Soft peer review: Social software and distributed scientific evaluation

Dario TARABORELLI *

Department of Psychology
University College London
Gower Street
London WC1 6BT
United Kingdom
d.taraborelli@ucl.ac.uk

Abstract The debate on the prospects of peer-review in the Internet age and the increasing criticism leveled against the dominant role of impact factor indicators are calling for new measurable criteria to assess scientific quality. Usage-based metrics offer a new avenue to scientific quality assessment but face the same risks as first generation search engines that used unreliable metrics (such as raw traffic data) to estimate content quality. In this article I analyze the contribution that social bookmarking systems can provide to the problem of usage-based metrics for scientific evaluation. I suggest that collaboratively aggregated metadata may help fill the gap between traditional citation-based criteria and raw usage factors. I submit that bottom-up, distributed evaluation models such as those afforded by social bookmarking will challenge more traditional quality assessment models in terms of coverage, efficiency and scalability. Services aggregating user-related quality indicators for online scientific content will come to occupy a key function in the scholarly communication system.

Keywords peer review; rating; impact factor; citation analysis; usage factors; scholarly publishing; social bookmarking; collaborative annotation; online reference managers; social software; web 2.0; tagging; folksonomy.

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. [2] The debate tackled in particular the issue of what measurable indicators are available to estimate the value of scientific knowledge production.

The motivations behind this debate are manifold, but they are partly related to the problem of the explosion of scientific content available in the World Wide Web. The massive availability of scientific content in the Internet is challenging the role academic journals had in the past as privileged vehicles for scientific communication and as filters of scientific quality: the Web has been actually paving the way to new forms of scientific evaluation (such as open peer review or open peer commentary [12]) that were not conceivable as such in the past. More dramatically, the Web is blurring the traditional distinction between content that has been selected because of peer review (what we may refer to as *a priori* scientific quality assessment) and content whose quality is determined by other criteria after its selection for publication (or *a posteriori* scientific quality assessment). Even if the importance of rigorous pre-publication selection criteria as a condition to secure scientific quality has hardly been compared to that of post-publication impact assessment, models such as Paul Ginsparg's two-tiered selection [8] have already started challenging the monolithic distinction between *a priori* and *a posteriori* evaluation.

Impact factor [7] has undoubtedly become the de facto standard to measure a posteriori scientific significance in many areas of research, but it has been challenged by several authors calling for more accurate or alternative indicators. [3, 9] 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 discussion 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 inflated 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 [16])

Whereas the search engine literature has long since acknowledged that hits or raw usage data provide a poor measure of popularity (let alone quality), there has been relatively little work on potential usage-related metrics that could complement

traditional quality indicators such as impact factor in the field of scientific literature. [5, 15]

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. [18] It is worth reporting some of the results of this survey:

- the majority of publishers are supportive of the UF concept, appear to be willing, in principle to participate in the calculation and publication of UFs, and are prepared to see their journals ranked according to UF;
- there is a diversity of opinion on the way in which UF should be calculated, in particular on how to define the following terms: total usage, specified usage period, and total number of articles published online. Tests with real usage data will be required to refine the definitions for these terms;
- there is not a significant difference between authors in different areas of academic research on the validity of journal Impact Factors as a measure of quality;
- the great majority of authors in all fields of academic research would welcome a new, usage-based measure of the value of journals;
- UF, were it available, would be a highly ranked factor by librarians, not only in the evaluation of journals for potential purchase, but also in the evaluation of journals for retention or cancellation;
- publishers are, on the whole, unwilling to provide their usage data to a third party for consolidation and for calculation of UF. The majority appear to be willing to calculate UFs for their own journals and to have this process audited;
- there are several structural problems with online usage data that would have to be addressed for UFs to be credible. Notable among these is the perception that online usage data is much more easily manipulated than is citation data.

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. It should be noted, though, that the scope of this survey was limited to the study of *access data* of online resources. Whereas usage statistics (such as those collected by the COUNTER project) certainly provide valuable information to estimate the popularity of online resources, it is arguable whether they will be able to correctly represent quality or scientific authority. In particular, it is debatable whether they will be able to overcome the major issues that afflicted search engine research over the last decade, which led it to abandon raw traffic data in favor of more accurate, scalable and spam-resistant criteria for quality assessment.

