$, to be less than 10^{-2} for $\alpha = 0.85$. The number of iterations increases to 44 for $\alpha = 0.90$.

An Alternative Way to Compute PageRank

Although Brin and Page originally defined PageRank as a solution to the eigensystem $\pi G = \pi$, the problem can be restated as a linear system. Recall, $G = \alpha S + (1-\alpha)\,\mathbb{1}v$. Transforming $\pi G = \pi$ to $0 = \pi - \pi G$ gives:

$$\begin{aligned}0 &= \pi - \pi G \\ &= \pi I - \pi(\alpha S + (1-\alpha)\,\mathbb{1}v) \\ &= \pi(I - \alpha S) - (1-\alpha)\,(\pi\mathbb{1})v \\ &= \pi(I - \alpha S) - (1-\alpha)v\end{aligned}$$

The last equality follows as above from the fact that π is a probability distribution vector, so $\pi\mathbb{1} = 1$. Thus

$$\pi(I - \alpha S) = (1-\alpha)v,$$

which means π solves a linear system with coefficient matrix $I - \alpha S$ and right-hand side $(1-\alpha)v$. Since the matrix $I - \alpha S$ is nonsingular [19], the linear system has a unique solution. For more details on viewing PageRank as the solution of a linear system, see [8, 10, 15, 19].

[2] Though not required, the restriction is often made that the personalization vector v and the dangling node vector w have all positive entries that sum to 1 instead of all non-negative entries that sum to 1. Under this restriction, the PageRank vector also has all positive entries that sum to 1.

Table 2. Modeling Surfer Behavior for the Directed Graph in Figure 2

	Damping Factor (α)	Personalization Vector (v)	Google Matrix (G)	Page Rank Vector ($\approx \pi$)	Ordering of Nodes (1 = Highest)
Model 1	0.85	$\left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right)$	$\begin{pmatrix} \frac{3}{80} & \frac{71}{80} & \frac{3}{80} & \frac{3}{80} \\ \frac{3}{80} & \frac{3}{80} & \frac{71}{80} & \frac{3}{80} \\ \frac{37}{80} & \frac{3}{80} & \frac{3}{80} & \frac{37}{80} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$	(0.21 0.26 0.31 0.21)	(3 2 1 3)
Model 2	0.85	(1 0 0 0)	$\begin{pmatrix} \frac{3}{20} & \frac{17}{20} & 0 & 0 \\ \frac{3}{20} & 0 & \frac{17}{20} & 0 \\ \frac{23}{40} & 0 & 0 & \frac{17}{40} \\ \frac{29}{80} & \frac{17}{80} & \frac{17}{80} & \frac{17}{80} \end{pmatrix}$	(0.30 0.28 0.27 0.15)	(1 2 3 4)
Model 3	0.95	$\left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right)$	$\begin{pmatrix} \frac{1}{80} & \frac{77}{80} & \frac{1}{80} & \frac{1}{80} \\ \frac{1}{80} & \frac{1}{80} & \frac{77}{80} & \frac{1}{80} \\ \frac{39}{80} & \frac{1}{80} & \frac{1}{80} & \frac{39}{80} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$	(0.21 0.26 0.31 0.21)	(3 2 1 3)
Model 4	0.95	(1 0 0 0)	$\begin{pmatrix} \frac{1}{20} & \frac{19}{20} & 0 & 0 \\ \frac{1}{20} & 0 & \frac{19}{20} & 0 \\ \frac{21}{40} & 0 & 0 & \frac{19}{40} \\ \frac{23}{80} & \frac{19}{80} & \frac{19}{80} & \frac{19}{80} \end{pmatrix}$	(0.24 0.27 0.30 0.19)	(3 2 1 4)

Google's Toolbar PageRank

The PageRank score of a webpage corresponds to an entry of the PageRank vector, π. Since π is a probability distribution vector, all elements of π are non-negative and sum to one. Google's toolbar includes a PageRank display feature that provides "an indication of the PageRank" for a webpage being visited [5]. The PageRank scores on the toolbar are integer values from 0 (lowest) to 10 (highest). Although some search engine optimization experts discount the accuracy of toolbar scores [25], a Google webpage on toolbar features [4] states:

> PageRank Display: Wondering whether a new website is worth your time? Use the Toolbar's PageRank™ display to tell you how Google's algorithms assess the importance of the page you're viewing.

Results returned by Google for a search on Google's toolbar PageRank reveal that many people pay close attention to the toolbar PageRank scores. One website [1] asserts that website owners have become addicted to toolbar PageRank.

Although Google does not explain how toolbar PageRank scores are determined, they are possibly based on a logarithmic scale. It is easy to verify that few webpages receive a toolbar PageRank score of 10, but many webpages have very low scores.

Two weeks after creating Table 1, I checked the toolbar PageRank scores for the top ten results returned by Google for the query "coffee." Those scores are listed in Table 3. They reveal a point worth emphasizing. Although PageRank is an important component of Google's overall ranking of results, it is not the only component. Notice that *https://www.dunkindonuts.com* is the ninth result in Google's top ten list. There are six results considered more relevant by Google to the query "coffee" that have lower toolbar PageRank scores than *https://www.dunkindonuts.com*. Also, Table 1 shows that both Yahoo! and MSN returned *coffeetea.about.com* and *en.wikipedia.org/wiki/Coffee* in their top ten listings. The toolbar PageRank score for both webpages is 7; however, they appear in Google's listing of results at 18 and 21, respectively.

