

## 1 PageRank Papers

The original Brin and Page papers are [56], [57], and [55]. Some others on the topic of PageRank are [21], [43], [44], [83], [218], [117], [118], [213], [42], [246], [191], [236], [145], [119], [147], [146], [161], [120], [136], [80], [125], [11], [93], [233], [201], [211], [210], [22], [63]. A good survey of web searching with PageRank summary is [20]. Other survey papers are: [159], [158]. Modeling web surfing with the 'back button': [92]. Rank aggregation methods: [88]. Popular audience books and articles about Google: [45], [166], [60]. Google file system: [109], [28].

Updating PageRank papers: [65], [198], [197], [156], [157], [134], [235].

Search engine optimizers' guides to PageRank: [9], [76], [219], [220]

## 2 HITS and its modified versions Papers

The original HITS paper is [150]. Others are [163], [83], [82], [40], [125], [72], [73], [12], [74], [62], [96], [97], [208], [64], [148], [77], [210], [103], [102]. Good survey paper: [51]. HITS for bibliometrics: [54]. The search engine Teoma uses HITS: [225].

## 3 Stability of Ranking Vectors

Studied in: [65], [198], [197], [162], [164]

## 4 LSI Papers

The first LSI paper [85] was followed by [34], [35], [37], [81], [78], [244], [245], [165], [140], [241], [247], [142], [143], [151], [131], [141], [47], [46], [155].

## 5 Similarity Measures

Some new similarity measures are proposed in: [137], [107], [208].

## 6 Stats and Structure of Web

Several papers study the structure of the web and statistics regarding the web: [58], [4], [3], [129], [7], [1], [232], [38], [2], [5], [6], [39], [95], [48], [19], [206], [202], [68], [99], [201], [22]. Compressing the web graph: [216], [214], [215].

## 7 Laplacian and Graph Theory

The eigenvalues of the Laplacian of an undirected graph provide a surprising amount of information about the structure of the associated matrix. Do such findings extend to directed graphs, like those underlying a Markov chain? See [190], [189], [100], [101].

## 8 Clustering

Clustering is a difficult and important issue for IR. See [239], [52], [237], [79], [78], [122], [30], [106], [36], [124], [100], [101], [185], [69], [230], [122].

## 9 Markov Chains

Classic Markov chain papers are [105], [113], [66], [179], [180], [184], [226], [87], [115], [231], [177], [130], [29], [133], [132], [224], [178], [182], [181], [175], [176], [169], [121], [67], [18], [193], [86].

## 10 Complex Networks and Small World Networks

The World Wide Web appears to be a complex network with scale-free structure. Some papers about complex networks and their signature properties are: [196], [170], [217], [59], [205], [13], [114], [171], [49], [50], [91], [195], [194], [144], [89], [203], [204], [149], [70], [71], [187], [186], [238], [90], [138], [243], [98], [25], [139], [26], [24].

## 11 Books

Useful books for IR researchers are [61], [183], [229], [240], [228], [227], [200], [23], [32], [33], [104], [112], [152], [242], [135], [221], [223] [188]. [154], [75], [238], [192], [27], [31], [24], [84], [110], [123].

## 12 Miscellaneous

Some papers that are harder to classify: [41], [168], [212, 160] deal with relevance feedback and user-tailored queries. [173] and [94] combine search engine results into one ranked list and [174] post-processes major search engine results annotating retrieved documents with topic information [172]. Spammer's guide to PageRank: [219]. Meta-search engines: [53]. TREC-5 conference: [116]. Medline data set: [8]. An ML info. retrieval model: [199]. IR and genetic algorithm: [209]. Inquirus 2 metasearch engine [111, 160]. Sampling pages uniformly from the Web [222]. Using maximum likelihood and training sets to detect "hidden" links between groups, such as for government intelligence applications: [153]. Using a training set to classify documents and incorporating text content and link information: [108]. The link

prediction problem: [167], [207]. Several papers by same authors on matrix perturbation and spectral decomposition: [17], [14], [15], [16]. Google is the Popular News: [234], [10]. Searching news sites: [127]. Web search challenges: [126], [128].

## References

- [1] Caslon Analytics net metrics and statistics guide. <http://www.caslon.com.au/metricsguide.htm>.
- [2] Search engine watch, March 1998. <http://www.searchenginewatch.com/>.
- [3] Cyveillance: sizing the internet. webpage, July 2000. <http://www.cyveillance.com/web/us/corporate/white-paper.htm>.
- [4] *Books in Print*. R.R. Bowker, New York, 2001.
- [5] *Ulrich's International Periodicals Directory*. Bowker, New York, 2001.
- [6] How much information? webpage, 2002. <http://www.sims.berkeley.edu/research/projects/how-much-info/int>
- [7] How much information, 2003. <http://www.sims.berkeley.edu/how-much-info-2003>.
- [8] Medlars test collection, December 2003. Available at <http://www.cs.utk.edu/lisi/>.
- [9] PageRank explained. Web Rank Info. Optimize your rankings at Google! <http://www.webrankinfo.com/english/pagerank>, 2003.
- [10] Why does my page's rank keep changing? Google PageRank information. <http://www.google.com/webmasters/4.html>, 2003.
- [11] Serge Abiteboul, Mihai Preda, and Gregory Cobena. Adaptive on-line pageimportance computation. In *Twelfth International World Wide Web (WWW12)*, 2003.
- [12] Dimitris Achlioptas, Amos Fiat, Anna R. Karlin, and Frank McSherry. Web search via hub synthesis. In *IEEE Symposium on Foundations of Computer Science*, pages 500–509, 2001.
- [13] Lada A. Adamic. The small world web. In S. Abiteboul and A.-M. Vercoustre, editors, *Proceedings of 3rd European Conference on Research and Advanced Technology for Digital Libraries, ECDL*, number 1696, pages 443–452. Springer-Verlag, 1999.
- [14] Rafikul Alam and Shreemayee Bora. Effect of linear perturbation on spectra of matrices.
- [15] Rafikul Alam and Shreemayee Bora. A geometric approach to sensitivity analysis of eigenvalues and eigendecompositions.
- [16] Rafikul Alam and Shreemayee Bora. On stable eigendecompositions of matrices.
- [17] Rafikul Alam and Shreemayee Bora. Stability of eigenvalues and spectral decompositions under linear perturbation.
- [18] David Aldous. Random walks on finite groups and rapidly mixing markov chains. In A. Dold and B. Eckmann, editors, *Lecture Notes in Mathematics*, volume 986, pages 243–297. Springer-Verlag, 1983.
- [19] Brian Amento, Loren Terveen, and Will Hill. Does authority mean quality? predicting expert quality ratings of web documents. In *The Twenty-Third Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2000.

