PageRank for Ranking Authors in Co-citation Networks

Ying Ding and Erjia Yan

School of Library and Information Science, Indiana University, 1320 East 10th Street, Bloomington, IN 47405-3907. E-mail: {dingying, eyan}@indiana.edu

Arthur Frazho

School of Aeronautics and Astronautics, Purdue University, 701 West Stadium Avenue, ARMS 3201, West Lafayette, IN 47907-2045. E-mail: arthur.e.frazho.1@purdue.edu

James Caverlee

Department of Computer Science, Texas A&M University, TAMU 3112, College Station, TX 77843-3112. E-mail: caverlee@cs.tamu.edu

This paper studies how varied damping factors in the PageRank algorithm influence the ranking of authors and proposes weighted PageRank algorithms. We selected the 108 most highly cited authors in the information retrieval (IR) area from the 1970s to 2008 to form the author co-citation network. We calculated the ranks of these 108 authors based on PageRank with the damping factor ranging from 0.05 to 0.95. In order to test the relationship between different measures, we compared PageRank and weighted PageRank results with the citation ranking, h-index, and centrality measures. We found that in our author co-citation network, citation rank is highly correlated with PageRank with different damping factors and also with different weighted PageRank algorithms; citation rank and PageRank are not significantly correlated with centrality measures; and h-index rank does not significantly correlate with centrality measures but does significantly correlate with other measures. The key factors that have impact on the PageRank of authors in the author co-citation network are being co-cited with important authors.

Introduction

PageRank is a key Web information retrieval algorithm and apparently plays an important part in Google's Web search success. PageRank works on the premise that the importance of any Web page can be judged by the pages that link to it (Brin & Page, 1998). It assumes Web hyperlinks to be "trust votes" and ranks search results based on these links interlinking them. PageRank provides additional insights into traditional keyword-based ranking, and further inspires a variety of new ranking algorithms proposed to improve PageRank (Langville & Meyer, 2006).

There is a long history of citation research originating from the 1940s that judges the quality of a paper based on the number of times it has been cited. Before the advent of the Web, printed journals, magazines, and conference proceedings were the main publication channels for academic scholars. Citations, functioning as hyperlinks, interlink scholarly publications and form scholarly graphs. Citation analysis constructs a useful way to analyze and rank documents, authors, or journals (Small, 1973; White & McCain, 1998). In their pioneering study, Brin and Page (1998) emphasized the important link between PageRank and citation analysis, stating "Academic citation literature has been applied to the web, largely by counting citations or backlinks to a given page. This gives some approximation of a page's importance or quality. PageRank extends this idea by not counting links from all pages equally, and by normalizing by the number of links on a page." (p. 109).

In bibliometrics, the number of citations is an indicator used to measure the impact of scientific publications. This measurement, however, has one drawback in that it does not count the importance of the citing papers: a citation from a groundbreaking, highly cited work (Maslov & Redner, 2008). This limitation can be alleviated by the PageRank algorithm, since it gives higher weights to the publications that are highly cited (e.g., publications having more inlinks) and also to papers cited by a few highly cited papers (e.g., publications linked by a few "important" papers). PageRank is therefore chosen as a complementary method to citation analysis, which allows us to identify publications referenced by highly cited articles.

Received January 17, 2009; revised April 8, 2009; accepted June 8, 2009

^{© 2009} ASIS&T • Published online 13 July 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21171

Two features of the original PageRank algorithm motivate the work in this paper. First, the damping factor, which represents the probability that a Web surfer randomly jumps from one page to another page, plays a fundamental role in the PageRank algorithm. It allows the use of a fixed point algorithm to compute the vector corresponding to the PageRank. The damping factor in PageRank is set at 0.85, meaning that there is an 85% chance that a random surfer will follow the links provided by the present page. This leaves a 15% chance that a random surfer will jump to a completely new page that has no links from the previously surfed pages.

Second, the original PageRank algorithm evenly distributes the PageRank score of one node¹ among its outbound links. Even though Google's PageRank algorithm has proven to be very successful, this even distribution of weights to the outbound links does not necessarily reflect the actual circumstances of the real world, as not all outbound links should have the same importance. Differences in the quality of Web pages, researchers, and journals do exist. Most of these domains reflect the power-law distribution phenomena, which indicate that a few nodes are important and the majority of nodes have very low importance. Hence, adding weights to the PageRank algorithm to reflect this observation has recently begun to attract some research interest.

The co-citation network contains three components: nodes that can be authors (e.g., author co-citation networks), papers (e.g., paper co-citation networks), and journals (e.g., journal co-citation networks); edges which link nodes when these nodes are co-cited; and co-citation frequencies of pairs of nodes as weights for the edges. The co-citation network forms an undirected and weighted graph. Although PageRank is originally designed for directed graphs, it can be applied to undirected graphs (Perra & Fortunato, 2008). In this paper we use the PageRank algorithm with different damping factors to rank authors based on an author co-citation network, where authors are nodes, co-cited links are edges, and the author co-citation frequencies are weights for the edges. We also propose weighted PageRank algorithms, test these based on the same co-citation network, and compare the results with other traditional bibliometric and social network measures. This research places special emphasis on whether different values of damping factors in PageRank algorithm can affect author ranks, and whether the proposed weighted PageRank will rank authors differently comparing it with the results of other related ranking methods, such as traditional citation analysis (e.g., citation ranks, h-index), and social network analysis (e.g., centrality measures).

This study has several limitations: (1) it tests the correlations among weighted PageRank algorithms, the original PageRank algorithm, centrality measures, and citation ranks, but it does not investigate which method is "better" than others; (2) applying PageRank algorithms to undirected graphs is justified based on the existing literature, the details about how to apply PageRank to the author co-citation network is discussed in the Methodology part, but the sound mathematical grounding and proof are not included in this study. This paper is organized as follows: The first section introduces the topic; the second section discusses related works; the third section explains the methods used in this paper; the fourth section illustrates the results; the fifth section compares these results with other related methods in bibliometrics and social network analyses; and the final section concludes the research and identifies future works.

Related Work

A page having a high PageRank has many pages pointing to it, or a page with high PageRank points to it. Intuitively, pages that are highly cited are worth browsing and pages that are cited by high PageRank pages are also worth reading. PageRank handles both cases recursively by propagating weights through the link structure of the Web (Maslov & Redner, 2008).

Here we present a brief review of the PageRank algorithm. To begin, we say that x is a *Markov vector* in \mathbb{R}^{v} if x is a column vector of length v and the sum of all the components of x equals one. Moreover, T is a *Markov matrix* on \mathbb{R}^{v} if T is a $v \times v$ matrix whose entries are all positive and each column sums to one. The PageRank algorithm can be viewed as a state space system of the form:

$$x(n+1) = dTx(n) + b \tag{1}$$

The state x(n) is a Markov vector of length v, and T is a Markov matrix on \mathbb{R}^{ν} , while d, the damping factor, is a scalar in [0, 1). Furthermore, b is a vector of length v consisting of all positive numbers that sum to 1-d. The initial condition x(0) is a Markov vector in \mathbb{R}^{ν} . This guarantees that the state x(n) is also a Markov vector. In PageRank, the components T_{ik} of T represent the probability that a random surfer will move from Webpage k to Webpage j following one of the links in Webpage k. The damping factor d is the probability that a random surfer will follow one of the links on the present page to go to another Webpage. 1-d is the probability that the surfer will open up a new page independent of the links on the existing page. The *jth* component $x_i(n)$ of x(n)is the probability that the random surfer is at the *jth* node at time *n*. The *jth* component b_i of *b* is the probability that the random surfer will choose the *jth* Webpage at random without following any links. Because T is a Markov matrix, all the eigenvalues for T are contained in the closed unit disc, which guarantees that dT is stable.

The choice of damping factor is empirical. Google PageRank sets its damping factor to 0.85, which offers a speedy convergence of the power method (Brin & Page, 1998). Boldi,

¹Graphs contain nodes and edges. Nodes of graphs are also called vertexes of graphs, which can be authors in coauthorship networks, Websites in hyperlinked Web graphs, or publications in cited paper graphs. Edges play a role in linking different nodes; for example, edges in coauthorship networks are collaboration relations (e.g., writing papers together) between two authors; edges in Web graphs are linking relations (e.g., one Website links to another Website) between two Websites; edges in paper graphs can be citing relations (e.g., one paper cites another paper) between two papers.

Santini, and Vigna (2005) provided the mathematical analysis of PageRank when the damping factor d changes, finding that, contrary to popular belief, for real-world graphs, values of damping factor d close to 1 do not give a more meaningful ranking than other high damping factors. When d takes a value close to 1, it slows down the convergence of the power method. This places much greater emphasis on the link structure of the Web and much less on the random tendencies of surfers (Langville & Meyer, 2006). When d = 0.99 and d = 0.85, their corresponding PageRank can be vastly different. Pretto (2002) found that when d changes, the top section of the ranking changes only slightly, while the lower section of the ranking varies dramatically. As d becomes smaller, the impact of the probability part of b in Formula (1) increases, while the part of Formula (1) that represents the topology of the graph decreases.

PageRank provides a computationally simple and effective way to evaluate the relative importance of publications beyond citation counts. Researchers use weights to solve the issue of even distribution of PageRank scores among the nodes' outbound links. There are several ways to give weights to PageRank. According to Formula (1), weights can be applied to T or b. The majority of the weighted PageRank literature focuses on adding weights to T by using various normalizations or application-based recalculations of T. Xing and Ghorbani (2004) added weights to the links based on the reference pages, differentiating between inbound and outbound link weight. Their simulation tests show that weighted PageRank performs better than the original PageRank in terms of returning larger numbers of relevant pages to a given query. Aktas, Nacar, and Menczer (2004) proposed a weighted PageRank based on user profiles so that weights are added to certain Internet domains users prefer. Yu, Li, and Liu (2004) added the temporal dimension to the PageRank formula and called it "TimedPageRank." They used citations and publications to define the TimedPageRank formula by weighting each citation according to its date. Bollen, Rodriguez, and Van de Sompel (2007) proposed weighted PageRank for ranking journals and compared them with their ISI journal Impact Factors (IF). They identified discrepancies between ISI IF and the weighted PageRank in the top 10 rankings, especially for computer science. Walker, Xie, Yan, and Maslov (2007) proposed a CiteRank algorithm by introducing two parameters: the inverse of the average citation depth and the time constant of the bias toward more recent publications. CiteRank thus shows current popularity, whereas PageRank corresponds to what may be termed its "lifetime achievement awards" (similar to h-index). Liu, Bollen, Nelson, and Sompel (2005) defined AuthorRank, a modification of PageRank that considers link weights among the coauthorship links. Other work aimed at improving PageRank in the context of author ranking includes Sidiropoulos and Manolopoulos (2005) and Fiala, Rousselot, and Ježek (2008).

Based on the literature review, we did not find related research on analyzing different damping factors for ranking authors based on co-citation networks. Most of the damping factor analyses focus on improving the performance of information retrieval. There are studies using PageRank in bibliometric and scientometric research, but few of them concentrate on analyzing the impact of different values of damping factors for ranking purposes. Also, while most of the weighted PageRank studies try to add weights to the Tmatrix in Formula (1), few of them place special attention on weighting the b part. This paper thus aims to fill in these gaps by testing different values of damping factors ranging from 0.05 to 0.95 for ranking authors in an author co-citation network, proposing the weighted PageRank algorithms that add weights to both T matrix and b part of the Formula (1), and comparing the results with other traditional methods from bibliometrics and social network analysis.

Methodology

Data Collection

Information Retrieval (IR) was selected as the testing field. Papers and their citations were collected from Web of Science (WOS) in the period from the 1970s to 2008. Based on carefully selected search terms related to the IR area (achieved by checking the Library Congress Subject Heading [LCSH] and consulting several domain experts), we searched for the following terms, including their plurals and spelling variations: INFORMATION RETRIEVAL, INFORMA-TION STORAGE and RETRIEVAL, QUERY PROCESS-ING, DOCUMENT RETRIEVAL, DATA RETRIEVAL, IMAGE RETRIEVAL, TEXT RETRIEVAL, CONTENT BASED RETRIEVAL, CONTENT-BASED RETRIEVAL, DATABASE QUERY, DATABASE QUERIES, QUERY LANGUAGE, OUERY LANGUAGES, and RELEVANCE FEEDBACK. In total, we collected 15,370 papers with 341,871 citations. These citation records only contain the first author, year, source, volume, and page number. We selected the most highly cited authors-those with more than 200 citations from the citation records-to arrive at 108 authors (some authors have the same number of citations). We calculated the co-citation frequency among each pair of authors, forming a 108-by-108 matrix.

PageRank for Undirected Graphs

PageRank is a graph-based ranking algorithm to determine the importance of a vertex within a graph by considering both its inbound links and outbound links. Although PageRank was originally designed for directed graphs, it can be applied to undirected graphs (Mihalcea, 2004, Perra & Fortunato, 2008). In an undirected graph, the out-degree of a vertex is equal to its in-degree. PageRank for undirected graphs has been used in computational linguistics, such as text summarization (Mihalcea, 2004), sentence extraction (Wang, Liu, & Wang, 2007), and word sense disambiguation (Mihalcea, Tarau, & Figa, 2004).

Adjacency matrix A is the basic matrix of a graph with the element A_{ij} equal to 1 if node *i* and *j* are connected by a link and 0 if they are not. If the network is a directed graph,



FIG. 1. One author co-citation network example.

the adjacency matrix is not necessarily symmetric (Perra & Fortunato, 2008). We used author co-citation networks where nodes present authors, edges represent the co-cited relations among authors, and the weights of the edges represent the co-citation frequency among these authors. This graph is undirected and weighted. For instance, if author j and author k are co-cited three times together, then it is interpreted in the author co-citation graph as the node of author j and the node of author k having one edge in between with the weight of 3. In this case, our entry for the co-citation matrix A is $A_{jk} = A_{kj} = 3$. We set all the diagonal elements of A to zero. T is the Markov matrix, which is a transition probability matrix obtained by normalizing all the columns of A to sum to one. Although an author co-citation network is an undirected graph, the normalized A matrix (also called transition matrix T) is not symmetric (see Figure 1). The "in-degree" and "out-degree" of a node is thus no longer the same and the "direction" of the links is indirectly indicated.

For example, there are four authors (A, B, C, D). A and B are co-cited 3 times, B and D 5 times, A and D 4 times, and C and D 1 time. The author co-citation graph is shown in Figure 1. It also displays the corresponding co-citation matrix A and its normalized derivation (a transition matrix T). Nodes with more links and high weights will get high PageRank scores. This can be interpreted in the author co-citation graph that authors co-cited with many other authors (which indicates "more links"), and co-cited many times with these authors (which indicates "high weights") will obtain high PageRank scores. Also, nodes linked with important nodes will get high PageRank scores. This means that in the author co-citation graph, authors who are co-cited with important authors will get high PageRank scores as well. Applying the PageRank algorithm to rank authors based on author cocitation networks considers the topology of the co-citation graphs.

Damping Factor

The damping factor d is the probability that the random surfer will follow a link on the existing Webpage. The random surfer, however, still has a (1-d) chance to start a completely new page. In author co-citation network, a high damping factor means that the random surfer has a high chance of following the internal co-citation links, and a low chance of starting with a totally new author (this author has no co-cited links with previously clicked authors). Choosing a different d can result in different ranking results (Boldi et al., 2005). A low damping factor means that every author has more or less the same chance (the probability roughly equals to 1/N where N is number of nodes in the graph) of being selected by the random surfer. Choosing d close to 1 (but not too close to 1) should give a "better" PageRank, but will also significantly increase the computing complexity (Boldi et al., 2005).

