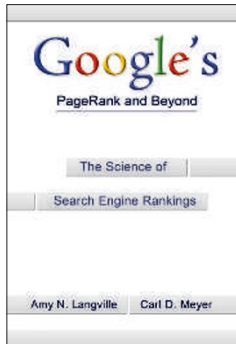


Under the bonnet



GOOGLE'S PAGERANK AND BEYOND: THE SCIENCE OF SEARCH ENGINE RANKINGS BY AMY N. LANVILLE & CARL D. MEYER

Princeton Univ. Press: 2006. 232 pp. \$35

Google is the buzzword on everyone's lips, but it is PageRank, the system of rating the relative importance of every page in cyberspace, that is the secret of the ubiquitous search engine's success. Developed by Sergey Brin and Larry Page, the principle on which PageRank operates is simply stated — a page is important if it is pointed to by other important pages. In practice, calculating the importance of every one of the more than eight billion pages in Google's database is a gargantuan task. This book attempts to tell the story of the mathematical engine that does it, and begins, oddly enough, in Tsarist Russia.

In 1906, the mathematician Andrei Markov wrote a seminal paper on the properties of an infinite random sequence of numbers, constructed such that the probability of a given number occurring at a given point in the sequence is determined only by the value of the number that immediately precedes it. To understand the relevance to PageRank of such a sequence — known as a Markov chain — consider a Generation-X-er who walks into a cybercafé and begins surfing the web aimlessly. He just clicks links at random, starting from whatever webpage was on the screen when he first sat down. If he does this indefinitely, the sequence of pages he visits will have all of the attributes of a Markov chain. More importantly, the distribution of pages within this chain will contain information about the structure and connectivity of the web. The theory of Markov chains and the tools of linear algebra provide a powerful framework for understanding, and for efficiently extracting, this information.

If I were taking, or teaching, a course in linear algebra today, this book would be a godsend. My first experience of linear algebra at university was not a good one. The scope of the course that introduced me to it was so narrow — restricted to the calculation of vector transformations in two and three

dimensions — that my overriding impression was that it invoked an unnecessary degree of abstraction to do something that was essentially trivial. That of course changed when I discovered, semesters later, its application to modern physics. But at the time, simply being introduced to the idea that these tools could be used to manage a matrix calculation of order eight-billion would have made its content seem much more relevant, and a heck of a lot more interesting.

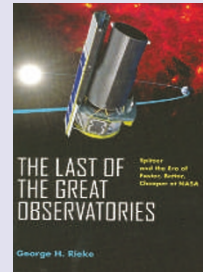
Despite this, this book falls into the common trap of trying to be too many things to too many people. The preface suggests that the intended audience includes both the general and the technical science reader. Although it makes a worthy attempt, the book never really gets close to pulling this off. The main motivation of the authors seems to have been to convey a sense of the beauty of linear algebra. But what they provide seems to be more of a nuts and bolts description, like an owner's manual, of what's under the PageRank bonnet. To justify marketing it to a broader audience, it is sprinkled with a handful of tantalizing anecdotes about the broader issues that have influenced its development — such as the unscrupulous 'link spammers' who exploit weakness in the PageRank calculation for commercial ends; or the amusing practice of 'Googlebombing', the most famous instance of which ranks George W. Bush's autobiography as the top entry for the search term 'failure'. But these are all too superficial and brief to satisfy the general reader.

Ultimately, the book I would have preferred to read is the book I suspect the authors would have preferred to have written — about the beauty of mathematics, made relevant by its application to PageRank, rather than a book about PageRank footnoted with mathematical ideas, inadequately explored.

Ed Gerstner

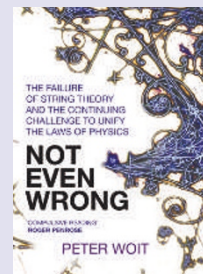
Ed Gerstner is a Senior Editor of Nature Physics

On our bookshelf



The Last of the Great Observatories: Spitzer and the Era of Faster, Better, Cheaper at NASA
by George H. Reike
University of Arizona
Press: 2006. 228 pp. \$40.

The Spitzer Space Observatory, known as the 'Infrared Hubble', was launched in 2003. Reike takes us behind the scenes, revealing the political, financial and scientific mood-swings over two decades.



Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Law of Physics
by Peter Woit
Jonathan Cape, London:
2006. 274 pp. £18.99.

A history of how string theory became the alpha male of theoretical particle physics, despite its inability to make any predictions or be disproved.