ON THE MATHEMATICS OF RANKING UNIVERSITIES AND SCIENTIFIC PRODUCTS

ANDREI MARCUS

Abstract. This is an introductory survey of some mathematical aspects of ranking. We discuss two main topics: aggregating indicators and Arrow's imposibility theorem, and Google's PageRank and some of its derivatives based on citation analysis.

1. Introduction

Quantitative metrics are rather superficial choices for assessing the research output of an individual scholar. Methods like counting papers and citations, calculating impact factors and h-indices, or looking at Eigenfactor Scores (described below) are not adequate compared with what should be the gold standard: reading the scholars publications and talking to experts about his or her work. Such opinions have been many times formulated, see Ewing, Adler and Taylor [17] and also the references given there, Goldreich [22], and even by proponents of automated methods, such as Bini, Del Corso, and Romani [7]. But many scholars, librarians, historians of science, editors, and other individuals are also interested in larger scale questions that require assessing hundreds or thousands of articles by hundreds or thousands of authors. As the number of scientific journals and papers is increasing at an almost exponential rate, and if a library can afford only a limited number of subscriptions, which journals should the librarian choose? The burden of similar decisions affects researchers, funding agencies, university administrators, reviewers. It can be very difficult and costly to give an in-depth evaluation of the research, therefore aggregate bibliometric statistics, regarded as indirect indicators of quality, can be useful.

We start with an example which should be much easier that ranking universities, that is, ranking athletes in combined events.

The **decathlon** is an olympic event with an arbitrary scoring system, and thus personal performance and records can be broken as new scoring tables are adopted.

Under the original 1912 scoring tables, *Akilles Järvinen* of Finland finished second in both the 1928 and 1932 Olympics, but the new scoring system introduced in 1934 gave Jarvinen higher converted totals than both the men he lost to. The table were updated again in 1050 and in 1962. In 1984, at the Los Angeles Olympic Games, Great Britain's *Daley Thompson* missed the world record by one point on the 1962 tables. The tables were changed later in 1984 and Thompsons score in Los Angeles converted to

Date: August 13, 2009.
Key words and phrases. Arrow's imposibility theorem, citation analysis, Impact Factor, Perron-Frobenius theorem, Google PageRank, Eigenfactor, university rakings.

a best-ever mark. The current 1984 table shows the following benchmark levels needed to earn 1000, 900, 800, and 700 points in each sport.

Event	1000 pts	900 pts	800 pts	700 pts	Units
100m	10.395	10.827	11.278	11.756	Seconds
Long Jump	7.76	7.36	6.94.1	6.51	Meters
Shot Put	18.4	16.79	15.16	13.53	Meters
High Jump	2.20	2.10	1.99	1.88	Meters
400m	46.17	48.19	50.32	52.58	Seconds
110m Hurdles	13.8	14.59	15.419	16.29	Seconds
Discus Throw	56.17	51.4	46.59	41.72	Meters
Pole Vault	5.28	4.96	4.63	4.29	Meters
Javelin Throw	77.19	70.67	64.09	57.45	Meters
1500m	233.79	247.42	261.77	276.96	Seconds

Even more puzzling is the **modern pentathlon**, which includes five events: 10 metre air pistol shooting, épée fencing, 200 m freestyle swimming, show jumping over a 350 to 450 m course with 12 to 15 obstacles, and a 3 km cross-country run.

Why is it so difficult to aggregate a number of indicators and to establish "the best" or "the most complete" athlete in the world? Choosing a winner is a problem in the *social choice theory.* It is usually formulated in terms of voting systems.

2. Arrow's impossibility theorem

The need to aggregate preferences occurs in many different disciplines: in welfare economics, where one attempts to find an economic outcome which would be acceptable and stable; in decision theory, where a person has to make a rational choice based on several criteria; and most naturally in voting systems, which are mechanisms for extracting a decision from a multitude of voters' preferences.

