Soft Peer Review

Social Software and Distributed Scientific Evaluation

Dario Taraborelli

Centre for Research in Social Simulation
Department of Sociology
University of Surrey

COOP 08
May 21, 2008

Social software and scientific significance

- Is social software changing how we think of scientific quality and impact?
- Can social software provide answers to the challenges faced by the traditional system of scientific evaluation?
- The role of web-based collaborative tools in producing distributed representations of scientific significance

Overview

Beyond peer review?

Collaborative metadata on the scientific literature

Production and consumption of evaluative representations

Traditional indicators of scientific quality

Large debate on the future of peer review (*Nature, 2006*).

- accuracy
- neutrality
- robustness
- timeliness
- scalability

scalability: ability to cope with an increasingly large mass of written scientific production.

Need of scalable, timely and easily digestible *proxies* of scientific quality.

New indicators of scientific quality

Massive online availability of scientific content: new forms of scientific evaluation.

The Web is blurring the distinction between:

- content assessed via peer review (**a priori** scientific quality assessment)
- content assessed by more distributed criteria after publication (**a posteriori** scientific quality assessment).

(Hybrid systems (ArXiv), open peer review, fluid publication)

From author-dependent to reader-dependent indicators

A posteriori criteria. How to complement citation-based measures of scientific significance (such as *impact factor*)?

Role of **usage factors**: towards more reader-dependent indicators of impact.

> *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)*

Usage factors

UK Serials Group's report on online UF: feasibility of implementing usage factors as a way to measure scientific impact. (Shepherd 2007)

- the majority of publishers are supportive of the UF concept and prepared to see journals ranked according to UF
- diversity of opinion on the way in which UF should be calculated.
- no significant difference between authors in different areas of academic research on the validity of journal impact factors as a measure of quality
- majority of authors would welcome new, usage-based measure of the value of journals
- several structural problems with online usage data for UFs to be credible (robustness against manipulation compared to citation data).

UF robustness

Little effort to move beyond plain indicators of traffic-based popularity (download rates) in state-of-the-art literature on UF (Harnad 2007, Armbruster 2008, Bollen et al. 2008)

Web 2.0 is missing from the picture!

10+ years of search engine research: hits or raw traffic data provide poor measure of authority and impact.

Social search: Benefits of integrating metrics from social software (Yanbe et al. 2007, Bao et al. 2007).

Social bookmarking services

Social bookmarking: costless, scalable and more robust metrics of scientific impact than raw hits or other usage-based statistics.

del.icio.us

citeulike

Connotea

Social bookmarking services

An item filed in an online reference manager (e.g. a journal article) is associated with a list of metadata (*tags, ratings, annotations*) compiled by the user when saving the item in her library.

The Web's hidden order
Commun. ACM, Vol. 44, No. 9. (September 2001), pp. 55-60.
by Adamic LA, Huberman BA
posted to authority linking metrics web web_epistemology by dartar as ★★

Social software and collaborative metadata

Online reference managers allow such metadata to be aggregated across users.

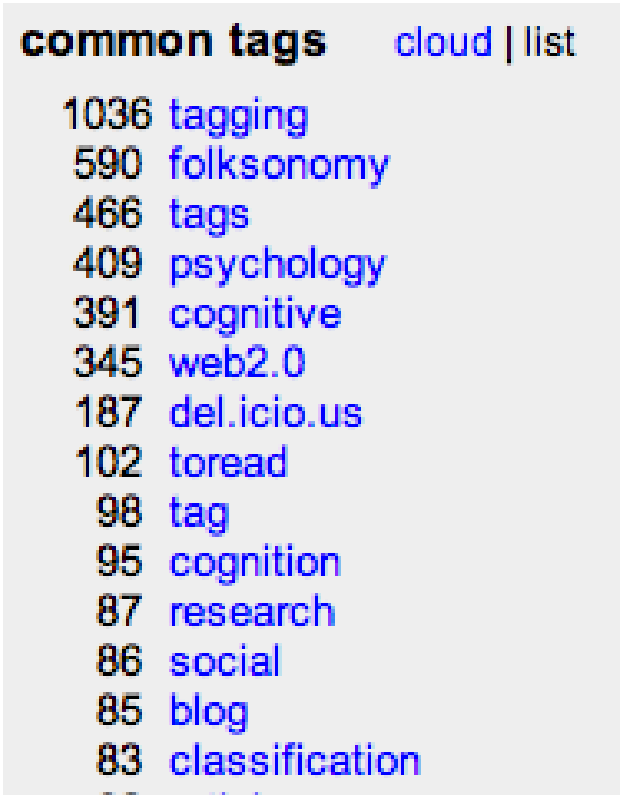
- Scarce interest of metadata taken at individual level
- Rich indicators and metrics when aggregated across the whole user community.

