A review of ranking problems in scientometrics and informetric

Non Peer-reviewed author version

EGGHE, Leo (2000) A review of ranking problems in scientometrics and informetric.
In: Scientific research evaluation and norms: proceedings. p. 2-15..

Handle: http://hdl.handle.net/1942/795

A REVIEW OF RANKING PROBLEMS IN SCIENTOMETRICS AND INFORMETRICS

by

L. EGGHE

LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium[1]
and
UIA, Informatie- en Bibliotheekwetenschap, Universiteitsplein 1,
B-2610 Wilrijk, Belgium
e-mail : legghe@luc.ac.be

ABSTRACT

The paper highlights problems with the ranking of average scores of scientific groups. These can be considered in a wide sense - e.g. scientific disciplines formed by journals or authors and where one studies e.g publication or citation scores. Other scores as aging or prices (for journals) are also possible as well as other groups (e.g. a country, studied w.r.t. a scientific discipline).

It is so that, depending on the used scoring device, the relative importance of such groups can increase or decrease. It is even so that rankings can increase (i.e. the group becoming less important) while their relative score increases.

The use of geometric averages, to overcome this problem, is explained and it is strongly advised to use it.

---

[1]Permanent address

I. Introduction

Let us introduce the problem by fixing a concrete case of a group and of a score, namely the case of a scientific subject (or discipline) (as e.g. defined by ISI) and their citation scores.

The Science Citation Index/Journal Citation Reports, (SCI/JCR) and the Social Science Citation Index/Journal Citation Reports, (SSCI/JCR) classify covered journals into subject categories. Part IV of the JCR, gives a listing of journals ranked by impact factor within category. Using these lists, it is not difficult to compute the average impact factor of ISI-journals of a subject category. Yet, it is of more importance to know the impact of a whole subject category, i.e. the global impact factor. Indeed, a comparison between the observed impact and the average impact of those subcategories in which a research group is active, is one of the most representative indicators in the evaluation of research groups [1], [2]. If every journal would publish the same number of articles the average impact factor would clearly be the same as the global impact factor. Yet, in [3], see also [4], we have shown that this is usually not the case. Indeed, [4] is, to the best of our knowledge, the first article which explicitly drew attention to the difference between the average impact factor of a subfield and its global impact factor.

Furtheron in this article we will use the notions 'subfield' and 'impact' in a general sense. Subfields do not have to coincide with subcategories of the JCR (although, of course, this is one of the important applications) and the impact is not necessarily the 'official' Garfield impact factor. It could as well be a generalized impact factor in the sense of [5], [6]. So, by the term 'impact', denoted as I, we will mean the quotient of a number of citations, denoted as C, by the corresponding number of publications, denoted as P (over a well-defined period). If we want to stress the fact that citations are a function of publications, we write C(P), and similarly for the impact :

$$I(P) = \frac{C(P)}{P} \tag{1}$$

Note that C(P) is an increasing function of P. The set of all journals under consideration (the 'subfield') is called a meta-journal [3], [5]. The impact of the i-th journal of this meta-journal is then denoted as I_i. The average impact factor, denoted as AIF, of a meta-journal consisting of N journals is defined as :

$$AIF = \frac{1}{N}\sum_{i=1}^{N}\frac{C_i}{P_i} = \frac{1}{N}\sum_{i=1}^{N} I_i \qquad (2)$$

The global impact factor of this same meta-journal, denoted as GIF, is then defined as :

$$GIF = \frac{\sum_{i=1}^{N} C_i}{\sum_{i=1}^{N} P_i} = \frac{\mu_C}{\mu_P} \qquad (3)$$

where μ_C is the average number of citations :

$$\mu_C = \frac{\sum_{i=1}^{N} C_i}{N} \qquad (4)$$

and μ_P is the average number of publications :

$$\mu_P = \frac{\sum_{i=1}^{N} P_i}{N} \qquad (5)$$

Finally, the ratio GIF/AIF will be denoted as ρ :

$$\frac{GIF}{AIF} = \rho \qquad (6)$$

II. The relation between ρ=GIF/AIF and the slope of the regression line of I(P) over P

That GIF can be smaller than AIF or reverse, is shown by the next theorem.

