Trying to quantify scientific success

Pitfalls of Bibliometrics

KAROL ŻYCZKOWSKI
Department of Physics, Astronomy and Applied Informatics, Jagiellonian University
Center for Theoretical Physics, Polish Academy of Sciences
karol@tatry.if.uj.edu.pl
Prof. Karol Życzkowski is a physicist, co-author of the square-root voting system for the Council of the European Union, a skiing guide to the Tatra Mountains and a monograph on "Geometry of Quantum States"

AGNIESZKA POLLO
Department of Physics, Astronomy and Applied Informatics, Jagiellonian University
Center for Theoretical Physics, Polish Academy of Sciences
Academia magazine, Polish Academy of Sciences
apollo@fuw.edu.pl
Dr. Agnieszka Pollo is an observational cosmologist, studying the clustering and evolution of galaxies

Identifying the best scientists among the mediocre is straightforward when given the perspective of centuries, or at least decades, of hindsight. The problem of evaluating contemporary researchers is far more complex

Scientific practice almost by its very definition involves wandering through the unknown. How can we tell whose research will end up finding useful applications, and who is just wasting both time and money?

In search of sponsors

There's no getting away from it - the most crucial issue is of course obtaining research funding. Typical scientists, both in the past and today, are not eccentric millionaires ready to spend a fortune on their experiments, or favorites of wealthy rulers willing to finance any kind of research. These days, most scientists are employed by numerous research institutes and universities, financed either from public funds or by various foundations. Governments and private sponsors have a stake in financing science; without having to fully understand the discipline, they need to know in advance how much the researchers they invest in are "worth". On the other hand, since scientists compete with one another for various types of research funding, they certainly want to know what the assessment criteria are in order to improve their chances.

The last century has seen a dramatic shift in the number of people working in scientific research, plus surging numbers of scientific disciplines, published results, journals and publications. As such, no scientist is able to keep up-to-date with all current issues, even within just a single discipline. Since not even the opinions of experts are reliable, how can scientific research and its worth be evaluated objectively?

There are no perfect solutions to this problem. The most popular method currently in use involves assessing bibliometric data, although all methods continue to arouse controversy.

Quantifying research value

The fundamental task of a scientist is to conduct research; their results are then published in a specialist journal in order to make them available to wider audiences. But how can the significance of such a publication be assessed by the scientific community? Simplifying things greatly, one might assume that the more frequently a publication is cited, the greater its impact. Computerized databases can easily track how many papers a given scientist has published, and how frequently his or her works have been cited in subsequent publications by other authors. Such data serves as a basis for various "scientometric" indicators intended to provide a simple and objective tool for assessing the quality of a scientific journal.

Many experts stress that the significance of each single citation is not really so clear-cut, and therefore statistics based on such citations are not as objective as is claimed by their advocates.

Agnieszka Pollo

Anyone who follows specialist literature can easily note the steady stream of new journals being established in many scientific disciplines. This increases competition, while the growing ranks of researchers keep filling the increasingly thick volumes with articles of varying quality. While editors strive to maintain the high quality of their journals, publishers rarely consider the complex details of the merits of any given publication and generally only pay attention to its *Impact Factor (IF)*. In order to increase a journal's market value, publishers frequently put pressure on the editorial team to increase the *IF* rather than improve the journal's quality. Although researchers frequently do not agree with this approach, the reality is that these days few can afford the luxury of completely ignoring citations-counting and scientometric tools.

And so, since bibliometric data is almost a necessity, it should at least be used responsibly, by applying the various indicators for the purpose they were designed to serve.

