Guillaume Cabanac
guillaume.cabanac@univ-tlse3.fr
@gcabanac

March 29th, 2015

In Praise of Interdisciplinary Research through Scientometrics

http://bit.ly/birCabanac2015

Workshop on Bibliometric-enhanced Information Retrieval (BIR)

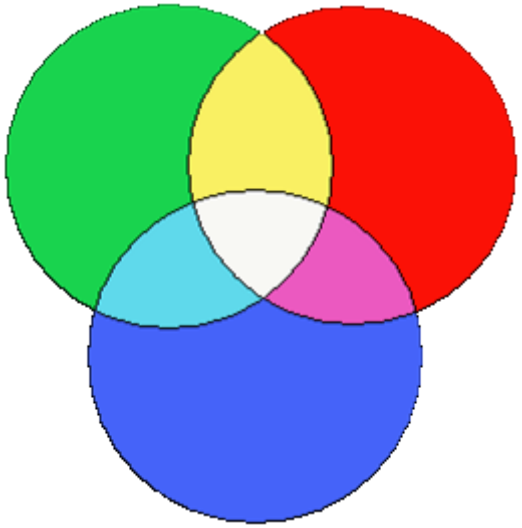
My Research Interests

Information Science

- Information Retrieval
- Digital Libraries

Psychology of Science

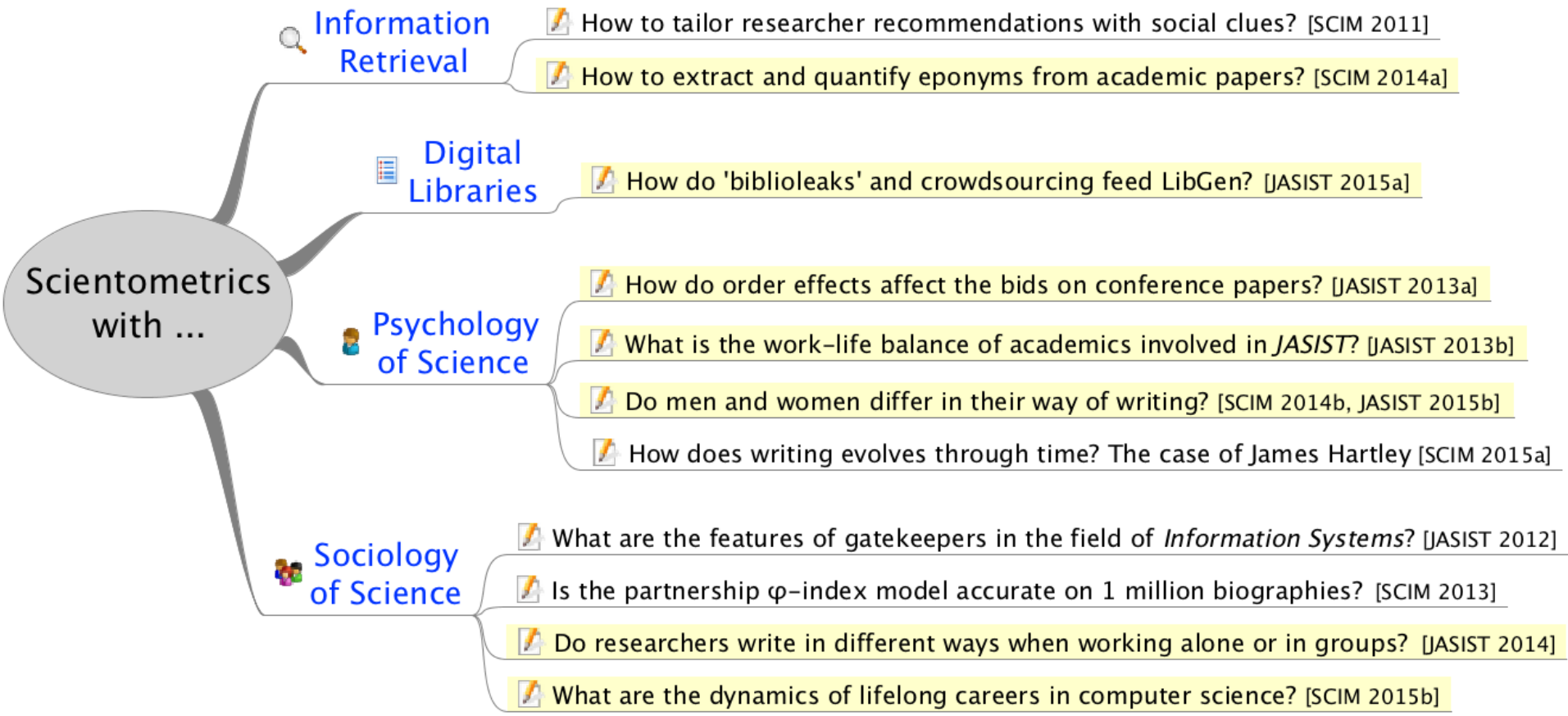
- Cognitive biases
- Academic writing



Sociology of Science

- Gatekeeping
- Collaboration

Case Studies



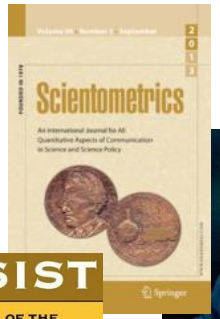
What's in my Toolkit?

ISI Web of Knowledge

Journal Citation Reports®

ScienceDirect

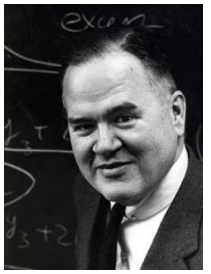
dblp
computer science bibliography



Price

Merton

SOFA



Tukey

ORACLE®

I ♥ #!/bin/bash

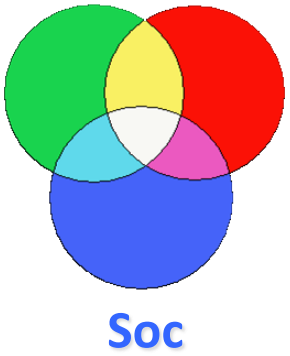
Java™

HtmlUnit

Hartley

Hubert

Milard



Information Science

- Information Retrieval
- Digital Libraries

Study 1

Eponymy and Scientific Heroes

Cabanac, G. (2014). Extracting and quantifying eponyms in full-text articles. *Scientometrics*, *98*, 3, 1631–1645. doi:10.1007/s11192-013-1091-8

Eponymy and Scientific Heroes

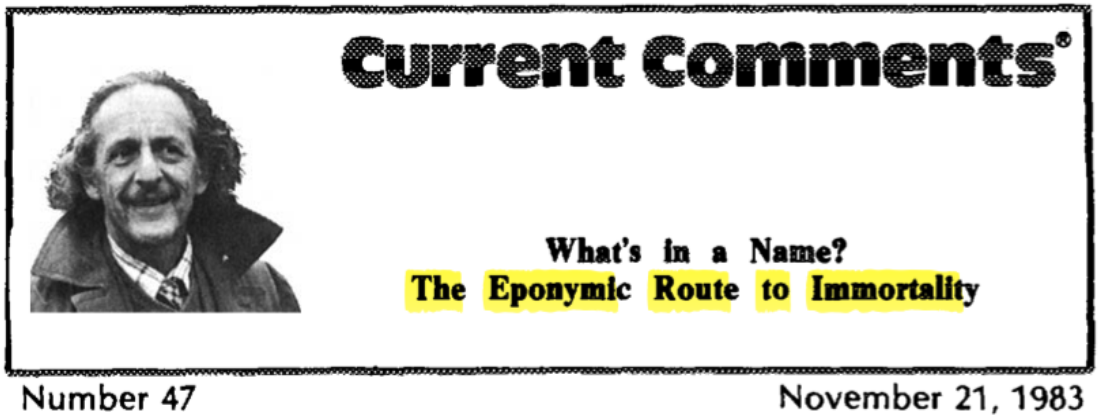
Robert K. Merton
1910 - 1992

"**mnemonic** and **commemorative** device" (p. 121)

Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1(1), 115–126.

"the practice of **affixing the name of the scientist** to all or part of what he has found, as with the Copernican system, Hooke's law, Planck's constant, or Halley's comet" (p. 643)

Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*, 22(6), 635–659.



"Eponyms remind us that science and scholarship are the **work of dedicated people**." (p. 393)

Eponymy and Scientific Heroes

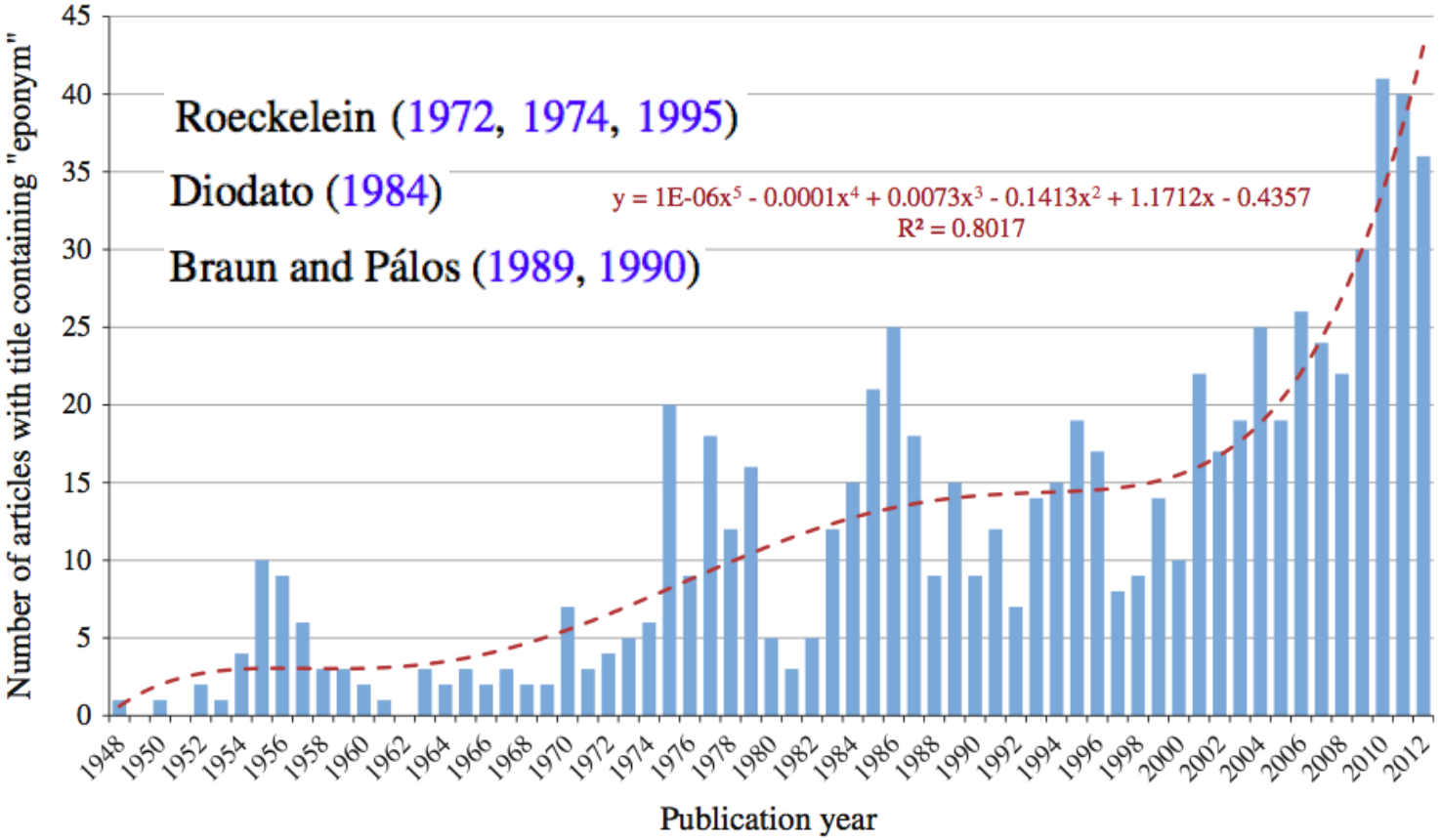


Fig. 1 The number of articles published each year about eponyms has been increasingly growing, with a sharper increase during the last decade. These 743 records were retrieved from the *Web of Knowledge* on April 1st, 2013 by searching for "eponym*" in the Title field. Note that this figure underestimates the literature of eponyms, as articles pre-dating 1948 or lacking this word in their titles were not retrieved (e.g., Merton 1942)

Eponymy and Scientific Heroes

- Related Concepts
  - The Reward System of Science (Merton, 1942, 1957)
  - Obliteration by Incorporation (Merton, 1988; McCain, 2011, 2012)
  - Non-indexed Eponymal Citedness (Száva-Kováts, 1994)

- Extraction and Quantification of Eponyms from Full-text
  - Show eponimised scientists in my field?
  - Update existing dictionaries of eponyms?
  - Spot new research trends?

I have long worshiped the eponym as one of the last vestiges of humanism remaining in an increasingly numeralized and computerized society.

(Robertson 1972)

Eponymy and Scientific Heroes

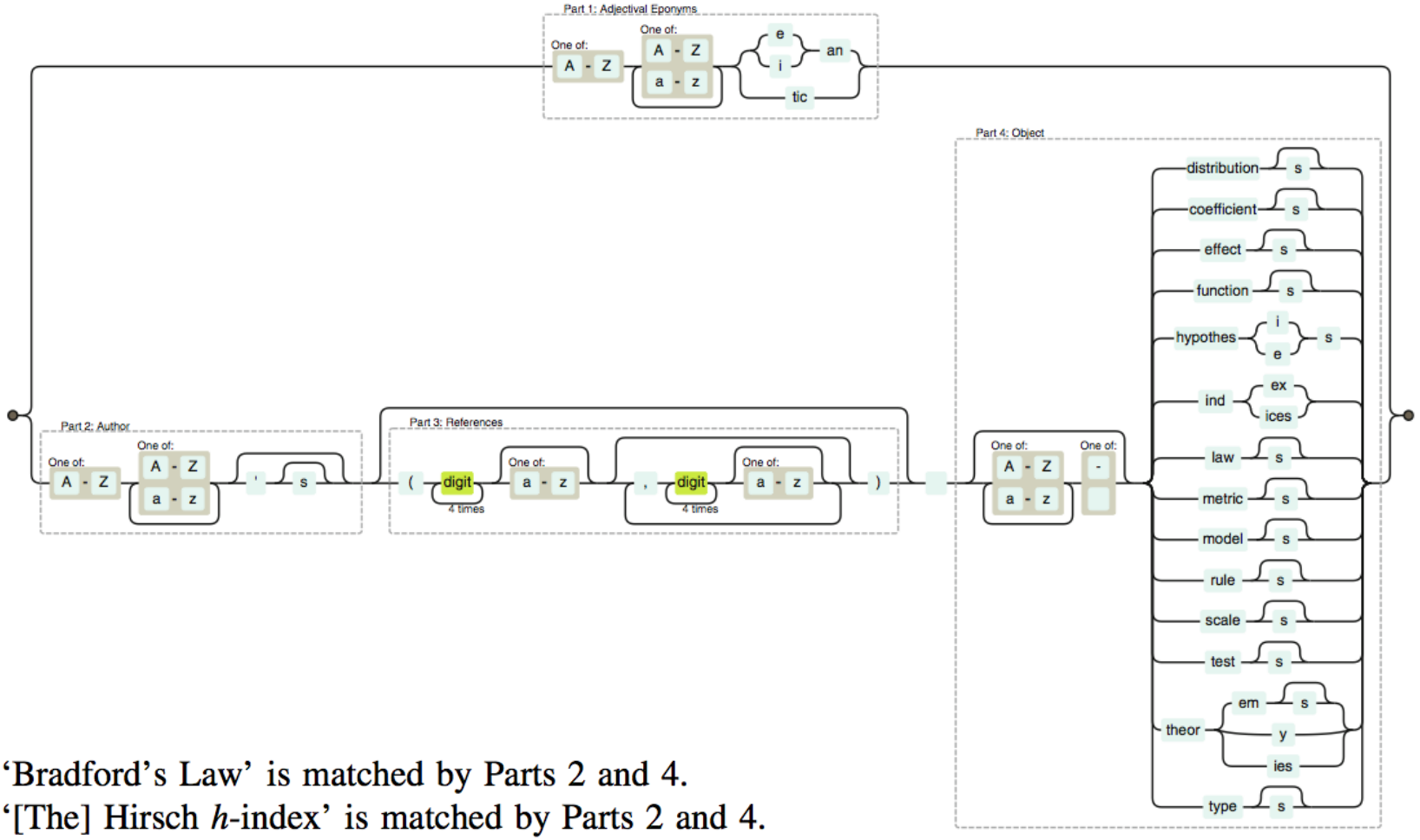


Fig. 2 Syntax diagram of the regular expression used in Listing 1 to extract eponyms from text. The upper sub-expression (i.e., Part 1) matches adjectival eponyms (e.g., 'Mertonian'), whilst the lower sub-expression (i.e., Parts 2–4) matches nominal eponyms, such as 'Vinkler's (2010a, 2013) π_v-index.' This diagram was produced by http://www.regexper.com

'Bradford's Law' is matched by Parts 2 and 4.
'[The] Hirsch *h*-index' is matched by Parts 2 and 4.
'Vinkler's (2010a, 2013) π_v-index' is matched by Parts 2–4.

Eponymy and Scientific Heroes

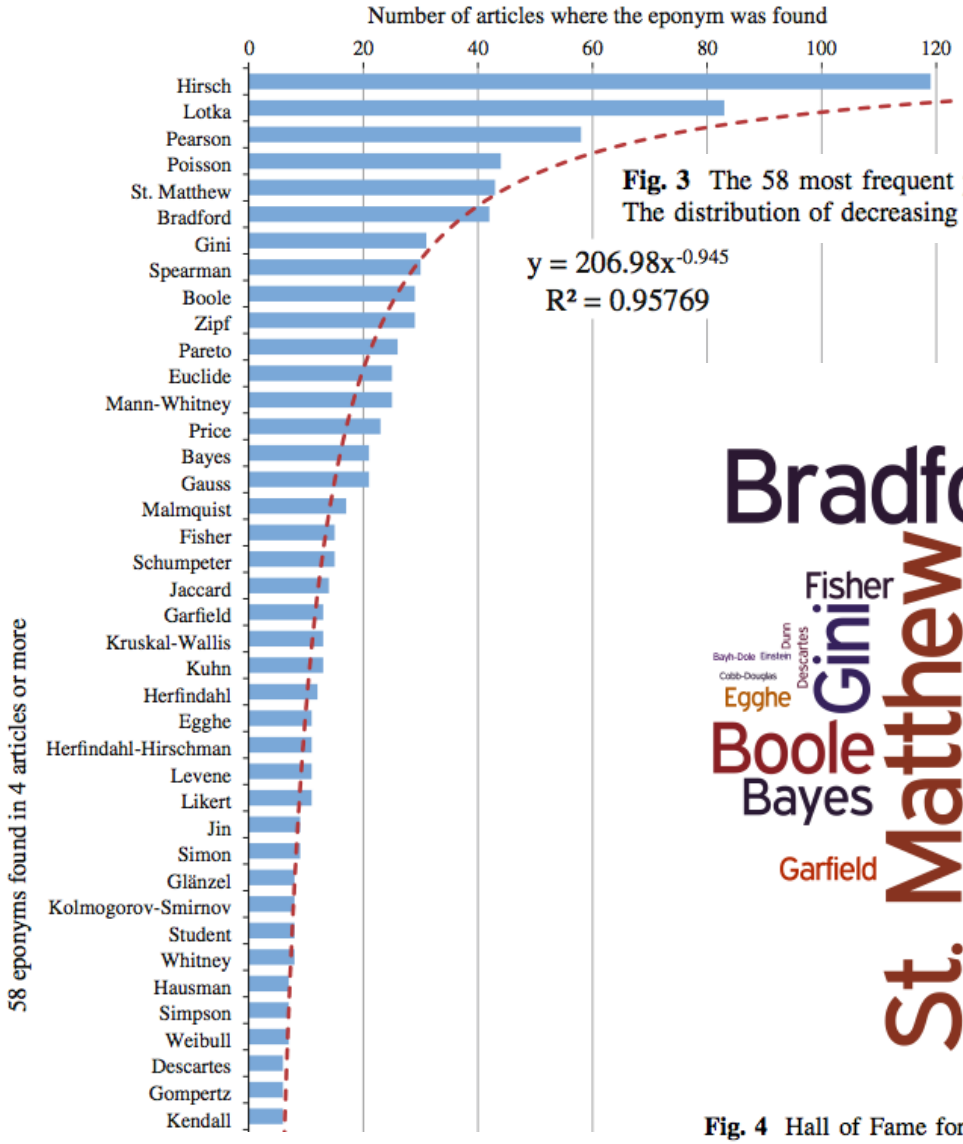
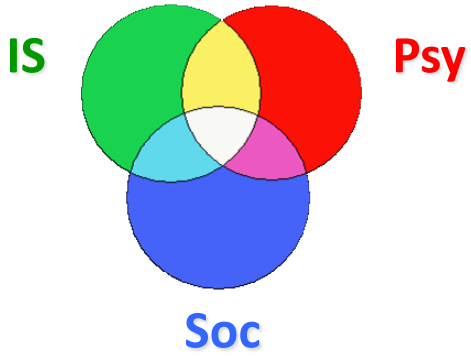


Fig. 3 The 58 most frequent person names cited in 821 *Scientometrics* articles published in 2010–2013. The distribution of decreasing name occurrences fits a power law ($R^2 = 0.9577$)

Fig. 4 Hall of Fame for the eponimised persons extracted from 821 *Scientometrics* articles published in 2010–2013 (see Fig. 3). This word cloud was produced by http://www.wordle.net



Information Science

- Information Retrieval
- Digital Libraries

Study 2

Bibliogifts = Bibioleaks + Crowdsourcing ?

Cabanac, G. (forthcoming). Bibliogifts in LibGen? A study of a text-sharing platform driven by biblioleaks and crowdsourcing. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23445

J Med Internet Res 2014;16(4):e112
DOI: 10.2196/jmir.3331

Viewpoint

Is Biblioleaks Inevitable?

