



## Identifying the most important research—is there more to life than Impact Factors?

By Matthew Cockerill

Citation-metrics have played an important role in identifying the best research ever since the pioneering work of Eugene Garfield in the 1960s.

Whereas standard bibliographic indexing services allow users to search for relevant works by keyword, a citation-tracking service differs in that it also includes the citation list of articles, and identifies which articles cite which other articles. This allows the most highly cited work on a given topic to be easily identified, which is useful because a highly-cited piece of work tends to be an important one. For this reason, the founders of Google cite Eugene Garfield's work on bibliometrics as a key inspiration for their 'PageRank' algorithm, which highlights those web pages that are most linked to (i.e. cited) by other web pages.

The problem with looking at the number of citations of a given article to get an idea of its importance, however, is that because of the timescale of the publication process, it takes a year or two following publication for citations to appear. In the fast moving world of research, this is often inadequate. For this reason, the most heavily used metrics based on citation information are not article metrics, but journal metrics that identify which journals have, on average, the most highly cited articles. Such journal metrics have come to play a hugely important, but controversial, role in the evaluation of research.

### Developments in citation tracking services

Although the information provided by citation tracking services is immensely valuable, it has also traditionally been extremely expensive to compile. Manually entering the reference information for 50 citations per paper, for hundreds of thousands of papers, was a massive undertaking. As a result, for more than 30 years, the field of citation tracking has been dominated by the Institute for Scientific Information (ISI), founded by Eugene Garfield, and now owned by Thomson Corporation.

The Thomson-ISI "Impact Factor" journal citation metric has become a *de facto* standard, playing a critical role when authors select a publication outlet for their research. The perception (and sometimes the reality) is that evaluation by potential employers and funders will focus less on the research itself than on the quality of the journal in which it was published, and that Impact Factor ends up serving a proxy for the quality of the research [1].

As journals have moved online however, the situation has started to change. It is no longer necessary to manually key-in bibliographic citation data since in most cases publishers already have this data in digital form. This means it is now relatively simple to build a citation indexing service if you can persuade publishers to make their data available for indexing. As such indexing will tend to encourage citation of the work concerned, and will tend to bring more readers to the work, many publishers are willing to make their citation data available for indexing in this way.

Google has used this approach to create its free bibliographic service, Google Scholar, which incorporates citation tracking. While the search functionality offered by Google Scholar is basic compared to the Thomson-ISI Web of Science service, it is nevertheless impressive for a free service. Several studies have found that Google Scholar's coverage is comparable to Thomson-ISI's in many areas [2]. The data in Google Scholar is not entered by human indexers, but is entirely harvested by computers from online journal web pages and data feeds. Google Scholar does not yet generate journal-metrics based on the data that it has compiled, but such metrics, analogous to the PageRank that Google calculates for web sites, would be of great interest if created.

Scopus, meanwhile, is a commercial service launched by Elsevier in late 2004 as an alternative to Web of Science, offering welcome competition for Thomson-ISI, which should help stimulate innovation. A useful example of such innovation is the Author ID feature that has recently been introduced by Scopus. The Author ID system aims to disambiguate multiple authors who have the same name, using sophisticated algorithms. This goes some way towards solving a problem that will be familiar to anyone who has ever tried to find all papers published by a particular author named Smith (or Yamamoto).

Online journal systems such as HighWire, Science Direct, PubMed Central and BioMed Central also offer basic facilities for navigating forwards and backwards through a web of citations, within the journals which are hosted within that particular system, and in some cases to other systems too, thanks to the CrossRef system.

CrossRef is a collaborative initiative between publishers, which focused initially on ensuring that the reference list of an article on one publisher's site could include links to the full text of the cited articles, even when those articles

>>>

## >>> Identifying the most important research

appeared on another publisher's site. Recently CrossRef has also encouraged publishers to deposit full reference list data. This then allows the cross-publisher linking system to be extended to 'forward linking' from earlier articles to the subsequent articles which have gone on to cite them.

Lastly, several non-commercial services such as CiteSeer and Citebase make use of material that is openly accessible on the web to provide a citation tracking service.

### Alternatives to citation tracking services

For all the importance of citation tracking, however, it is not a full solution to the problem of identifying the most important papers.

As already mentioned, article-level citation information takes time to emerge, while journal-level citation information can misleadingly imply that all research in a given journal is of equal interest. There is therefore a need for other, better ways to highlight the most important articles as soon as possible following publication.

Online journals naturally suggest one obvious metric that can be used—the number of downloads that an article receives. Other things being equal, the more downloads an article receives, the more likely it is to be of broad interest. Many publishers, including BioMed Central, regularly make available 'most viewed' lists, to highlight hot articles.

BioMed Central has taken this approach one step further by introducing a 'highly accessed' logo which appears on any BioMed Central article that has received an unusually large number of accesses, considering its age and the journal in which it was published. Roughly 10% of articles receive such a 'highly accessed' logo, which remains permanently associated with the article so that the author can refer to it on their résumé.

Although download metrics are popular and useful, they should be treated with caution as they can be vulnerable to manipulation. BioMed Central devotes significant effort to ensure that robot accesses and other suspicious patterns of activity do not distort these measures.

While citation and download metrics attempt to provide an objective measure of an article's importance, subjective measures can also play just as important a role in highlighting key articles. Traditionally, the main way in which subjective evaluations of research publications would be conveyed would be through review literature, ranging from broad annual surveys of a field, to tightly focused mini-reviews. The problem with review articles, however, is that they generally appear months after the research concerned has been published, whereas readers really want to know about the very latest research.