- [20] Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke, and Sriram Raghavan. Searching the web. *ACM Transactions on Internet Technology*, 2001.
- [21] Arvind Arasu, Jasmine Novak, Andrew Tomkins, and John Tomlin. PageRank computation and the structure of the web: experiments and algorithms. In *The Eleventh International WWW Conference*, May 2002.
- [22] Konstantin Avrachenkov and Nelly Litvak. Decomposition of the google pagerank and optimal linking strategy. Technical report, INRIA, January 2004.
- [23] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, New York, 1999.
- [24] Albert-Laszlo Barabasi. *Linked: The New Science of Networks*. Plume, 2003.
- [25] Albert-Laszlo Barabasi, Reka Albert, and Hawoong Jeong. Attack and error tolerance of complex networks. *Nature*, 406:378, 2000.
- [26] Albert-Laszlo Barabasi, Reka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281:69–77, 2000.
- [27] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition*. SIAM, Philadelphia, PA, 1994.
- [28] Luiz Andre Barroso, Jeffrey Dean, and Urs Holzle. Web search for a planet: The google cluster architecture. *IEEE Micro*, pages 22–28, 2003.
- [29] Michele Benzi and Miroslav Tuma. A parallel solver for large-scale Markov chains. *Applied Numerical Mathematics*, 41:135–153, 2002.
- [30] Pavel Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002.
- [31] Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Academic Press, Inc., 1979.
- [32] Michael W. Berry, editor. *Computational Information Retrieval*. SIAM, Philadelphia, 2001.
- [33] Michael W. Berry and Murray Browne. *Understanding Search Engines: Mathematical Modeling and Text Retrieval*. SIAM, Philadelphia, 1999.
- [34] Michael W. Berry, Z. Drmac, and Elizabeth R. Jessup. Matrices, vector spaces and information retrieval. *SIAM Review*, 41:335–362, 1999.
- [35] Michael W. Berry and R. D. Fierro. Low-rank orthogonal decompositions for information retrieval applications. *Journal of Numerical Linear Algebra with Applications*, 1(1):1–27, 1996.
- [36] Michael W. Berry, Bruce Hendrickson, and Padma Raghavan. Sparse matrix reordering schemes for browsing hypertext. In J. Renegar, M. Shub, , and S. Smale, editors, *Lectures in Applied Mathematics (LAM)*, volume 36, pages 99–123. American Mathematical Society, 1996.
- [37] Michael W. Berry and Gavin W. O’Brien. Using linear algebra for intelligent information retrieval. *SIAM Review*, 37:573–595, 1998.
- [38] Michael W. Berry, P. Wang, and J. Bownas. Website query analysis: trend and behavior detection. In *Second SIAM Conference on Data Mining*, April 2002.

- [39] Krishna Bharat and Andrei Broder. Estimating the relative size and overlap of public web search engines. In *The Seventh International WWW Conference*, New York, 1998. Elsevier Science.
- [40] Krishna Bharat and Monika R. Henzinger. Improved algorithms for topic distillation in hyperlinked environments. In *21st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 104–111, 1998.
- [41] Krishna Bharat, Farzin Maghoul, and Raymie Stata. The term vector database: fast access to indexing terms for webpages. *Computer Networks*, 33:247–255, 2000.
- [42] Krishna Bharat and George A. Mihaila. When experts agree: using non-affiliated experts to rank popular topics. *ACM Transactions on Information Systems*, 20(1):47–58, 2002.
- [43] Monica Bianchini, Marco Gori, and Franco Scarselli. PageRank: A circuital analysis. In *The Eleventh International WWW Conference*, May 2002.
- [44] Monica Bianchini, Marco Gori, and Franco Scarselli. Inside pagerank. *ACM Transactions on Internet Technology*, 2003. In press.
- [45] Nancy Blachman, Eric Fredricksen, and Fritz Schneider. *How to Do Everything with Google*. McGraw-Hill, 2003.
- [46] Katarina Blom. *Information retrieval using the singular value decomposition and Krylov subspaces*. PhD thesis, University of Chalmers, January 1999.
- [47] Katarina Blom and Axel Ruhe. Information retrieval using very short Krylov sequences. In *Computational Information Retrieval*, pages 41–56, 2001.
- [48] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. Structural properties of the African web. In *The Eleventh International WWW Conference*, May 2002.
- [49] Bela Bollobas and Oliver Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1(1):1–35, 2003.
- [50] Bela Bollobas, Oliver Riordan, J. Spencer, and G. Tusnady. The degree sequence of a scale free random graph process. *Random Structures and Algorithms*, 18:279–290, 2001.
- [51] Allan Borodin, Gareth O. Roberts, Jeffrey S. Rosenthal, and Panayiotis Tsaparas. Finding authorities and hubs from link structures on the world wide web. *World Wide Web*, pages 415–429, 2001.
- [52] Rodrigo A. Botafogo and Ben Shneiderman. Identifying aggregates in hypertext structures. In *UK Conference on Hypertext*, pages 63–74, 1991.
- [53] P. Bradley. Multi-search engines - a comparison. webpage, January 2002. <http://www.philb.com/msengine.htm>.
- [54] U. Brandes and T. Willhalm. Visualizing bibliographic networks with a reshaped landscape metaphor. In *4th Joint Eurographics - IEEE TVCG Symposium on Visualization (VisSym '02)*, pages 159–164. ACM Press, 2002.
- [55] Sergey Brin, Rajeev Motwani, Lawrence Page, and Terry Winograd. What can you do with a web in your pocket? *Data Engineering Bulletin*, 21:37–47, 1998.
- [56] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 33:107–117, 1998.