Arrows impossibility theorem, or Arrows paradox, tells that no voting system can convert the ranked preferences of individuals into a community-wide ranking while also meeting a certain set of reasonable criteria with three or more discrete options to choose from. These criteria are called *unrestricted domain*, *non-imposition*, *non-dictatorship*, *Pareto efficiency*, and *independence of irrelevant alternatives.*

The theorem is named after economist Kenneth Arrow (a co-recipient of the 1972 Nobel Prize in Economics), who proved the theorem in his Ph.D. thesis and in the original paper titled *A Difficulty in the Concept of Social Welfare* (see Arrow [2]). Arrow's theorem is related to the *Condorcet voting paradox*, formulated in the times of the French Revolution. There are also variants of this theorem, such as the Gibbard-Satterthwaite theorem. For detailed discussions we refer to Geanakoplos [21], Reny [30], Taylor [35], and Taylor and Pacelli [36].

Let me state the theorem in a more formal manner. Assume that we need to extract a preference order on a given finite set $A = \{a, b, c, \dots\}$ of at least three options (alternatives, outcomes). Each individual in the society (or equivalently, each decision criterion) consisting of at least two memebers gives a particular order of preferences on the set of options. We are searching for a preferential voting system, called a *social welfare function*, which transforms the set of preferences into a single global societal preference order. The theorem makes the following assumptions, which are considered to be reasonable requirements of a *fair voting method*:

Non-dictatorship. The social welfare function should account for the wishes of multiple voters. It cannot just mimic the preferences of a single voter. Formally speaking, the function is a *dictatorship* by individual n, if for every pair a and b, society strictly prefers a whenever n strictly prefers a to b.

Unrestricted domain (or universality). The social welfare function should account for all preferences among all voters to yield a unique and complete ranking of societal choices. Thus: the voting mechanism must account for all individual preferences; it must do so in a manner that results in a complete ranking of preferences for society; it must deterministically provide the same ranking each time voters' preferences are presented the same way.

Independence of irrelevant alternatives. The social welfare function should provide the same ranking of preferences among a subset $\{a, b\}$ of options as it would for a complete set of options. Changes in individuals' rankings of irrelevant alternatives (ones outside the subset $\{a, b\}$) should have no impact on the societal ranking of the subset $\{a, b\}$.

Positive association of social and individual values (monotonicity). If any individual modifies his or her preference order by promoting a certain option, then the societal preference order should respond only by promoting that same option or not changing, never by placing it lower than before. An individual should not be able to hurt an option by ranking it higher.

Non-imposition (or citizen sovereignty). Every possible societal preference order should be achievable by some set of individual preference orders. In mathematical terms, the social welfare function is surjective.

Arrow's theorem says that:

if the decision-making body has at least two members and at least three options to decide among, then it is impossible to design a social welfare function that satisfies all these conditions at once.

A later (1963) version of Arrow's theorem was be obtained by replacing the monotonicity and non-imposition criteria with

Pareto efficiency or unanimity. Society puts alternative a above b whenever every individual puts a above b.

This version is stronger (it has weaker conditions) - since monotonicity, non-imposition, and independence of irrelevant alternatives together imply Pareto efficiency, whereas Pareto efficiency, non-imposition, and independence of irrelevant alternatives together do not imply monotonicity.

Arrow's theorem shows that it is impossible to devise a ranking method that is fully satisfactory. The study on theoretical and empirical grounds of aggregation procedures is one of the tasks of the so called Multi-Criteria Decision Making (MCDM) (see [3], [16]).

3. Ranking universities

How does Arrow's theorem apply to the problems which torment the academia?

"The Centre for Science and Technology Studies, Leiden University, has developed a new ranking system entirely based on its own bibliometric indicators. ... on the basis of the same data and the same technical and methodological starting points, different types of impact-indicators can be constructed, for instance one focusing entirely on impact, and another in which also scale (size of the institution) is taken in to account. Rankings based on these different indicators are not the same, although they originate from exactly the same data. Moreover, rankings are strongly influenced by the size-threshold used to define the set of universities for which the ranking is calculated."

This is what can be read on the website [45] devoted to the Leiden Ranking 2008. I confess that I've read these lines with amusement, because this is exactly what one can expect in the light of Arrow's theorem.