Faculty of 1000 is an opinion-driven literature awareness service which attempts to solve this problem. The "Faculty Members" do not write full-blown reviews. Instead, they regularly contribute short, structured 'evaluations' of

recent research articles that have caught their attention. The Faculty of 1000 website automatically aggregates many such evaluations, to create a literature awareness service which can rapidly identify important articles. As well as providing a numeric rating (the F1000 Factor), the system lets users see which faculty members have rated the article as of special interest, and why. A "Hidden Jewels" section of the website calls attention to articles which received high ratings from Faculty Members, despite having been published in relatively obscure journals.

### Researchers doing it for themselves?

Faculty of 1000 depends on a set of experts within each field, whose reputation adds authority to their evaluations.

Other, less regulated approaches are possible however. Recently there has been an explosion of activity in areas such as "social networking", and "user-generated content". Wikipedia is just one example that shows how a surprisingly high quality end result can be produced by taking advantage of network effects, and opening up systems to anyone motivated to contribute.

CiteULike and Connotea are two very similar sites which take this approach. Each site allows users to conveniently bookmark their favourite scholarly articles, and to "tag" those articles according to topic. In this way, these systems simply provide a convenient bibliographic record-keeping system, similar to software like EndNote. What is special about these sites, however, is that by default, the information about which user has tagged which article is shared with other users of the site. This creates a rich web of information that can be used to assist in navigating the literature. For example, users can see which articles in a given field have been tagged by the most users. They can also see which topics are hot by



Figure 1  
CiteULike's popular tags

## Identifying the most important research

reviewing the list of most highly-used tags (Figure 1). If you find that other users seem to have similar interests to you, you can even add them to a “watch list” so that if they tag an article, your attention is automatically called to it. These systems are still in their infancy, but they show a great deal of promise. By taking advantage of so-called “semantic web” technology, it may even be possible that one day, systems such as CiteULike might act not only as a personal filing system, but could also evolve to be a personal knowledge management system, with the ability to distil facts from many different papers and relate them to one another, seeking patterns of correlation or contradiction.

More immediate benefits are promised by approaches based on another modern web phenomenon – blogging. Postgenomic and Mixed States are two sites which serve the needs of biologists and physicists respectively. They each collate data from dozens of scientific web logs, and use the assembled data to identify research articles attracting the attention of scientifically inclined bloggers. The relevance and comprehensiveness of the information delivered by these sites is clearly dependent on the quality and quantity of the blogs on which they are based. However, with more and more scientists blogging, the usefulness of such services, especially for casual browsing, is only likely to increase.

### Matthew Cockerill

BioMed Central, London, UK  
matt@biomedcentral.com  
www.biomedcentral.com

### References

1. Seglen Per O. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997;314:498–502.
2. Pauly D, Stergiou KI. Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google's Scholar service. *Ethics in Science and Environmental Politics* 2005; 33-35  
<http://www.int-res.com/articles/esep/2005/E65.pdf>

### Websites

- <http://scientific.thomson.com/products/wos/>
- <http://www.scopus.com>
- <http://www.citeser.org>
- <http://www.citebase.org>
- <http://www.crossref.org>
- <http://www.facultyof1000.com>
- <http://www.connotea.org>
- <http://www.citeulike.org>
- <http://www.postgenomic.com>
- <http://mixedstates.somethingsimilar.com>

## Special TWS subscription rates for MET members

The European Medical Writers Association (EMWA) has extended to The Mediterranean Editors and Translators (MET) members a special subscription rate of €30 for a year's subscription (4 issues) to *The Write Stuff*. To subscribe and for further information, please contact EMWA at [info@emwa.org](mailto:info@emwa.org). You will need to provide your MET membership number.

## Upcoming events of interest

### International Communication—Promising Practices (METM 06)

27<sup>th</sup> – 28<sup>th</sup> October 2006, Barcelona, Spain

The 2nd meeting of the recently formed association Mediterranean Editors and Translators (METM) will include a mix of panel discussions, presentations, training workshops and plenary discussions targeted at editors, translators, linguists and oral communication coaches who work in the Euro-Mediterranean area.

Go to <http://www.metmeetings.org/pagines/metm06.htm> for more information.

### EMWA Autumn conference

16<sup>th</sup> - 18<sup>th</sup> November 2006, Brussels, Belgium

Professional training on a broad spectrum of topics for medical communicators.

Go to <http://www.emwa.org/ConfAut06/Introduction.html> for more information.

### Practical Solutions for Filing Variations 2006 (IIR)

29<sup>th</sup> - 30<sup>th</sup> November 2006, Hilton London Green Park, London,

Gain an in-depth understanding of the most up-to-date and successful regulatory strategies for filing compliant variations.

Go to [www.iir-events.com/IIR-conf/PTI/Default.aspx?EventSector=49](http://www.iir-events.com/IIR-conf/PTI/Default.aspx?EventSector=49) for more information.

### Compiling eCTD Variations (a post-conference workshop to the above)

1<sup>st</sup> December 2006, Hilton London Green Park, London

SAVE! EMWA members receive up to £300 discount by booking before the 20th of September 2006. Quote VIP code: CQ5030ELTWS.

### Registration of Pharmaceuticals in the EU 2006 (IIR)

11<sup>th</sup> - 14<sup>th</sup> December 2006, Corinthia Towers Hotel, Prague, Czech Republic

A practical course addressing current requirements for Marketing Authorisation applications in the modular CTD format. It will also review electronic submissions (eCTD).

SAVE! EMWA members receive a 10% discount by quoting: CQ5017EMWA.

Go to <http://www.informa-ls.com/registrationeu/> for more information.