- [57] Sergey Brin, Lawrence Page, R. Motwami, and Terry Winograd. The PageRank citation ranking: bringing order to the web. Technical report, Computer Science Department, Stanford University, 1998.
- [58] Andrei Broder, Ravi Kumar, and Marzin Maghoul. Graph structure in the web. In *The Ninth International WWW Conference*, May 2000.
- [59] Mark Buchanan. *Ubiquity: The History of Science ... Or Why the World is Simpler than We Think*. Crown Publishers, New York, 2001.
- [60] Tara Calishain and Rael Dornfest. *Google Hacks: 100 Industrial-Strength Tips and Tricks*. O'Reilly, 2003.
- [61] Steven Campbell and Carl D. Meyer. *Generalized Inverses of Linear Transformations*. Pitman, San Francisco, 1979.
- [62] Soumen Chakrabarti, Byron Dom, David Gibson, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Spectral filtering for resource discovery. In *ACM SIGIR workshop on Hypertext Information Retrieval on the Web*, 1998.
- [63] Yen-Yu Chen, Qingqing Gan, and Torsten Suel. I/o-efficient techniques for computing PageRank. In *Proceedings of the eleventh international conference on Information and knowledge management (CIKM'02)*, pages 549–557, Virginia, USA, 2002.
- [64] Zheng Chen, Jidong Wang, Liu Wenyin, and Wei-Ying Ma. A unified framework for web link analysis. 2002.
- [65] Steve Chien, Cynthia Dwork, Ravi Kumar, and D. Sivakumar. Towards exploiting link evolution.
- [66] Grace E. Cho and Carl D. Meyer. Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra and its Applications*, 335(1–3):137–150, 2001.
- [67] Grace E. Cho and Carl D. Meyer. Aggregation/disaggregation errors for nearly uncoupled markov chains. Technical report, NCSU Tech. Report #102301, 2003.
- [68] Junghoo Cho and Hector Garcia-Molina. The evolution of the web and implications for an incremental crawler. In *Proceedings of the Twenty-sixth International Conference on Very Large Databases*, 2000.
- [69] HwaJeong Choi and Daniel B. Szyld. Application of threshold partitioning of sparse matrices to markov chains. In *IEEE International Computer Performance and Dependability Symposium IPDS'96*, pages 158–165, September 1996.
- [70] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdown. *Physical Review Letters*, 85, 2000.
- [71] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Breakdown of the internet under intentional attack. *Physical Review Letters*, 86, 2001.
- [72] David Cohn and Huan Chang. Learning to probabilistically identify authoritative documents. In *Proceedings of the 17th International Conference on Machine Learning*, pages 167–174, Stanford, CA, 2000.
- [73] David Cohn, Huan Chang, and Andrew McCallum. Creating customized authority lists. In *Proceedings of the 17th International Conference on Machine Learning*, Stanford, CA, 2000.
- [74] David Cohn and Thomas Hofmann. The missing link: a probabilistic model of document content and hyperlink connectivity. *Advances in Neural Information Processing Systems*, 13, 2001.