By analyzing empirical data, Gaufriau and Larsen [20] show that the rankings of countries research output based on number of publications or citations heavily depend on the counting method, and that rankings based on different counting methods cannot be compared. They compared the *Whole Counting* method, when full credit for a publication is given to a country when at least one of the authors is from that country, and *Fractional Counting*, when a country receives a fraction of full credit for a publication equal to the fraction of authors to the publication coming from that country. Gaufriau and Larsen also recommend that all rankings of countries and institutions in the should be based on Fractional Counting.

A study done by Dehon, McCathie and Verardi [11] with the aid of the so-called *principal component analysis* (a mathematical technique that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components) of the Shanghai ARWU ranking reveals that two different and uncorrelated aspects of academic research are measured and aggregated: *overall research output* and *top researchers*. This implies that the relative weight given by the rankers to these two factors determines to a large extent the final ranking.

The same Shanghai ARWU ranking is analyzed by Billaut, Bouyssou and Vincke [6] from the perspective of Multiple Criteria Decision Making. Their conclusions are quite critical: criteria used by authors of ARWU are only loosely connected with what they intended to capture; the ranking involves several arbitrary parameters and many micro-decisions that are not documented (the raw data is not publicly available, hence it cannot be checked); the aggregation technique used is flawed; in general, weighted sum would is a poor way to aggregate criteria; the aggregation technique that is used is nonsensical.

These authors, and several others, such as van Raan [37], [38], and Florian [18] also point out to additional serious methodological problems of the Shanghai ranking, but we refrain here from discussing this.

Perhaps the dependence on the methodology is not such a huge problem if the rankings are made by or for tabloid newspapers. But the amusement disappears when governments and funding agencies jump into the game. Let me make it clear that no one, not even Arrow's theorem, says that every ranking or voting system is totally flawed. For instance, it is clear for the naked eye that the American Ivy League universities are better than the Romanian universities by any sensible measure. One doesn't need savant methodologies and painstaking (and error-prone) data gathering to conclude that. But if one believes the data published in recent years, the three oldest Romanian universities, Bucharest, Iaşi and Cluj, have quite similar performance. I'm pretty certain that for any prescribed order of them, a reasonable ranking method which produces that order can be devised. This means that if the government decides to give additional funding to only one of them, then somebody powerful enough ultimately has to impose its will. The problem is still there even if one ranks departments or programs instead universities. One mathematics department can be better in Algebra, the other one in Geometry, and we can go even deeper.

Recent analyses and opinions on various teaching&learning quality and research assessments, as well as university rankings can be found in Evans [15], Ewing, Adler and Taylor [17], Hicks [24], Ortega and Aguillo [28], and van Vught [39].

4. Citation analysis: ranking academics, journals, and scholarly papers

Most of automatic methods rely on citation analysis. Peer review is not always practicable, so some questions, such as bibliometric analysis of the relative influence of the full contents of a journal can only be answered by a large-scale quantitative approach. For these questions, citation data can be useful, and one should make the best possible use of it. As librarians work to meet increasing subscription prices with increasingly constrained subscription budgets, powerful measures of journal influence and journal value may usefully supplement expert opinion and other sources of information in making difficult decisions about journal holdings. There are, of course, plausible assumptions underlying the use of citation analysis as a heuristic.

There exist several databases which offer citation statistics, such as ISI's JCR, Scopus, American Mathematical Society's Mathematical Reviews, Citeseer, etc. In this section, after briefly discussing the JCR based journal Impact Factor, we present some more recent ranking methods based on Google's PageRank algorithm. This new approach is used by EigenFactor [42], SCImago [46], RedJasper [46], and it is based on the idea that not all the citations are equal.

4.1. **Impact factor.** A journal's impact factor is a measure invented by Eugene Garfield [19] and counts the number of times that articles published in a census period cite articles published during an earlier target window. The impact factor as calculated by *Thomson Scientific* has a one year census period and uses the two previous years for the target window. Stated more formally, let n_t^i be the number of times in year t that the year $t-1$ and $t-2$ volumes of journal i are cited. Let a_t^i be the number of articles that appear in journal i in year t. The impact factor IF_t^i of journal i in year t is

$$\mathrm{IF}_t^i = \frac{n_t^i}{a_{t-1}^i + a_{t-2}^i}.$$

Many authors consider that the time window of two years is too small to be relevant for many disciplines, especially in the case of social sciences, but also mathematics.

