Bibliography
Webgraph papers

*

April 5, 2005

References

[1] Daniel M. Abrams and Steven H. Strogatz. Modelling the dynamics of language death. *Nature*, 424:900, August 21 2003.

[2] 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.

[3] Lada A. Adamic. The small world web. In Serge Abiteboul and Anne-Marie Vercoustre, editors, *Proceedings of the 3rd European Conf. Research and Advanced Technology For Digital Libraries, ECDL*, volume 1696, pages 443–452. Springer-Verlag, 1999.

[4] Lada A. Adamic. Zipf, power-laws, and pareto - a ranking tutorial. Technical report, 2000.

[5] Lada A. Adamic and Bernardo A. Huberman. Power law distribution of the World Wide Web. *comment, Science*, 287:2115a, 2000.

[6] Lada A. Adamic, Rajan M. Lukose, and Bernardo A. Huberman. *Local search in unstructured networks*, chapter 13, pages 295–317. Wiley-VCH, Berlin, 2002.

[7] Lada A. Adamic, Rajan M. Lukose, Amit R. Puniyani, and Bernardo A Huberman. Search in power-law networks. *Physical Review E*, 64:46135–46143, October 2001.

[8] Eytan Adar. Guess: The graph exploration system. manual/tutorial, HP Labs, February 16 2005. version 0.1 (alpha).

[9] Jafar Adibi, Hans Chalupsky, Marko Grobelnik, Dunja Mladenic, and Natasa Milic-Frayling. Kdd-2004 workshop report: Link analysis and group detection (linkkdd-2004). *ACM SIGKDD Explorations Newsletter*, 6(2):136–139, December 2004.

[10] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, *Proceedings of the 20th International Conference Very Large Data Bases, Vldb*, pages 487–499. Morgan Kaufmann, 1994. 584 Citations.

[11] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 171–180, 2000.

[12] William Aiello, Fan R. K. Chung, and Linyuan Lu. Random evolution in massive graphs. In *IEEE Symposium On Foundations Of Computer Science*, pages 510–519, 2001.

[13] R. Alberich, J. Miro-Julia, and F. Rossello. Marvel universe looks almost like a real social network. Technical report, arxiv.org, 2002.

[14] Réka Albert and Albert-László Barabási. Dynamics of complex systems: Scaling laws for the period of boolean networks. *Physical Review Letters*, 84(24):5660–5663, June 2000.

[15] Réka Albert and Albert-László Barabási. Topology of evolving networks: local events an universality. *Physical Review Letters*, 85(24):5234–5237, December 2000.

[16] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–98, 2002.

[17] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *Science*, 401:130–131, Septmeber 1999.

[18] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world wide web. *Nature*, 401(6749):130–131, September 1999.

[19] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *Nature*, 401:130–131, September 1999.

[20] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Mean-field theory for scale-free random networks, 1999.

[21] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, July 2000.

[22] James Allan. Automatic hypertext link typing. In *Proceedings Of The 7th ACM Conference On Hypertext*, pages 42–52, 1996.

[23] Elvind Almaas, Rajendra V. Kulkarni, and David Stroud. Characterizing the structure of small-world networks. *Physical Review Letters*, 88(9):098101–1–098101–4, 2002.

[24] Luís A. Nunes Amaral, A. Scala, Marc Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97:11149–11152, 2000.

[25] Brian Amento, Loren Terveen, and Will Hill. Does authority mean quality? predicting expert quality ratings of web documents. In *Proceedings Of The 23rd Annual International ACM (SIGIR) Conference On Research And Development In Information Retrieval*, WWW Information Retrieval, pages 296–303, 2000.

[26] Corin R. Anderson, Pedro Domingos, and Daniel S. Weld. Relational markov models and their application to adaptive web navigation. In *Proceedings of the 17th International Joint Conference on Knowledge Discovery and Data Mining*, pages 143–152, 2002.

[27] Arvind Arasu, Jasmine Novak, Andrew Tomkins, and John Tomlin. Pagerank computation and the structure of the web: Experiments and algorithms. In *Proceedings Of The Eleventh International World Wide Web Conference, Honolulu, Hawaii, Usa, 7–11 May 2002*, 2002.

[28] Yossi Azar, Amos Fiat, Anna R. Karlin, Frank Mcsherry, and Jared Saia. Spectral analysis of data. In *ACM Symposium On Theory Of Computing*, pages 619–626, 2001.

[29] Franco Bagnoli and Michele Bezzi. Small world effects in evolution. *Physical Review E*, 64(2):021914–1–021914–9, August 2001.

[30] Justin Balthrop, Stephanie Forrest, M. E. J. Newman, and Matthew M. Williamson. Technological networks and the spread of computer viruses. *Science*, 304(5670):527–529, April 23 2004.

[31] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine Learning*, 56:89–113, 2004.

[32] Ziv Bar-Yossef, Alexander Berg, Steve Chien, Jittat Fakcharoenphol, and Dror Weitz. Approximating aggregate queries about web pages via random walks. In *VLDB 2000, Proceedings of 26th International Conference on Very Large Data Bases, September 10-14, 2000, Cairo, Egypt*, pages 535–544, 2000.

[33] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[34] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272(2):173–187, 1999.

[35] Albert-László Barabási, Réka Albert, Hawoong Jeong, and J. B. Brockman. Power-law distribution of the world wide web. *response, Science*, 287:2115a, 2000.

[36] Albert-László Barabási, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, A. Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311(4):590–614, 2002.

[37] Albert-László Barabási, Erzsebet Ravasz, and Tamas Vicsek. Deterministic scale-free networks. *Physica A*, 299(4):559–564, 2001.

[38] L.-A. Barabási. *Linked: The New Science of Networks*. Perseus Pr, New York, 2002.

[39] A. D. Barbour and Gesine Reinert. Small worlds. *Random Structures and Algorithms*, 19(1):54–74, 2001.

[40] Andrea Baronchelli and Vittorio Loreto. Data compression approach to information extraction and classification. Technical report, arxiv.org, August 20 2004.

[41] Alain Barrat, Marc Bethélemy, and Alessandro Vespignani. Modeling the evolution of weighted networks. *Physical Review E*, 70:066149–1–066149–12, 2004.

[42] Alain Barrat and M. Weigt. On the properties of small-world network models. *The European Physical Journal B*, 13:547–560, 2000.

[43] Marc Barthélémy and Luís A. Nunes Amaral. Small-world networks: evidence for a crossover picture. *Physical Review Letters*, 82:3180–3183, 1999.

[44] M. Bauer and D. Bernard. A simple asymmetric evolving random network. Technical report, arxiv.org, 2002.

[45] R. J. Baxter. *Exactly Solved Models in Statistical Mechanics*. Academic Press, London, 1982.

[46] Theodore C. Belding. Nobility and stupidity: Modeling the evolution of class endogamy. Technical report, Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI, June 3 2004.

[47] R. Belew. *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW*. Cambridge University Press, Cambridge, 2000.

[48] E. A. Bender and E. R. Canfield. The asymptotic number of labelled graphs with given degree sequences. *Journal of Combinatorial Theory A*, 24:296–307, 1978.

[49] Johannes Berg and Michael Lässig. Correlated random networks. Technical report, arxiv.org, 2002.

[50] Donna Bergmark. Collection synthesis. In *Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries*, pages 253–262, 2002.

[51] Americo T. Bernardes, Dietrich Stauffer, and Janos Kertész. Election results and the sznajd model on barabási network. *The European Physical Journal B*, 25:123–127, 2002.

[52] Michael W. Berry, Susan T. Dumais, and Gavin W. O'Brien. Using linear algebra for intelligent information retrieval. Technical Report UT-CS-94-270, University of Tennessee, 1994.

[53] Upinder S. Bhalla and Ravi Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283:339–340, January 15 1999.

[54] Krishna Bharat and Andrei Broder. A technique for measuring the relative size and overlap of public web search engines. In *Proceedings Of The 7th World Wide Web Conference (WWW7)*, pages 379–388, Brisbane, Australia, Apr 1998.

[55] Krishna Bharat, Andrei Broder, Monika Henzinger, Puneet Kumar, and Suresh Venkatasubramanian. The connectivity server: Fast access to linkage information on the web. *Computer Networks and ISDN Systems*, 30(1–7):469–477, April 1 1998.

[56] Krishna Bharat and Monika R. Henzinger. Improved algorithms for topic distillation in a hyperlinked environment. In *Proceedings Of (SIGIR)-98, 21st ACM International Conference On Research And Development In Information Retrieval*, pages 104–111, Melbourne, Au, 1998.

[57] Piotr Bialas, Zdzislaw Burda, and Des Johnston. Condensation in the backgammon model. *Nuclear Physics B*, 493(3):505–516, 1997.

[58] Monica Bianchini, Marco Gori, and Franco Scarselli. Inside google's web page scoring system, 2001.

[59] Ginestra Bianconi. Mean-field solution of the ising model on a barabási-albert network. Technical report, arxiv.org, 2002.

[60] Ginestra Bianconi and Albert-László Barabási. Bose-einstein condensation in complex networks. *Physical Review Letters*, 86(24):5632–5635, June 11 2001.

[61] Ginestra Bianconi and Albert-László Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(4):436–442, 2001.

[62] Sven Bilke and Carston Peterson. Topological properties of citation and metabolic networks. *Physical Review E*, 64(3):036106–1–036106–5, September 2001.

[63] William James Bluestein. *Hypertext versions of journal articles: computer aided linking and realistic human evaluation.* PhD thesis, University of Western Ontario, 1999.

[64] Avrim Blum and Shuchi Chawla. Learning from labeled and unlabeled data using graph mincuts. In *Proceedings of the 18th International Conf. on Machine Learning*, pages 19–26, citeseer.ist.psu.edu/article/blum01learning.html, 2001. Morgan Kaufmann, San Francisco, CA.

[65] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. *Computational Learning Theory*, pages 92–100, 1998.

[66] B. Bollobás. *Modern Graph Theory.* Springer, New York, 1998.

