Suffering, bearing and loving peer review

ovvero

Inferno, purgatorio e paradiso nel peer review

Francisco Grimaldo
Departament d'Informàtica
Universitat de València

cost
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY

The Peer Review (PR) process

Is a cornerstone of science (publications, grants, projects)

Scrutinizes scientific contributions and decides about their fate

Determines how resources are allocated

Deserves to be studied and evaluated

Previous findings: bias, failures, disagreement, dissatisfaction…

Suffering

Inferno

Roma, caput mundi …

Agent-based models of Peer Review

Understand scientific behavior in idealized situations

Develop of alternative models

… while fighting against the lack of empirical data …

Two opposing forces

Simplify to the extreme:

Swarm intelligence
Complex systems, economics, sociology
Synchronous, interleaved, stepwise…
Oversimplification?

Make it as complex as you can

Cognitive intelligence
Philosophy, psychology, sociology
Asynchronous, concurrent, behavioral-driven…
Over-fitting and validation?

The Wyndham Lewis' Manifesto, Blast Magazine, 1914

"We fight first on one side, then on the other, but always for the SAME cause, which is neither side or both sides and ours.

Mercenaries were always the best troops.

We [engineers] *are primitive Mercenaries in the Modern World."*

LABSS-ISTC-CNR: PR-M

GECS – U. Brescia: Squazzoni – Gandelli (remix)

Squazzoni – Gandelli, JASSS, 2013

Studied the effect or reciprocity on the quality of PR.

<u>Extensions:</u>

Reviewing behaviours (fair, unreliable & strategic)
Editorial policies (author-referee matching)

<u>Two new research questions:</u>

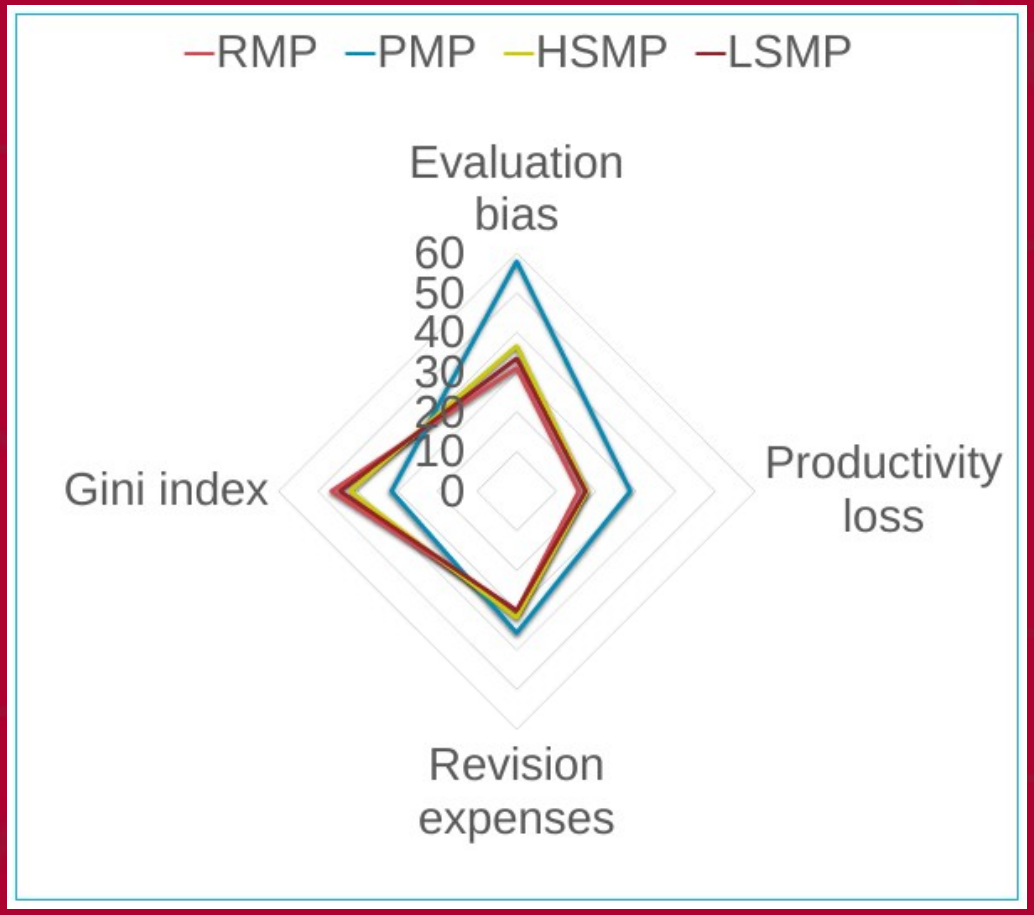
Impact of strategic behaviors on quality and efficiency?
Effects of matching scientists based on academic status?

Effects of cheating behaviours

Scenario	Evaluation bias	Productivity loss	Reviewing expenses
Weak selection (75% published submissions)			
Random behaviour	16.51 %	7.68 %	25.98 %
Cheating	20.07 %	4,91 %	21.34 %
Medium-level selection (50% published submissions)			
Random behaviour	25.27 %	14.98 %	30.77 %
Cheating	56.63 %	28.02 %	32.21 %
Strong selection (25% published submissions)			
Random behaviour	29.42 %	15.00 %	29.42 %
Cheating	70.86 %	34.72 %	35.24 %

When competition is pushed too hard. An agent-based model of strategic behaviour of referees in peer review. J.B. Cabotà, F. Grimaldo, F. Squazzoni. ECMS. 2013.

Indexes disequilibrium in local competition



Do editors have a silver bullet? An agent-based model of peer review. J.B. Cabotà, F. Grimaldo, F. Squazzoni. ECMS. 2014.

Positive effects of matching policies

	Eval. bias	Prod. loss	Rev. exp.	Gini. index	Cheat. perc.
Cheating					
RMP	70.86	34.72	35.24	0.28	0.27
PMP	51.97	25.69	35.19	0.33	0.25
HSMP	61.95	29.81	34.60	0.30	0.19
LSMP	73.00	36.92	34.86	0.29	0.32
Glass ceiling					
RMP	70.35	34.70	34.56	0.29	0.34
PMP	58.02	28.56	35.64	0.32	0.38
HSMP	65.88	32.26	35.23	0.30	0.37
LSMP	68.21	34.47	34.29	0.29	0.36

Do editors have a silver bullet? An agent-based model of peer review. J.B. Cabotà, F. Grimaldo, F. Squazzoni. ECMS. 2014.

Some first conclusions

Possible editorial counteractions to reduce the impact of misbehavior by matching authors/referees according to their track record

Bias reduction in competitive scenarios by assigning referees of similar or higher quality

Marching policies can also be detrimental:

Stratification of scientists in competing groups
In presence of niches of competition

The PR-M model

Main differences

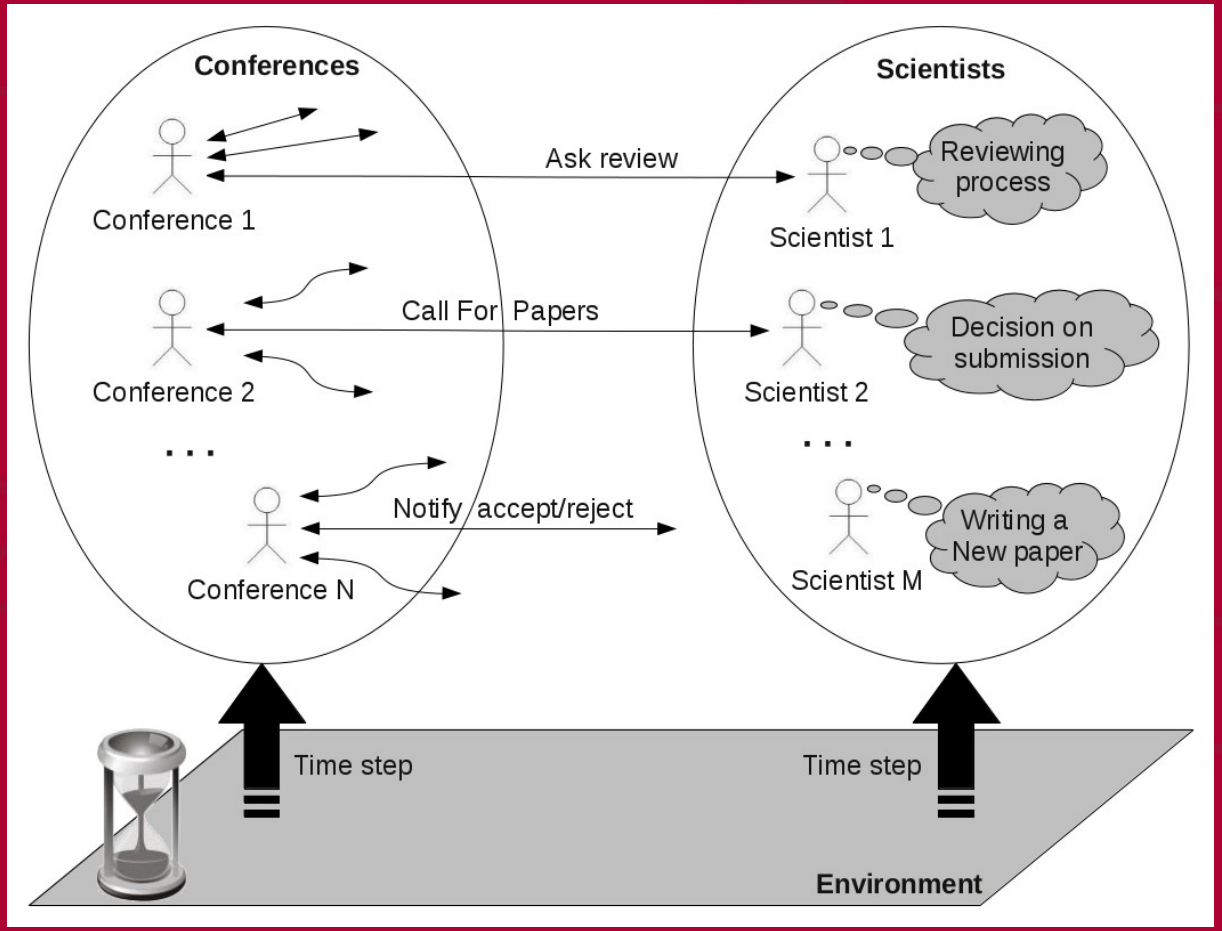
Multivalued reviewing

Acceptance policy

Reviewing image

Account for disagreement

PC selection



Two more research questions

Can the PR system ensure quality, in the face of variable reviewing skills or strategic behaviors, thanks to some selection process of the PC composition that leans on disagreement control?

Is the rational strategy really detrimental?

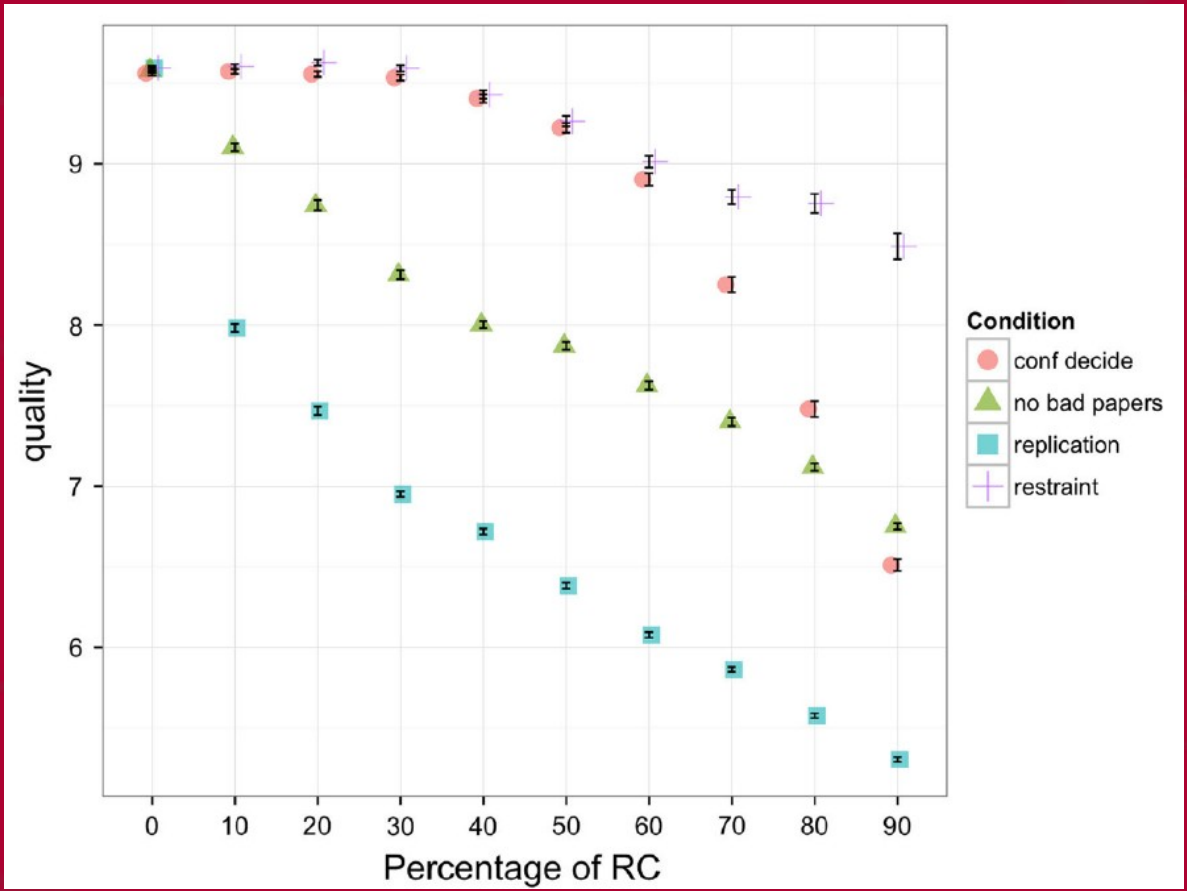
In which sense and under which circumstances?