Powerful solution to collect large sets of collaborative metadata on scientific literature.

1. Semantic indicators

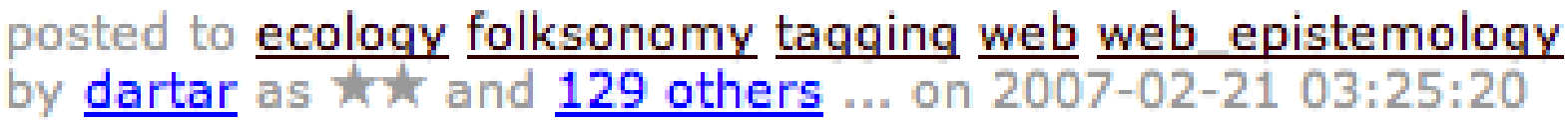
Collaboratively aggregated tags can be used to extract **semantic similarity** measures

Richer semantic descriptors than those originally provided by the author.



2. Impact and popularity indicators

Number of users who bookmarked the same item: effective metric to identify highly popular publications within a given community.



Robustness of bookmarking behaviour as an indicator of impact

- bookmarks require user registration whereas hits can be artificially inflated via robots;
- a bookmark indicates a single, intentional action performed by a user displaying interest for a publication;

Failure of attempts at ranking impact on the basis of explicit user rating.

3. Hotness

Metrics to identify **short-term impact**, or emerging trends within a given research community.

Similar criteria adopted in citation analysis - impact measured on a temporal scale:

- *High Immediacy*: frequency of citations an article receives within a specific time frame
- *Cited Half-Life*: estimate of how long an article is perceived as relevant in the field

Social bookmarking services can provide instant representations of *what's hot* within a specified time frame.

4. Collaborative annotation

Collaborative annotation introduced by platforms such as Naboj (collaborative annotations of arXiv preprints) or electronic journals (such as Philica).

- Online reference managers do not require specific incentives for notes and reviews to be produced
- Natural behaviour of users as opposed to costly open peer reviewing proposals (see Nature's pilot experiment - Greaves et al. 2006)

Aggregation of notes is a robust strategy to build large sets of evaluative representations of the scientific literature insofar as individual annotations are added for private, non-communicational use.

The role of collaborative evaluation in scholarly communication

Scope and limits of this approach:

- individual user credentials in collaborative system are not guaranteed compared to traditional assessment criteria;
- the system is not completely immune to self-promotion and gaming until it reaches critical mass;

Far more reliable source of proxy indicators than raw UF.

Requirements for scientific quality assessment systems

Criteria for any candidate system alternative to traditional peer review (**Jennings 2006**):

- 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
- It must be *resistant to 'gaming'* by authors.

Production and consumption of evaluative representations

Social bookmarking can be used to extract large-scale, affordable and timely indicators of scientific significance from user behaviour without the need of specific incentives.

Under which conditions can they compete with more traditional metrics of scientific quality?

- Correlation with standard indicators (e.g. citation data - see Brody, Harnad and Carr 2006)
- Critical user mass to be reached;
- Evaluative representations of scientific significance aggregated from social bookmarking to be redistributed as consumable affordances for other users;

Some take home messages

Towards alternative **usage-dependent metrics** to assess scientific significance:

- Scalable evaluation systems require *proxies*
- Social annotation allows large-scale, incentive-free creation of evaluative metadata
- Robustness of ranking metrics produced in non-communicational contexts (e.g. *bookmarking behaviour*)
- Collection and redistribution of evaluative representations as key to reach critical mass

Acknowledgments

Funding

Marie Curie Fellowship `MEIF-CT-2006-024460`

Cognition in Structured Electronic Environments.

Thanks to:

- Stevan Harnad
- Christophe Heintz
- Kevin Emamy, Richard Cameron (*CiteULike*)
- Ian Mulvany (*Nature Connotea*)
- *AcademicProductivity.com* readers