CRA Snowbird Workshop:

Deconstructing Current Models of Publications in CS
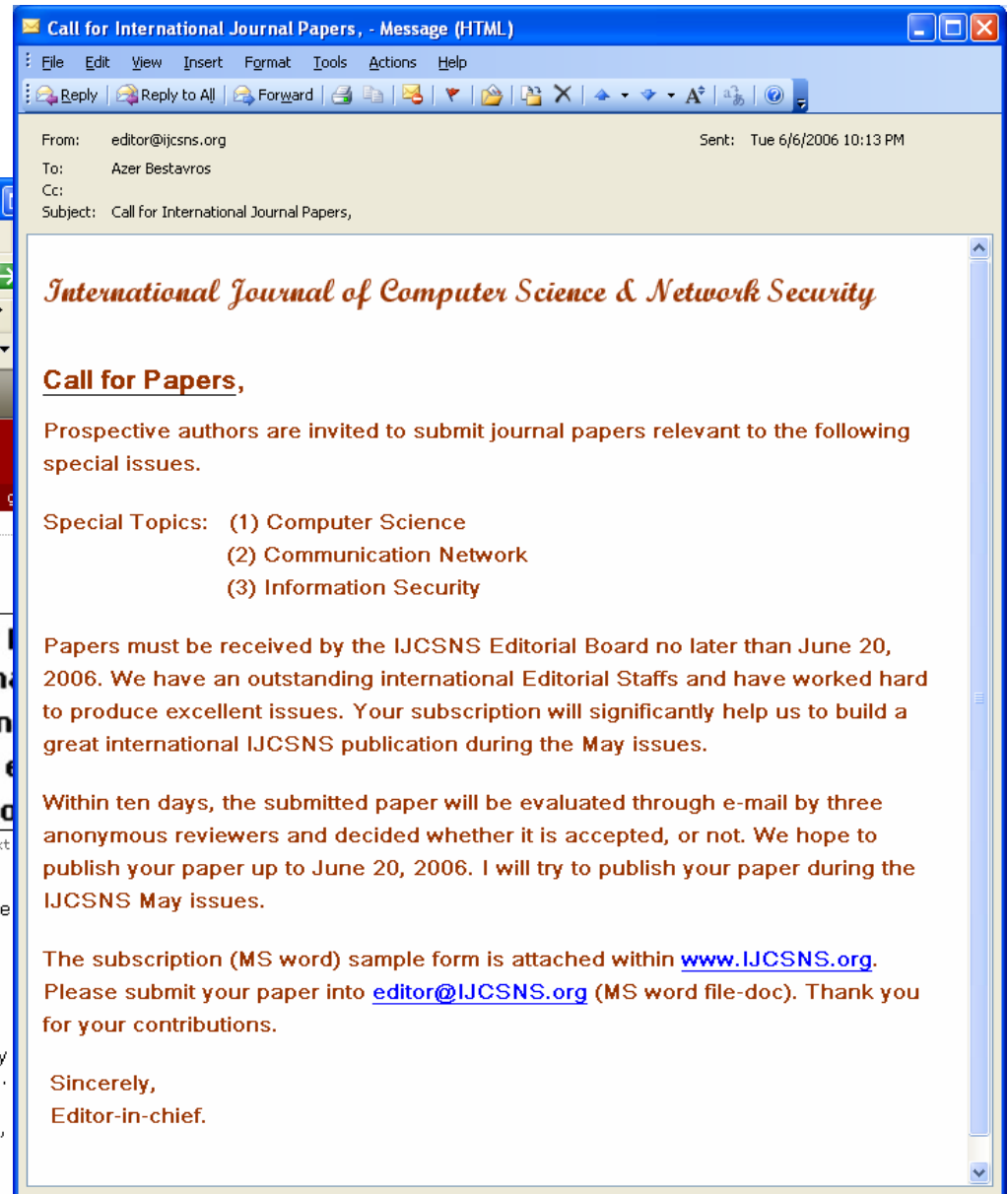
Azer Bestavros

Computer Science Department

Boston University

Scope

- Interplay between various stakeholders, which raises issues such as
    - The tenure pressure of publish or perish
    - 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 the above aspects on the scientific record of CS research

- The role that organizations such as CRA or NSF could or should play

Worried?

File   Edit   View   Favorites   Tools   Help

Address   http://news.b

Google   :cept random papers   G Search   PageRank   msn   Links

BBC NEWS | Americas ...

BBC

Home   News   Sport   Radio   TV   Weather   Languages

BBC NEWS

○ UK version   ● International version   About the versions | Low g

Last Updated: Friday, 15 April, 2005, 16:39 GMT 17:39 UK

✉ E-mail this to a friend   🖶 Printable version

Prank fools US science conference

A collection of computer-generated gibberish in the form of an academic paper has been accepted at a scientific conference, to the delight of hoaxers.

Three US boffins built a programme designed to create research papers with random text, charts and diagrams.

rant liberty formic
eath occlusive divina
l kind of medication
dministrate haggle e
pace lamentation co

The bogus paper uses random text

Two bogus papers were submitted to a computing conference in Florida, and one of them was accepted.

One of the hoaxers said the fake paper was designed to expose the lack of standards at academic gatherings.

The paper has the nonsense headline "Rooter: A Methodology for the Typical Unification of Access Points and Redundancy".