Adam G Dunn[1], PhD; Enrico Coiera[1], MBBS, PhD; Kenneth D Mandl[2,3], MD, MPH

Through a concerted effort, hackers gain access to the databases of six publishers that together control access to the majority of subscription-based biomedical journal articles. This group makes copies of every article from every journal and releases them into the public domain. Subsets of articles are mirrored in anonymous peer-to-peer networks, creating a decentralized and multiply-redundant repository that is accessible to any human or computer algorithm. The repository grows when its users begin to add new and missing articles, creating a self-sustaining system of frictionless, free, and universal access to published research. While there would be recourse against offenders and while the wider academic community may be unlikely to embrace illicit activity, a robust international article-sharing underground is created. Academics in wealthy countries generally enjoy the privilege of institutional subscriptions to many journals, but articles that require payment to read or download (paywalled) are largely beyond the reach of everyone else and there is a substantial motivation to access this new resource.

The potential for this form of guerrilla open access is rarely discussed
of recent cyber-attacks against commercial and government interests. L
of illegal music file sharing on Napster and the massive releases of gov
The War Logs and global surveillance disclosures, can force these issu
into the mainstream public debate.

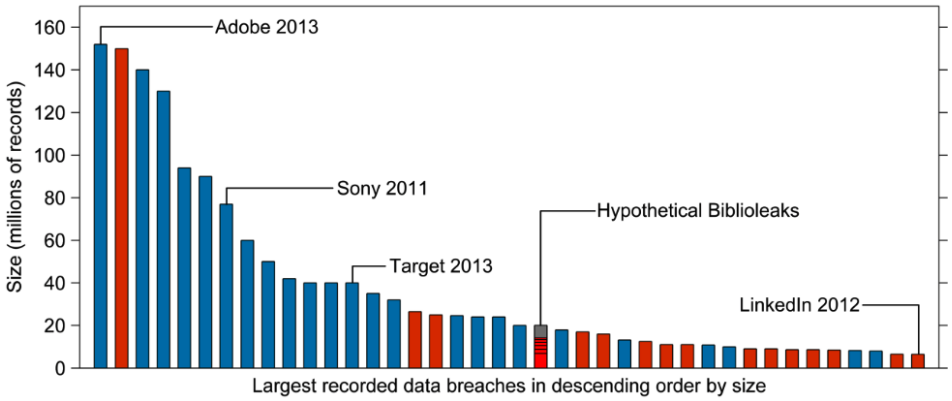
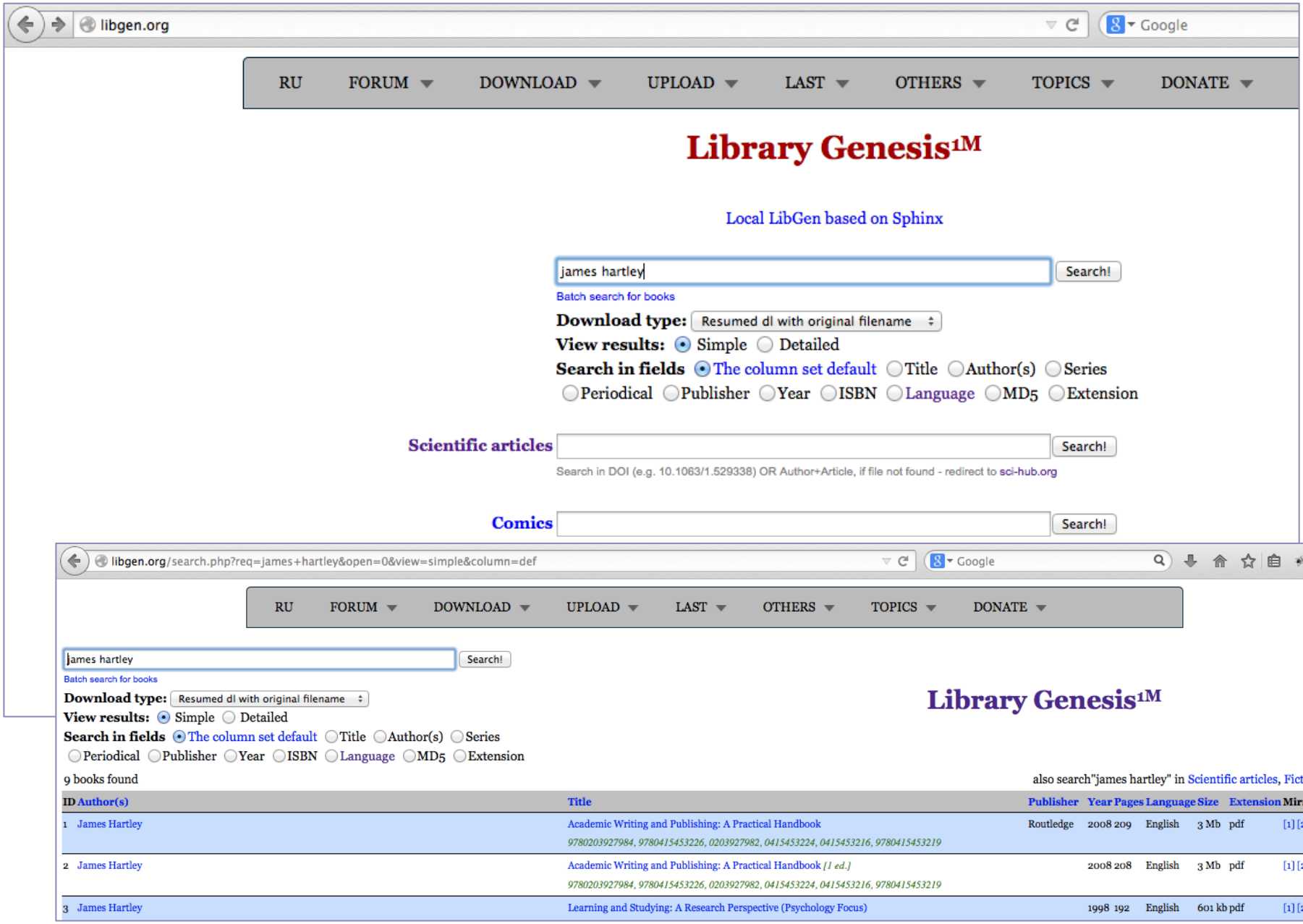
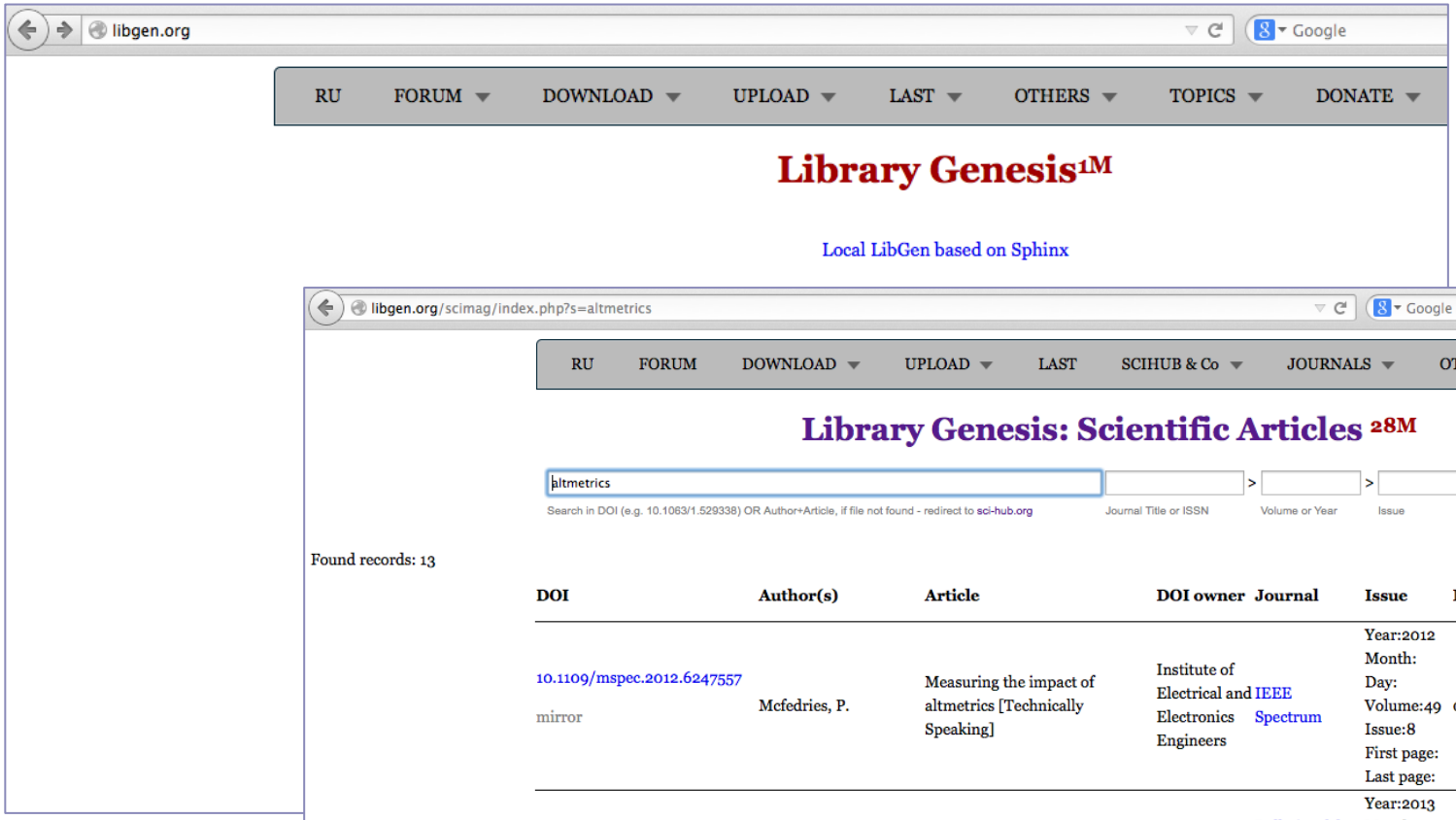


Figure 2. Largest recorded data breaches by number of records (accessed 7 January 2014). Hacks are in blue, all other breach types in orange (eg, stolen/lost disks)–compared to a hypothetical breach equivalent to the numbers of articles indexed by PubMed for which full-text versions require a subscription or payment to access. The proportions associated with the 6 largest publishers (sampled from outgoing PubMed links on 7 January 2014) make up 72% of these inaccessible articles (in red).


libgen.org
Google
RU FORUM DOWNLOAD UPLOAD LAST OTHERS TOPICS DONATE
Library Genesis1M
Local LibGen based on Sphinx
james hartley
Search!
Batch search for books
Download type: Resumed dl with original filename
View results: Simple Detailed
Search in fields The column set default Title Author(s) Series
Periodical Publisher Year ISBN Language MD5 Extension
Scientific articles
Search in DOI (e.g. 10.1063/1.529338) OR Author+Article, if file not found - redirect to sci-hub.org
Comics
libgen.org/search.php?req=james+hartley&open=0&view=simple&column=def
Library Genesis1M
9 books found
also search"james hartley" in Scientific articles, Fiction, Comics
ID Author(s) Title Publisher Year Pages Language Size Extension Mirrors Edit
1 James Hartley Academic Writing and Publishing: A Practical Handbook Routledge 2008 209 English 3 Mb pdf [1] [2] [3] [4] [edit]
9780203927984, 9780415453226, 0203927982, 0415453224, 0415453216, 9780415453219
2 James Hartley Academic Writing and Publishing: A Practical Handbook [1 ed.] 2008 208 English 3 Mb pdf [1] [2] [3] [4] [edit]
9780203927984, 9780415453226, 0203927982, 0415453224, 0415453216, 9780415453219
3 James Hartley Learning and Studying: A Research Perspective (Psychology Focus) 1998 192 English 601 kb pdf [1] [2] [3] [4] [edit]



DOI	Author(s)	Article	DOI owner	Journal	Issue	ISSN	Size, kB	Other
10.1109/mspec.2012.6247557 mirror	Mcfedries, P.	Measuring the impact of altmetrics [Technically Speaking]	Institute of Electrical and Electronics Engineers	IEEE Spectrum	Year:2012 Month: Day: Volume:49 Issue:8 First page: Last page:	0018-9235(p)	526kb	MD5 Errors Edit BibTeX
10.1002/bult.2013.1720390406 mirror	Mounce, Ross	Open access and altmetrics: Distinct but complementary	Wiley Blackwell (John Wiley & Sons)	Bulletin of the American Society for Information Science and Technology	Year:2013 Month:04 Day: Volume:39 Issue:4 First page:14 Last page:17	0095-4403(p) 1550-8366(e)	185kb	MD5 Errors Edit BibTeX Abstract
10.1002/bult.2013.1720390405 mirror	Piwowar, Heather; Priem, Jason	The power of altmetrics on a CV	Wiley Blackwell (John Wiley & Sons)	Bulletin of the American Society for Information Science and Technology	Year:2013 Month:04 Day: Volume:39 Issue:4 First page:10 Last page:13	0095-4403(p) 1550-8366(e)	94kb	MD5 Errors Edit BibTeX Abstract

Biblioleaks + Crowdsourcing = Bibliogifts ?

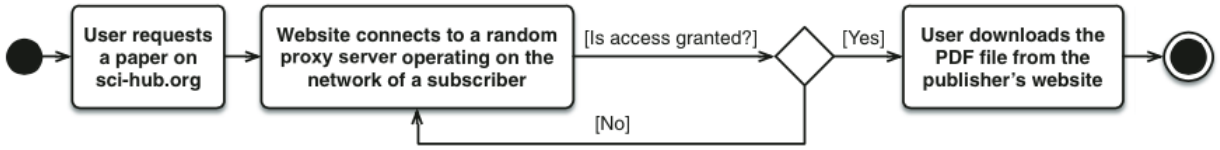
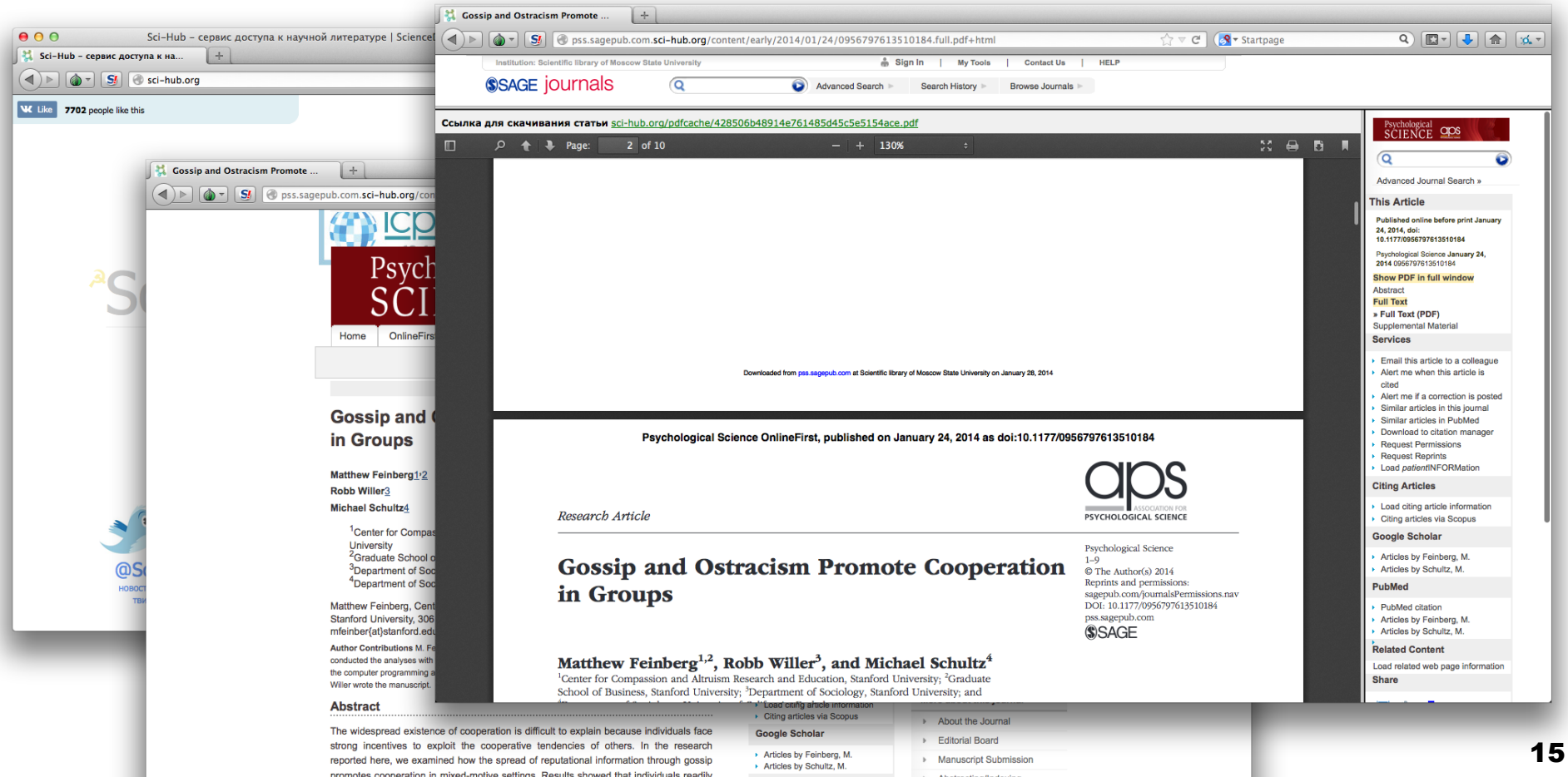
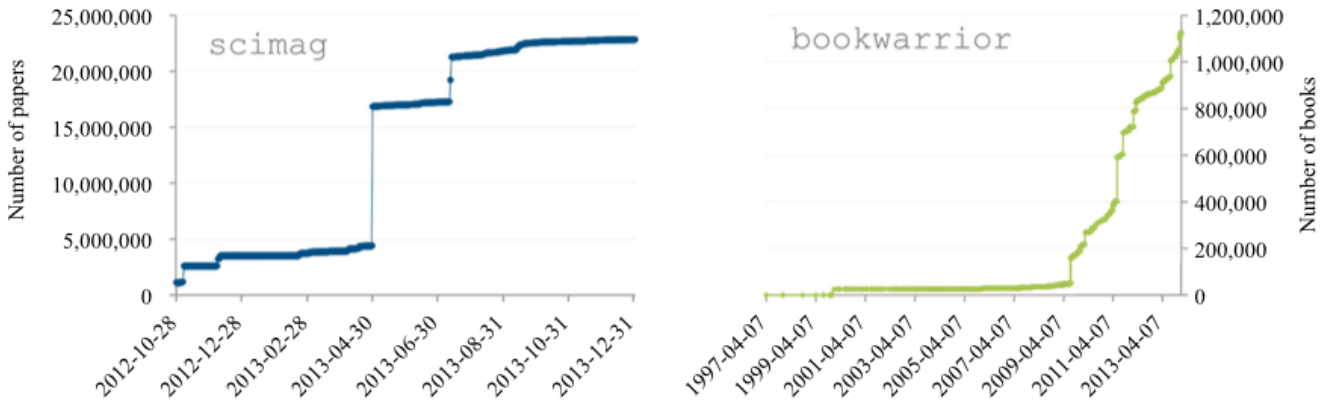


FIG. 9. This activity diagram shows how Sci-Hub serves PDF files of articles through proxy servers operating on the networks of worldwide institutions. The user randomly switches proxy servers until he or she gets connected to a proxy entitled to download the PDF of the requested article.



Biblioleaks + Crowdsourcing = Bibliogifts ?



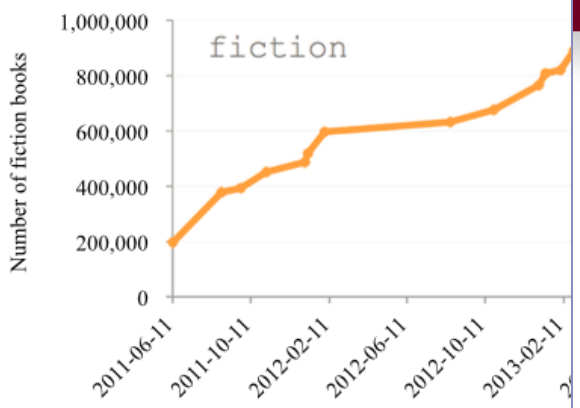


FIG. 1. Growth of the number of educational and recreational mat

JSTOR Evidence in United States vs. Aaron Swartz

OVERVIEW | SUMMARY OF EVENTS | DOCUMENTS

July 30, 2013

JSTOR is a not-for-profit digital library. Our mission is to expand access to scholarly knowledge worldwide and to preserve it for future generations. Between late September 2010 and early January 2011, someone downloaded 4.8 million articles from JSTOR—about 80% of our entire database—using the MIT network. People around the world now know that the person responsible was Aaron Swartz, that the United States government opted to prosecute Mr. Swartz, and that the criminal case came to an end with Mr. Swartz's tragic suicide in January 2013. At the time of the downloading, we had no idea who was responsible.