Since a high PageRank score for a webpage does not guarantee that the webpage appears high in the listing of search results, search engine optimization experts emphasize that "on the page" factors, such as placement and fre-

Table 3. Toolbar PageRank Scores for the Top Ten Results Returned by *www.google.com* for April 10, 2006, Search Query "coffee"

Order	Google's Top Ten Results	Toolbar PageRank
1	www.starbucks.com	7
2	www.coffeereview.com	6
3	www.peets.com	7
4	www.coffeegeek.com	6
5	www.coffeeuniverse.com	6
6	www.coffeescience.org	6
7	www.gevalia.com	6
8	www.coffeebreakarcade.com	6
9	https://www.dunkindonuts.com	7
10	www.cariboucoffee.com	6

quency of important words, must be considered when developing good webpages. Even the news media have started making adjustments to titles and content of articles to improve rankings in search engine results [21]. The fact is most search engine users expect to find relevant information quickly, for any topic. To keep users satisfied, Google must make sure that the most relevant webpages appear at the top of listings. To remain competitive, companies and news media must figure out a way to make it there.

Want to Know More?

For more information on PageRank, see the survey papers by Berkhin [10] and Langville and Meyer [19]. In addition, the textbook [20] by Langville and Meyer provides a detailed overview of PageRank and other ranking algorithms.

ACKNOWLEDGMENTS

Many people reviewed this article, and I thank each of them. In particular, I thank Ilse Ipsen and Steve Kirkland for encouraging me to write this article. I thank Ilse Ipsen and my fellow "Communicating Applied Mathematics" classmates, Brandy Benedict, Prakash Chanchana, Kristen DeVault, Kelly Dickson, Karen Dillard, Anjela Govan, Rizwana Rehman, and Teresa Selee, for reading and re-reading preliminary drafts. Finally, I thank Jay Wills for helping me find the right words to say.

REFERENCES

[1] *www.abcseo.com/seo-book/toolbar-google.htm*, Google Toolbar PageRank.

[2] *http://www.google.com/corporate/history.html*, Google Corporate Information: Google Milestones.

[3] *http://www.google.com/technology/index.html*, Our Search: Google Technology.

[4] *http://www.google.com/support/toolbar/bin/static.py?page=features.html&hl=en*, Google Toolbar: Toolbar Features.

[5] *http://toolbar.google.com/button_help.html*, Google Toolbar: About Google Toolbar Features.

[6] *http://www.uspto.gov/main/patents.htm*, United States Patent and Trademark Office official website.

[7] *http://www.webrankinfo.com/english/seo-news/topic-16388.htm*, January 2006, Increased Google Index Size?

[8] Arvind Arasu, Jasmine Novak, Andrew Tomkins, and John Tomlin, *PageRank Computation and the Structure of the Web: Experiments and Algorithms*, Technical Report, IBM Almaden Research Center (November 2001).

[9] John Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, Penguin Group, New York, 2005.

[10] Pavel Berkhin, A survey on PageRank computing, *Internet Mathematics* **2** (2005), no. 1, 73–120.

[11] Celeste Biever, Rival engines finally catch up with Google, *New Scientist* **184** (2004), no. 2474, 23.

[12] Sergey Brin and Lawrence Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* **33** (1998), 107–117.

[13] Germund Dahlquist and Åke Björck, *Numerical Methods in Scientific Computing*, vol. II, SIAM, Philadelphia, to be published, http://www.math.liu.se/~akbjo/dqbjch9.pdf.

[14] Deborah Fallows, Search engine users, Pew Internet & American Life Project Report, January 2005.

[15] David Gleich, Leonid Zhukov, and Pavel Berkhin, Fast Parallel PageRank: A Linear System Approach, Technical report, WWW2005.

[16] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996.

[17] Zoltán Gyöngyi, Hector Garcia-Molina, and Jan Pedersen, *Combating Web spam with TrustRank*, Proceedings of the 30th International Conference on Very Large Databases, Morgan Kaufmann, 2004, pp. 576–587.

[18] Taher H. Haveliwala and Sepandar D. Kamvar, *The Second Eigenvalue of the Google Matrix*, Technical report, Stanford University, 2003.

[19] Amy N. Langville and Carl D. Meyer, Deeper inside PageRank, *Internet Mathematics* **1** (2004), no. 3, 335–380.

[20] ——, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.

[21] Steve Lohr, This boring headline is written for Google, *The New York Times*, April 2006.

[22] Anuj Nanavati, Arindam Chakraborty, David Deangelis, Hasrat Godil, and Thomas D'Silva, *An investigation of documents on the World Wide Web*, http://www.iit.edu/~dsiltho/Investigation.pdf, December 2004.

[23] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd, The PageRank citation ranking: Bringing order to the Web, Technical report, Stanford University, 1998.

[24] Lee Rainie and Jeremy Shermak, Big jump in search engine use, Pew Internet & American Life Project Memo, November 2005.

[25] Chris Ridings and Mike Shishigin, PageRank uncovered, Technical Paper for the Search Engine Optimization Online Community.

[26] Stefano Serra-Capizzano, Jordan canonical form of the Google matrix: A potential contribution to the PageRank computation, *SIAM J. Matrix Anal. Appl.* **27** (2005), no. 2, 305–312.

[27] Danny Sullivan, Nielsen NetRatings search engine ratings, *Search Engine Watch*, January 2006.

[28] Danny Sullivan and Chris Sherman, *Search Engine User Attitudes*, iProspect.com, Inc., May 2005.