Journal of Documentation, Vol. 54, No. 5, December 1998

RESEARCH BRIEF

RANKING OF SCIENTISTS – A NEW APPROACH

B.K. SEN

Visiting Professor, Faculty of Computer Science & Information Technology
University of Malaya, 50603 Kuala Lumpur, Malaysia

T.A. PANDALAI
ARUNA KARANJAI

National Centre on Bibliometrics, INSDOC, New Delhi - 110067, India

A formula for the ranking of scientists based on diachronous citation counts is proposed. The paper generalises the fact that the citation generation potential (CGP) is not the same for all papers, it differs from paper to paper, and also to a certain extent depends on the subject domain of the papers. The method of ranking proposed in no way replaces peer review. It merely acts as an aid for peers to help them arrive at a better judgement.

INTRODUCTION

One of the off-shoots of the Renaissance was the emergence of learned societies in Europe from the second half of the sixteenth century. Accademia Secretorum Naturae, established in 1560 in Naples by della Porta, was the earliest of such learned societies. Thereafter, Accademia dei Lincei was founded in Rome in 1603, Accademia del Cimento in Florence in 1651, the Royal Society of London in 1662, Académie des Sciences in France in 1666, and so on [1].

Another offshoot of the Renaissance was the emergence of periodicals which in many cases came about by the need felt by the members of the learned societies to disseminate the results of their research. *Le Journal des Sçavans*, the first scientific periodical in the real sense of the term, was started in January 1665 in France, followed by *Philosophical Transactions* that appeared in May 1665 in England. Afterwards periodicals appeared in various parts of Europe and by the end of the century there were as many as thirty-five titles comprising twenty journals, eight proceedings, two reviews, two almanacs, and three others [2].

When the journals started receiving research articles from various scientists, the need arose to verify the research results and the claims made. For example, in October 1776 when Antoni van Leeuwenhoek sent a letter to the Royal Society of London describing his discovery of 'little animals' with the crude microscope he invented, many members of the Royal Society simply could not believe it thereby obliging Leeuwenhoek to procure 'written attestations to the reliability of his observations from ministers, jurists, and medical men' [3]. To avoid the recurrence

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

of such an awkward situation, gradually the system of peer review developed which continues to date. Peer review is not restricted only to deciding the suitability of a paper for publication in a journal but also to judging the worthiness of a scientist for an award or for the membership of a society. Peer review, though considered possibly the best method of judgement for such purposes, is not always beyond criticism. At times, even the committee deciding the Nobel Prize each year has been criticised for improper selection of the candidate!

The Council of Scientific and Industrial Research, one of the most eminent scientific organisations of India, instituted an award in memory of its first director and architect the late Dr Shanti Swaroop Bhatnagar in 1957 [4]. The award, popularly known as the Bhatnagar Award, is given every year for notable contributions by Indians in the field of mathematical sciences, physical sciences, chemical sciences, earth sciences, biological sciences, medical sciences and engineering sciences based on the last five years' contributions of the scientists. Since the very beginning the awardees have been selected by a panel of peers comprising the top ranking scientists of the country. As can be expected, there are different panels of peers for different subjects. At the end of the 1980s, at the instance of the then Director General of CSIR, Dr A.P. Mitra, FRS, citation analysis of the contributions of the nominated candidates was introduced along with peer review to decide the suitability of the candidate for the award.

The idea of ranking scientists according to the number of citations was mooted by Garfield [5] way back in 1968, when he drew up a list of the fifty most cited scientists (Table 1) using the *Science Citation Index* database of 1967 and predicted that some of the future Nobel laureates would be from that unique list of fifty. Within barely a year, the prediction came true as M. Gellmann (ranked 6th in the list) and D.H.R. Barton (ranked 41st in the list) won the Nobel Prize respectively for physics and chemistry in 1969. In later years H.C. Brown, G. Herzberg, N.F. Mott, C. Deduve, U.S. von Euler, P.J. Flory figuring in the list also won the Nobel Prize. It may be noted that L.D. Landau, L. Pauling, J.C. Eccles, R.S. Mulliken, F. Jacob, all appearing in the list, won the Prize earlier. The list containing as many as thirteen Nobel laureates gave enough indication of the effectiveness of citation analysis in the evaluation of scientists and scientific papers.