Online access data belong to a family of traditional ranking metrics that were recently challenged by the so called Web 2.0 revolution and by the diffusion of social software and socially aggregated web metrics. Surprisingly, little has been done to date to understand how to combine the benefits of social network analysis with scientific quality assessment in light of the new forms of collaboration allowed by Web 2.0 services.

The question I aim to address in this paper is the following: is there any kind of measurable indicator to bridge the gap between citation analysis and impact factor on the one hand and raw access data on the other hand in order to provide efficient measures of scientific quality 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. I review in particular the case of social bookmarking systems targeted at the academic community such as Nature's [Connotea](#) and [CiteULike](#) 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 socially 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 metrics based on these metadata are likely to outperform more traditional evaluation processes in terms of coverage, speed and efficiency. Social 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 very large scale.

Traditional peer review has been criticized 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 ratings based on collaborative metadata 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 of users. 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.

The following is a list of areas in which I expect metrics from social bookmarking services targeted at the academic community to challenge traditional quality assessment indicators.

2.1   Semantic relevance

A widely acknowledged application of tags as collaborative metadata is their use as semantic descriptors. Tagging is the most popular example of how social software (at least according to its defendants) helped overcome the limits of traditional, top-down approaches to content categorization. Collaboratively aggregated tags can be used to extract similarity patterns, for automatic clustering or to improve the quality of search engine results.[4, 19]

In the case of academic literature, tags can provide extensive lists of keywords for scientific papers, often more accurate and descriptive than those originally added by the author. Figures 1. and 2. compare keywords respectively used by the author and by the community of users to describe a popular article about tagging, ordered by the number of users who added a specific tag in their bookmarks as a descriptor for the article.

A cognitive analysis of tagging save this
to tags tagging cognitive cognition folksonomy facets ... saved by 1911 other people ... on sept 27, 2005

Figure 1: List of keywords for a popular article on "tagging" as compiled by the author, from Del.icio.us.

common tags   cloud | list

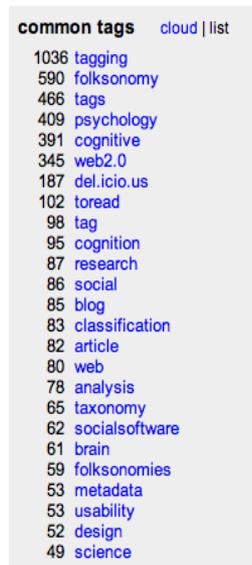|	
1036	tagging
590	folksonomy
466	tags
409	psychology
391	cognitive
345	web2.0
187	del.icio.us
102	toread
98	tag
95	cognition
87	research
86	social
85	blog
83	classification
82	article
80	web
78	analysis
65	taxonomy
62	socialsoftware
61	brain
59	folksonomies
53	metadata
53	usability
52	design
49	science

Figure 2: Distribution of collaboratively aggregated keywords for the same article as in figure 1, from Del.icio.us.

Similar lists can be found in CiteULike or Connotea, although neither of these services seem to have realized so far how important it is to rank tags by the number of users who applied them to a specific item. Measuring tag density per item in social

software is possibly the most reliable strategy to estimate the semantic relevance of an item without relying on expert feedback. Services allowing to aggregate keywords compiled by multiple users to describe scientific references are then in the best position to become providers of virtually cost-free, collaboratively aggregated semantic metadata for large sets of scientific articles and to challenge more traditional and costly top-down categorization approaches.

2.2 Popularity

Another fundamental type of metadata that can be extracted by social bookmarking systems is popularity indicators. Looking at how many users bookmarked an item in their personal reference library can provide a reliable measure of the popularity of that item within a given community. Understandably, academically oriented services (like CiteSeer, Web of Science or Google Scholar) have focused so far on citation analysis, which is the standard indicator of a paper's authority in the bibliometric tradition. I anticipate that popularity indicators from online reference managers will eventually become a factor as crucial as citation analysis for evaluating scientific content.

This may sound paradoxical if we consider that complex authority measures were introduced precisely to avoid the typical biases of raw usage-based popularity indicators. 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 spam-resistant) kind of action 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.

