

## *Short Communication*

### Citation generation potential

B. K. Sen

Project Investigator- Indian National Science Academy,  
80 Shivalik Apartments, Alaknanda, New Delhi – 110019,  
Email: bksen1938@gmail.com

*Received: 14 November 2012, revised 21 May 2013*

The article traces the history of the first use of the term citation potential and citation generation potential (CGP) and provides their definitions. Discusses how the terms differ from one another in terms of meaning. Identifies the factors such as intrinsic quality, rank of a scientist, subject, document type and mega-authorship that are responsible for the CGP. Explains how the CGP of a paper, a journal or a scientist can be determined and used.

**Keywords:** Citations, Citation generation potential

#### **Introduction**

Citing documents is a universal phenomenon. Ever since the advent of Science Citation Index in 1963, citations have attained greater importance. Using citations, impact factor of journals, immediacy index, H index, etc., are being determined. In many cases citation counts are being used for selecting scientists at research institutions, awarding them fellowships, prizes, promotions, and so on.

It has been seen that not all documents receive equal number of citations. Some receive hundreds of citations and some others no citation at all. Some articles start receiving citations soon after its publication or even before its publication with such indications as 'In press', 'Communicated' etc. On the other hand there are articles that start receiving citations after the elapse of a few years. The question arises what are the factors that help generate citations? Here is an attempt to explain some of them

#### **The terms**

The term citation generation potential or CGP in short, was used by Sen et al<sup>1</sup> in 1998. There is no indication in Web that the term was used before by anyone else. Garfield<sup>2</sup> used the term citation potential way back in 1979 to mean the average number of citations per paper in a given field. Moed<sup>3</sup> defined citation potential 'as average number of cited references per paper in the subject field' It is to be

noted that Garfield used in his definition the expression 'citations per paper' and Moed used the expression 'cited references per paper'. Moed further says that citation potential indicates how frequently papers in a subject field cite other papers<sup>4</sup>. The term citation generation potential is quite different from citation potential. It means the capability of a document to generate citations. The mathematical explanation of the terms follows.

**Citation potential** – Let us suppose on a particular day 200 papers on supergravity are found in a database. The number of references appended to these papers range from 10 to 25 and total 3600. Hence, the citation potential of supergravity papers is  $3600/200 = 18$ . The citation potential will be considerably reduced in case the papers cited more than once is considered only a single paper. Moed takes only one to three-year old papers for his source normalized impact per paper (SNIP)

**Citation generation potential** – Suppose a paper on hadron published in 2008 has received 40 citations by 2012. The CGP of the paper is  $40/5 = 8$ . It takes a single paper into account.

#### **Factors responsible**

Several factors responsible for the generation of citations contribute to the citation generation potential (CGP). Some of the factors are discussed below.

Not all publications generate equal number of citations. By and large, papers in high impact journals generate larger number of citations. It is not necessary that all the papers of high impact journals will generate large number of citations. Even a few papers published in a very high impact journal may not receive any citations at all. It, therefore, becomes essential to know the factors that lead to the generation of citations.

#### **Intrinsic quality**

If we go through the papers that had generated large number of citations, then it will be seen that all the papers have some intrinsic qualities. Either they have reported some path-breaking discoveries, new methods of turning out a product, an outstanding process, a new explanation of an old phenomenon, or something similar.

Of course, controversial papers sometimes generate a good number of citations. Some people may dispute the claim made in the paper. Others may support it. This process leads to the generation of a number of citations.

Review papers may also generate a good number of citations since these papers are scanned by scientists to know what had already been done in their field of interest. Generally, they quote these papers in their research works leading to the generation of a large number of citations.

#### *Rank of a scientist*

Papers by scientists such as Nobel laureates, generally generate good number of citations. Scientists consider the area for which the Nobel Prize has been awarded fertile and start working on it and cite the papers by the Nobel laureate. Many a time a researcher cites the papers of a senior scientist, his/her research guide, and so on.

#### *Invisible college*

Scientists working on the same field the world over form some sort of a network. They exchange their ideas in the corridor meetings of the conferences, through emails, blogs, face books, and so on. Generally, they cite one another in their papers whenever the occasion arises.

#### *Subject*

CGP depends on the subject as well. A paper on a promising cancer drug can generate thousands of citations, while a paper on superalgebra cannot generate more than a few. The reason is simple. The more the number of workers, the more will be the number of papers as well as the citations. Obviously, the less number of workers will generate less number of papers and on average less number of citations.

#### *Document type*

CGP varies according to types of documents. Generally it is high for the research periodicals of international standard, compared to a thesis or conference document.

It can be explained with the concept of visibility. Compared to a research paper in a standard international journal accessible in Internet, the visibility of a thesis will be quite low as it is mostly unavailable in the Internet. Neither it is printed. Usually a copy of it is kept in the library among thousands of other theses and mostly escapes notice. Hence, a thesis will generally have less visibility and

its CGP will be low. Similar will the case of a conference paper. Some conference papers are never published. Conference proceedings publish the papers if received on time. Generally these proceedings are published keeping in view the number of participants, and distributed amongst the participants. Once the conference is over, the copies are generally become unavailable. As a result, their visibility becomes very poor leading to poor CGP. Moreover, scientists do not value conference papers very much thinking that any finding of value will appear in research periodicals.

