

Journal-Ranking.com: An Online Interactive Journal Ranking System

Andrew Lim, Hong Ma, Qi Wen, Zhou Xu

Department of Industrial Engineering and Logistics Management
Hong Kong University of Science and Technology, Clear Water Bay, Kowloon
{iealim,hongm,vendy,xuzhou}@ust.hk

Brenda Cheang, Bernard Tan, Wenbin Zhu

Red Jasper Limited
Hong Kong Science Park, Shatin, No. 5 Science Park West Ave, Hong Kong
{brendakarli, btan, zhuwb}@redjasper.com

Abstract

Journal-Ranking.com is perhaps the first online journal ranking system in the world which allows any individual to conduct citation analyses among more than 7000 academic journals with regards to his/her own interests. Besides classical statistics on citation quantity, the system also allows taking citation quality into consideration after extending on the well known PageRank method, which has been successfully applied to evaluate the impact of web pages by Google. *Journal-Ranking.com* was developed by Red Jasper Limited and the Hong Kong University of Science and Technology. The system underwent a soft launch in June 2006 and has been successfully deployed in 2007. The website has registered an average of a million hits per month since its official launch. This paper explains the development and design of *Journal-Ranking.com* and its impacts to the global research community.

Introduction

The global research community has long and fervently been debating the topic of journal impact. Well, just what is the impact of a journal? Today, the Science Citation Index (SCI) recognizes over 7,000 journals and just about every active researcher ritualistically looks to the ranking of a journal's impact when considering where to send their papers to for publishing as it is largely considered a reflection on the quality of their research. Consequently, the sheer volume of available journals renders it pivotal for researchers to accurately gauge a journal's impact because publishing their works in established journals would have significant influence on peer recognition, promotion opportunities, and so on.

In essence, there are two major approaches to evaluating journal impact. One is based on expert surveys, and the other on citation analysis (ISI, 2006). The latter, also

widely known as the ISI's Impact Factor (IF), is often considered the more objective approach since it measures the impact of a journal by its usefulness to other journals, which avoids any subjective opinions. However, ISI's approach in citation analysis has come under heavy criticism for not factoring citation quality since a citation from a prestigious journal shares the same weight as one from an unremarkable journal.

There are a handful of websites that release rankings of journals in specific areas of research. For example, the *ISI Web of Science (WOS)* publishes a journal citation report every year (ISI, 2006), in which journals are categorized into fields, and are ranked within each field according to their average number of citations received per paper. Audiences can only access the static reports with no right to change the structure of categories, or the criteria for ranking journals.

The limitations of current ranking sites inspired our team to develop *Journal-Ranking.com*, a system which would allow users to interactively configure their ranking interests, as well as provide a more reasonable method to evaluate a journal's impact.

We believe that *Journal-Ranking.com* is perhaps the first online interactive journal ranking system in the world. The system was completed and subsequently underwent a soft launch in June of 2006 where we invited some colleagues to review the system. When we were able to verify our results with our colleagues, we officially launched the service in mid January 2007.

The primary objective of *journal-ranking.com* is to attract university professors, research scholars, and students to evaluate journal impact based on their own interests. The evaluation results calculated by the system are closer to the perceptions of scholars compared with ISI's IF and yet, the ranking results are derived based on objective analyses over journals across citations.

Journal-Ranking.com has since attracted an average of a million hits per month since the official launch and membership has grown to about 3,000 with an average of between 50 to 100 new members per day.

This paper describes the steps involved in the development of *Journal-Ranking.com* from its conception to its completion, along with a breakdown of the inner workings of each portion of the system.

Our team spent a great deal of time reviewing our ranking results with many esteemed colleagues from many differing fields. We are pleased to reveal that our results, when compared with expert surveys, have shown that there are significant correlations.

The Methodology Used

The methodology we adopted involves modifying a classical scientific model known as eigenvector analysis although it is more widely known as Google's PageRank algorithm (Page, et. al. 1999).

Google's PageRank algorithm, applied in Google's search engine in 1998, was developed at Stanford University by Larry Page and Sergey Brin. It is essentially a link analysis algorithm which assigns a numerical weight to each element in a given set of articles. The purpose of assigning a weight to each element is to measure its relative importance within a particular set of articles.