Disagreement control (effectiveness)

Conference	% Initial disagreements	% Final disagreements	% Disagreement reduction
National	18.25	-	-
Summer School	10.71	-	-
International	5.41	-	-
Intl. Core C	5.0	-	-
Intl. Core B	0.0	-	-
Hom-0%RC	4.3	2.9	32.6
Hom-10%RC	6.1	4.5	26.2
Hom-30%RC	11.9	5.6	52.9
Het-0%RC-LQ	4.7	3.6	23.4
Het-0%RC-MQ	3.4	1.7	50.0
Het-0%RC-HQ	4.2	3.8	9.5
Het-10%RC-LQ	9.4	4.2	55.3
Het-10%RC-MQ	8.6	5.5	36.1
Het-10%RC-HQ	5.2	2.4	53.9
Het-30%RC-LQ	46.0	11.8	74.4
Het-30%RC-MQ	16.0	6.1	61.9
Het-30%RC-HQ	3.9	2.8	28.2

A simulation of disagreement for control of rational cheating in peer review. F. Grimaldo, M. Paolucci. Advances in Complex Systems. 2013.

Effects of rational cheaters



Mechanism change in a simulation of peer review: from junk support to elitism. M. Paolucci, F. Grimaldo. Scientometrics. 2014.

Some more conclusions

PR outcomes are sensitive to how scientists identify their competitors

Editorial counteractions to reduce the impact of referee misbehavior:

Avoid matching amongst peers under local competition
Select referees considering disagreements

PR and strategic behavior show a complex interaction:

It can cause a quality collapse or …
even a slight quality increase depending on the mechanisms

Bearing

Purgatorio

PEERE Cost Action

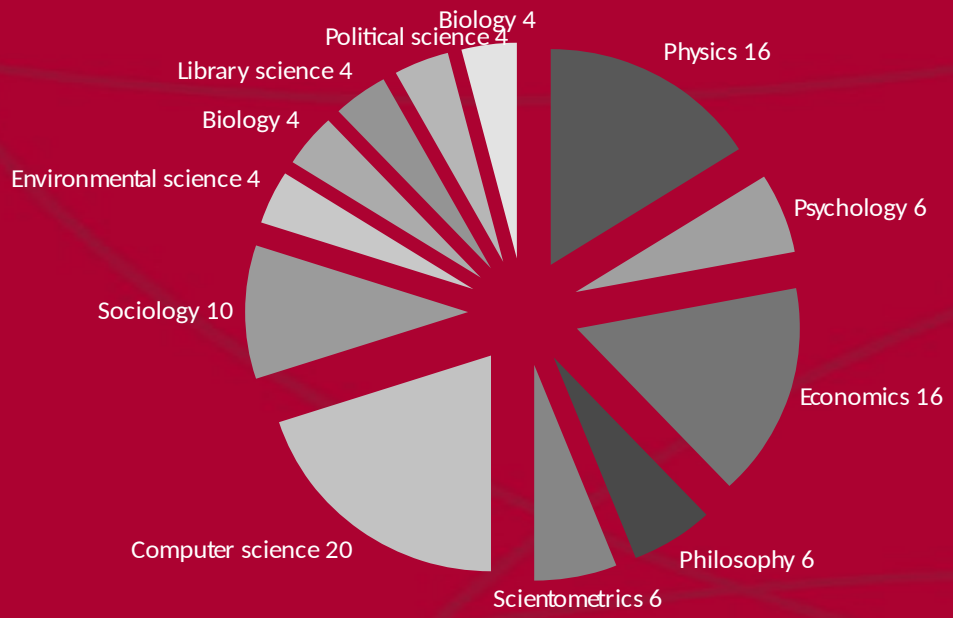
PEERE

PEERE aims to improve efficiency, transparency and accountability of peer review

The objectives of this Action are:

- **to analyze peer review in different scientific areas by integrating quantitative and qualitative research and incorporating recent experimental and computational findings;**
- **to evaluate implications of different models of peer review and to explore new incentive structures, rules and measures to improve collaboration in all stages of the peer review process;**
- **to involve science stakeholders in data sharing and testing initiatives;**
- **to define collaboratively a joint research agenda that points to an evidence-based peer review reform**

PEERE



Chair of the Action:
Prof Flaminio SQUAZZONI (IT)

Vice Chair of the Action:
Dr Francisco GRIMALDO (ES)

Science officer of the Action:
Dr Giuseppe LUGANO

Administrative officer of the Action:
Ms Anja VAN DER SNICKT

Participations

Country	Date
Austria	10/01/2014
Belgium	29/01/2014
Bulgaria	05/02/2014
Croatia	14/02/2014
Czech Republic	19/06/2014
Denmark	10/02/2014
Estonia	23/01/2014
Finland	24/01/2014
France	05/12/2013
fYR Macedonia	24/09/2014
Germany	07/01/2014
Greece	11/12/2013
Hungary	23/01/2014
Ireland	29/01/2014
Israel	28/11/2013
Italy	30/12/2013
Malta	02/01/2014
Netherlands	23/12/2013
Norway	23/01/2014
Poland	10/12/2013
Portugal	25/01/2014
Serbia	17/06/2014
Slovenia	20/03/2014
Spain	27/11/2013
Sweden	31/03/2014
Switzerland	16/01/2014
United Kingdom	22/11/2013
Total: 27	

Research questions

What research work on peer review has been done so far?

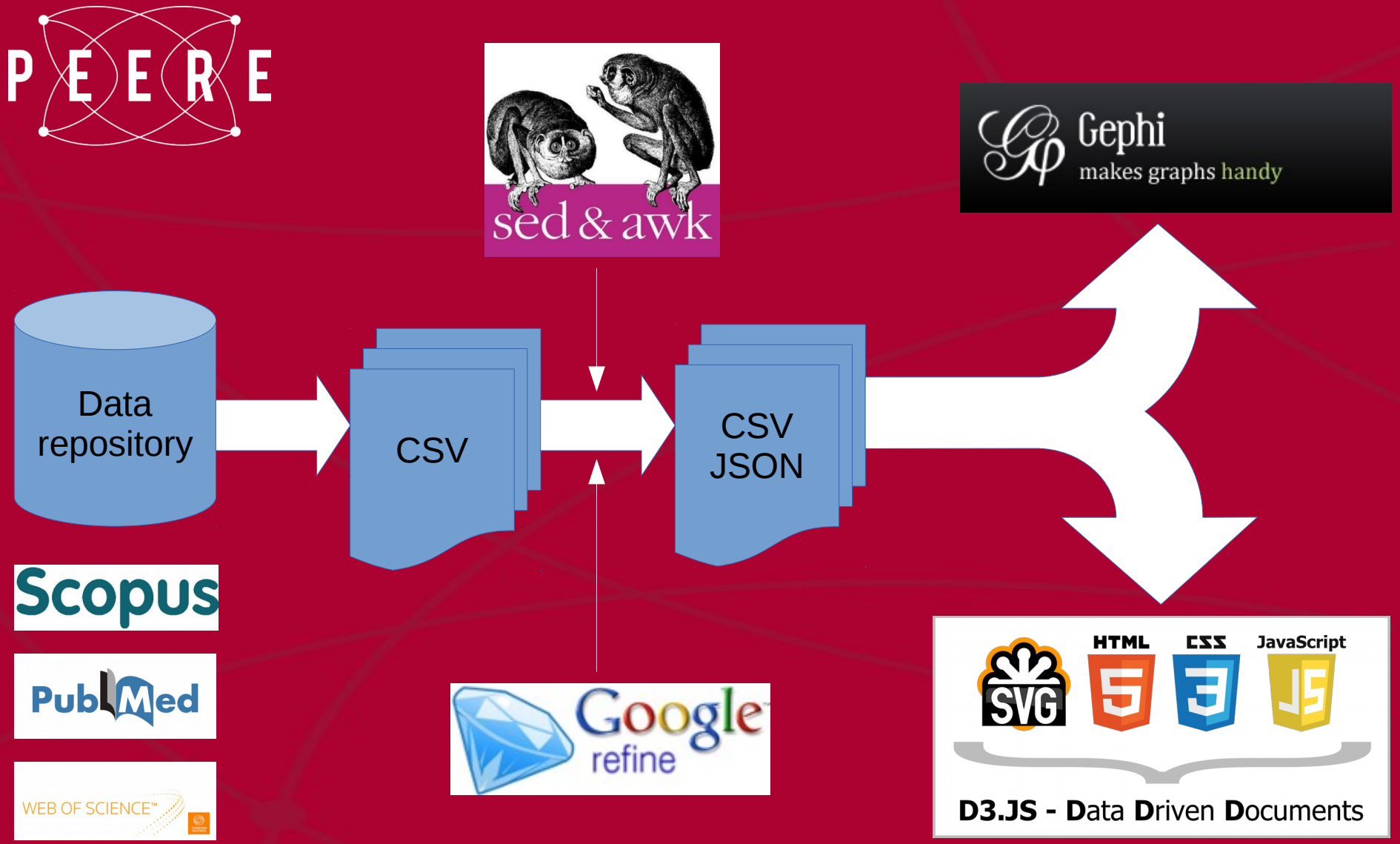
What has been the evolution of scientific publications about PR?

What kinds of documents are we speaking about?

Which is the community behind this research?

How does this community look like?