Garfield's list is based on the citations received by a scientist in one year; the citations considered here are synchronous. Moreover, the list encompasses scientists from all fields. Our case was slightly different. The citation profiles we were compiling listed the citations received over a number of years, i.e. diachronous citations, and we were intending to rank the scientists field by field instead of producing a single ranked list like Garfield. The question arose as to whether Garfield's method of ranking would be applicable in our case. When Garfield's method of ranking scientists was tried, the inadequacy of the method for our purpose became evident. In a number of cases it was observed that several scientists received more or less same number of citations from a widely varying number of papers. The productivity of scientists even within the same subject and period (five years) varied widely from a meagre four papers to over forty papers. The citations received by papers belonging to the same field also varied widely. Some papers drew only one or two citations in five years, and many drew ten or more citations just in one year. This led us to the premise that a scientist, say A, receiving p number of citations

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

Table 1. *Fifty most cited scientists of 1967**

Rank	*Scientist*	*Total times cited*	*Rank*	*Scientist*	*Total times cited*
1	Lowry D H	2921	26	Eliel EL	721
2	Chance B	1374	27	Streiweiser A	717
3	Landau L D (62 Phys)	1174	28	Mulliken RS (66 Chem)	712
4	Brown H C (79 Chem)	1150	29	Jacob F (65 Med)	711
5	Pauling L (54 Chem)	1063	30	Born M	710
6	Gell-Mann M (69 Phys)	942	31	Brachet J	706
7	Cotton FA	940	32	Winstein S	702
8	Pople JA	933	33	Albert A	687
9	Bellamy IJ	906	34	Luft JH	674
10	Snedecor GW	904	35	Deduve C (74 Med)	673
11	Boyer PD	893	36	Von Euler US (70 Med)	668
12	Baker BR	876	37	Fieser Lf	666
13	Kolthoff IM	853	38	Huisgen R	661
14	Herzberg G (71 Chem)	842	39	Novikoff AE	655
15	Fischer F	826	40	Goodwin TW	643
16	Seitz F	822	41	Barton DHR (69 Chem)	632
17	Djerassi C	801	42	Fisher RA	631
18	Bergmeyer HU	754	43	Bates DR	627
19	Webwer G	750	44	Flory PJ (74 Chem)	626
20	Reynolds ES	748	45	Stahl E	626
21	Mott NF (77 Phys)	741	46	Dewar MJS	619
22	Eccles JC (63 Med)	737	47	Gilman H	618
23	Feigi F	729	48	Folch JZ	618
24	Freud S	727	49	Dische Z	614
25	Pearse AGE	726	50	Glick D	609

*Information within brackets following a name (added by us) indicates the year of winning the Nobel Prize by the scientist, and the subject.

from *m* number of papers deserves higher ranking than scientist B receiving *p* number of citations from *n* number of papers where *n*>>*m*. Let us illustrate this with a concrete case. In Table 2 it is seen that KCR received eighty-seven citations from four papers and SGA received eighty-six citations from twenty-one papers. According to our premise KCR deserves higher ranking as he received the citations from a very small number of papers compared to SGA. In other words, the papers by KCR were found to be good papers in the sense that they attracted a good number of citations. The premise warranted a new method of ranking.

When we started evaluating forty-odd research laboratories of CSIR conducting research in almost all major disciplines of science in 1988 we observed that the performances of laboratories are better revealed by the mean impact factor rather than the total impact factor [6]. By the same analogy we thought that the mean citation score would be a better method of ranking of scientists. Accordingly the formula given below was tried for ranking the scientists.

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

$$R = c/p \tag{1}$$

Where R stands for rank;
c stands for total number of citations; and
p stands for total number of cited papers.

It is to be noted that Irvine and Martin in their evaluation of CERN and other facilities also laid emphasis on mean citation score [7].

