Forget citations for unbiased research evaluation

By Emre Sarigöl, René Pfitzner, Ingo Scholtes, Antonios Garas and Frank Schweitzer

Researchers at the Chair of Systems Design, ETH Zurich, Switzerland

Published on **EuroScientist**: www.euroscientist.com

Subheading: A recent study of the social dimension of science publishing shows evaluating research quality using scholarly publications citations is not free of bias

Quantitative measures are increasingly used in scientific performance evaluation—be it for research institutions, departments, and individual scientists. Measures like the absolute or relative number of published research articles are frequently applied to quantify the *productivity* of scientists. To measure the *impact* of research, citation-based measures like the total number of citations, the number of citations per published article or the *h*-index, have been proposed. Proponents of such citation-based measures or rankings argue that they allow to quantitatively and objectively asses the *quality* of research. They have thus been encouraging their use as simple proxies for the *success* of scientists, institutions or even whole research fields.

There is an intriguing idea that, by means of citation metrics, the task of assessing research quality can be "outsourced" to the *collective intelligence* of the scientific community. This has resulted in citation-based measures becoming increasingly popular among research administrations and governmental decision makers. Such measures are thus used as one criterion in the evaluation of grant proposals and research institutes or in hiring committees for faculty positions.

Read this post online: http://www.euroscientist.com/forget-citations-for-unbiased-research-evaluation/

EuroScience | 1, Quai Lezay-Marnésia | F-67000 Strasbourg | France
Tel +33 3 8824 1150 | Fax +33 3 8824 7556 | office@euroscience.org | **www.euroscience.org**

Social factors influence

Considering the potential impact for the careers of – especially young – scientists, it is reasonable to take a step back and ask a simple question: To what extent do *social factors* influence the number of citations of their articles?

Arguably, this question challenges the perception of science as a systematic pursuit for objective truth. Ideally, it should be free of personal beliefs, biases or social influence. On the other hand, quoting Werner Heisenberg, "*science is done by humans.*" It would therefore be surprising if specifically scientific activities were free from the influences of social aspects.

Often the term "social influence" has a negative connotation. However, we do not think that social influence in science necessarily stems from malicious or unethical behaviour, such as e.g. nepotism, prejudicial judgments, discrimination or in-group favouritism.

The truth is that scientists operate in the context of increasing amount of published research articles. And they have a limited ability to keep track of potentially relevant works. Hence, we suspect that a growing importance of social factors in citation behaviour is due to natural mechanisms of *social cognition* and *social information filtering*

Scholarly citation behaviour

In a study recently published in EPJ Data Science, we address this issue by studying the influence of social structures on scholarly citation behaviour. Using a data set comprising more than 100,000 scholarly publications by more than 160,000 authors, we extract time-evolving co-authorship networks. And we use them as a simple proxy for the evolving social (collaboration) network of the scientific discipline *computer science*.

Based on the assumption that the notion of centrality , or importance, of scientists in the co-authorship network is indicative for the *visibility* of their work. We then study to what extent the "*success*" of research articles in terms of citations can be predicted. To do so, we only used knowledge about the embedding of authors in the co-authorship network *at the time of publication*.

Our prediction method is based on a classification method, called random forest, and utilises a set of complementary network centrality measures. We find strong evidence for our hypothesis that authors whose papers are highly cited in the future have – on average – a significantly higher centrality in the co-authorship network at the time of publication.

Remarkably, we are able to predict whether an article will belong to the 10% most cited articles with a precision of 60%. We argue that this result quantifies the existence of a *social bias*, manifesting itself in terms of visibility and attention, and influencing measurable citation "success" of researchers. The presence of such a social bias threatens the interpretation of citations as *objectively awarded esteem*, which is the justification for using citation-based measures as universal proxies of *quality* and *success*.

Considering our finding of strong statistical dependence between social centrality and citation success, one could provocatively state the following: if citation-based measures were to be good proxies for scientific success, so should be measures of centrality in the social network. We assume that not many researchers would approve having the quality of their work be evaluated by means of such measures.

We would like to emphasise that by this we *do not* want to join in the line of – sometimes remarkably uncritical – proponents of citation-based evaluation techniques. Instead, we hope to contribute to the discussion about the manifold influencing factors of citation measures and their explanatory power concerning scientific success. Especially, we *do not* see our contribution in the development of automated success prediction techniques. Should they be widely adopted, they could possibly have devastating effects on the general scientific culture and attitude.

Highlighting social influence mechanisms, we rather think that our findings are an important contribution to the ongoing debate about the meaningfulness and use of citation-based measures. We further hope that our work contributes to a better understanding of the multi-faceted, complex nature of citations and citation dynamics, which should be a prerequisite for any reasonable application of citation-based measures.

Read this post online: http://www.euroscientist.com/forget-citations-for-unbiased-research-evaluation/

EuroScience | 1, Quai Lezay-Marnésia | F-67000 Strasbourg | France
Tel +33 3 8824 1150 | Fax +33 3 8824 7556 | office@euroscience.org | **www.euroscience.org**

This article is adapted with permission from a study called '**Predicting scientific success based on coauthorship networks**,' which was published in open access as part of a thematic series entitled: **Scientific Networks and Success in Science**.

Photo credit: jairoagua

Read this post online: http://www.euroscientist.com/forget-citations-for-unbiased-research-evaluation/

EuroScience | 1, Quai Lezay-Marnésia | F-67000 Strasbourg | France
Tel +33 3 8824 1150 | Fax +33 3 8824 7556 | office@euroscience.org | **www.euroscience.org**