PR as a field of research: An empirical analysis on literature trends


PEERE
sed & awk
Gephi
makes graphs handy
Data repository
CSV
CSV
JSON
Scopus
PubMed
WEB OF SCIENCE
Google refine
HTML
CSS
JavaScript
SVG
D3.JS - Data Driven Documents

PEERE

Data source repositories

Time period: 1969 – 2014
Searched term: Peer review

Scopus:
- Article title
- Author keywords

PubMed:
- MeSH: Medical Subject Headings
- ["Peer Review, Research"]

Web of Science:
- Title

Away from the numbers…

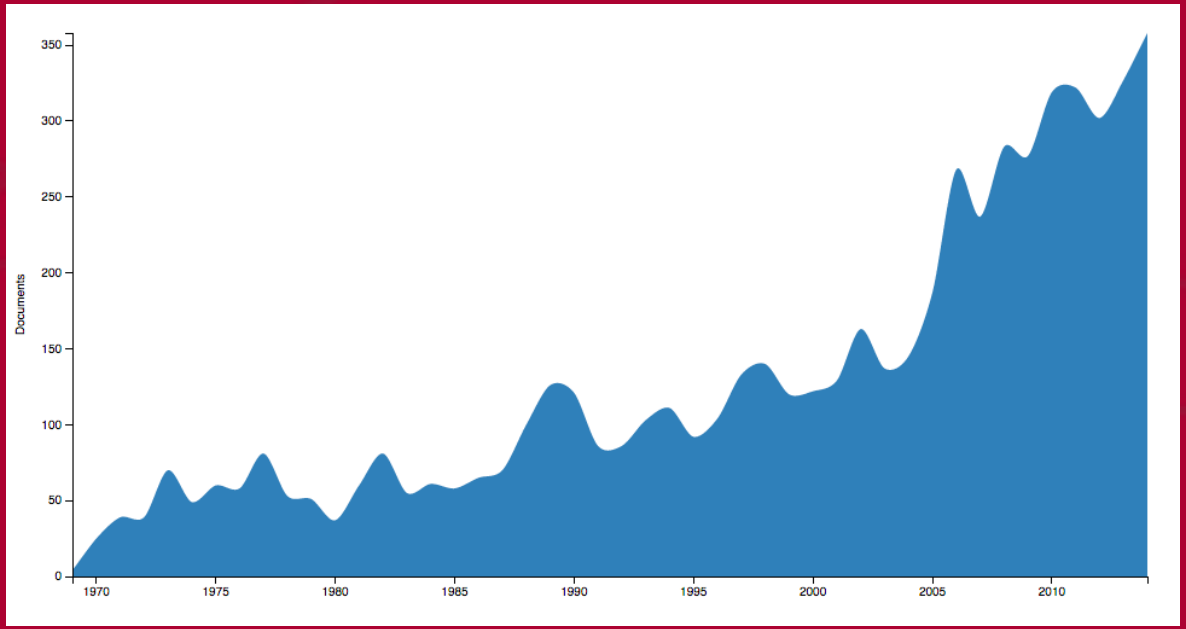
Data extraction:
May 1st, 2015

Baseline:
Scientific docs written in English
~47.5 million documents

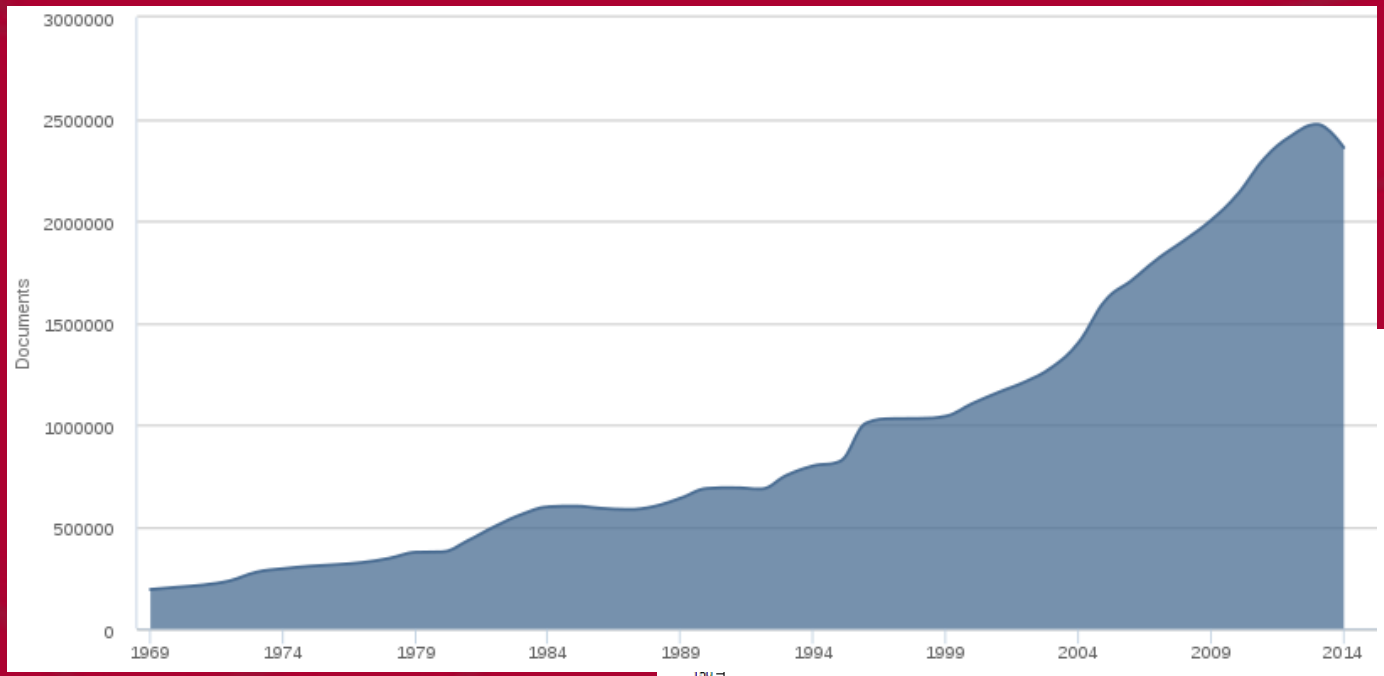
Clean up:
Unreliability of "Abstract"
Unreliability of "Indexed Keywords"
Avoid duplicates
~5900 docs (Scopus),
~5200 docs (PubMed & WoS)

Evolution (papers published/year)

Scopus

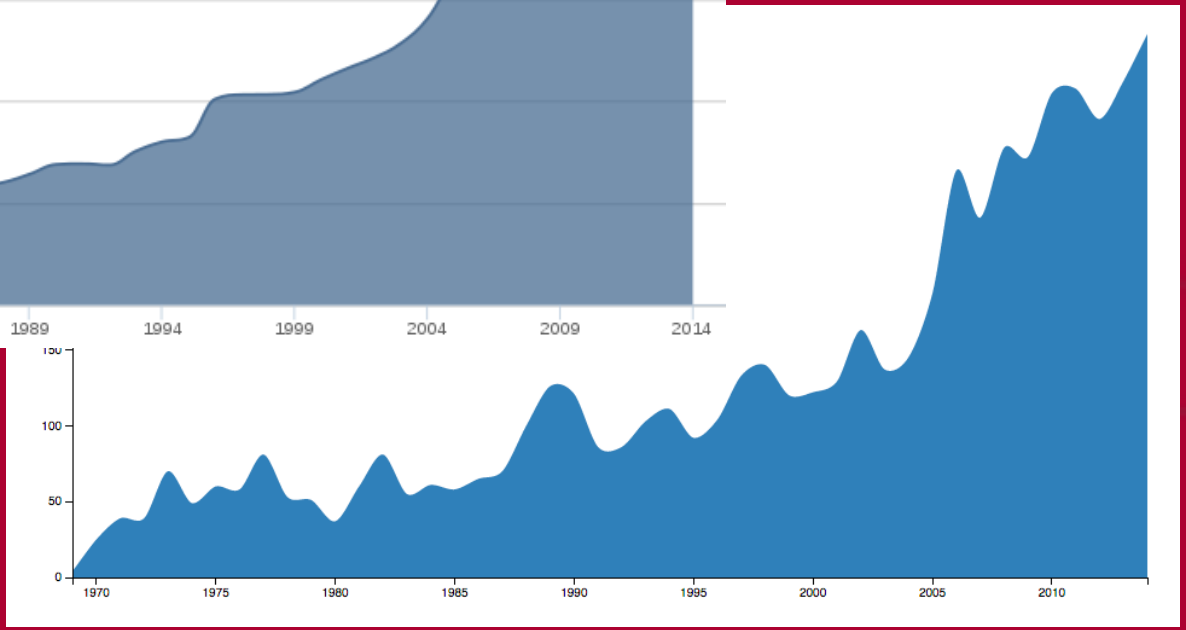


Evolution (papers published/year)

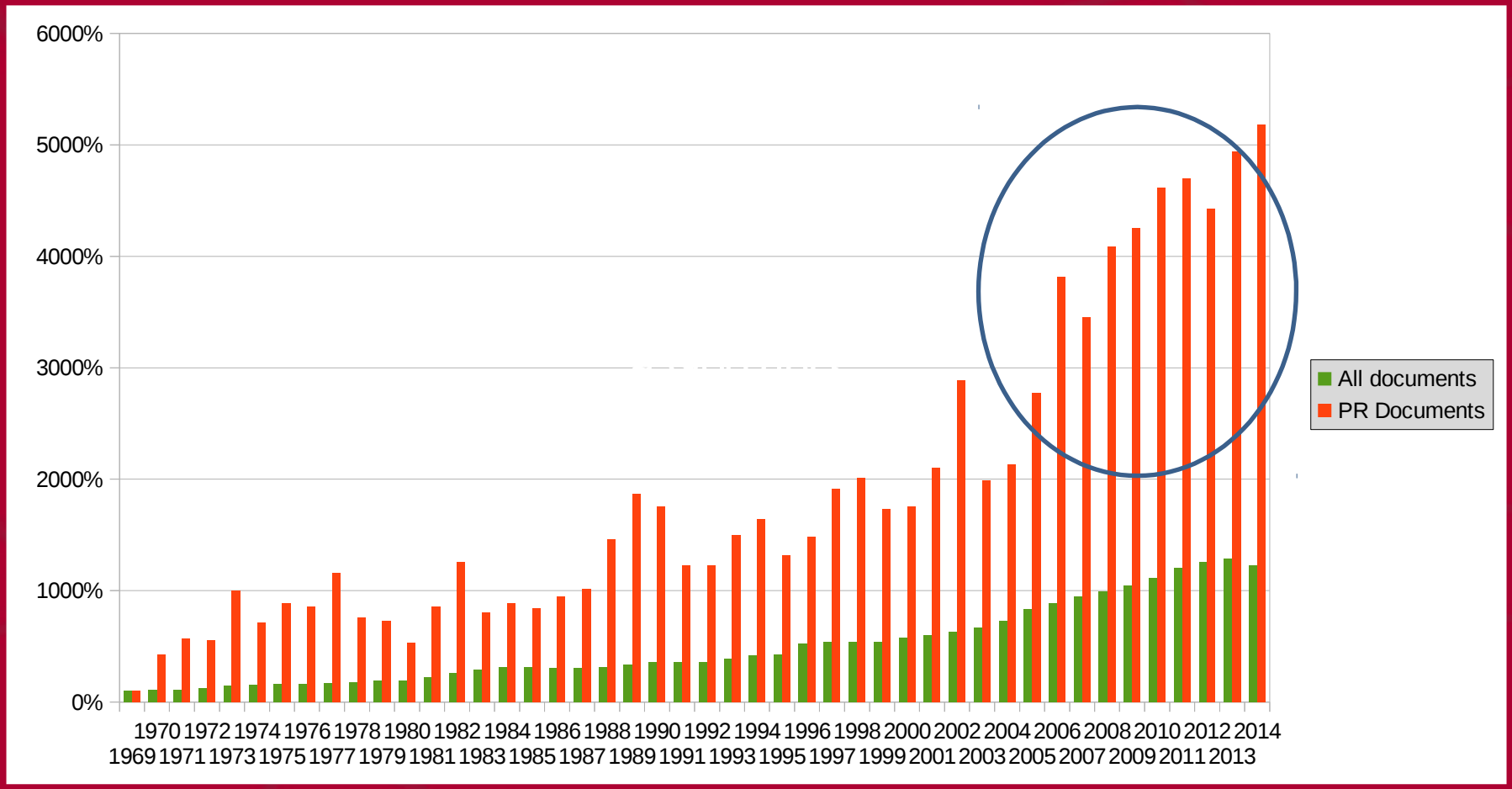


Scopus

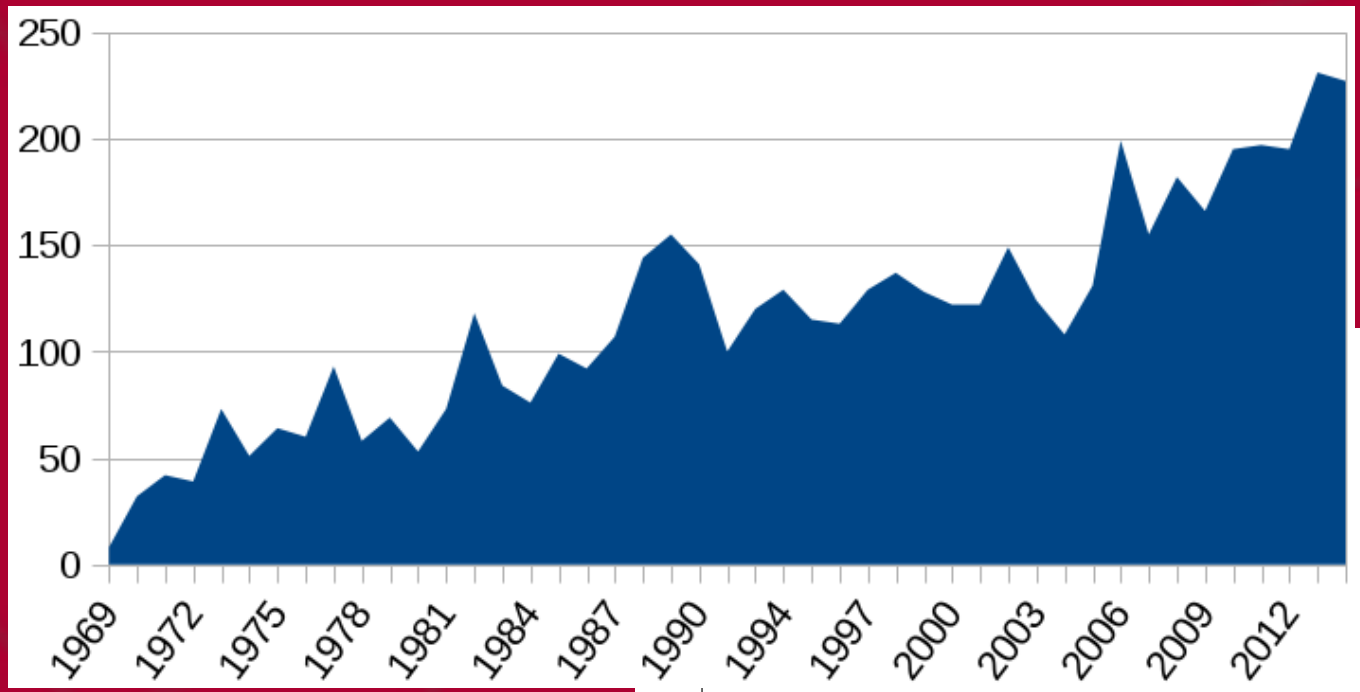
Papers in English



Growth (n° of papers published)

