Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 971–976

Assessment of the EPQ probability parameter for scientific articles publishing

Rafał Rumin
AGH University of Science and Technology,
Al. Mickiewicza 30, 30-059 Kraków, Poland
E-mail: rumin@agh.edu.pl

Piotr Potiopa
AGH University of Science and Technology,
Al. Mickiewicza 30, 30-059 Kraków, Poland
E-mail: ppotiopa@zarz.agh.edu.pl

***Abstract*—This work presents the analysis of evaluation concerning the articles that are send to publication in academic journals, basing on additional parameters not resulting from essential value of the research work. Currently, majority of article verification algorithms is oriented on the selection of such works that are potentially more strongly influencing the international position of journal. For that purpose, editorial offices, and also reviewers, apply multi-criterion parametric evaluations and accepted parameters have often very subjective character. Presented work makes an attempt to identify used criterion functions i.e. defining evaluation parameters. These parameters were divided onto categories and there was proposed their preliminary verification basing on statistical analysis of already published articles in individual journals. Each parameter has attributed weight function, which allows to defined its impact on the total evaluation of article, and also adaptation of formula to any academic journal. Weight functions will be determined with usage of neural networks or genetic algorithms, aiming to their individual adaptation to particular journal.**

I. Introduction

PREVIOUS investigations over the evaluation of academic journals, cause the continuous improvement of algorithms by which the articles published in these journals are verified. This results from endless aspiration of journals to obtain maximum of points in created rankings (Philadelphia List, Impact Factor, quoting indicators etc.) [2-4].

New appearing methods of the evaluation of journals and modifications of already existing, cause that the essential evaluation of the article can be unsettled in the interest of the parametric evaluation forced by publisher[8-16].

In effect, innovative publications can be inadequately evaluated or not published due to their wrong preparation. Introduced and described below coefficients of scientific articles parametrization are supposed for the task to determine an influence of these subjective parameters on the evaluation of articles in individual journals. Furthermore, there were presented series of factors which, if they will be taken into consideration during writing of scientific articles , have a chance to increase probability of obtaining positive review and in effect the acceptance of publication in renowned journals. In the further process of research works, there is planned realization of automatic information system, which role will be verification of the working version of article, before sending it to the journal and the definition of the probability of obtaining high parametric evaluation. Described parametric evaluation will determine the coefficient EPQ – Estimated Paper Quality. This coefficient will be helpful for scientists who concentrate mainly over essential, and less over the editorial part of their scientific article. The low value of EPQ should induce the author to analyze and supplement his publication before sending article to editorial office of chosen earlier journals.

II. Evaluation of journals

Academic journals are subjected to continuous verification through evaluation of published articles influence on environment of scientists. There exist many parameters evaluating the parametric quality of the journal, here are some of them [20-26]:

- Impact Factor (IF)
- Relative Citation Rates (RCR) / Journal to Field Impact Score (JFIS)
- Article Influence (AI)
- SCImago Journal Rank (SJR)
- Source-Normalized Impact per Paper (SNIP).

Each of them characterizes different factors which influence final evaluation of journals. Different journal evaluation criteria cause the inhomogeneity in resultant rankings. Furthermore, algorithms of evaluation are subjected to continuous changes aiming to the most reliable definition of publications quality. From this reason, the aim of publishing companies, instead of valuing scientific publications having less 'popular' character (though substantially equally good whether even much better), can be wish of achievement of as highest parametric coefficients evaluating other of their publications.

The most popular is Impact Factor (IF) (1). It counts all quoting from particular calendar year, and it divides them by amount of "cited" publications from last two years (C)

978-1-4673-4471-5/$25.00

$$IF = \frac{B}{C} \quad (1)$$

Other indicators, although they also reflect the parametric quality rating of journal, are not so popular.

It can be accepted that, as higher evaluation of given journal in the ranking, the article published in it, has a chance to obtain greater range, and consequently receiving greater quantity of quotations. It seems that there exists the conformity of business among the journal and author of the article, however this concerns only wishes of obtaining as maximum quotations quantity at other publishers through the large number of scientists.

Wanting to check our chances for the publication in given journal, often we set incorrect question - *will this journal publish my article?*

To show existing dependences and conflicts of interests between author and editor, one ought to set himself the question:

How my article will help the journal to obtain better position in the ranking (more points in the parametric evaluation of journals)?

