Qualitative Citation Analysis

Frits Vaandrager

Institute for Computing and Information Sciences
Radboud University Nijmegen

SIREN, 6 October 2005

Acknowledgements

Thanks to

- SIREN organizers
- Members advisory board NWO project "Developing bibliometric indicators of research performance in Computer Science"
- Henk Moed
- Frans Birrer

... but views presented in this talk are my own.

Perspective: prospective "victim" of citation analysis.

Outline

Outline

Outline

Outline

The Quest for Excellence

> *"All research at the Radboud University Nijmegen is leading. In external research assessments it is rated at least very good (4) on the criteria quality, productivity, vitality and relevance."*

Strategic Plan Radboud University Nijmegen 2005-2009

Defining Productivity

"The committee assesses academic productivity by relating the output (the number of publications in total and in each category) to the input of human resources."

Assessment of Research Quality, Protocol 1998, VSNU

Productivity Definition of Committee Steels

$$Productivity = \frac{Output}{Input}$$

where

- $Output$ = number of academic publications, with weight 2 for PhD theses and books
- $Input$ = full time equivalents in research, not counting PhD students

1 research monograph of 760pp = 2 journal papers of 2pp
1 postdoc = 2.5 faculty

Lessons Learned

- It is very tempting to use science indicators that are "automatically" computed, even for people who should know
- Indicators often unreliable
- Shirking of responsibility

Citation Analysis for Computer Science?

There is *some* correlation between (perceived) quality of researchers and the number of citation to their work. E.g.,

- All Dutch groups that were rated excellent on one or more criteria in previous assessment have representative in Top50 of most cited Dutch Computer Scientists according to Citeseer.
- All researchers in Top16 working in Dutch institute belonged to group that was rated excellent on one or more criteria.

Less correlation between Citeseer list and outcome last assessment.

Practical Problems (apart from Conceptual Ones)

- Citeseer and Google Scholar full of "junk" and mistakes; criterion for inclusion not clear
- ISI incomplete

E.g., paper Lynch, Segala & Vaandrager on *Hybrid I/O Automata* cited 142× in Citeseer, 175× in Google Scholar and 6× in ISI

For many CS researchers coverage ISI better than Citeseer

Ongoing CWTS Study

Goal:
Methodological study aimed at development of bibliometric indicators of research performance in various subfields of Computer Science

Basic approach:
Extension of ISI database with publications from LNCS, ACM and (selection of) IEEE. Study citations to publications of Dutch Computer Scientists from period 1996-2001.

Which Citations Should Count?

- If we only count citations from ISI, LNCS, ACM and part of IEEE, for some groups coverage will be lower than for others.
  Coverage AI? Computational linguistics? CS is intrinsically multidisciplinary
- Citations to technical reports?
  No!? How about 1084 citations to Plotkin's DAIMI report in Google Scholar?
- Citations to papers in e.g. EATCS Bulletin?
  No!? How about 266 citations to Coalgebra Tutorial by Jacobs & Rutten?
- Citations to papers in obscure workshop?
  No!? Still proceedings of such workshops are often published as LNCS

What Should We Count? (cnt)

- Citations to books?

  No!? How about 1256 citations to Barendregt's book in Google Scholar?
- Citations to course notes?

  No!? But I am proud if my papers are discussed in courses!
- Self citations?

  No!? But what if other group builds on idea that someone contributed in joint paper? Isn't this research impact?
- Should all citations have equal weight?

How To Define "Performance Indicators"?

- Total number of citations?

  Hard to find; may be inflated by small number of big 'hits', which may not be representative of the individual if he/she is coauthor with many others
- Citation per paper?

  Hard to find; penalizes high productivity; in CS often paper first appears in proceedings, then as TR, then in journal
- Hirsch index?

  Difficult to compare individuals with different scientific age

Typical Arguments Used in Discussion

- *"True quality insensitive for precise definition"*
  A genius will surface irrespective of precise definition.
  But for most of us, choice of definition determines score 5, 4 or 3, i.e. life or death according to RU strategic plan.
- *"CS discipline should grow up"*
  Grow up to become what? Physicists?
  What if Dutch CS grows up and rest of the world doesn't?

Should We Use Citation Analysis At All?

Yes! In fact, we already do. ISI, Citeseer and Google Scholar are used widely for evaluation purposes and help to improve quality of evaluations

Key advantage of Citeseer and Google Scholar:

direct access to citing paper

Citeseer also provides context of citations.

Browsing for one hour through citations to researcher (or group) already provides a lot of insight!

Vision

Citeseer like system that allows one to

- Enter set of target (cited) publications of individual or group
- Specify well-defined set of source (citing) publications, e.g., LNCS, ACM, IEEE, Elsevier,..
- Browse on-line through citing papers, both in and outside well-defined set
- Quantitative information *only* as catalyst for thought
- Open!

Qualitative versus Quantitative

Assuming that evaluation committee does a proper job (!), qualitative citation analysis avoids some problems quantitative approach

- Goal displacement
- Manipulation of own citation scores

Conclusions

- Technology to realize vision exists (but can be improved)
- Main problems are political/financial
- System in interest international research community
- NWO/Universities should sponsor projects in this area
- With most publications on-line CS can play role of forerunner
- Don't wait for perfect system before using it

Conclusions (cnt)

- As consequences of research assessments are getting so big, their quality should be improved accordingly
- Duplication of work should be avoided
- Evaluation committees should properly justify their judgment