Counting citations gives the possibility to define various measures, such as the h-index, the m-index, the g-index, the $g1$-index (see Hirsch [25], Egghe [14], Costas and Bordons [10], and Ewing, Adler and Taylor [17] for definitions and discussions. The study from a statistical point of view of these indicators and their countless variations (see Deineko and Woegingerb [12]) is one of the main preoccupations of specialized journals like *Scientometrics* and *Journal of Informetrics.*

For instance, Althouse et al. [1] study how impact factors vary across fields and over time. By decomposing average impact factors into four contributing components – field growth, average number of cited items per paper, fraction of citations to papers published within two years, and fraction of citations to JCR-listed items – the sources of this variation are determined. It turns out that an increasing number of citations in the reference lists of published papers is the greatest contributor to impact factor inflation over time. Differences in the fraction of citations to JCR-indexed literature is the greatest contributor to differences across fields, though cross-field differences in impact factor are also influenced by differences in the number of citations per paper and differences in the fraction of references that were published within two years. By contrast, the growth rate of the scientific literature and cross-field differences in net size and growth rate have very little influence on impact factor inflation or on cross-field differences in impact factor. Vanclay [40] argues that the journal IF does not deal evenly with journals to which citations accrue quickly over a confined period and journals to which citations accrue slowly over an extended period, because of the 2-year target window.

4.2. **Google's PageRank and the Eigenfactor algorithm.** The papers Bergstrom [4], and Bergstrom, West and Wiseman [5] describe a new metric for the assessment of journal quality that is based on the use of the PageRank (see Brin and Page [9], Langville and Meyer [27]), the algorithm used by the Google search engine. PageRank uses the Perron-Frobenius theorem (it asserts that a real square matrix with positive entries has a unique largest real eigenvalue and that the corresponding eigenvector has strictly positive components.) to provide the relative ranking of various web sites. This theorem is now used to rank football teams, generate schedules for sports teams and to establish tennis rankings. A similar use of PageRank algorithm is to rank academic doctoral programs based on their records of placing their graduates in faculty positions Schmidt and Chingos [33]. In PageRank terms, academic departments link to each other by hiring their faculty from each other (and from themselves).

As Bollen et al. [8] point out, "*Google's PageRank algorithm computes the status of a web page based on a combination of the number of hyperlinks* *that point to the page and the status of the pages that the hyperlinks originate from. By taking into account both the popularity and the prestige factors of status, Google has been able to avoid assigning high ranks to popular but otherwise irrelevant web pages.*" Note that PageRank was inspired by old papers on citation analysis, such as de Solla Price [13].

Whereas the IF is calculated simply on the basis of static citation rates and publication numbers, the Eigenfactor approach is to rank journals much as Google ranks web pages. While Google uses the network of hyperlinks on the Web, we use citations in the academic literature as provided by ISI's Journal Citation Reports (JCR). The aim is to identity the most "influential" journals, where a journal is considered to be influential if it is cited often by other influential journals. This seems circular, but in fact can iteratively calculate the importance of each journal in the citation network by a simple mathematical algorithm.

This iterative ranking method tries to account for the fact that one citation from a high-quality journal may be more valuable than multiple citations from less important publications. The importance of a citation is measured by the influence of the citing journal divided by the total number of citations appearing in that journal. This intends to remedy the differences across disciplines and journals in the propensity to cite other papers. For example, a citation from a review article that has references to large numbers of papers counts for less than a citation from a research article that cites only papers that are relevant for the argumentation.