The answer is dependent on many factors (the general diagram of dependence among publication publishers and authors is presented on Fig. 1), which can subjectively influence the evaluation of article, aside from its essential value.

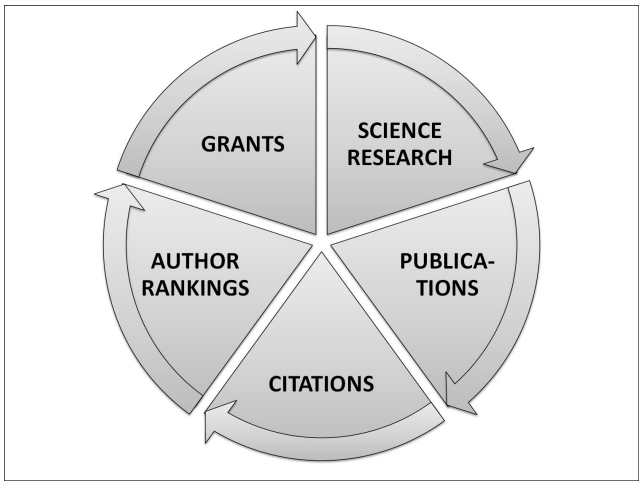

Fig 1. The influence of research work elements on financing and publishing of research

A. The evaluation of Authors

Authors are also subjected to parametric evaluation, targeting the verification of their achievements through the creation of ranking reflecting their contribution to the development of given field. One of main parameters applied in relation to authors of publication is proposed in year 2005 the Hirsch index (h-index) [1]. As easily can be envisaged, such evaluation can be sensitive on manipulation on the side of several cooperating with themselves authors, who mutually will quote their works (aside from their essential contribution into researches). The parametric evaluation of publication issues from the category of scientific research. Scientific researches require financings, and one of the popular sources of learning financing are exploratory grants. To obtain the financing there is expected that the scientist will carry out planned investigations and their effect will have visible influence on given exploratory field. How is measured such influence?

Most readable measure are publications and their quotations. From this reason, scientists who have a suitably high Hirsch index, are treated as trustworthy to commit to them public money on carried researches. Readable dependencies appear between the financing of research, with quantity of publication, and with their quotations which put themselves greater chance for future financial resources.

III. Parametric evaluation of the Article

Every article, except the essential value, can be described by a group of parameters defining its quality from the interest of journal point of view. Here appears mentioned conflict of interests between publishing houses, and authors [17].

In the evaluation of article, the essential value can be estimated by additional parameters defining the range of carried researches, e.g. the article containing theoretical models can classified lower than articles containing, except the theory, also simulation models. As the best will be evaluated articles containing the experimental verification carried researches. Separately, enough high classification can have articles containing rich and complex reviews of the literature from the given field, because this type of articles are quoted often many times. This results from the specific approach of scientists to carried researches and wishes of using elaborated earlier literature review - which often requires a lot of time, and belongs to „little attractive" researches.

Thereby, that at the evaluation of articles value nobody can foresee how often he will be quoted in the future. In the simplification it can be assumed, that in the initial phase of article analysis, each has an evaluation for the essential value on the same level. Since the quantity of elaborated article future quotations cannot be influenced, it can be influenced whom the author quotes in his own publication. This way the quantity of "gained' quotations from the journal's point of view, can be controlled . The issue here is the period of time in which journals are subjected to evaluation in rankings. For the calculation of Impact Factor, there are taken into consideration last 2 years, what means that the auto quotation of other articles which appeared in the same publishing house within a period of last 2 years, have a positive influence on IF indicator increase. Therefore, the publishing house will be willingly promoting art-

icles which already show quotations from their own journal, what is a method to obtaining higher place in the ranking. However, if there exists a group of journals given by the common institution, then cross quotations of other journals belonging to the same publisher are also added value. Here arises a threat regarding the reliability of published articles, because one can apply the mechanism which would permit ranking speculations between journals. Following the paragraphs of this article contain the case study describing such situations.