The formula we used for ranking was working all right. However, the problem arose when we observed that in certain cases two different scientists were getting the same rank with the value of R being the same, though the number of papers contributed by them varied widely. For example, a scientist C receiving twenty citations out of four papers published in five years gets the same rank as that of the scientist D who gets one hundred citations from twenty papers published during the same period as in both the cases citations per paper average five. It was felt that scientist D deserves a higher ranking than scientist C, because not all papers of scientists draw an equal number of citations which we observed from citation profiles of hundreds of scientists we compiled at the National Centre of Bibliometrics, a Sectoral Information Centre of NISSAT (National Information System for Science and Technology), established at INSDOC, New Delhi in 1988. We observed that the papers of a scientist (when the number is reasonably high) can be divided into four categories on the basis of citations they receive, i.e. (i) a few papers receive a large number of citations, (ii) a larger number receives a smaller number of citations, (iii) a substantial number receive a very small number of citations, and (iv) the remainder receive no citations at all. Hence, it was reasoned that of the twenty papers of D receiving one hundred citations, a few may draw twenty or more citations and some may not draw any citations at all. In reality, such cases were observed and it was decided to rank D above C receiving twenty citations *in toto*. For this, a new formula had to be evolved. Experimentation started with different formulas. Some of these are:

$$R = \frac{(c - s)^{3/2}}{p} \tag{2}$$

$$R = \frac{(c - s/2)^2}{p} \tag{3}$$

$$R = \frac{(c - s)^2}{p} \tag{4}$$

where R = the rank score
c = total number of citations received by the papers
s = self citations
p = number of cited papers.

The last formula served our purpose best. The limitations faced in ranking the scientists with the same number of citations per paper was sufficiently reduced with the formula. Mathematically, it can be shown as follows:

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

Suppose the scientist E received h number of citations (omitting self citations) out of b number of papers; and the scientist F received f number of citations (omitting self citations) out of a number of papers, where $h>f$, and $b>a$. Let the average citations per paper in both the cases be equal, hence

$$\frac{h}{b} = \frac{f}{a} = c \text{ (say)} \tag{5}$$

Self citations were not considered for deciding the rank following the advice of some peers.

Now, let us take the rank score of E and F respectively as R_1 and R_2.

According to equation 4,

$$R_1 = \frac{h^2}{b} = ch \because h/b = c \tag{6}$$

$$R_2 = \frac{f^2}{a} = cf \because f/a = c \tag{7}$$

We know $h>f$, hence $ch>cf$, $\therefore R_1>R_2$

This brings about the difference in ranking and gives the deserving scientist the higher ranking.

Let us demonstrate it with a simple example. Taking the values of h, b, f, and a as 100, 25, 60 and 15, we find $R_1 = 400$ and $R_2 = 240$ even though the average number of citations in both the cases is 4.

The ranking of scientists using the above formula has been quite effective. The peer review deciding the award and the ranking of scientists using the above formula have shown a close match (Table 3). In most cases, the peers have selected those scientists for awards who ranked among the first ten in the list. In certain cases, there has been an exact match. For almost a decade the method of ranking has been used with success. While reviewing, a peer uses the ranked list citation profile of each scientist where he can see the citations received by each individual paper as well as the summary of the citation profiles of all the papers. The peers to whom we had occasion to talk felt that the aforesaid items help them a lot in the objective evaluation of a scientist.

CONCLUSION

Our observations at the National Centre of Bibliometrics may be summarised thus:

(1) When the contributions of a scientist in terms of research papers is reasonably high, say, one hundred or more, his papers can be grouped in four distinct categories on the basis of citations they receive:

(i) a small number of papers receive a large number of citations. The phenomenon which we started observing since late 1980s has also been observed by Per Seglen in 1992 [8];