Whether social bookmarking popularity data are better indicators than access-based factors to measure the scientific significance of an article within a given academic community is an empirical question. It has been shown that ratings for scientific articles aggregated from an online community of biologists (F1000) strongly correlate with their impact factor. [1] A comparison of the distribution of citations, the distribution of popularity indicators in social bookmarking services and access-based figures for a representative sample of papers would provide a very much needed contribution to the understanding of how good different kinds of usage-related metrics are at predicting scientific impact. [see for instance 6, 11]

Interestingly, a number of social bookmarking systems such as Del.icio.us have started realizing the strategic importance of redistributing popularity data they collect. Del.icio.us recently introduced the possibility of displaying on external websites popularity indicators based on the number of users who filed a specific URL in their bookmarks. Similar ideas have been in circulation for years (consider for example Google's PageRank indicator or Alexa's traffic ranking in their browser toolbars) but it seems that social bookmarking systems have not fully acknowledged the importance of redistributing the metadata they collect.

posted to ecology folksonomy tagging web web_epistemology
by dartar as ★★ and 129 others ... on 2007-02-21 03:25:20

Figure 3: Popularity indicators for an article in CiteULike.

Connotea, CiteULike and similar services should consider giving back to content providers (from which they borrow bibliographic metadata) the ability to display the popularity indicators they produce. When this happens, it is not unlikely that publishers may start displaying popularity indicators on their websites (e.g. *Article X was bookmarked by 10,234 readers*) to promote their content.

2.3   Hotness

"Hotness" can be described as an indicator of short-term scientific significance, a useful measure to identify emerging research trends within specific communities. Mapping popularity distributions on a temporal scale is actually common practice. Indicators such as ISI Impact Factor are systematically complemented with time-dependent metrics: *High Immediacy* on the one hand describes the frequency of citations an article receives within a specific timeframe, which allows to identify journals that are good at providing cutting edge information; *Cited Half-Life* on the other hand can be used to estimate how long an article is perceived as relevant in the field. Similar criteria are used by social software services (such as Del.icio.us, Technorati or Flickr) to determine *what's hot* in the last days of activity.

Online reference managers have recently started to look at such indicators. As of its current implementation, CiteULike measures "hotness" by explicitly asking users to vote for articles they like. The goal—CiteULike developer Richard Cameron explains—is to "catch influential papers as soon as possible after publication". There are several reasons to believe that explicit votes may not be the best way to capture emerging trends within a given academic community. Relying on votes (whether they are combined or not with other metrics) is a questionable strategy to measure time-related popularity information from users, insofar as most users who use social bookmarking services are unlikely to cast a vote if they do not see its immediate benefits, whereas a large part of those users who actively vote may do so for opportunistic reasons. In order to provide reliable figures, popularity indicators should rely on patterns that are implicitly generated by user behavior: the best way to know what a community of users likes is definitely not to *ask* them, but to aggregate meaningful patterns from the natural behavior of users who joined a given service. Hopefully online reference management services will soon realize the importance of extracting measures of time dependent popularity in an implicit and automatic way: most mature social software projects solved this issue by avoiding the use of explicit votes.

2.4 Collaborative annotation

One of the most understated (and in my opinion, most promising) aspects of online reference managers is the ability they provide to collaboratively annotate content. Collaborative annotation functionality was introduced by platforms 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 an activity individual users naturally engage in 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. [10]

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 any explicit incentive or call for commentary.

Could such annotations be used to extract meaningful metadata at a large scale for the purpose of measuring scientific quality and impact? The obvious reason why bottom-up, collaborative annotation cannot be compared, in this respect, with traditional refereeing is that the expertise of the reviewer cannot be directly measured. The crucial question is then to understand if there is a viable strategy to make collaborative annotation more reliable while maintaining the advantages of social software.

Weighting user contributions by independently assessed authority is an issue that was recently brought to public attention by the Citizendium vs. Wikipedia debate. The solution proposed by the Citizendium founder to externally check the academic credentials of contributors is definitely a good solution to secure the quality of contributions against inaccuracy, abuse and vandalism. But the question remains open whether this approach is scalable without specific incentives. I suggest in what follows an alternative solution that could be implemented in online reference management systems to combine some features of anonymous refereeing with the benefits afforded by social software.