[67] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled random graphs. *The European Journal of Combinatorics*, 1(311–316), 1980.

[68] Béla Bollobás. *Random Graphs.* Academic Press, London, Uk, 1985.

[69] Béla Bollobás and Wenceslas Fernandez de la Vega. The diameter of random regular graphs. *Combinatorica*, 2(2):125–134, 1982.

[70] Béla Bollobás and Oliver Riordan. The diameter of a scale-free random graph. Preprint, 2003.

[71] Béla Bollobás and Oliver Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1(1):1–35, 2003.

[72] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gabor Tusnady. The degree sequence of a scale-free random graph process. *Random Structure Algorithms*, 18:279–290, 2001.

[73] Philip Bonacich. Power and centrality: A family of measures. *American Journal Of Sociology*, 92(5):1170–1182, March 1987.

[74] Abraham Bookstein and Don Swanson. Probabilistic models for automatic indexing. *Journal Of The American Society For Information Science*, 25(5):118–132, September/October 1974.

[75] Stefan Bornholdt and Holger Ebel. World-wide web scaling exponent from simon's 1955 model. *Physical Review E*, 64(3):035104–1–035104–4, September 2001.

[76] Stefan Bornholdt and Thimo Rohlf. Topological evolution of dynamical networks: Global criticality from local dynamical rules. *Physical Review Letters*, 84(26):6114–6117, June 26 2000.

[77] Stefan Bornholdt and Heinz Georg Schuster, editors. *Handbook of Graphs and Networks*. Wiley-VCH, Berlin, 2002.

[78] Allan Borodin, Gareth O. Roberts, Jeffrey S. Rosenthal, and Penayiotis Tsaparas. Finding authorities and hubs from link structures on the world wide web. In *Tenth International World Wide Web Conference*, 2001.

[79] Indrani Bose. Biological networks. Technical report, arxiv.org, 2002.

[80] Rodrigo A. Botafogo and Ben Shneiderman. Identifying aggregates in hypertext structures. In *Proceedings of the Third Annual ACM conference on Hypertext*, pages 63–74, San Antonio, Texas, USA, 1991. ACM Press.

[81] Jean-Philippe Bouchaud and Marc Mézard. Wealth condensation in a simple model of economy. *Physica A*, 282(4):536–545, 2000.

[82] Jean-Philippe Bouchaud and Marc Potters. *Theory of Financial Risks: From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, 2000.

[83] Justin Boyan, Dayne Freitag, and Thorsten Joachims. A machine learning architecture for optimizing web search engines. In *Proceedings Of The AAAI Workshop On Internet-Based Information Systems*, 1996.

[84] Tim Bray. Measuring the web. In *Proceedings of the fifth international World Wide Web conference on Computer networks and ISDN systems*, pages 993–1005, 1996.

[85] Brian E. Brewington and George Cybenko. How dynamic is the web? *Computer Networks (Amsterdam, Netherlands: 1999)*, 33(1–6):257–276, 2000.

[86] Brian E. Brewington and George Cybenko. Keeping up with the changing web. *IEEE Computer*, 33(5):52–58, 2000.

[87] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[88] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Journal of Computer Networks (Amsterdam)*, 33(1–6):309–320, Jun 2000.

[89] Andrei Z. Broder, Steven C. Glassman, Mark S. Manasse, and Geoffrey Zweig. Syntactic clustering of the web. In *Selected papers from the sixth international conference on World Wide Web*, pages 1157–1166, 1997.

[90] Andrei Z. Broder, Robert Krauthgamer, and Michael Mitzenmacher. Improved classification via connectivity information. In *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, pages 576–585. Society for Industrial and Applied Mathematics, 2000.

[91] James H. Brown, Vijay K. Gupta, Bai-Lian Li, Bruce T. Milne, Carla Restrepo, and Geoffrey B. West. The fractal nature of nature: Power laws, ecological complexity and biodiversity. *Philosophical Transactions of the Royal Society of London B*, 357(1421):619–626, May 29 2002.

[92] Mark Buchanan. *Nexus: Small Worlds and the Groundbreaking Science of Networks*. W. W. Norton & Co., New York, 2002.

[93] Zdzislaw Burda, Joao D. Correia, and Andre. Krzywicki. Statistical ensemble of scale-free random graphics. *Physical Review E*, 64(4):046118–1–046118–9, October 2001.

[94] Zdzislaw Burda, D. Johnston, Jerzy Jurkiewicz, M. Kaminski, Maciej A. Nowak, Gabor Papp, and Ismail Zahed. Wealth condensation in pareto macro-economics. *Physical Review E*, 65:026102–1–026102–4, 2002.

[95] Declan Butler. Souped-up search engines. *Nature*, 405(6783):112, 2000.

[96] Pável Calado, Berthier Ribeiro-Neto, Nivio Ziviani, Edleno Moura, and Ilmério Silva. Local versus global link information in the web. *ACM Transactions On Information Systems (Tois)*, 21(1):42–63, 2003.

[97] Duncan S. Callaway, John E. Hopcroft, Jon M. Kleinberg, M. E. J. Newman, and Steven H. Strogatz. Are randomly grown graphs really random? *Physical Review E*, 64(4):041902–1–041902–7, October 2001.

[98] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25):5468–5471, December 18 2000.

[99] Antoni Calvó-Armengol and Yves Zenou. Job matching, social network and word-of-mouth communication. Technical report, arxiv.org, 2001.

[100] J. M. Carlson and J. Doyle. Highly optimized tolerance: Robustness and design in complex systems. *Physical Review Letters*, 84(11):2529–2532, March 13 2000.

[101] J. M. Carlson and John Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E*, 60(2):1412–1427, August 1999.

[102] Jeromy Carrière and Rick Kazman. Webquery: Searching and visualizing the web through connectivity. In *Proceedings Of The Sixth International World Wide Web Conference*, pages 701–711, 1997.

[103] Rich Caruana, Thorsten Joachims, and Lars Backstrom. Kdd-cup 2004: results and analysis. *SIGKDD Explorations Newsletter*, 6(2):95–108, 2004.

[104] Soumen Chakrabarti, Martin Van Den Berg, and Byron Dom. Focused crawling: A new approach to topic-specific web resource discovery. *Computer Networks (Amsterdam, Netherlands: 1999)*, 31(11–16):1623–1640, 1999.

[105] Soumen Chakrabarti, Byron Dom, Rakesh Agrawal, and Prabhakar Raghavan. Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies. *The VLDB Journal*, 7(3):163–178, 1998.

[106] Soumen Chakrabarti, Byron Dom, David Gibson, Jon Kleinberg, S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Hypersearching the web. *Scientific American*, 280(6):54–60, June 1999.

[107] Soumen Chakrabarti, Byron Dom, David Gibson, Jon Kleinberg, Prabhakar Raghavan, and Sridhar Rajagopalan. Automatic resource list compilation by analyzing hyperlink structure and associated text. In *Proceedings Of The 7th International World Wide Web Conference*, 1998.

[108] Soumen Chakrabarti, Byron E. Dom, and Piotr Indyk. Enhanced hypertext categorization using hyperlinks. In Laura M. Haas and Ashutosh Tiwary, editors, *Proceedings Of SIGMOD-98, ACM International Conference On Management Of Data*, pages 307–318, Seattle, Us, 1998. ACM Press, New York, Us.

[109] Soumen Chakrabarti, Byron E. Dom, S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins, David Gibson, and Jon Kleinberg. Mining the web's link structure. *Computer*, 32(8):60–67, 1999.

[110] Soumen Chakrabarti and Byron Edward Dom. Feature diffusion across hyperlinks, April 1998.

[111] Soumen Chakrabarti, David A. Gibson, and Kevin S. McCurley. Surfing the web backwards. *Computer Networks (Amsterdam, Netherlands: 1999)*, 31:1679–1693, 1999.

[112] Soumen Chakrabarti, Mukul Joshi, and Vivek Tawde. Enhanced topic distillation using text, markup tags, and hyperlinks. In W. Bruce Croft, David J. Harper, Donald H. Kraft, and Justin Zobel, editors, *Proceedings*

Of The 24th Annual International ACM (SIGIR) Conference On Research And Development In Information Retrieval ((SIGIR)-01), pages 208–216, New York, Sep 2001. ACM Press.

[113] Soumen Chakrabarti, Mukul M. Joshi, Kunal Punera, and David M. Pennock. The structure of broad topics on the web. In *Proceedings of the eleventh international conference on World Wide Web*, pages 251–262, 2002.

[114] Soumen Chakrabarti, Kunal Punera, and Mallela Subramanyam. Accelerated focused crawling through online relevance feedback. In *WWW, Hawaii*. ACM, May 2002.

[115] Qian Chen, Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott J. Shenker, and Walter Willinger. The origin of power laws in internet topologies revisited. In *Proceedings Of IEEE Infocom 2002*, 2002.

[116] Zheng Chen, Shengping Liu, Liu WenYin, Geguang Pu, and Weiying Ma. Building a web thesaurus from web link structure. In *Proceedings of the 26th Annual ACM SIGIR Conference On Research and Development In Information Retrieval*, pages 49–55, July 2003.

[117] Steve Chien, Cynthia Dwork, Ravi Kumar, Daniel R. Simon, and D. Sivakumar. Link evolution: Analysis and algorithms. *Internet Mathematics*, 1(3):277–304, 2003.

[118] Junghoo Cho and Hector Garcia-Molina. Synchronizing a database to improve freshness. In *ACM SIGMOD Record*, pages 117–128, 2000.

[119] Junghoo Cho and Hector Garcia-Molina. Parallel crawlers. In *Proceedings of the eleventh international conference on World Wide Web*, pages 124–135, 2002.

[120] Junghoo Cho and Hector Garcia-Molina. Estimating frequency of change. In *ACM Transactions on Internet Technology*, volume 3, pages 256–290. ACM Press, 2003.