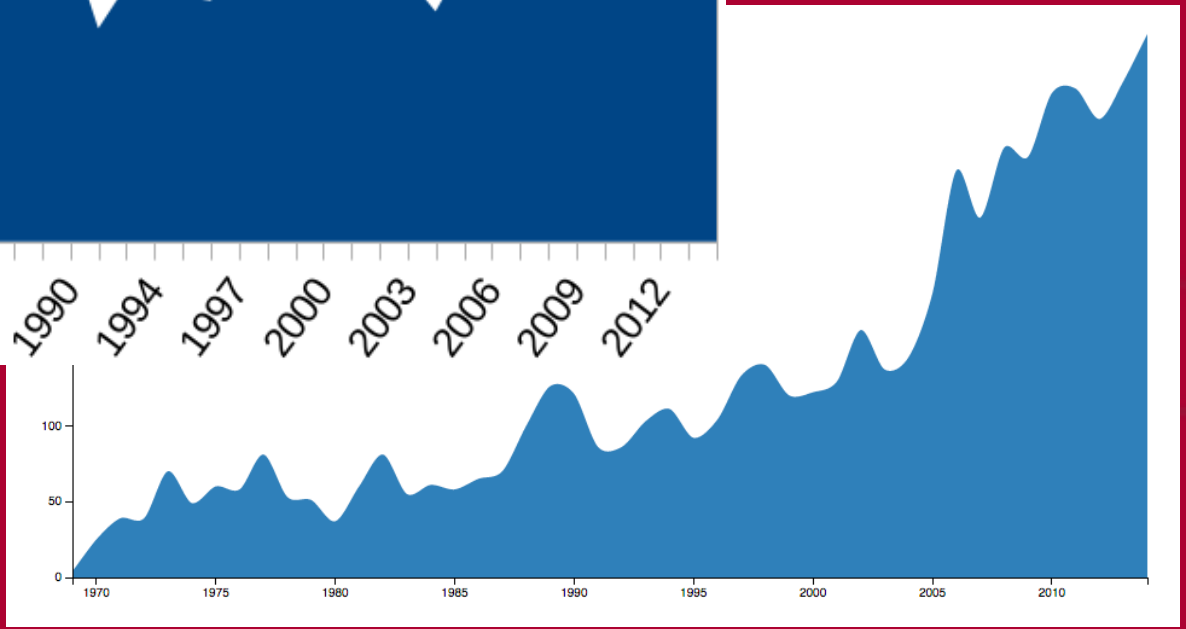


Evolution (papers published/year)



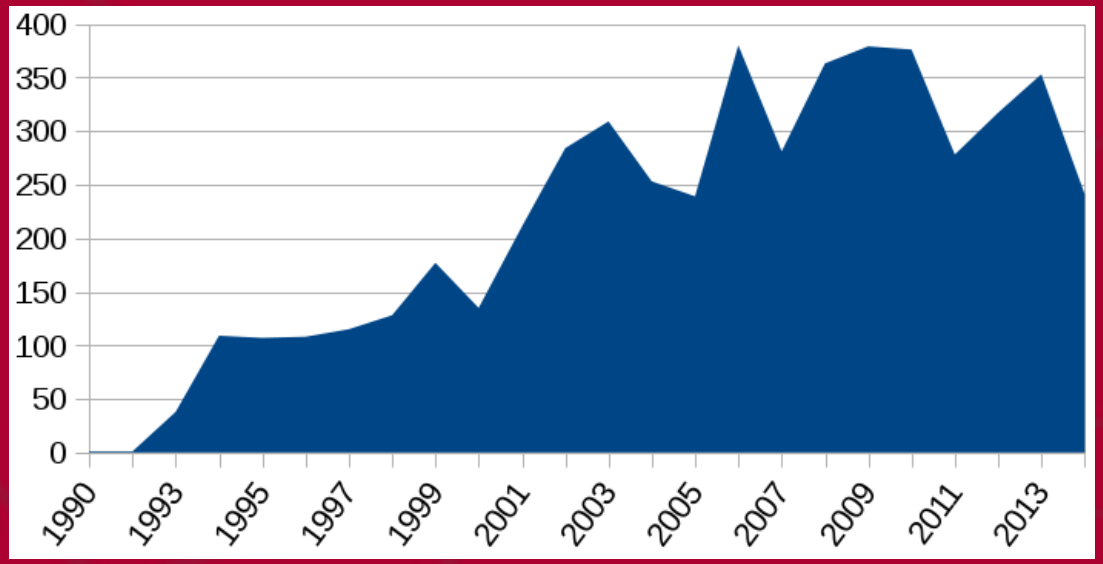
Web of Science

Scopus



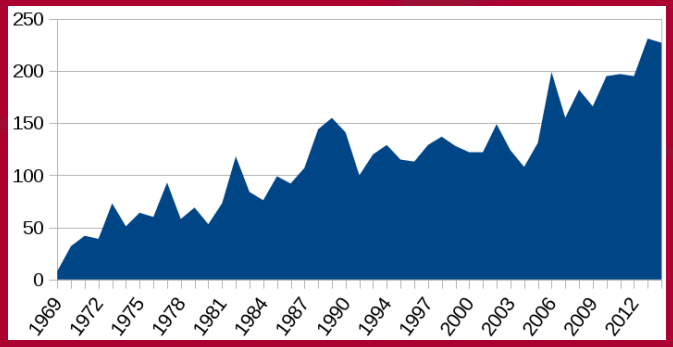
Evolution (papers published/year)

PubMed

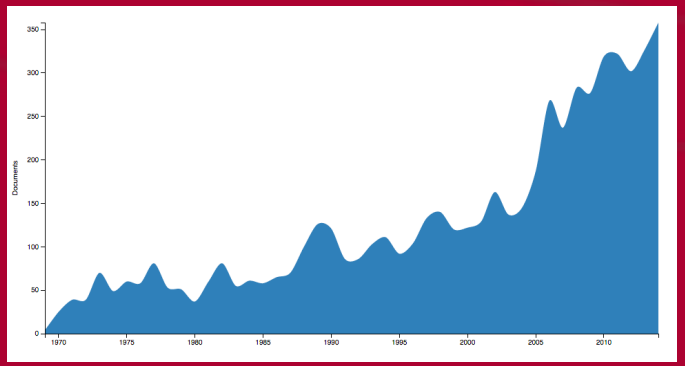


International Congress on Peer Review and Biomedical Publication
2013, 2009, 2005, 2001, 1997, 1993 and 1989

Web of Science



Scopus



Three datasets

Time period: 1969 – 2014
Searched term: Peer review
Scopus: Article title or Author keywords

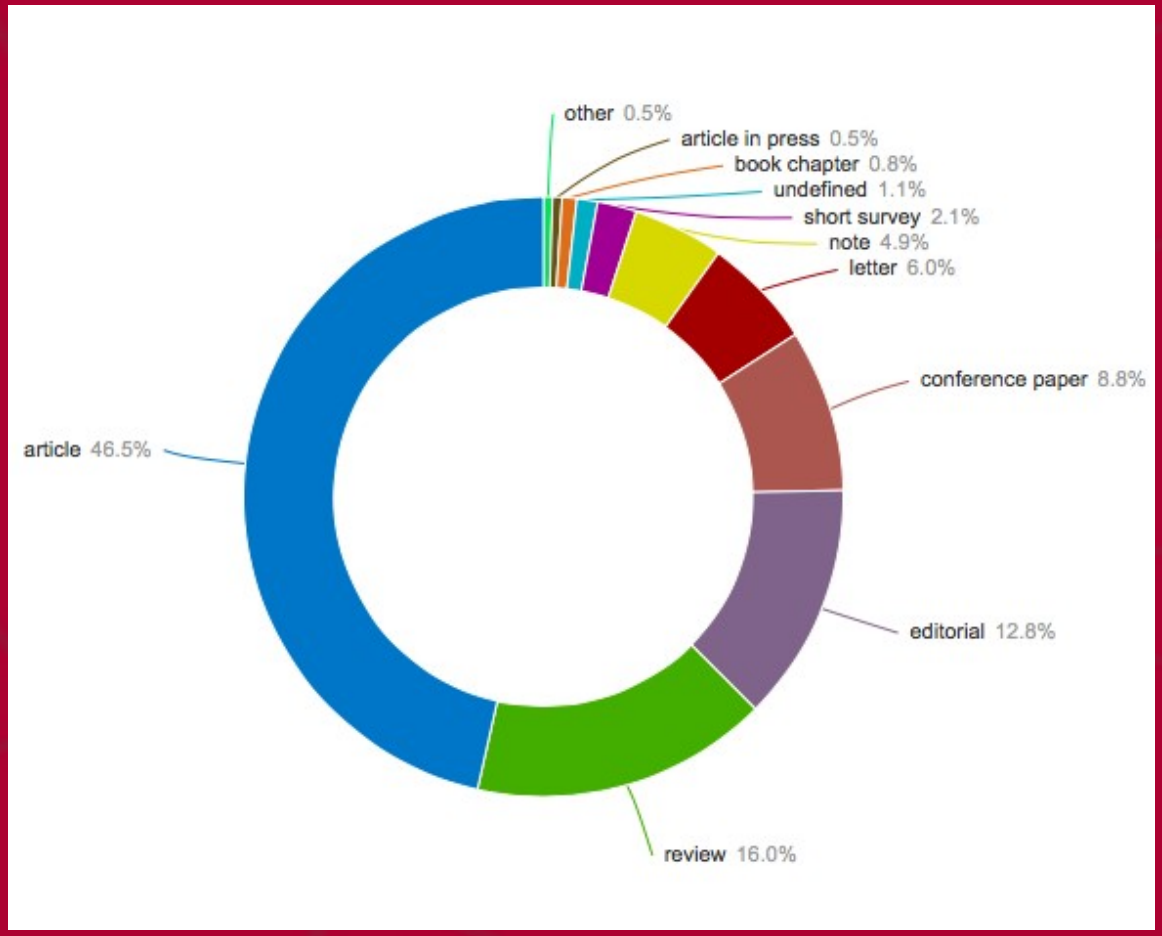
DS1 (5918 documents):
Any kind of document

DS2 (3281 documents):
Document types: article or conference paper

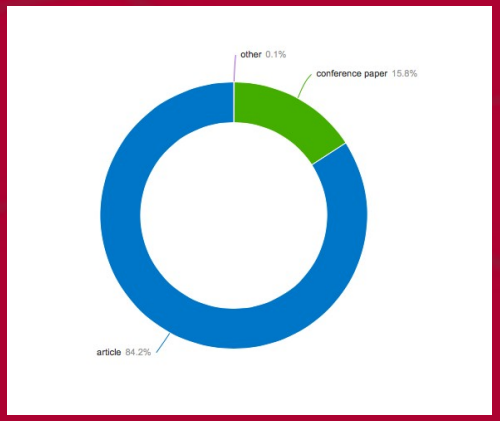
DS3 (1602 documents):
Documents not tagged as “Medicine”