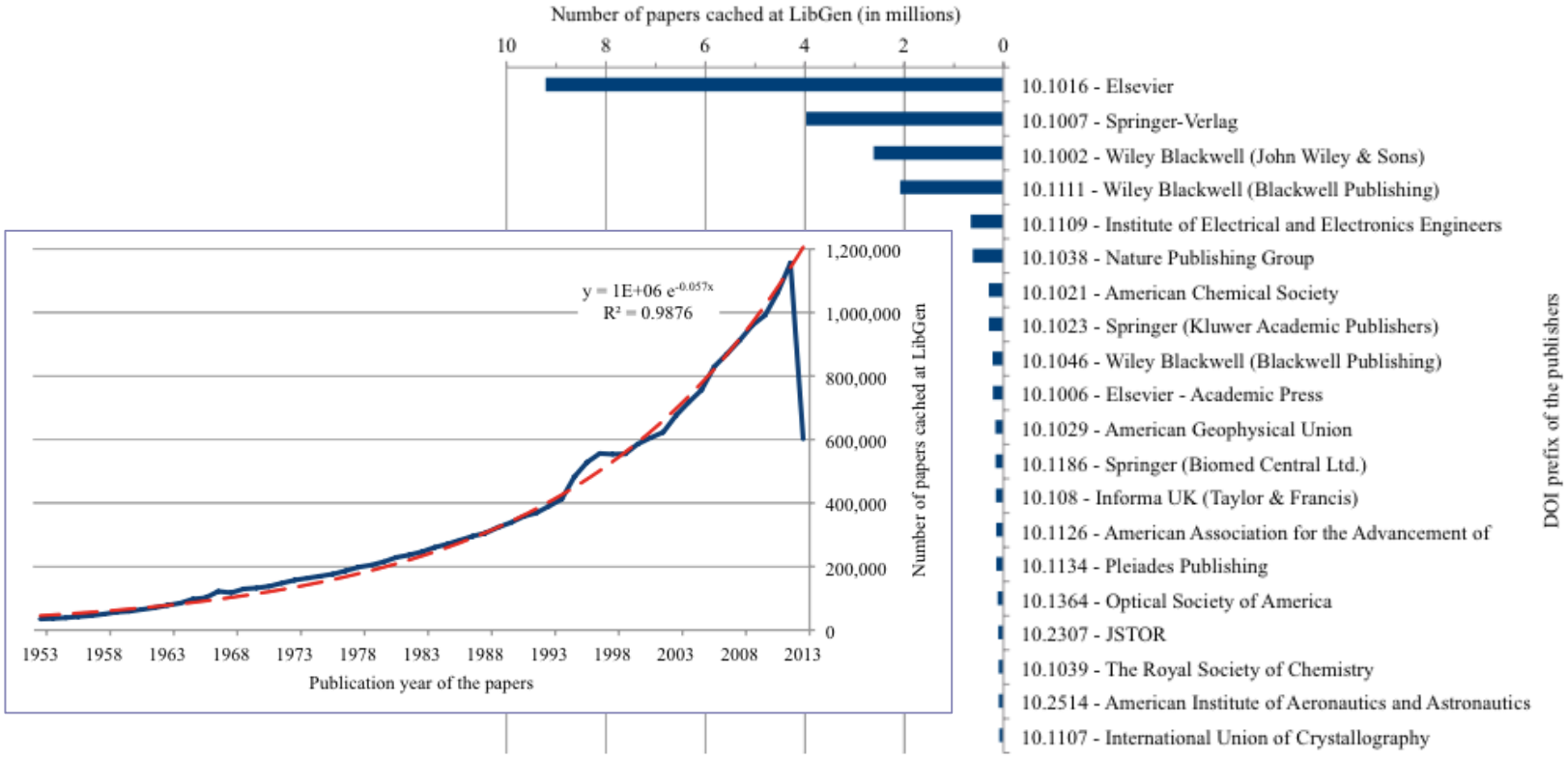
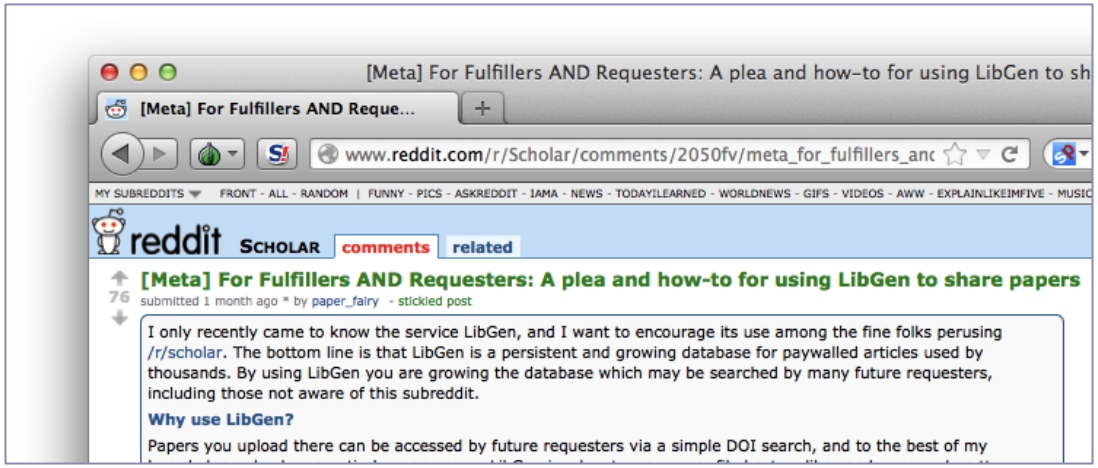


FIG. 3. Distribution of the 22,829,088 articles available from LibGen, for each registrant (DOI prefix and associated publishing house). Only the top 20 publishers in number of cached articles are shown, totaling 94% of all cached articles.

TABLE 4. Number of articles available at LibGen versus registered with a DOI at CrossRef.

Publisher	Number of papers		Coverage of LibGen (%)
	LibGen	CrossRef	
All	22,829,088	63,580,196	36
Elsevier	9,579,795	12,398,807	77
Springer	4,504,256	8,538,817	53
Wiley	4,973,954	6,848,146	73

Bibliogifts as Rogue Open Access?



A**** C******** @t******** 22h

In #icanhazpdf tweets: inc. link, lead author, date, your email address. Perhaps field too (e.g. 'physics'). Delete request once fulfilled.

Expand Reply Retweet Favourite More

A** M******* @A********** 26 Nov

#icanhazpdf "@K*********: "Symbiogenesis: Mechanisms, Evolutionary Consequences, and Systematic Implications" annualreviews.org/doi/abs/10.114..."

ACRL 2015 March 25–28, 2015, Portland, Oregon

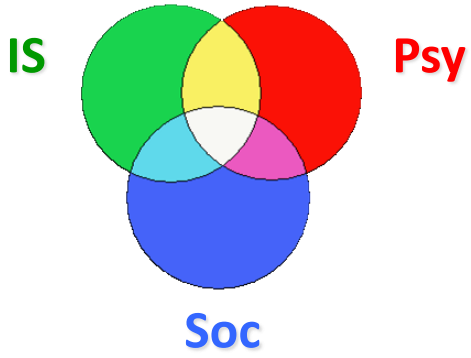
Bypassing Interlibrary Loan Via Twitter: An Exploration of #icanhazpdf Requests

Carolyn Caffrey Gardner and Gabriel J. Gardner

FIG. 10. Example of two tweets published with the #icanhazpdf hashtag on Twitter (user accounts are anonymized). The first tweet recalls informal guidelines when requesting articles. The second tweet requests an article by mentioning its title and hyperlink.

Khabsa M, Giles CL (2014) The Number of Scholarly Documents on the Public Web. PLoS ONE 9(5): e93949. doi:10.1371/journal.pone.0093949

27 million (24%) are freely available since they do not require a subscription or payment of any kind.



Psychology of Science

- Cognitive biases
- Academic writing

Study 3

Order Bias in Conference Bids

Cabanac, G., & **Preuss**, T. (2013). Capitalizing on order effects in the bids of peer-reviewed conferences to secure reviews by expert referees. *Journal of the American Society for Information Science and Technology*, *64*, 2, 405–415. doi:10.1002/asi.22747

PUBLIC OPINION QUARTERLY, FALL, 1954

Why an Order Effect

By SAM L. BECKER

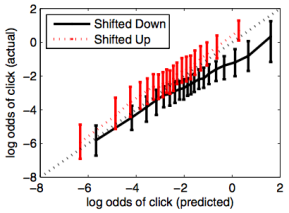
The author of this article contends that determining the effect of position on an item in a check-list question is much more complicated than is ordinarily assumed. Among the factors which he thinks must be considered are; the setting of the interview, the length of the list, and the length of the questionnaire.

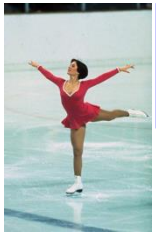
Sam. L. Becker is Associate Professor in the Radio-TV-Film Division of the Speech Department at the University of Iowa.

(Figure 1, it would appear that the later on a check-list that a program type is listed, the less the chance that someone will select it as one of his five favorites. In general, the chances of a program type being chosen as a "favorite" seem to improve steadily as the program type moves from the sixteenth to the first position on the check-list.

Miller, J.M., & Krosnick, J.A. (1998). The impact of candidate name order on election outcomes. Public Opinion Quarterly, 62(3), 291–330. doi:10.1086/297848

Craswell, N., Zoeter, O., Taylor, M., & Ramsey, B. (2008). An experimental comparison of click position-bias models. WSDM ('08): Proceedings of the first ACM international conference on web search and data mining (pp. 87–94). New York: ACM.





Bruine de Bruin, W. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. Acta Psychologica, 118(3), 245–260. doi:10.1016/j.actpsy.2004.08.005

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. Psychological Science, 20(11), 1309–1312. doi:10.1111/j.1467-9280.2009.02453.x

Conferences Affected by a Submission-Date bias?

- Peer-review

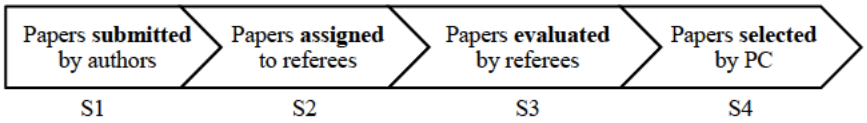


Fig. 1 Stages of the peer review process for conference paper selection

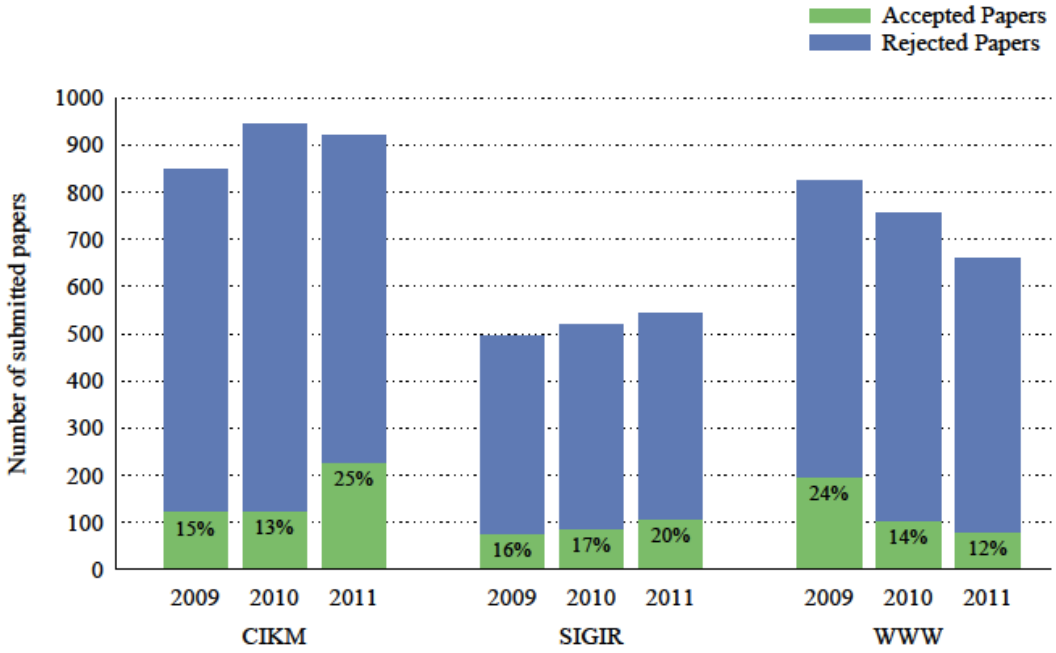


Fig. 2 Number of papers submitted to three leading Computer Science conferences held between 2009 and 2011. Accepted papers are shown in green, while rejected papers are shown in blue. (Source: the ACM Digital Library http://dl.acm.org)

The Submission-Date bias

- Dataset from the ConfMaster conference management system

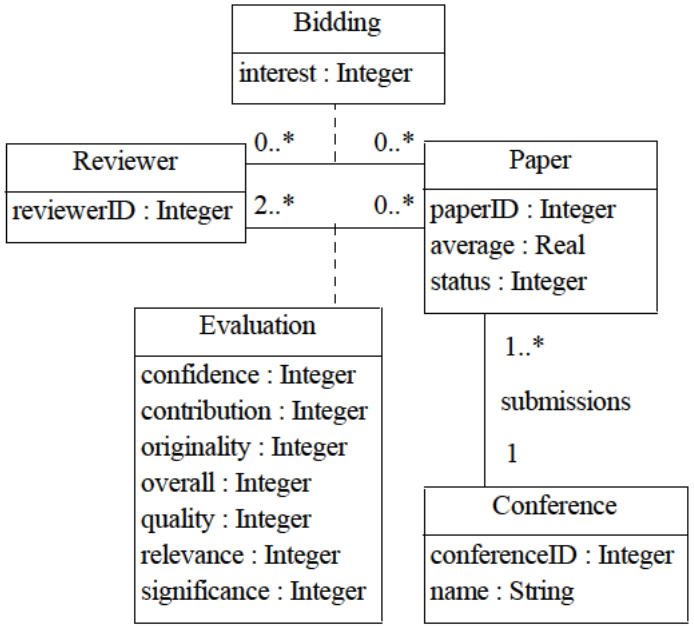


Fig. 3 UML class diagram modeling data generated by peer-reviewed conferences supported by a bidding process

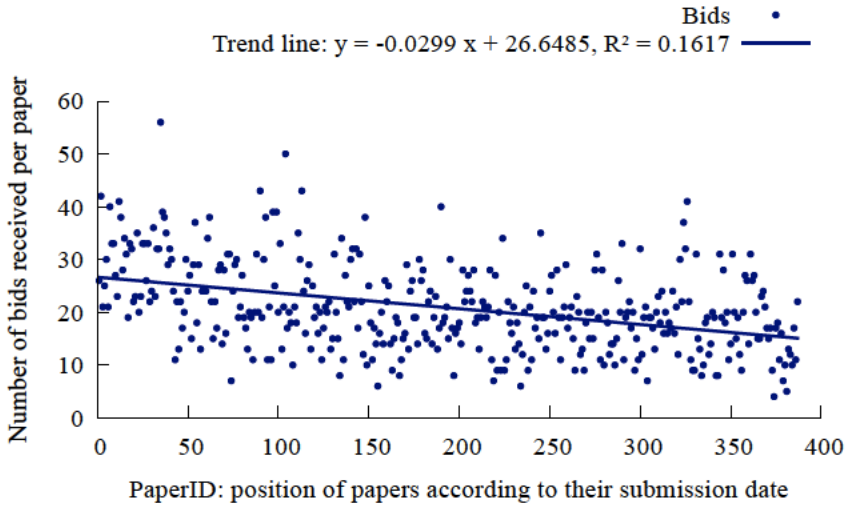


Fig. 4 Scatter plot for conference number 3903 (see Appendix A) showing the number of bids received per paper position, as well as the trend line for the data points (linear least squares regression with associated coefficient of determination R^2)

The Submission-Date bias

- Influence of submission date on bids

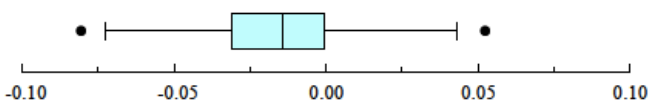


Fig. 8 Box plot of trend line slopes for paper bids (see Figure 7)

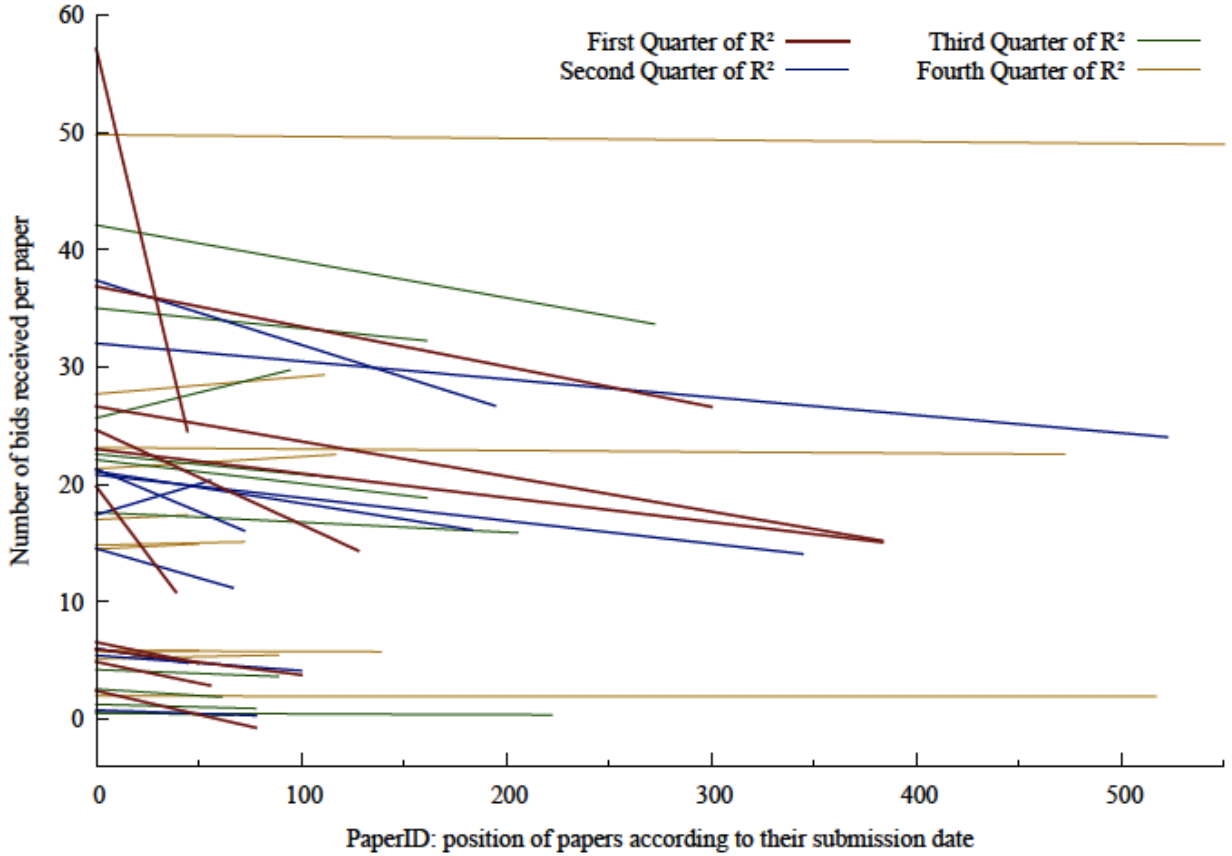


Fig. 7 Trend lines for bids given to papers submitted to the 42 peer-reviewed conferences. Line width is proportional to goodness of fit (R^2). Note that the x axis is cut at position 550 for readability concerns (only one conference has 831 papers).

Referee Assignment Matters

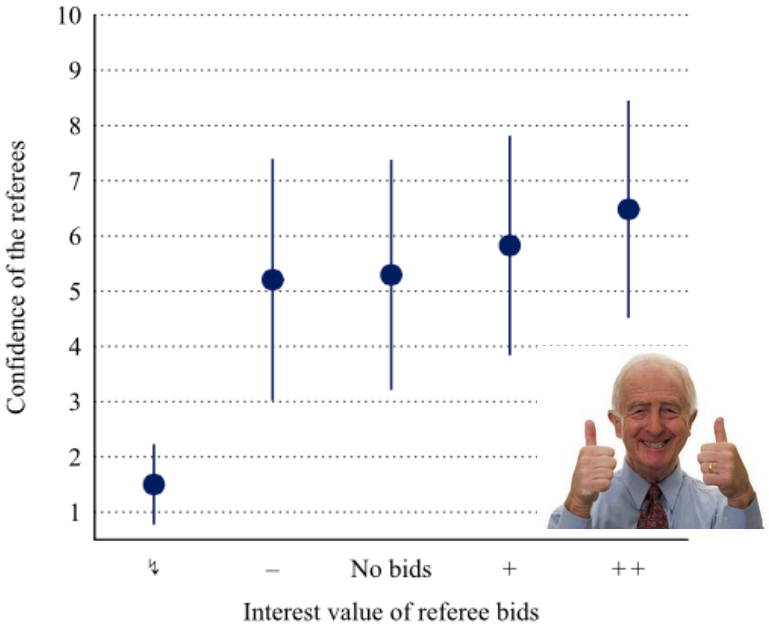


FIG. 9. Confidence of referees as stated in the reviews they made for papers they bid on (with a given interest value) and were assigned to. The mean (μ) and the standard deviation (σ) of confidence values are showed by bullets (μ) and bars (μ ± σ).