IV. EPQ - the coefficient of the parametric evaluation of article

To show series of factors participating in the evaluation of given article, there was proposed the EPQ coefficient (Estimated Paper Quality). It can be presented as weighted mean of individual parameters, with suitably assorted weights functions. The value of parameters is standardized so that it contains itself in range from 0 to 1. This type of method descends from Churchman and Ackoff (1954) researches, under the name SAW (Simple Additive Weighting) [18-19]. SAW is one of more popular solutions in MADM type (Multi-Attribute Decision Making) problems of the which undoubtedly is a problem described in work. Elaborated process of EPQ calculation is similar to above methods, however there appear differences in designating of individual parameters. Differences are caused by different way of P_i parameters values determination.

$$EPQ = \frac{1}{n}\sum_{i=1}^{n} P_i * w_i \qquad (2)$$

where P_i is appropriate parameter of evaluation, with following index n appointed , and w_i is weight for given parameter. Below in the table (cf. Table I) there is presented list of parameters together with their asserted values and ranges. All parameters P_i are situated in the same range: $P \in [0,1]$.

V. The example of the EPQ calculation

The definition of the exact value EPQ does not decide about "the success" and the publication of the given magazine article. This will permit however finishing up and improvement of the editorial part which could not take into account mentioned above factors influencing decision of editors and reviewers. Elaborated in such way system using the informatics network , will permit quick definition of the article modification. Outwardly, it will enable based on obtained result EPQ to propose the alternative academic journal which parameters answer to the result. There was calculated the value EPQ for example of publication based on the *Matlab* software.

VI. SEO, Hirsch index and Impact Factor

A. The similarity of the Hirsch Index and Impact Factor to Page Rank, and threats resulting from Black Hat SEO methods?

The growth of the Hirsch index and IF is strictly dependent on the quantity of given author publication quotations. This model can be compared to the published ranking of websites (*PR - Page Rank*) used in Google search engine [5-7]. The similarity refers to the quantity of quotations which correspond to quantities of returnable links indicating given page of data sources.

There are known general methods of influencing the algorithm of search engine in this way, so that the indicated page will be higher in the SERP ranking (Search Engine Results Page). This methods are divided on so called white and black. White Hat SEO - means the positioning of the website in compliance with official guidelines of search engines, what should result in better page adaptation to Webcrawler's and engines of search engines requirements. Good preparation of the website facilitates quick indexing of it in the search engine base of data, however increasing number of valuable references to page (gained naturally and resulting from its popularity and uniqueness) permits its positioning and obtaining of high place in the SERP ranking. As valuable references are acknowledged links from pages about high PR which are often visited by users (e.g. thematic, community websites). There also exists Black Hat SEO which is characterized using all possible gaps in the search engine, for the purpose of raising the ranking of given website. Such effects are achieved through the manipulation with the quantity of returnable links and their "artificial" addition through generating large quantity of pages with links. So many of manipulation methods is the necessity of continuous algorithms change of search and qualitative selection of websites.

From obvious reasons, exact parameters of the algorithm are not revealed for the purpose of their protection before the manipulation. There can be only estimated general dependencies and on their base there can be created algorithms improving the position of website in ranking of searches. Methods of rankings creating e.g. PR and IF, and also H- index, cause the risk of appearing methods taken from SEO, which in the artificial way will manipulate results of mentioned above rankings. Probably there is no possibility of obtaining 100% reliable and objective ranking not burdened with the above risk.

From this reason, the essential evaluation of publication can be shaken, in the interest of the parametric evaluation. This can cause the reverse to intended effect i.e. these rankings will promote less ambitious scientific discoveries, but artificially will overvalue indexes across the elaboration of their manipulation method. Below there is presented case study, which in the mental experiment could result with