Kinds of documents per dataset

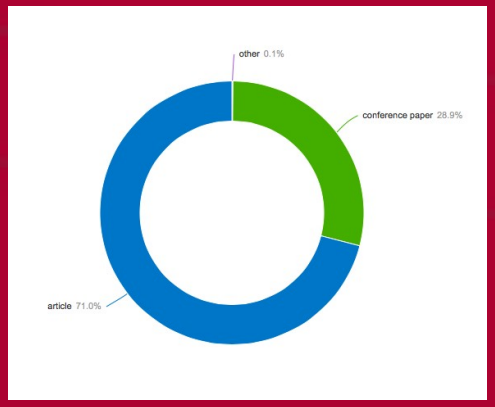
DS1



DS2



DS3



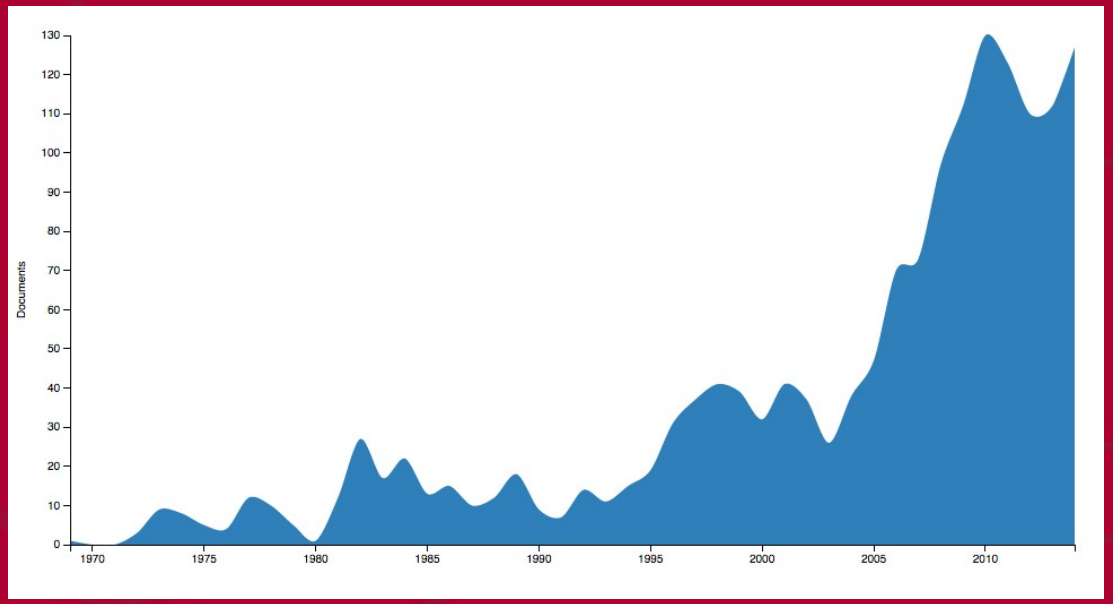
PEERE

From which disciplines?

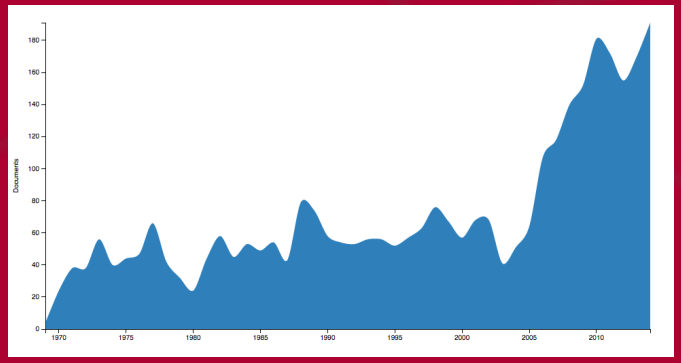
Subject	DS1	Subject	DS2	Subject	DS3
Medicine	50.7	Medicine	51.3	Social Sciences	35.1
Social Sciences	18.3	Social Sciences	20.1	Computer Science	23.8
Nursing	8.1	Computer Science	11.8	Engineering	14
Computer Science	7.9	Nursing	8.7	Nursing	10.5
Engineering	5.3	Engineering	6.89	Management	5.1
Multidisciplinary	5.2	Biochem / Mol Bio	3.16	Psychology	5
Biochem / Mol Bio	4.6	Psychology	3.04	Mathematics	4.8
Management	3.7	Management	2.68	Arts & Humanities	4.4
Health Professions	3.1	Health Professions	2.59	Agri/Bio Sciences	3.2
Agri/Bio Sciences	2.8	Mathematics	2.53	Env. Science	3.09
Other	23	Other	19.7	Other	27.2
Undefined	2.4	Undefined	3.58	Undefined	7.35

Evolution (papers published/year)

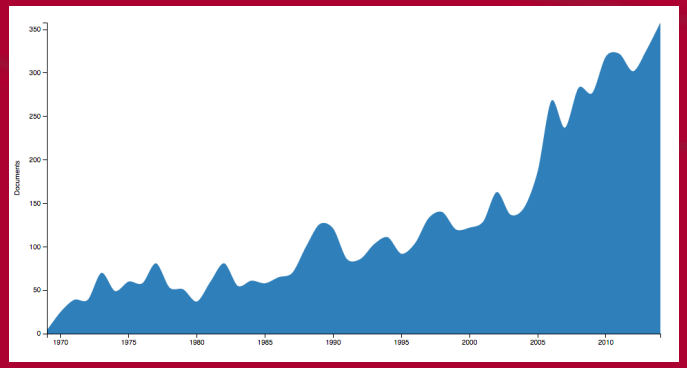
DS3



DS2



DS1



Editorials & research in Medicine do not rule the field in a different way.

Which journal to look at for interesting peer review references

Journal name (DS1)	*PR citations*	*Total citations*	*% of PR citations*
Journal of the american medical association (JAMA)	2150	124822	1.72
British Medical Journal (BMJ)	987	85434	1.15
Nature	750	590324	0.13
Science	741	537035	0.14
Scientometrics	730	5129	14.23
The New England Journal of Medicine	426	257469	0.17
Lancet	418	176528	0.24
Annals of internal medicine	283	47309	0.6
American psychologist	281	16533	1.7
Academic medicine: journal of the Association of American Medical Colleges	243	9372	2.6

Where to publish your research if it is not in the field of “Medicine”

Journal / Proceedings book (DS3)	Count
ASEE annual conference and exposition, conference proceedings	37
Scientometrics	36
Australian clinical review	31
Federal register	28
Lecture notes in computer science	26
Frontiers In Education conference proceedings	25
Professional psychology: research and practice	23
Journal of informetrics	18
Research evaluation	16
Nature	16

What papers should I read?

Publication (DS1)	Citations
Validation of an index of the quality of review articles	413
Nepotism and sexism in peer-review	372
Publication prejudices: An experimental study of confirmatory bias in the peer review system	269
Peer-review practices of psychological journals: The fate of published articles, submitted again	255
The power of gifts: Organizing social relationships in open source communities	186
The philosophical basis of peer review and the suppression of innovation	177
The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation	174
Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises	170
Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomised trial	161
Effect on the quality of peer review of blinding reviewers and asking them to sign their reports. A randomized controlled trial	149

Papers performing data analysis about Peer Review

Search for the terms:

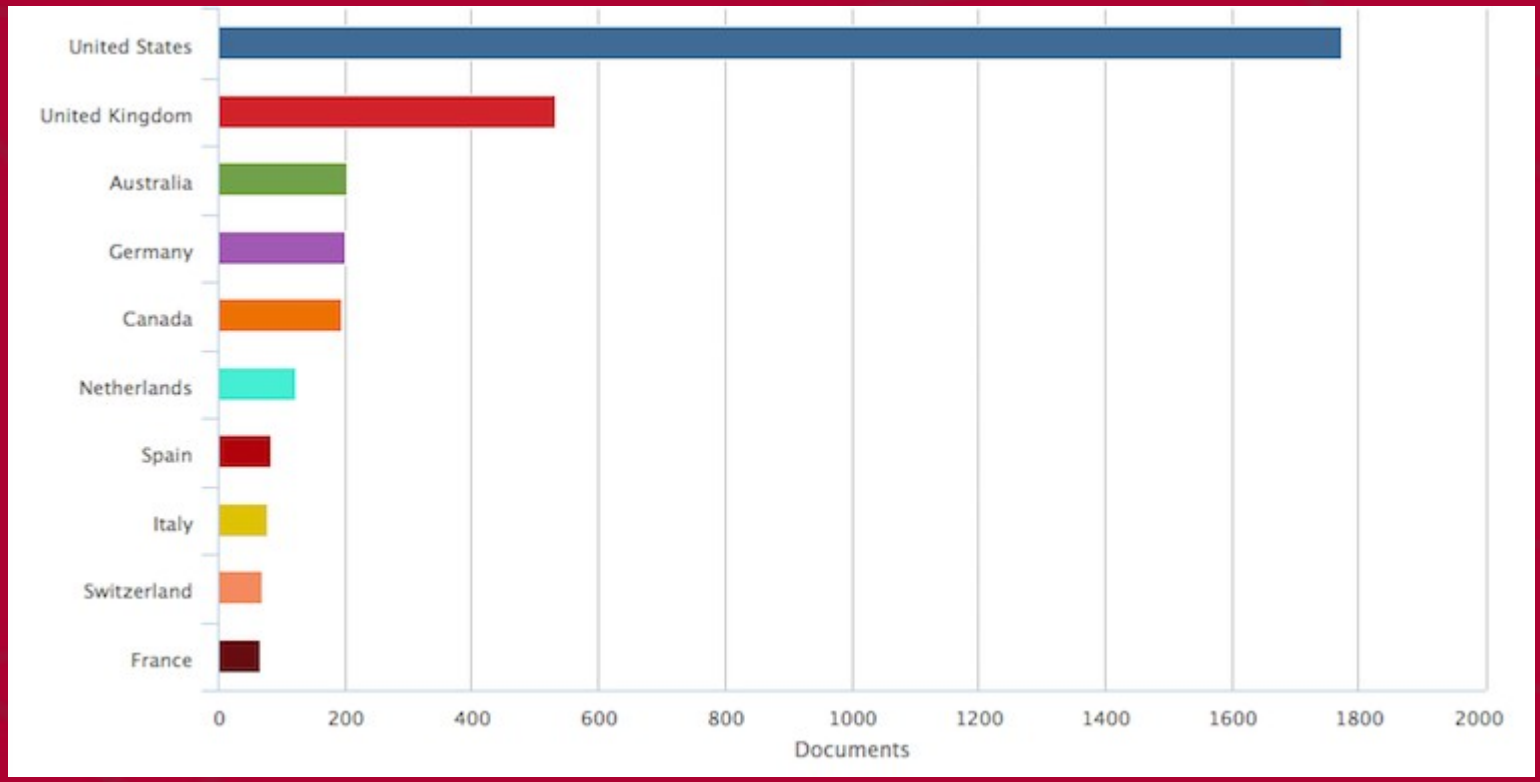
Data, survey, quantitative analysis, empirical analysis, experiment

In the fields:

Title, abstract, author keywords & indexed keywords

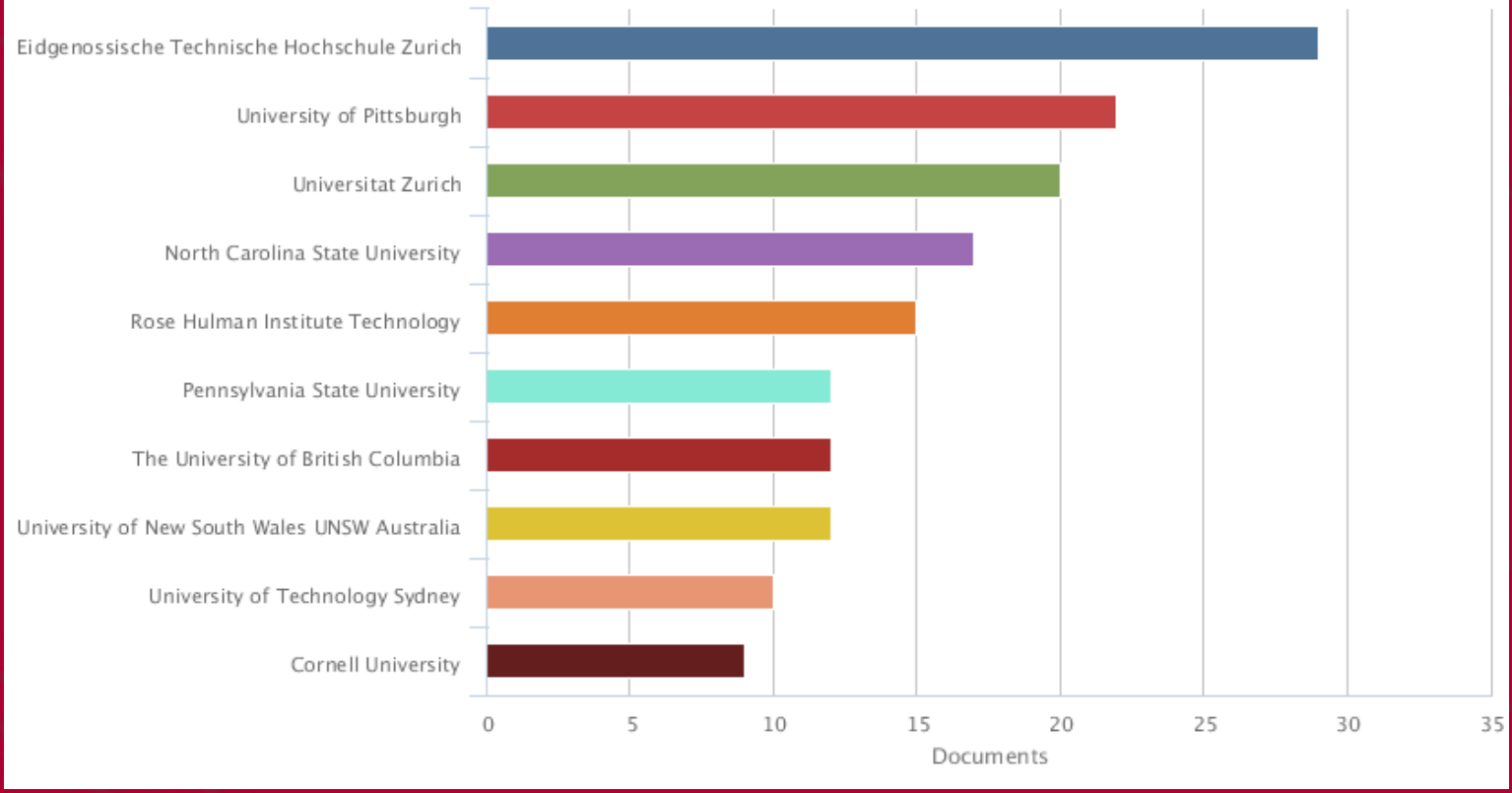
Only ~1100 out of ~6000 contain the targeted terms

Where is research work coming from?



Results for DS1 (DS2 & DS3 are almost identical)

From which institution?



Results for DS3
(DS1 & DS2 interleave medical centers after 3rd position)

Who are the more cited authors?

Authors (DS1)	Citations	Authors (DS2)	Citations	Authors (DS3)	Citations
smith r	509	bornmann l	294	bornmann l	270
bornmann l	500	daniel h-d	211	daniel h-d	194
godlee f	492	godlee f	195	schunn cd	108
daniel h-d	337	smith r	182	cho k	107
black n	327	black n	131	marsh hw	102
rennie d	287	rennie d	119	godlee f	95
evans s	283	evans s	114	smith r	94
van rooyen s	280	schunn cd	108	van raan afj	88
jefferson t	247	marsh hw	107	merton rk	78
davidoff f	244	cho k	107	harnad s	78

Who are the more social authors?

Author name	*Total*	*Avg.*	*Author*	*Total*	*Avg.*	*Author*	*Total*	*Avg.*
bornmann-l	64	1,49	bornmann-l	43	1,48	bornmann-l	40	1,48
daniel-h-d	54	1,59	daniel-h-d	39	1,5	daniel-h-d	36	1,5
rennie-d	35	1,75	jolly-b	27	6,75	carlson-p	24	2,18
bowie-p	35	3,18	cox-j	27	6,75	gehringer-ef	18	1,29
flanagin-a	31	2,38	bowie-p	27	3,38	yoshida-n	17	5,67
marusic-a	30	3,75	mckay-j	21	3,5	kula-rg	17	5,67
godlee-f	30	3,33	lough-m	21	3,5	iida-h	17	5,67
bloos-f	28	7	southgate-l	20	5	hamasaki-k	17	5,67
jolly-b	27	6,75	malicki-m	20	5	fujiwara-k	17	5,67
cox-j	27	6,75	godlee-f	19	3,17	mutz-r	16	2,29

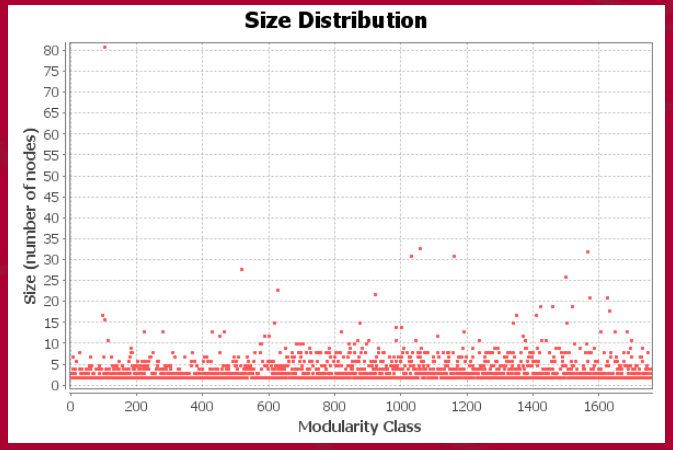
PEERE

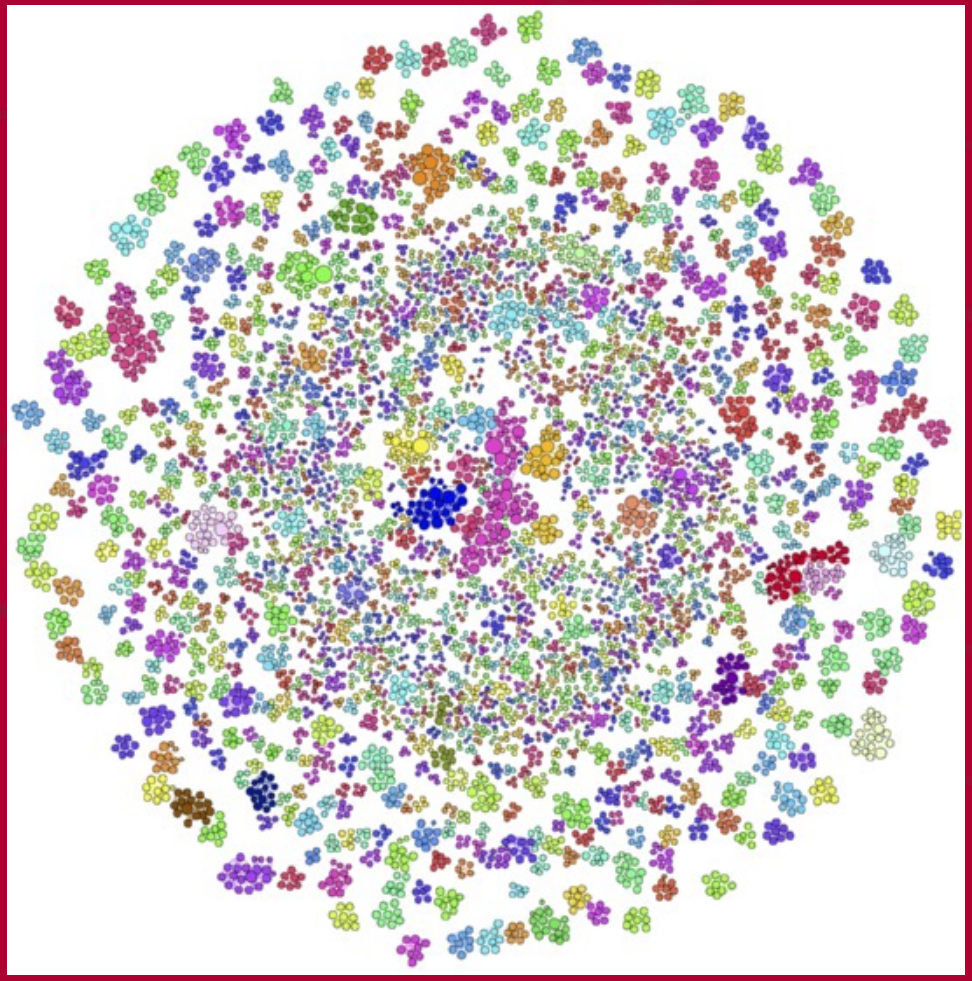
How are they interconnected?

Co-authorship network

Fragmented field

6775 authors
1760 communities





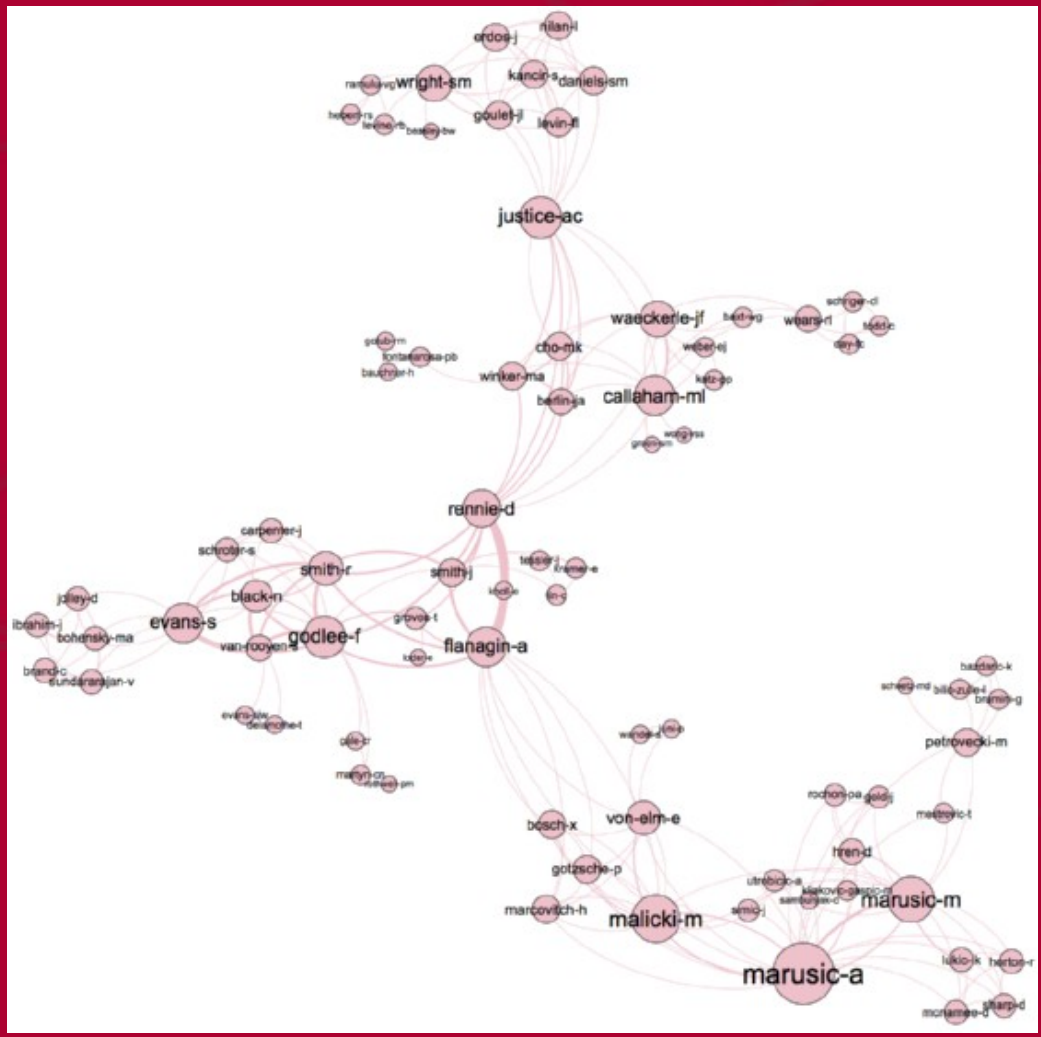
Where's Wally?