- [75] P. J. Courtois. *Decomposability*. Academic Press, New York, 1977.
- [76] Phil Craven. Google’s PageRank explained. Web Workshop. <http://www.webworkshop.net/>, 2003.
- [77] Brian D. Davison, Apostolos Gerasoulis, Konstantinos Kleisouris, Yingfang Lu, Hyun ju Seo, Wei Wang, and Baohua Wu. Discoweb: applying link analysis to web search. In *Eighth International World Wide Web Conference*, May 1999.
- [78] Inderjit S. Dhillon. Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42(1/2):143–175, 2001.
- [79] Inderjit S. Dhillon. Efficient clustering of very large document collections. In Robert L. Grossman, editor, *Data Mining for Scientific and Engineering Applications*, pages 357–388. Kluwer Academic Publishers, 2001.
- [80] Michelangelo Diligenti, Marco Gori, and Marco Maggini. Web page scoring systems for horizontal and vertical search. In *WWW2002*, Honolulu, Hawaii, USA, 2002.
- [81] Chris Ding. A similarity-based probability model for LSI. In *The Twenty-Second ACM SIGIR ‘99 Conference*, pages 59–65, 1999.
- [82] Chris Ding, Xiaofeng He, Parry Husbands, Hongyuan Zha, and Horst Simon. Link analysis: hubs and authorities on the World Wide Web. Technical Report 47847, Lawrence Berkeley National Laboratory, May 2001.
- [83] Chris Ding, Xiaofeng He, Hongyuan Zha, and Horst Simon. PageRank, HITS and a unified framework for link analysis. In *Proceedings of the 25th ACM SIGIR Conference*, pages 353–354, Tampere, Finland, August 2002.
- [84] Ian S. Duff, A. M. Erisman, and J. K. Reid. *Direct Methods for Sparse Matrices*. Oxford Science Publications, 1997.
- [85] Susan T. Dumais. Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments and Computers*, 23:229–236, 1991.
- [86] Dundee. *A Numerical Analyst Looks at the “Cutoff Phenomenon” in Card Shuffling and Other Markov Chains*, 1997.
- [87] Dundee proceedings. *A numerical analyst looks at the ‘cutoff phenomenon’ in card shuffling and other Markov chains*, 1997.
- [88] Cynthia Dwork, Ravi Kumar, and Moni Naor and D. Sivakumar. Rank aggregation methods for the web. In *Tenth International World Wide Web (WWW10)*, 2001.
- [89] Holger Ebel, Latz-Ingo Mielsch, and Stefan Bornholdt. Scale-free topology of e-mail networks. *Physical Review*, E 66, 2002.
- [90] Paul Erdos and Alfred Renyi. On random graphs I. *Math. Debrecen*, 6:290–297, 1959.
- [91] K. A. Eriksen and M. Hornquist. Scale-free growing networks imply preferential attachment. *Physical Review*, E 65, 2001.
- [92] Ronald Fagin, Anna R. Karlin, Jon Kleinberg, Prabhakar Raghavan, Sridhar Rajagopalan, Ronitt Rubinfeld, Madhu Sudan, and Andrew Tomkins. Random walks with ‘back buttons’. In *32nd ACM Symposium on Theory of Computing*, 2000.
- [93] Ronald Fagin, Ravi Kumar, Kevin S. McCurley, Jasmine Novak, D. Sivakumar, John A. Tomlin, and David P. Williamson. Searching the workplace web. In *Twelfth International World Wide Web (WWW12)*, 2003.

- [94] Ronald Fagin, Ravi Kumar, and D. Sivakumar. Comparing top  $k$  lists. In *ACM SIAM Symposium on Discrete Algorithms*, 2003.
- [95] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.
- [96] Ayman Farahat, Thomas Lofaro, Joel C. Miller, Gregory Rae, F. Schaefer, and Lesley A. Ward. Modifications of kleinberg’s HITS algorithm using matrix exponentiation and web log records. In *ACM SIGIR Conference*, pages 444–445, September 2001.
- [97] Ayman Farahat, Thomas Lofaro, Joel C. Miller, Gregory Rae, and Lesley A. Ward. Existence and uniqueness of ranking vectors for linear link analysis. In *ACM SIGIR Conference*, September 2001.
- [98] Illes J. Farkas, Imre Derenyi, Albert-Laszlo Barabasi, and Tamas Vicsek. Spectra of real-world graphs: Beyond the semicircle law. *Physical Review E*, 64, 2001.
- [99] Dennis Fetterly, Mark Manasse, Marc Najork, and Janet L. Wiener. A large-scale study of the evolution of web pages. In *Twelfth International World Wide Web (WWW12)*, 2003.
- [100] Miroslav Fielder. Algebraic connectivity of graphs. *Czechoslovak Mathematics Journal*, 23(98):298–305, 1973.
- [101] Miroslav Fielder. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematics Journal*, 25(100):619–633, 1975.
- [102] Francois Fouss, Jean-Michel Renders, and Marco Saerens. Some relationships between kleinberg’s hubs and authorities, correspondence analysis, and the Salsa algorithm. In *Proceedings of the 3rd IEEE International Conference on the Data Mining (ICDM)*, pages 521–524, 2003.
- [103] Francois Fouss, Jean-Michel Renders, and Marco Saerens. Some relationships between kleinberg’s hubs and authorities, correspondence analysis, and the Salsa algorithm. In *Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data (JADT 2004)*, pages 445–455, 2004.
- [104] W.B. Frakes and Ricardo Baeza-Yates. *Information Retrieval: Data Structures and Algorithms*. Prentice Hall, Englewood Cliffs, NJ, 1992.
- [105] Robert E. Funderlic and Carl D. Meyer. Sensitivity of the stationary distribution vector for an ergodic Markov chain. *Linear Algebra and its Applications*, 76:1–17, 1986.
- [106] Glenn Fung. A comprehensive overview of basic clustering algorithms. May 2001.
- [107] Prasanna Ganesan, Hector Garcia-Molina, and Jennifer Widom. Exploiting hierarchical domain structure to compute similarity. Technical Report 2001-26, Stanford University Database Group, June 2001.
- [108] Lise Getoor, Eran Segal, Ben Taskar, and Daphne Koller. Probabilistic models of text and link structure for hypertext classification. In *IJCAI Workshop on Text Learning: Beyond Supervision*, August 2001.
- [109] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The google file system. In *Proceedings of the nineteenth ACM symposium on Operating systems principles*, pages 29–43, New York, USA, 2003.
- [110] James Gillies and Robert Cailliau. *How the Web was born : the story of the World Wide Web*. Oxford University Press, 2000.
- [111] Eric J. Glover, Steve Lawrence, Michael D. Gordon, William P. Birmingham, and C. Lee Giles. Web search – your way. *Communications of the ACM*, 1999. accepted for publication.