Weighted PageRank

The PageRank state equation Formula (1) consists of two parts: dTx(n) + b. The first part dTx(n) models the linking structure (through weight distribution) of citation networks, and the second part *b* represents the equal probability of getting cited. To add weights to PageRank, we not only add weights to the *T* matrix through the normalization of the column (which converts the adjacency matrix into an asymmetric matrix), but to the *b* part as well. Nodes with high publications or citations will therefore have a high probability of getting cited. We propose our weighted PageRank in Formula (2):

$$PR_W(p) = (1-d)\frac{w(p)}{\sum_{i=1}^N w(p_i)} + d\sum_{i=1}^k \frac{PR(p_i)}{C(p_i)}$$
(2)

where W(p) is the weight matrix for the nodes in the graph. Each node may have different weights, which can be the number of publications of the node, or the number of citations of the node. This approach can be generalized for any other meaningful weights, e.g., h-index of the node, and centrality measures of the node. In this study we tested Formula (2) with the weights as the publications of the authors and the citations of the authors. We used MatLab to calculate the PageRank and weighted PageRank for these 108 authors based on the author co-citation network.

Results and Discussion

This section contains two parts. The first part studies PageRank with different damping factors and the second part deals with different weighted PageRank algorithms.

Damping Factors

Figure 2 shows the ranks of the top 20 highly cited authors. The X axis represents different damping factors starting from 0.005 and then ranging from 0.05 to 0.95 in steps of 0.1. The Y axis represents the ranking positions ranging from Rank 1 to Rank 55. We found that ranking positions are quite different when the damping factor *d* changes. As expected, when d = 0.005, the ranking is very different when compared to the rest. It shows that when *d* is low, each author has more or less the same PageRank score, which is close to 1/108 = 0.0093. Even for this case, Salton G (p = .0096), Abiteboul S (p = 0.0094) from INRIA (France), and Robertson SE (p = 0.0094) from Microsoft Research



FIG. 2. Ranking of top 20 highly cited authors based on PageRank with different damping factors.

Cambridge (UK) are ranked as the top three authors. Their PageRank scores are much higher than average (average p = 0.0093). If the damping factor *d* is very low and certain authors are still ranked highly, it suggests that these authors really have a great influence in the network.

Salton G and Robertson SE remain very stable for their ranking throughout different damping factors, while the ranking position of Abiteboul S drops from Rank 3 (d = 0.45) to Rank 13 (d = 0.95) when the value of the damping factor increases. Based on the theory of PageRank algorithm, this phenomenon can be explained as follows: although Abiteboul S is co-cited with many other authors, his author co-citation frequency with Salton G is very low (only 14 times). When the value of damping factor increases, the citing structure becomes dominant, which means that the number of cocitations with Salton G can increase the PageRank score significantly. Abiteboul S is not co-cited frequently with Salton G, therefore his PageRank drops as d increases.

Belkin NJ and Van Rijsbergen CJ stay stable with ranking Rank 3, 4, or 5 throughout different damping factors. But the rank of Rui Y decreases from Rank 6 to Rank 15, because he is only co-cited with Abiteboul S 4 times, Robertson SE 14 times, and Salton G 228 times. While he is more often co-cited with Smith JR 593 times (ranked Rank 11 based on citation counting), Cox IJ 335 times (ranked Rank 29 based on citation counting), and Smeulders AW 237 times (ranked Rank 31 based on citation counting), with the exception of Smith JR, none of these are significant nodes in this author co-citation network. Spink A's rank increases from Rank 13 to Rank 8. Spink A is significantly co-cited with Saracevic T (1,092 times), Belkin NJ (741 times), and Jansen BJ (577 times). Her co-citations with the top three highly cited authors are Salton G (317 times), Robertson SE (101 times), and Abiteboul S (4 times). She is ranked Rank 9 by citation counting, but her ranking changes with different damping factor: Rank 13 (d = 0.05), Rank 12 (d = 0.35), Rank 11 (d = 4.55), Rank 9 (d = 0.85), and Rank 8 (d = 0.95). She is well connected to the important authors (the top five highly cited authors: Salton G, Robertson SE, Abiteboul S, Belkin NJ, and Van Rijsbergen CJ). When the link structure of the network is considered (when d is high), her rank improves. The same is true for Jones KS, who is ranked Rank 10 based on citations. While considering the topology of the network, she is ranked Rank 7. One might infer that she is co-cited frequently with other important authors, and after checking the data this turns out to be true. The top three co-cited authors with her are Salton G (1447 times), Robertson SE (897 times), and Van Rijsbergen CJ (483 times).

Smith JR drops from Rank 8 (d = 0.05) to Rank 16 (d = 0.95). Following the same analysis, we find that the top three co-cited authors with Smith—Rui Y (593 times), Pentland A (341 times), and Flickener M (319 times)—are not "important" authors. The rank of Faloutsos C also drops from Rank 10 to Rank 26 as *d* increases. The top three co-cited authors with him are Berchtold D (391 times), Salton G (200 times), and Smith JR (170 times). Harman D is quite stable ranking either Rank 11 or Rank 12, where her top three co-cited authors are Salton G (912 times), Robertson

SE (463 times), and Voorhees EM (457 times). Voorhees EM remains stable as well. By following the previous examples, we can assume that his top three co-cited authors (Salton G [745 times], Robertson SE [328 times], and Harman D [279 times]) should include some of the top highly cited authors. The rank of Flickner M drops significantly from Rank 18 to Rank 34. His co-citation status with the top three highly cited authors are Salton G (83 times), Robertson SE (5 times), and Abiteboul S (8 times), while Flickner M's top three highly co-cited authors are Smith JR (319 times), Rui Y (265 times), and Pentland A (222 times). This explains why his ranking drops. Bates MJ stays stable. Codd EF changes from Rank 14 to Rank 38, while his citation rank is Rank 17. Codd EF is co-cited with Salton G (56 times), Robertson SE (6 times), and Abiteboul S (286 times). His top three co-cited authors are Abiteboul S (286 times), Date CJ (238 times), and Ullman JD (213 times). The ranking for Baezayates R is unusual: his citation rank is Rank 18 yet all his PageRank scores are very low, ranging from Rank 34 to Rank 55. He is co-cited with Salton G (302 times), Robertson SE (94 times), and Abiteboul S (62 times). His top three co-cited authors are Salton G (302 times), Fuhr N (96 times), and Robertson SE (94 times). His co-citations with the rest of 107 authors are evenly distributed. The main reason for his high citation rank may be his important book, Modern Information Retrieval, published in 1999 by Addison-Wesley. This book is ranked Rank 4 among most cited IR books. The other top three highly cited books are Salton G's Introduction to Modern Information Retrieval (1983), Automatic Information Organization and Retrieval (1968), and Van Rijsbergen CJ's Information Retrieval (1979). Fuhr N is also very stable, as is Jain AK. In order to find out whether different document types (mainly books or journal papers) will attract more citations in our 108 author set, we checked the top 100 highly cited documents with the authors coming from this 108 author set. We found that journal papers attract more citations (around twice as many) than books. These highly cited journal papers are not being necessarily review articles. They are mainly seminal papers that offer innovations on systems or methods.

Table 1 summarizes the changes of the authors and their PageRank with different damping factors. If the author's top three co-cited authors include some of the top five highly cited authors, their PageRank rankings either remain stable or increase when the damping factor increases (e.g., Spink A, Croft WB, and Voorhees EM). Author rank dropping in PageRank can be partially interpreted as their co-citation with the top five highly cited authors are not significant compared with their top three co-cited authors (e.g., Abiteboul S, Rui Y, Smith JR, Faloutsos C, Flickner M, and Codd EF).

Weighted PageRank

For the weighted PageRank, we first use the citations of the 108 authors as weights. Figure 3 shows the scatterplot of the weighted PageRank on citation (denoted as PR_c) of 108 authors with different damping factors ranging from 0.05 to 0.95. The X axis is the ranked top 108 highly cited authors with the number representing the rank of the author based on citations. It indicates that the weighted PageRank on citations does not vary in relation to different damping factors (with average Spearman r = 0.90, p < 0.01). Also, the ranks of the authors, based on weighted PageRank on citation, converge with the ranks of the authors, based on citation counts.