Please refer to the section "Uses of AI Technology" for a more detailed description of our methodology

Formulation of the Ranking Problem

In order to support scenario analyses, the system first needs to evaluate the journal impact based on the citations network (Wasserman, Faust, 1994). Thus an instance for assessing journal impact will consist of the following components:

1. A complete set of n journals, denoted by $J = \{1, \dots, n\}$, for ranking.
2. The size of each journal j in J , denoted by a_j , which represents the number of articles published annually in journal j
3. A time period $[s, t]$ for citations for study, i.e., only citations from articles that are published from year s to year t will be considered in our analyses
4. A n times n citation matrix, denoted by $C = [c_{i,j}]$, whereby $c_{i,j}$ indicates the frequency of citations referred by articles in journal i to articles in journal j
5. A core journal set W , which is a subset of J , consisting of m journals for which we are interested in assessing the impact.

6. Two parameters to indicate the ranking preference, which include α for the weight for self-citation and β for external citation.

Based on the instances described above, the ranking system is mapped to an m dimension vector s , which indicates the impact of m journals. The sum of every dimension in s is normalized to be one.

The Usage of Journal-Ranking.com

The core function of the system is to process user queries and generate ranking results accordingly.

A typical process for creating a rank list is as follows:

1. Create a category structure (e.g., the hierarchical subject category used by Science Citation Index)

The system aims to provide convenient tools for users to create new category structures either from scratch or based on existing structures and enables saving the customized structures for later use.

The rationale behind this is because sometimes the category structures defined by SCI may not suit a particular application. For example, a computer science department is tasked to evaluate the impact of a journal. Citations from biology may be irrelevant and so, should be ignored. Citations from math, on the other hand, may be related and hence, of some importance. So in such cases, a new category structure can be created to include only relevant journals.

2. Setting parameters

The model used by the system distinguishes citations from the same journal, from the same category and from other categories. Users are able to set different weights for the three types of citations when adjusting the contributions of these citations to the final standing of a journal. The system offers a user interface to create parameter settings and allows users to store them for later use.

From one viewpoint, it might be easier to be cited by articles from the same journal (especially if the citing article is authored by the same group of people) than by articles from other disciplines. From another viewpoint, one might believe that citations from the same area indicate more relevance. Regardless of which viewpoint a user takes, this function literally empowers users with the opportunity to tinker with the various parameters.

Additionally, the system allows users to specify a period, such that only citations occurring within the period of interest is considered.

3. Setting a journal list as a filter

The system allows users to specify a list of journals, such that only journals in the given list will appear in the final result. This allows users to analyze the relative standing of journals in the specified list.

Note: The results generated by only considering citations among journals in a list L (this can be done by defining a category structure that consists of only journals in L) is different from the results generated by first ranking all journals (using the SCI category structures that consists of all categories and journals) and then, take the relative ranking of those appearing in list L (by specify a filter in this step). For example, a journal, J1, has very few citations from L, but many citations from other journals; another journal, J2, has many citations from L, but very few citations from other journals. J1 will be ranked lower than J2 if we only consider citations from L; however, J1 may be ranked much lower than J2 if we consider citations from all other journals.

This function used in conjunction with customized category structures offers users extensive flexibility.

4. Compute the ranking using our citation model (please refer to formulation and uses of AI Technology)
5. Sort the output by rank, by name, by various impact factors, by ISI IF, to analyze the result

The system presents the filtered ranking results in a table and allows users to sort the results in ascending or descending order of different columns by clicking the header. The default result is sorted by the rank of the journal.

6. Scenario analyses

By combining all of the above functions, we were able to analyze several interesting scenarios. For example, we were able to

- Find out the relative ranking of all computer science journals; consider their contributions to all journals over a specified period. This can be achieved by using the default category structure (includes all SCI defined categories and journals), set equal weights for all types of citations, and use the journal list containing all computer science journals as the filter.
- Focus on a particular area, such as computer science and ranking journals according to their impacts to computer science journals. This can be achieved by creating a category structure that contains only one category - computer science and a related journal.

- Find out the effects of journal self citation. This can be achieved by setting the weight of self citations to zero and the weight of inter-category citations to be one and intra-category citations to be one.
- Compare your results against the ISI IF. We generalized the way ISI IF is computed by providing 3 additional indicators namely B2 (2004), B4 (2004) and B6 (2004). B2 (2004) is the average number of citations received by papers published in the journal in the two preceding years before 2004 (i.e. 2003, 2002) from all the papers published in 2004. This is commonly known as the ISI Impact Factor or IF. B4 (2004) refers to the four preceding years before 2004, and B6 (2004) refers to the six preceding years before 2004.

