

Deconstructing Publication Models and Practices in CS

Azer Bestavros

Computer Science Department
Boston University

Telefonica Research, Barcelona, Spain
November 20, 2008

Stakeholders

- Authors
- Employers
- Research communities
- Conference organizers
- Commercial publishers
- Professional organizations
- Libraries
- On-line brokers and search engines
- Funding agencies and tax payers
- Not mutually exclusive!

Azer Bestavros

Deconstructing Publication Models and Practices in CS

2

Scope

- Issues raised by interplay between stakeholders:
 - The pressure of publish or perish
 - The tradeoff between publishing and patenting
 - The diminishing quality of reviews
 - The role of and effects on funding
 - The value and impact of conference versus journal publications
 - The perception and reality of cliquishness of top-ranked conferences
 - The premise and impact of open-access publications
 - Conferences as money-making propositions
- Impact of above on scientific record of CS research

Azer Bestavros

Deconstructing Publication Models and Practices in CS

3

Worried?

Azer Bestavros

Deconstructing Publication Models and Practices in CS

4

Worried?

Read before you cite!
[Simkin&Roychowdhury:2002]

Copied citations create renowned papers?
[Simkin&Roychowdhury:2005]

Unread citations

As a matter of fact, scientists often don't read what they reference, but copy citations from literature lists used in other papers instead. You can do this and get away with until one day you copy a citation, which enters in it a DNA of someone else's misprint. In this case you can be identified and brought to justice (similar to how biological DNA evidence helps to connect criminals, who committed more serious offences than yours).

The model of random-citing scientists (see cond-mat/0510150) was inspired by Fermi's insight and justified by the aforementioned repeated statements. It is as follows: when a scientist is writing a manuscript he picks three random papers, cites them, and also copies a quarter of their references. The model accounts quantitatively for empirically observed citation distributions (see Fig. 2). Simple mathematical probability not genius, can explain why some papers are cited a lot more than the other.

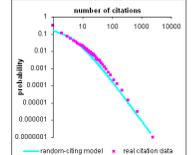
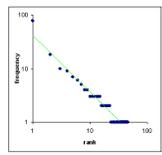


Figure 1. Distribution of suspects to citations to one renowned paper, ranked according to frequency of their repetition. (This figure is from cond-mat/0510150)

Figure 2. Outcome of the model of random-citing scientists compared to actual citation data. (This figure is from cond-mat/0510150)

Source:
<http://www.ee.ucla.edu/~simkin>

Azer Bestavros

Deconstructing Publication Models and Practices in CS

5

Symptoms of a Problem?

- Average number of publications on the CVs of faculty candidates is an order of magnitude larger than it was 20 years ago
- Reviews are mostly ignored in final version of proceedings, yet we insist on defending the significance of conference papers
- Reinventing the wheel across disciplines in CS is paramount (and embarrassing)
- Percentage of citations in archival journal and proceedings to non-archival works (TRs, web pages, ...) is alarming
- At least 25% of the citations in publications at top venues were never read by the authors citing them!
- 17 unsolicited emails from CRA community prior to workshop I ran on this subject at Snowbird 2006, many 1-1 side discussions at Snowbird, and dozens of followup emails afterwards!

Azer Bestavros

Deconstructing Publication Models and Practices in CS

6

A Vicious Cycle?

- Top-tier venues are increasingly selective – when acceptance rates are in the single digits, many decent papers get rejected
- Misguided bean-counting approaches to evaluating scholarship (for hiring & tenure) creates demand for less competitive venues
- Profiteers are more than happy to create even more venues (conferences and journals) to service that demand
- More publication venues result in the “re-submit until accepted somewhere” syndrome
- The proliferation of conferences and journals (some for good reasons) adds to the reviewing load on the community
- Lower-quality reviews mean that junk gets accepted (even in top-tier venues), polluting the scientific record
- A polluted scientific record means that the burden of weeding it falls on those with the least experience (the students)

Current Reviewing Models

- Number of reviews:
 - Fixed: Each paper gets k reviews
 - Variable: number of reviews proportional to rank: e.g., two reviews for bottom 50%, three for next 25%, and four for next 12.5%
- Number of Rounds:
 - One: All reviews are done independently
 - More: Reviews in early rounds are available in the next round
- Delegation:
 - PC member may ask an outside reviewer to submit a review
 - PC member must review all submissions
- Blindness:
 - Blind review: Reviewers are anonymous; authors are not
 - Double blind: Reviewers are anonymous; authors are anonymous
 - Blind authorship: Authors are anonymous; reviewers are not (!)