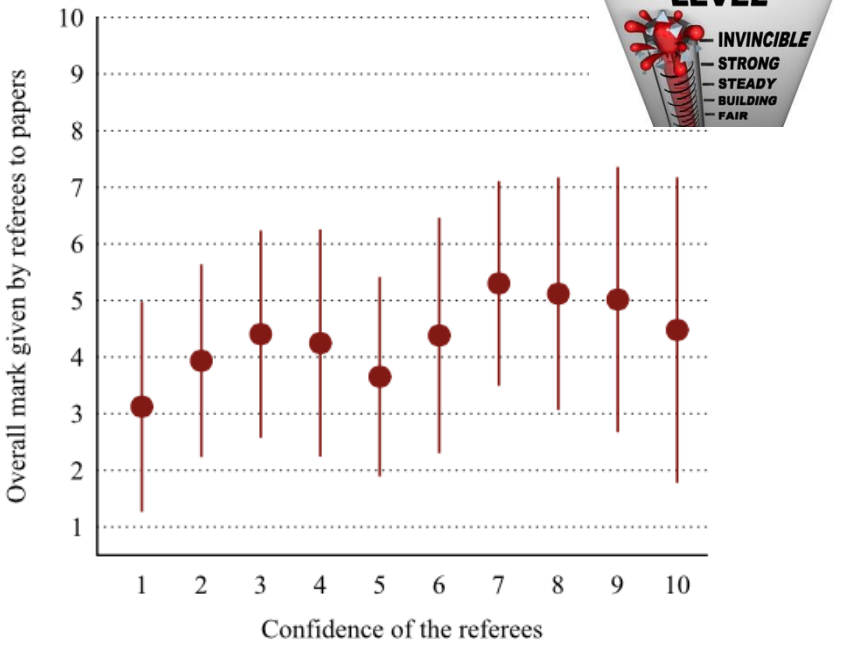
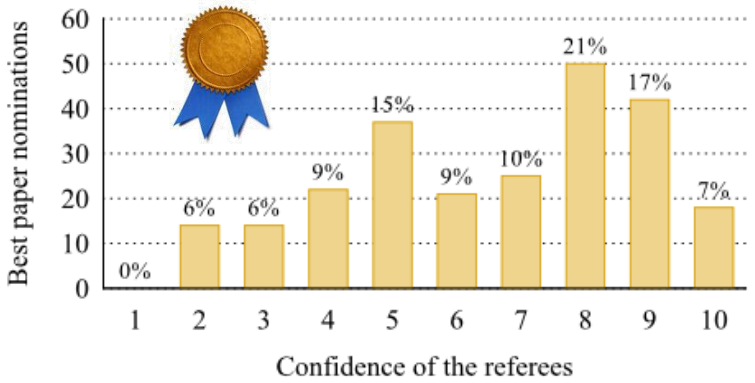
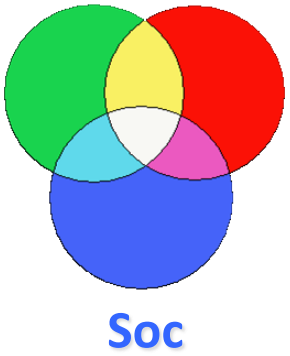


FIG. 12. Overall mark given to papers by referees according to their confidence ($N = 19{,}108$ reviews). The mean (μ) and the standard deviation (σ) of overall mark values are showed by bullets (μ) and bars (μ ± σ).

FIG. 13. Distribution of the 243 best paper nominations (among 19,108 reviews) with respect to referee confidence.



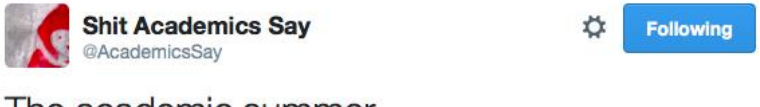


Psychology of Science

- Cognitive biases
- Academic writing

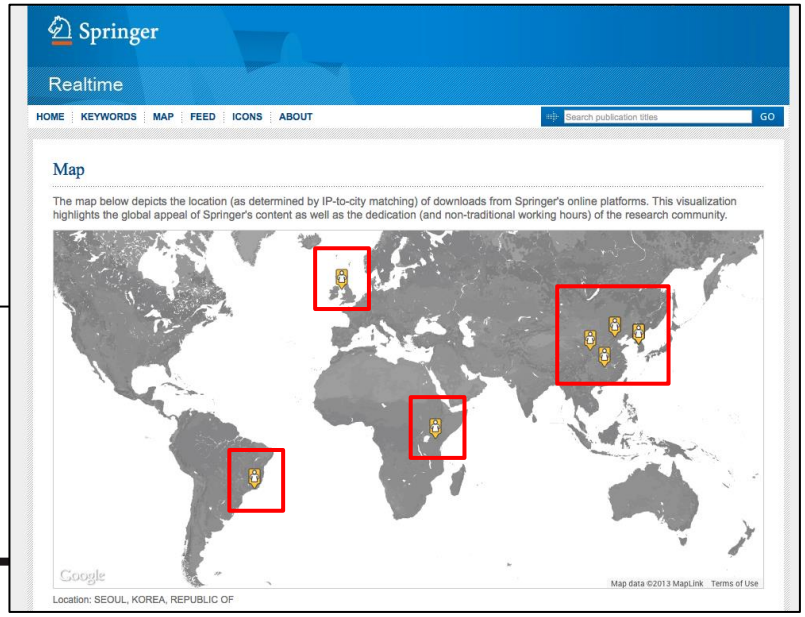
Study 4

Issues of Work-Life Balance



Cabanac, G., & **Hartley**, J. (2013). Issues of work-life balance among *JASIST* authors and editors [Brief Communication]. *Journal of the American Society for Information Science and Technology*, *64*, 10, 2182–2186. doi:10.1002/asi.22888

Issues of Work-life Balance



Journal of Informetrics 6 (2012) 655–660

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Exploring scientists' working timetable: Do scientists often work overtime?

Xianwen Wang[a,b,c,*], Shenmeng Xu[a,b,1], Lian Peng[a,b,1], Zhi Wang[a,b,1], Chuanli Wang[a,b,1], Chunbo Zhang[a,b], Xianbing Wang[d]

[a] WISE Lab, Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian 116085, China
[b] School of Public Administration and Law, Dalian University of Technology, Dalian 116085, China
[c] DUT-Drexel Joint Institute for the Study of Knowledge Visualization and Scientific Discovery, Dalian University of Technology, Dalian 116085, China
[d] School of Control Science and Engineering, Dalian University of Technology, Dalian 116085, China

ARTICLE INFO

Article history:
Received 7 June 2012
Received in revised form 9 July 2012
Accepted 9 July 2012

Keywords:
Scientist
Realtime
Working habits
Springer
Work–family conflict

ABSTRACT

A novel method is proposed to monitor and record scientists' working timetable. We record the downloads information of scientific papers real-timely from Springer round the clock, and try to explore scientists' working habits. As our observation demonstrates, many scientists are still engaged in their research after working hours every day. Many of them work far into the night, even till next morning. In addition, research work also intrudes into their weekends. Different working time patterns are revealed. In the US, overnight work is more prevalent among scientists, while Chinese scientists mostly have busy weekends with their scientific research.

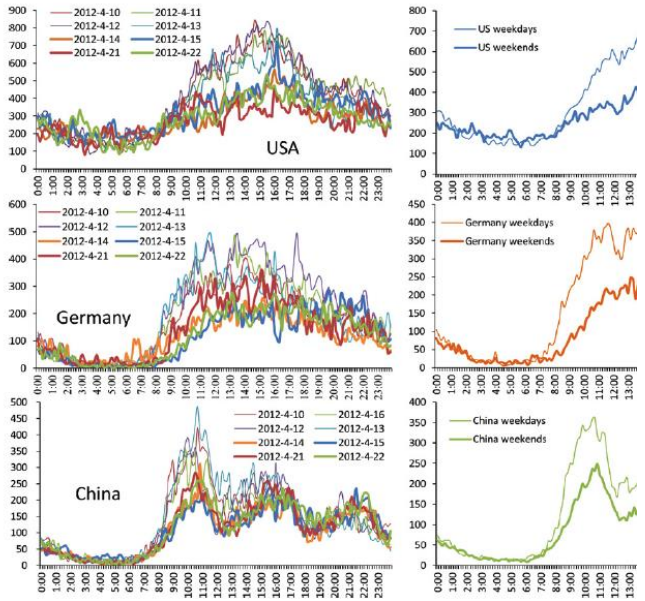


Fig. 4. Downloads of the top 3 countries during weekdays and weekends.

Issues of Work-life Balance

CORRESPONDENCE
NATURE|Vol 450|20/27 December 2007

Come all ye scientists, busy and exhausted. O come ye, O come ye, out of the lab

Richard J. Ladle†, Ana C. M. Malhado†, Peter A. Todd*
† Oxford University Centre for the Environment, South Parks Road, Oxford OX1 3QY, UK
*Department of Biological Sciences, National University of Singapore, 117543 Singapore

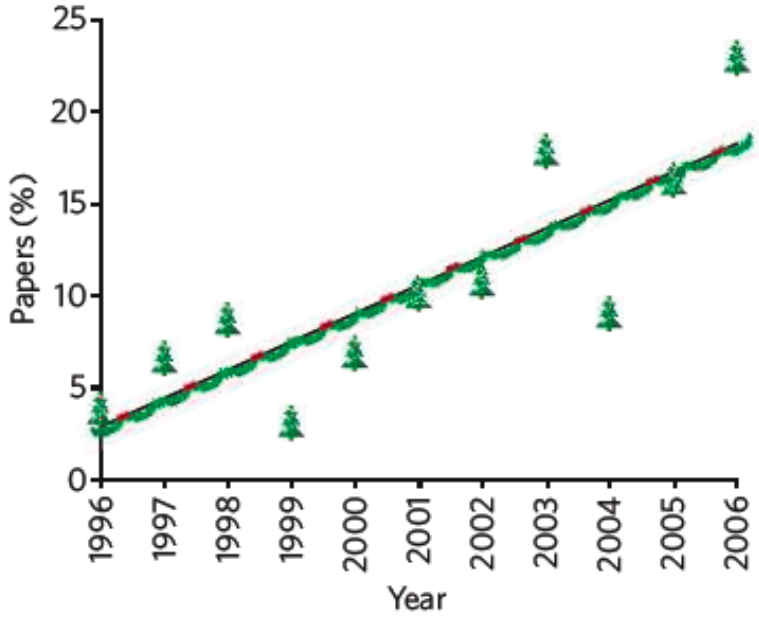


Figure 1 | Can't see the wood for the trees? Take a break. Proportion of published papers submitted on 25 December relative to mean number submitted on the 25th of the month (excluding weekends) for all other months in that calendar year. $R^2 = 0.69$.

Issues of Work-life Balance

Using the shape recovery method to evaluate indexing techniques

Guillermo Oyarce

Article first published online: 14 MAY 2008

DOI: 10.1002/asi.20830

Issue

Journal of the American Society for Information Science and Technology

Volume 59, Issue 9, pages 1479–1492, July 2008

Additional Information (Show All)

How to Cite | Author Information | Publication History

Author Information

University of North Texas, P.O. Box 276, Denton, TX 76202

Email: Guillermo Oyarce (oyga@unt.edu; goyarce@gmail.com)

Publication History

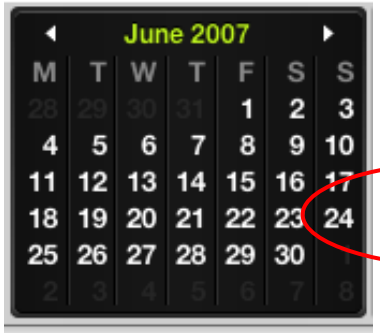
Issue published online: 9 JUN 2008

Article first published online: 14 MAY 2008

Manuscript Accepted: 1 JAN 2008

Manuscript Revised: 31 DEC 2007

Manuscript Received: 24 JUN 2007



Sunday!

Issues of Work-life Balance

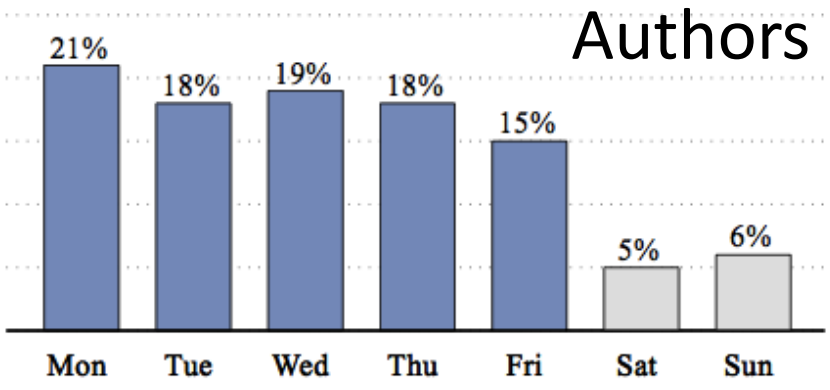
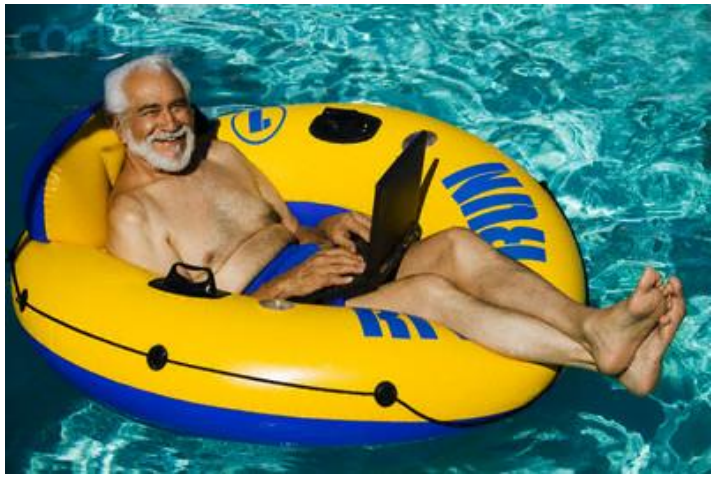


FIG. 1. Percentages of new submissions posted by authors by days of the week (percentages are rounded).



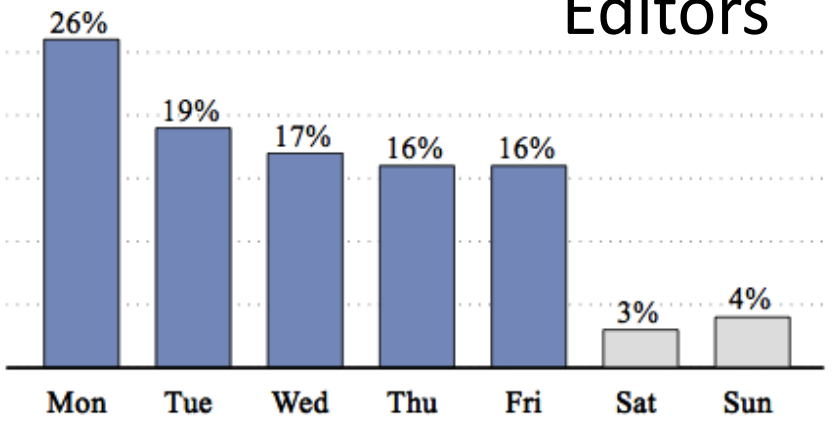


FIG. 3. Composite data from two *JASIST* editors showing the percentages of manuscript acceptances by days of the week (percentages are rounded).

Issues of Work-life Balance

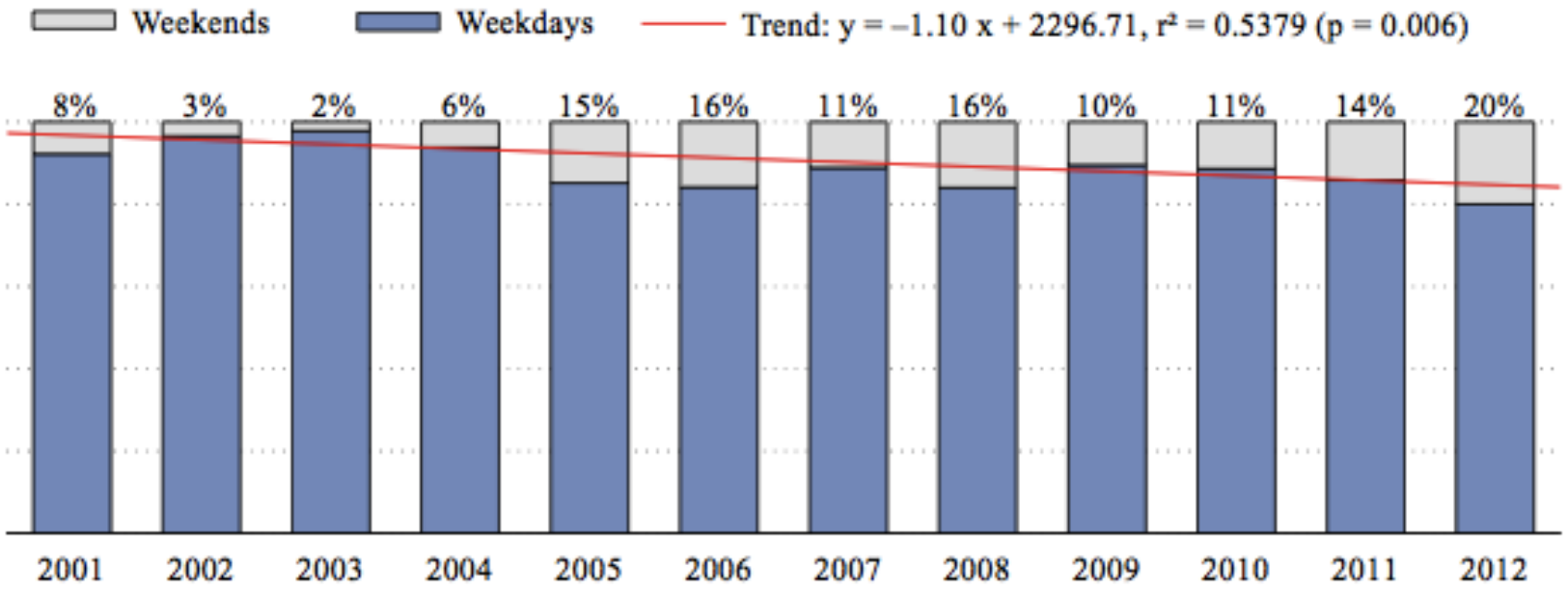
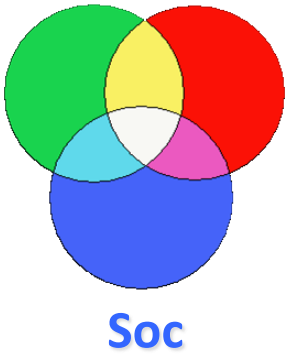


FIG. 4. Evolution of the percentages of new submissions and final submissions posted by authors during weekdays and weekends between 2001 and 2012.

We hope that our findings will raise an awareness of the steady increases in work among scientists before it affects our work–life balance even more.

[...] raise readers' awareness of these problems and how they might affect them. But it is hard not to forget that, for some:

Work is play when it's something you like.
Andy Warhol (1928–1987)



Psychology of Science

- Cognitive biases
- Academic writing

Studies 5 and 6

Differences in Academic Writing:

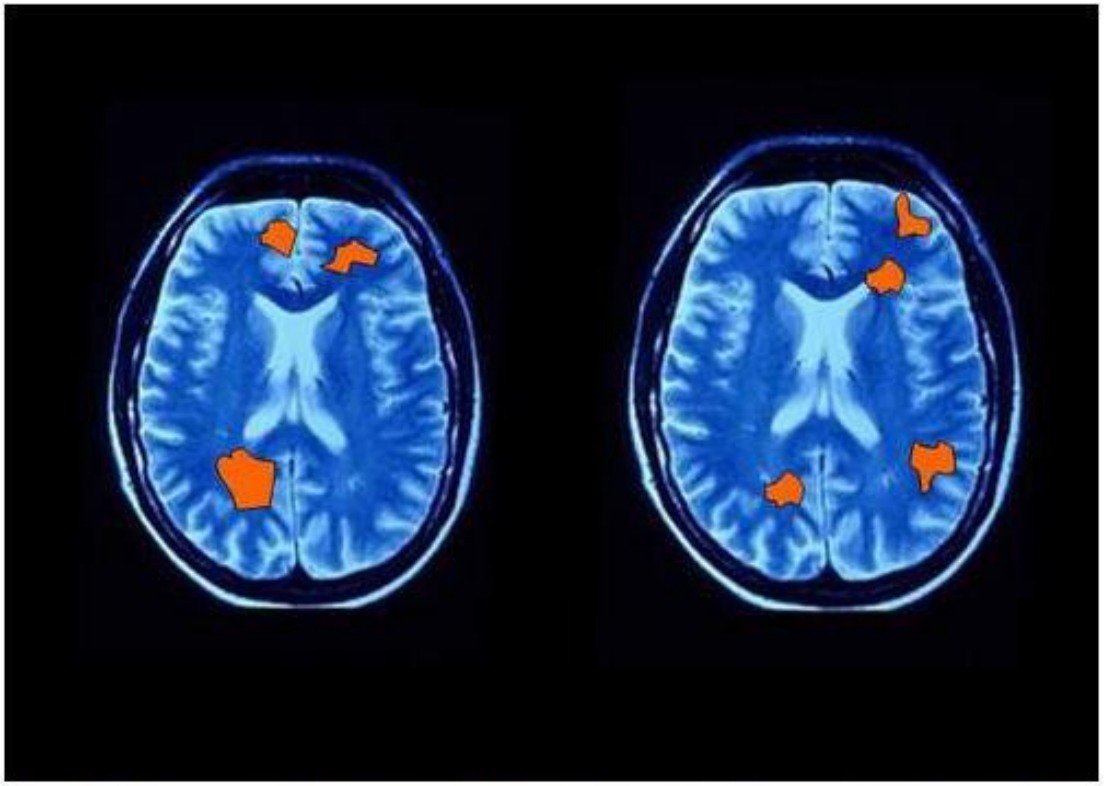
Gender & Solo/Multiauthor

Hartley, J., & **Cabanac**, G. (2014). Do men and women differ in their use of tables and graphs in academic publications? *Scientometrics*, *98*, 2, 1161-1172. doi:10.1007/s11192-013-1096-3

Cabanac, G., **Hubert**, G., & **Hartley**, J. (2014). Solo versus collaborative writing: Discrepancies in the use of tables and graphs in academic articles. *Journal of the American Society for Information Science and Technology*, *65*, 4, 812–820. doi:10.1002/asi.23014

Academic Writing – Gender Study

- Theory in Psychology (1960's)
  - **Men** are more **spatially** and **mathematically** oriented than women
  - **Women** are more **verbally** oriented than men



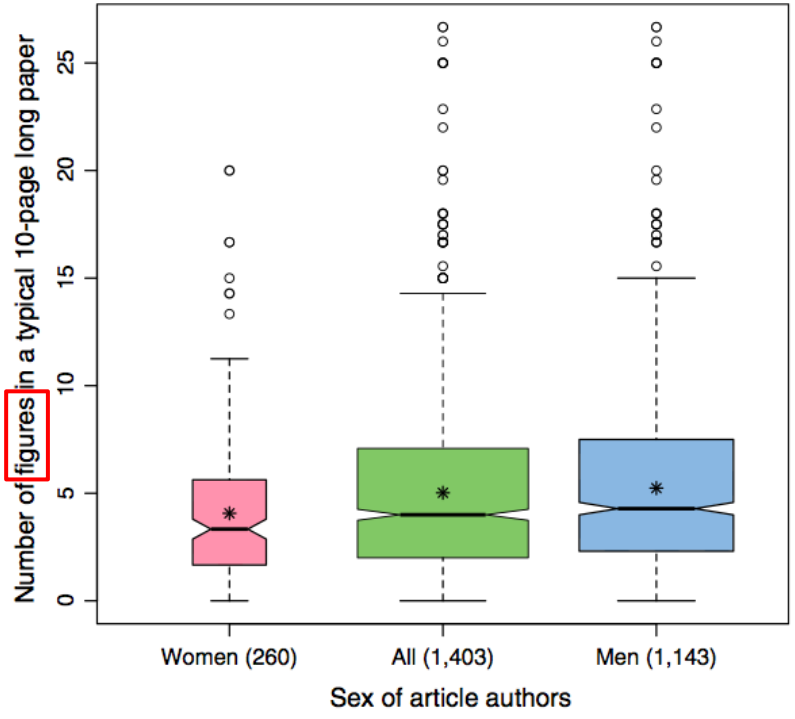
http://www.kaheel7.com/eng/images/stories/2(5).jpg

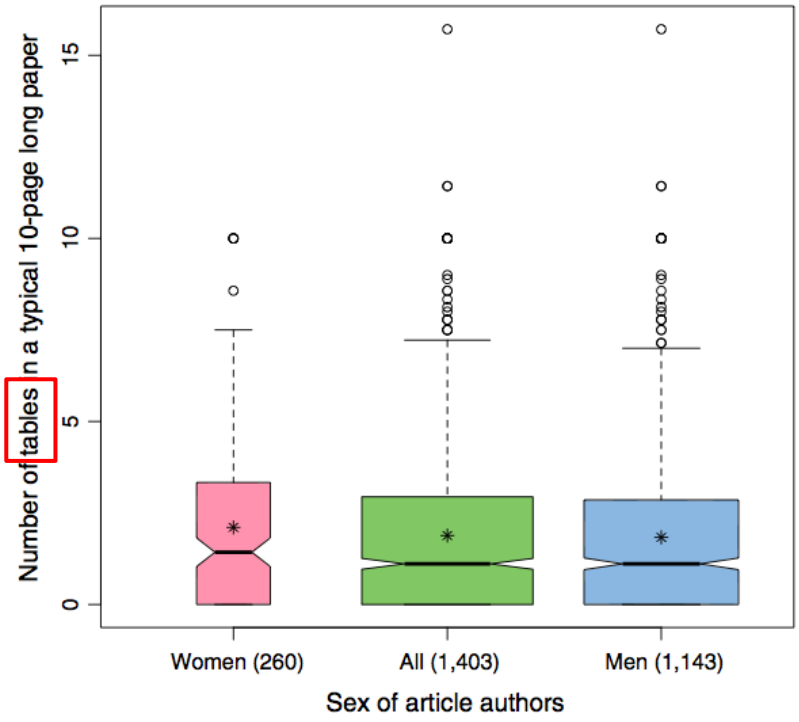
When both men and women are performing the same task, different areas
In theirs brain activation

Academic Writing – Gender Study

- Our findings on 1,403 single-authored articles in science
  - Men use **26% more figures** than Women ($p < 0.001$)
  - Men use **11% more tables** than Women ($p = 0.102$)

… but is the difference practical?