[121] Junghoo Cho, Hector García-Molina, and Lawrence Page. Efficient crawling through URL ordering. *Computer Networks and ISDN Systems*, 30(1–7):161–172, 1998.

[122] Junghoo Cho and Alexandros Ntoulas. Effective change detection using sampling, 2002.

[123] Kim Christensen, Raul Donangelo, Belita Koiller, and Kim Sneppen. Evolution of random networks. *Physical Review Letters*, 81(11):2380–2383, September 14 2000.

[124] Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, Providence, RI, 1997.

[125] Fan R. K. Chung and Linyuan Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–114, 2003.

[126] Fan R. K. Chung, Linyuan Lu, and Van Vu. The spectra of random graphs with given expected degrees. *Internet Mathematics*, 1(3):257–275, 2003.

[127] Kenneth W. Church. Empirical estimates of adaptation: The chance of two noriegas is closer to $p/2$ than p^2. In *Proceedings of the International Conference On Computational Linguistics*, Saarbruecken, Germany, August 2000.

[128] Kenneth W. Church and William Gale. Poisson mixtures. *Natural Language Engineering*, 1(2):163–190, 1995.

[129] Kenneth W. Church and Patrick Hanks. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29, 1990.

[130] Philipp Cimiano and Steffen Staab. Learning by googling. *SIGKDD Explor. Newsl.*, 6(2):24–33, 2004.

[131] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70:066111–1–066111–6, December 6 2004.

[132] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Physical Review Letters*, 85(21):4626–4628, November 20 2000.

[133] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances In Neural Information Processing Systems*, volume 10. The MIT Press, 1998.

[134] David Cohn and Huan Chang. Learning to probabilistically identify authoritative documents. In *Proceedings of the 17th International Conference On Machine Learning*, pages 167–174. Morgan Kaufmann, San Francisco, Ca, 2000.

[135] David Cohn and Thomas Hofmann. The missing link - a probabilistic model of document content and hypertext connectivity. In *Neural Information Processing Systems 13*, 2001.

[136] Colin Cooper and Alan Frieze. A general model of web graphs, 2001.

[137] Colin Cooper and Alan Frieze. Crawling on simple models of web graphs. *Internet Mathematics*, 1(1):57–90, 2003.

[138] Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew K. McCallum, Tom M. Mitchell, Kamal Nigam, and Seán Slattery. Learning to construct knowledge bases from the world wide web. *Artificial Intelligence*, 118(1/2):69–113, 2000.

[139] Dragos Cvetcović, M. Doob, and H. Sachs. *Spectra of Graphs*. Cambridge University Press, Cambridge, 1979.

[140] Dary J. Daley and Joe M. Gani. *Epidemic Modeling*. Cambridge University Press, Cambridge, Uk, 1999.

[141] J. Davidsen, H. Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, 88:128701, March 8 2001.

[142] Brian D. Davison. Recognizing nepotistic links on the web. In *AAAI-2000 Workshop On Artificial Intelligence For Web Search*, pages 23–28, Austin, Texas, July 2000. AAAI Press.

[143] Brian D. Davison. Topical locality in the web. In *Research and Development In Information Retrieval ((SIGIR))*, pages 272–279, 2000.

[144] M. Argollo de Menezes, Cristian F. Moukarzel, and T. J. P. Penna. First-order transition in small-world networks. *Europhysics Letters*, 50(5):574–579, June 1 2000.

[145] Derek J. de Solla Price. Networks of scientific papers. *Science*, 149(3683):510–515, July 30 1965.

[146] Jeffrey Dean and Monika R. Henzinger. Finding related pages in the world wide web. *Computer Networks (Amsterdam, Netherlands: 1999)*, 31(11–16):1467–1479, 1999.

[147] Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by latent semantic analysis. *Journal Of The American Society Of Information Science*, 41(6):391–407, 1990.

[148] Inderjit S. Dhillon, James Fan, and Yuqiang Guan. Efficient clustering of very large document collections. In R. Grossman, G. Kamath, and R. Naburu, editors, *Data Mining for Scientific and Engineering Applications*. Kluwer Academic Publishers, 2001.

[149] Inderjit S. Dhillon and Dharmendra S. Modha. Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42(1/2):143–175, 2001.

[150] Devanshu Dhyani, Wee Keong Ng, and Sourav S. Bhowmick. A survey of web metrics. *ACM Computing Surveys (CSUR)*, 34(4):469–503, 2002.

[151] Michelangelo Diligenti, Frans Coetzee, Steve Lawrence, C. Lee Giles, and Marco Gori. Focused crawling using context graphs. In *26th International Conference On Very Large Databases, Vldb 2000*, pages 527–534, Cairo, Egypt, 10–14 September 2000.

[152] Stephen Dill, S. Ravi Kumar, Kevin S. McCurley, Sridhar Rajagopalan, D. Sivakumar, and Andrew Tomkins. Self-similarity in the web. In *The Vldb Journal*, pages 69–78, 2001.

[153] Peter Sheridan Dodds, Duncan J. Watts, and Charles F. Sabel. Information exchange and the robustness of organizational networks. *Proceedings of the National Academy of Sciences of the USA*, 100(21):12516–12521, October 14 2003.

[154] Pedro Domingos and Michael Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3):103–130, 1997.

[155] William P. Doran, Nicola Stokes, Eamonn Newman, John Dunnion, and Joe Carthy. A hybrid statistical/linguistic model for generating news story gists. In *SIGIR '04: Proceedings of the 27th annual international conference on Research and development in information retrieval*, pages 464–465, New York, NY, USA, 2004. ACM Press.

[156] S. N. Dorogovstev, A. V. Goltsev, and J. F. F. Mendes. Ising model on networks with an arbitrary distribution of connections. *Physical Review E*, 66:016104–1–016104–5, July 8 2002.

[157] S. N. Dorogovstev and J. F. F. Mendes. Evolution of networks with aging of sites. *Physical Review E*, 62(2):1842–1845, August 2000.

[158] S. N. Dorogovstev and J. F. F. Mendes. Exactly solvable small-world networks. *Europhysics Letters*, 50(1):1–7, April 2000.

[159] S. N. Dorogovstev and J. F. F. Mendes. Scaling behaviour of developing and decaying networks. *Europhysics Letters*, 52(1):33–39, October 2000.

[160] S. N. Dorogovstev and J. F. F. Mendes. Effect of the accelerated growth of communications networks on their structure. *Physical Review E*, 63(2):025101–1–025101–4, February 2001.

[161] S. N. Dorogovstev and J. F. F. Mendes. Language as an evolving word web. *Proceedings of the Royal Society of London B*, 268(1485):2603–2606, December 22 2001.

[162] S. N. Dorogovstev and J. F. F. Mendes. Scaling properties of scale-free evolving networks: Continuum approach. *Physical Review E*, 63(5):056125–1–056125–19, May 2001.

[163] S. N. Dorogovstev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51(4):1079–1187, June 1 2002.

[164] S. N. Dorogovstev, J. F. F. Mendes, and A. N. Samukhin. Growing networks with heritable connectivity of nodes. Technical report, arxiv.org, 2000.

[165] S. N. Dorogovstev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Physical Review Letters*, 85(21):4633–4636, November 20 2000.

[166] S. N. Dorogovstev, J. F. F. Mendes, and A. N. Samukhin. WWW and internet models from 1955 till our days and the "popularity is attractive" principle. Technical report, arxiv.org, 2000.

[167] S. N. Dorogovstev, J. F. F. Mendes, and A. N. Samukhin. Generic scale of 'scale-free' networks. *Physical Review E*, 63(6):062101–1–062101–4, June 2001.

[168] S. N. Dorogovstev, J. F. F. Mendes, and A. N. Samukhin. Giant strongly connected component of directed networks. *Physical Review E*, 64(2):025101–1–025101–4, August 2001.

[169] S. N. Dorogovstev, J. F. F. Mendes, and A. N. Samukhin. Modern architecture of random graphs: Constructions and correlations. Technical report, arxiv.org, 2002.

[170] S. N. Dorogovstev, J. F. F. Mendes, and A. N. Samukhin. Principles of statistical mechanics of random networks. Technical report, arxiv.org, 2002.

[171] Petros Drineas, Alan Frieze, Ravi Kannan, Santosh Vempala, and V. Vinay. Clustering in large graphs and matrices. In *Soda: ACM-SIAM Symposium On Discrete Algorithms (A Conference On Theoretical And Experimental Analysis Of Discrete Algorithms)*, 1999.

[172] Petros Drineas, Alan Frieze, Ravi Kannan, Santosh Vempala, and V. Vinay. Clustering large graphs via the singular value decomposition. *Machine Learning*, 56(1-3):9–33, 2004.

[173] Ted E. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, March 1993.

[174] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the tenth international conference on World Wide Web*, pages 613–622, 2001.

[175] Holger Ebel, Lutz-Ingo Mielsch, and Stefan Bornholdt. Scale free topology of e-mail networks. *Physical Review E*, 66(3):035103–1–035103–4, September 30 2002.

[176] Jean-Pierre Eckmann and Elisha Moses. Curvature of co-links uncovers hidden thematic layers in the world wide web. Technical report, arxiv.org, 2001.

[177] Ali Eftekhari. Fractal geometry of literature: First attempt to shakespeare's works. Technical report, arxiv.org, August 17 2004.

[178] Leo Egghe and Ronald Rousseau. *Introduction to Infometrics: Quantitative Methods in Library, Documentation and Information Science*. Elsevier, Amsterdam, 1990.

[179] David Eppstein and Joseph Wang. A steady state model for power graph laws. Technical Report DM/0204001, arxiv.org, 2002.

[180] P. Erdös and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–291, 1959.

[181] P. Erdös and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[182] Guler Ergun. Human sexual contact network as a bipartite graph. *Physica A*, 308(4):483–488, 2002.

[183] Guler Ergun and G. J. Rodgers. Growing random networks with fitness. *Physica A*, 303(2):261–272, 2002.

[184] Stephen Eubank, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, and Nan Wang. Structural and algorithmic aspects of massive social networks. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 718–727, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics.

