CS 685 The Structure of Information Networks February 4, 2005

Reaction Paper 1

Due Date: February 18, 2005

The first reaction paper will cover the topics of the genesis of the web, the global structure of the web, and link-based methods in web search; the topics are defined by the list of papers below.

Recall that a reaction paper will cover two related papers associated with this segment of lectures; at least one of these must not be on the given list. The reaction paper should be about three pages of discussion related to the papers. The discussion should go beyond a mere summary of the two papers. For instance, you might discuss some weaknesses in the papers, and suggest alternative approaches; formulate different hypotheses than the papers and suggest experiments that could be used to validate the hypotheses; discuss related work from a different field that the authors were not aware of, or a potential application of the work to a different field.

The papers listed below are the same as those listed on the web site under items 1 & 2; links to these papers can be found on the web site.

- Overview and global web structure.
  - V. Bush. *As We May Think.* Atlantic Monthly, July 1945.
  - World Wide Web Consortium. *A Little History of the World Wide Web, 1945-1995.*
  - R. Albert, H. Jeong, A. Barabasi. *Diameter of the World-Wide Web.* Nature 401:130-131, 1999.
  - A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener. *Graph structure in the web.* 9th International World Wide Web Conference, May 2000.
  - S. Dill, R. Kumar, K. McCurley, S. Rajagopalan, D. Sivakumar, A. Tomkins. *Self-similarity in the Web.* 27th International Conference on Very Large Data Bases, 2001.
  - Krishna Bharat, Andrei Broder, Monika Henzinger, Puneet Kumar, and Suresh Venkatasubramanian. *The Connectivity Server: fast access to linkage information on the Web.* Proc. 7th International World Wide Web Conference, 1998. Also, Computer Networks and ISDN Systems, 30:469-477, 1998.
- The HITS and PageRank algorithms
  - J. Kleinberg. *Authoritative sources in a hyperlinked environment.* Journal of the ACM, 46:604-632, 1999.
  - S. Brin and L. Page. *The Anatomy of a Large-Scale Hypertextual Web Search Engine.* Proc. 7th International World Wide Web Conference, 1998. Also Computer Networks and ISDN Systems, 30:107-117, 1998.

- Connections with citation analysis
  - E. Garfield. *Citation analysis as a tool in journal evaluation.* Science, 178:471-479, 1972.
  - G. Pinski, F. Narin. Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. Information Processing and Management, 12(1976), pp. 297–312.
- Comparisons of indegree and PageRank-based link methods
  - B. Amento, L. Terveen, and W. Hill. *Does "Authority" Mean Quality? Predicting Expert Quality Ratings of Web Documents.* 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2000.
  - T. Upstill, N. Craswell, and D. Hawking. *Predicting fame and fortune: PageRank or indegree?* Proceedings of the 8th Australasian Document Computing Symposium, 2003.
- Probabilistic models for link analysis
  - D. Cohn and H. Chang *Probabilistically Identifying Authoritative Documents.* 17th International Conference on Machine Learning, 2000.
  - R. Lempel, S. Moran. *The Stochastic Approach for Link-Structure Analysis (SALSA) and the TKC Effect.* 9th International World Wide Web Conference, May 2000.
  - A. Borodin, J. S. Rosenthal, G. O. Roberts, P. Tsaparas, *Finding Authorities and Hubs From Link Structures on the World Wide Web.* 10th International World Wide Web Conference, May 2001.
- Issues in and extensions of link-based methods
  - B. D. Davison. *Recognizing Nepotistic Links on the Web.* AAAI Workshop on Artificial Intelligence for Web Search, 2000.
  - K. Bharat and M. R. Henzinger. *Improved algorithms for topic distillation in a hyperlinked environment.* 21st International Conference on Research and Development in Information Retrieval (SIGIR 1998).
  - S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. *Mining the link structure of the World Wide Web.* IEEE Computer, August 1999.
  - D. Cohn, T. Hofmann, *The Missing Link – A Probabilistic Model of Document Content and Hypertext Connectivity.* Advances in Neural Information Processing Systems (NIPS) 13, 2000.
  - S. Chakrabarti, M. M. Joshi and V. B. Tawde. *Enhanced topic distillation using text, markup tags, and hyperlinks.* 24th Annual International Conference on Research and Development in Information Retrieval, 2001.

  - D. Rafiei, A. Mendelzon. *What is this Page Known for? Computing Web Page Reputations.* Proc. WWW9 Conference, Amsterdam, May 2000
  - P. Domingos, M. Richardson. *The Intelligent Surfer: Probabilistic Combination of Link and Content Information in PageRank.* Advances in Neural Information Processing Systems 14, 2002.
  - T. H. Haveliwala. *Topic-Sensitive PageRank.* 11th International World Wide Web Conference, 2002.

- Uses of anchortext in page-finding tasks
  - N. Craswell, D. Hawking, and S. Robertson. *Effective site finding using link anchor information.* 24th International Conference on Research and Development in Information Retrieval (SIGIR 2001), 250-257, 2001.
  - N. Eiron, and K. McCurley. *Analysis of anchor text for web search.* 26th International Conference on Research and Development in Information Retrieval (SIGIR 2003), 459-460 (abstract), 2003.
  - T. Upstill, N. Craswell, and D. Hawking. *Query-independent evidence in home page finding.* ACM Transactions on Information Systems, 21:286-313, 2003.

- Related papers
  - A. Broder. *A taxonomy of web search.* SIGIR Forum, 36:3-10, 2002.