Academic Writing – Study of Collaborations

- The "Friday" Hypothesis
  - When writing in groups, it is harder to agree on text than on figures/graphs
  - More Figures and Graphs in multi-author papers?

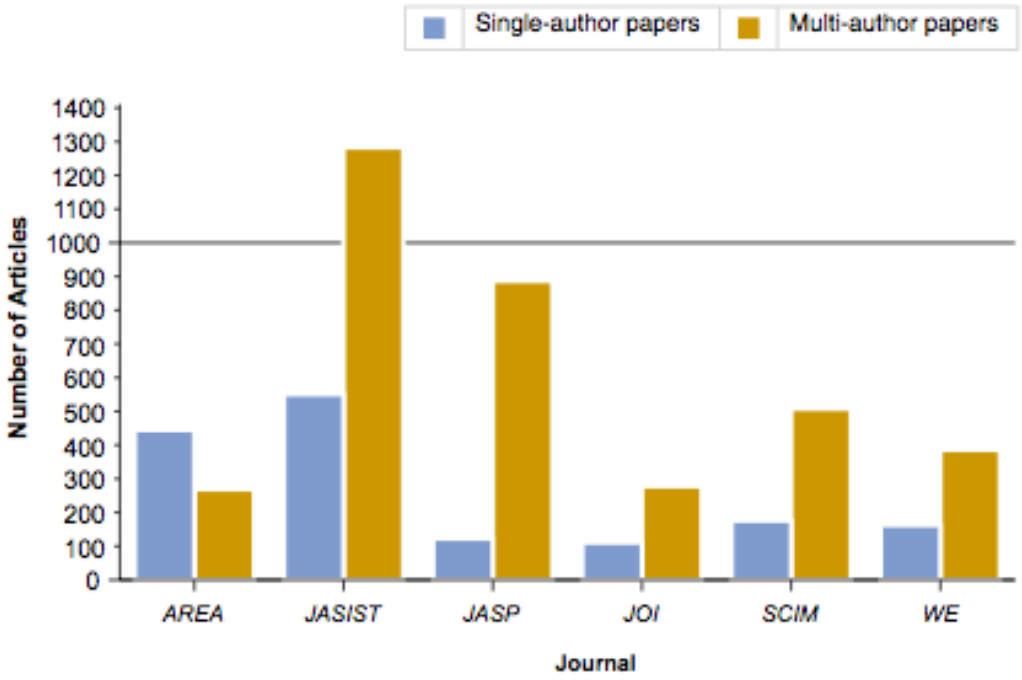


FIG. 1. Distribution of single-author and multiauthor articles for the six journals under study. All journals but one publish more multiauthor articles than single-author articles (between 12% and 30%). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Academic Writing – Study of Collaborations

- More tables in multi-author vs. single-author papers

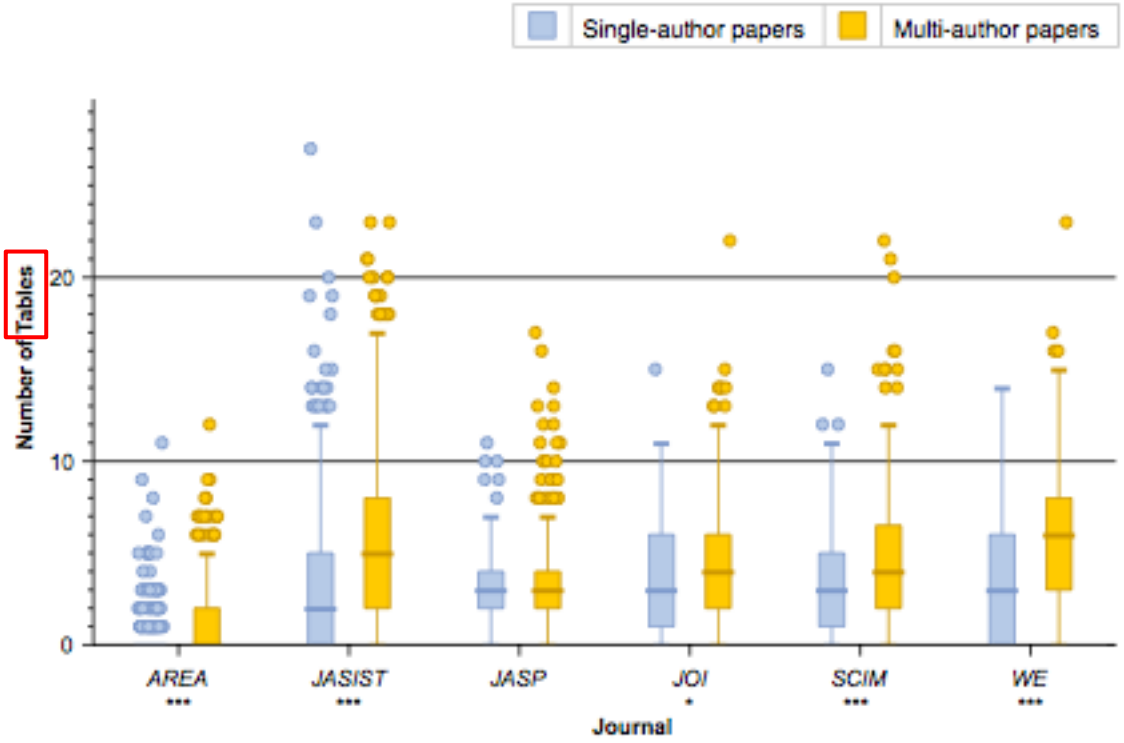


FIG. 3. These box plots show the number of tables in single-author versus multiauthor articles. Visual inspection and significance tests (*$p < .05$, **$p < .01$, ***$p < .001$) show that there are more tables in multiauthor articles when compared with single-author articles for five of six journals. H1 is thus supported.

Academic Writing – Study of Collaborations

- More figures in multi-author vs. single-author papers

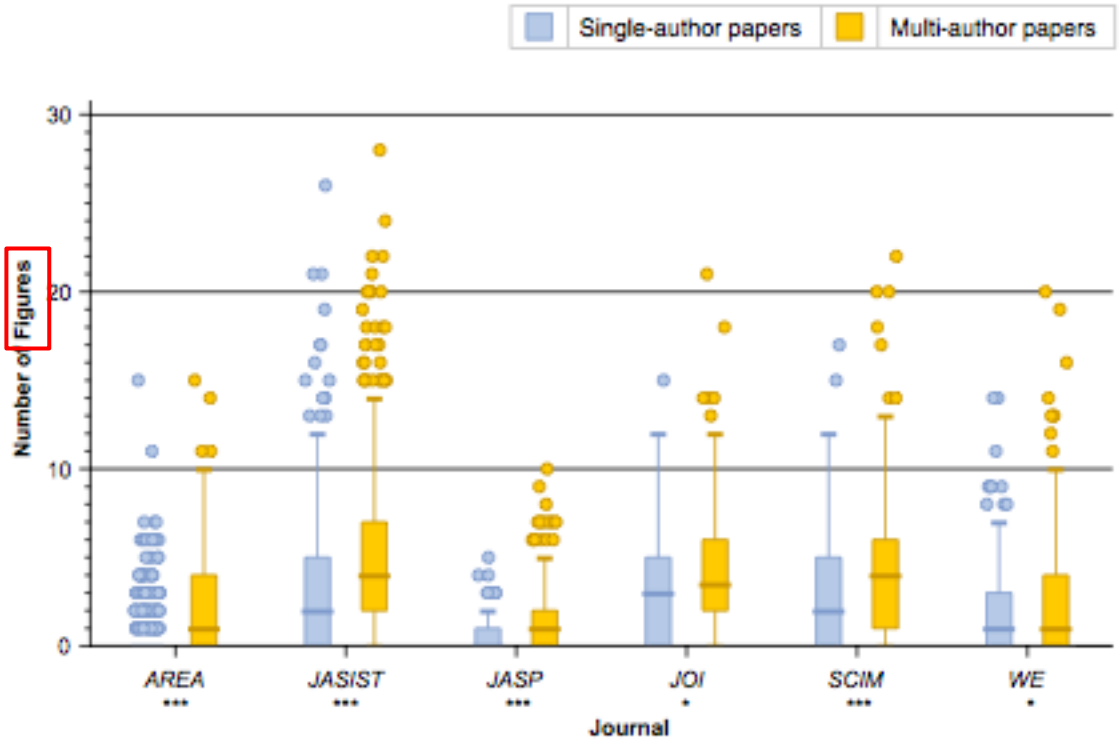
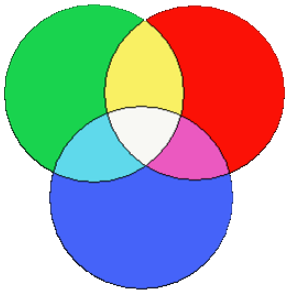


FIG. 5. These box plots show the number of figures in single-author versus multiauthor articles. Visual inspection and significance tests (*$p < .05$, **$p < .01$, ***$p < .001$) show that there are more figures in multiauthor articles when compared with single-author articles for all journals. H3 is thus supported. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



Sociology of Science

- Gatekeeping
- Collaboration

Study 7

Lifelong Publication Dynamics in CS

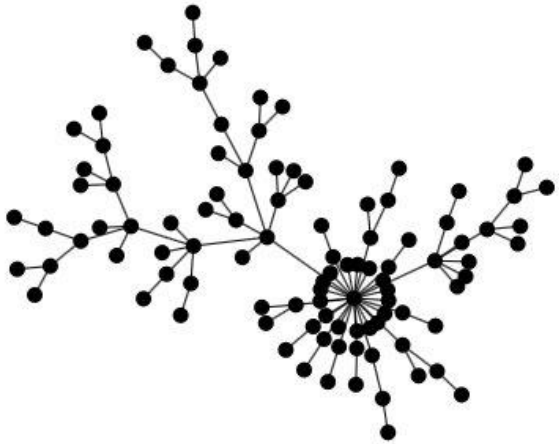
Cabanac, G., **Hubert**, G., & **Milard**, B. (2015). Academic careers in Computer Science: continuance and transience of lifetime co-authorships. Scientometrics, *102*, 1, 135–150. doi:10.1007/s11192-014-1426-0

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected. A model based on these two ingredients reproduces the observed stationary scale-free distributions, which indicates that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.

SCIENCE VOL 286 15 OCTOBER 1999



Physica A 311 (2002) 590–614

Evolution of the social network of scientific collaborations

A.L. Barabási[a,b,*], H. Jeong[a], Z. Néda[a,b,c], E. Ravasz[a], A. Schubert[d], T. Vicsek[b,e]

these quantities. The results indicate that the network is scale-free, and that the network evolution is governed by preferential attachment, affecting both internal and external links. However, in

(i) *New nodes*: For a new author, that appears for the first time on a publication, preferential attachment has a simple meaning. It is more likely that the first paper will be co-authored with somebody that already has a large number of co-authors (links) that with somebody less connected. As a result "old" authors with more links will increase

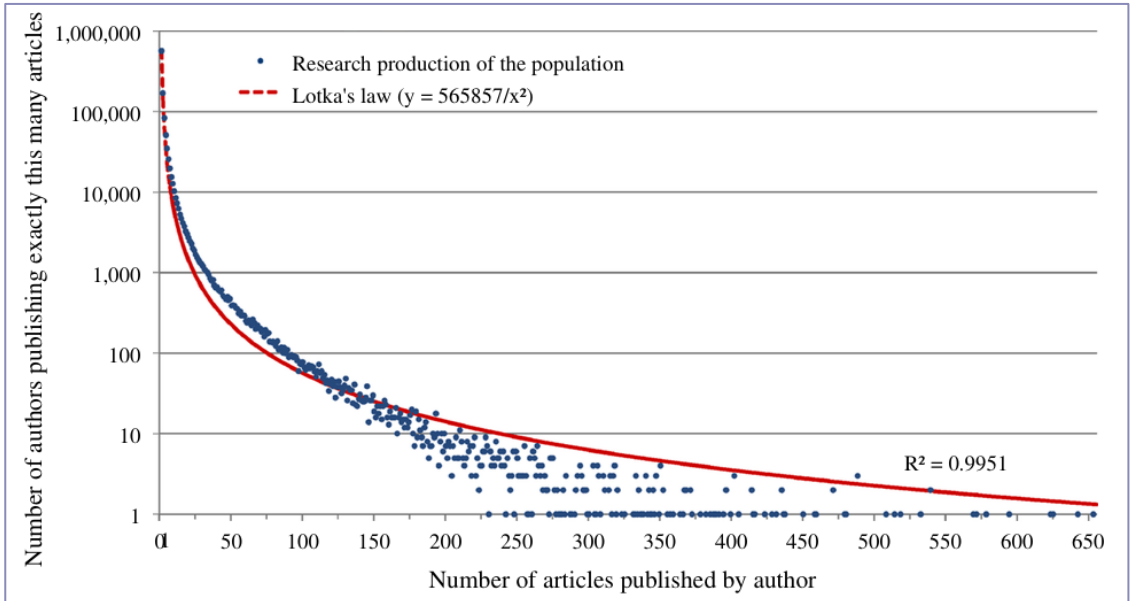


Fig. 2 Research production of the population of articles ($N = 1{,}870{,}054$), as recorded in the DBLP

The population

Source: https://projects.groept.be/~emedia

The sampled "fiftysomethings"



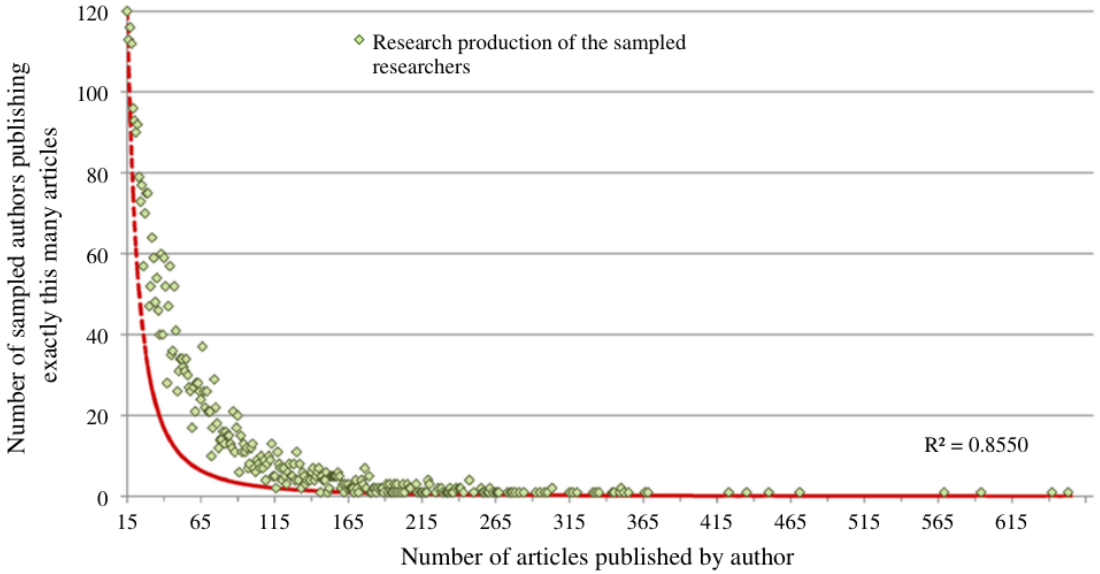


Fig. 4 Research production of the 3,860 sampled researchers, in terms of conference papers and journal articles ($N = 209{,}377$), as recorded in the DBLP. This distribution fits the Lotka's law ($R^2 = 0.8550$)

Transience and Continuance

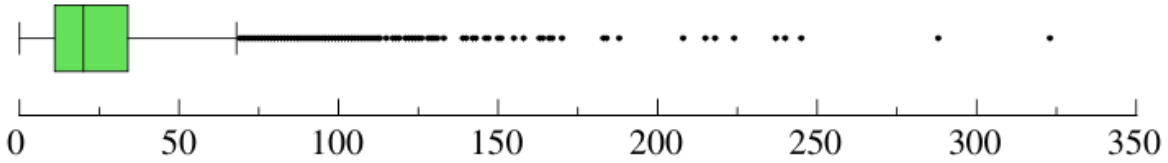


Fig. 5 Average number of co-authors for the researchers under study ($N = 3{,}860$)

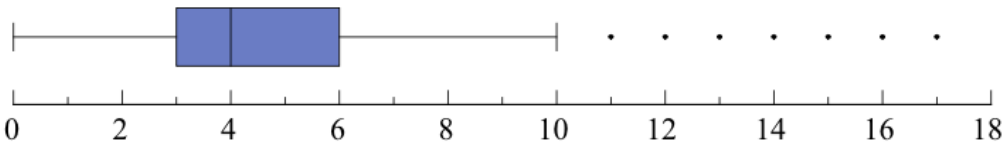


Fig. 6 Partnership ability φ-index of the 3,860 researchers under study showing the distribution of the number of φ co-authors per researcher

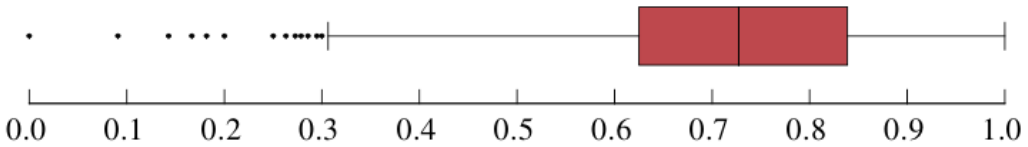


Fig. 7 Average share of transient co-authors per researcher under study who had collaborators ($N = 3,830$)

