Philippe Hünenberger
ETH Zürich

Bibliometry: The wizard of O's

Imagine there is once a shortage of building space in Switzerland. This affects everyone, individuals and businesses, including restaurants. As the space must be reserved for the best places only, the authorities need an *objective* criterion to make an *optimal* choice, and an *obvious* definition for the quality of a restaurant is the frequency F at which the average citizen eats there. *Optimal*? *Objective*? *Obvious*? The three O's – I'll come back to that ... Anyway, F-factors are evaluated to five significant digits and analyzed using the most modern computer programs. Poorly performing places are closed and systematically replaced by better ones. Owing to this selection process, the cooks themselves end up considering the F-factor as the ultimate measure of professional success in the branch. And after a few years of this policy, one realizes with great surprise that the gastronomic landscape of the country has been reduced to cafeterias and fast foods. Indeed, the average citizen eats there more often than at four-star restaurants. Far-fetched? Well, I sometimes have the feeling that with the use and abuse of bibliometric impact factors to monitor academic research, we are letting the wizard of O's take us precisely down this path.

Optimal?

First "O" to question: In times of shortage, is *optimization* really the sole or even a good strategy and if yes, at which spatial (group) and temporal (planning) scales? Scientific research is a collective endeavor, and the best teams seldom consist of clones of an optimal individual. In addition, optimizing for short-term return is not the same as planning for sustainability and long-term effectiveness. Nature did very well 65 million years ago to have set aside some mammalians in case of, although they were definitely less performing than the dinosaurs at the time. In one word, diversifying is as important a strategy as optimizing. Thinkers, inspirers, nurturers and logisticians are as needed in scientific teams as pure individualistic communicators, recruited on the sole basis of their personal publication metrics. You do not make a winning soccer team with eleven top goal-scorers. And you do not make a successful "fellowship of the ring" with nine copies of Aragorn.

Another problem of optimization at all costs is that it is not compatible with risk-taking. As a rule of thumb, if you want a percentage P of true discoveries in research (and of orthogonal thinkers in scientific teams), you need to also accept a percentage P of unsuccessful efforts (and of poor scientists), the remaining being average incremental or fashionable research. By trying to optimize the percentage P towards zero in a no-risk strategy, one merely ensures that 100% of the research will be mediocre, while claiming very loud that it is top-level. Is this really what we want?

Objective?

Second "O" to question: Can scientific quality actually be measured by an *objective* criterion? As a scientist, I have the greatest respect for objective (reproducible) data. We are a theoretical-chemistry group, so we actually produce terabytes of it on a weekly basis. But this data alone does not make us any smarter. The real scientific talent is in the questions we formulate, in the design of clever experiments to address them, and in the analysis and interpretation of the results to formulate insightful answers. None of this is objectively measurable and, actually, none of this has anything to do with bibliometry whatsoever.

There is another interesting parallel between my work and bibliometric assessment. The interpretation of raw scientific results often relies on the reduction of very high-dimensionality problems (our terabytes of data) to one-dimensional indicators (a handful of functions shown in the figures of a scientific article). These projections must be selected carefully and are meant to facilitate an understanding of the process, given the limited capabilities of the human brain and language. However, both the selection and the interpretation of these indicators, two highly subjective processes, still require a deep knowledge of the mechanics and chemistry of the system. Treating these indicators as pure black-box outputs can be extremely misleading. The same holds for bibliometric indices. Although they represent some objective one-dimensional projection of the academic research process, their interpretation makes no sense and their use in decision-making is very dangerous for anyone who is blind to the underlying complexity. For this reason, it is essential that science managers keep in mind how scientific research works in practice, not on a flowchart but on the ground.

Let's return to the restaurant analogy. How would a gastronomic guide proceed to evaluate quality? They would send small teams of experts to taste the food, a procedure akin to peer reviewing. This procedure is tedious, time-consuming, expensive, demanding in terms of personal competence, and partly subjective. And (yes!) it does involve an emotional component. Yet, in many ways, it is far better to reason objectively based on subjective expert assessments, than to reason subjectively (without being aware of it) based on objective but irrelevant one-dimensional indicators. Ultimately, the probing instrument for a complex high-dimensional process must itself be complex and high-

dimensional. When I hear an exciting scientific talk or I read a high-quality scientific article, I know it is good simply because I feel thrilled and inspired. And more often than the opposite, my colleagues feel just the same. But for some reason, although my computer can spit out a wealth of irrelevant bibliometric data about the author, it stubbornly refuses to share my enthusiasm.

