Soft peer review: Social software and distributed scientific evaluation

Dario Taraborelli
University College London
Gower Street
London WC1E 6BT
United Kingdom
d.taraborelli@ucl.ac.uk

1 Beyond peer review: usage-based metrics and scientific quality assessment

A large debate has addressed in recent years the peer-review model of scientific assessment, questioning, among others, its ability to be affordable, accurate, timely, objective, and efficient at detecting fraud.[1] The debate tackled in particular the issue of what measurable indicators are available to estimate the value of scientific knowledge production. *Impact factor* [4] has undoubtedly become the *de facto* standard to measure quality in scientific production in many fields, but it has been challenged by several authors calling for more accurate or alternative indicators.[2, 5]

The necessity of new assessment strategies to overcome the limits of traditional peer review and the need of new metrics to complement impact factor indicators has become the object of a lively debate in the literature. In the field of Open Access, projects such as CiteBase or OpCit have been introduced to enable tracking popularity metrics such as the number of views or downloads per article and to explore the relationship between usage and impact for free online papers. Harnad observes that *usage-based metrics* are increasingly perceived by the scientific community as a necessary complement to traditional peer review as an indicator of scientific significance:

> a new potential measure of on-line impact, not available in the on-paper era, is *usage*, in the form of hits. This measure is noisy [in that] it can be inated by automated web-crawlers, short-changed by intermediate caches, abused by deliberate self-hits from authors, and undiscriminating between nonspecific site-browsing and item-specific reading) (...), [but] seems to have some signal-value too, partly correlated with and partly independent of citation impact. (S. Harnad, cit. in McKiernan [8])

A first milestone in this sense is a report published by the UK Serials Group on online usage factors (UF), whose objective was to obtain an initial assessment of the feasibility of developing and implementing journal usage factors as a criterion to measure scientific quality.[9] The results of this survey clearly show that usage-based metrics, as a way to complement traditional peer review, are perceived as a major need by several actors (authors, librarians, publishers) in the scientific communication system.

Online access data belong to a family of traditional metrics that have been recently challenged by the so called Web 2.0 revolution and by the diffusion of social software. Surprisingly, little has been done to date to understand how to combine the benefits offered by social software (such as social network analysis and human computation) to leverage scientific quality assessment strategies.

In this paper I address the following question: is there any solution to bridge the gap between traditional indicators such as citation analysis and impact factor on the one hand and raw access data on the other hand in order to provide efficient and reliable indicators of scientific significance as it is perceived by the academic community?

I will argue that social software (in particular social bookmarking systems) offer a unique opportunity to provide costless and accurate metrics that may become in the long run more relevant to measure scientific impact than raw hits or other forms of usage-based statistics. In this paper I review the case of *social bookmarking systems* targeted at the academic community such as CiteULike or Nature's Connotea and discuss the challenges traditional scientific evaluation processes face when compared with these systems.

2 Social software and collaborative metadata

Online reference managers are extraordinary productivity tools: they allow users to file scientific references from online databases and easily access, annotate, categorize and share these references with collaborators. It would be a mistake, though, to take this as their primary interest for the academic community. As it is often the case for social software services, online reference managers are becoming powerful and costless solutions to collect large sets of metadata, in this case *collaborative aggregated metadata on scientific literature*.

An item in an online bookmarking system (e.g. a paper from an academic journal) is described by a list of tags, ratings, annotations compiled by the user when filing the item in his or her library. Online reference managers allow such metadata to be aggregated from the entire user community. Taken at the individual level, these metadata are hardly of any interest, but at a large scale they are likely to outperform more traditional evaluation processes in terms of coverage, speed and efficiency. Collaborative metadata cannot offer the same guarantees as standard selection processes (insofar as they do not rely on experts' reviews and are less immune to biases and manipulations). However, they are an interesting solution for producing virtually costless evaluative representations of scientific knowledge at a large scale.

Traditional peer review has been criticised on various grounds but possibly the major threat it is currently facing is *scalability*, i.e. the ability to cope with an increasingly large number of scientific paper submissions, which - given the limited number of available reviewers and time constraints on the publication cycle - results in a relative smaller and smaller acceptance rate for high quality journals.

Although metrics based on collaborative metadata (what we may call *soft peer review*) will never replace hard evaluation models such as traditional peer review, they are in a good position to outperform them in terms of efficiency and scalability, at least as soon as they reach critical mass. When this happens and as soon as their potential is fully acknowledged, I anticipate that academic content providers (including publishers, scientific portals and bibliographic databases) will be urged to integrate metadata from social software services.