Figure 4 shows the weighted PageRank of the top 20 highly cited authors and its different damping factors. We see some discrepancy between weighted PageRank and the citation ranks. Notably, Abiteboul S, Rui Y, Faloutsos C, Flickner M, and Baezayates R rank lower in weighted PageRank than their corresponding citation ranks, and all of them have the corresponding drops identified in Table 1. The top three highly cited authors among these 108 authors are Salton G (average PR = 0.04357 vs. average PR_c = 0.0841), Robertson SE (average PR = 0.0224 vs. average PR_c = 0.0325), and Abiteboul S (average PR = 0.0160 vs. average PR_c = 0.0217). All of their PR_c have significant increases, which can be partially explained by the weight matrix, as the "citation effect" has been counted twice—once in the *b* part and once in the *T* part of Formula (1).

We test another weighted PageRank using the publications of the 108 authors (we only count the publications in which the 108 authors are the first authors) as the weights. Figure 5 shows the scatterplots of the weighted PageRank of 108 authors with different damping factors ranging from 0.05 to 0.95. It indicates that the weighted PageRank on publications (denoted as PR_p) does not vary according to different damping factors (with Spearman r = 0.9, p < 0.01). But the ranks of the authors based on weighted PageRank on publications are not consistent with the ranks of the authors based on citation counts.

Figure 6 shows the weighted PageRank on publication of the top 20 highly cited authors with different damping factors. Again, we see some discrepancy between weighted PageRank on publication ranks and the citation ranks. Notably, Abiteboul S, Rui Y, Faloutsos C, Flickner M, and Codd EF rank lower in the weighted PageRank on publication than their corresponding citation ranks, and all of them have the corresponding drops identified in Table 1. While Belkin NJ, Spink A, Harman D, Bates MJ, and Fuhr N's weighted PageRank on publication ranks are higher than their corresponding citation ranks, all of them either increased or stayed stable in Table 1. Based on the publication ranks, the top three highly productive authors are Salton G (average PR = 0.0436 vs. average $PR_p = 0.0778$), Spink A (average PR = 0.0133 vs. average $PR_p = 0.0382$), and Chen HC (average PR = 0.0084 vs. average $PR_p = 0.0242$). All of them have significant increases when comparing their original PageRank with weighted PageRank on publication.

Comparison With Other Measures

In social network analysis there are already welldeveloped methods to measure the importance of the nodes in the network (Freeman, 1979). Several studies ranked authors via centrality values and compared them with scientometric

| | Co-cited | with the top 3 highly | cited authors | | Top 3 co-cited authors | | |
|----------------------|----------|-----------------------|---------------|---------------------|--------------------------|-------------------------|---------------------------------|
| Highly cited authors | Salton G | Robertson SE | Abiteboul S | Top 1 | Top 2 | Top 3 | PageRank |
| SALTON G | 12337 | 2343 | 68 | Robertson SE (2343) | Van Rijsbergen CJ (1832) | Croft WB (1768) | Stable (always No. 1) |
| ROBERTSON SE | 2343 | 3006 | 14 | Salton G (2343) | Van Rijsbergen CJ (982) | Jones SK (897) | Stable (No. $3 \text{ or } 2$) |
| ABITEBOUL S | 14 | 68 | 2482 | Hull R (533) | Buneman P (487) | Ullman JD (400) | Drop (No.2 to 13) |
| BELKIN NJ | 788 | 595 | 11 | Saracevic T (1041) | Ingwersen P (828) | Salton G (788) | Stable (No. 4 or 3) |
| VANRIJSBERGEN CJ | 1832 | 982 | 12 | Salton G (1832) | Robertson SE (982) | Croft WB (634) | Stable (No. 5 or 4) |
| RUIY | 228 | 14 | 4 | Smith JR (593) | Cox IJ (335) | Smeulders AW (237) | Drop (No.6 to 15) |
| SARACEVIC T | 507 | 443 | L | Belkin NJ (1041) | Spink A (1092) | Ingwersen P (579) | Increase (No. 7 to 5) |
| CROFT WB | 1768 | 756 | 5 | Salton G (1768) | Robertson SE (756) | Jones SK (441) | Increase (No. 9 to 6) |
| SPINK A | 317 | 101 | 4 | Saracevic T (1092) | Belkin NJ (741) | Jansen BJ (577) | Increase (No. 13 to 8) |
| JONES KS | 1447 | 897 | L | Salton G (1447) | Robertson SE (897) | Van Rijsbergen CJ (483) | Increase (No. 12 to 7) |
| SMITH JR | 176 | 19 | 4 | Rui Y (593) | Pentland A (341) | Flickener M (319) | Drop (No.8 to 16) |
| FALOUTSOS C | 200 | 21 | 19 | Berchtold D (391) | Salton G (200) | Smith JR (170) | Drop (No.10 to 26) |
| HARMAN D | 912 | 463 | 11 | Salton G (912) | Robertson SE (463) | Voorhees EM (457) | Stable (No. 11 or 12) |
| VOORHEES EM | 745 | 328 | 5 | Salton G (745) | Robertson SE (328) | Harman D (279) | Increase (No. 43 to 17) |
| FLICKNER M | 83 | 5 | 8 | Smith JR (319) | Rui Y (265) | Pentland A (222) | Drop (No.18 to 34) |
| BATES MJ | 342 | 248 | 0 | Belkin NJ (725) | Saracevic T (489) | Fidel R (434) | Stable (No. 15 or 16) |
| CODD EF | 56 | 9 | 286 | Abiteboul S (286) | Date CJ (238) | Ullman JD (213) | Drop (No.14 to 38) |
| BAEZAYATES R | 302 | 94 | 62 | Salton G (302) | Fuhr N (96) | Robertson SE (94) | Drop (No.34 to 55) |
| FUHR N | 653 | 472 | 98 | Salton G (653) | Robertson SE (472) | Van Rijsbergen CJ (362) | Stable (No. 18 or 19) |
| JAIN AK | 133 | 5 | 2 | Rui Y (210) | Smith JR (153) | Swain MJ (145) | Drop (No.42 to 49) |
| | | | | | | | |

TABLE 1. Analysis of authors and their changes of PageRank.

Notes. Author co-citation frequencies are different when compared with author's citation frequencies. The co-citation frequency multiplies two authors' appearance in one paper, which leads to higher number of frequency than the authors' singular citation counts.

JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY—November 2009 2235 DOI: 10.1002/asi



FIG. 3. Scatterplot of weighted PageRank on citations of 108 authors in different damping factors.



FIG. 4. The top 20 highly cited authors' weighted PageRank on citations and its different damping factors.

2236 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY—November 2009 DOI: 10.1002/asi



FIG. 5. Scatterplot of weighted PageRank on publications of 108 authors in different damping factors.



FIG. 6. The top 20 highly cited authors' weighted PageRank on publication and its different damping factors.

measures (Liu et al., 2005; Yin, Kretschmer, Hanneman, & Liu, 2006). The centrality measures the position of each vertex in the network, and directly associates with theories in social sciences as such "weak ties" and "social capital". They contain three major measures: degree centrality, betweenness centrality, and closeness centrality. Degree centrality is equal to the number of ties (connections) that a vertex has with other vertices. Closeness emphasizes the distance of a vertex to all others in the network by focusing on the geodesic distance from each vertex to all others. Betweenness centrality is based on the number of shortest paths passing through a vertex.

We selected the top 20 highly cited authors and calculated their corresponding ranks by using PageRank with different damping factors, their citation ranks, and their ranks on three centrality measures (see Table 2). We also calculated the Spearman correlation among these rankings, with results shown in Table 3.