- [112] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1996.
- [113] Gene H. Golub and Carl D. Meyer. Using the  $QR$  factorization and group inverse to compute, differentiate and estimate the sensitivity of stationary probabilities for Markov chains. *SIAM Journal on Algebraic and Discrete Methods*, 17:273–281, 1986.
- [114] Ramesh Govindan and Hongsuda Tangmunarunkit. Heuristics for internet map discovery. In *IEEE INFOCOM 2000*, pages 1371–1380, Tel Aviv, Israel, March 2000.
- [115] Winfried K. Grassmann, M. I. Taskar, and Daniel P. Heyman. Regenerative analysis and steady state distributions for Markov chains. *Operations Research*, 33(5):1107–1116, 1985.
- [116] D. Harman and Ellen Voorhees. Overview of the fifth Text REtrieval Conference (TREC-5). In *Information Technology: The Fifth Text REtrieval Conference (TREC-5)*, pages 1–28, Gaithersburg, MD: NIST, November 1996.
- [117] Taher H. Haveliwala. Efficient computation of PageRank. Technical report, Computer Science Department, Stanford University, 1999.
- [118] Taher H. Haveliwala. Topic-sensitive PageRank. In *The Eleventh International WWW Conference*, May 2002.
- [119] Taher H. Haveliwala and Sepandar D. Kamvar. The second eigenvalue of the google matrix. Technical report, Stanford University, 2003.
- [120] Taher H. Haveliwala, Sepandar D. Kamvar, and Glen Jeh. An analytical comparison of approaches to personalizing pagerank. Technical report, Stanford University, 2003.
- [121] M. Haviv. Aggregation/disaggregation methods for computing the stationary distribution of a markov chain. *SIAM J. Num. Anal.*, 22:952–966, 1987.
- [122] Xiaofeng He, Hongyuan Zha, Chris Ding, and Horst Simon. Web document clustering using hyper-link structures. *Computational Statistics and Data Analysis*, 45:19–45, 2002.
- [123] Kevin Hemenway and Tara Calishain. *Spidering Hacks: 100 Industrial-Strength Tips and Tricks*. O’Reilly, 2003.
- [124] Bruce Hendrickson and Tamara G. Kolda. Partitioning sparse rectangular and structurally nonsymmetric matrices for parallel computation. *SIAM Journal of Scientific Computing*, 21(6):2048–2072, 2000.
- [125] Monika Henzinger. Link analysis in web information retrieval. *IEEE Data Engineering Bulletin*, 23(3):3–8, 2000.
- [126] Monika Henzinger. Algorithmic challenges in web search engines. *Journal of Internet Mathematics*, 1(1):115–126, 2003.
- [127] Monika Henzinger, Bay-Wei Chang, Brian Milch, and Sergey Brin. Query-free news search. In *Twelfth International World Wide Web (WWW12)*, 2003.
- [128] Monika Henzinger, Rajeev Motwani, and Craig Silverstein. Challenges in web search engines. In *SIGIR Forum*, 2002.
- [129] Monika R. Henzinger, Hannes Marais, Michael Moricz, and Craig Silverstein. Analysis of a very large altavista query log. Technical Report 1998-014, Digital SRC, October 1998.

- [130] Daniel P. Heyman and Dianne P. O’Leary. What is fundamental for Markov chains: First passage times, fundamental matrices, and group generalized inverses. In William J. Stewart, editor, *Numerical Solution of Markov Chains (NSMC’95)*, pages 151–161. Kluwer Academic, 1995.
- [131] M. K. Hughey and Michael W. Berry. Improved query matching using kd-trees, a latent semantic indexing enhancement. *Information Retrieval*, 2:287–302, 2000.
- [132] Jeffrey J. Hunter. Generalized inverses, stationary distributions and mean first passage times with applications to perturbed Markov chains. *Research Letters in the Information and Mathematical Sciences*, 3:99–116, 2002.
- [133] Jeffrey J. Hunter. Stationary distributions and mean first passage times of perturbed Markov chains. *Research Letters in the Information and Mathematical Sciences*, 3:85–98, 2002.
- [134] Ilse C. F. Ipsen and Steve Kirkland. Convergence analysis of an improved PageRank algorithm. December 2003.
- [135] D. L. Isaacson and R. W. Madsen. *Markov Chains: Theory and Applications*. John Wiley and Sons, New York, 1976.
- [136] Glen Jeh and Jennifer Widom. Scaling personalized web search. Technical report, Stanford University, 2002.
- [137] Glen Jeh and Jennifer Widom. SimRank: A measure of structural-context similarity. In *Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume 11, Edmonton, Alberta, Canada, July 2002.
- [138] Hawoong Jeong, Zoltan Nda, and Albert-Lszl Barabasi. Measuring preferential attachment for evolving networks. *Europhysics Letters*, 61:567–572, 2003.
- [139] Hawoong Jeong, B. Tombor, Reka Albert, Z. Oltvai, and Albert-Laszlo Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.
- [140] Eric P. Jiang and Michael W. Berry. Solving total least squares problems in information retrieval. *Linear Algebra and its Applications*, 316:137–156, 2000.
- [141] Fan Jiang and Michael L. Littman. Approximate dimension equalization in vector-based information retrieval. In *The Seventeenth International Conference on Machine Learning*, pages 423–430, 2000.
- [142] K. Sparck Jones. A statistical interpretation of term specificity and its applications in retrieval. *J. Documentation*, 28:11–21, 1972.
- [143] William P. Jones and George W. Furnas. Pictures of relevance: A geometric analysis of similarity measures. *Journal of American Society for Information Retrieval*, 38:420–442, 1987.
- [144] K. I. Kahng, B. Khang, and D. Kim. Spectra and eigenvectors of scale-free networks. *Physical Review*, E 64, 2001.
- [145] Sepandar D. Kamvar and Taher H. Haveliwala. The condition number of the pagerank problem. Technical report, Stanford University, 2003.
- [146] Sepandar D. Kamvar, Taher H. Haveliwala, and Gene H. Golub. Adaptive methods for the computation of pagerank. Technical report, Stanford University, 2003.
- [147] Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, and Gene H. Golub. Exploiting the block structure of the web for computing pagerank. Technical report, Stanford University, 2003.