Zooming in on the biggest community

81 authors (1.2%)

Ana Marusic (WG1 leader) is the most collaborative

She is also ranked 3rd with respect to *centrality*



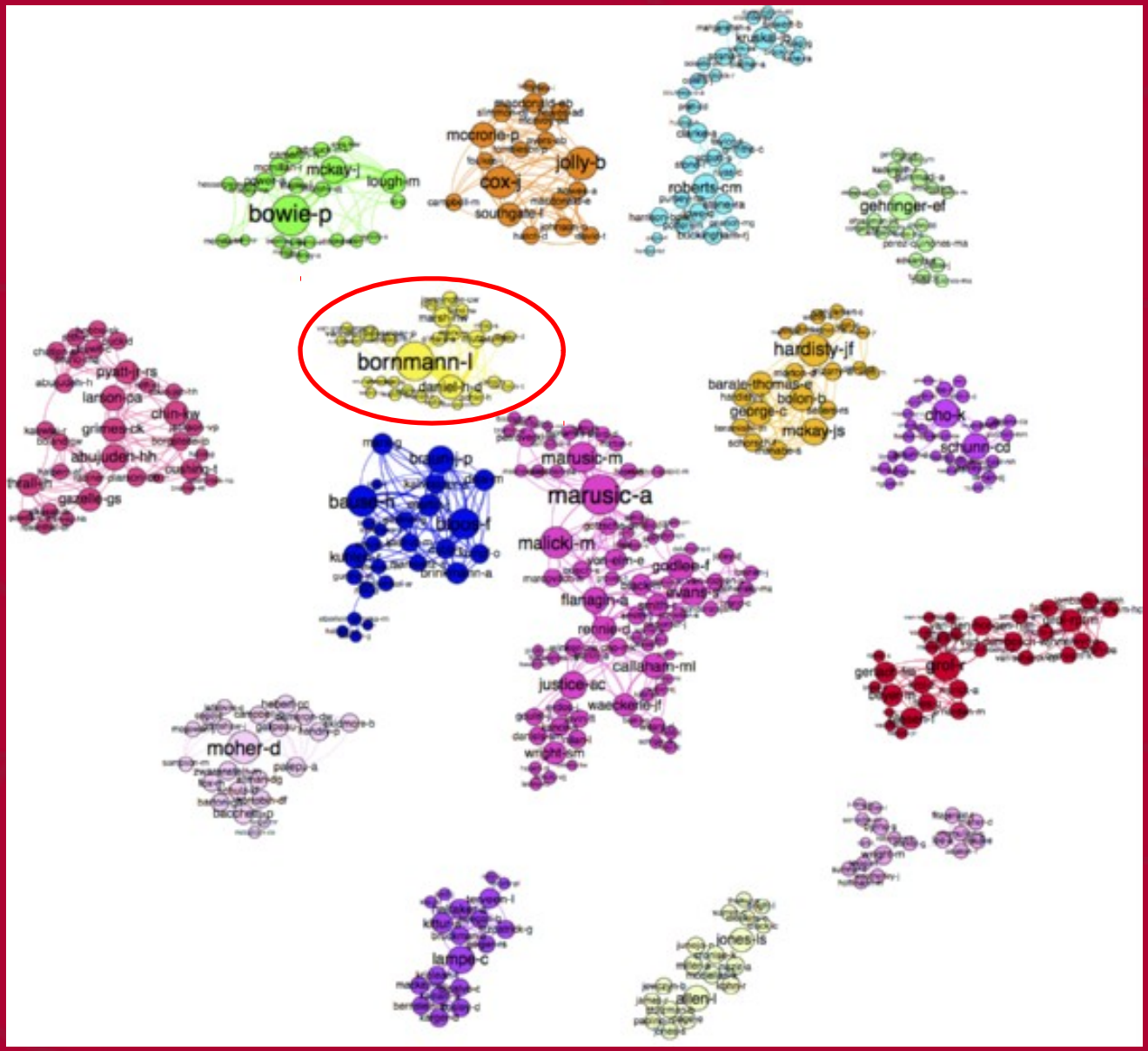
cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

PEERE

Who are the more central authors?

Author (DS1)	Centrality	Author (DS2)	Centrality	Author (DS3)	Centrality
rennie-d	814	flanagin-a	237	lampe-c	49
flanagin-a	765	rennie-d	208	cho-k	31
marusic-a	541.2	godlee-f	188	carlson-p	25
malicki-m	531	malicki-m	175.5	lee-r	24
justice-ac	517	bowie-p	118.5	van-leeuwen	24
godlee-f	320.7	marusic-a	104.5	moed-hf	24
marusic-m	311.2	smith-j	84	trevisan-m	24
bowie-p	220.5	evans-s	75	bornmann-l	16
wright-sm	200	grol-r	57.3	kittur-a	16
martin-j	169.5	mckay-j	56.5	gehringer-ef	16

cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

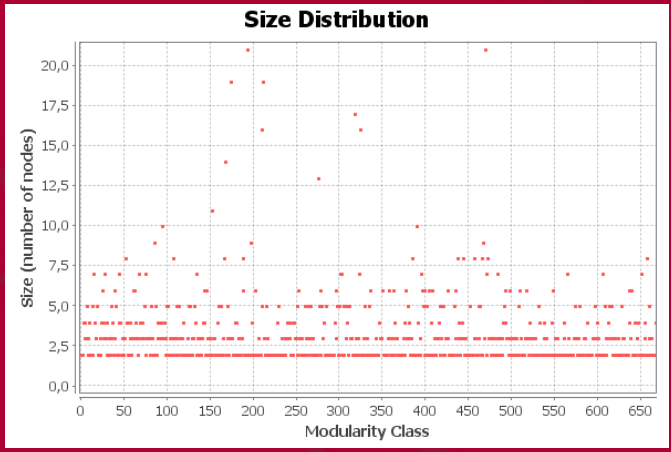


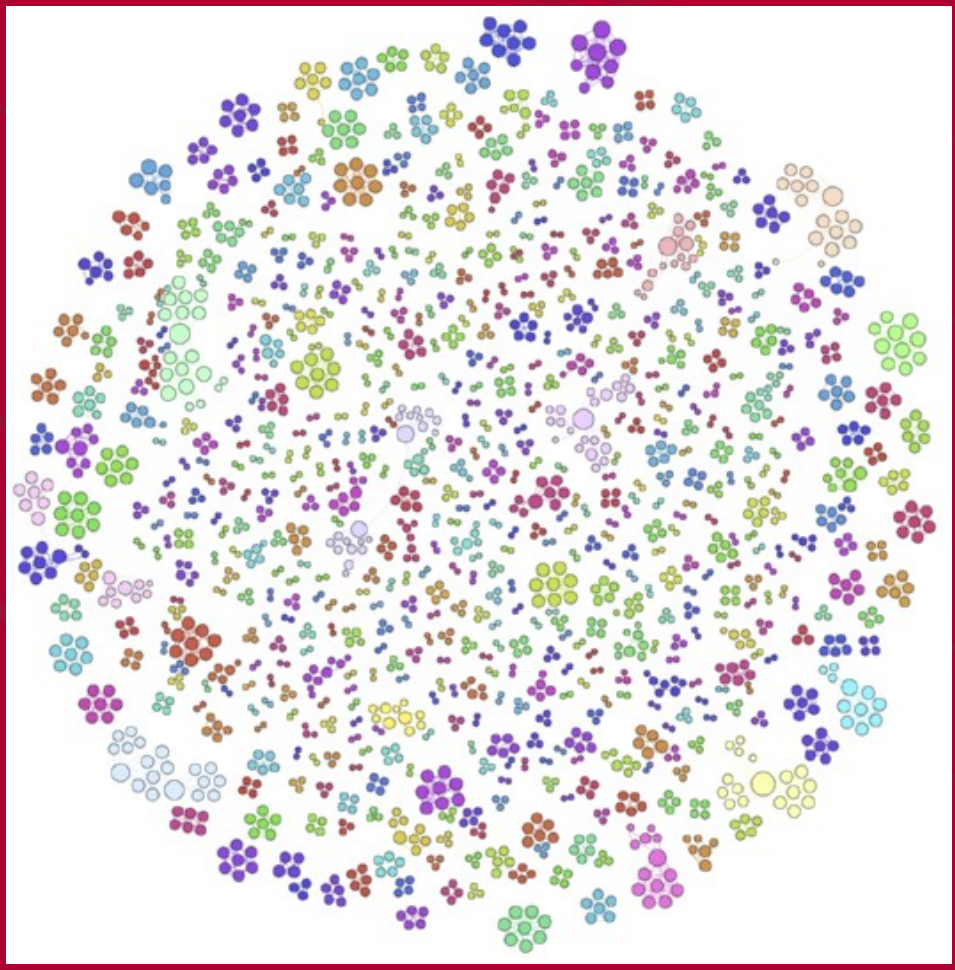
Any other "Wallies" around?

Things can be even worse in DS3

Co-authorship network

The communities have less than 5 researchers



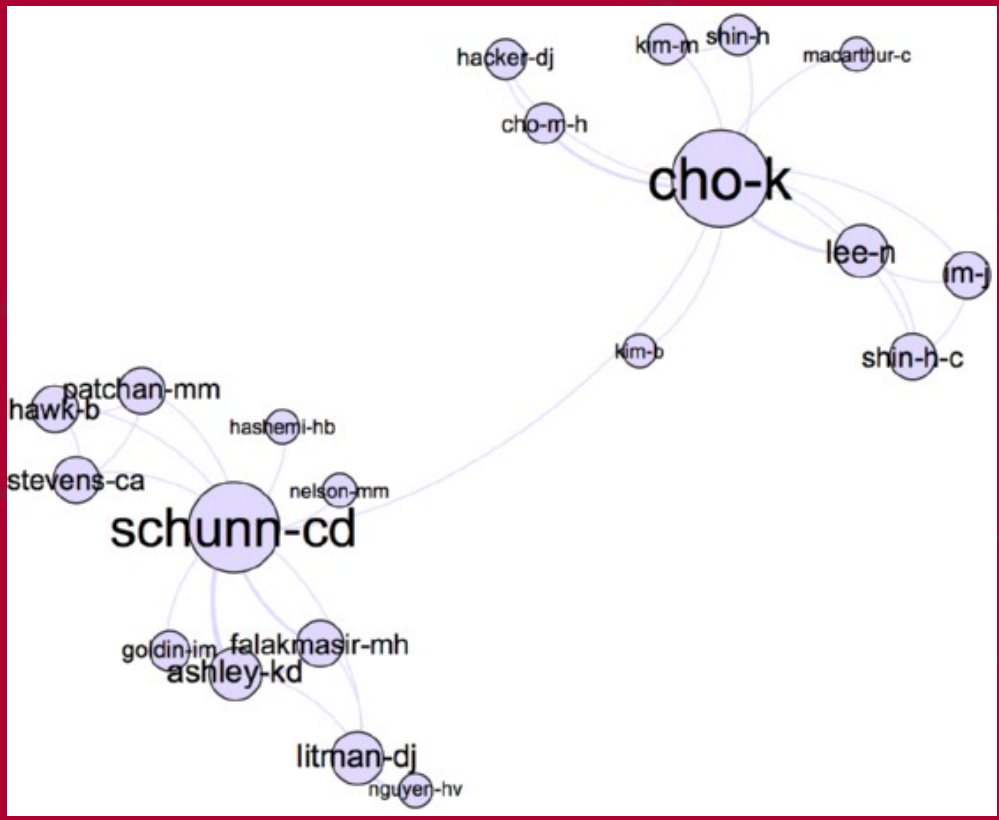


From which discipline is Wally from?