According to the Spearman correlation analysis, citation rank is highly correlated with PageRank with different damping factors (average Spearman r = 0.941, p < 0.01). Figure 7 shows the scatterplot of the rank of PageRank with the damping factor from 0.05 to 0.95 versus citation rank. It is consistent with the result found by Pretto (2002) that demonstrates

TABLE 2. Author Ranks based upon various methods.

| Rank | 0.05 | 0.15 | 0.25 | 0.35 | 0.45 | 0.55 | 0.65 | 0.75 | 0.85 | 0.95 | Citation | Degree | Betweenness | Closeness |
|------------------|------|------|------|------|------|------|------|------|------|------|----------|--------|-------------|-----------|
| SALTON G | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ROBERTSON SE | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 11 | 21 | 10 |
| ABITEBOUL S | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 5 | 8 | 13 | 3 | 4 | 7 | 71 |
| BELKIN NJ | 4 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 4 | 22 | 24 | 20 |
| VANRIJSBERGEN CJ | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 5 | 5 | 13 | 4 |
| RUI Y | 6 | 6 | 6 | 6 | 7 | 8 | 9 | 9 | 10 | 15 | 6 | 24 | 31 | 23 |
| SARACEVIC T | 7 | 8 | 7 | 7 | 6 | 6 | 6 | 6 | 5 | 5 | 7 | 25 | 51 | 25 |
| CROFT WB | 9 | 9 | 9 | 9 | 9 | 7 | 7 | 7 | 6 | 6 | 8 | 15 | 14 | 14 |
| SPINK A | 13 | 13 | 13 | 12 | 12 | 11 | 11 | 11 | 9 | 8 | 9 | 57 | 93 | 57 |
| JONES KS | 12 | 10 | 10 | 10 | 10 | 10 | 8 | 8 | 7 | 7 | 10 | 17 | 30 | 18 |
| SMITH JR | 8 | 7 | 8 | 8 | 8 | 9 | 10 | 10 | 11 | 16 | 11 | 38 | 32 | 35 |
| FALOUTSOS C | 10 | 11 | 11 | 13 | 13 | 13 | 15 | 16 | 19 | 26 | 12 | 12 | 2 | 11 |
| HARMAN D | 11 | 12 | 12 | 11 | 11 | 12 | 12 | 12 | 12 | 11 | 13 | 3 | 14 | 3 |
| VOORHEES EM | 43 | 16 | 16 | 17 | 17 | 17 | 17 | 18 | 17 | 17 | 14 | 20 | 38 | 19 |
| FLICKNER M | 18 | 18 | 19 | 20 | 20 | 21 | 21 | 24 | 25 | 34 | 15 | 16 | 11 | 15 |
| BATES MJ | 15 | 17 | 17 | 15 | 16 | 16 | 16 | 15 | 15 | 12 | 16 | 78 | 81 | 78 |
| CODD EF | 14 | 14 | 14 | 16 | 18 | 19 | 20 | 22 | 28 | 38 | 17 | 74 | 23 | 74 |
| BAEZAYATES R | 34 | 43 | 45 | 48 | 47 | 50 | 53 | 54 | 55 | 55 | 18 | 6 | 16 | 6 |
| FUHR N | 19 | 19 | 20 | 19 | 19 | 18 | 18 | 19 | 18 | 21 | 19 | 2 | 9 | 2 |
| JAIN AK | 42 | 34 | 40 | 39 | 39 | 40 | 40 | 42 | 45 | 49 | 20 | 39 | 33 | 38 |



FIG. 7. Scatterplots of the ranks of different damping factors versus citation ranks for the top 20 highly cited authors (with rectangle as citation rank and diamond as damping factor). X axis represents the authors who are numbered based on the ranks of their citations, and Y axis represents ranking.

| TABLE 3. T | The Spearman c | orrelation anal | ysis of differe | nt damping fac | ctors. | | | | | | | | | |
|------------------------|-------------------|------------------|------------------|------------------|------------------|-------------------|-------------|-------------|-------------|-------------|-------------|--------------------|-------------|--------------------|
| | d = 0.05 | d = 0.15 | d = 0.25 | d = 0.35 | d = 0.45 | d = 0.55 | d = 0.65 | d = 0.75 | d = 0.85 | d = 0.95 | Citation | Degree | Betweenness | Closeness |
| d = 0.05 | 1.000 | 0.968 | 0.970 | 0.968 | 0.958 | 0.941 | 0.926 | 0.914 | 0.835 | 0.749 | 0.931 | 0.316^{a} | 0.329^{a} | 0.132^{a} |
| d = 0.15 | 0.968 | 1.000 | 0.998 | 0.988 | 0.983 | 0.967 | 0.956 | 0.946 | 0.878 | 0.785 | 0.958 | 0.269^{a} | 0.240^{a} | 0.113 ^a |
| d = 0.25 | 0.970 | 0.998 | 1.000 | 0.989 | 0.986 | 0.971 | 0.962 | 0.952 | 0.887 | 0.797 | 0.964 | 0.271 ^a | 0.235^{a} | 0.114 ^a |
| d = 0.35 | 0.968 | 0.988 | 0.989 | 1.000 | 760.0 | 0.985 | 0.976 | 0.971 | 0.916 | 0.841 | 0.959 | 0.262^{a} | 0.187^{a} | 0.101 ^a |
| d = 0.45 | 0.958 | 0.983 | 0.986 | 0.997 | 1.000 | 0.991 | 0.983 | 0.979 | 0.929 | 0.857 | 0.962 | 0.271^{a} | 0.171^{a} | 0.111^{a} |
| d = 0.55 | 0.941 | 0.967 | 0.971 | 0.985 | 0.991 | 1.000 | 0.995 | 0.992 | 0.959 | 0.895 | 0.965 | 0.278^{a} | 0.171^{a} | 0.143^{a} |
| d = 0.65 | 0.926 | 0.956 | 0.962 | 0.976 | 0.983 | 0.995 | 1.000 | 0.997 | 0.970 | 0.911 | 0.961 | 0.290^{a} | 0.175^{a} | 0.155 ^a |
| d = 0.75 | 0.914 | 0.946 | 0.952 | 0.971 | 0.979 | 0.992 | 0.997 | 1.000 | 0.980 | 0.931 | 0.952 | 0.271^{a} | 0.145^{a} | 0.156^{a} |
| d = 0.85 | 0.835 | 0.878 | 0.887 | 0.916 | 0.929 | 0.959 | 0.970 | 0.980 | 1.000 | 0.973 | 0.916 | 0.244^{a} | 0.068^{a} | 0.197^{a} |
| d = 0.95 | 0.749 | 0.785 | 0.797 | 0.841 | 0.857 | 0.895 | 0.911 | 0.931 | 0.973 | 1.000 | 0.836 | 0.221^{a} | 0.019^{a} | 0.233^{a} |
| Citation | 0.931 | 0.958 | 0.964 | 0.959 | 0.962 | 0.965 | 0.961 | 0.952 | 0.916 | 0.836 | 1.000 | 0.286^{a} | 0.198^{a} | 0.137^{a} |
| Degree | 0.316^{a} | 0.269^{a} | 0.271^{a} | 0.262^{a} | 0.271^{a} | 0.278^{a} | 0.290^{a} | 0.271^{a} | 0.244^{a} | 0.221^{a} | 0.286^{a} | 1.000 | 0.836 | 0.842 |
| Betweenness | 0.329^{a} | 0.240^{a} | 0.235^{a} | 0.187^{a} | 0.171^{a} | 0.171^{a} | 0.175^{a} | 0.145^{a} | 0.068^{a} | 0.019^{a} | 0.198^{a} | 0.836 | 1.000 | 0.652 |
| Closeness | 0.132^{a} | 0.113^{a} | 0.114^{a} | 0.101^{a} | 0.111^{a} | 0.143^{a} | 0.155^{a} | 0.156^{a} | 0.197^{a} | 0.233^{a} | 0.137^{a} | 0.842 | 0.652 | 1.000 |
| ^a Not signi | ficantly correlat | ted. For the res | t, they are sig. | nificantly corre | elated at the .0 |)1 level (1-taile | .(be | | | | | | | |

when *d* changes, the top ranks stay stable while the low ranks vary. To a slight degree, the citation rank is closest to PageRank with damping factor d = 0.55, which is consistent with the result of Chen et al. (2007). Their empirical study shows that scientific papers usually follow a shorter path, an average of two links. They set up their damping factor as 0.5. Ma, Guan, and Zhao (2008) also set their damping factor to 0.5, using PageRank algorithm to evaluate the research influence for several countries in the fields of biochemistry and molecular biology. Yet both PageRank and citation rank are not significantly correlated with centrality measures (average Spearman r = 0.197, p > 0.05). In summary, in the author cocitation graph, through the analysis of the top 20 highly cited authors, we found that:

- Citation rank is highly correlated with PageRank with different damping factors; and
- Citation rank and PageRank are not significantly correlated with centrality measures.