Application Description

The system supports both static and dynamic query modes. The static query mode allows users to quickly access rank lists computed using a set of predefined rules; the dynamic query mode allows users to freely customize the ranking according to their own rules and settings.

It is expected that many users will begin using the tool by issuing simple queries. The system not only provides 27 sets of parameter settings, it also pre-computes and stores the results corresponding to the 27 default settings. The 27 default rankings utilize the category structures defined by SCI and include all journals. Users have 3 choices for the weight of self citations (0, 0.5, 1), 3 choices for the weight of inter-category citations (0, 0.5, 1) (where the weight of intra-category citations is set to 1) and 3 choices for the period of consideration: from start to year 2004, from 1995 to year 2004 or from year 2000 to year 2004. Users may choose to display all journals or only journals from one particular category (such as computer science). These types of common queries are translated into direct SQL queries and processed very quickly by the system.

For users in need of greater customization, the system can allow users to create their own category structures, parameter settings and process the queries by dynamically invoking the ranking engine (that implements the algorithm described in the model) to compute the result. The computed result is stored using the category structures and parameter settings as key. The computation of a customized ranking takes a few seconds to a few minutes depending on the number of journals involved and the parameter settings. When users repeat the queries (with different filters), the saved results are fetched and returned directly.

To enhance user experience:

1. The system tries to reuse the input from users as much as possible to minimize the user's efforts expended on the input.
 - a) The system defines default category structures and journal lists so that new category structures and journal lists can be created with small modifications of default values
 - b) User inputs such as category structure, parameter settings and filters are remembered so that they can be recalled and reused in different queries using a drop down list.
2. The system offers interactive "wild-card aware" full text search capabilities. Users can use key words to find lists of matching journals or journal lists. The search result appears in the result boxes. The search function is implemented using Web 2.0 technology so that page refreshing is unnecessary when updating the results to offer users a more responsive user interface
3. The system offers Drag and Drop support for editing customized journal lists (or category structures). Journals appearing in a search result can be dragged and dropped onto the journal list with the click of a mouse.

Hardware & Software

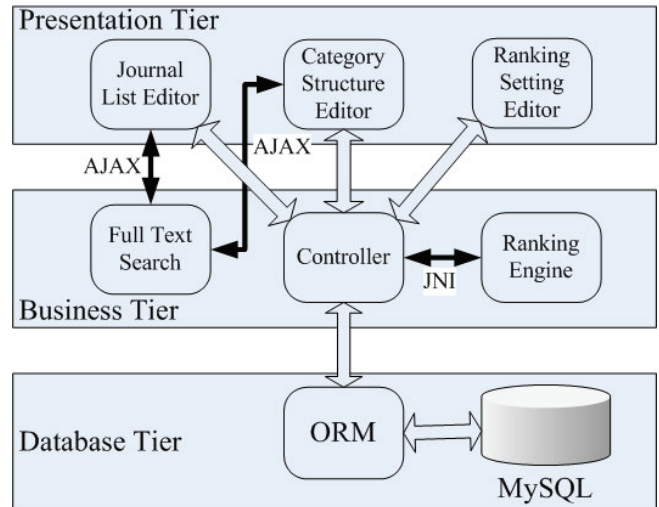
The system was coded in *Java 1.5* using the *Eclipse 3.1* software and *My SQL* databases. The central machine used was a Dell PC with a *Pentium-4 2.80GHZ CPU* and *1.00 GB RAM* memory.

System Design

The system is a typical 3-tier implementation consisting the web presentation tier, business logic tier and database tier (Reese, 1997). The web presentation was implemented using JSP/servlet technology with Java Script and AJAX for interactive web 2.0 features. The business logic was implemented on top of the spring framework. We used hibernate to map our object model to relational databases and we used the MySQL database server to store all users' data and computed ranking results.

The presentation tier handles the user interfaces, collects inputs from users and displays the results to users. The business logic tier processes the input data, fetch required data from the database tier and compute ranking. The business logic layer also ensures that application-level transactions are performed atomically and requests from different users are properly isolated to prevent interferences. The database tier provides an abstraction for managing the persistent data storage, reducing the complexity of this process.