We describe the computation the 2006 Eigenfactor scores. From the Thompson Scientific JCR dataset, one extracts a citation matrix, $C = (c_{i,j})$ for the 7611 ISI-linked science and social science journals, where $c_{i,j}$ is the number of citations from 2006 articles in journal j to articles in journal i published in 2001-2005. We fill in with zeros the diagonal of C to ignore journal self-citations. We construct a normalized version of C, named $H = (h_{i,j})$, normalized by the column sums, i.e.: the number of outgoing citations from each journal:

$$h_{i,j} = \frac{c_{i,j}}{\sum_k c_{k,j}}.$$

We also compute an article vector a, where a_i is the number of articles published by journal i over the five-year target window, divided by the total number of articles published by all source journals over the same five-year window. Some of the journals listed in the matrix H will be dangling nodes journals that do not cite any other journals. Any column of the matrix H that has all 0 entries is a dangling node; we replace all such columns in H with the vector a to produce a new modified matrix H'. This is a stochastic matrix by construction. H' corresponds to a *random walk* on the scientific literature. From this, we construct a new stochastic matrix, P:

$$P = \alpha H' + (1-\alpha)a\mathbf{e}^T,$$

where $\mathbf{e}^T$ is a row vector of all 1's (and ^{T} means transposition), and thus $a\mathbf{e}^T$ is a matrix with identical columns a. This corresponds to a process which follows the literature with probabilities $1-\alpha$ and "teleports" to a random journal with weights proportional to the number of articles published by a journal. As in the case of Google's PageRank, the value $\alpha = 0.85$ is used. Define the vector π^* as the leading eigenvector of P, which corresponds to the fraction of time spent at each journal in P. These fractions serve as weights of journal influence. The *Eigenfactor* score, EF, is defined as

$$\mathrm{EF} = 100 \frac{H\pi^*}{\sum_i [H\pi^*]_i}.$$

The *Article Influence* score AI_i for each journal i is a measure of the per-article citation influence of the journal. The *Article Influence* score is calculated as

$$\mathrm{AI}_i = 0.01 \frac{\mathrm{EF}_i}{a_i},$$

where EF_i is the *Eigenfactor* score for journal i, and a_i is the i-th entry of the normalized article vector. This measure is more directly comparable to ISIs Impact Factor.

As pointed out in Ewing, Adler and Taylor [17], a weakness of *Eigenfactor* is that all citations from articles in a given journal j to articles published in j during the preceding five years are discarded as "self-citations". But these are not "self-citations" in any normal sense of the word, and the data from the Mathematical Reviews Citations database [48] suggests that about one third of all citations are discarded in this way.

4.3. **Integrated models for ranking.** Bini, Del Corso, and Romani proposed in [7] an integrated ranking of authors, journals, papers, areas, and institutions, with the intent to design a tunable method to capture the different needs. They note that previously, the ranking of journals was usually based on citations, while the ranking of papers and authors follows from the rank of the journals where the research is published. The basis of their method is again the PageRank algorithm [9], and they also build on the work of other authors such as Bollen [8], Bergstrom [4], Palacios-Huerta and Volij [29].

The general principle of this approach is the "mutual reinforcement between papers, journals, authors":

- A paper is important if published in an important journal but also if cited by important papers and authored by important authors.
- An author is important if she has important co-authors and has written important papers published in important journals.
- A journal is important if collects citations from important journals, publishes important papers by important authors.

The authors described and analyze three models: the *One-class model*, made up by Papers only, the *Two-class model*, made up by Papers and Authors, and the *Three-class model*, made up by Papers, Authors and Journals.

Let us briefly describe the One-class model. One can represent the citation process as a graph and hence as a binary matrix. A main assumption is that receiving a citation is always good!

The notation is as follows: n is the number of papers; $C = (c_{i,j})$ is the $n \times n$ citation matrix, that is,

$$c_{i,j} = \begin{cases} 1 & \text{if paper } i \text{ cites paper } j, \\ 0 & \text{otherwise;} \end{cases}$$

let $\mathbf{e} := (1, 1, \ldots, 1)^T$, $\mathbf{d} := C\mathbf{e}$, hence $d_i = \sum_{j=1}^{n} c_{i,j} e_j$, and assume that $d_i \neq 0$, for $i = 1, \ldots, n$. Finally, let

$$P = (p_{i,j}) := \mathrm{diag}(\mathbf{d})^{-1} C.$$

Then P is *row stochastic*, that is, $P\mathbf{e} = \mathbf{e}$, or equivalently, $\sum_j p_{i,j} = 1$ for all i. If P is irreducible (it is not similar to a block upper triangular matrix via a permutation), by the *Perron-Frobenius theorem* it exists unique a vector $\pi = (\pi_i) > 0$ such that $\sum_i \pi_i = 1$ and $\pi^T = \pi^T P$, that is

$$\pi_j = \sum_{i=1}^{n} \pi_i \frac{c_{i,j}}{d_i},$$

(which means that each paper equally distributes its importance among all the cited papers).