- [148] Hung-Yu Kao, Ming-Syan Chen, Shian-Hua Lin, and Jan-Ming Ho. Entropy-based link analysis for mining web informative structures. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 574–581. ACM Press, 2002.
- [149] Jeffrey O. Kephart, Gregory B. Sorkin, David M. Chess, and Steve R. White. Fighting computer viruses. *Scientific American*, November:88–93, 1997.
- [150] Jon Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46, 1999.
- [151] Tamara G. Kolda and Dianne P. O’Leary. A semi-discrete matrix decomposition for latent semantic indexing in information retrieval. *ACM Transactions on Information Systems*, 16:322–346, 1998.
- [152] Robert R. Korfhage. *Information Storage and Retrieval*. Wiley, New York, 1997.
- [153] Jeremy Kubica, Andrew Moore, Jeff Schneider, and Yiming Yang. Stochastic link and group detection. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 798–804. AAAI Press/MIT Press, July 2002.
- [154] V. G. Kulkarni. *Modeling and Analysis of Stochastic Systems*. Chapman and Hall Publishers, London, 1995.
- [155] Amy N. Langville. The linear algebra behind search engines. *Journal of Online Mathematics and its Applications*, 2003. Submitted in June 2003.
- [156] Amy N. Langville and Carl D. Meyer. Updating PageRank using the group inverse and stochastic complementation. Technical Report crsc02-tr32, North Carolina State University, Mathematics Department, CRSC, 2002.
- [157] Amy N. Langville and Carl D. Meyer. Updating the stationary vector of an irreducible Markov chain. Technical Report crsc02-tr33, N. C. State, Mathematics Dept., CRSC, 2002.
- [158] Amy N. Langville and Carl D. Meyer. A survey of eigenvector methods of web information retrieval. *The SIAM Review*, 2003. Accepted in December 2003.
- [159] Amy N. Langville and Carl D. Meyer. Deeper inside pagerank. *Internet Mathematics Journal*, 2004. Accepted in February 2004.
- [160] Steve Lawrence. Context in web search. *IEEE Data Engineering Bulletin*, 23(3):25–32, 2000.
- [161] Chris Pan-Chi Lee, Gene H. Golub, and Stefanos A. Zenios. Partial state space aggregation based on lumpability and its application to pagerank. Technical report, Stanford University, 2003.
- [162] Hyun Chul Lee and Allan Borodin. Perturbation of the hyperlinked environment. In *Proceedings of the Ninth International Computing and Combinatorics Conference*, July 2003.
- [163] Ronny Lempel and S. Moran. The stochastic approach for link-structure analysis (SALSA) and the TKC effect. In *The Ninth International WWW Conference*, May 2000.
- [164] Ronny Lempel and Shlomo Moran. Rank-stability and rank-similarity of link-based web ranking algorithms in authority-connected graphs. In *Second Workshop on Algorithms and Models for the Web-Graph (WAW 2003)*, Budapest, Hungary, May 2003.
- [165] Todd A. Letsche and Michael W. Berry. Large-scale information retrieval with LSI. *Informatics and Computer Science*, pages 105–137, 1997.
- [166] Michael S. Malone. The complete guide to Googlemania. *Wired*, 12.03, 2004.
- [167] J. N. Manjunatha, K. R. Sivaramakrishnan, R. K. Pandey, and M. Narasimha Murthy. Citation prediction using time series approach kdd cup 2003 (task 1). *SIGKDD Explorations*, 5(2):152–153, 2003.