I discuss in this paper a list of indicators for which I expect social bookmarking services targeted at the academic community to challenge traditional quality assessment processes:

1. Semantic Relevance. Collaboratively aggregated tags can be used to extract similarity patterns, for automatic clustering or to improve the quality of search engine results.[3, 10] In the case of the scientific literature, tags can provide extensive lists of keywords for scientific papers, often more accurate and descriptive than those originally added by the author.
2. Popularity. Measuring the number of users bookmarking an item in their personal reference library can provide a reliable measure of the popularity of that item within a given community. Social bookmarking data are likely to provide more robust indicators than usage factors insofar as they result from the intentional behavior of users interested in marking an item for future use rather than from pure navigation patterns. *Bookmarking an item* is a much more relevant (and virtually more robust) usage pattern to estimate user interest than merely *following a link*. In this sense, social bookmarking systems are likely to provide accurate figures on papers that are frequently read and cited in a given area of science.
3. Hotness. Hotness can be described as an indicator of short-term popularity, a useful measure to identify emerging trends within specific communities. Mapping popularity distributions on a temporal scale is actualy common practice. Indicators such as ISI Impact Factor are complemented with time-dependent metrics (*High Immediacy, Cited Half Life*). Online Reference Managers can similarly be used to extract short-term popularity indicators that are much more reliable than raw usage factors.
4. Collaborative annotation One of the most understated (and possibly most promising) aspects of online reference managers is the ability they provide to collaboratively annotate content. Collaborative annotation functionality was introduced by pilot projects such as Naboj (a service allowing collaborative annotations of arXiv preprints) or electronic journals such as Philica, allowing open peer commentaries on the articles they feature. A distinctive feature of online reference managers is that they do not require specific incentives for notes and reviews to be produced, since annotating references is something individual users naturally do when filing a reference in their library. The issue of incentives

and of the cost related to reviewing content for free has actually been one of the major obstacles to the diffusion of open peer review systems, witness the failure of Nature's pilot experiment in 2006.[6] Collecting annotations from users of online reference managers, on the other hand, looks like a more viable strategy precisely because these annotations are generated spontaneously. Online reference managers allow users to add public notes and short reviews to items they bookmark, which in turn can be used to automatically aggregate collaborative lists of annotations without explicit incentives or call for commentary.

3 The role of collaborative evaluation in scientific knowledge communication

The potential of social bookmarking to provide relatively unbiased metrics and to answer the need of more accurate and scalable indicators of scientific significance has been underestimated in the current debate on usage factors. Compared to raw access-data, social bookmarking metrics are likely to provide better proxies to estimate the impact of scientific papers in the academic community insofar as they are aggregated from much more specific and incentive-free usage patterns: the act of *bookmarking* an item as opposed to the act of simply *following a link* or downloading a paper.

In the long run, I expect these bottom-up, distributed processes to become more and more valuable to the academic community and traditional publishers to acknowledge the necessity of integrating metadata collected through social software. This will be possible as soon as collaborative annotation services reach critical mass and start developing facilities (ideally programmable interfaces or API) to expose the data they collect and feed them back to potential consumers (publishers, individual users or other services). Whether this distributed, *soft peer review* will oust more traditional assessment models is a question that can only be answered by considering the conditions that any candidate system alternative to peer review should meet.[7]

The future role of social bookmarking systems, as I envision this, is not dissimilar to that of *mashup services*, as intermediate providers–between information producers and information consumers–of aggregated metadata. If they succeed in doing this, they will come to occupy a crucial function in the system of scientific knowledge production and challenge traditional approaches to scientific quality assessment.

References

[1] Peer review and fraud. *Nature*, 444(7122):971–972, December 2006. doi: 10.1038/444971b. URL http://dx.doi.org/10.1038/444971b.

[2] The impact factor game. *PLoS Medicine*, 3(6), June 2006. doi: 10.1371/journal.pmed.0030291. URL http://dx.doi.org/10.1371/journal.pmed.0030291.

[3] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 501–510, New York, NY, USA, 2007. ACM Press. ISBN 9781595936547. doi: 10.1145/1242572.1242640. URL http://portal.acm.org/citation.cfm?id=1242572.1242640.

[4] E. Garfield. The agony and the ecstasy— the history and meaning of the journal impact factor. September 2005. URL http://garfield.library.upenn.edu/papers/jifchicago2005.pdf.

[5] W. Glanzel. Journal impact measures in bibliometric research. pages 171–193, 2002. URL http://www.ingentaconnect.com/content/klu/scie/2002/00000053/00000002/00400216.

[6] S. Greaves, J. Scott, M. Clarke, L. Miller, T. Hannay, A. Thomas, and P. Campbell. Nature's trial of open peer review. *Nature*, December 2006. doi: doi:10.1038/nature05535. URL http://www.nature.com/nature/peerreview/debate/nature05535.html.

[7] C. G. Jennings. Quality and value: The true purpose of peer review. *Nature*, 2006. doi: doi:10.1038/nature05032. URL http://www.nature.com/nature/peerreview/debate/nature05032.html.

[8] G. McKiernan. Peer review in the internet age: Five (5) easy pieces. *Against the Grain*, 16(3):52–55, June 2004. URL http://www.public.iastate.edu/~gerrymck/DraftFive.htm.

[9] P. T. Shepherd. Final report on the investigation into the feasibility of developing and implementing journal usage factors. Technical report, May 2007. URL http://www.uksg.org/sites/uksg.org/files/Final%20Report%20on%20Usage%20Factor%20project.pdf.

[10] Y. Yanbe, A. Jatowt, S. Nakamura, and K. Tanaka. Can social bookmarking enhance search in the web? In *JCDL '07: Proceedings of the 2007 conference on Digital libraries*, pages 107–116, New York, NY, USA, 2007. ACM Press. ISBN 9781595936448. doi: 10.1145/1255175.1255198. URL http://portal.acm.org/citation.cfm?id=1255175.1255198.