TABLE I.
DEFINING PARAMETERS FOR CALCULATION THE EPQ INDICATOR

No.	Parameter P_i	Meaning of value substituted to P_i	Formula on P_i	Range	Initial weight w_i	Range of weight
1	P_1	H – authors Hirsch index	$P_1=(1-\frac{1}{1-H})*(w_1)$	H=[0:inf]	1	[0:1]
2	P_2	I – the quantity of authors indexed publications	$P_2=(1-\frac{1}{1-I})*(w_2)$	I=[0:inf]	1	[0:1]
3	P_3	C – quantity of authors indexed quotations	$P_3=(1-\frac{1}{1+C})*(w_3)$	C=[0:inf]	1	[0:1]
4	P_4	S - degree/ the scientific title of the author (none/engineer/MSc/the doctor/assistant professor/professor)	$P_4=(1-\frac{1}{1+20*S})*(w_4)$	S=[0:5]	1	[0:1]
CONTENT RATING OF ARTICLE						
5	P_5	Gaussian distribution calculated basing on the quantity of all quotations contained by author in the article, where: d - height of the Gaussian curve top, x - quantity of all quotations contained by author in the article, σ - standard deviation of Gaussian distribution, μ - expected value, equal average quantity of quotations devolving on one article in the given journal, a - quotations devolving on one article (k) in the given journal.	$P_5=(d*e^{\frac{-(x-\mu)^2}{2\sigma^2}})*(w_5)$ $\sigma=\sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(x_i-\mu)^2}$ $\mu=\frac{1}{k}\sum_{i=1}^{k}a_i$	d=[1] x=[0:inf] σ=[0:inf] μ=[0:inf] a=[0:inf] k=[0:inf]	1	[0:1]
6	P_6	A - the quantity of quotations coming from archival numbers of the same journal to which the publication is submitted	$P_6=(1-\frac{1}{1-A})*(w_6)$	A=[0:inf]	1	[0:1]
7	P_7	B - the quantity of quotations coming from archival numbers of remaining journals belonging to the same publishing house to which publication is submitted	$P_7=(1-\frac{1}{1+B})*(w_7)$	B=[0:inf]	1	[0:1]
8	P_8	The indicator of the publication originality. O - the quantity of similar articles earlier published by the author. D - the sum of "duplicates", measured by the coefficient of similarity of genuine text and small pictures between previous articles of the author, and with his current publication	$P_8=(\frac{1}{1+O})*(w_8)$ $O=\sum_{i=1}^{n}D_i$	O=[0:inf] D=[0:inf]	1	[0:1]
9	P_9	R_d - the quantity of quoted publications of the current editor of journal to which publication is submitted	$P_9=(1-\frac{1}{1+R_d})*(w_9)$	R_d=[0:inf]	1	[0:1]
10	P_{10}	R_c - the quantity of quoted publications of current reviewer of journal to which publication is submitted	$P_{10}=(1-\frac{1}{1+R_c})*(w_{10})$	R_c=[0:inf]	1	[0:1]
OTHER PARAMETERS						
11	P_{11}	J - the quantity of authors publications quoted by current editor or reviewer of the journal to which publication is submitted	$P_{11}=(1-\frac{1}{1+J})*(w_{11})$	[0:inf]	1	[0:1]
12	P_{12}	K - the quantity of authors common publication articles and current editor or reviewer of journal to which publication is submitted	$P_{12}=(1-\frac{1}{1+K})*(w_{12})$	[0:inf]	1	[0:1]
13	P_{13}	Z - quantity of elements from the range carried researches (the form of survey): review, theory, model, simulation, experiment, lack/other.	$P_{13}=(1-\frac{1}{1+20+Z})*(w_{13})$	[0:5]	1	[0:1]

"artificial" increasing of IF for the journal, or with 'artificial" increasing of the H-index for given scientist.

B. *How to create Journal with IF=100 (Case Study I)?*

In the after-mentioned mental experiment we establish that one publishing house can belong to several academic journals having similar character, or there was undertaken cooperation between publishing houses for the purpose of one common journal strong promotion. The first journal

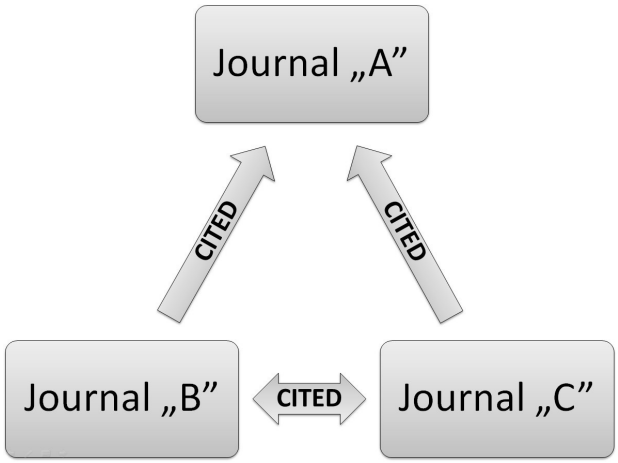

Fig 2. The diagram of quotations hinterland building for promoted journal.

„A" will be promoted, however remaining „B", „C" will constitute the hinterland with place for journal "A" quotations.

In this case arises classical system of hinterland with links (in our case - with quotations), known among computer scientists dealing with the positioning of websites (so called SEO - Search Engine Optimization). The example of building of quotations hinterland is presented on "Fig. 2". On the assumption that the journal "A" will have few articles in one publishing-cycle , then remaining journals can force writing for them authors, to quote several articles from journal "A". So extortionate ranking of promoted journal can have other advantages.

VII. The methodology of designating weights

Particularly essential from the usage of EPQ indicator point of view, is the possibility of weights definition w_i in way compatible to parametric evaluations applied by the given journal. The large number of academic journals causes different approach to the parametric evaluation of accepted to editorial office and the review of article. Basing on the data from previous years, considering all publications printed within the framework of one publishing-title , we are able to determine weights of individual parameters individually for the given journal.

For that purpose we will use neural networks with the feedback which will learn to recognize the influence of given parameter on the positive acceptance of article to the publication. In case of the analysis, already printed publications, we will subordinate the quantity of published articles from the value of individual parameters. The more articles will have e.g. the high parameter P6, the greater influence on the printing of publication has the quantity of archival articles quotations laded from the same journal.

VIII. Application realizing EPQ designating

For the purpose of individual parameters designation, we will use the access to individual databases, among others: SCOPUS, WEB OF KNOWLEDGE, and others. Application in the first instance will collect data: ref. of author, quotations, journal, publication, and then made the evaluation of parametric sent publication. Based on this evaluation, it can propose suggestions ref. introductions of changes in the article, or present proposal of the alternative journal to which the parametric evaluation was in order better. The system architecture may be built based on the client-server methodology what is presented on Fig. 3.

IX. The architectural schema

On the after mentioned Fig. 3 there is presented the general architectural schema of the system.

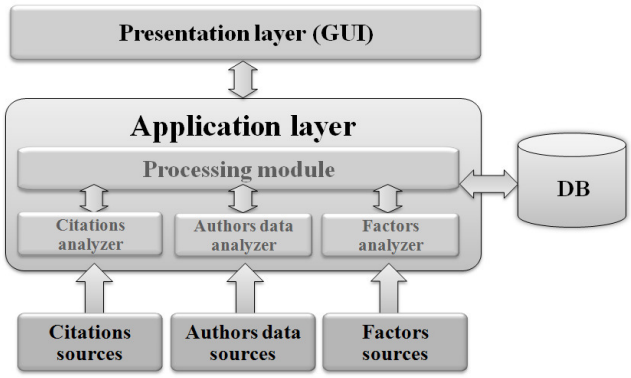

Fig 3. The architecture of proposed information system designating the EPQ coefficient.

In presented architecture system we distinguish:

1. *Presentation layer* - layer of the application responsible for the presentation of results and communication with user, receiving data from user (proposed article, survey for the author)
2. *Application layer* - layer responsible for the resumption of data and processing of results, consists of:
   - *citations analyzer* (module processing the quotation categorizing and counting quotations of authors works.

- *authors data analyzer* (module processing data of authors (also reviewers and editors), checking relations of author with journals across quotations as well as categorizing his achievement)
- *factors analyzer* (module being supposed for the task to process available data sources information of used in the algorithm coefficients for journals and authors)

1. *DB* - layer of database recording source data and results of calculations with application layer, permitting caching of data sources in the situation when data don't need to be refreshed at every operation of weight-coefficients calculation weight- coefficients.
2. *Sources* - layer of gaining data from chosen sources dividing into sources of quotations gaining ("citations sources"), given authors ("authors the date sources") and coefficients used in the algorithm of EPQ count ("factors sources").

The system architecture in case of further development can be calibrated because the module of processing may receive partial results of calculations (weights of component parameters) from individual modules which can be find on separate instances of servers. Every module of gaining data can have the separate database in which will store received results of the data sources indexing In case of presentation layer, the system can communicate with software of the *thin client* type in case of approach users (authors of articles) and with the software of the *fat client* type in case of the administrator who can control work of the processing module (settings control).

III. Conclusion

As this is the elaboration of the preliminary concept of articles parametric evaluation across proposing of the EPQ parameter , only verification on figures will permit the definition of its real effectiveness in the classification of articles to individual academic journals. Methodology is based on foundations that the substantially good article can be worse evaluated, due to remaining factors on which reviewers and editors of journals pay attention. Elaborated system, targets proper verifying and correction of article before delivering to publishing houses. This will permit to carry essential research on equally high level, and to regard of subjective 'expectations' from the side of publishing house in relation to the author. So improved article has greater chances for printing in the renowned journal, what can positively rebound on future publications of many authors.

References

[1] J. E. Hirsch, “An index to quantify an individual’s scientific research output.” *Proc. Nat. Acad, Sci.* (PNAS), 2005, vol. 102, nr 46, s. 16569-16572.

[2] L. Egghe, R. Rousseau, “Introduction to informetrics”. Amsterdam: Elsevier, 1990.

[3] H. F. Moed, M. Vriens, “Possible inaccuracies occurring in citation analysis”. *Journal of Information Science*, 1989, 15, 95-107.

[4] R. Todorov, W. Glanzel, “Journal citation measures: A concise review”. *Journal of Information Science*, 1988, I4. 47-65.

[5] S. Brin, L. Page. “The anatomy of a large-scale hypertextual web search engine”. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*, pages 107–117. Elsevier Science Publishers B. V., 1998.

[6] L. Page, S. Brin, R. Motwani, T. Winograd, “The PageRank citation ranking: Bringing order to the web”. Technical Report, Stanford Digital Library Technologies Project, 1998.

[7] S. Maslov, S. Redner, “Promise and Pitfalls of Extending Google’s PageRank Algorithm to Citation Networks”, *Journal of Neuroscience* 2008, 28, 11103.

[8] D. Zhou, S. A. Orshanskiy, H. Zha, C. L. Giles, “Co-Ranking Authors and Documents in a Heterogeneous Network”, *INTERNATIONAL CONFERENCE ON DATA MINING*, 2007.

[9] P. Chen, H. Xie, S. Maslov, S. Redner. “Finding Scientific Gems with Google”. *J.INFORMET.*, 1:8, 2007.

[10] E. Garfield. “Citation analysis as a tool in journal evaluation”. *Science*, 178(60):471–479, November 1972.

[11] X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel. “Coauthorship networks in the digital library research community”. *arXiv.org:cs/0502056*, 2005.

[12] M. Bianchini, M. Gori, and F. Scarselli. “Inside pagerank”. *ACM Trans. Inter. Tech.*, 5(1):92–128, 2005.

[13] De Moya, F. “The SJR indicator: A new indicator of journals’ scientific prestige”, 2009, Arvix. (arxiv.org/abs/0912.4141)

[14] E. Garfield, “The history and meaning of the journal impact factor”. JAMA 2006;295:90-3.

[15] M. Amin, M. Mabe, “Impact factors: Use and Abuse”. *Perspectives in Publishing, 2000 - October, 1,* 1-6

[16] Mu-Hsuan Huang, Wen-Yau Cathy Lin, "The influence of journal self-citations on journal impact factor and immediacy index", *Online Information Review*, 2012, Vol. 36 Iss: 5, pp.639 – 654

[17] C. Wenneras, A. Wold ”Nepotism and sexism in peer-review”. *Nature*. 1997 May 22;387(6631):341-3.

[18] C.W. Churchman, R.L. Ackoff, “An approximate measure of value”. *Journal of the Operational*, *Research Society of America*, 1954, **2**(2), 172–187.

[19] D. Widayanti , S. Oka, S. Arya, “Analysis and Implementation Fuzzy Multi-Attribute Decision Making SAW Method for Selection of High Achieving Students in Faculty Level” *IJCSI International Journal of Computer Science Issues*, Vol. 10, Issue 1, No 2, January 2013, ISSN 1694-0784

[20] SNIP and SJR at Journal Metrics (www.journalmetrics.com)

[21] SCImago Journal Rank (SJR) (http://www.scimagojr.com/)

[22] Source-Normalized Impact per Paper (SNIP) (www.journalindicators.com)

[23] Impact Factor (IF) (http://thomsonreuters.com/products_services/science/free/essays/impact_factor)

[24] h-index (http://help.scopus.com/robo/projects/schelp/h_hirschgraph.htm)

[25] Article Influence (AI) (www.eigenfactor.org)

[26] Relative Citation Rates (RCR)/Journal to Field Impact Score (JFIS)