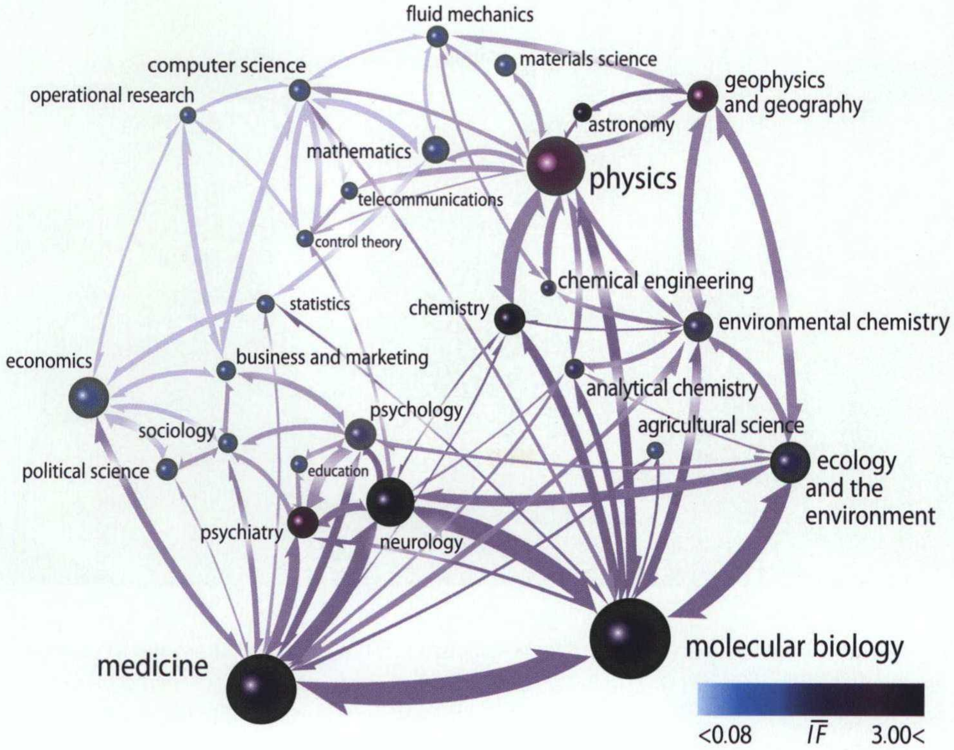
Impact factor and the *h*-index

Eugene Garfield designed the notion of the impact factor (*IF*) as a quantitative description of the impact of a given scientific journal. It describes the ratio of the number of citations obtained by papers published in a given journal in the two preceding years to the total number of articles published that year in the journal. This number takes into account the effect over a short period, since the *IF* is only affected by citations appearing in literature up to two years following the publication of the original article. Garfield's *IF* was originally intended to be used for medical and natural sciences, and is well adapted to their specificities. However, journal publishers, librarians, and administrators assessing scientific research apply it to many other disciplines, such as mathematics, where it would actually make more sense to count citations going back 5 or even 10 years after publication, due to a slower circulation of results in this discipline. It should be stressed that the *IF* does not take into account the number of authors, self-citations or the specificity of each discipline, and its value is easy to manipulate. It should also be remembered that the *IF* was devised to assess scientific journals, and as such cannot be used to evaluate the quality of individual articles published in a given journal.

The *h*-index, proposed by Jorge Hirsch in 2005 based on the distribution of citations received by papers published by a given researcher, is a very different quantity. A scientist has an index *h* if he or she has published h papers, each of which has been cited in other papers at least h times. As such it reflects both the number of publications, and the number of citations of these articles.

The number of scientific publications has grown rapidly during the last century

Trying to quantify scientific success



A graph of citations between different scientific disciplines, and mean *IF* value for the given discipline; circle size corresponds to the number of journals indexed in each field

Paweł Adamów after Althouse, West, Bergstrom & Bergstrom (2009)

Biology vs. mathematics

A common methodological mistake is using bare bibliometric indicators to compare researchers working in different scientific disciplines, even though the values of the indicators depend entirely on the scientific discipline. As such, direct comparison of bibliometric index values for mathematics and biology is inappropriate, as is stating that a given medical journal is better than a geological one due to its higher *IF*.

In order to correctly interpret such numerical values, it is necessary to understand the working habits within given scientific communities. For example, although computer scientists produce a high volume of papers, they are frequently put off by the slow publication procedures of specialist journals, preferring instead to publish in regular conference proceedings, some of which are not indexed in databases. The situation is even more complicated in the humanities, whose scholars frequently publish valuable papers in their native language in niche journals. In order to conduct any meaningful comparisons between the indices (describing journals, researchers or research institutions), one has to use relative numbers - looking at the values of the indices in relation to the mean in a given discipline.