Figure 1. System Architecture



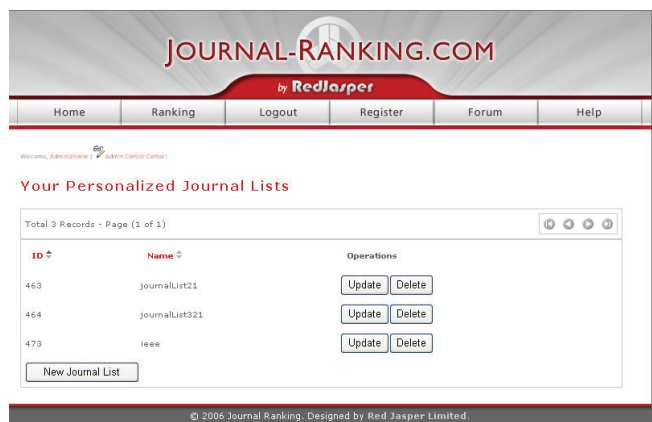
The following subsections describe the system's special modules such as, Journal List Editor, Interactive Full Text Search, etc:

Journal List Editor

This is a standard CRUD+Listing type of web application with two pages, namely listing page and updating page.

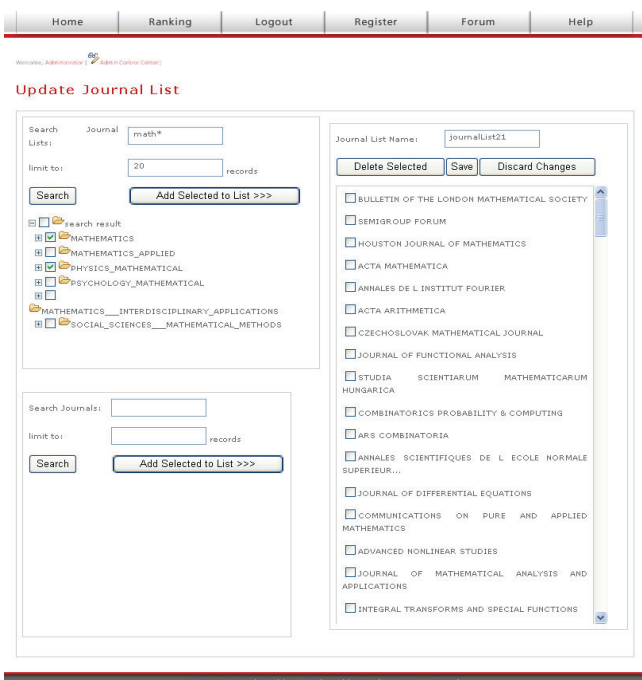
The listing page displays all users who defined the journal list items, one row per item (see figure 2). Each item has an associated Update button and Delete button. At the end of the listing page there is a New Journal List button. When the "update" button is clicked, the corresponding journal list object is loaded and displayed in an update page (see figure 3) for users to modify the attributes. When the "delete" button is clicked, the item is deleted, and the listing page is refreshed. When this is completed, an empty new journal list is added, and users are prompted to update the journal list page to modify the attributes.

Figure 2. Listing Journal Lists



To update the journal list page (Figure 3), there are three areas of focus. The upper half of the left hand side of the screen capture allows users to search journal lists using keywords. The lower half of the left hand side allows users to search individual journals by keywords. The right hand side displays the journal list under editing. The user can search for existing journal lists using key words and/or search for exist, users can drag search results to the current journal list under editing. If a journal list is dragged and dropped to the right hand side, it means all journals in that list are added as parts of current list; if a journal is dragged and dropped to the right hand side, then the journal is added to the current list. If the user finds it difficult to drag and drop, he/she can select the interested journal lists (or journals) and click Add Selected to List button to add the selected item to the current list. To delete unwanted items from the current list, users may select some journals in the current list and click the delete button to delete them (the update page will be reloaded to reflect the changes). When editing is done, the user can click the save button to save the changes to database and go back to the listing pages, or click discard changes which will directly go to the listing page without saving the changes.

Figure 3 Updating Journal List



Implementing Interactive Full Text Search

The search function provided on updating the journal list page (Figure 3) was implemented using AJAX. AJAX is a set of JavaScripts running on the client side (i.e., inside browser on client's desktop). When users click the search button, a java script function is called that will send an

asynchronous request to the server, pass the query to the server and wait for the server to process the query. The advantage of using AJAX here is two-fold:

1. Only raw input and output is communicated between the client and server in binary format, the size of data transferred is much smaller than a formatted html page as is done on conventional web applications. This results in boosting the throughput.
2. It allows interactive manipulation of the results where the user can drag and drop items between search results and the editing lists. The change is immediately reflected, which is much more responsive than a full page reload as experienced in conventional web applications.