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

Table 2. *Ranking of scientists according to rank score*

Scientists	*No. of papers (no. of cited papers shown within brackets) P (p)*	*No. of citations (no. of self citations shown within brackets) c(s)*	*Rank score (in descending order) $(c\text{-}s)^2/p$*
YSH	12(8)	149(11)	2380.50
KCR	4(3)	87(3)	2352.00
DPD	10(9)	117(8)	1320.11
DDR	22(18)	98(4)	490.89
AKE	35(23)	134(15)	615.70
SGA	21(15)	86(8)	405.60
MLN	29(16)	79(8)	315.06
NVM	22(16)	70(14)	196.00
VGK	42(25)	79(11)	184.96
BCS	7(5)	35(7)	156.80
BKT	34(19)	55(4)	136.89
PKD	32(17)	47(8)	126.75
SKD	13(8)	45(4)	120.13
KNS	39(8)	33(2)	120.13
DCY	31(14)	49(10)	108.64
VDD	44(13)	43(6)	105.31
KSH	20(8)	29(10)	45.13
GRN	11(6)	18(2)	42.67
ASN	7(2)	10(1)	40.50
AKN	20(7)	20(4)	36.57

Table 3. *Rank of awardee scientists*

	Rank of awardee scientists in various subjects as per the formula						
Year of Award	*Math Sci*	*Phys Sci*	*Chem Sci*	*Earth Sci*	*Bio Sci*	*Med Sci*	*Engng Sci*
1991	2nd, 3rd	2nd, 4th	1st, 7th	1st, 8th	3rd, 15th	4th*	3rd*
1992	4th*	10th, 14th	4th, 9th	2nd*	10th, 17th	1st, 2nd	12th*
1993	6th, 7th	3rd, 9th	16th, 23rd	1st*	-	2nd*	1st, 3rd

*One award only

(ii) a larger number receive a smaller number of citations;
(iii) a substantial number receive a very small number of citations; and
(iv) the remainder receives no citations at all.

We have a feeling that the citation profiles of some of the scientists at least would follow a Bradford distribution [9]. We shall try to verify this in future.

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

(2) The citation profiles of research papers lead us to generalise the fact that the citation generation potential (CGP) is not the same for all papers, it differs from paper to paper, and also to a certain extent depends on the subject to which the papers belong. Research papers in mathematics usually generate a smaller number of citations compared to the papers of almost any other scientific field, whereas research papers in biochemistry or molecular biology usually generate a large number of citations. In other words, the mean citation score of papers in mathematics is usually found to be less than the papers in physics, chemistry, biology, medicine, and so on.

(3) The method of ranking in no way replaces peer review. It acts as an aid at the hands of the peers to help them arrive at a better decision.

ACKNOWLEDGEMENT

We are all grateful to Professor T. Viswanathan, Director, Insdoc, New Delhi, and the first author is also grateful to Professor Mashkuri Yaacob, Dean, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur for providing all facilities to complete the work. We are also grateful to the referees for their valuable suggestions which have helped simplify our mathematical proof of the vital equation and improve our paper.

REFERENCES

1. Sedgwick, W.T. and Tyler, H.W. *A short history of science.* New York: Macmillan, 1939.
2. Cronick, D.A. *A history of scientific and technical periodicals: the origins and development of the scientific and technological press from 1665–1790.* New York: Scarecrow Press, 1962.
3. Heniger, J. Leeuwenhoek, Antoni van. *Dictionary of Scientific Biography.* New York: Scribners, 1973, v.8, p.127.
4. *India – a Reference Annual* 1981, p.82.
5. Garfield, E. and Martin, M.V. 135th Meeting, American Association for the Advancement of Science, Dallas, 1968.
6. Sen, B.K. and Shailendra, K. Evaluation of recent scientific research output by bibliometric method. Paper presented at the International Conference on Science Indicators for Developing Countries, Paris, 15–19 October 1990. Also published in *Scientometrics*, 23(1), 1992, 31–46.
7. Martin, B.R. and Irvine, J. Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12, 1983, 61–90.
8. Seglen, Per O. The skewness of science. *Journal of the American Society for Information Science*, 43, 1992, 628–638.
9. Bradford, S.C. *Documentation.* London: Crosby Lockwood, 1949.

(*Revised version received 5 October 1998*)

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib