EDITORIAL **Open Access**

Scientific networks and success in science

Frank Schweitzer[*]

[*]Correspondence:
fschweitzer@ethz.ch
Chair of Systems Design, ETH
Zürich, Weinbergstrasse 58, Zürich,
8092, Switzerland

1 Editorial

There is not only an emerging new science about data, there is also new data about science, in particular about scientific activities related to publications, collaborations, and citations. EPJ Data Science, with its focus on the digital traces generated in techno-socio-economic systems, aims to provide a platform for this research on science, to facilitate the *science of science*. We are interested in this topic not simply because we want to analyse new data sets, but also because, as scientists, we constitute the subject of this research and we are affected by the conclusions derived from it. In fact, quantitative measures of scientific output and success in science already impact the evaluation of researchers and the funding of proposals, hence the future of science and *our* future in science. Thus, it is appropriate to ask whether such quantitative measures convey the right information and what insights might be missing. This regards in particular the role of social networks in the promotion of scientists and/or scientific ideas. On a higher level, these new data sets also provide a closer look into the *evolution* of science *per se*. Analysing large-scale collaboration and citation networks allows to better understand the (de)fragmentation of science along classical borders, how new research topics emerge, and how new ideas and methods spread across disciplines.

In our thematic series "*Scientific networks and success in science*", we start with a first batch of three works that highlight how a large-scale analysis of bibliographic data can help us to better understand the complex social processes in science.

In their paper "Inequality and cumulative advantage in science careers: a case study of high-impact journals" Alexander Petersen and Orion Penner address important questions related to the social mechanism in science. Using longitudinal data that covers publications in high-impact journals between 1970 and 2005, they measure the evolution of inequality in terms of the distribution of "scientific success". They argue that the distributions found are consistent with a strong cumulative advantage by which the initial success of individuals is amplified. Quantitative evidence for this feedback effect is provided in terms of continuously decreasing waiting times between consecutive publications in the highest impact journals.

Addressing an issue of significant practical relevance when studying bibliographic data sets, in their paper "Exploiting citation networks for large-scale author name disambiguation", Christian Schulz and co-authors show how the name disambiguation problem can be addressed by means of an analysis of citation networks. They develop an algorithm which is based on the observation that two papers written by the same author are expected to be more similar in the citation network, than two papers written by different authors which

happen to have the same name. They demonstrate the power of their method using data from Web of Science, a data set which has previously been shown to be prone to name ambiguation issues. As such, this work should be of significant interest to anyone analysing bibliographic data in an attempt to say something about individual scientists.

Studying the influence of social structures on citation-based measures, in their paper "Predicting scientific success based on co-authorship networks" Emre Sarigöl and co-authors address the question whether quantitative, citation-based measures of scientific impact should actually be seen as "objective". Combining supervised machine learning and network analysis techniques in an innovative way, they show that the position of scientists in the collaboration network alone is - to a surprisingly large degree - indicative for the future citation success of their papers. Clearly, the results of this study should make us think twice whenever we are tempted to use citation-based measures to quantify scientific impact.

The three papers in this series are excellent demonstrations of how data science can contribute to the *science of science.* They not only provide us with new methods to study large-scale bibliographic data, they also show how we can use these methods to gain new insights into the complex social processes at work in science. Certainly, one may argue that the presence of cumulative advantage mechanisms or socially-biased citations is *per se* not surprising. However, leveraging on large data sets and state-of-the-art statistical analysis techniques, the authors have shown that we are now able to *validate* and *quantify* these phenomena, empirically. The ability to test and invalidate hypotheses about the social processes at work in the academic community opens broad perspectives for the science of science. Furthermore, the detected patterns of social influence should also raise significant attention of stakeholders such as funding agencies, editorial boards or academic institutions. Hence, there are good reasons to continue with this thematic series on *Scientific networks and success in science.*