Renewal of Collaborations

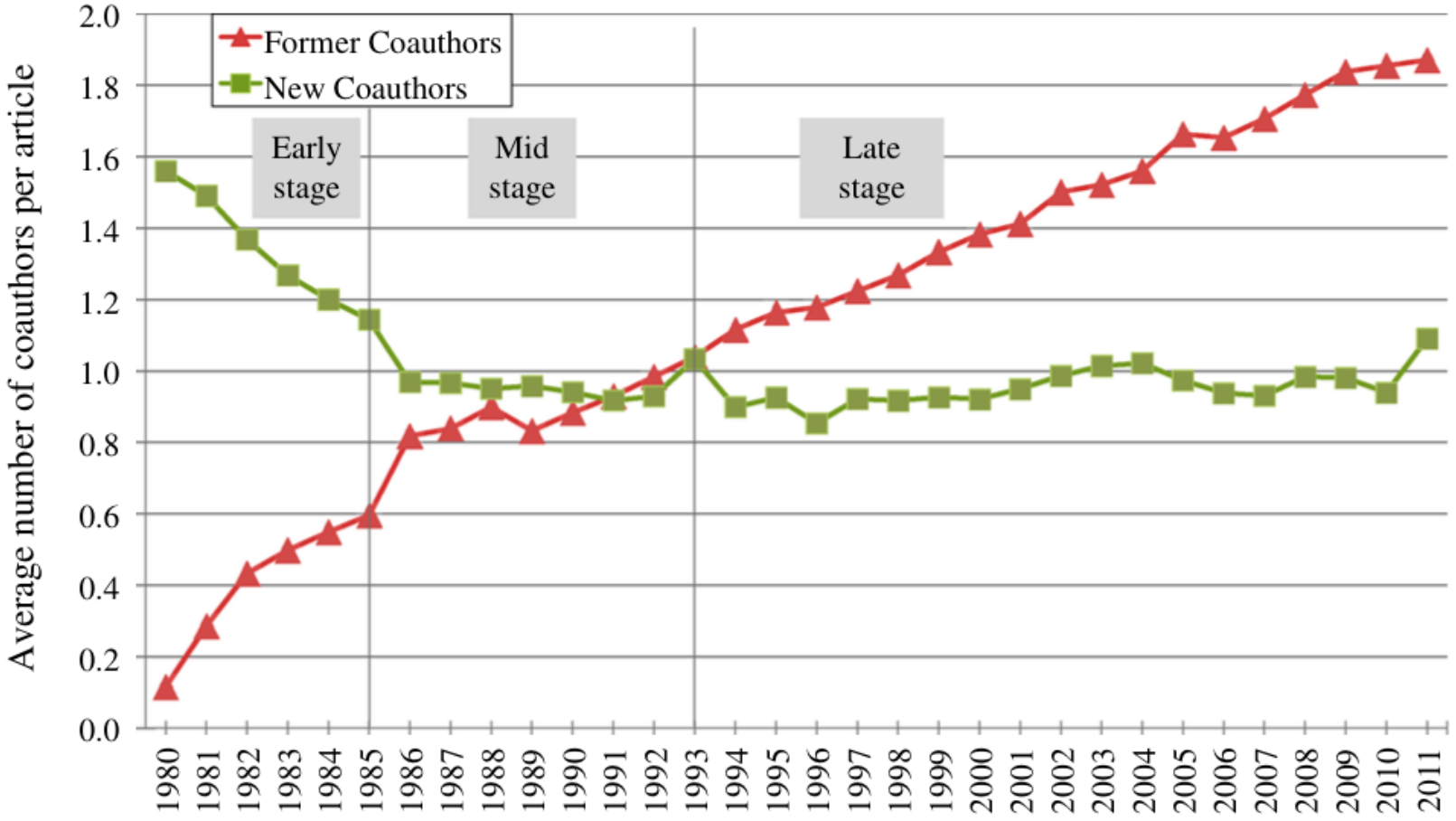


Fig. 10 Evolution of the balance between Former and New co-authors for the 3,860 researchers under study

Success Breads Success and Homophily

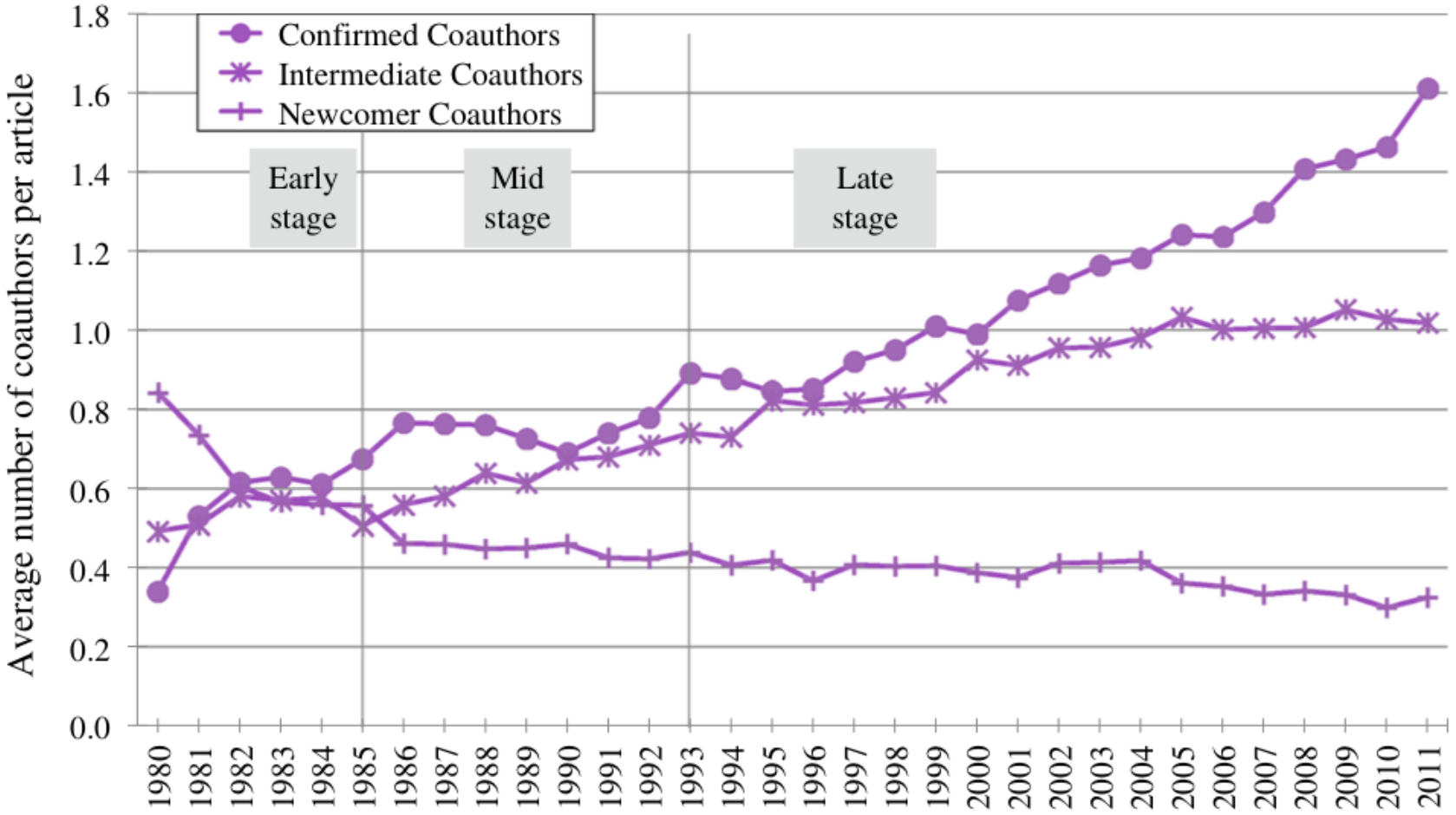
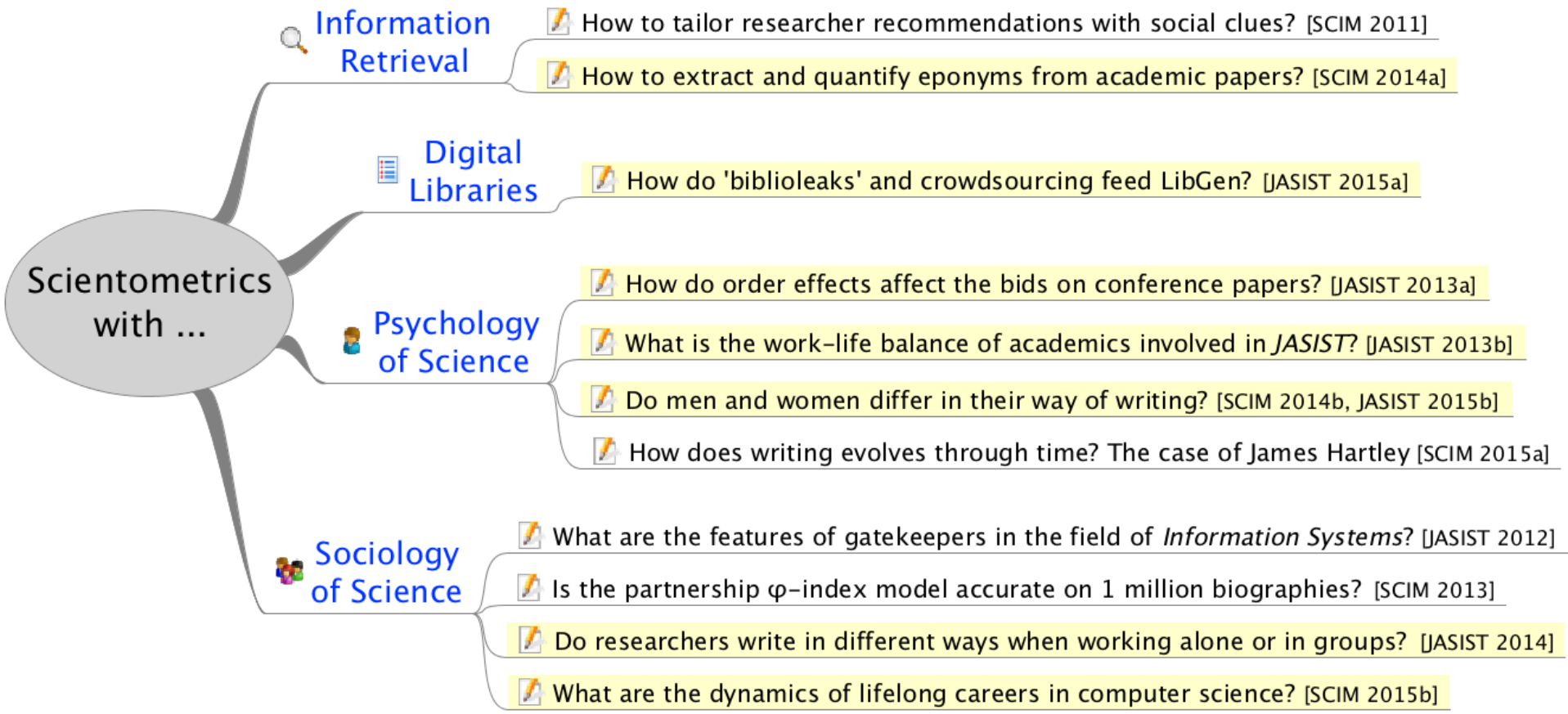


Fig. 11 Evolution of the balance between Newcomer, Intermediate, and Confirmed co-authors for the 3,860 researchers under study

Case Studies



BIR@ECIR: Revitalising IR ↔ Bibliometrics

Inf Retrieval (2014) 17:412–429
DOI 10.1007/s10791-013-9232-5

INFORMATION RETRIEVAL IN THE INTELLECTUAL PROPERTY DOMAIN

The effect of citation analysis on query expansion for patent retrieval

Parvaz Mahdabi · Fabio Crestani

JDOC
70,1

52

Received 17 October 2012

Assigning publications to multiple subject categories for bibliometric analysis

An empirical case study based on percentiles

Lutz Bornmann

Bibliometric evolution: Is the *journal of the association for information science and technology* transforming into a specialty Journal?

Jeppe Nicolaisen[1] and Tove Faber Frandsen
2

Article first published online: 17 JUN 2014
DOI: 10.1002/asi.23224

Issue

Information Processing &

Volume 51, Issue 2, March 2015, Pag

Breadth and depth of citati

Siluo Yang, Ruizhen Han

OIR
38,3

348

Received 3 July 2013
First revision approved
25 August 2013

Decrease in free computer science papers found through Google Scholar

Lee A. Pedersen
University Library, Brown University, Providence, Rhode Island, USA, and
Julie Arendt
VCU Libraries, Virginia Commonwealth University, Richmond, Virginia, USA

Information Sciences

Volume 221, 1 February 2013, Pages 245–261

Using semi-structured data for assessing research paper similarity

Germán Hurtado Martín[a, b], Steven Schockaert[c], Chris Cornelis[b, d], Helga Naessens[a]

World Wide Web (2015) 18:253–263
DOI 10.1007/s11280-013-0226-4

ELM-based name disambiguation in bibliography

Donghong Han · Siqi Liu · Yachao Hu · Bin Wang · Yongjiao Sun

Foundations and Trends® in Information Retrieval > Vol 7 > Issue 1

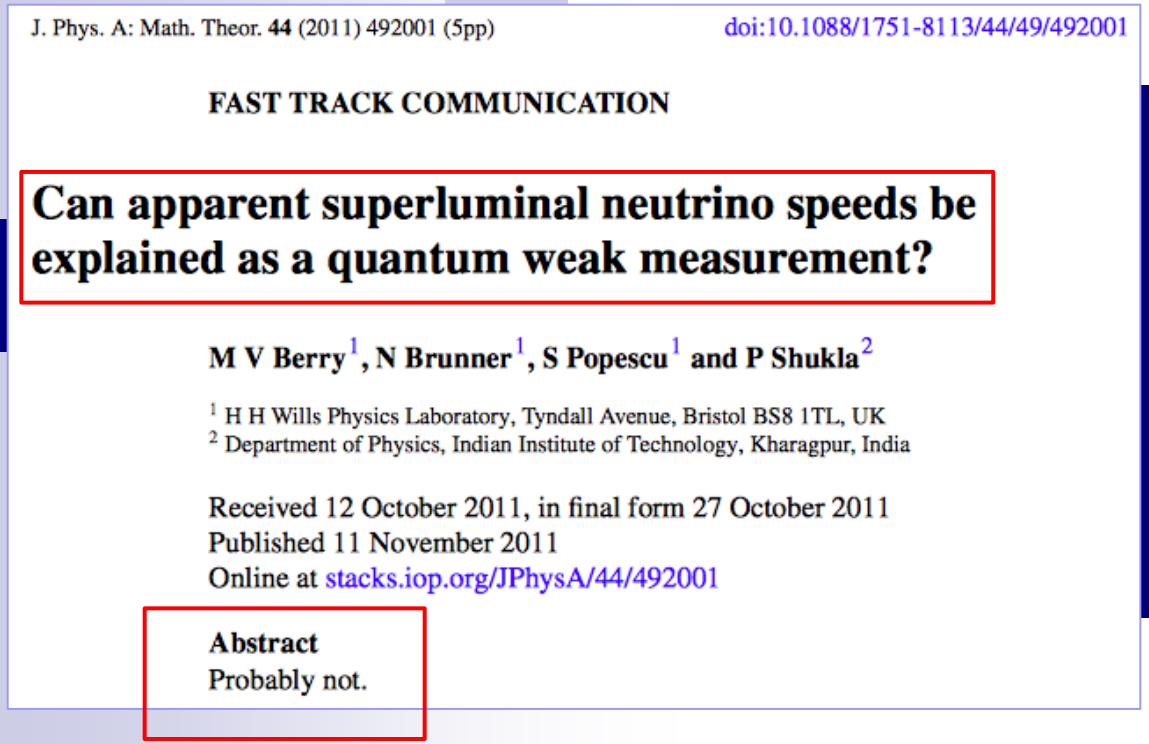
Patent Retrieval

Mihai Lupu, Vienna University of Technology, Austria, lupu@ifs.tuwien.ac.at

Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles

Journal of Information Science
2014, Vol. 40(1) 115–124

DOI: 10.1177/0165551513511393
jis.sagepub.com
SAGE



J. Phys. A: Math. Theor. **44** (2011) 492001 (5pp) doi:10.1088/1751-8113/44/49/492001

FAST TRACK COMMUNICATION

Can apparent superluminal neutrino speeds be explained as a quantum weak measurement?

M V Berry[1], **N Brunner**[1], **S Popescu**[1] **and P Shukla**[2]

[1] H H Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK
[2] Department of Physics, Indian Institute of Technology, Kharagpur, India

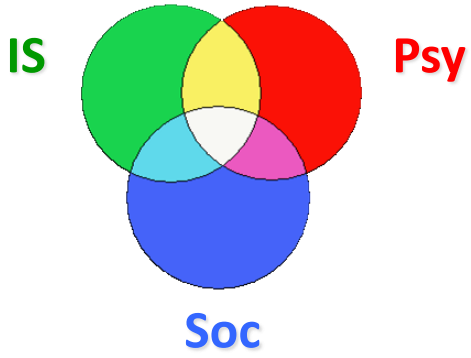
Received 12 October 2011, in final form 27 October 2011
Published 11 November 2011
Online at stacks.iop.org/JPhysA/44/492001

Abstract
Probably not.

Thanks!

http://www.irit.fr/~Guillaume.Cabanac

Twitter: @gcabanac



Information Science

- Information Retrieval
- Digital Libraries

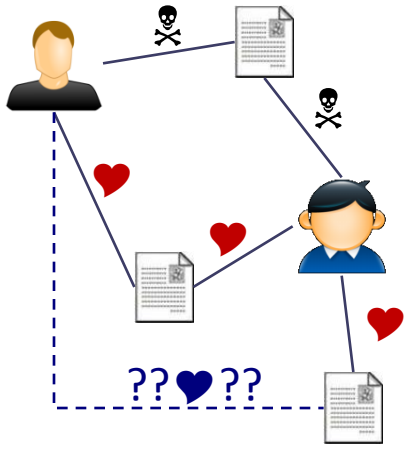
Study 8

How to Recommend Researchers According to their Research Topics and Social Clues?

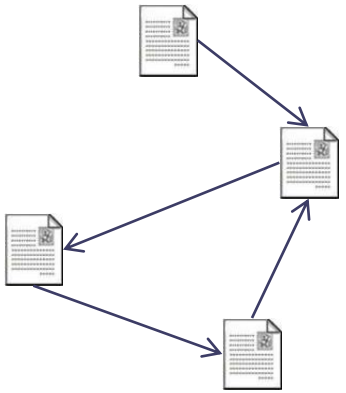
Cabanac, G. (2011). Accuracy of inter-researcher similarity measures based on topical and social clues. *Scientometrics*, *87*, 3, 597–620. doi:10.1007/s11192-011-0358-1

Recommendation of Literature (McNee et al., 2006)

- Collaborative filtering
  - Principle: mining the preferences of researchers
    → *those who liked this paper also liked…*
  - ☹ Snowball effect / fad
  - ☹ Innovation?
  - ☹ Relevance of theme?

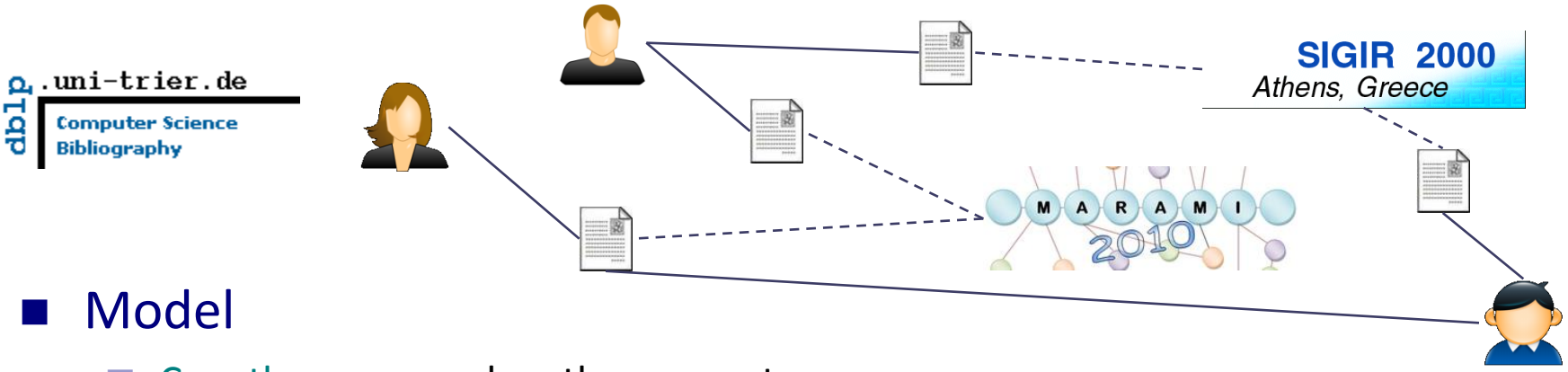


- Cognitive filtering
  - Principle: mining the contents of articles
    → profile of resources (researcher, articles)
    → citation graph

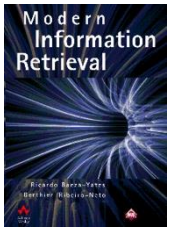


- Hybrid approach

Foundations: Similarity Measures Under Study



- Model
  - Coauthors — graph authors ↔ auteurs
  - Venues — graph authors ↔ conferences / journals
- Social similarities
  - Inverse degree of separation — *length of the shortest path*
  - Strength of the tie — *number of shortest paths*
  - Shared conferences — *number of shared conference editions*
- Thematic similarity
  - Cosine on Vector Space Model $\mathbf{d}_i = (w_i^1, \ldots, w_i^n)$ built on titles (doc / researcher)



Computing Similarities with Social Clues

- Task of literature review
  - Requirement topical relevance
  - Preference social proximity (meetings, project…)

⇒ re-rank topical results with social clues

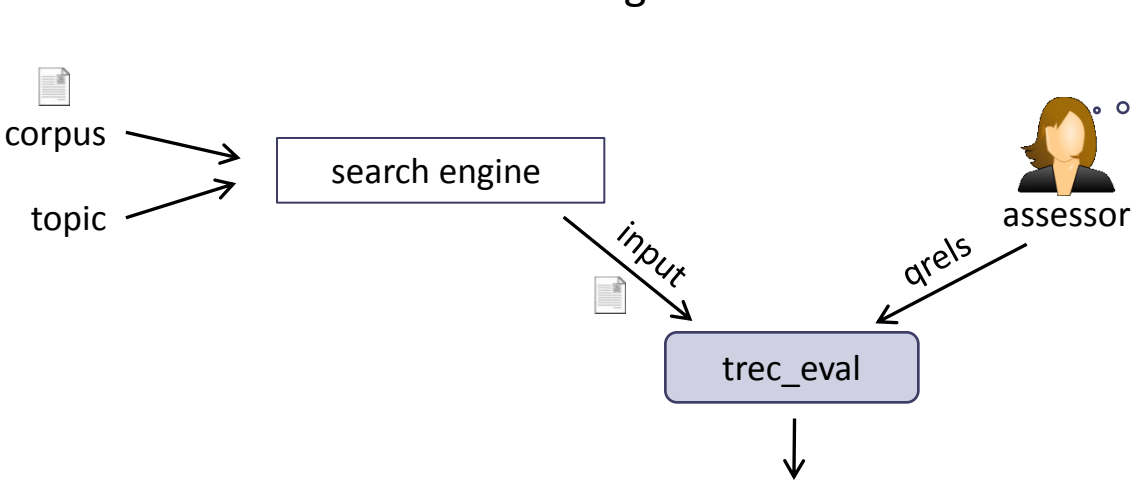
- Combination with CombMNZ (Fox & Shaw, 1993)

- Final result: list of recommended researchers

Evaluation Design

- Comparison of recommendations and researchers' perception
  - Q1 : Effectiveness of topical (only) recommendations?
  - Q2 : Gain due to integrating social clues?