In order to further test the correlation among these measures, we extended our test to include all 108 authors. We took the original PageRank (denoted as PR) with damping factor as 0.15 (to emphasize the equal chance of being cited), 0.55 (as the short path in citation graph is around two links), and 0.85 (to stress the graph topology); weighted PageRank on citation (PR c) with damping factor as 0.15, 0.55, and 0.85; weighted PageRank on publication (PR_p) with damping factor as 0.15, 0.55, and 0.85; and centrality measures (degree, betweenness, closeness), h-index and citation. We compared their ranking similarity by using Spearman correlation coefficient (see Table 4). Citation rank is highly correlated with weighted PageRank on citation, less so with original PageRank, and least with weighted PageRank on publication. It has low correlation with centrality measures and h-index, but the correlations are still significant (average Spearman r = 0.403, p < 0.01). H-index is not highly correlated with the rest of the measures, making it a unique measure for ranking authors. Centrality measures have low correlation with original PageRank and weighted PageRank (except that betweenness is not significantly correlated with weighted PageRank on publication). This indicates that centrality can measure different perspectives of the author impact than other measures. Original PageRank and weighted PageRank on citation are highly correlated, while weighted PageRank on publication has relatively low correlation with the other two PageRank measures.

Figure 8 shows the scatterplots of different measures on the author co-citation network with the total 108 authors. Figure 8a shows the original PageRank versus citation rank, which reflects the high correlation among them with consistency in the top ranks and more diversity in the low ranks. Figure 8b plots correlation between weighted PageRank on citation versus citation ranks. Figure 8c displays the convergence between weighted PageRank on publication versus citation ranks (average Spearman r = 0.585 and p < 0.01). Figure 8d illustrates the correlation among different centrality measures with citation ranks. Although they are quite

| TABLE 4. | Spearman corr | elation coeffi | cients for dif | ferent author ra | nking methods. | | | | | | | | | |
|---|-------------------------------------|---------------------------------|----------------------------------|------------------|--------------------|---------------|--------------|--------------|----------------------|-------------|-----------------|-------------|-------------|----------|
| | PR(.15) | PR(.55) | PR(.85) | PR_c(.15) | PR_c(.55) | PR_c(.85) | PR_p(.15) | PR_p(.55) | PR_p(.85) | Degree | Betweenness | Closeness | h-index | Citation |
| PR(.15) | 1.000 | 0.971 | 0.904 | 0.883 | 0.926 | 0.880 | 0.476 | 0.558 | 0.647 | 0.288 | 0.303 | 0.258 | 0.182^{a} | 0.819 |
| PR(.55) | 0.971 | 1.000 | 0.971 | 0.874 | 0.958 | 0.950 | 0.526 | 0.627 | 0.737 | 0.276 | 0.225 | 0.247 | 0.159^{b} | 0.791 |
| PR(.85) | 0.904 | 0.971 | 1.000 | 0.847 | 0.957 | 0.992 | 0.597 | 0.713 | 0.846 | 0.330 | 0.171^{a} | 0.305 | 0.146^{b} | 0.757 |
| PR_c(.15) | 0.883 | 0.874 | 0.847 | 1.000 | 0.955 | 0.868 | 0.554 | 0.629 | 0.706 | 0.452 | 0.384 | 0.426 | 0.240 | 0.982 |
| PR_c(.55) | 0.926 | 0.958 | 0.957 | 0.955 | 1.000 | 0.966 | 0.588 | 0.686 | 0.794 | 0.397 | 0.285 | 0.371 | 0.211^{a} | 0.895 |
| PR_c(.85) | 0.880 | 0.950 | 0.992 | 0.868 | 0.966 | 1.000 | 0.627 | 0.742 | 0.876 | 0.389 | 0.188^{a} | 0.365 | 0.160^{a} | 0.784 |
| PR_p(.15) | 0.476 | 0.526 | 0.597 | 0.554 | 0.588 | 0.627 | 1.000 | 0.975 | 0.857 | 0.240 | -0.039^{b} | 0.220^{a} | 0.242 | 0.525 |
| PR_p(.55) | 0.558 | 0.627 | 0.713 | 0.629 | 0.686 | 0.742 | 0.975 | 1.000 | 0.941 | 0.296 | -0.023^{b} | 0.278 | 0.227 | 0.586 |
| PR_p(.85) | 0.647 | 0.737 | 0.846 | 0.706 | 0.794 | 0.876 | 0.857 | 0.941 | 1.000 | 0.393 | $0.035^{\rm b}$ | 0.378 | 0.182^{a} | 0.644 |
| Degree | 0.288 | 0.276 | 0.330 | 0.452 | 0.397 | 0.389 | 0.240 | 0.296 | 0.393 | 1.000 | 0.771 | 0.980 | 0.126^{b} | 0.472 |
| Betweenness | 0.303 | 0.225 | 0.171 ^a | 0.384 | 0.285 | 0.188^{a} | -0.039^{b} | -0.023^{b} | 0.035^{b} | 0.771 | 1.000 | 0.745 | 0.145^{b} | 0.420 |
| Closeness | 0.258 | 0.247 | 0.305 | 0.426 | 0.371 | 0.365 | 0.220^{a} | 0.278 | 0.378 | 0.980 | 0.745 | 1.000 | 0.121^{b} | 0.446 |
| h-index | 0.182^{a} | 0.159^{b} | 0.146^{b} | 0.240 | 0.211 ^a | 0.160^{a} | 0.242 | 0.227 | 0.182^{a} | 0.126^{b} | 0.145^{b} | 0.121^{b} | 1.000 | 0.272 |
| Citation | 0.819 | 0.791 | 0.757 | 0.982 | 0.895 | 0.784 | 0.525 | 0.586 | 0.644 | 0.472 | 0.420 | 0.446 | 0.272 | 1.000 |
| ^a Correlation i ^b Correlations | is significant a are not signifi | t the .05 level cant; for the 1 | l (1-tailed). rest, correlati | on is significan | t at the .01 leve | 1 (1-tailed). | | | | | | | | |

scattered, they still have significant correlation with average Spearman r = 0.446 and p < 0.01.

In summary, based on the author co-citation graph, with the test of these 14 different measures on the total 108 authors, we identified the following:

- Citation rank is significantly correlated with all these 14 measures, especially with various PageRank algorithms;
- Original PageRank and different weighted PageRank are significantly correlated;
- Centrality measures have low correlation with citation and various PageRank algorithms; and
- H-index is not significantly correlated with centrality measures.

Conclusion

In this study we selected the 108 most highly cited authors in the field of IR from the 1970s to 2008 to form the author cocitation network. We calculated the ranks of these 108 authors based on PageRank with damping factor ranging from .05 to .95. Furthermore, we compared the PageRank result with the result of citation ranking and centrality ranking.

We found that the citation rank is close to PageRank with d = .55 (Spearman r = .965, p < .01). In general, the citation rank is similar to PageRank (with average Spearman r = .941, p < .01). This means that PageRank and citation rank share similar results. Both citation rank and PageRank are quite different when compared to centrality ranks (average Spearman r is below .197, p > .05). This means that they are not significantly correlated.

We also introduced two different weighted PageRank algorithms: one weighted on citations and the other weighted on publication. Weighted PageRank on citation converges with the citation rank, while weighted PageRank on publication shows some discrepancies with citation ranks. We further compared the selected 14 measures based on the total 108 author set. We found that citation ranks are highly correlated with various PageRank algorithms; original PageRank and weighted PageRank algorithms are correlated; and centrality measures and h-index show different perspectives of measures when compared with citation ranks and various PageRank algorithms. The key factors that have impact on the PageRank of authors in the author co-citation network are being co-cited with important authors.