[185] Alex Fabrikant, Elias Koutsoupias, and Christos H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In *ICALP: Annual International Colloquium On Automata, Languages and Programming*, 2002.

[186] Ronald Fagin, Anna R. Karlin, Jon Kleinberg, Prabhakar Raghavan, Sridhar Rajagopalan, Ronitt Rubinfeld, Madhu Sudan, and Andrew Tomkins. Random walks with the back button. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 484–493, 2000.

[187] Chris Faloutsos and Stavros Christodoulakis. Signature files: an access method for documents and its analytical performance evaluation. *ACM Transactions on Information Systems*, 2(4):267–288, 1984.

[188] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *Sigcomm*, pages 251–262, 1999.

[189] Illes J. Farkas, Imre Derényi, Albert-László Barabási, and Tamas Vicsek. Spectra of "real-world" graphs: Beyond the semi-circle law. *Physical Review E*, 64(2):026504–1–026504–5, August 2001.

[190] Illes J. Farkas, Hawoong Jeong, Tamas Vicsek, Albert-László Barabási, and Zoltan N. Oltvai. The topology of the transcription regulatory network in the yeast, *Saccharomyces cerevisiae*. *Phsycia A*, 318(4):601–612, 2003.

[191] Gary Flake, Steve Lawrence, and C. Lee Giles. Efficient identification of web communities. In *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 150–160, Boston, MA, August 20–23 2000.

[192] Gary W. Flake, Kostas Tsioutsiouliklis, and Robert E. Tarjan. Graph clustering techniques based on minimum cut trees. Technical Report 2002-06, NEC, Princeton, NJ, 2002.

[193] Gary William Flake, Steve Lawrence, C. Lee Giles, and Frans Coetzee. Self-organization of the web and identification of communities. *IEEE Computer*, 35(3):66–71, 2002.

[194] Daniela Florescu, Alon Y. Levy, and Alberto O. Mendelzon. Database techniques for the world-wide web: A survey. *SIGMOD Record*, 27(3):59–74, 1998.

[195] P. J. Flory. Statistical thermodynamics of random networks. *Proceedings of the Royal Society of London A*, 351(1666):351–378, November 19th 1976.

[196] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, March 1977.

[197] Mark E. Frisse. Searching for information in a hypertext medical handbook. In *Proceedings of the ACM conference on Hypertext and Hypermedia*, pages 57–66, Chapel Hill, North Carolina, USA, 1987. ACM Press.

[198] Henryk Fuks and Anna T. Lawniczak. Performance of data networks with random links. *Mathematics and Computers in Simulation*, 51(2):101–117, December 22 1999.

[199] Eugene Garfield. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, 122(3159):108–111, 1955.

[200] Eugene Garfield. Citation analysis as a tool in journal evaluation. *Science*, 178(4060):471–479, 1972.

[201] Eugene Garfield. *Citation Indexing: Its Theory and Application in Science*. Wilay, New York, 1979.

[202] Ralph Garner. *A computer oriented, graph theoretic analysis of citation index structures*. Drexel University Press, Philadelphia, PA, 1967.

[203] Alexander Gelbukh and Grigori Sidorov. Zipf and Heaps Laws' coefficients depend on language. In *Proceedings of the 2001 Conference on Intelligent Text Processing and Computational Linguistics*, pages 332–335, 2001.

[204] David Gibson, Jon M. Kleinberg, and Prabhakar Raghavan. Inferring web communities from link topology. In *UK Conference On Hypertext*, pages 225–234, 1998.

[205] E.N. Gilbert. Random graphs. *Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[206] Nigel Gilbert. A simulation of the structure of academic science. *Sociological Research Online*, 2(2), 1997.

[207] Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Similarity search in high dimensions via hashing. In *Proceedings of the 25th International Conference on Very Large Data Bases*, pages 518–529. Morgan Kaufmann Publishers Inc., 1999.

[208] Michelle Girvan and M.E.J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Science USA*, 99:8271–8276, 2002.

[209] M. Gitterman. Small-world phenomena in physics: The ising model. *Journal of Physics A: Mathematical and General*, 33(47):8373–8382, December 1 2000.

[210] Oren Glickman and Ido Dagan. Identifying lexical paraphrases from a single corpus: A case study for verbs. Technical report, arxiv.org, December 25 2003.

[211] Kwang-Il Goh, Byungnam Kahng, and Dong-Hee Kim. Spectra and eigenvectors of scale-free networks. *Physical Review E*, 64(5):051903–1–051903–5, November 2001.

[212] Kwang-Il Goh, Byungnam Kahng, and Dong-Hee Kim. Pluctuation-driven dynamics of the internet topology. *Physical Review Letters*, 88(10):108701, March 11 2002.

[213] Kwang-Il Goh, E. S. Oh, Hawoong Jeong, Byungnam Kahng, and Dong-Hee Kim. Classification of scale-free networks. *Proceedings of the National Academy of Sciences of the USA*, 99(20):12583–12588, October 1 2002.

[214] A. V. Goltsev, S. N. Dorogovstev, and J. F. F. Mendes. Critical phenomena in networks. Technical report, arxiv.org, 2002.

[215] Mark Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[216] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83:1420–1443, 1978.

[217] Russell D. Gray and Quentin D. Atkinson. Language-tree divergence times support the anatolian theory of indo-european origin. *Nature*, 426:435–439, November 27 2003.

[218] Gianluigi Greco, Sergio Greco, and Ester Zumpano. Web communities: Models and algorithms. *World Wide Web: Internet and Web Information Systems*, 7(1):59–82, 2004.

[219] Stephen J. Green. Building hypertext links in newspaper articles using semantic similarity. In *Proceedings Of Third Workshop On Application Of Natural Language To Information Systems (Nldb '97)*, pages 178–190, Vancouver, June 1997.

[220] Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 491–501, New York, NY, USA, 2004. ACM Press.

[221] X. Guardiola, Albert Díaz-Guilera, Conrad J. Perez, Alex Arenas, and Mateu Llas. Modelling diffusion of innovations in a social network. Technical report, arxiv.org, 2002.

[222] R. Guimerá, L. Danon, A. Daz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68:065103–1–065103–4, 2003.

[223] Jin Han and Wei Li. How structure affects power-law behavior. Technical report, arxiv.org, 2002.

[224] Mor Harchol-Balter, Tom Leighton, and Daniel Lewin. Resource discovery in distributed networks. In *Proceedings of the eighteenth annual ACM symposium on Principles of distributed computing*, pages 229–237, 1999.

[225] Taher Haveliwala. Topic-sensitive PageRank. In *Proceedings Of The Eleventh International World Wide Web Conference*, Honolulu, Hawaii, May 2002. ACM.

[226] David Hawking, Ellen Voorhees, Nick Craswell, and Peter Bailey. Overview of the TREC8 web track. In *Proceedings Of The Eighth Text Retrieval Conference (TREC-8)*, Gaithersburg, Maryland, November 1999.

[227] David Heckerman, David Maxwell Chickering, Christopher Meek, Robert Rounthwaite, and Carl Myers Kadie. Dependency networks for inference, collaborative filtering, and data visualization. *Journal of Machine Learning Research*, 1:49–75, 2000.

[228] Bruce Hendrickson and Robert Leland. A multi-level algorithm for partitioning graphs. In *Supercomputing*, 1995.

[229] Monika R. Henzinger. Link analysis in web information retrieval. *IEEE Data Engineering Bulletin*, 23(3):3–8, 2000.

[230] Monika R. Henzinger. Algorithmic challenges in web search engines. *Internet Mathematics*, 1(1):115–126, 2003.

[231] Monika R. Henzinger, Allan Heydon, Michael Mitzenmacher, and Marc Najork. Measuring index quality using random walks on the web. In *Proceeding of the eighth international conference on World Wide Web*, pages 1291–1303, 1999.

[232] Monika R. Henzinger, Allan Heydon, Michael Mitzenmacher, and Marc Najork. On near-uniform url sampling. *Journal of Computer Networks (Amsterdam)*, 33(1–6):295–308, June 2000.

[233] Michael Hersovici, Michal Jacovia, Yoelle S. Maareka, Dan Pelleg, Menachem Shtalhaima, and Sigalit Ura. The shark-search algorithm–an application:tailored web site mapping. In *Proceedings Of The 7th World Wide Web Conference (WWW7)*, Brisbane, Australia, April 1998.

[234] Allan Heydon and Marc Najork. Mercator: A scalable, extensible web crawler. *World Wide Web*, 2(4):219–229, 1999.

[235] Thomas Hofmann. Probabilistic latent semantic analysis. In *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 289–296, San Francisco, CA, 1999. Morgan Kaufmann Publishers.

[236] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual ACM Conference on Research and Development in Information Retrieval*, pages 50–5–7, Berkeley, California, August 1999.

[237] Thomas Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42(1-2):177–196, 2001.

[238] Thomas Hofmann and Jan Puzicha. Unsupervised learning from dyadic data. Technical Report TR-98-042, University of California Berkely, Berkeley, CA, 1998.

[239] Thomas Hofmann and Jan Puzicha. Latent class models for collaborative filtering. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 688–693. Morgan Kaufmann Publishers Inc., 1999.

[240] Thomas Hofmann, Jan Puzicha, and Michael I. Jordan. Learning from dyadic data. In *Proceedings of the 1998 conference on Advances in neural information processing systems II*, pages 466–472. MIT Press, 1999.

[241] Ramin Homayouni, Kevin Heinrich, Lai Wei, and Michael W. Berry. Gene clustering by latent semantic indexing of MEDLINE abstracts. *Bioinformatics*, 21(1):104–115, 2005.

[242] C. H. Hubbell. An input-output approach to clique identification. *Sociometry*, 28(4):377–399, December 1965.

[243] Bernardo A. Huberman. *The Laws of the Web*. MIT Press, Cambridge, MA, 2001.

[244] Bernardo A. Huberman and Lada A. Adamic. Growth dynamics of the World-Wide Web. *Nature*, 401(6749), September 9 1999.