Output

The output of query is a table, every row is a journal. Rows are sorted by rank by default. Every column specifies an attribute about the journal such as name, rank, impact factors computed by our model, ISI impact factors and number of articles in the journal. By clicking the header of a column will sort the result by that column (click once with sort in descending order indicated by a dark downwards arrow in the column header and click one more time in order to sort again but in ascending order indicated by a dark upward arrow in the column header). By default the table is sorted by Rank in ascending order.

Citation Analysis Engine

The citation analysis model was implemented in C and integrated into the backend using JNI. The algorithm used to compute the ranking is computationally intensive, and to be as fast as possible, we chose to program in C. The backend will essentially pre-process the data required and then invoke the engine via java as the native interface.

Uses of AI Technology

The major objective of *journal-ranking.com* is to evaluate journal impact in an objective manner. Citation analysis is purported to be an objective measurement based on the viewpoint that the influence of a journal (and its articles) is determined by their usage to other journals (and articles), and where their usage can be reflected by the citations that they have received. However, using only citation quantity is considered to be biased to a certain degree, because a widely-held notion is that citation quantity does not equally represent citation quality. Logic tells us that a citation by a paper published in a prestigious journal should far outweigh a citation by a paper published in an unremarkable journal.

We then applied eigenvector analysis to evaluate journal impact by taking citation importance into consideration. The method has been widely applied in image processing, pattern recognition, as well as web page rankings (such as Google's PageRank algorithm) (Page, et. al. 1999) and journal rankings (Palacios-Huerta and Volij 2004; Pinski and Narin, 1976; Bollen, et. al. 2006).

However, before applying eigenvector analysis (EA), we transformed the citation matrix C by weights in order to reflect the ranking preference. For each frequency of citation $c_{i,j}$, if i and j are equal, it is weighted by α ; if i is outside the core journal, then it is weighted by β .

To apply the EA method, we first calculate a proportional citation matrix P based on the given citation matrix C , where $p_{i,j} = c_{i,j} / \sum_{k \in J} c_{i,k}$, which indicates the proportion of

citations sent from journal i to journal j . The vector s for journal impact can thus be obtained by calculating the eigenvector of the matrix P^T , corresponding to its largest eigenvalue. In other words, s is a solution to the following equations:

$$\begin{cases} s_j = \sum_{i \in J} p_{i,j} s_i, \text{ for } j \in J \\ \sum_{j \in J} s_j = 1 \end{cases}$$

Here the journal impact is defined by the sum of the product of the citation proportion from each journal and the impact of the citing journal, which reflects the citation quality for assessing the journal impact.

Lastly, the EA method is implemented via a recurrent method, whereby values of s are regularly updated by the equations above until it converges into a stable vector. After testing through 7000+ journals, the total running time is less than 10 seconds. For personal users, the set of journals for ranking usually focuses on a particular research area, whose size of journals is far less than 7000. In this case, the ranking engine can accomplish the job almost instantaneously.

Application Use and Payoff

Journal-ranking.com will provide a platform to continuously improve the rating of journals. This, we believe, is important to the advances of research. Since the launch of this site, we have received an average of a million hits per month and approximately 3,000 users signed up as members. The membership numbers increase by between 50 to 100 daily so the statistics recorded here captures only the current state of events. These encouraging statistics will allow the founding company, Red Jasper Limited, to seek governmental and commercial support to expand this as well as related services so as to

eventually compete with ISI and other university ranking web portals.

As a secondary benefit, there are many more visitors to our company website that make enquiries about our products.

Application Development and Deployment

Journal-ranking.com is a project jointly orchestrated by Red Jasper Limited and the HKUST. The project commenced at the beginning of 2006 as the company felt that there was untapped potential in offering standardized objective online ranking services. *Journal-Ranking.com* is the first of such projects that offers a more flexible, easy to use service to rank various academic journals in the academic community. It was built with the intention of improving the outdated ISI index.

The HKUST team collected data in early 2006 from publishers of journals and the ISI's Web of Science Database. For each journal, the frequency of its recent citations appeared in 2004 and references to every other cited journal were collected, including self citations. In order to provide a default structure for journal categories, we also used the journal categories that were defined by the ISI Web of Science. Based on such a default setting, it was more convenient for users to modify category structures based on their own interests.