- IR experiments: Cranfield paradigm (TREC…)
  - Does the search engine retrieve relevant documents?

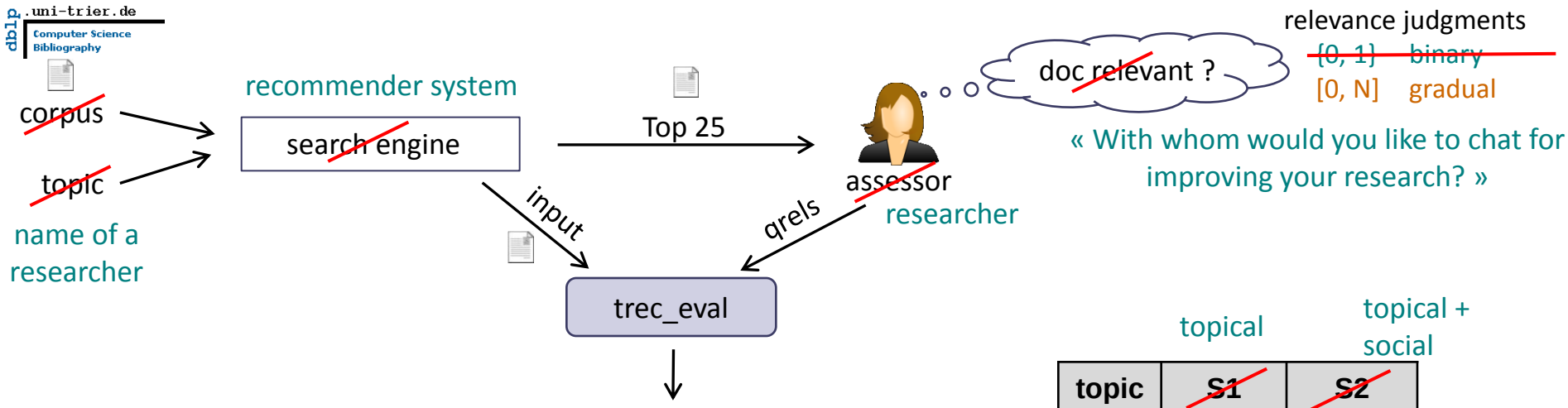


topic	S1	S2
1	0.5687	0.6521
…	…	…
50	0.7124	0.7512
avg	0.6421	0.7215

improvement +12.3 %
significativity $p < 0.05$ (paired t-test)

Evaluating Recommendations

- Adaptation of the Cranfield paradigm (TREC…)
  - Is the ~~search engine~~ rec. sys. Retrieving relevant ~~documents~~ researchers?

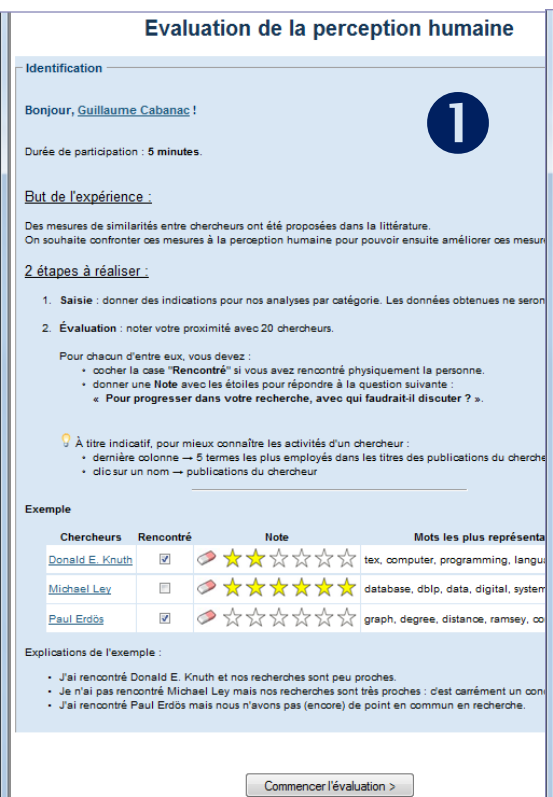


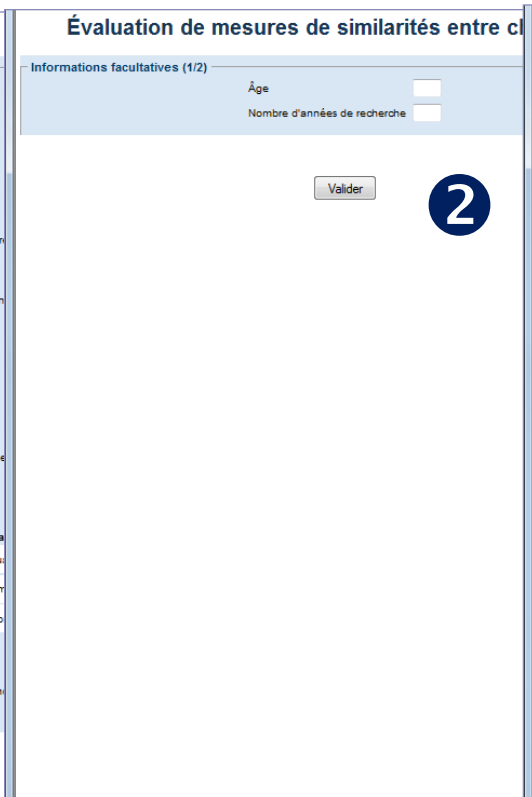
	topical	topical + social
topic	~~S1~~	~~S2~~
1	0.5687	0.6521
…	…	…
#subjects ~~50~~	0.7124	0.7512
avg	0.6421	0.7215

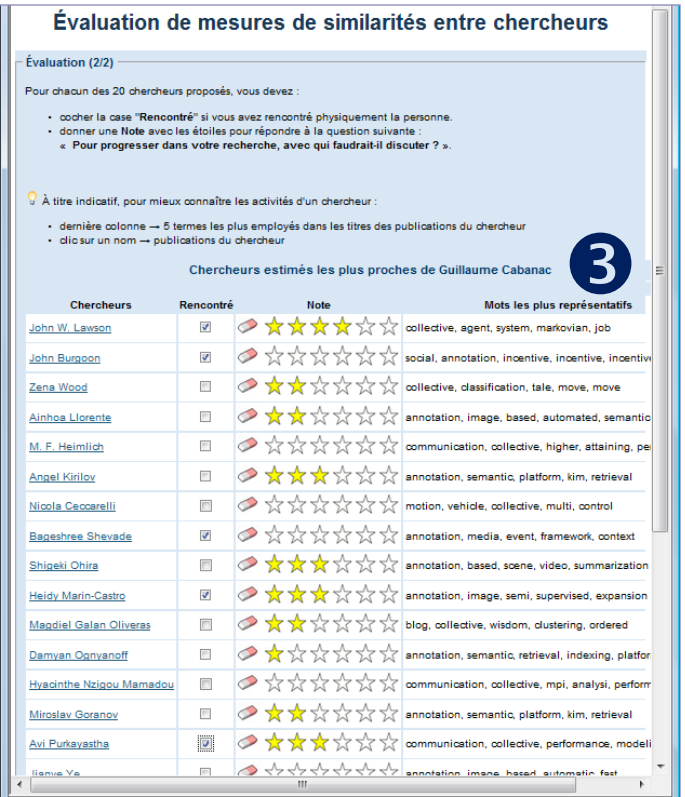
improvement +12.3 %
significativity $p < 0.05$ (paired t-test)

Experiment

- Features
  - Data: dblp.xml (713 MB = 1.3M publications for 811,787 researchers)
  - Subjects: 90 researchers-contacts LinkedIn contacted by mail
    74 researchers began to fill the questionnaire. 71 completed it
- Interface for assessing recommendations

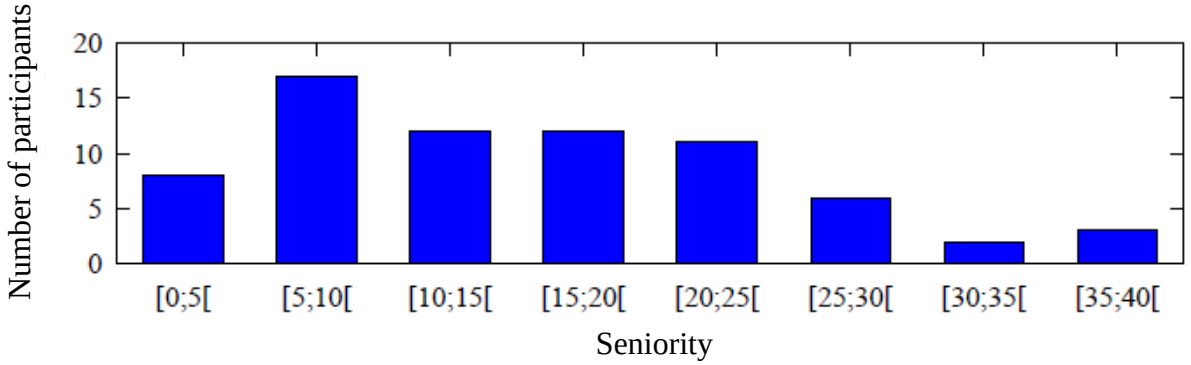




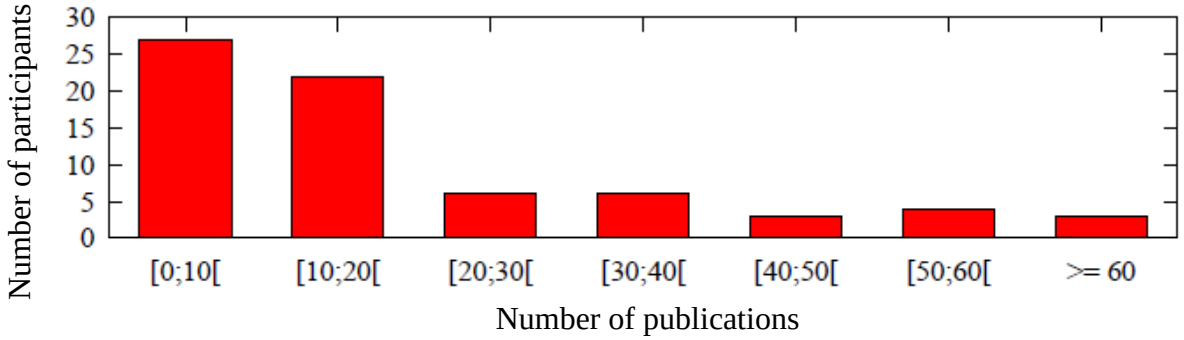


Experiments: Profile of the Participants

- Experience of the 71 subjects Mdn = 13 years

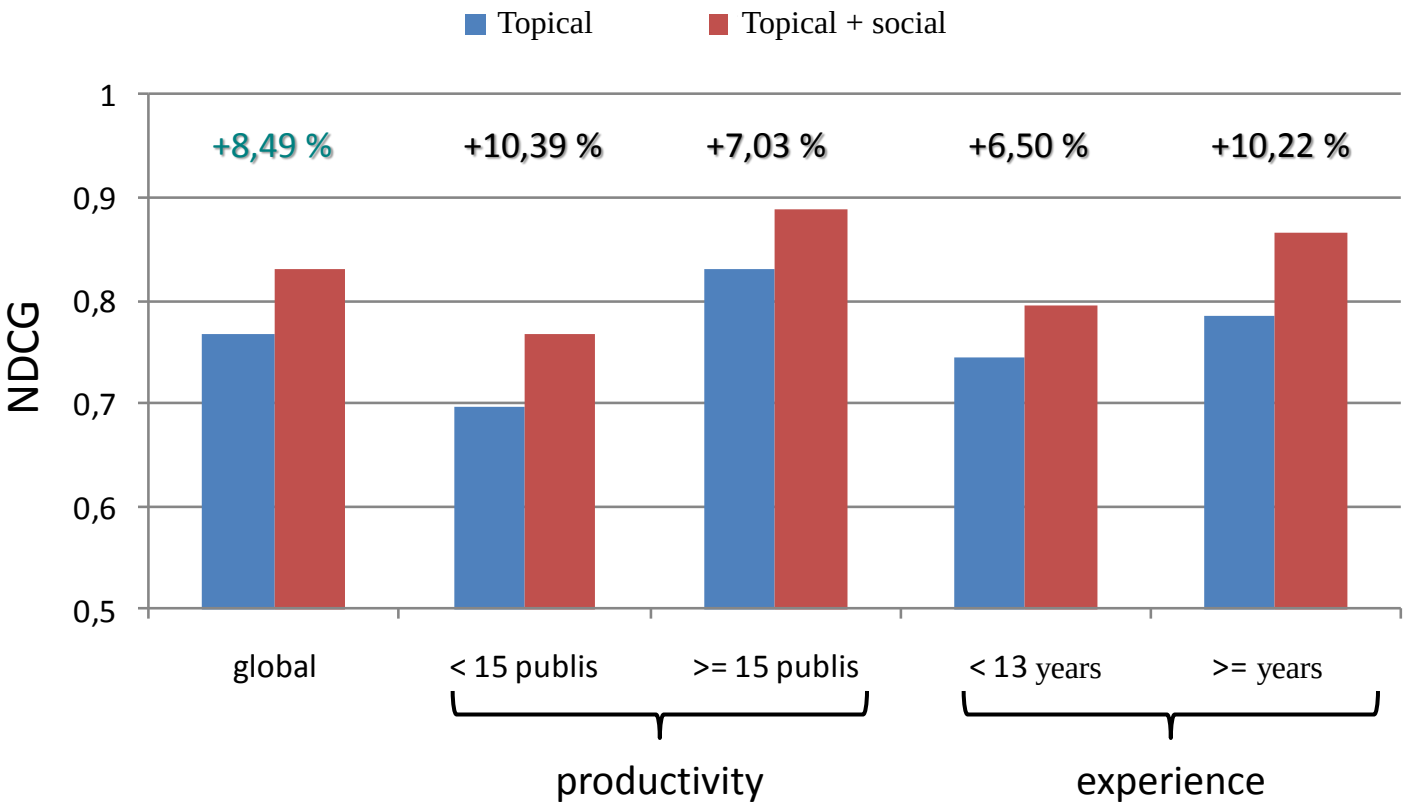


- Productivity of the 71 subjects Mdn = 15 publications

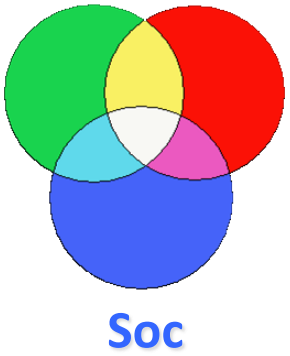


Empirical Validation of our Hypothesis

- Strong baseline ⇒ effective approach based on VSM



- +8.49 % = significant improvement ($p < 0.05$; $n = 70$) of topical recommendations by social clues



Sociology of Science

- Gatekeeping
- Collaboration

Study 9

Research Landscape from the Perspective of Gatekeepers in Information Systems

---

Cabanac, G. (2012). Shaping the landscape of research in information systems from the perspective of editorial boards: A scientometric study of 77 leading journals. *Journal of the American Society for Information Science and Technology*, *63*, 5, 977–996. doi:10.1002/asi.22609

Landscape of Research in Information Systems

- The gatekeepers of science

Landscape of Research in Information Systems

- The 77 core peer-reviewed IS journals in the WoS

Table 2 Leading 77 IS journals ranked by decreasing 5YJIF. Categories A, B, C, and D are delimited by the quartiles of the 5YJIF distribution

Rank	*JCR* abbreviated title	Publisher	5YJIF	Category	Rank	*JCR* abbreviated title	Publisher	5YJIF	Category
1	*Mis Quart*	U. Minnesota	9.208	A	40	*IEEE Secur Priv*	IEEE	1.830	C
2	*VLDB J*	Springer	6.987	A	41	*Wirel Netw*	Springer	1.784	C
3	*ACM T Inform Syst*	ACM	5.774	A	42	*Inform Retrieval*	Springer	1.752	C
4	*J Am Med Inform Assn*	BMJ	5.199	A	43	*Mobile Netw Appl*	Springer	1.725	C
5	*IEEE T Wirel Commun*	IEEE	4.534	A	44	*Comput Secur*	Springer	1.718	C
6	*Data Min Knowl Disc*	Springer	4.432	A	45	*Comput Netw*	Elsevier	1.610	C
7	*IEEE Pervas Comput*	IEEE	4.395	A	46	*World Wide Web*	Springer	1.564	C
8	*Inform Manage-Amster*	Elsevier	4.297	A	47	*Distrib Parallel Dat*	Springer	1.543	C
9	*J ACM*	ACM	4.200	A	48	*Eur J Inform Syst*	Palgrave	1.534	C
10	*IEEE T Mobile Comput*	IEEE	3.956	A	49	*Int J Coop Inf Syst*	World Scientific	1.468	C
11	*IEEE T Inform Theory*	IEEE	3.936	A	50	*Inform Syst Manage*	Taylor & Francis	1.436	C
12	*IEEE T Knowl Data En*	IEEE	3.691	A	51	*Inform Software Tech*	Elsevier	1.426	C
13	*IEEE T Depend Secure*	IEEE	3.649	A	52	*Geoinformatica*	Springer	1.396	C
14	*J Chem Inf Model*	ACS	3.631	A	53	*J Vis Commun Image R*	Elsevier	1.389	C
15	*IEEE Network*	IEEE	3.529	A	54	*Int J Inf Tech Decis*	World Scientific	1.379	C
16	*ACM T Database Syst*	ACM	3.290	A	55	*Internet Res*	Emerald	1.346	C
17	*J Manage Inform Syst*	M.E. Sharpe	3.215	A	56	*Inform Syst Front*	Springer	1.298	C
18	*Inform Sciences*	Elsevier	3.089	A	57	*J Intell Inf Syst*	Springer	1.207	C
19	*Enterp Inform Syst*	Elsevier	3.085	A	58	*Comput J*	Oxford	1.194	C
20	*Int J Med Inform*	Elsevier	3.061	A	59	*Online Inform Rev*	Emerald	1.111	D
21	*Decis Support Syst*	Elsevier	2.842	B	60	*Comput Commun Rev*	ACM	1.079	D
22	*ACM T Web*	ACM	2.813	B	61	*Acta Inform*	Springer	1.072	D
23	*ACM T Auton Adap Sys*	ACM	2.707	B	62	*Comput Commun*	Elsevier	1.012	D
24	*J Inf Technol*	Palgrave	2.664	B	63	*Int J Distrib Sens N*	Taylor & Francis	0.882	D
25	*J Strategic Inf Syst*	Elsevier	2.531	B	64	*Inform Process Lett*	Elsevier	0.877	D
26	***J Am Soc Inf Sci Tec***	**Wiley**	**2.480**	**B**	65	*Informatica-Lithuan*	IOS Press	0.854	D
27	*IEEE T Multimedia*	IEEE	2.372	B	66	*Multimedia Syst*	Springer	0.852	D
28	*Int J Geogr Inf Sci*	Taylor & Francis	2.303	B	67	*J Org Comp Elect Com*	Taylor & Francis	0.851	D
29	*Inform Syst*	Elsevier	2.302	B	68	*IEEE Syst J*	IEEE	0.825	D
29	*Knowl Inf Syst*	Springer	2.302	B	69	*J Res Pract Inf Tech*	ACS	0.752	D
31	*IEEE T Inf Technol B*	IEEE	2.268	B	70	*Multimedia Tools Appl*	Springer	0.712	D
32	*SIGMOD Rec*	ACM	2.224	B	71	*J Signal Process Sys*	Springer	0.578	D
33	*Inform Process Manag*	Elsevier	2.106	B	72	*Sci China Ser F*	Springer	0.473	D
34	*Wirel Commun Mob Com*	Wiley	2.069	B	73	*Bell Labs Tech J*	Wiley	0.459	D
35	*Data Knowl Eng*	Elsevier	2.053	B	74	*RAIRO-Theor Inf Appl*	EDP Sciences	0.451	D
35	*J Database Manage*	IGI Global	2.053	B	75	*J High Speed Netw*	IOS Press	0.442	D
37	*IEEE Multimedia*	IEEE	2.020	B	76	*J Inf Sci Eng*	Academia Sinica	0.390	D
38	*J Inf Sci*	Sage	1.996	B	77	*KSII T Internet Inf*	KSII	0.200	D
39	*Requir Eng*	Springer	1.907	B					