[245] Bernardo A. Huberman, Peter L. Pirolli, James E. Pitkow, and Rajan M. Lukose. Strong Regularities in World Wide Web Surfing. *Science*, 280(5360):95–97, 1998.

[246] Ramon Ferrer i Cancho, Christiaan Janssen, and Ricard V. Solé. The topology of technology graphs: Small world patterns in electronic circuits. *Physical Review E*, 64(4):046119–1–046119–5, October 2001.

[247] Ramon Ferrer i Cancho and Ricard V. Solé. The small-world of human language. *Proceedings of the Royal Society of London B*, 268(1482):2261–2265, November 7 2001.

[248] Ramon Ferrer i Cancho and Ricard V. Solé. Two regimes in the frequency of words and the origins of complex lexicons: Zipf's law revised. *Journal of Quantitative Linguistics*, 8(3):165–173, December 2001.

[249] Ramon Ferrer i Cancho and Ricard V. Solé. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788–791, February 4 2003.

[250] Takashi Ito, Kosuke Tashiro, Shigeru Muta, Ritsuko Ozawa, Tomoko Chiba, Mayumi Nishizawa, Kiyoshi Yamamoto, Satoru Kuhara, and Yoshiyuki Sakaki. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between yeast proteins. *Proceedings of the National Academy of Sciences of the USA*, 97(3):1143–1147, February 1 2000.

[251] Sanjay Jain and Sandeep Krishna. A model for the emergence of cooperation, interdependence and structure in evolving networks. *Proceedings of the National Academy of Sciences of the USA*, 98(2):543–547, January 16 2001.

[252] S. Janson, D. E. Knuth, T. Luczak, and B. Pittel. The birth of the giant component. *Random Structures and Algorithms*, 4(233–358), 1993.

[253] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley, New York, 2000.

[254] Hawoong Jeong, Zoltan Néda, and Albert-László Barabási. Measuring preferential attachment for eveolving networks. Technical report, arxiv.org, 2001.

[255] Hawoong Jeong, B. Tombor, Réka Albert, Zoltan N. Oltvai, and Albert-László Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–653, October 5 2000.

[256] Sune Norhoj Jespersen and Alexander Blumen. Small-world networks: Links with long-tailed distributions. *Physical Review E*, 62(5):6270–6274, November 2000.

[257] Emily M. Jin, Michelle Girvan, and M. E. J. Newman. Structure of growing social networks. *Physical Review E*, 64(4):046132–1–046132–8, October 2001.

[258] Thorsten Joachims. A statistical learning model of text classification for support vector machines. In *SIGIR 2001: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, September 9-13, 2001, New Orleans, Louisiana, USA*, volume 24, pages 128–136. ACM, 2001.

[259] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings Of The ACM Conference On Knowledge Discovery and Data Mining (Kdd)*. ACM, 2002.

[260] Jürgen Jost and M. P. Joy. Evolving networks with distance preferences. Technical report, arxiv.org, 2002.

[261] Byungnam Kahng, Y. Park, and Hawoong Jeong. Robustness of the in-degree exponent for the world wide web. Technical report, arxiv.org, 2001.

[262] Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, and Gene H. Golub. Exploiting the block structure of the web for computing PageRank. Preprint, March 2003.

[263] Richard M. Karp, Christian Schindelhauer, Scott Shenker, and Berthold Vöcking. Randomized rumor spreading. In *IEEE Symposium on Foundations of Computer Science*, pages 565–574, citeseer.ist.psu.edu/karp00randomized.html, 2000.

[264] S. A. Kauffman. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford University Press, Oxford, 1995.

[265] David Kempe, Jon Kleinberg, and Alan Demers. Spatial gossip and resource location protocols. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 163–172, 2001.

[266] M. Kessler. Bibliographic coupling between scientific papers. *American Documentation*, 14:10–25, 1963.

[267] Peter D. Killworth and H. Russell Bernard. The reversal small-world experiment. *Social Networks*, 1(2):159–192, 1978.

[268] Beom Jun Kim, Chang No Yoon, Seugn Kee Han, and Hawoong Jeong. Path finding strategies in scale-free networks. *Physical Review E*, 65(2):027103–1–027103–4, January 23 2001.

[269] Hyun-Joo Kim, Youngki Lee, In-Mook Kim, and Byungnam Kahng. Scale -free networks in financial correlations. Technical report, arxiv.org, 2001.

[270] J. Kim, P. L. Krapivsky, Byungnam Kahng, and S. Redner. Infinite-order percolation and giant fluctuations in a protein interaction network. Technical report, arxiv.org, 2002.

[271] Osame Kinouchi, Alexandre S. Martinez, Gilson F. Lima, G. M. Lourenco, and Sebastian Risau-Gusman. Deterministic walks in random networks: An application to thesaurus graphs. Technical report, arxiv.org, 2001.

[272] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings Of The 32nd Acm Symposium On Theory Of Computing*, 2000.

[273] Jon Kleinberg. Small-world phenomena and the dynamics of information. *Advances in Neural Information Processing Systems (NIPS)*, 14, 2001.

[274] Jon Kleinberg. Detecting a network failure. *Internet Mathematics*, 1(1):37–56, 2003.

[275] Jon Kleinberg and Steve Lawrence. The structure of the web. *Science*, 294(5548):1849–1850, 2001.

[276] Jon M. Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 599–608, 1997.

[277] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal Of The ACM*, 46(5):604–632, 1999.

[278] Jon M. Kleinberg. Hubs, authorities, and communities. *ACM Computing Surveys*, 31(4es):Article No. 5, December 1999.

[279] Jon M. Kleinberg. Navigation in a small world. *Nature*, 406:845, 2000.

[280] Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins. The Web as a graph: Measurements, models and methods. In *Proceedings of the 5th Annual International Computing and Combinatorics Conference*, 1999.

[281] Jon M. Kleinberg and Eva Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. In *IEEE Symposium on Foundations of Computer Science*, pages 14–23, 1999.

[282] Judith Kleinfeld. Could it be a big world after all? The 'six degrees of separation' myth. Forthcoming, Society, 2002, April 12 2001.

[283] Konstantin Klemm and Victor M. Eguíluz. Growing scale-free networks with small-world behavior. *Physical Review E*, 65(5):057102–1–057102–4, May 8 2002.

[284] Konstantin Klemm and Victor M. Eguíluz. Highly clustered scale-free networks. *Physical Review E*, 65(3):036123–1–036123–5, March 2002.

[285] Daphne Koller and Mehran Sahami. Hierarchically classifying documents using very few words. In Douglas H. Fisher, editor, *Proceedings Of ICML-97, 14th International Conference On Machine Learning*, pages 170–178, Nashville, TN, 1997. Morgan Kaufmann Publishers, San Francisco, Us.

[286] Charles Korte and Stanley Milgram. Acquaintance networks between racial groups: application of the small world method. *Journal of Personality and Social Psychology*, 15:101–108, 1970.

[287] Wessel Kraaij, Jian-Yun Nie, and Michel Simard. Embedding web-based statistical translation models in cross-language information retrieval. *Computational Linguistics*, 29(3):381–419, September 2003.

[288] P. L. Krapivsky and S. Redner. Organization of grwoing random networks. *Physical Review E*, 63(6):066123–1–066123–14, June 2001.

[289] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85(21):4629–4632, November 20 2000.

[290] P. L. Krapivsky, G. J. Rodgers, and S. Redner. Degree distributions of growing networks. *Physical Review Letters*, 86(23):5401–5404, June 4 2001.

[291] Valdis Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.

[292] Andries Kruger, C. Lee Giles, Frans Coetzee, Eric Glover, Gary Flake, Steve Lawrence, and Cristian Omlin. DEADLINER: Building a new niche search engine. In *Ninth International Conference on Information and Knowledge Management, CIKM 2000*, pages 272–281, Washington, DC, November 6–11 2000.

[293] Andre Krzywicki. Defining statistical ensembles of random graphs. Technical report, arxiv.org, 2001.

[294] Rajendra V. Kulkarni, E. Almaas, and David Stroud. Evolutionary dynamics in the bak-sneppen model on small-world networks. Technical report, arxiv.org, 1999.

[295] Rajendra V. Kulkarni, E. Almaas, and David Stroud. Exact results and scaling properties of small-world networks. *Physical Review E*, 61(4):4268–4271, April 2000.

[296] L. Kullmann and Janos Kertész. Preferential growth: Exact solution of the time-dependent distributions. *Physical Review E*, 63(5):051112–1–051112–7, May 2001.

[297] L. Kullmann and Janos Kertész. Preferential growth: Solution and application to modelling stock market. *Physica A*, 299(1):121–126, October 1 2001.

[298] L. Kullmann, Janos Kertész, and K. Kaski. Time dependent cross correlations between different stock returns: A directed network of influence. *Physical Review E*, 64(5):057105–1–057105–3, November 2001.

[299] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins, and Eli Upfal. Stochastic models for the web graph. In *Focs: IEEE Symposium On Foundations Of Computer Science (Focs)*, citeseer.ist.psu.edu/article/kumar00stochastic.html, 2000.

[300] S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins, and Eli Upfal. The web as a graph. In *Proceedings of the 19th ACM Symposium on Principles of Database Systems*, pages 1–10, 2000.

[301] S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Extracting large-scale knowledge bases from the web. In *The VLDB Journal*, pages 639–650, 1999.

[302] S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Trawling the web for emerging cyber-communities. *Computer Networks (Amsterdam, Netherlands: 1999)*, 31(11–16):1481–1493, 1999.

[303] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, pages 282–289. Morgan Kaufmann, San Francisco, CA, 2001.

[304] Karim R. Lakhani and Eric von Hippel. How open source software works: "free" user-to-user assistance. *Research Policy*, 32:923–943, 2003.

[305] Amy N. Langville and Carl D. Meyer. Deeper inside pageRank. *Internet Mathematics*, 1(3):335–380, 2003.