#### *Mega-authorship*

Nowadays papers by more than 10 authors are but common. At times, the number of authors of a paper touches 50 or even more.

The phenomenon of self citation is universal and at times account for 10% of the total citations or even more. Suppose the paper P has been contributed by 50 authors. If all these authors contribute one paper each in near future and cite the paper P, then it will have 50 citations within a pretty short period.

#### **Methodology for determining CGP of a Document**

For determining the CGP of a paper, citations received by a paper within five years of its publication is to be taken into account. Then the total number of citations is to be divided by five. This will give the average number of citations per year, i.e. citation generation potential of the paper. Thus, the number of citations generated by a given paper per year is the unit of CGP, calculated on the basis of five-year cumulation.

Here a question may arise, why to take five years?

Suppose a paper has been published in Ukrainian language in December 2010 in printed form. The paper by and large will remain unknown to the world till its abstract appears in some international abstracting services, which are generally in English. The appearance of the abstract for the first time may take about a year or more. When the scientist cites the abstract in his own paper, it may take one more year for the paper to be published. Thus, two years will elapse before the citation appears in a publication. This period will be more for the non-Anglophone scientists to cite the paper. If one intends to get the full length article translated. It may take in some cases more than a year to get the translation. By the time a researcher gets the translation it is already three years. One more year will pass before the citation of the full length paper appears. Thus we see already four years

are gone. Papers that are published in highly less known journals are also cited very late. In this connection mention may be made of a study by Kessler and Heart<sup>5</sup> which has shown that if a paper is not cited at all within five years of its publication it is unlikely to be cited thereafter. This justifies the five year period for calculating the CGP. A very highly cited paper of an Indian scientist published in 1984 received its first citation in 1989. By now, the paper has received more than 1,000 citations<sup>6</sup>.

From 6<sup>th</sup> year onwards yearly citation rate may be taken into account to see whether the CGP is rising or declining.

#### **CGP Graph**

It should be drawn considering the citations received from the very first year. In some cases the number of citations may be zero. In such cases the graph will start from zero. Gradually it will rise. Then it may reach a plateau region. Thereafter it will start declining. In the course of years it may or may not reach zero. The book *Origin of Species* by Charles Darwin published in 1859 is cited still today. It cannot be predicted whether or not the CGP graph of this book will ever touch zero.

#### **CGP of a research periodical**

Here, if the journal is new, five-year period may be taken to determine the CGP of the journal. Dividing the number of citations received by the citable items during the preceding five years by five one will get the CGP. From the sixth year, citation rate per year will be the CGP. CGP for other periodicals or publications may also be determined in the same way.

In this case also the initial portion of the graph will be quite similar to the graph of a paper. Thereafter the graph will change depending on the ceasing, continuance, change of character, etc of the periodical.

#### **CGP of a scientist**

If the scientist is new in the field, say publishing papers for the last few years, determination of CGP should not be undertaken. It may be determined after a period of five years. In this case we are to take into account the total number of citations received by the papers of the scientist published during the past five years. The number is to be divided by five. From sixth year, citation rate per year will be the CGP.

#### **Uses of CGP**

It will be useful for employing a scientist. If the CGP graph of the scientist is rising, then the employer will get a clear indication about the potential of the scientist. Of course the employer will have to see the extent of contribution by the scientist in the two-authored or multi-authored publications.

It will also be of great help for the librarians in the selection of periodicals. If the CGP of a periodical is declining or pretty low, a librarian may decide to drop the periodical.

#### **Limitations of CGP**

CGP cannot be determined in quantitative terms soon after the publication of a research or review paper or any other publication. A minimum time is to be given for the generation of optimum number of citations.

If a scientist is having mostly two-authored or multi-authored publications, then the CGP may not reflect the genuine scenario.

#### **Conclusion**

In the realm of scientific research, citations have been used an indicator for various measures. In this article, attributes that can lead to generation of citations have been discussed and how the CGP can be determined is given. The evaluation of the method given will be carried out with actual examples and it is hoped the concept will pave way for others also to apply the method given using actual data.

#### **References**

- 1 Sen B K, Pandalai T A and Karanjai A, Ranking of scientists – a new approach, *Journal of Documentation*, 54 (5) (1998) 622–628.
- 2 Garfield E, *Citation Indexing. Its theory and application in science, technology and humanities*, (Wiley; New York) 1979.
- 3 Moed H F, Measuring contextual citation impact of scientific journals, *Journal of Informetrics*, 2010.doi:10: 1016j.joi 2010.01.002
- 4 Moed H F, A new journal citation impact measure that compensates for disparities in citation potential among research areas, *Annals of Library and Information Studies*, 57 (3) (2010) 271-277.
- 5 Kessler M M, Heart F F, Concerning the probability that a given paper will be cited. TP-TM-80S (MIT; Cambridge) 1962.
- 6 Kakkar P, Das B and Viswanathan P N, A modified spectrophotometric assay of superoxide-dismutase, *Indian Journal of Biochemistry and Biophysics*, 21 (2) (1984) 130-132.