Theorem II.1 [8]. If r_p denotes the slope of the regression line of I(P) over P, then

$$r_p > 0 \leftrightarrow \rho > 1 \tag{7}$$

and also

$$r_p = 0 \leftrightarrow \rho = 1 \tag{8}$$

Proof. We will only show (7).

$$\rho > 1 \tag{9}$$

$$\leftrightarrow \sum_{i=1}^{N} \frac{C_i}{P_i} < N.\frac{\mu_C}{\mu_P} \tag{10}$$

On the other hand, the slope r_P of the regression line of I(P) over P is larger than 0

$$\leftrightarrow N \sum_{i=1}^{N} P_i \cdot \frac{C_i}{P_i} - \left(\sum_{i=1}^{N} P_i\right)\left(\sum_{i=1}^{N} \frac{C_i}{P_i}\right) > 0 \tag{11}$$

(see e.g. [7], p. 66)

$$\leftrightarrow \mu_C > \mu_P \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{C_i}{P_i} \tag{12}$$

$$\leftrightarrow \sum_{i=1}^{N} \frac{C_i}{P_i} < N \cdot \frac{\mu_C}{\mu_P} . \qquad \square \tag{13}$$

Corollary II.1 [8].
(i) If I(P) is decreasing in P, then $\rho < 1$
(ii) If I(P) is increasing in P, then $\rho > 1$.

That GIF<AIF can be as well as the opposite GIF>AIF is not a bad property in itself since the comparison is made for the same group. It would be bad, of course, if the above described differences between GIF and AIF could lead to interchange of rankings between different groups. This can indeed be obtained, cf. [9]. We will illustrate this (very bad) property in the framework of country scores w.r.t. author-counts, as a second field of application.

III. Country scores w.r.t. author's publications.

It is an intricate problem to measure multiple authorship or citation counts to papers with multiple authors. Another aspect related with this is the measurement of country-scores, as obtained from the authors' data.

One might be convinced that it suffices to stipulate in a clear way what type of measurement one uses in order to have an unambiguous study or article on this problem. Unfortunately, this is not so. Let us first indicate four different types of measurement of author (country, institutions,...) scores in case of multi-authored papers.

(1) First author counting : only the first of the N authors (N=1,2,3,...) of a paper is given credit and the credit given is 1, cf. [7], [10]. The method is sometimes called straight counting.

(2) Total counting : each of the N authors receives a credit 1, cf. [11]. This counting method is also called normal (or standard) counting, see also [7].

(3) Fractional counting : each of the N authors receives a score of $\frac{1}{N}$, [7],[11],[12],[13],[14]. This counting method is sometimes called adjusted counting.

(4) Proportional counting : If an author has rank R in a paper with N authors (R=1,2,...,N), then he/she receives a score of

$$\frac{2}{N}\left(1-\frac{R}{N+1}\right) . \tag{14}$$

This formula for the credit in case of proportional counting is obtained by dividing the absolute weights N+1-R by the sum of all ranks :

$$1+2+...+N = \frac{N(N+1)}{2} ,$$

cf. [15].

Here we focus on the counting methods (2), (3) and (4). It is our feeling that the first method falls outside the methodology of study of the other counting procedures, yet it would be interesting to have a stochastic model in this case. We do not consider it further on.

We are interested in several questions. First of all we want to know how the relative score of an author (country,...) is affected by the counting procedure. Indeed, only relative scores are important : absolute scores are incomparable since the totality of attributed weights is different from one counting method to another. So we want to know the individual scores of each author (country,...) within a certain counting method, divided by the totality of scores that are given in this counting method.