[306] Ray R. Larson. Bibliometrics of the world wide web: An exploratory analysis of the intellectual structure of cyberspace. In *Annual Meeting Of The American Society For Information Science*, 1996.

[307] Vito Latora and Massimo Marchiori. Is the boston subway a small-world network? Technical report, arxiv.org, 2002.

[308] Steve Lawrence. Context in web search. *IEEE Data Engineering Bulletin*, 23(3):25–32, 2000.

[309] Steve Lawrence. Online or invisible? *Nature*, 411(6837):521, May 31 2001.

[310] Steve Lawrence and C. Lee Giles. Context and page analysis for improved web search. *IEEE Internet Computing*, 2(4):38–46, 1998.

[311] Steve Lawrence and C. Lee Giles. Searching the World Wide Web. *Science*, 280(5360):98–100, 1998.

[312] Steve Lawrence and C. Lee Giles. Accessibility of information on the web. *Nature*, 400(6740):107–109, July 8 1999.

[313] Steve Lawrence and C. Lee Giles. Searching the web: General and scientific information access. *IEEE Communications*, 37(1):116–122, 1999.

[314] Steve Lawrence, C. Lee Giles, and Kurt Bollacker. Digital libraries and autonomous citation indexing. *IEEE Computer*, 32(6):67–71, 1999.

[315] Timothy Robert Leek. Information extraction using hidden markov models. Master's thesis, UC San Diego, 1997.

[316] Ronny Lempel and Schlomo Moran. The stochastic approach for link-structure analysis (SALSA) and the Tkc Effect. *Computer Networks (Amsterdam, Netherlands: 1999)*, 33(1–6):387–401, 2000.

[317] Ronny Lempel and Schlomo Moran. SALSA: the stochastic approach for link-structure analysis. *ACM Transactions on Information Systems*, 19(2):131–160, 2001.

[318] Ronny Lempel and Aya Soffer. PicASHOW: Pictorial authority search by hyperlinks on the web. *ACM Transactions On Information Systems*, 20(1):1–24, Jan 2002.

[319] Richard E. Lenski, Charles Ofria, Robert Pennock, and Christoph Adami. The evolutionary origin of complex features. *Nature*, 423:139–144, May 8 2003.

[320] Mark Levene, Trevor Fenner, George Loizou, and Richard Wheeldon. A stochastic model of evolution on the web. *Computer Networks*, 39(3):277–287, June 21 2002.

[321] Wentian Li. Random texts exhibit Zipf's law-like word frequency distribution. *IEEETIT: IEEE Transactions on Information Theory*, 38(6):1842–1845, November 1992.

[322] Erez Lieberman, Christoph Hauert, and Martin A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433:312–316, January 20 2005.

[323] Fredrik Liljeros, Christofer R. Edling, Luis A. Nunes Amaral, H. Eugene Stanley, and Yvonne Aberg. The web of human sexual contacts. *Nature*, 411(6840):907, June 21 2001.

[324] Alun L. Lloyd and Robert M. May. How viruses spread among computers and people. *Science*, 292(5520):1316–1317, May 18 2001.

[325] A. J. Lotka. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, 16(12):317–323, June 19 1926.

[326] Thomas Lux and Michele Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–499, February 11 1999.

[327] Clifford Lynch. Searching the internet. *Scientific American*, 276(3):52–57, March 1997.

[328] Benoit B. Mandelbrot. *The Fractal Geometry of Nature*. Freeman, New York, 1983.

[329] Massimo Marchiori. The quest for correct information on the Web: Hyper search engines. *Computer Networks and ISDN Systems*, 29(11):1225–1235, September 1997.

[330] Sergei Maslov and Kim Sneppen. Specificity and Stability in Topology of Protein Networks. *Science*, 296(5569):910–913, 2002.

[331] Sergei Maslov, Kim Sneppen, and Alexei Zaliznyak. Pattern detection in complex networks: Correlation profile of the internet. Technical report, arxiv.org, 2002.

[332] Sergei Maslov and Yi-Cheng Zhang. Extracting hidden information from knowledge networks. *Physical Review Letters*, 87(24):248701–1–248701–4, December 10 2001.

[333] Nisha Mathias and Venkatesh Gopal. Small-worlds: How and why. *Physical Review E*, 63(2):021117–1–021117–12, February 2001.

[334] Sebastian M. Maurer and Bernardo A. Huberman. Competitive dynamics of web sites. Technical report, arxiv.org, 2000.

[335] Robert M. May and Alun L. Lloyd. Infection dynamics on scale-free networks. *Physical Review E*, 64(6):066112–1–066112–4, December 2001.

[336] Katherine W. McCain. Core journal networks and cocitation maps in the marine sciences: tools and information management in interdisciplinary research. In *Proceedings of the 55th annual meeting on Celebrating change : information management on the move*, pages 3–7. American Society for Information Science, 1992.

[337] Andrew McCallum and Kamal Nigam. A comparison of event models for naive Bayes text classification. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, pages 41–48, 1998.

[338] Andrew McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. Building domain-specific search engines with machine learning techniques. In *Proceedings of the AAAI-99 Spring Symposium On Intelligent Agents In Cyberspace, 1999.*, 1999.

[339] Alberto Medina, Ibrahim Matta, and John Byers. On the origin of power laws in internet topologies. *Computer Communications Review*, 30(2):18–28, April 2000.

[340] Sergey Melnik, Sriram Raghavan, Beverly Yang, and Hector Garcia-Molina. Building a distributed full-text index for the web. *ACM Transactions on Information Systems*, 19(3):217–241, 2001.

[341] Filippo Menczer. ARACHNID: Adaptive retrieval agents choosing heuristic neighborhoods for information discovery. In *Machine Learning: Proceedings of the Fourteenth International Conference*, pages 227–235, 1997.

[342] Filippo Menczer. Links tell us about lexical and semantic web content. Technical Report Technical Report Computer Science Abstract Cs.Ir/0108004, arxiv.org, August 2001.

[343] Filippo Menczer. Growing and navigating the small world web by local content. *Proceedings Of The National Academy Of Sciences*, 99(22):14014–14019, 2002.

[344] Filippo Menczer and Richard K. Belew. Adaptive retrieval agents: Internalizing local context and scaling up to the web. *Machine Learning*, 39(2/3):203–242, 2000.

[345] Weiyi Meng, Clement T. Yu, and King-Lup Liu. Building efficient and effective metasearch engines. *ACM Computing Surveys*, 34(1):48–89, 2002.

[346] Adam Meyerson, Liadan O'Callaghan, and Serge Plotkin. A k-median algorithm with running time independent of data size. *Machine Learning*, 56:61–87, 2004.

[347] Milena Mihail and Christos H. Papadimitriou. On the Eigenvalue power law. In *Proceedings of RANDOM 2002.*, 2002.

[348] Stanley Milgram. The small world problem. *Psychology Today*, 1(1):60–67, 1967.

[349] Ron Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[350] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. Preprint, 2001.

[351] Michael Mitzenmacher. Dynamic models for file sizes and double Pareto distributions. *Internet Mathematics*, 1(3):305–333, 2003.

[352] Michael Molloy and Bruce Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability, and Computing*, 7(3):295–305, September 1998.

[353] Jose M. Montoya and Ricard V. Solé. Topological properties of food webs: From real data to community assembly models, 2001. Working Papers of Santa Fe Institute, 01-11-069.

[354] Jose M. Montoya and Ricard V. Solé V. Small world patterns in food webs. *Journal of Theoretical Biology*, 214(3):405–412, February 7 2002.

[355] Christopher Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678–5682, May 2000.

[356] Christopher Moore and M. E. J. Newman. Exact solution of site and bond percolation on small-world networks. *Physical Review E*, 62(5):7059–7064, November 2000.

[357] Stephen Morris. Contagion. *Review of Economic Studies*, 67:57–58, 2000.

[358] Stefano Mossa, Marc Barthélémy, H. Eugene Stanley, and Luis A. Nunes Amaral. Truncation of power law behaviour in "scale-free" network models due to information filtering. *Physical Review Letters*, 88(13):138701, March 14 2002.

[359] Frederick Mosteller and David L. Wallace. *Inference and Disputed Authorship: The Federalist*. Addison-Wesley, Massachusetts, 1964.

[360] Terutaka Nabeshima and Yukio-Pegio Gunji. Zipf's law in phonograms and weibull distribution in ideograms: comparison of english with japanese. *Biosystems*, 73(2):131–139, 2004.

[361] Marc Najork and Janet L. Weiner. Breadth-first search crawling yields high-quality pages. In *Proceedings Of The 10th World Wide Web Conference (WWW7)*, 2001.

[362] Olfa Nasraoui, Bamshad Mobasher, Brij Masand, and Bing Liu. Webkdd 2004: web mining and web usage analysis post-workshop report. *SIGKDD Explorations Newsletter*, 6(2):147–151, 2004.

[363] M. E. J. Newman. Models of the small world. *Journal of Statistical Physics*, 101(3):819–841, November 2000.

[364] M. E. J. Newman. Small worlds: The structure of social networks. Technical report, arxiv.org, 2000.

[365] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):025102–1–025102–4, August 2001.

[366] M. E. J. Newman. Ego-centered networks and the ripple effect. Technical report, arxiv.org, 2001.

[367] M. E. J. Newman. Exact solutions of epidemic models on networks, 2001. Working Papers of Santa Fe Institute, 01-12-073.

[368] M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings Of The National Academy Of Sciences*, 98(2):404–409, 2001.

[369] M. E. J. Newman. Who is the best connected scientist? A study of scientific coauthorship networks. *Physical Review E*, 64(1):016132–1–016132–7, July 2001. Scientific collaboration networks. Part II. Shortest paths, weighted networks, and centrality.

[370] M. E. J. Newman. Who is the best connected scientist? A study of scientific coauthorship networks. *Physical Review E*, 64(1):016131–1–016131–8, July 2001. Scientific collaboration networks. Part I. Network construction and fundamental results.