The vector π is called the *Perron vector* of P, and by definition, π_i is the *rank of paper i*.

Some problems may occur in this model. In general P, as defined above, is not irreducible. Moreover, there may exist dangling nodes, i.e., papers which cite no papers, so that C may have null rows and P cannot be constructed. Even when P is irreducible it may be periodic. To remedy these problems, a so-called *dummy paper* is introduced, which cites and is cited by all the existing papers except by itself. The dummy paper collects the importances of all the papers and redistributes them uniformly to all the subjects by creating no privileges.

The authors say that their method is flexible enough to meet multiple goals, but they also acknowledge that one should use automatic ranking only when strictly necessary, that there are parameters that have to be agreed upon at a "political" level, and that peer review cannot be replaced.

4.4. **Discussion.** Automated rankings and evaluation of research based on citation analysis have some advantages: they are easy to calculate, time aware, objective, etc. (although "objectivity is disputed). Among the weaknesses are the following:

- a citation is not always a trusting vote;
- data source and coverage are subject of dispute;
- citation gathering can be a very slow process;
- it depends on the area of research the way how authors choose the papers to cite;
- in the same journal there are articles with different citation rates;
- the ranking doesn't always agree with the widely accepted journal's reputation.

The use of PageRank represents an improvement of the ISI's IF when one is interested less in the general popularity of a journal and more in its "value". But the complexity of PageRank algorithms can be a pitfall because the final results are harder to understand. In return, proponents of PageRank methods often quote Albert Einstein: "*Everything in life should be as simple as possible, but no simpler*". A compromise proposal would be the use of the so called Y-factor of a journal

$$Y(j) = \mathrm{IF}(j) \times \mathrm{EF}(j),$$

which rates journals by using an equal measure of both IF and PageRank.

Despite Google's commercial success, nobody claims that PageRank is the best possible algorithm for ranking web pages. One can only say that most users wild find relevant results for most of their searches at the top of the list. Other search engines use different algorithms which are not bad at all, so Google's domination of the search market has many reasons. It is also worth mentioning that the details of the ranking methodology are kept secret, to avoid manipulation. A company named "SearchKing" which used to make money by creating link farms suddenly lost its high ranking in 2002, and sued Google in trying to obtain a disclosure of the algorithm (see Langville and Meyer [27]). However, the courts upheld Google's right to keep its software confidential. This aspect should be compared to one of the university ranking principles formulated at the 2006 IREG meeting:

Berlin principle no. 6. *Thou shalt be transparent regarding the methodology used for creating the rankings.*

(Contrary to this requirement, experiments like Thomson JIF and Shanghai ARWU ranking appear to be irreproducible, and much of their raw data is not publicly available, see Rossner, Van Epps and Hill [31], Florian [18]. Yet they are often used to decide academic employment and funding.)

It is important to note that the currently available methods for the evaluation of the quality of scientific papers and the status of the journals that publish these papers are undergoing a period of re-evaluation.

5. Concluding remarks

By human nature, "rankings are here to stay", as it is often said, despite the fact that the very notion of "the best university" is illusory. Rankings are important marketing tools, but can also have various unintended and perverse consequences.

Arrow's theorem tells us that there is no such thing as "the best ranking method", that achieves everything we dream of. No metric of scholarly impact represents a perfect solution, and different mathematical models may be useful in different circumstances.

Ranking universities is not a "scientific" exercise, but a journalistic one, which is fine. It can be motivated by the right of public to be informed. Of course, only the raw data could be published, letting the interested members of public to analyze and decide, but then probably the newspapers wouldn't sell too well.