On Reviewing Processes and Impact

- When acceptance rates drop below some ungodly levels, it is often the case that other factors come into play in accept/reject decisions, e.g., prior exposition by PC/panel members to the work (also sometimes diagnosed as indicative of cliquishness). What is the “best response” from an author’s perspective?
- When acceptance rates drop below some ungodly levels, there is a risk that good or fresh (but perhaps risky) work may not even be submitted to top venues. What are the implications?
- Why not have conferences accept all papers that are “acceptable” but only accept for presentation the best of these (or even a random set)?

On Reviewing Load and Quality

- The quality of reviews is ultimately tied to reviewers’ “load”. Arguably, the ease of on-line submissions has escalated the number of submissions – especially due to the “*submission recycling until accepted*” phenomenon. Are there publication models that are effective in dealing with this?
- Why are reviews of accepted papers not made public? Would this mitigate the issue of review quality? Would this be an incentive for reviewers who may get citations for their reviews? There has been some interesting experiments here!
- Why are submissions to a conference not made public? After all, patent filings are made public before patents are issued. Wouldn’t this act as a natural filter that mitigates the review load issue and establish a fair basis for citing the earliest incarnation of an idea – not to mention solving problems of double submissions, submission recycling, ...

On the Dangers of Bean Counting

- There is a perception of a correlation between the quality of the work accepted in a conference and the acceptance rate in that conference. Yet, the argument can be made that acceptance rate is much more about the size of a community (the denominator of acceptance rate) and not the scientific merit of accepted papers (the numerator of the acceptance rate). The same issue applies to using citation counts.
- Arguably, a significant number of citations listed in scientific papers are not even read by those who cite them. This casts doubts on the value of citation indices such as CiteSeer or Google Scholar. Shouldn’t we actively discredit misconceptions such as number of citations = quality, and acceptance rate = quality?

On the Health of CS Scientific Record

- With the ease of on-line publications, prevalence of all-you-can-eat on-line digital libraries, and availability of tools such as CiteSeer and Google Scholar, should we worry about what future researchers will have to do to identify the proper scientific record?
- What could organizations such as NSF or CRA do to uphold the quality of the scientific record? Should they step in? Are the issues with CS really that much different from other scientific fields to warrant this?
- What could an organization such as NSF or CRA do to inform the tenure and promotion processes? Should they step in? Are the issues with CS really that much different from other scientific fields to warrant this?

On Ownership and Pricing

- What “value” does the name of a publisher add to a publication? Does that value justify the “price” charged?
- Who owns (or should own) the right to disseminate results of scientific research? Authors? Employers? Reviewers or editors? Publishers? Sponsors? Should that right be exclusive? Why (or why not)?
- In the spirit of the Bayh-Dole Act of 1980, why should tax-payers have to pay to access publications of federally-funded research?
- Conferences and journals often state that work published in these venues should not have appeared before. Why? What constitutes “work”? Is it the idea, the text, the experiments, or?

On Open Access

- Given that open-access on-line venues of publications are relatively new phenomena, we see hesitance from prospective authors to consider publishing their works in such venues. What are the reason for such hesitance?
- Given the low-barrier-to-entry for on-line open-access publication venues, it could be argued that the proliferation of such venues could further diminish the signal-to-noise ratio of the scientific record, with many negative implications (this is not much different from bloggers eating the lunch of major media outlets).
- Revenues from publications are vital to the financial viability of organizations such as the IEEE/ACM. Could organizations such as ACM or IEEE survive if open access is adopted? What if they don't? What do we loose?

On Role of Professional Organizations

- There is a perception that as long as a conference is profitable, an organization such as IEEE/ACM will continue to sponsor it. Would IEEE or ACM actively fund a conference or venue that loses money?
- Does the ACM or IEEE-CS worry about the quality of its publications? What processes are in place to assess the success or impact of IEEE/ACM conferences?
- In an age of on-line and open access, why should the CS community look for brand names such as IEEE or ACM or Usenix or SIAM or AAAI?