It was accepted for the World Multiconference on Systemics, Cybernetics and Informatics (WMSCI), due to be held in the city of Orlando in July.

News Front Page

Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Science/Nature
Technology
Entertainment

Have Your Say
In Pictures
Country Profiles
In Depth
Programmes
RELATED BBC SITES
SPORT
WEATHER
ON THIS DAY
NEWSWATCH

🌐 Internet

Call for International Journal Papers, - Message (HTML)

File   Edit   View   Insert   Format   Tools   Actions   Help

Reply   Reply to All   Forward

From: editor@ijcsns.org   Sent: Tue 6/6/2006 10:13 PM
To: Azer Bestavros
Cc:
Subject: Call for International Journal Papers,

International Journal of Computer Science & Network Security

Call for Papers,

Prospective authors are invited to submit journal papers relevant to the following special issues.

Special Topics: (1) Computer Science
                  (2) Communication Network
                  (3) Information Security

Papers must be received by the IJCSNS Editorial Board no later than June 20, 2006. We have an outstanding international Editorial Staffs and have worked hard to produce excellent issues. Your subscription will significantly help us to build a great international IJCSNS publication during the May issues.

Within ten days, the submitted paper will be evaluated through e-mail by three anonymous reviewers and decided whether it is accepted, or not. We hope to publish your paper up to June 20, 2006. I will try to publish your paper during the IJCSNS May issues.

The subscription (MS word) sample form is attached within www.IJCSNS.org. Please submit your paper into editor@IJCSNS.org (MS word file-doc). Thank you for your contributions.

Sincerely,
Editor-in-chief.

Deconstructing the current models of publications in CS

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 it until one day you copy a citation, which carries 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 convict criminals, who committed more serious offences than yours).
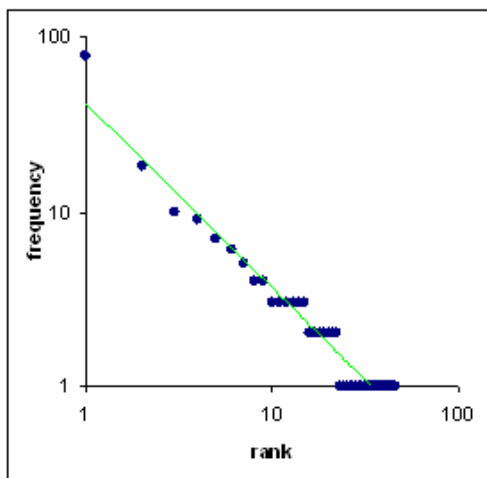
Figure 1. Distribution of misprints in citations to one renowned paper, ranked according to frequency of their repetition. (This figure is from cond-mat/0212043)

The **model of random-citing scientists** (see cond-mat/0305150) was inspired by Fermi's insight and justified by the aforementioned repeat misprints. 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 distribution (see Fig. 2). Simple mathematical probability, not genius, can explain why some papers are cited a lot more than the other.
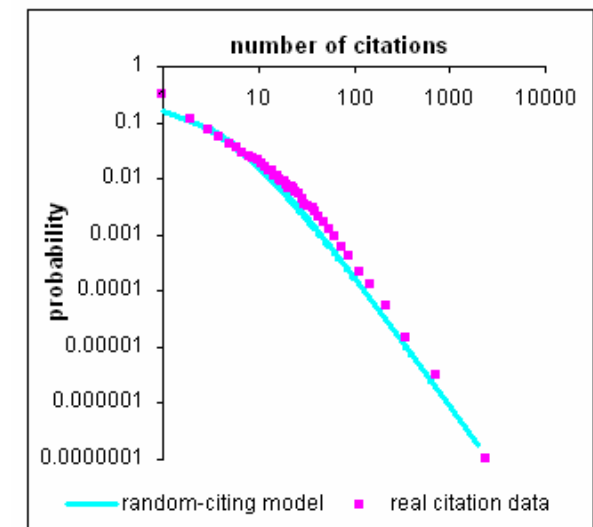
Figure 2. Outcome of the model of random-citing scientists compared to actual citation data. (This figure is from cond-mat/0305150)

Deconstructing the current models of publications in CS

Symptoms of a Problem?

- Average number of publications on the CVs of faculty candidates is an order of magnitude larger than it was 15 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 Snowbird and many 1-1 side discussions at Snowbird!

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)

Stakeholders

- Authors
- Research communities
- Conference organizers
- Commercial publishers
- Professional organizations
- Libraries
- On-line brokers and search engines
- Funding agencies and tax payers

- Not mutually exclusive!

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, do you 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). Comments?

- Are you interested in seeing organizations such as ACM or IEEE adopt an open access policy? If they do, do you think that the "raison d'etre" for publications by independent groups would vanish? Is that good or bad?

On Role of Professional Organizations

- How important are revenues from publications to the financial viability of organizations such as the IEEE/ACM?

- 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 looses 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?

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). Comments?

- 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. Comments?

- 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?

- 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. Comments?

- 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 students will have to do to identify the proper scientific record?

- What could an organization 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?

CRA Snowbird Workshop:

Deconstructing Current Models of Publications in CS

Panelists

Azer Bestavros @ Boston U (Moderator)

Jerry Engel @ IEEE CS

Michael Pazzani @ Rutgers U & NSF

Jennifer Rexford @ Princeton U

Moshe Vardi @ Rice U

John White @ ACM