Besides citation data and category structures, the HKUST team also collected detailed information about journals, such as the number of articles published in a journal every year, the impact factor of a journal, etc. Such information was useful in the calculation for assessing journal impact.

The development of the model for ranking began in parallel with the data collection. The model is validated both theoretically and by real data in the first half of the year and refined throughout the next half of the year based on detailed analyses of the output generated by the model from selected categories of Computer Science. The model was refined so that the results generated would largely be consistent with common perception and other rankings with a few exceptions. The refined model was then submitted to Operations Research and Management Science for validation. It was found in the process that some data were disconnected from the rest, and special techniques were used to address such issues.

The Red Jasper development team designed the web based user-interface almost simultaneously as when the HKUST team completed a relatively stable model. The team consulted typical users including faculty members, research fellows and postgraduate students in the IELM department of HKUST to uncover common interests and use cases on journal ranking. The team then initiated a soft launch when they unified the requirements and built a functional prototype using JSP technology, inviting the same group of users to test our prototype. The common

feedback was that the conventional based web user interface was rather troublesome when composing new journal lists and constructing new category structures. Following their suggestions, we now offer full text search capabilities when selecting the list of journals of interest from the global database, and then use the web 2.0 technology (AJAX in particular) to implement an interactive user interface that enables users to create new journal lists by dragging and dropping. It took 2 months to collect the requirements and to implement the system.

Like most web projects, *journal-ranking.com* has undergone a series of functional testing, load testing and refinements after the implementation phase. The whole process took approximately six months.

Following the refinement, we opened up the system for more user testing (invited testers only). This time, the users' behaviors were closely monitored and we found that although most users performed some customized ranking, more often, they used predefined configurations offered by the system. Initially the system offered three default configurations where one treated all citations with equal weight, one treated self citations within the same journal as of zero weight, and one treated self citations within journals to be half of inter journal citations. From this observation, we tried to optimize the system by pre-computing and storing the ranking results for the default configurations and also extended the default configurations from 3 to 27 (3 adjustable factors, each provides 3 default values). It turned out that the system's throughput was greatly improved and user experience was also improved. We took another 3 months to collect feedback and further refined the system.

The system was officially launched in mid Jan 2007. Invitation emails were initially sent to faculty members in Hong Kong Universities and we since received a tremendous amount of responses.

Maintenance

Presently, *journal-ranking.com* is hosted in the HKUST. However, the hardware and software are being managed by the Red Jasper development team. The team is also responsible for monitoring the forum feedback and improving the system accordingly.

Red Jasper is committed to providing further support in offering journal ranking services to the academia. Meanwhile, Red Jasper is also planning to extend the ranking services to paper level ranking and to offer performance evaluation services based on journal ranking and paper ranking in the future.

Conclusion

This paper briefly summarizes the development of an online interactive journal ranking project and illustrates how AI techniques, although simple, could be adopted innovatively to automate the ranking of journals.

We described the motivations for the project, as well as, the importance of an open, customizable journal ranking service. Journal ranking services allow the academic community to experiment with the ranking of journals using different parameter settings. It creates an alternative platform to the dated ISI index, and enables the exchange of opinions on how to improve the techniques for journal ranking.

Though *journal-ranking.com* is recently launched, the substantial amount of hits it has received thus far has at least proved that journal ranking is of great interest to many academicians/researchers and many of them have their opinions about it. We believe that this is a monumental step toward revising the current standard of measuring the impact of journals.

References

- J. Bollen, M. A. Rodriguez, H. V. de Sompel, 2006. *Journal Status*, *Scientometrics* 69(3).
- ISI, 2006. *Journal citation report by ISI web of knowledge*, <http://portal.isiknowledge.com>.
- L. Page and S. Brin and R. Motwani and T. Winograd, 1999. *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report, Stanford University.
- I. Palacios-Huerta, O. Volij, 2004. *The Measurement of Intellectual Influence*, *Econometrica* 72, 963-977.
- G. Pinski, F. Narin, 1976. *Citation influence for journal aggregates of scientific publications: Theory, with Applications to the Literature of Physics*, *Information Processing and Management* 12, 297-312.
- G. Reese. 1997. *Database Programming with JDBC and Java*, O'Reilly.
- S. Wasserman, K. Faust, 1994, *Social Network Analysis: Methods and Applications*, chapter 5.3 Centrality and Prestige: Directional Relations, 205-124, Cambridge University Press.