A possible bottom-up solution to the problem of ranking contributions by authority would be to introduce user rating as a function of their perceived expertise as measured by the user community. Asking users to directly rate each other is definitely not a viable approach: as in the case of "hotness" measures based on explicit votes, mutual user rating is an easily biasable strategy. Indirectly rating expertise by rating anonymous contributions looks like a much more robust solution. Assuming that users massively annotate references they file in their library and accept to make these notes public, notes from multiple users can be easily aggregated and displayed for each item. Notes could then be displayed anonymously to other users, who would have the possibility to save a note in their own library if they consider it useful. This behavior (i.e. importing someone else's annotation) could be taken as an indirect

positive rating for the author of the note, whose overall score would result from the number of anonymous contributions she wrote that other users imported.

These ratings can then be calculated on a per-topic basis. Suppose user A has a large number of positive ratings for comments she posted on papers tagged with tag *dna*: this will be a bottom-up indicator of her expertise for the *dna*; topic within the user community. User A will then have different degrees of expertise for topics *tag1*, *tag2*, *tag3*, as a function of how useful other users found her anonymous annotations to papers tagged respectively with *tag1*, *tag2*, *tag3*.

This is just an example of how valuable information could be extracted from collaborative annotations by adding an indirect rating layer within online reference management systems. Allowing indirect rating of annotations posted via anonymous contributions would allow implementing at a large scale a sort of *soft peer review* system. This in turn would allow social software services to aggregate much larger sets of evaluative metadata about scientific literature than traditional reviewing models will ever be able to provide at a large scale.

3 The role of collaborative evaluation in scientific knowledge production

I reviewed a number of ways in which social software metrics might help bridge the gap between traditional quality indicators and raw usage factors, thus answering to the need of more accurate metrics to evaluate scientific significance. The potential of social bookmarking to provide relatively unbiased metrics is 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 usage patterns: the act of *bookmarking* an item as opposed to the act of simply *following a link* or *downloading a paper*. Obviously, there is no guarantee that bookmarking is spam-free, and social bookmarking immune to self-promotion gaming, but there are several reasons to believe it is far more reliable as a proxy than mere usage data:

- bookmarks require user registration whereas usage data can be artificially inflated via robots;
- a bookmark indicates a single action by a user whereas in the general case there is no way to understand how many hits are generated by different users or by the same user visiting the same resource several times;

In this sense, social bookmarking systems offer a unique opportunity to provide a class of usage-related indicators of scientific quality that look more robust than any other kind of bottom up solutions to this problem.

How will traditional top-down quality assessment cope with the diffusion of new forms of distributed scientific evaluation? Whether 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.

Former Nature editor Charles G. Jennings [14] summarizes the basic requirements for a scientific quality assessment system as follows (emphasis mine):

- It must be *reliable* it must predict the significance of a paper with a level of accuracy comparable to or better than the current journal system.
- It must produce a recommendation that is easily *digestible*, allowing busy scientists to make quick decisions about what to read. [...]
- It must be *economical*, not only in terms of direct costs such as web operations, but also in terms of reviewer time invested.
- It must *work fast*. The peer review system produces clear-cut decisions relatively quickly (in part because editors pester reviewers to deliver their reports), whereas many forms of communal assessment such as the emergence of a statistically significant pattern of citations or expert recommendations are likely to be slow and gradual by comparison. [...]
- It must be *resistant to 'gaming'* by authors. Of course, savvy authors already know how to work the current system, but the separation of powers between editors and anonymous reviewers does I believe preserve some integrity to the process.

Understanding whether evaluations enabled by social bookmarking meet these criteria is beyond the scope of the present discussion. Quantitative analyses will have to compare how peer review and distributed evaluation processes perform as competing scientific assessment systems against the above benchmarks.

It is noteworthy, however, that on top of reliability requirements, several of the conditions suggested by Jennings explicitly refer to the sustainability of evaluation systems. It is not implausible in the long run scientific evaluation systems, in order to be sustainable, will have to become independent of scientific dissemination systems (e.g. scholarly journals run by academic publishers). Evaluation and dissemination can be regarded as two distinct functions in the scientific communication system [17] that are currently fulfilled by the same actors, i.e. peer-reviewed journals. There seems to be no reason to exclude that the relationship between evaluation and dissemination systems may change in the future under the pressure of new technologies. This is particularly likely in a situation in which scientific content and metadata about this content are massively available online, thus favoring the development of third party services.