Citation networks, formed by publications and the citations, are different from hyperlink networks, formed by the Websites and hyperlinks. Unlike hyperlinks, a paper cannot update its references after it is published, and it can only cite published works. Citation networks thus introduce a time order which makes the aging effects more important than those on the Web. PageRank is a form of "lifetime contribution award," as there is no time feature in this algorithm. In the future, we would like to add the time dimension to the PageRank algorithm for ranking authors in order to reflect the dynamics of their contributions and their changes of importance over a certain period.

2240 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY—November 2009 DOI: 10.1002/asi



FIG. 8. Scatterplots of different weighted PageRank and centrality measures versus citation rank based on the author co-citation network with the total 108 authors. X axis represents the authors who are numbered based on their citation ranks. Y axis represents the ranks.

Acknowledgments

The authors thank two anonymous reviewers for valuable comments and suggestions which helped shape and improve this paper. We also thank Russell Duhon for PageRank and centrality-related discussions and support.

References

- Aktas, M.S., Nacar, M.A., & Menczer, F. (2004, August). Using hyperlink features to personalize Web search. In Proceeding of 6th International Workshop on Knowledge Discovery from the Web (WebKDD 2004), Seattle, WA.
- Boldi, P., Santini, M., & Vigna, S. (2005). PageRank as a function of the damping factor. In A. Ellis & T. Hagino (Chairs), Proceedings of 14th International World Wide Web Conference, (pp. 557–566). New York: ACM Press.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30, 107–117
- Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google. Journal of Informetrics, 1, 8–15.
- Fiala, D., Rousselot, F., & Ježek, K. (2008). PageRank for bibliographic networks. Scientometrics, 76(1), 135–158.
- Freeman, L.C. (1979). Centrality in social networks. Conceptual clarification. Social Networks, 1, 215–239.

- Langville, A.N., & Meyer, C.D. (2005). Deeper inside PageRank. Internet Mathematics, 1(3), 335–380.
- Langville, A.N., & Meyer, C.D. (2006). Google's PageRank and beyond: The science of search engine rankings. Princeton, NJ: Princeton University Press.
- Liu, X., Bollen, J., Nelson, M.L., & Sompel, H.V. (2005). Co-authorship networks in the digital library research community. Information Processing and Management, 41, 1462–1480.
- Ma, N., Guan J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. Information Processing and Management, 44, 800–810.
- Maslov, S., & Redner, S. (2008). Promise and pitfalls of extending Google's PageRank algorithm to citation networks. Journal of Neuroscience, 28(44), 11103–11105.
- Mihalcea, R. (2004, July). Graph-based ranking algorithms for sentence extraction, applied to text summarization. Paper presented at the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004), Barcelona, Spain.
- Mihalcea, R., Tarau, P., & Figa, E. (2004, August). PageRank on semantic networks with application to word sense disambiguation. Paper presented at the 20th International Conference on Computational Linguistics, Geneva, Switzerland.
- Perra, N., & Fortunato, S. (2008). Spectral centrality measures in complex networks. Physical Review E 78(3), 036107-1–036107-10.
- Pretto, L. (2002). A theoretical analysis of PageRank. In Proceedings of the Ninth International Symposium on String Processing and Information Retrieval. (pp. 131–144). New York: Springer.

- Sidiropoulos, A., & Manolopoulos, Y. (2006). A generalized comparison of graph-based ranking algorithms for publications and authors. Journal of Systems and Software, 79(12), 1679–1700.
- Small, H. (1973). Cocitation in science literature: New measures of relationship between two documents. Journal of the American Society for Information Science, 24(4), 265–269.
- Walker, D., Xie, H., Yan, K.K., & Maslov, S. (2007). Ranking scientific publications using a model of network traffic. Journal of Statistical Mechanics, 6, P06010.
- Wang, J., Liu J., & Wang, C. (2007). Keyword extraction based on PageRank. Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, 4426, 857–864.
- White, H.D., & McCain, K.W. (1998). Visualizing a discipline: An author cocitation analysis of information science 1972-1995. Journal of the American Society for Information Science, 49, 327–355.
- Xing, W., & Ghorbani, A.A. (2005). Information domain modeling for adaptive Web systems. Web Intelligence, 10, 684–687.
- Yin, L., Kretschmer, H., Hanneman, R.A., & Liu, Z. (2006). Connection and stratification in research collaboration: An analysis of the COLLNET network. Information Processing and Management, 42, 1599–1613.
- Yu, P.S., Li, X., & Liu, B. (2004, May). On the temporal dimension of search. In Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters, New York, NY.

Appendix. Ranks of the 108 authors by PageRank, centrality measures and citation.

| Author | PR(.15) | PR(.55) | PR(.85) | Degree | Betweenness | Closeness | Citation | h-index |
|------------------|---------|---------|---------|--------|-------------|-----------|----------|---------|
| SALTON G | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 |
| ROBERTSON SE | 2 | 2 | 2 | 11 | 21 | 10 | 2 | 34 |
| VANRIJSBERGEN CJ | 5 | 5 | 4 | 5 | 13 | 4 | 3 | 39 |
| ABITEBOUL S | 3 | 4 | 8 | 4 | 7 | 68 | 4 | 25 |
| BELKIN NJ | 4 | 3 | 3 | 22 | 24 | 20 | 5 | 39 |
| RUI Y | 6 | 8 | 10 | 24 | 30 | 23 | 6 | 80 |
| SARACEVIC T | 8 | 6 | 5 | 25 | 50 | 24 | 7 | 14 |
| CROFT WB | 9 | 7 | 6 | 15 | 15 | 14 | 8 | 14 |
| JONES KS | 10 | 10 | 7 | 17 | 29 | 17 | 9 | 103 |
| SPINK A | 13 | 11 | 9 | 55 | 89 | 54 | 10 | 11 |
| SMITH JR | 7 | 9 | 11 | 38 | 31 | 37 | 11 | 7 |
| BAEZAYATES R | 43 | 50 | 55 | 6 | 16 | 5 | 12 | 54 |
| FALOUTSOS C | 11 | 13 | 19 | 12 | 3 | 11 | 13 | 25 |
| HARMAN D | 12 | 12 | 12 | 3 | 14 | 3 | 14 | 54 |
| BATES MJ | 17 | 16 | 15 | 75 | 77 | 74 | 15 | 25 |
| VOORHEES EM | 16 | 17 | 17 | 20 | 37 | 19 | 16 | 87 |
| FLICKNER M | 18 | 21 | 25 | 16 | 11 | 16 | 17 | 93 |
| ULLMAN JD | 21 | 22 | 30 | 85 | 33 | 85 | 18 | 2 |
| CODD EF | 14 | 19 | 28 | 71 | 23 | 72 | 19 | 93 |
| BORGMAN CL | 32 | 32 | 27 | 54 | 64 | 52 | 20 | 23 |
| FUHR N | 19 | 18 | 18 | 2 | 9 | 2 | 21 | 64 |
| JAIN AK | 34 | 40 | 45 | 39 | 32 | 38 | 22 | 39 |
| CHANG SK | 25 | 30 | 40 | 27 | 8 | 26 | 23 | 30 |
| ZADEH LA | 74 | 70 | 63 | 7 | 4 | 6 | 24 | 5 |
| COOPER WS | 15 | 14 | 14 | 52 | 72 | 51 | 25 | 57 |
| SWAIN MJ | 27 | 29 | 37 | 81 | 87 | 81 | 26 | 93 |
| BOOKSTEIN A | 22 | 15 | 13 | 72 | 88 | 71 | 27 | 30 |
| STONEBRAKER M | 20 | 24 | 33 | 87 | 43 | 88 | 28 | 57 |
| LANCASTER FW | 36 | 28 | 23 | 51 | 58 | 50 | 29 | 57 |
| INGWERSEN P | 24 | 20 | 16 | 61 | 92 | 60 | 30 | 39 |
| PENTLAND A | 23 | 23 | 29 | 48 | 42 | 47 | 31 | 14 |
| SMEULDERS AWM | 47 | 47 | 57 | 70 | 67 | 70 | 32 | 18 |
| AGRAWAL R | 57 | 79 | 90 | 34 | 18 | 33 | 33 | 18 |
| EGENHOFER MJ | 64 | 80 | 95 | 92 | 47 | 92 | 34 | 18 |
| SWANSON DR | 39 | 33 | 26 | 86 | 104 | 86 | 35 | 11 |
| KOHONEN T | 81 | 85 | 84 | 23 | 27 | 21 | 36 | 9 |
| BLAIR DC | 49 | 46 | 38 | 68 | 70 | 67 | 37 | 64 |
| BUCKLEY C | 33 | 34 | 31 | 18 | 28 | 15 | 38 | 51 |
| CHEN HC | 41 | 45 | 47 | 42 | 41 | 41 | 39 | 3 |
| ELLIS D | 28 | 26 | 21 | 60 | 82 | 58 | 40 | 34 |
| MARCHIONINI G | 42 | 42 | 35 | 43 | 62 | 42 | 41 | 51 |
| FAGIN R | 29 | 35 | 44 | 35 | 10 | 35 | 42 | 7 |
| MANJUNATH BS | 54 | 53 | 59 | 76 | 80 | 76 | 43 | 11 |
| YU CT | 40 | 31 | 24 | 14 | 6 | 13 | 44 | 39 |
| PORTER MF | 87 | 76 | 65 | 32 | 34 | 31 | 45 | 91 |
| FIDEL R | 30 | 27 | 22 | 50 | 81 | 49 | 46 | 39 |
| HARTER SP | 26 | 25 | 20 | 90 | 107 | 90 | 47 | 39 |
| KIM W | 35 | 37 | 51 | 93 | 39 | 93 | 48 | 64 |