[371] M. E. J. Newman. Assortative mixing in networks. Technical report, arxiv.org, 2002.

[372] M. E. J. Newman. Random graphs as models of networks. Technical report, arxiv.org, 2002.

[373] M. E. J. Newman. The spread of epidemic disease on networks. *Physical Review E*, 66(16):016128–1–016128–11, July 2002.

[374] M. E. J. Newman, Stephanie Forrest, and Justin Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):035101–1–035101–4, September 2002.

[375] M. E. J. Newman, I. Jensen, and R. M. Ziff. Percolation and epidemics in a two-dimensional small world. *Physical Review E*, 65(2):021904–1–021904–7, February 2002.

[376] M. E. J. Newman, Christopher Moore, and Duncan J. Watts. Mean-field solution of small-world networks. *Physical Review Letters*, 84(14):3201–3204, April 3 2000.

[377] M. E. J. Newman and Duncan J. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(5):341–346, December 6 1999.

[378] M. E. J. Newman and Duncan J. Watts. Scaling and percolation in the small-world network model. *Physical Review E*, 60(6):7332–7342, December 1999.

[379] M.E.J. Newman, Steven H. Strogatz, and Duncan J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118–1–026118–17, July 24 2001.

[380] Andrew Y. Ng, Alice X. Zheng, and Michael I. Jordan. Link analysis, eigenvectors and stability. In *International Joint Conferences on Artificial Intelligence*, pages 903–910, 2001.

[381] 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.

[382] Kamal Nigam and Rayid Ghani. Analyzing the effectiveness and applicability of co-training. In *Proceedings of the ninth international conference on Information and knowledge management*, pages 86–93. ACM Press, 2000.

[383] Kamal Nigam, John Lafferty, and Andrew McCallum. Using maximum entropy for text classification. In *Proceedings of Machine Learning for Information Filtering Workshop*, 1999.

[384] Kamal Nigam, Andrew Kachites McCallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39(2-3):103–134, 2000.

[385] Martin A. Nowak, Natalia L. Komarova, and Partha Niyogi. Computational and evolutionary aspects of language. *Nature*, 417:611–617, June 6 2002.

[386] Martin A. Nowak, Joshua B. PLotkin, and Vincent A. A. Jansen. The evolution of syntactic communication. *Nature*, 404:495–498, March 30 2000.

[387] Zoltan N. Oltvai and Albert-László Barabási. SYSTEMS BIOLOGY: Life's Complexity Pyramid. *Science*, 298(5594):763–764, 2002.

[388] Edward T. O'Neill, Patrick D. McClain, and Brian F. Lavoie. A methodology for sampling the world wide web. *Annual Review of OCLC Research*, 1997.

[389] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, Stanford University, Stanford, Ca, Usa, November 11 1998.

[390] Gopal Pandurangan, Prabhakara Raghavan, and Eli Upfal. Using page rank to characterize web structure. In *8th Annual International Computing and Combinatorics Conference (COCOON)*, 2002.

[391] Christos H. Papadimitriou, Prabhakar Raghavan, Hisao Tamaki, and Santosh Vempala. Latent semantic indexing: A probabilistic analysis. *JCSS: Journal Of Computer and System Sciences*, 61(2):217–235, October 2000.

[392] Kishore Papineni. Why inverse document frequency? In *Proceedings of the North American Association for Computational Linguistics*, pages 25–32, 2001.

[393] Romualdo Pastor-Satorras, Eric Smith, and Ricard V. Solé V. Evolving protein interaction networks through gene duplication, 2002. Working Papers of Santa Fe Institute, 02-02-008.

[394] Romualdo Pastor-Satorras, Alexei Vázquez, and Alessandro Vespignani. Dynamical and correlation properties of the internet. *Physical Review Letters*, 87(25):258701–1–258701–4, December 17 2001.

[395] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic searching in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, April 2 2000.

[396] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117–1–066117–8, June 2001.

[397] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65(3):035108–1–035108–4, March 2002.

[398] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemics and immunization in scale-free networks. Technical report, arxiv.org, 2002.

[399] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex-networks. *Physical Review E*, 65(3):036104–1–036104–8, March 2002.

[400] 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.

[401] Richard Perline. Zipf's law, the central limit theorem, and the random division of the unit interval. *Physical Review E*, 54(1):220–223, 1996.

[402] T. K. Philips, D. F. Towsley, and J. K. Wolf. On the diameter of a class of random graphs. *IEEE Transactions on Information Theory*, 36(2):285–288, 1990.

[403] Stuart L. Pimm, John Lawton, and Joel Cohen. Food web patterns and their consequences. 350(6320):669–674, April 25 1991.

[404] Peter Pirolli, James Pitkow, and Ramana Rao. Silk from a sow's ear: extracting usable structures from the web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 118–125. ACM Press, 1996.

[405] Jay M. Ponte and W. Bruce Croft. A language modeling approach to information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 275–281. ACM Press, 1998.

[406] Alexandrin Popescul, Lyle H. Ungar, David M. Pennock, and Steve Lawrence. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, pages 437–444. Morgan Kaufmann Publishers Inc., 2001.

[407] Denise Pumain. Scaling laws and urban systems, August 2003. Working Papers of Santa Fe Institute, 04-02-002.

[408] Amit R. Puniyani and Rajan M. Lukose. Growing random networks under constraints. Technical report, arxiv.org, 2001.

[409] Amit R. Puniyani, Rajan M. Lukose, and Bernardo A. Huberman. Intentional walks on scale-free small worlds. Technical report, arxiv.org, 2001.

[410] Davood Rafiei and Alberto O. Mendelzon. What is this page known for? computing web page reputations. *Computer Networks (Amsterdam, Netherlands: 1999)*, 33(1–6):823–835, Jun 2000.

[411] S. Redner. How popular is your paper? an empirical study of the citation distribution. *European Physical Journal B*, 4(2):131–134, 1998.

[412] Jason Rennie and Andrew McCallum. Using reinforcement learning to spider the web efficiently. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 335–343. Morgan Kaufmann Publishers Inc., 1999.

[413] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186. ACM Press, 1994.

[414] Mathew Richardson and Pedro Domingos. The intelligent surfer: Probabilistic combination of link and content information in page rank. In *Advances In Neural Information Processing Systems 14*. MIT Press, 2002.

[415] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 61–70, 2002.

[416] Matei Ripeanu, Ian Foster, and Adriana Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal*, 6(1):50–57, January/February 2002.

[417] Stephen E. Robertson and S. Walker. Some simple effective approximation to the 2-Poisson model for probabilistic weighted retrieval. In W. B. Croft and C. J. Van Rijsbergen, editors, *SIGIR 94, Proceedings Of The Seventh International Conference On Research And Development In Information Retrieval*, pages 232–241. Springer–Verlag, 1994.

[418] L. M. Sander, C. P. Warren, and I. M. Sokolov. Epidemics, disorder, and percolation. Technical report, arxiv.org, January 21 2003.

[419] Sunita Sarawagi and V. G. Vinod Vydiswaran. Learning to extract information from large domain-specific websites using sequential models. *SIGKDD Explor. Newsl.*, 6(2):61–66, 2004.

[420] Nima Sarshar, P. Oscar Boykin, and Vwani Roychowdhury. Scalable percolation search in power law networks. Technical report, arxiv.org, June 7 2004.

[421] Ramesh Sarukkai. Link prediction and path analysis using markov chains. In *Proceedings of the Ninth Annual World Wide Web Conference*, volume 33, 2000.

[422] Jacques Savoy. An extended vector-processing scheme for searching information in hypertext systems. *Information Processing and Management*, 32(2):155–170, 1996.

[423] Hinrich Schütze and Craig Silverstein. A comparison of projections for efficient document clustering. In *Proceedings Of The 20th Annual ACM SIGIR Conference On Research and Development In Information Retrieval*, pages 74–81, July 1997.

[424] John Scott. *Social Network Analysis: A Handbook*. Sage Publications, London, 1979.

[425] John R. Seeley. The net of reciprocal influence. *Canadian Journal of Psychology*, 3(4):234–240, 1949.

[426] Rituparna Sen and Mark H. Hansen. Predicting web users' next access based on log data. *Journal of Computational and Graphical Statistics*, 12(1):143–155, March 2003.

[427] Kristie Seymore, Andrew McCallum, and Ronald Rosenfeld. Learning hidden markov model structure for information extraction. In *AAAI'99 Workshop on Machine Learning for Information Extraction*, 1999.

[428] Upendra Shardanand and Pattie Maes. Social information filtering: algorithms for automating "word of mouth". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217. ACM Press/Addison-Wesley Publishing Co., 1995.

[429] John E. Shore and Rodney W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minium cross-entropy. *IEEE Transactions on Information Theory*, 26(1):26–37, January 1980.

[430] Z. Silagadze. Citations and the zipf-mandelbrot's law. *Complex Systems*, 11(487–499), 1997.

[431] Craig Silverstein, Monika Henzinger, Hannes Marais, and Michael Moricz. Analysis of a very large altavista query log. Technical Report 1998-014, Digital SRC, 1998. http://gatekeeper.dec.com/pub/DEC/SRC/technical-notes/abstracts/src-tn-1998-014.html.

[432] Herbert A. Simon. On a class of skew distribution functions. *Biometrika*, 42(4):425–440, December 1955.

[433] Herbert A. Simon. *Models of Man.* Wiley, New York, 1957.

[434] Amit Singhal and Marcin Kaszkiel. A case study in web search using trec algorithms. In *Proceedings of the tenth international conference on World Wide Web*, pages 708–716. ACM Press, 2001.

[435] Frantisek Slanina and Miroslav Kotrla. Random networks created by biological evolution. *Physical Review Letters E*, 62(5):6170–6177, November 2000.

[436] Noam Slonim, , and Naftali Tishby. Document clustering using word clusters via the information bottleneck method. In *Research and Development in Information Retrieval*, pages 208–215, 2000.

[437] Henry Small. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24:265–269, 1973.