- [168] Massimo Marchiori. The quest for correct information of the web: hyper search engines. In *The Sixth International WWW Conference*, April 1997.
- [169] R. B. Mattingly and C. D. Meyer. Computing the stationary distribution vector of an irreducible markov chain on a shared-memory multiprocessor. *Numerical Solution Of Markov Chains, Probability: Pure and Applied*, Ed. by W. J. Stewart, 8:491–508, 1991.
- [170] Robert M. May and Alun L. Lloyd. Infection dynamics on scale-free networks. *Physical Review*, E 64, 2001.
- [171] Alberto Medina, Ibrahim Matta, and John Byers. On the origin of power laws in internet topologies. *Computer Communications Review*, 30(2):18–28, 2000.
- [172] Alberto O. Mendelzon and Davood Rafiei. What do the neighbours think? computing web page reputations. *IEEE Data Engineering Bulletin*, 23(3):9–16, 2000.
- [173] Alberto O. Mendelzon and Davood Rafiei. An autonomous page ranking method for metasearch engines. In *The Eleventh International WWW Conference*, May 2002.
- [174] Alberto O. Mendelzon and Davood Rafiei. Topic: measuring webpage reputation. webpage, September 2002. <http://www.cs.toronto.edu/db/topic/about.html>.
- [175] C. D. Meyer and G. W. Stewart. Derivatives and perturbations of eigenvectors. *SIAM J. Numer. Anal.*, 25:679–691, 1988.
- [176] Carl D. Meyer. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Rev.*, 17:443–464, 1975.
- [177] Carl D. Meyer. An alternative expression for the mean first passage matrix. *Linear Algebra and its Applications*, 22:41–47, 1978.
- [178] Carl D. Meyer. Analysis of finite Markov chains by group inversion techniques. *Recent Applications of Generalized Inverses, Research Notes in Mathematics, Pitman*, Ed. S. L. Campbell, 66:50–81, 1982.
- [179] Carl D. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Review*, 31(2):240–272, 1989.
- [180] Carl D. Meyer. Uncoupling the Perron eigenvector problem. *Linear Algebra and its Applications*, 114/115:69–74, 1989.
- [181] Carl D. Meyer. The character of a finite Markov chain. *Linear Algebra, Markov Chains, and Queueing Models, IMA Volumes in Mathematics and its Applications*, Ed., C. D. Meyer and R. J. Plemmons, Springer-Verlag, 48:47–58, 1993.
- [182] Carl D. Meyer. Sensitivity of Markov chains. *SIAM J. Matrix Anal. Appl.*, 15:715–728, 1994.
- [183] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, 2000.
- [184] Carl D. Meyer and James M. Shoaf. Updating finite Markov chains by using techniques of group matrix inversion. *Journal of Statistical Computation and Simulation*, 11:163–181, 1980.
- [185] Violeta Migallon, Jose Penades, and Daniel B. Szyld. Experimental study of parallel iterative solutions of Markov chains with block partitions. In Brigitte Plateau, William J. Stewart, and Manuel Silva, editors, *Numerical Solutions of Markov Chains (NSMC'99)*, pages 96–110. Prensas Universitarias de Zaragoza, 1999.
- [186] Milena Mihail, Christos Gkantsidis, and Ellen Zegura. Spectral analysis of internet topologies. In *Proceedings of Infocom 2003*, 2003.

- [187] Milena Mihail and Christos H. Papadimitriou. On the eigenvalue power law. In *RANDOM 2002*, Harvard, MA, 2002.
- [188] D. L. Minh. *Applied Probability Models*. Brooks/Cole Publishing Co., Duxbury Imprint, Pacific Grove, CA, 2001.
- [189] Bojan Mohar. The Laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk, editors, *Graph Theory, Combinatorics, and Applications*, pages 871–898. Wiley, 1991.
- [190] Bojan Mohar. Some applications of Laplace eigenvalues of graphs. In G. Hahn and G. Sabidussi, editors, *Graph Symmetry: Algebraic Methods and Applications*, pages 225–275. Kluwer Academic, 1997.
- [191] Cleve Moler. The world’s largest matrix computation. *Matlab News and Notes*, pages 12–13, October 2002.
- [192] Cleve B. Moler and Kathryn A. Moler. *Numerical Computing with MATLAB*. SIAM, 2003.
- [193] National Academy of Sciences. *The cutoff Phenomena in Finite Markov Chains*, volume 93, 1996.
- [194] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review*, E 64, 2001.
- [195] M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98:404–409, 2001.
- [196] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–255, 2003.
- [197] Andrew Y. Ng, Alice X. Zheng, and Michael I. Jordan. Link analysis, eigenvectors and stability. In *Seventh International Joint Conference on Artificial Intelligence*, 2001.
- [198] Andrew Y. Ng, Alice X. Zheng, and Michael I. Jordan. Stable algorithms for link analysis. In *Proceedings of the 24th Annual International ACM SIGIR Conference*. ACM, 2001.
- [199] Kenney Ng. A maximum likelihood ratio information retrieval model. In *The Eighth Text REtrieval Conference (TREC-8)*, 1999.
- [200] J. R. Norris. *Markov Chains*. Cambridge University Press, 1997.
- [201] Gopal Pandurangan, Prabhakara Raghavan, and Eli Upfal. Using PageRank to Characterize Web Structure. In *8th Annual International Computing and Combinatorics Conference (COCOON)*, 2002.
- [202] Gopal Pandurangan, Prabhakara Raghavan, and Eli Upfal. Using PageRank to characterize web structure. In *8th Annual International Computing and Combinatorics Conference (COCOON)*, 2002.
- [203] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 2001.
- [204] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Physical Review*, E 65, 2002.
- [205] Vern Paxson and Sally Floyd. Why we don’t know how to simulate the internet. In *Winter Simulation Conference*, pages 1037–1044, 1997.