Bibliometrics in practice

The imperfections of the bibliometric tools listed here have motivated attempts to devise new, more precise tools that take many additional factors into account. However, it is unlikely that a single, universal numerical index for reliable assessment of scientific research will ever be created. Therefore one can use in parallel several indicators as auxiliary data supporting the peer review process. All indices should be applied reasonably, for the purpose they were designed to serve. For example, *IF* - devised for assessing journals - must not be used for evaluating the quality of a given article or author. Likewise, the *h*-index, created for comparing the achievements of different researchers within a single discipline, should not be used for comparing the quality of different journals or research produced by various institutes, since it is highly dependent on the size of the institution. Applying the *h*-index to comparing a post-doctoral researcher with a professor with many years experience is equally misleading, since the index value increases with the author's age and experience.

Use wisely!

In a world dominated by decision-makers demanding hard numbers, what should researchers do? Most of all, keep a level head. They should be diligent in their research, write good papers, and publish them in respected journals. They should cite all publications appropriately, following the conventions in their discipline, while trying not to worry about the values of their bibliometric indices: a good researcher's work will usually come to be reflected in high-quality parameters, regardless of the index used. Finally, they should not let themselves be dragged into shallow games of artificially inflating the values of specific indices describing their work, which is simply a waste of time and energy.

Reviewers wield a lot of power as the current system stands, therefore they should be expected to use their broad knowledge of a subject reasonably when evaluating applications for research funding or prizes. Bibliometric data should only be used to support other information, and cannot replace peer review. If it is absolutely necessary to evaluate a given paper's impact on the scientific community, its total citations should be used, rather than the *IF* of the journal where it was published. For more recent publications, an author's impact factor can be used, defined as the mean number of citations of the author's papers dating back between 3-5 years, within two years after their publication.

Bibliometrics and funding

The greatest responsibility still rests with those responsible for managing research. Scientific research is a multidimensional process; therefore sensibly describing it in terms of a single factor is impossible, as is expecting to develop a single perfect bibliometric index. It is better to use several such indicators in parallel, together with a healthy dose of common sense. When comparing the values of indicators from different scientific disciplines, we should also remember to scale them against the mean values in each field and the time interval.

Also worth supporting is the idea of using quantified indices in which the scientists being evaluated have a say in how their work gets assessed. One example is the application procedure for grants from the European Research Council, where every applicant selects his or her 10 most important publications from a given time period and states how many times each one has been cited.

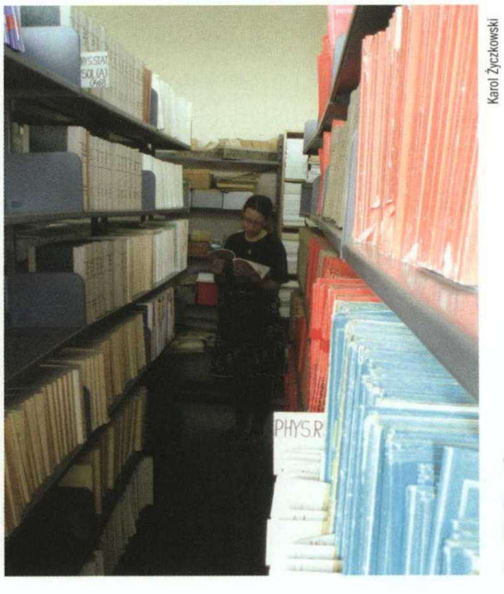
The current application process for a grant from Poland's a National Science Center requires the submission of at least 5 publications. This absence of any maximum number for the most important publications (conference papers, grants) to be considered seems to be an oversight: requesting a list of the best 10 publications allows experts to form an opinion on the author's credentials, whereas setting no upper limit encourages applicants to needlessly create more work for themselves and the reviewers.

This article can be concluded with an appeal to everyone involved in the process of evaluating and financing research: do not rely on scientometric indicators alone, but use your common sense as well! ■

Further reading:

Hirsch J. E. (2005). An index to quantify an individual's scientific research output. *PNAS, 102*, 16569-16572.

Życzkowski K. (2010). Citation graph, weighted impact factors and performance indices. *Scientometrics, 85*, 301-315.


Karol Życzkowski

The *impact factor (IF)* of a given journal is not an indication of the quality of articles published in it