The emergence of search engines as universal quality assessment institutions to orient users in content selection is the result of the pressure put on the system by the explosion of content and by the need of efficient and scalable solutions to cope with this explosion. In this sense, search engine have come to occupy a crucial epistemic function between knowledge producers and knowledge consumers in the World Wide Web. [13]

Scientific knowledge transmission may face the same destiny in an even more dramatic way. I proposed a few ways in which social bookmarking and collaborative annotation systems could be used to extract large-scale indicators of scientific quality from user behavior without the need of specific incentives. 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).

The future role of social bookmarking systems (as I envision this) is not dissimilar from that of mashup services, as intermediate providers–between information producers and information consumers–of aggregated metadata. To quote the conclusions of an article on the future of the mashup economy:

> [y]ou don't have to have your own data to make money off of data access. Right now, there's revenue to be had in acting as a one-stop shop for mashup developers, essentially sticking yourself right between data providers and data consumers.

A similar reason could justify a strong presence of these services in the scientific communication system. If they succeed in doing this, they will come to occupy a crucial function in the system of scientific knowledge production and challenge the traditional approaches to scientific assessment.

References

[1] Revolutionizing peer review? *Nat Neurosci*, 8(4):397–397, April 2005. doi: 10.1038/nn0405397. URL http://dx.doi.org/10.1038/nn0405397.

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

[3] 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.

[4] 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://dx.doi.org/10.1145/1242572.1242640.

[5] J. Bollen, H. Van de Sompel, J. A. Smith, and R. Luce. Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing & Management*, 41(6):1419–1440, December 2005. doi: 10.1016/j.ipm.2005.03.024. URL http://dx.doi.org/10.1016/j.ipm.2005.03.024.

[6] T. Brody, S. Harnad, and L. Carr. Earlier Web usage statistics as predictors of later citation impact. J. Am. Soc. Inf. Sci. Technol., 57(8):1060–1072, June 2006. ISSN1532-2882. doi: 10.1002/asi.v57:8. URL http://dx.doi.org/10.1002/asi.v57:8.13

[7] E. Garfield. The agony and the ecstasy— the history and meaning of the jour nal impact factor. In *International Congress on Peer Review And Biomedical Pub lication*, Chicago, September 2005. URL http://garfield.library.upenn.edu/ papers/jifchicago2005.pdf.

[8] P. Ginsparg. Can peer review be better focused. *Science & Technology Libraries*, 22 (3-4):5–17, January 2004. doi: 10.1300/J122v22n03 02. URL http://people.ccmr.cornell.edu/~ginsparg/blurb/pg02pr.html.

[9] W. Glanzel. Journal impact measures in bibliometric research. *Scientometrics*, 53(2): 171–193, 2002. URL http://www.ingentaconnect.com/content/klu/scie/2002/ 00000053/00000002/00400216.

[10] 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.

[11] S. Harnad. Open access scientometrics and the uk research assessment exercise. In D. Torres-Salinas and H. F. Moed, editors, *11th Annual Meeting of the International Society for Scientometrics and Informetrics*, volume 11, pages 27–33, 2007. URL http://eprints.ecs.soton.ac.uk/13804/.

[12] S. Harnad. *Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals*, pages 103–118. MIT Press, 1996. URL http://eprints.ecs.soton.ac.uk/2900/.

[13] C. Heintz. Web search engines and distributed assessment systems. *Pragmatics & Cognition*, 14(2):387–409, 2006.

[14] 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.

[15] M. Jensen. The new metrics of scholarly authority. *The Chronicle*, June 2007. URL http://chronicle.com/free/v53/i41/41b00601.htm.

[16] 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.

[17] H. Roosendaal and P. Geurts. Forces and functions in scientific communication. In *Cooperative Research Information Systems in Physics*, Oldenburg, Germany, August 1997. URL http://www.physik.unioldenburg.de/conferences/crisp97/roosendaal.html.

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

[19] 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://dx.doi.org/10.1145/ 1255175.1255198.