- [206] David M. Pennock, Gary W. Flake, Steve Lawrence, Eric J. Glover, and C. Lee Giles. Winners don't take all: characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences*, 99(8):5207–5211, 2002.
- [207] Claudia Perlich, Foster Provost, and Sofus Kacskassy. Predicting citation rates for physics papers: Constructing features for an ordered probit model. *SIGKDD Explorations*, 5(2):154–155, 2003.
- [208] Donnacha Phelan and Nicholas Kushmerick. A descendant-based link analysis algorithm for web search. 2002.
- [209] F. Picarougne, N. Monmarch, A. Oliver, and G. Venturini. Web mining with a genetic algorithm. In *The Eleventh International WWW Conference*, May 2002.
- [210] Luca Pretto. *Link Analysis Techniques for ranking webpages*. PhD thesis, University of Padua, 2002.
- [211] Luca Pretto. A theoretical analysis of PageRank. In *Proceedings of the Ninth International Symposium on String Processing and Information Retrieval*, pages 131–144, Lisbon, Portugal, September 2002.
- [212] Jesus U. Quevedo, Ana G. Covarrubias, and S. H. Huang. Improving retrieval by querying and examining prestige. In *The Eleventh International WWW Conference*, May 2002.
- [213] Davood Rafiei and Alberto O. Mendelzon. What is this page known for? computing webpage reputations. In *The Ninth International WWW Conference*, pages 823–835. Elsevier Science, May 2000.
- [214] Sriram Raghavan and Hector Garcia-Molina. Compressing the graph structure of the web. In *Proceedings of the IEEE Conference on Data Compression*, pages 213–222, March 2001.
- [215] Sriram Raghavan and Hector Garcia-Molina. Towards compressing web graphs. In *Proceedings of the IEEE Conference on Data Compression*, pages 203–212, March 2001.
- [216] Sriram Raghavan and Hector Garcia-Molina. Representing web graphs. In *Proceedings of the 19th IEEE Conference on Data Engineering*, Bangalore, India, March 2003.
- [217] Sid Redner. How popular is your paper? an empirical study of the citation distribution. *European Physics Journal*, B 4:131–134, 1998.
- [218] Matthew Richardson and Petro Domingos. The intelligent surfer: probabilistic combination of link and content information in PageRank. *Advances in Neural Information Processing Systems*, 14, 2002.
- [219] Chris Ridings. PageRank explained: everything you've always wanted to know about PageRank. Black Box Group: Rank Write Roundtable. <http://www.rankwrite.com/>, May 2002.
- [220] Chris Ridings and Mike Shishigin. PageRank uncovered. September 2002.
- [221] Andrei Rogers. *Matrix Analysis of Interregional Population Growth and Distribution*. University of California Press, Berkeley and Los Angeles, 1968.
- [222] Paat Rusmevichientong, David M. Pennock, Steve Lawrence, and C. Lee Giles. Methods for sampling pages uniformly from the World Wide Web. In *AAAI Fall Symposium on Using Uncertainty Within Computation*, pages 121–128, 2001.
- [223] Gerard Salton and C. Buckley. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1983.

- [224] Kimberly Ann Flagg Sellers. Iterative methods for computing mean first passage times of Markov chains. Master's thesis, University of Maryland, 1998.
- [225] Chris Sherman. Teoma vs. google, round 2. *Silicon Valley Internet*, 2002. <http://dc.internet.com/news/print.php/1002061>.
- [226] Herbert A. Simon and Albert Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29:111–138, 1961.
- [227] G. W. Stewart. *Introduction to Matrix Computations*. Academic Press, Inc., 1973.
- [228] G. W. Stewart and Ji-guang Sun. *Matrix Perturbation Theory*. Academic Press, Inc., 1990.
- [229] William J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.
- [230] William J. Stewart and Tugrul Dayar. Comparison of partitioning techniques for two-level iterative solvers on large, sparse Markov chains. *SIAM Journal on Scientific Computing*, 21(5):1691–1705, 2000.
- [231] William J. Stewart and W. Wu. Numerical experiments with iteration and aggregation for Markov chains. *ORSA Journal on Computing*, 4(3):336–350, 1992.
- [232] Danny Sullivan. Searches per day. *Search Engine Watch*, 2003. <http://searchenginewatch.com/reports/article.php/2156461>.
- [233] John A. Tomlin. A new paradigm for ranking pages on the world wide web. In *Twelfth International World Wide Web (WWW12)*, 2003.
- [234] M. Totty and M. Mangalindan. Cat and mouse, as google becomes web's gate keeper, sites fight to get in. *Wall Street Journal*, February 26, CCLXI(39):1–11, 2003.
- [235] Ah Chung Tsoi, Gianni Morini, Franco Scarselli, and Markus Hagenbuchner. Adaptive ranking of web pages. In *Twelfth International World Wide Web (WWW12)*, 2003.
- [236] Twelfth International World Wide Web Conference. *Extrapolation Methods for Accelerating PageRank Computations*, 2003.
- [237] Stijn Marinus van Dongen. *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, May 2000.
- [238] Duncan J. Watts. *Small Worlds*. Princeton University Press, 1999.
- [239] Ron Weiss, Bienvenido Velez, Mark A. Sheldon, Chanathip Namprempre, Peter Szilagy, Andrzej Duda, and David K. Gifford. Hypursuit: A hierarchical network search engine that exploits content-link hypertext clustering. In *Seventh ACM Conference on Hypertext*, pages 180–193, 1996.
- [240] James H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, 1965.
- [241] Dian I. Witter and Michael W. Berry. DOWDATING the latent semantic indexing model for conceptual information retrieval. *The Computer Journal*, 41(1):589–601, 1998.
- [242] Richard S. Wurman. *Information Anxiety*. Doubleday, New York, 1989.
- [243] Soon-Hyung Yook, Hawoong Jeong, and Albert-Laszlo Barabasi. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences*, 99:13382–13386, 2002.
- [244] Hongyuan Zha, Osni Marques, and Horst D. Simon. A subspace-based model for information retrieval with applications in latent semantic indexing. *Lecture Notes in Computer Science*, 1457:29–42, 1998.

- [245] Hongyuan Zha and Horst D. Simon. On updating problems in latent semantic indexing. *SIAM Journal on Scientific Computing*, 21(2):782–791, 1999.
- [246] Dell Zhang and Yisheng Dong. An efficient algorithm to rank web resources. *Computer Networks*, 33:449–455, 2000.
- [247] Xiaoyan Zhang, Michael W. Berry, and Padma Raghavan. Level search schemes for information filtering and retrieval. *Information Processing and Management*, 37:313–334, 2001.