Automated rankings can be manipulated. This is why Google's true methodology for ranking webpages is a closely guarded secret.

Various ranking methodologies, even if carefully devised, are at most statistically relevant. For instance, if a university library subscribes to journals by selecting them acording to a PageRank type algorithm, it is very likely that most of the faculty will be satisfied.

Comparing two individuals for a serious purpose is a different matter. When evaluating individual papers or researchers, or even individual journals or universities, there is no substitute for reading and understanding the work.

Official rankings made for the purpose of differentiated funding are the result of political decisions. Their intricate methodologies are mere justifications, but this is where the real fun of ranking begins!

References

[1] B.M. Althouse, J.D. West, T. Bergstrom and C.T. Bergstrom. *Differences in Impact Factor Across Fields and Over Time.* Journal of the American Society for Information Science and Technology, 60 (2009) 27–34.

[2] K.J. Arrow. *A Difficulty in the Concept of Social Welfare.* The Journal of Political Economy, **58** (1950), 328–346.

[3] D. Bouyssou, T. Marchant, M. Pirlot, A Tsoukiàs and Ph. Vincke. *Evaluation and Decision Models with Multiple Criteria. Stepping stones for the analyst.* Springer, New York 2006.

[4] C.T. Bergstrom. *Eigenfactor: Measuring the value of prestige of scholarly journals.* C&RL News, **68** (2007), 314–316.

[5] C.T. Bergstrom, J.D. West and M.A. Wiseman. *The Eigenfactor metrics*, Journal of Neuroscience 28(45) (2008), 11433–11434.

[6] J.-C. Billaut, D. Bouyssou and Ph. Vincke. *Should you believe in the Shanghai ranking? An MCDM view.* preprint, hal-00388319, version 2 – 15 July 2009.

[7] D.A. Bini, G.M. Del Corso and F. Romani. *Evaluating scientific products by means of citation-based models: a first analysis and validation.* Electronic Transactions on Numerical Analysis **33** (2008), 1–16.

[8] J. Bollen, M.A. Rodriguez and H. Van de Sompel. *Journal status.* Scientometrics **69** (2006) (arxiv.org:cs.DL/ 0601030, DOI: 10.1007/s11192-006-0176-z)

[9] S. Brin and L. Page. *The Anatomy of a Large-Scale Hypertextual Web Search Engine.* Comput. Networks ISDN Systems, **30** (1998), 107–117.

[10] R. Costas and M. Bordons. *The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level.* J. Informetrics **1** (2007), 193–203.

[11] C. Dehon, McCathie and V. Verardi. *Uncovering excellence in academic rankings: a closer look at the Shanghai ranking.* Scientometrics (2009), in press.

[12] V.G. Deineko and G.J. Woegingerb. *A new family of scientific impact measures: The generalized Kosmulski-indices.* Scientometrics, 80 (2009), 821–828.

[13] D.J. de Solla Price. *Networks of scientific papers.* Science **169** (1965), 510–515.

[14] L. Egghe. *Theory and practice of the g-index.* Scientometrics **69** (2006), 131–152.

[15] M. Evans. *Killing Thinking: The Death of the Universities.* Continuum, London-New York, 2004.

[16] J. Figueira, S. Greco and M. Ehrgott eds. *Multiple Criteria Decision Analysis: State of the Art Surveys.* Springer, Boston 2005.

[17] J. Ewing, R. Adler and P. Taylor. Joint IMU/ICIAM/IMS-Committee on Quantitative Assessment of Research. *A report on Citation Statistics*, 2008. http://www.mathunion.org/publications/report/citationstatistics/.

[18] R. Florian. *Irreproducibility of the results of the Shanghai academic ranking of world universities.* Scientometrics, **72** (2007), 25–32.

[19] E. Garfield. *Citation Analysis as a Tool in Journal Evaluation.* Science **178** (1972), 471–479.

[20] M. Gauffriau and P.O. Larsen. *Counting methods are decisive for rankings based on publication and citation studies.* Scientometrics, **64** (2005), 85–93.

[21] J. Geanakoplos. *Three brief proofs of Arrow's impossibility theorem.* preprint, Yale University, 2001.

