self to them entirely. *Récoltes et Semailles* explains much more clearly than his letters how he came to feel that doing mathematics, while in itself a pursuit of extraordinary richness and creativity, was less important than turning towards aspects of the world which he had neglected all his life: the outer world, with all of what he perceived as the dangers of modern life, subject as it is to society's exploitation and violence, and the inner world, with all its layers of infinite complexity to be explored and discovered. And, apart from the sporadic bursts of mathematics of the 1980s and early 1990s, he chose to devote the rest of his life to these matters, while Serre continued to work on mathematics, always sensitive to the excitement of new ideas, new areas, and new results. In some sense, the difference between them might be expressed by saying that Serre devoted his life to the pursuit of beauty, Grothendieck to the pursuit of truth.

Equipe Analyse Algébrique
Universite de Paris 6
175 rue du Chiraleret
75013 Paris
France
e-mail: Leila@math.jussieu.fr

Google's PageRank and Beyond: The Science of Search Engine Rankings

by Amy N. Langville and Carl D. Meyer

PRINCETON, NEW JERSEY, AND OXFORD, PRINCETON UNIVERSITY PRESS, 2006, 234 PP., $35.00/£22.95, ISBN-13; 978-0-691-12202-1

REVIEWED BY PABLO FERNÁNDEZ

First Scene: California, 1998. Jon Kleinberg, a young scientist working at IBM's Almaden Research Center in Silicon Valley, is presenting his HITS (Hypertext Induced Topic Search) algorithm. Almost simultaneously, at nearby Stanford University, two Computer Science doctoral students, Sergey Brin and Larry Page, are putting the finishing touches on their PageRank algorithm, a future core component of Google's search engine. Both projects are based on an innovative idea: using the hyperlink structure of the web to enhance a search engine's results.

Second Scene: Madrid-Mountain View, California, 2006 (same characters). Kleinberg, now a professor at Cornell University, is receiving the Rolf Nevanlinna Prize at the International Congress of Mathematicians in Madrid. Brin and Page, sitting at their offices in Google's headquarters, are plotting some new *google-gadgets* to delight the immense host of *googlemaniacs.*

Not a lot of time separates these two scenes, but it was enough to completely revolutionize the way we use the web. Google has become the standard for web searching. Its impact has moved far beyond the world of t echnology; it pervades our daily lives. While Kleinberg's work has not been so successfully commercially developed, it has earned him the recognition of the mathematical community with the aforementioned Nevanlinna Prize, given "for outstanding contributions in mathematical aspects of Information Science."

The book under review gives a comprehensive overview of the state-of-the-art technology of web search engines, putting special emphasis, as the title suggests, on ranking procedures.

A search engine is designed to perform several tasks. First, it collects the information contained in the myriad pages of the web ("crawling," in the jargon). Next, all of this information is stored, compressed, and processed to build content indexes. Finally comes the interaction with the user. When a user types a query the engine must find in the indexes the pages that contain relevant information and must show this outcome as an ordered list. Now comes the key point, perhaps the most important ingredient of the search process: In which order should the information be displayed? It is desirable, even essential, that most of the time the user be able to find the most relevant information in, say, the first 10 or 20 displayed pages.

The scale of the problem is enormous. There are billions of pages on the web, with an estimated average page size of 500 Kb, and all of these figures are increasing day by day. The web is also dynamic: most pages change their contents on a regular basis and millions of pages are added (or disappear) each year. The engine, of course, must respond to queries in real time!

Considering this, it is remarkable that the mathematical ideas behind the ranking algorithms that make effective searching possible require only some basic tools from linear algebra.

Remember our basic question: after the relevant pages for a query have been selected, we need to assign a *score* to them to determine the order in which to display them. Part of this score is related to the position of the query term within the document (in the title, in the body) or, for combined searches, to the distance between the terms in the text. Each search engine has different rules for assigning this "content score." But there is another score, to be combined with the former score, to make up the overall score, which should not be query dependent, but should reflect the global relevance of each page. This is the "popularity score," given by the ranking schemes that are the focus of our book.

The idea is simple: just regard the hyperlinks as recommendations, with two extra comments: (1) the status of the recommender is important, and (2) the recommendation should drop in weight if the recommender is too generous giving them. In short, a web page is important (gets a high popularity score) if it is pointed to by other important (high ranked) pages. All these features can be formalized with a fairly simple mathematical model.

Let us imagine that there are n pages in the web, and let us consider the nxn adjacency matrix H for the internet, viewed as a directed graph. This internet graph has web pages as vertices and hyperlinks as edges joining these vertices. The (i,j) entry of H is 1, if there is a directed edge from vertex i to vertex j, and 0 otherwise. To fulfil demand (2) above, PageRank considers a (row) normalized version, H', of this matrix, in which all the entries of each row are divided by the total sum of the row (the number of out-links of the page). The resulting matrix is (almost) *stochastic*, so its entries may be viewed as *probabilities*. A nice interpretation of this formulation is a random surfer traveling along the web, following the links between nodes, and uniformly choosing his next destination from among the links in the current node. It turns out that the ranking list we need is just the stationary vector of the Markov chain associated to matrix H'. In fact, some adjustment must be done in order to have stochasticity, because H could have zero rows (pages with no out-links). In these cases we can replace these 0-rows with $1/n$ entries. Call this adjusted and normalized matrix J.