Obvious?

Third "O" to question: Is there really an *obvious* relationship between scientific quality and bibliometric indicators? There are two aspects to this question: What do we consider to be quality in science and to which extent do bibliometric indicators characterize this quality? Already the first question is difficult, and there is a wide spectrum of opinions, from the most idealistic to the most utilitarian. Ultimately, we do science because it is in our genes of *Homo Sapiens Sapiens*: The urge to understand how the world functions and to apply this understanding for adjusting the world to our needs. So, maybe we can agree that scientific quality is related to the successful *acquisition* or *application* of new knowledge. This is already a two-dimensional space, *i.e.* beyond the realm of one-dimensional functions. And actually, bibliometric indicators do not even belong to this space as they exclusively focus on the *transmission* of knowledge, *i.e.* they are at best indirectly influenced by its acquisition and application. As a result, they probe scientific quality in a direct way neither from the idealistic nor from the utilitarian perspective.

Nowadays, the basic bibliometric currency unit (BCU) is one citation of one of your articles in an article of a peer scientist. And the basic assumption chain is something like: (1) your scientific quality is proportional to your number of quality papers; (2) the quality of a paper is proportional to the number of peer scientists who find it good; (3) the extent to which a peer scientist finds your paper good is proportional to the number of BCUs she/he gives you; (4) each time a peer scientist gives one or more BCUs to one of your papers, it means she/he has read it (I mean, past the title) and considered good; (5) all peer scientists have an equal probability to have seen any of your papers, before deciding whether they would give you a BCU or not. The first statement can arguably serve as a definition, with the already questionable corollary that a scientist who does not publish at all is automatically to be regarded as a bad scientist. For example, according to such a narrow definition and because he did not leave any writing of his own, Socrates would rank as an appalling philosopher. With the possible exception of this first one, no single statement in the list above is correct. Just check once in details where you (or a colleague) collect your BCUs. This is a sobering up experiment! And since there are many ways to generate BCUs artificially, I am wondering when it will become possible to buy them on the internet and what will be the resulting parity to the dollar. I will not detail the specific shortcomings of a given measure (*e.g.* *h*-index). All scientists who know how things work in practice can give you many examples of their shortcomings. My objection is not technical (how could we improve the index), it is fundamental: No numerical index whatsoever can measure scientific quality!

A self-reinforcing system

One of the main problems of the bibliometry-based evaluation system is that it is self-reinforcing. There is a well-known effect in sociology (self-categorization theory) called the social proof. In this particular instance, it states that if a certain *F*-factor, which may well be largely irrelevant, becomes the main criterion for accessing a given social elite (*e.g.* researcher position at a university, awardee of a prestigious grant), researchers, who are also humans after all, will spontaneously tend to first passively accept (compliance) and then actively believe (internalization) that the *F*-factor *is* the real measure of their fitness for this elite, *i.e.* of their true talent. As an older generation of scientists (those who also knew the pre-bibliometric times) gives way to a newer generation (those who obtained their positions thanks to their bibliometric fitness), compliance progressively gives way to internalization. More and more researchers show interest (and pride) for their bibliometric indices, compare their values to those of their peers, and work at boosting them as efficiently as possible. Questions about the goal of science and the true nature of scientific quality fade in the background, as they seem to be less immediately relevant.

To fight against the raise of the bibliometric dictatorship, I see a primary role for established scientists, those who no longer need to prove their quality and still have (some) freedom to comply or not with the current fashion. Comparatively, younger scientists are more on a tight leash, as it is made clear to them that bibliometric performance is the key to their academic future. But if one no longer finds critical thinkers in the universities, where will one find them? So, maybe we should all switch off our computers for a moment and take the time to think: what we do, why we do it, and whether it is good to keep doing it this way. The wizard of O's is no real wizard as everyone knows, merely an illusionist. And if we let him do his thoroughly absurd job till the end, we are going to be known to the future generations as the civilization of fast-food science.

Prof. Dr. Philippe Hünenberger
ETH Zürich
Laboratory of Physical Chemistry
HCI G233
CH-8093 Zurich
Switzerland

Phone +41-44-632 5503
phil@igc.phys.chem.ethz.ch
http://www.csms.ethz.ch

Citation: Hünenberger P: Bibliometry: The wizard of O's. *Infozine* **2016,** Special Issue 1, 9–10, DOI: 10.3929/ethz-a-010745184

Published: December 12, 2016