(Continued)

Appendix. (Continued)

| Author | PR(.15) | PR(.55) | PR(.85) | Degree | Betweenness | Closeness | Citation | h-index |
|----------------|----------|----------|----------|--------|-------------|-----------|----------|----------|
| DEERWESTER S | 77 | 81 | 73 | 37 | 61 | 34 | 49 | 93 |
| JANSEN BJ | 53 | 57 | 53 | 36 | 60 | 36 | 50 | 57 |
| FOX EA | 44 | 41 | 32 | 30 | 55 | 29 | 51 | 30 |
| BANCILHON F | 31 | 39 | 49 | 104 | 98 | 104 | 52 | 93 |
| DATE CJ | 62 | 71 | 77 | 91 | 38 | 91 | 53 | 107 |
| NIBLACK W | 55 | 54 | 62 | 46 | 35 | 45 | 54 | 93 |
| MAWY | 37 | 36 | 41 | 82 | 73 | 82 | 55 | 34 |
| VOORHEES E | 73 | 69 | 66 | 67 | 93 | 66 | 56 | 103 |
| BUNEMAN P | 58 | 62 52 | /1 | 94 | 54 | 94 | 57 | 5/ |
| GKAEFE G | 45 | 52 | 0/ | 100 | /8 | 100 | 58 | 83 |
| | 50 | 40 | 39 70 | 59 | 94 | 57 | 59 | 07 25 |
| VANGY | 90 | 91 | 86 | 64 | 75 | 63 | 61 | 23 70 |
| BERNERSI EE T | 108 | 108 | 108 | 29 | 20 | 28 | 62 | 103 |
| GUDIVADA VN | 65 | 72 | 76 | 58 | 20 40 | 20 59 | 63 | 93 |
| GARFIELDE | 106 | 106 | 104 | 78 | 100 | 3) 77 | 64 | 1 |
| MILLER GA | 99 | 100 | 100 | 10 | 19 | 9 | 65 | 34 |
| SHNEIDERMAN B | 88 | 90 | 85 | 13 | 5 | 12 | 66 | 5 |
| JOACHIMS T | 95 | 98 | 98 | 33 | 51 | 32 | 67 | 91 |
| MAIER D | 38 | 38 | 50 | 103 | 85 | 103 | 68 | 39 |
| HEARST MA | 79 | 82 | 75 | 31 | 56 | 30 | 69 | 74 |
| HERSH WR | 103 | 103 | 99 | 88 | 105 | 87 | 70 | 34 |
| KUHLTHAU CC | 82 | 61 | 48 | 96 | 106 | 96 | 71 | 74 |
| ROCCHIO JJ | 69 | 64 | 56 | 21 | 46 | 22 | 72 | 107 |
| BRIN S | 75 | 88 | 91 | 8 | 17 | 7 | 73 | 93 |
| HAWKING D | 52 | 60 | 58 | 45 | 63 | 44 | 74 | 87 |
| WANG JZ | 89 | 78 | 79 | 98 | 103 | 98 | 75 | 39 |
| GUTTMAN A | 51 | 66 | 78 | 69 | 25 | 69 | 76 | 106 |
| WONG SKM | 56 | 44 | 34 | 41 | 49 | 40 | 77 | 39 |
| CARSON C | 59 | 63 | 69 | 89 | 71 | 89 | 78 | 93 |
| CERI S | 93 | 93 | 94 | 102 | 91 | 102 | 79 | 25 |
| LEWIS DD | 67 | 73 | 68 | 40 | 57 | 39 | 80 | 87 |
| HERSH W | 66 | 67 | 60 | 66 | 84 | 65 | 81 | 57 |
| KADECKI I | 68 | 49 | 36 | 84 | 83 | 84 | 82 | /0 |
| SCHAMBER L | /1 | 55 56 | 42 | 57 | 90 45 | 50 25 | 83 | 83 |
| SWIEATON AF | 12 | 30 86 | 40 | 20 | 43 | 23 | 04 95 | /4 10 |
| LAWRENCE S | 04 06 | 80 05 | 80 | 47 | 32 101 | 40 | 83 | 10 |
| DEL RIMBO A | 90 63 | 95 65 | 74 | 65 | 53 | 79 64 | 80 87 | 74 80 |
| IMIFI INSKI T | 104 | 105 | 106 | 105 | 96 | 105 | 88 | 54 |
| KWOK KL | 83 | 75 | 61 | 79 | 97 | 78 | 89 | 83 |
| BERTINO E | 92 | 92 | 93 | 44 | 12 | 43 | 90 | 18 |
| XU J | 97 | 96 | 89 | 56 | 36 | 55 | 91 | 70 |
| CHAUDHURI S | 101 | 104 | 105 | 74 | 26 | 75 | 92 | 30 |
| BERCHTOLD S | 91 | 87 | 92 | 101 | 86 | 101 | 93 | 93 |
| HUANG J | 80 | 84 | 83 | 95 | 74 | 95 | 94 | 18 |
| CRESTANI F | 98 | 97 | 82 | 49 | 65 | 48 | 95 | 57 |
| GUTING RH | 78 | 94 | 102 | 106 | 99 | 106 | 96 | 64 |
| MARON ME | 85 | 58 | 43 | 62 | 76 | 61 | 97 | 74 |
| WILLETT P | 105 | 101 | 96 | 63 | 95 | 62 | 98 | 23 |
| CHAKRABARTI S | 60 | 83 | 87 | 53 | 44 | 53 | 99 | 14 |
| MEADOW CT | 86 | 68 | 52 | 73 | 68 | 73 | 100 | 64 |
| COX IJ | 76 | 74 | 72 | 80 | 79 | 80 | 101 | 39 |
| HULL R | 46 | 43 | 54 | 107 | 102 | 107 | 102 | 74 |
| FRAKES WB | 102 | 102 | 101 | 28 | 48 | 27 | 103 | 83 |
| BEERI C | 48 | 51 | 64 | 99 | 66 | 99 | 104 | 64 |
| MULLER H | 100 | 99 | 103 | 83 | 69 | 83 | 105 | 51 |
| NIELSEN J | 107 | 107 | 107 | 19 | 22 | 18 | 106 | 39 |
| GRUMBACH S | 94 | 89 | 97 | 108 | 108 | 108 | 107 | 80 |
| KOUSSOPOULOS N | 70 | 77 | 88 | 97 | 59 | 97 | 108 | 70 |

Notes. We took WOS data as it is and did not check the author's name variations. So here two people ended up with different names in WOS: Voorhees E vs. Voorhees EM and Hersh W vs. Hersh WR. H-index of authors is subject to the capability of disambiguation of authors with common names in the database.