But now we face a computational problem: How can we determine this vector? Remember the matrix is extremely large. The idea is now to make a second adjustment, to guarantee that the new matrix also be *primitive*. Then the stationary vector exists, is unique, and, as it is the eigenvector associated to the dominant eigenvalue of the matrix, can be calculated with a simple and fast numerical procedure such as the power method. The reader should be aware that the Perron–Frobenius Theorem on positive (or non-negative) matrices plays a key role in all these arguments. The precise adjustment is given by

$$G = \alpha\, J + (1 - \alpha)\, \frac{1}{n\, \mathbf{e} \cdot \mathbf{e}^T}$$

where $\mathbf{e}$ stands for the vector of all ones and α is a number between 0 and 1. G is known as Google's matrix. PageRank output is just its dominant eigenvector, which can be obtained with an iterative method such as $\alpha^{k+1} = \alpha^k\, G$. In terms of the random surfer, this adjustment brings in a new possibility, namely that with probability $1 - \alpha$ the surfer gets bored of following the links and "teleports" to any page of the web. The reader could argue that this is an "artificial" matrix (for instance, Google's choice of the value of the teleportation constant, α, is around 0.85). And indeed it is, but it allows effective computation of the ranking vector and, above all, it works! The resulting output is incredibly good at assigning relevance to web pages.

So this is the model. Not too complicated, is it? Of course, there are many details to complete: computational aspects (such as convergence rates of the iterative scheme, sensitivities to the parameters), possible improvements of the numerical procedures (or the model itself) . . .

Most of this can be found in the fifteen chapters of the book under review. The first three chapters introduce the reader to the main features of web searching, including some review of the traditional methods of information retrieval. Chapters 4 to 10 deal with the mathematics behind Google's algorithm: the Markov chain model, numerical procedures, sensitivities to parameters, convergence issues, methods for updating the rankings, etc. All the mathematical concepts used in the book are treated in detail in the "Mathematical Guide" of Chapter 15: linear algebra, Markov chains, Perron–Frobenius Theory, etc. Chapter 11 includes a brief review of Kleinberg's HITS algorithm; other ranking methods are mentioned in Chapter 12. Chapter 13 discusses some questions related to the future of web information retrieval, including spam and personalized searches. I would have liked to see a more comprehensive discussion of ethical issues such as privacy and censorship. Considering that Google has become the standard source for information (you appear in Google or you are nothing!), these are really disturbing topics. But probably a whole new book could be written on this.

The book under review is excellently written, with a fresh and engaging style. The reader will particularly enjoy the "Asides" interspersed throughout the text. They contain all kind of entertaining stories, practical tips, and amusing quotes. "How do search engines make money?," "Google bombs," "The Google dance," and "The ghosts of search" are some of these stimulating asides. The book also contains some useful resources for computation: Pieces of Matlab code are scattered throughout the book, and Chapter 14 contains a guide to web resources related to search engines.

Despite the technical sophistication of the subject, a general science reader can enjoy much of the book—certainly Chapters 1 to 3, and also Chapter 13. With some basic knowledge of linear algebra, the description of the model (Chapters 4 and 5) can be followed without problem. Chapters 6 to 12 are more technical, and they are intended for experts. The authors provide a webpage (http://pagerankandbeyond.com/) that includes a list of errata for this edition.

Departamento de Matemáticas
Facultad de Ciencias
Universidad Autónoma de Madrid
Ciudad Universitaria de Cantoblanco
28049 Madrid, Spain
e-mail: pablo.fernandez@uam.es

Mathematical Form: John Pickering and the Architecture of the Inversion Principle

by Pamela Johnston (ed.)

With contributions by Mohsen Mostafavi, George L. Legendre, John Pickering, Chris Wise, John Silver, and John Sharp

LONDON: ARCHITECTURAL ASSOCIATION, 2006, 96 PP., £15.00, ISBN 978-1-902902-37-1

REVIEWED BY KIM WILLIAMS

John Pickering's models—or buildings—of geometrical forms, derived through the process of inversion, are the subject of this little book. The book is essentially a catalogue of the 2002 exhibit of his works in the gallery of London's Architectural Association. Pickering is an artist (the exhibit included his earlier studies of the human form, but this book does not), who at a certain point, in his own words, "became anti-nature" and dedicated himself to the derivation and visualization of complex three-dimensional geometrical forms. To create his forms, Pickering employs the "inversion principle," transformations of either plane or solid