[438] Ricard V. Solé, Romualdo Pastor-Satorras, Eric D. Smith, and Thomas Kepler. A model of large-scale proteome evolution. Working Papers of Santa Fe Institute, 01-08-041.

[439] Sorin Solomon and Sergei Maslov. Pareto laws in financial autocatalytic/multiplicative stochastic systems. Technical report, arxiv.org, 2000.

[440] Sorin Solomon and Peter Richmond. Stability of pareto-zipf law in non-stationary economics. In A. Kirman and J. B. Zimmerman, editors, *Economics with Heterogeneous Interacting Agents*, page 141. Springer, Berlin, 2001. Lecture Notes in Economics and Mathematical Systems.

[441] Didier Sornette and Rama Cont. Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *Journal of Physics I, France*, 7:431–444, 1997.

[442] Wataru Souma, Yoshi Fujiwara, and Hideaki Aoyama. Small-world effects in wealth distribution. Technical report, arxiv.org, 2001.

[443] Olaf Sporns and Rolf Koetter. Motifs in brain networks. *Public Library of Science*, 2(11):1910–1918, November 2004.

[444] Mark Steyvers and Joshua B. Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Submitted To Cognitive Science.

[445] Nicola Stokes, Eamonn Newman, Joe Carthy, and Alan F. Smeaton. Broadcast news gisting using lexical cohesion analysis. In Sharon McDonald and John Tait, editors, *Proceedings of the Document Understanding Conference (DUC), 2004*, volume 2997 of *Lecture Notes in Computer Science*, pages 209–222, Sunderland, UK, April 5-7 2004. Springer.

[446] Steven H. Strogatz. Exploring complex networks. *Nature*, 410(6826):268–276, March 15 2001.

[447] Bosiljka Tadić. Access time of an adaptive random walk on the world wide web. Technical report, arxiv.org, 2001.

[448] Bosiljka Tadić. Adaptive random walks on the class of web graph. *The European Physical Journal B*, 23(2):221–228, September 2001.

[449] Bosiljka Tadić. Dynamics of directed graphs: The World Wide Web. *Physica A*, 293(2):273–284, April 1 2001.

[450] Bosiljka Tadić. Temporal fractal structures: Origin of power-laws in the world wide web. Technical report, arxiv.org, 2001.

[451] Ben Tasker, Pieter Abbeel, and Koller Daphne. Discriminative probabilistic models for relational data. In *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence (UAI-02)*, pages 485–492, San Francisco, CA, 2002. Morgan Kaufmann Publishers.

[452] Linda Tauscher and Saul Greenberg. Revisitation patterns in world wide web navigation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 399–406. ACM Press, 1997.

[453] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[454] Constantino Tsallis and Marcio P. de Albuquerque. Are citations of scientific papers a case of nonextensivity? *The European Physical Journal B*, 13(4):777–780, February 2000.

[455] Peter D. Turney. Coherent keyphrase extraction via web mining. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 434–439, Acapulco, Mexico, August 9-15 2003. NRC-46496.

[456] Peter D. Turney. Human-level performance on word analogy questions by latent analysis. Technical Report NRC-47422, National Research Council of Canada: Institute for Information Technology, December 6 2004.

[457] Peter D. Turney and Michael L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):325–346, October 2003. NRC-47422.

[458] Peter Uetz, Loic Giot, Gerard Cagney, Traci A. Mansfield, Richard S. Judson, James R. Knight, Daniel Lockshon, Vaibhav Narayan, Maithreyan Srinivasan, Pascale Pochart, Alia Qureshi-Emili, Ying Li, Brian Godwin, Diana Conover, Theodore Kalbfleisch, Govindan Vijayadamodar, Meijia Yang, Mark Johnston, Stanley Fields, and Jonathan M. Rothberg. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae. Nature*, 403(6770):623–627, February 10 2000.

[459] L. Ungar and D. Foster. Clustering methods for collaborative filtering. In *Proceedings of the Workshop on Recommendation Systems*, 1998.

[460] K. P. Unnikrishnan, Ramasamy Uthurusamy, and Jiawei Han. The third sigkdd workshop on mining temporal and sequential data (kdd/tdm 2004). *SIGKDD Explor. Newsl.*, 6(2):152–152, 2004.

[461] Sergi Valverde, Ramon Ferrer i Cancho, and Ricard V. Solé. Scale-free networks from optimal design. Technical report, arxiv.org, 2002.

[462] Alexei Vázquez. Statistics of citation networks. Technical report, arxiv.org, 2001.

[463] Alexei Vázquez, A. Flammini, A. Maritan, and Alessandro Vespignani. Modeling of protein interaction networks. Technical report, arxiv.org, 2001.

[464] Alexei Vázquez, Romualdo Pastor-Satorras, and Alessandro Vespignani. Large-scale topological and dynamical properties of the internet. *Physical Review E*, 65(6):066130–1–066130–12, June 2002.

[465] Jean Véronis. HyperLex: Lexical cartography for information retrieval. *Computer Speech and Language*, 18(3):223–252, July 2004.

[466] Dmitri Volchenkov and Philippe Blanchard. An algorithm generating scale free graphs. Technical report, arxiv.org, 2002.

[467] Andreas Wagner. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Molecular Biology and Evolution*, 18(7):1283–1292, July 2001.

[468] Andreas Wagner. The connectivity of large genetic networks: Design, history, or mere chemistry?, 2003. Working Papers of Santa Fe Institute, 03-11-062.

[469] Andreas Wagner and David A. Fell. The small world inside large metabolic networks. *Proceedings of the Royal Society of London B*, 268(1478):1803–1810, September 7 2001.

[470] Jill Walker. Links and power: The political economy of linking on the web. In *Proceedings of the 13th Conference on Hypertext and Hypermedia*, 2002.

[471] C. P. Warren, L. M. Sander, and I. M. Sokolov. Geography in a scale-free network model. *Physical Review E*, 66:056105–1–056105–5, 2002.

[472] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods And Applications*. Cambridge University Press, Cambridge, Uk, 1994.

[473] Duncan J. Watts. *Small Worlds*. Princeton University Press, Princeton, NJ, 1999.

[474] Duncan J. Watts, Peter Sheridan Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296(5571):1302–1305, May 17 2002.

[475] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, June 4 1998.

[476] Martin Weigt and Alexander K. Hartmann. The number of guards needed by a museum: A phase transition in vertex covering of random graphs. *Physical Review Letters*, 84(26):6118–6121, June 26 2000.

[477] Ron Weiss, Bienvenido Vélez, and Mark A. Sheldon. HyPursuit: a hierarchical network search engine that exploits content-link hypertext clustering. In *Proceedings of the the seventh ACM conference on Hypertext*, pages 180–193. ACM Press, 1996.

[478] D. R. White and M. E. J. Newman. Fast approximation algorithms for finding node-independent paths in networks, 2001. Working Papers of Santa Fe Institute, 01-07-035.

[479] Harrison C. White. Search parameters for the small world problem. *Social Forces*, 49(2):259–264, December 1970.

[480] Richard J. Williams and Neo D. Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180–182, March 9 2000.

[481] Richard J. Williams, Neo D. Martinez, Eric L. Berlow, Jennifer A. Dunne, and Albert-László Barabási. Two degrees of separation in complex food webs, 2001. Working Papers of Santa Fe Institute, 01-07-036.

[482] Guirong Xue, Huajun Zeng, Zheng Chen, Weiying Ma, Hongjiang Zhang, and Chaojun Lu. Implicit link analysis for small web search. In *Proceedings of the 26th Annual ACM SIGIR Conference On Research and Development In Information Retrieval*, pages 56–63, July 2003.

[483] Soon-Hyung Yook, Hawoong Jeong, and Albert-László Barabási. Modeling the internet's large-scale topology. Technical report, arxiv.org, 2001.

[484] Soon-Hyung Yook, Hawoong Jeong, Albert-László Barabási, and Yuhai Tu. Weighted evolving networks. *Physical Review Letters*, 86(25):5835–5838, June 18 2001.

[485] H. Peyton Young. The diffusion of innovations in social networks. Working papers of Santa Fe Institute, 02-04-018.

[486] H. Peyton Young. Condorcet's theory of voting. *American Political Science Review*, 82(4):1231–1244, December 1988.

[487] Mohammed J. Zaki, Shinichi Morishita, and Isidore Rigoutsos. Report on biokdd04: workshop on data mining in bioinformatics. *SIGKDD Explorations Newsletter*, 6(2):153–154, 2004.

[488] Damian H. Zanette. Critical behavior of propagation on small-world networks. *Physical Review E*, 64(5):050901–1–050901–4, November 2001.

[489] Damian H. Zanette. Criticality of rumor propagation on small-world networks. Technical report, arxiv.org, 2001.

[490] Damian H. Zanette. Dynamics of rumor propagation on small-world networks. *Physical Review E*, 65(4):041908–1–041908–9, April 2002.

[491] Damian H. Zanette and Marcelo Kuperman. Effects of immunization in small-world epidemics. *Physica A*, 309(4):445–452, June 15 2002.

[492] Damian H. Zanette and Susanna C. Manrubia. Vertical transmission of culture and the distribution of family names. *Physica A*, 295(1):1–8, June 1 2001.

[493] Nouradine Zekri and Jean-Pierre Clerc. Statistical and dynamical study of disease propagation in a small world network. *Physical Review E*, 64(5):056116–1–056116–6, November 2001.

[494] Ruth Yuee Zhang, Laks V. S. Lakshmanan, and Ruben H. Zamar. Extracting relational data from html repositories. *SIGKDD Explorations Newsletter*, 6(2):5–13, 2004.

[495] Ya Zhang, Chao-Hsien Chu, Xiang Ji, and Hongyuan Zha. Correlating summarization of multi-source news with k-way graph bi-clustering. *SIGKDD Explorations Newsletter*, 6(2):34–42, 2004.

[496] George K. Zipf. *Human Behaviour and the Principle of Least Effort.* Addison-Wesley, Cambridge, MA, 1949.