Interesting (Actual) Experiments

- Separate the reviewing process from the program selection process, i.e., acceptance decisions (e.g., Global Internet 2007)
- Allow papers to be submitted along with reviews from prior submissions (e.g., VLDB – Sigmod'06, Sigmod → VLDB'07)
- Publish a paper as well as an editorial on it, based on reviews of the paper (e.g., HotNets 2004)
- Increasing use of shadow program committees (e.g., Co-Next 2008, SOSIP 2007, Sigcomm 2005)
- Ask authors if they agree ahead of time to make their papers and the reviews public (e.g., Global Internet 2007)
- Allow authors (of border-line papers) to submit rebuttals prior to PC meeting (e.g., Sigmod 2005) – generally not well received

Interesting (Daring) Experiments

- Accept all papers that are acceptable to the PC, but present only a subset (e.g., based on a theme, or a random subset)
- Publish all submissions in an on-line archive and allow public comments (including anonymous) to influence PC decisions
- Token-based acceptance with post-conference review for inclusion in archival proceedings and future token distribution
- Overlay conferences or journals allowing editorial board to bid for best publicly available work to be presented and published

How to Judge the Quality of a Venue?

- Check if you or people you respect referenced papers in that venue
- Check their impact rating (ratio of citations to citable items) – is it in the top 3 in your area?
- Check their ratio of attendees to papers – is it 20:1 or is it 1:1 ?
- Check their sponsors – is it a society or a commercial entity?
- Check their program committee membership
 - How many did you cite or where cited by a paper you liked?
 - How many are in institutions/labs you would love to join?
- Tong-in-Cheek:
 - Avoid conferences with grand names – e.g., “International”, ...
 - Avoid conferences that students cannot afford – e.g., “Cruise Ship” ...
 - ...

How to Respond to a Rejection?

- Do not complain; do not take it personally.
- Paper rejection means there were better papers in the subjective view of the PC.
- Read your submission before you read the reviews
- Put yourself in the reviewer's shoes – did you make it easy for them to see your contributions?
- The process is noisy, but not random – try to distill the message you get from the reviewers (including, it was the wrong venue).
- Do not compromise your idea of a good paper simply to incorporate feedback – not all feedback is good!
- Climb the ladder: you should only try to resubmit to equally-strong or stronger venues!

Good Practices

- You are what you read – be selective in what you read and cite
- Take risks (while you can) – think big; avoid tunnel vision
- Do not slice and dice your work to get more publications – your work will have less impact if you do
- Do not cite yourself unless absolutely necessary – when possible, refer to other people's work instead of your own
- Retain copyright to all your works

Good Practices

- Finish the paper a few weeks ahead; give it to people to read; do something else; and then come back to it
- Always issue your completed works as on-line technical reports prior to submission for review – even if venue is double-blind
- Always send your completed works prior to submission to experts or people you want to inform about your work
- Go on a tour to evangelize your work even before your work is out for review

Questions and Thoughts

- Is writing a good paper enough to get noticed or to get cited?
 - There is no substitute to evangelizing your work!
- Why is "originality" of work such a big deal? What is wrong with "incremental (but groundbreaking) work"?
 - These are means to an end, not an end by themselves! Cost of publication is negligible; the editorial process (if good) is expensive; these are "means" to streamline the editorial process.
 - Be careful not to under-sell your work as incremental! The result may be incremental, but the approach to get the result may be original.
- What is better, to accept a bad paper or to reject a good paper?
- What is the role of shepherding a paper? Is it worth the time of the authors? of the shepherds?

Questions and Thoughts

- Won't the lower-barrier entry of open-access publication lead to proliferation of low-quality publications?
 - There is no relationship between the business model and editorial process. Quality is a function of the editorial process.
- Should CS authors worry about the longevity of open-access journals?
 - Yes, but they should also question the longevity of commercial publishers!
 - Why is CS much more conservative about open-access publication than other scientific disciplines?
- Overlay journals are the future! Their value added is the stamp of approval of their editorial board.