Landscape of Research in Information Systems

- Exploratory data analysis

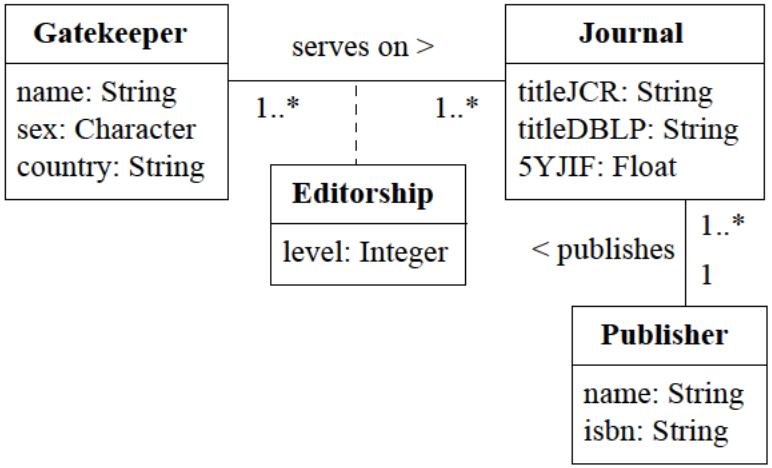


Fig. 1 UML class diagram for the dataset of editorial boards

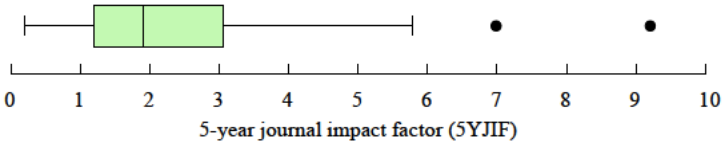


Fig. 2 Box plot of 5YJIF values for the 77 leading IS journals

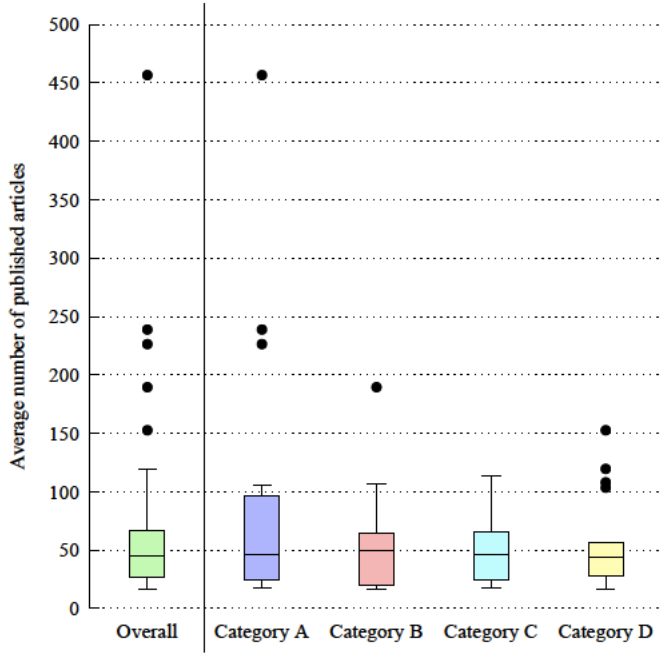


Fig. 4 Box plots of the average number of published articles per year for the 77 leading IS journals

Landscape of Research in Information Systems

- Exploratory data analysis

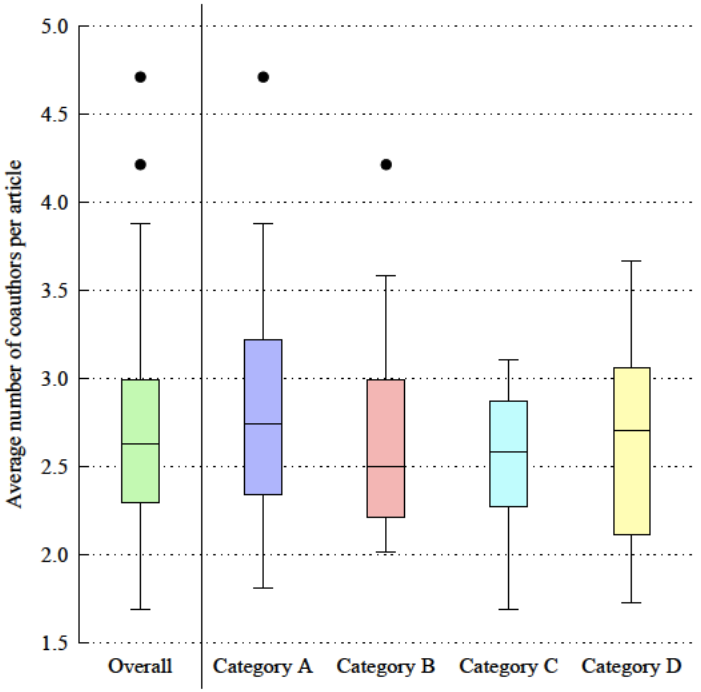


Fig. 6 Box plots of the number of coauthors per article, averaged for each of the 77 IS journals

Word cloud of the 100 most representative terms (highest TF·IDF) extracted from titles of articles published in the 77 leading IS journals

Landscape of Research in Information Systems

- Topical map of the IS field

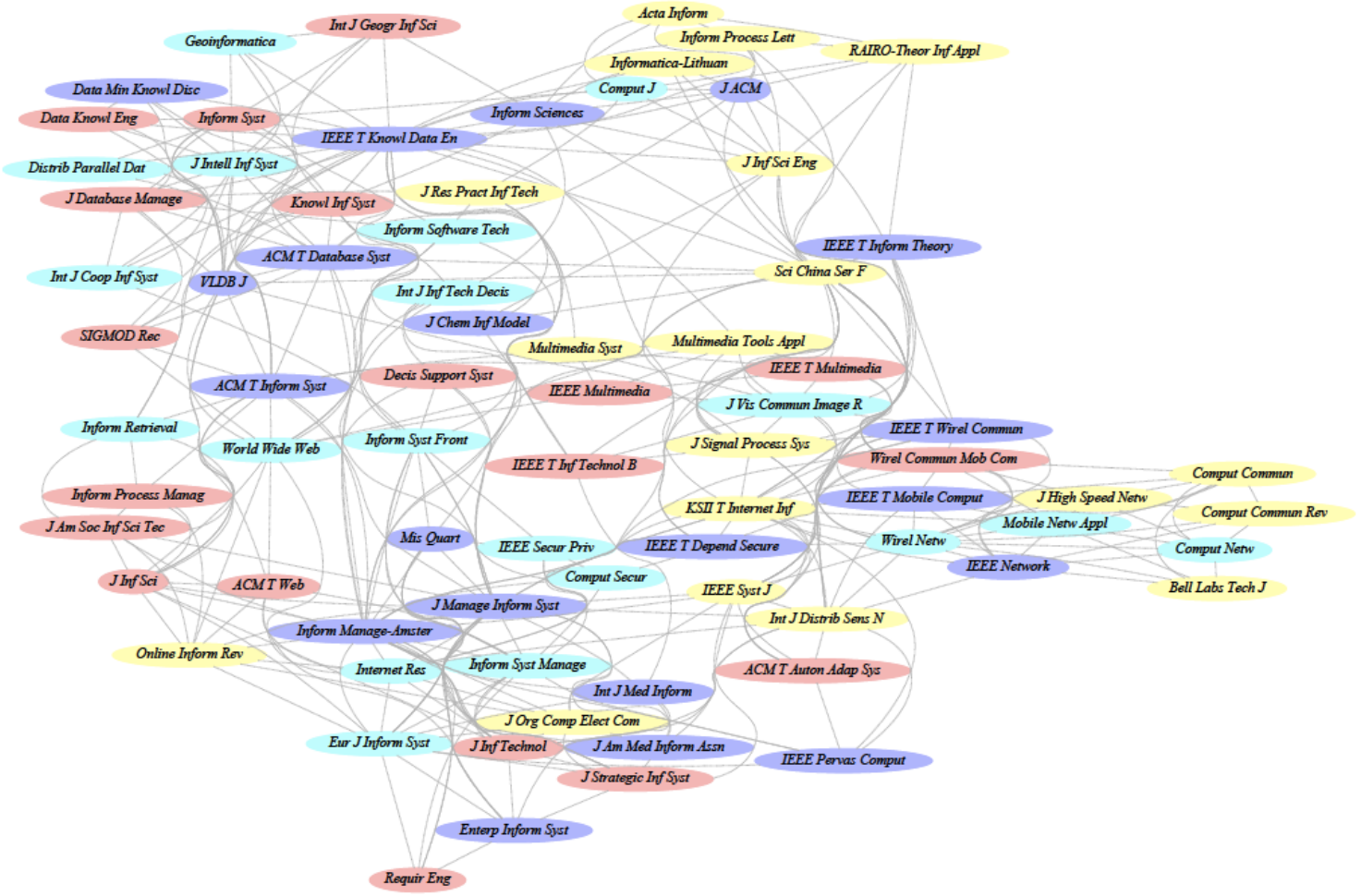


Fig. 9 Map of the 77 leading IS journals laid out according to their topic similarity. Each vertex of the graph represents a journal, whose category is shown by the color of the vertex (A , B , C , or D). An edge connects two vertices; its length is inversely proportional to similarity between the vertices. Vertices were laid out using Kamada and Kawai's (1989) force-directed placement algorithm.

Landscape of Research in Information Systems

- Most influential gatekeepers

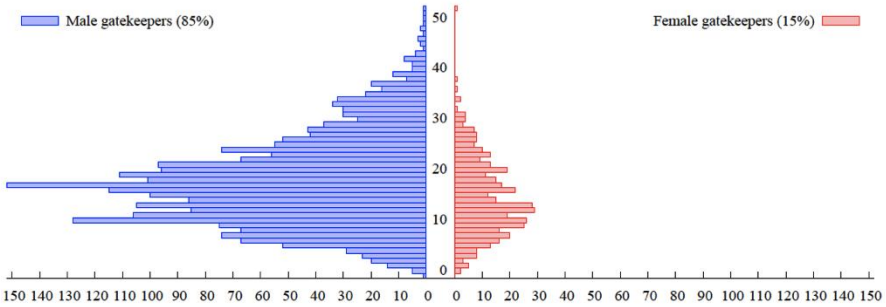


Fig. 12 Population pyramid of gatekeepers serving on the 77 leading IS journals, showing the distribution of seniority (i.e., number of years since first published scholarly article) with respect to gender

Table 4 Top 50 gatekeepers serving on the 77 leading IS journals in year 2011

Rank	Gatekeepers' involvement				Gatekeepers' weighted involvement			
	Gatekeeper	Country	Sex	#Journals	Gatekeeper	Country	Sex	Score
1	Elisa Bertino	us	f	8	Elisa Bertino	us	f	3.50
2	Andrew B. Whinston	us	m	5	Andrew B. Whinston	us	m	3.17
3	Athanasios V. Vasilakos	gr	m	5	Hsiao-Hwa Chen	tw	m	2.58
4	Benjamin W. Wah	us	m	5	Benjamin W. Wah	us	m	2.25
5	Qian Zhang	hk	f	5	Anthony S. Acampora	us	m	2.17
6	Anthony S. Acampora	us	m	4	Pericles Loucopoulos	uk	m	2.17
7	Edward A. Fox	us	m	4	Justin Zobel	au	m	2.08
8	Fabio Crestani	ch	m	4	Imrich Chlamtac	it	m	2.00
9	Hsiao-Hwa Chen	tw	m	4	Qian Zhang	hk	f	2.00
10	Johannes Gehrke	us	m	4	Fabio Crestani	ch	m	1.92
11	Justin Zobel	au	m	4	James R. Marsden	us	m	1.92
12	Kalle Lyytinen	us	m	4	Lotfi A. Zadeh	us	m	1.92
13	Lotfi A. Zadeh	us	m	4	Ricardo A. Baeza-Yates	cl	m	1.92
14	Matthias Jarke	de	m	4	Amit P. Sheth	us	m	1.83
15	Robert J. Kauffman	us	m	4	Beng Chin Ooi	sg	m	1.83
16	Sid L. Huff	nz	m	4	Mike P. Papazoglou	nl	m	1.83
17	Sudha Ram	us	f	4	Sudha Ram	us	f	1.83
18	Aidong Zhang	us	f	3	Leonid Libkin	uk	m	1.75
19	Amit P. Sheth	us	m	3	Marianne Winslett	us	f	1.75
20	Andrzej Skowron	pl	m	3	Robert J. Kauffman	us	m	1.75
21	Antonio Capone	it	m	3	Ugur Çetintemel	us	m	1.75
22	Athman Bouguettaya	au	m	3	Athanasios V. Vasilakos	gr	m	1.67
23	Beng Chin Ooi	sg	m	3	Clyde W. Holsapple	us	m	1.67
24	Bernard C. Y. Tan	sg	m	3	Gary J. Koehler	us	m	1.67
25	Blaize Horner Reich	ca	f	3	Kian-Lee Tan	sg	m	1.67
26	Bruce W. Weber	uk	m	3	Leonard Kleinrock	us	m	1.67
27	ChengXiang Zhai	us	m	3	Mischa Schwartz	us	m	1.67
28	Chris Jermaine	us	m	3	Mohsen Guizani	kw	m	1.67
29	Christina Fragouli	ch	f	3	Philip A. Bernstein	us	m	1.67
30	Colette Rolland	fr	f	3	Sid L. Huff	nz	m	1.67
31	Daniel Dajun Zeng	us	m	3	Wen-Lian Hsu	tw	m	1.67
32	David L. Olson	us	m	3	Witold Pedrycz	ca	m	1.67
33	Dominik Slezak	ca	m	3	Keng Siau	us	m	1.60
34	Douglas W. Oard	us	m	3	Edward A. Fox	us	m	1.58
35	Edie M. Rasmussen	us	f	3	Johannes Gehrke	us	m	1.58
36	Fabrizio Sebastiani	it	m	3	Minho Jo	kr	m	1.58
37	Gary J. Koehler	us	m	3	Bernard C. Y. Tan	sg	m	1.50
38	Hasan Pirkul	us	m	3	ChengXiang Zhai	us	m	1.50
39	Ian Ruthven	uk	m	3	Erol Gelenbe	uk	m	1.50
40	Iris Vessey	us	f	3	Ling Liu	us	f	1.50
41	James R. Marsden	us	m	3	Marek Rusinkiewicz	us	m	1.50
42	Javier Lopez	es	m	3	Nigel Davies	uk	m	1.50
43	Jayant R. Haritsa	in	m	3	Prabuddha De	us	m	1.50
44	Jiangchuan Liu	ca	m	3	Richard Baskerville	us	m	1.50
45	John C. Henderson	us	m	3	Srinivasan Keshav	ca	m	1.50
46	John Leslie King	us	m	3	Vijay K. Vaishnavi	us	m	1.50
47	Jon Crowcroft	uk	m	3	Matthias Jarke	de	m	1.48
48	Kar Yan Tam	hk	m	3	Sihem Amer-Yahia	us	f	1.42
49	Kian-Lee Tan	sg	m	3	Kalle Lyytinen	us	m	1.40
50	Leonard Kleinrock	us	m	3	Colette Rolland	fr	f	1.35

Landscape of Research in Information Systems

- Number of gatekeepers per country

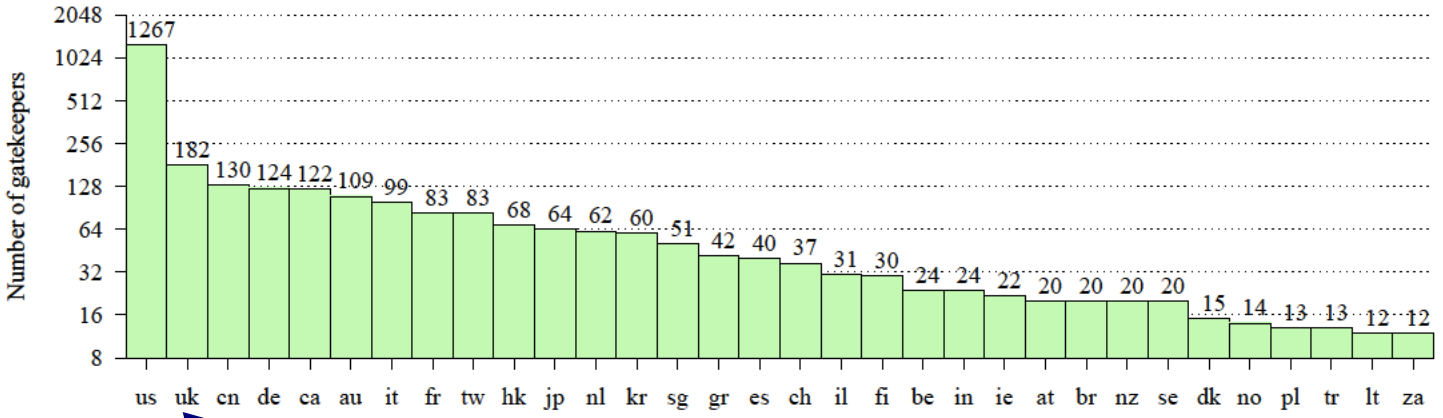


Fig. 15 Number of gatekeepers serving on the 77 leading IS journals for each country, as declared in their affiliation. Countries with less than 12 gatekeepers are not shown due to space limitations

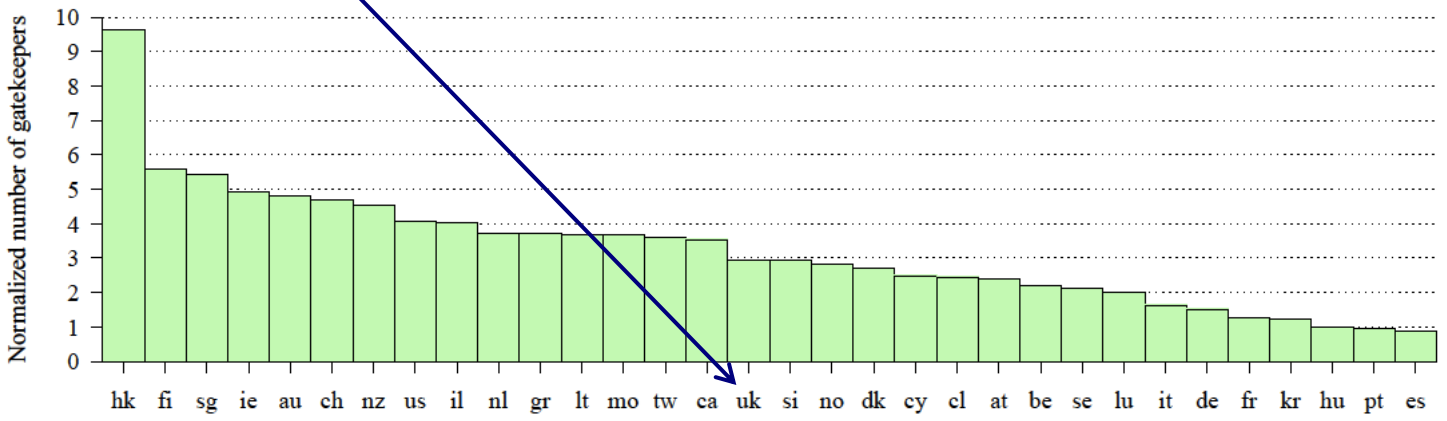


Fig. 16 Number of gatekeepers serving on the 77 leading IS journals for each country, normalized by the country's population. Due to space limitations, only the top 32 countries are displayed (same number of countries as in Figure 15)

Landscape of Research in Information Systems

- Geographic and gender diversity

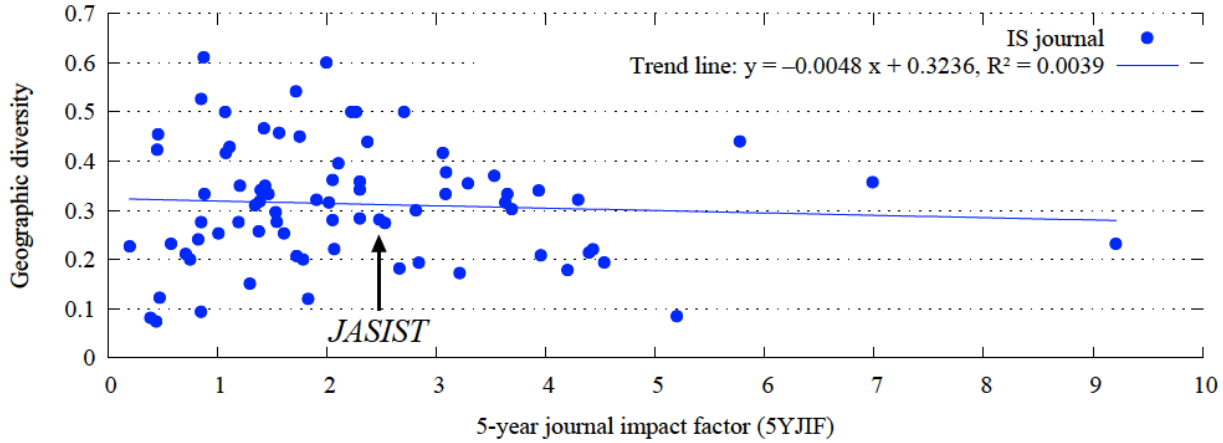


Fig. 19 Scatter plot showing the geographic diversity of gatekeepers in the 77 leading IS journals. Each data point represents a journal according to its 5YJIF (*x* axis) and its geographic diversity: the ratio of distinct countries from editors' affiliations over the number of editors serving on its editorial board (*y* axis)

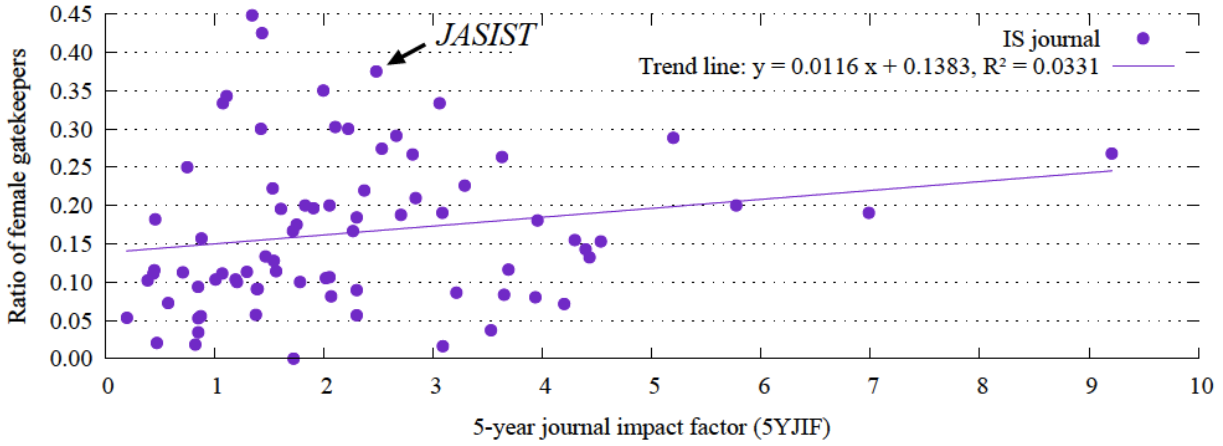


Fig. 21 Scatter plot showing the proportion of female gatekeepers in the boards of the 77 leading IS journals. Each data point represents a journal according to its 5YJIF (*x* axis) and the ratio of female over male gatekeepers serving on its board (*y* axis)


MORE
Coming Soon!