Linear Algebra and its Applications 435 (2011) 908–909

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Book Review

Review of: "Numerical algorithms for personalized search in self-organizing information networks" by Sep Kamvar, Princeton Univ. Press, 2010, 160pp., ISBN13: 978-0-691-14503-7

Sep Kamvar's monograph presents work from his 2004 thesis on how numerical algorithms are a critical component in search systems. Kamvar focuses on the new information networks arising in the late 1990s and early 2000s. When originally published, these results were groundbreaking and resulted in an NSF press release [1] and even a Slashdot listing. [2] The book is divided into two parts, one for each of the two types of distributed networks discussed: the world-wide web (Part 1) and peer-to-peer file sharing networks (Part 2). Chapter 1 includes a brief synopsis of these two parts; indeed, it is much like the following two paragraphs of this review.

Part 1 focuses on the PageRank algorithm, a technique used by the Google search engine to evaluate the quality of a web-page. Chapter 2 tersely introduces the overall PageRank idea, the underlying PageRank Markov chain, and how the power method computes the PageRank vector, which is the stationary distribution of the PageRank Markov chain. Chapter 3 presents a theorem on the eigenvalues of the transition matrix of the PageRank Markov chain – the so-called *Google matrix*. Using these eigenvalues, the next chapter analyses the condition number of a linear system to compute the PageRank vector. Chapters 5–7 develop three different techniques to accelerate computing the PageRank vectors. The first technique (Chapter 5) is an extension of classic extrapolation approaches such as Aitkin extrapolation and quadratic extrapolation, albeit customized to the PageRank problem, which reduces the wall-clock time by 30%. The second technique (Chapter 6) capitalizes on the different convergence behaviors of individual components of the PageRank vector. Some components converge after only a few iterations and eliminating computation associated with these components reduces the overall computation time. The final technique (Chapter 7) exploits block structure in the PageRank matrix of the web-graph to develop a multi-level approach. Kamvar's goal with these three techniques is to accelerate PageRank to enable a scalable personalized PageRank approach, where there is a PageRank vector for each user of the search engine. At the end of this chapter, he discusses how BlockRank achieves this goal.

Part 2 tackles search in peer-to-peer file sharing networks, like those used in the original-but-now-defunct Napster and Kazaa applications. In such systems, queries for information are propagated among connected clients. After a query produces a result, the clients form a new connection and transmit the information. Everything is decentralized in these networks. Consequently, reputable behavior, which is accurately responding to queries and file transmission requests, is a prized com-

[1] http://www.nsf.gov/od/lpa/news/03/pr0356.htm, last accessed on January 4, 2011.

[2] http://developers.slashdot.org/story/03/05/14/2117231/Compute-Googles-PageRank-5-Times-Faster, last accessed on January 4, 2011; Slashdot is a popular technological news source partially known for the propensity of its articles to cause outages on websites due to *too many* visitors – the "Slashdot" effect.

doi:10.1016/j.laa.2011.01.013

modity. Chapter 8 begins the exploration by describing a realistic simulation system for a peer-to-peer network; this system is used for the experimental evaluations in the remaining chapters. Using this simulator, Chapter 9 evaluates a distributed trust and reputation protocol called EigenTrust because of its relationship with an eigenvector. Chapter 10 introduces adaptive network topologies, whereby clients try to find and connect to similar peers. The idea is that a well defined adaptive topology with trust should move malicious peers (those that provide incorrect information) and free-riding clients (those that only take information) to the periphery of the system. Thorough experimental evaluations against a variety of different types of malicious behavior validate this protocol.

Not much of the text seems to have been edited from the 2004 thesis [4], or some of the original articles. As such, this book is most appropriate for those with some prior background in mathematical models and algorithms for ranking and reputation problems. Langville and Meyer's book on PageRank [5] would provide a good introduction. Indeed, Langville and Meyer devote Chapter 9 of their text to a brief discussion of precisely the three techniques to accelerate PageRank from Chapters 5 to 7.

The book is written in a pleasant style, with clearly stated results. For example, the proof in Chapter 3 proceeds from the perspective of graph theory, Markov chain theory, and eigenvalue analysis. I find this proof provides more insight into the problem structure than the purely algebraic proof found in Langville and Meyer. The experimental results from Part 2 are likewise thoroughly and clearly explained.

However, to compare the book against a broader, more recent, treatment is to miss the point. It represents the research perspective of the early 2000s, when the question of how to accelerate PageRank and manage distributed reputation were still open questions. It also represents a uniquely curated collection of research ideas, many of which had only been distributed as Stanford technical reports [1–3]. This means it will become an important, authoritative source for these topics. For example, I had not been aware of the A^d extrapolation procedure in Chapter 5, which appears to be a highly competitive algorithm for PageRank, or the author's work on adaptive P2P topologies, which would also apply to other distributed computations. Others researching network measures like PageRank and reputation measures like EigenTrust will likely feel similarly, and this book should find a home on their shelves.

As a postscript to this review, please let me note that the author is donating the proceeds from the book sales to the Gene Golub Memorial Fellowship and the Rajeev Motwani Foundation. Stanford Professors Golub and Motwani were two of Kamvar's collaborators and mentors, both of whom died unexpectedly.

David F. Gleich
Sandia National Laboratories,
Livermore, CA, USA
Tel.: +1 925 294 6769
E-mail address: dfgleic@sandia.gov

References

[1] Taher Haveliwala, Sepandar D. Kamvar, Dan Klein, Christopher Manning, Gene Golub, Computing PageRank using power extrapolation, Technical Report, Stanford University, 2003.
[2] Taher H. Haveliwala, Sepandar D. Kamvar, The second eigenvalue of the Google matrix, Technical Report 2003-20, Stanford University, March 2003 (accessed 08.04.07).
[3] Sepandar Kamvar, Taher Haveliwala, Christopher Manning, Gene Golub, Exploiting the block structure of the web for computing PageRank, Technical Report 2003-17, Stanford InfoLab, 2003.
[4] Sepandar D. Kamvar, Personalized Search in Self-organizing Information Networks, Ph.D. Thesis, Stanford University, June 2004.
[5] Amy N. Langville, Carl D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings, Princeton University Press, 2006.