Download counts – An early indicator for monitoring progress of science

Science rarely moves in an orderly manner. At the heart of all science there is the generation, dissemination and manipulation of data consistent with specific hypotheses, which may either be explicit or implicit. The progress of science is the idea that describes how science increases its problem-solving ability through the application of existing and new scientific method. The scientific method is a body of techniques for investigating phenomena and acquiring new (i.e. unknown) knowledge as well as for revising and integrating previously known knowledge. The progress of science would largely depend on acquiring and using new knowledge in systematic way.

As science is global, scientists from any corner of the globe can contribute to the progress of science. Scientific research is potentially open to a much wasteful duplication of human effort and credit for a new discovery goes entirely to the scientist who first announces it. This is the reason why the result of a research is diffused as quickly and widely as possible. Scientists write their findings as research papers to report and record their recent achievements and try to publish the same in high-impact journals to get enhanced visibility and recognition. Research papers constitute the primary scientific literature and represent the true and visible output of recent activities in every field of research. Scientific literature is cumulative in nature and every breakthrough discovery generally makes a huge body of existing literature of the field obsolete.

The use of citation analysis in monitoring the progress of science is not new. Garfield[1] was the pioneer in using citation frequency as a measure of research activity and performance. Small and Griffith[2] used it to identify the fields or specialties that make up the leading-edge scientific development. Fava and collaborators[3] showed how it can monitor the progress of research in clinical medicine. The discrete pieces of research work are mechanically identified by citation counts. High citations indicate that a researcher's work has been consulted and cited by many scientists, and thus had high research impact on the field. Conversely, the citation count of a particular paper does not necessarily reflect anything about its elegance, quality or relative importance to the progress of science.

In the print era, the publication and communication process was lengthy and time-consuming. An author submitted his research paper for publication in a journal of his choice. Then the editor of the journal sent the paper for peer-reviewing. After reviewing, the paper was either accepted/rejected or asked for conditional revision. The time that a paper took from the moment of acceptance until its publication might range anywhere from one month to one year or even longer. After publication of an issue of a journal, it took about three days to three months or even longer to reach other experts of the field depending on the distance of the subscribing institution from the place of publication and mode of delivery. Then a printed article was normally available to few select locations for consultation. Authors can cite a paper if they knew the paper and find it to be relevant to their own paper, and further, considered it to be important enough to cite explicitly. Again, the time that it took from the moment of consulted by experts cited by few of them and until those citing papers were published might range anywhere from one to two years or even longer.

In the Internet era, due to rapid growth of the web, the publication and communication process is much shortened. In fact, open access (OA) research papers are available for consultation as soon as the manuscripts are ready and deposited in OA archives. The use of OA papers can significantly accelerate the scientific communication process, but it lacks peer-reviewing. In most of the cases now, the researcher can consult a research paper through the web as soon as the same is accepted for publication such 'articles in press' enable access to the latest research. The number of downloads of a research paper is normally taken as an early indicator of its impact, because the digital version of a paper appears sooner and it gets additional time than its counterpart print version. Thus it is now possible to identify the impact of a research paper much earlier before it is published in a print journal.

The download counts also provide an early estimate of probable citation counts of research papers. Shin[4] showed that the impact of a journal increases when it transforms medium from paper to digital plus paper combination. Brody and collaborators[5] report that there is correlation between downloads and impact, particularly for high-impact papers. They are holding out the promise for the ability to predict high-impact papers much earlier than before. Zhao[6] compared the visibility of authors between the web (digital) and print journals. His study revealed that the author rankings by number of citations that resulted from *CiteSeer* data are highly correlated with that obtained from *Science Citation Index*.

The list of major developments in every field of research is the main component of the progress of science. There is no fundamental difference between download counts and citation counts in identifying such a list of major developments, but data of download counts are available much earlier than that of citation counts. In the changed scenario of web-publishing, the download counts can be taken as an early indicator (not waiting years for the citations to materialize) for monitoring and forecasting the progress of science. The use of this indicator can drastically reduce the time of research evaluation and can help review the ongoing research projects (if at all possible). Cockerill[7] rightly pointed out that citation tracking choices always keep scientists in the dark for a long time but download counts, in contrast, might enlighten them quickly.

1. Garfield, E., *Curr. Contents*, 31 January 1973, no. 5.
2. Small, H. G. and Griffith, B. C., *Sci. Stud.*, 1974, **4**, 17.
3. Fava, G. A., Guidi, J. and Sonino, N., *Psychother. Psychosom.*, 2004, **73**, 331.
4. Shin, E. J., *J. Inf. Sci.*, 2003, **29**, 527.
5. Brody, T., Harnad, S. and Carr, L., *J. Am. Soc. Inf. Sci. Technol.*, 2006, **57**, 1060.
6. Zhao, D., *Inf. Process. Manage.*, 2005, **41**, 1403.
7. Cockerill, M. J., *BMC Bioinf.*, 2004, **5**, 93.

HARI PRASAD SHARMA

Bengal Engineering and Science University, Shibpur,
Howrah 711 103, India
e-mail: sharma_hp@hotmail.com