We will give formulae for these relative scores. We will show, by formulae and by examples, that it is very well possible to have a larger score for author a than for author b in one counting method while the opposite is true in another method. As a consequence of this, the relative importance (expressed e.g. by the obtained ranks) is reversed by changing the counting procedure.

In addition to this - and this is almost a paradox - it is even possible to have an increase of the relative score of an author when going from one counting method to another and at the same time an increase of the rank of this author. In other words this author becomes more important according to his/her relative score and less important according to his/her rank, and this within the same counting procedure as compared with the first counting procedure!

III.1 Formulae for counting procedures

We will first fix some general notation and terminology. We will consider N articles and for each article $i=1,..,N$, let a_i denote the number of co-authors of article i. Let c be a country. Then $a_i(c)$ will denote the number of co-authors from country c in article i. The term "country" must be interpreted in a general way : it can be a real country name or the name of an institution or even an author's name. The latter case is then a special case of this general formalism, namely where $a_i(c)=0$ or 1. In general, for a real country or institution, $a_i(c)$ can be (in principle) any natural number (including zero). That is why we will use this more general framework. Clearly, $\forall i=1,\ldots,N$:

$$a_i = \sum_c a_i(c), \tag{15}$$

III.1.1 Total counts.

Every author receives a score of 1, hence the total score in this system is

$$W_2 = \sum_{i=1}^{N} a_i \tag{16}$$

By notation, the total score of country c is

$$W_2(c) = \sum_{i=1}^{N} a_i(c) \tag{17}$$

and hence, its relative score is

$$Q_2(c) = \frac{W_2(c)}{W_2} = \frac{\sum_{i=1}^{N} a_i(c)}{\sum_{i=1}^{N} a_i} \tag{18}$$

Note that (18) corresponds to GIF (formula (3)).

III.1.2 Fractional counts.

Since the total score for an article is 1, we have that the total score in this system is $W_3=N$. Each co-author receives a score $\frac{1}{a_i}$ in article i and hence, country c receives an absolute score of

$$\frac{a_i(c)}{a_i} \tag{19}$$

in article i

(there are $a_i(c)$ co-authors from country c).

The total score of country c is

$$W_3(c) = \sum_{i=1}^{N} \frac{a_i(c)}{a_i} \tag{20}$$

and hence, its relative score is

$$Q_3(c) = \frac{1}{N} \sum_{i=1}^{N} \frac{a_i(c)}{a_i} \tag{21}$$

Note that (21) corresponds to AIF (formula(2)).

III.1.3 Proportional counts.

We will not go further into the formula for proportional counts since it is intricate and since we do not need it further on. We will be able to draw conclusions for $Q_4(c)$ (the relative score of country c when proportional counts are used) by using the following theorem, which can be found in [9].

Theorem III.1 : For any system of collaboration between countries (institutions, authors,...) we can construct another one such that the proportional counting system in the latter one equals the fractional counting system in the former as well as in the latter one. Furthermore the total counting systems are also the same in both systems.

This theorem enables us to deduce results on the comparison of $Q_2(c)$ and $Q_4(c)$ whenever we have a result on the comparison of $Q_2(c)$ and $Q_3(c)$. Therefore we limit our attention to the comparison between total counting and fractional counting and obtain an analogous result on the comparison between total counting and proportional counting, without additional work.

III.2 Nontrivial examples of anomalies between scores of total and fractional counts and between scores of total and proportional counts.

We are interested to see whether an example exists in which (c, c' : countries, authors, institutions, ...) :

$$Q_2(c) > Q_2(c') \tag{22}$$

$$Q_3(c) < Q_3(c') \tag{23}$$

This would be a first sign of the ambiguity of these two counting methods : in the total counting system c is considered to be more important than c' (since we are dealing with relative scores : c occupies a larger fraction of the scores than c') ; in the fractional counting system, exactly the opposite would be true!

The paradox would be complete if the example could also give

$$Q_2(c) < Q_3(c) \tag{24}$$

Then (24) indicates that c has a smaller weight in the total counting system than in the fractional counting system, yet this smaller weight has a higher importance (as compared to c') than the larger weight in the respective counting systems, because of (22) and (23)!

If an example of the simultaneous occurence of (22), (23) and (24) can be given, this would then be a definitive condemnation of the used methodology and make many studies that draw conclusions on this basis (at least) doubtful. In this section we show that such examples indeed exist, albeit for relatively complex systems (of total number of articles).

Nevertheless we will be able to give examples in the simple case of 3 authors (or - equivalently - 3 countries for which we can even assume $a_i(c)=1$ for every $i=1,\dots,N$).

Based on the formulae of $Q_2(c)$ and $Q_3(c)$ in the previous section and on (22), (23) and (24) we are able to construct the examples. a,b,c, denote authors (or countries) and the rows denote publications.

First example

	a	b	c
1		x	
2		x	
3	x	x	
4	x	x	
5	x	x	
6	x	x	
7	x	x	
8	x	x	
9	x	x	
10	x	x	
11	x	x	
12	x	x	
13	x	x	
14	x	x	
15	x	x	
16	x		x
17	x		x
18	x		x
19	x	x	x
20	x	x	x
21	x	x	x
22	x	x	x
23	x	x	x
24	x	x	x
25	x	x	x
26	x	x	x
27	x	x	x
28	x	x	x
29	x	x	x
30	x	x	x
31	x	x	x

This results in the following values of Q_2 and Q_3.

Q_2	Q_3
$Q_2(a) = \underline{0.3973}$	$Q_3(b) = 0.4140$
$Q_2(b) = 0.3836$	$Q_3(a) = \underline{0.3978}$
$Q_2(c) = 0.2192$	$Q_3(c) = 0.1882$

Note that $Q_2(a) > Q_2(b)$, $Q_3(a) < Q_3(b)$ and $Q_3(a) > Q_2(a)$ as required.

We will now see if an example exists in which a, b and c do not collaborate all together.

We have the following solution.

	a	b	c
1	x		
2	x		
3	x		
4	x		
5	x		
6		x	
7		x	
8		x	
9		x	
10		x	
11		x	
12		x	
13			x
14	x	x	
15	x	x	
16		x	x
17		x	x
18		x	x
19		x	x
20		x	x
21	x		x
22	x		x
23	x		x
24	x		x
25	x		x
26	x		x
27	x		x
28	x		x

Now we have the Q_2 and Q_3 values

Q_2	Q_3
$Q_2(a) = \underline{0.3488}$	$Q_3(b) = 0.3750$
$Q_2(b) = 0.3256$	$Q_3(a) = \underline{0.3571}$
$Q_2(c) = 0.3256$	$Q_3(c) = 0.2679$

Note again that $Q_2(a) > Q_2(b)$, $Q_3(a) < Q_3(b)$ and $Q_3(a) > Q_2(a)$.

IV. A solution to the anomalies in ranking.

Basing ourselves on the results in [16] we have that the geometric "versions" of Q_2 and Q_3 are (denoted by Q_2^g, Q_3^g) :

By (18), Q_2 is

$$Q_2(c) = \frac{\text{arithmetic average of } \{a_1(c),\ldots,a_N(c)\}}{\text{arithmetic average of } \{a_1,\ldots,a_N\}} ,$$

hence

$$Q_2^g(c) = \frac{(a_1(c)a_2(c)\ldots a_N(c))^{\frac{1}{N}}}{(a_1a_2\ldots a_N)^{\frac{1}{N}}} \tag{25}$$

By (21), Q_3 is

$$Q_3(c) = \text{arithmetic average of } \left\{\frac{a_1(c)}{a_1},\ldots,\frac{a_N(c)}{a_N}\right\} ,$$

hence

$$Q_3^g(c) = \left(\frac{a_1(c)}{a_1} \cdot \frac{a_2(c)}{a_2} \cdots \frac{a_N(c)}{a_N}\right)^{\frac{1}{N}} . \tag{26}$$

It is now clear from (25) and (26) that

$$Q_2^g(c) = Q_3^g(c) \tag{27}$$

for any system and any c. Furthermore, if this is so, obviously all rankings in the total counting system are the same as in the fractional counting system. Hence all ambiguities are gone!

We leave it as an open problem to study the behavior of the proportional counting system in this context, i.e. the study of $Q_4^g(c)$.

The problems as discussed in this paper have many applications, not only in relation with citation scores or publication scores. In fact we can even give examples beyond the scope of informetrics and scientometrics.

V. More applications.

V.1. Price index

For an article, one counts the total number of references and the number of references which are given to articles not older than d years, where the year of publication is counted as year one. (Price [17] uses d=5). Its quotient is Price's index. Moed [18] uses the average Price index as a measure for the field, while Price himself used the global one. Of course, one can equally well calculate Price's index for a journal. The difference between the two approaches was discussed by Wouters and Leydesdorff [19]. They note that for the journal Scientometrics the average - five year - Price index was 0.514, while the global one was 0.43. From the fact that the global one is smaller than the average one, we conclude that the slope of the regression line of Price's index over the total number of references is decreasing. Price's index was further studied in [20].

V.2. Text to reference ratio

In [21], a number of journals was randomly selected and, for each article published during the years 1980 and 1987, the (estimated) number of words and the number of references were obtained. Its quotient yields the text to reference ratio of each article. As far as we could see, the authors did not state how they obtained the journal data ; namely, as an average of article data or as global text to reference ratios. We assume they used the global method.

V.3. Receptivity factor

In this application, the field is fixed. For every article under consideration, one collects the number of references and the number of references to articles written by fellow countrymen. This is done per country, where, again, it is possible to take an average or a global point of view. Dividing this result by the share of the country in the total output in the field yields the receptivity index for foreign literature (cf.[22]).

V.4. Journal prices

For every journal, one collects the number of published pages (in a year) and the subscription price. The quotient is the price per page. One could similarly calculate the price per character, or the price per citation, as a kind of 'value for money' indicator. Results can then be brought together per field or per publisher [23]. For some publishers, this seems to be a very controversial procedure [24, 25]!

In [4] one calculates a price per article. The authors give preference to the weighted, i.e. global, price per article of a discipline or faculty, as they also do for the case of the impact factor.

V.5. Ageing

For each article, one counts c_j, the number of references to articles which are j years old, $j=0,1,\ldots,10$. Here, the number 10 is used for convenience ; we further assume that none of the c_j is equal to zero. Then the ageing rate r_k of this article is determined as the average of the quotients c_{j+1}/c_j, $j=0,\ldots,9$. Note that other definitions of ageing are used in the literature. The above definition is used only as an example.

Then, to determine the ageing rate of a journal, one can use the average of the ageing rate of all articles (AAR : average ageing rate of a journal), or one can use a global approach, i.e. take the sum of all c_js, form quotients and then take the average of all quotients (GAR : global ageing rate of a journal).

V.6. Gross regional product (GRP)

Averages, as discussed in this article, occur in many fields. A well-known example from econometrics is the case that for every country one collects the gross national product (GNP) and the number of inhabitants. Its quotient yields the GNP per capita. When considering, for example, the GRP per capita in the European Union, one can take the average of all GNP per capita of every member country, or one could calculate the global GRP per capita. Again, we think that the second index is the more significant one.

For other applications (discipline influence score, fill-rates as measures of library performance) we refer to [16].

References

[1] R.E. De Bruin, H.F. Moed and E. Spruyt. Antwerpse analyses. Rapport ten behoeve van de bestuursorganen van de Universiteit Antwerpen, 1993.

[2] R. Rousseau. A scientometric study of the scientific publications of LUC. Report, 1995.

[3] R. Rousseau and G. Van Hooydonk. Journal production and journal impact factors. Journal of the American Society for Information Science, 47(10), p. 775-780, 1996.

[4] G. Van Hooydonk, R. Gevaert, G. Milis-Proost, H. Van De Sompel and K. Debackere. A bibliotheconomic analysis of the impact factors of scientific disciplines. Scientometrics, 30, p. 65-81, 1994.

[5] R. Rousseau. Citation distribution of pure mathematics journals. In : Informetrics 87/88. L. Egghe, R. Rousseau, (Eds.), Amsterdam, Elsevier, p. 249-262, 1988.

[6] R. Rousseau. A note on maximum impact factors. In : Information as a Global Commodity : Communication, Processing and Use, CAIS/ACSI '93, 21st Annual Conference, 11-14 July 1993, p. 120-125, 1993.

[7] L. Egghe and R. Rousseau. Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science, Elsevier, Amsterdam, 1990.

[8] L. Egghe and R. Rousseau. Average and global impact of a set of journals. Scientometrics, 36, p. 97-107, 1996.

[9] L. Egghe, R. Rousseau and G. Van Hooydonk. The behavior of scores obtained through total, fractional or proportional counting. Preprint 1998.

[10] J.R. Cole and S. Cole. Social stratification in Science. The University Press, Chicago, 1973.

[11] D. Lindsey. Production and citation measures in the sociology of science : the problem of multiple authorship. Social Studies of Science, 10, p. 145-162, 1980.

[12] D.J. De Solla Price . Letter to the Editor. Science, 212, p. 987, 1981.

[13] Q. Burrell and R. Rousseau. Fractional counts for authorship attribution : a numerical study. Journal of the American Society for Information Science, 46, p. 97-102, 1995.

[14] L. Egghe. Source-item production laws for the case that items have multiple sources with fractional counting of credits. Journal of the American Society for Information Science, 47, p. 730-748, 1996.

[15] G. Van Hooydonck. Fractional counting of multi-authored publications : consequences for the impact of authors. Journal of the American Society for Information Science, 48, p. 944-945, 1997.

[16] L. Egghe and R. Rousseau. Averaging and globalising quotients of informetric and scientometric data. Journal of information Science, 22, p. 165-170, 1996.

[17] D. De Solla Price. Citation measures of hard science, soft science, technology, and nonscience. In : C.E. Nelson and D.K. Pollack (eds.), Communication Among Scientists and Engineers, Heath, Lexington, MA, p. 3-22, 1970.

[18] H.D. Moed. Bibliometric measurement of research performance and Price's theory of differences among the sciences. Scientometrics, 15, p. 473-483, 1989.

[19] P. Wouters and L. Leydesdorff. Has Price's dream come true : is scientometrics a hard science ? Scientometrics, 31, p. 193-222, 1994.

[20] L. Egghe. The Price Index and Its relation to the mean and median reference age. Journal of the American Society for Information Science, 48, p. 564-573, 1997.

[21] A.E. Little, R.M. Harris and P.T. Nicholls. Text to reference ratios in scientific journals. In : L. Egghe and R. Rousseau (eds.), Informetrics 89/90, Elsevier, Amsterdam, p. 211-216, 1990.

[22] I.L. Herman. Receptivity to foreign literature : a comparison of UK and US citing behavior in librarianship and information science. Library and Information Science Research, 13, p. 37-47, 1991.

[23] H.H. Barschall and J.R. Arrington. Cost of physics journals : a survey. Bulletin of the American Physical Society, 33, p. 1437-1447, 1988.

[24] C. Holden. Gordon and Breach impanels a journal jury. Science, 249, p. 298-299, 1990.

[25] A.L. O'Neill. The Gordon and Breach litigation : a chronology and summary. Library Resources and Technical Services, 37, p. 127-133, 1993.