[22] O. Goldreich *On the evaluation of scientific work.* http://www.wisdom.weizmann.ac.il/ oded/on-eval.html.

[23] J.M. Henshaw. *Does measurement measure up? How numbers reveal and conceal the truth.* The Johns Hopkins University Press, 2006.

[24] D. Hicks. *Evolving regimes of multi-university research evaluation.* High Educ **57** (2009), 393–404.

[25] J.E. Hirsch. *An index to quantify an individuals scientific research output.* Proc. Natl. Acad. Sci. U.S.A. **102** (2006), 16569–16572.

[26] J.P. Keener. *The Mathematics of Ranking Schemes or Should Utah Be Ranked in the Top 25*? preprint, University of Utah, 2004.

[27] A.N. Langville and C.D. Meyer. *Google's PageRank and Beyond. The Science of Search Engine Rankings.* Princeton University Press, 2006.

[28] J.L. Ortega, I.F. Aguillo. *Mapping world-class universities on the web.* Information Processing and Management **45** (2009), 272–279.

[29] I. Palacios-Huerta and O. Volij. *The measurement of intellectual influence.* Econometrica **72** (2004). 963–977.

[30] P.J. Reny. *Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach.* Preprint, University of Chicago, 2000.

[31] M. Rossner, H. Van Epps, and E. Hill. *Show me the data.* The Journal of Cell Biology, **179** (2007), 1091–1092. http://jcb.rupress.org/cgi/content/full/179/6/1091.

[32] M. Rosvall and C.T. Bergstrom. *Maps of random walks on complex networks reveal community structure.* Proc. Natl. Acad. Sci. USA **105** (2008), 1118–1123.

[33] B.M. Schmidt and M.M. Chingos. *Ranking Doctoral Programs by Placement: A New Method.* PS: Political Science and Politics 40 (2007), 523–529.

[34] M.J. Stringer, M. SalesPardo, and L.A. Nunes Amaral. Effectiveness of journal ranking schemes as a tool for locating information. PLoS ONE **3** (2008): e1683 http://dx.doi.org/10.1371/journal.pone.0001683.

[35] A.D. Taylor. *Social Choice and the Mathematics of Manipulation.* Cambridge University Press, 2005.

[36] A.D. Taylor and A.M. Pacelli. *Mathematics and Politics. Strategy, Voting, Power and Proof.* Springer, New York 2008.

[37] A.F.J. van Raan. *Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods.* Scientometrics, **62** (2005), 133–143.

[38] A.F.J. van Raan. *Reply to the comments of Liu et al.* Scientometrics, **64** (2005), 111–112.

[39] F.A. van Vught ed. *Mapping the Higher Education Landscape. Towards a European Classification of Higher Education.* Springer, 2009.

[40] J.K. Vanclay. *Bias in the journal impact factor.* Scientometrics, 78 (2009), 3–12.

[41] J. West and C.T. Bergstrom. *Pseudocode for calculating Eigenfactor Score and Article Influence Score using data from Thomson-Reuters Journal Citations Reports.* preprint, University of Washington, 2008.

[42] Eigenfactor. *Ranking and mapping scientific knowledge.* http://www.eigenfactor.org.

[43] IREG-2. *Berlin Principles on Ranking of Higher Education Institutions.* Berlin May 20th 2006. http://www.ihep.org.

[44] Belgian Mathematical Society and National Committee of Mathematics. Symposium: The mathematics of ranking. October 15, 2008. http://www.cs.kuleuven.be/˜nalag/research/workshops/ranking/.

[45] The Leiden Ranking 2008. http://www.cwts.nl/ranking/LeidenRankingWebSite.html.

[46] SCImago Journal & Country Rank. http://www.scimagojr.com.

[47] Journal-Ranking.com by RedJasper. http://www.journal-ranking.com/ranking/web/index.html.

[48] Mathematical Reviews Publications database. http://www.ams.org/mathscinet/.

[49] http://www.wikipedia.org.

"Babeş-Bolyai" University, Faculty of Mathematics and Computer Science, Str. Mihail Kogălniceanu nr. 1, RO-400084 Cluj-Napoca, Romania