Zooming in on the biggest community

Kwangsu Cho
Cognitive science
Yonsei Univesity, Seoul

Christian D. Schunn
Psychology
University of Pittsburgh, USA



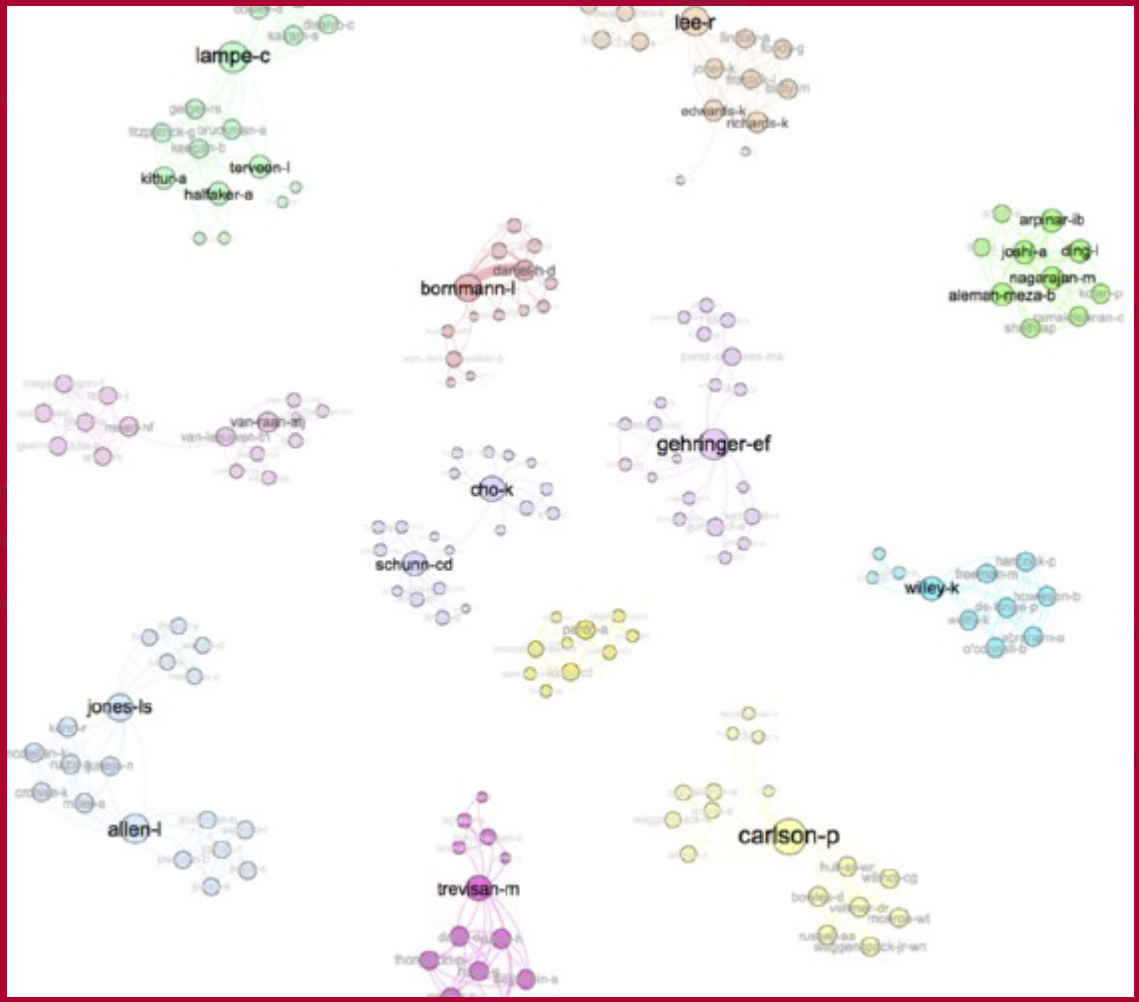
cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

Top 12 biggest communities

Communities with more than 10 authors

187 researchers

8.3% of the network



cost
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY

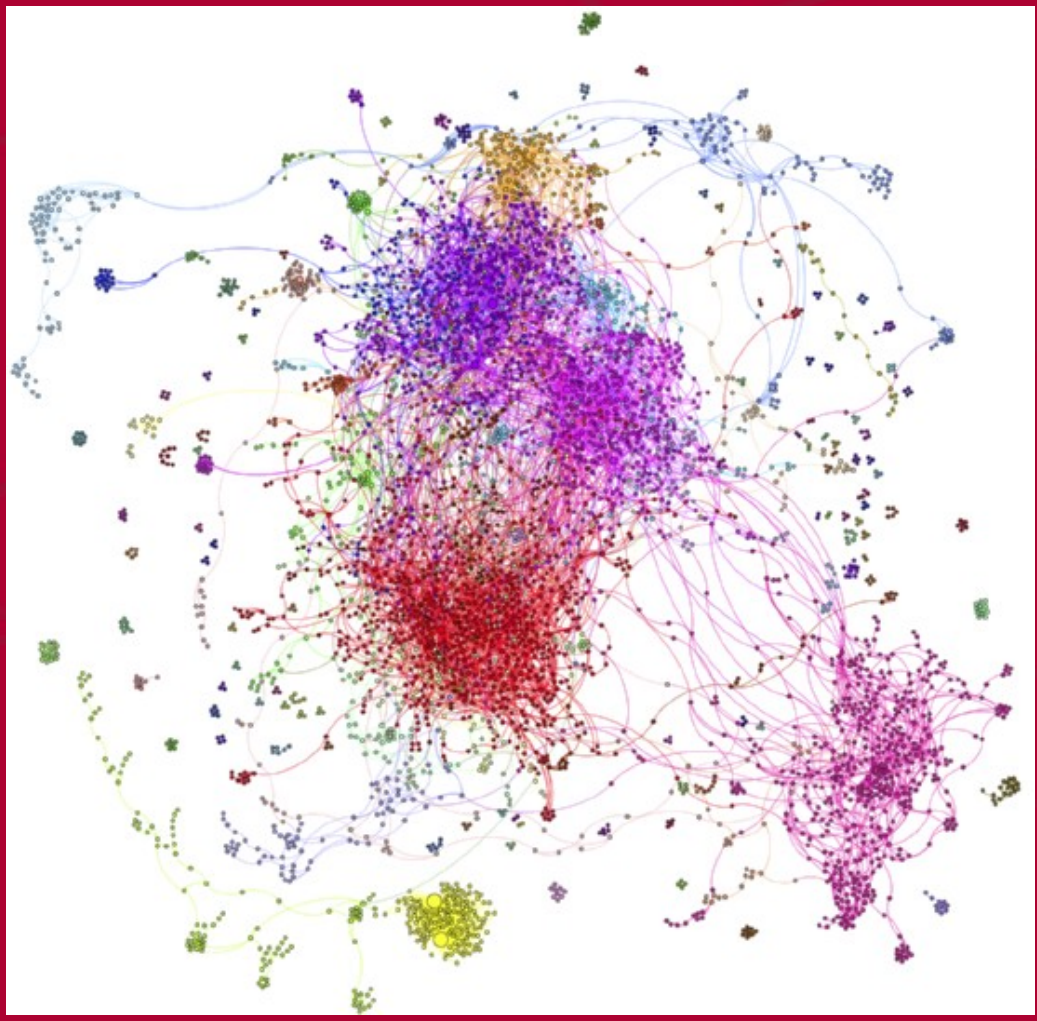
But there is still hope...

Citation network DS1

4316 nodes

132 clusters

The biggest community connects the 82.5% of the network.



… we live in the best of all possible worlds

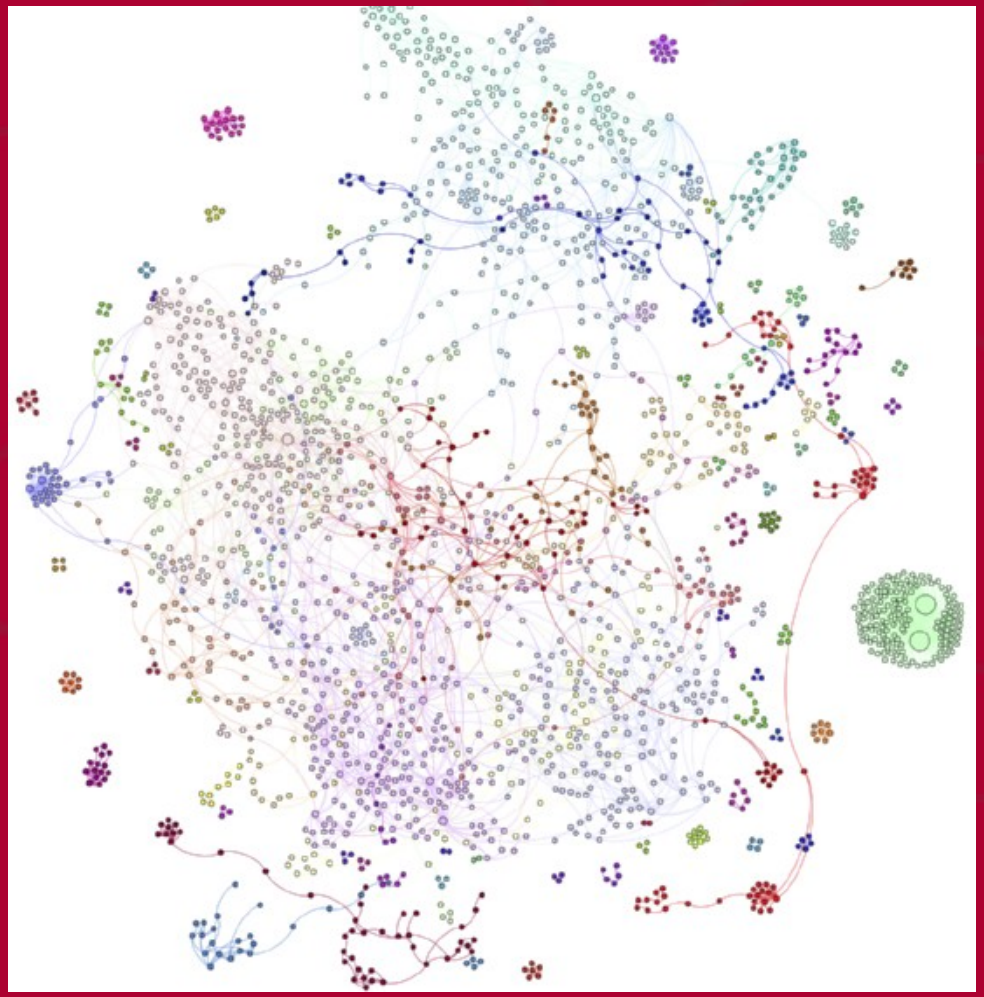
Citation network DS2

2229 nodes

95 clusters

The biggest community connects the 65.6% of the network.

We have even found a phantom!!



Meanwhile… the first PR dataset arrived!

JASSS: Journal of Artificial Societies and Social Simulation

Some raw data:

- 1272 paper submissions (1997-2012)
- 75% overall rejection rate
- Editorial decision within 60 days
- Average report time by referees 30 days
- Multidisciplinary

Empirical analysis of referees' attitudes in a multi-disciplinary journal

Sample data

Data included 915 submissions, information on 1819 authors and referees, i.e., authors (921), referees (317) and both authors and referees (581)

For each submission, we had: review acceptance/refusal, referee name and recommendation, time taken for reporting, text of the report, editorial decision

For each personal record, we had: disciplinary background, status (e.g. full/associate professor, private researcher or junior researcher), position in the journal, i.e., membership in the journal's editorial board, affiliation and a rating of referee quality assigned by the editor

Status	Number of requests	Number of refusals	Percentage
Professors	1178	145	12.31
Private researchers	162	19	11.73
Junior researchers	1222	112	9.7

referee status	Mean	Std. Err.	[95%	Conf. Interval]
full/associate professors	28.21	0.94	26.37	30.05
private researchers	27.63	2.09	23.53	31.73
junior researchers	25.23	0.74	23.78	26.69

referee status	Mean	Std. Err.	[95%	Conf. Interval]
full/associate professors	534.20	14.76	505.25	563.15
private researchers	532.57	36.23	461.50	603.64
junior researchers	622.98	16.17	591.27	654.70

Disciplinary background of the referees	Number of requests	Number of refusals	Percentage
Humanities	152	12	7.89
Social sciences	430	46	10.70
Behavioral sciences	199	26	13.07
Physics	162	26	16.05
Environmental sciences	122	15	12.30
Computer sciences/engineering	742	53	7.14
Math	42	4	9.52
Geography	76	8	10.53
Medicine	35	1	2.86
Economics	400	64	16.00
Management	199	21	10.55

background	revision time	Std. Err.	[95% Conf.	Interval]
Humanities	23.05	2.21	18.72	27.38
Social Sciences	27.07	1.49	24.15	29.99
Behavioral Sciences	27.57	1.93	23.78	31.36
Physics	31.26	2.99	25.40	37.12
Environmental Sciences	31.85	2.66	26.63	37.07
Computer Sciences / Engineering	24.75	0.92	22.95	26.55
Math	15.50	2.32	10.95	20.05
Geography	31.39	3.26	25.01	37.78
Medicine	19.38	3.95	11.63	27.13
Economics	29.31	1.68	26.01	32.60
Management	26.24	1.93	22.45	30.02

background	revision length	Std. Err.	[95% Conf.	Interval]
Humanities	551.2	45.9	461.1	641.4
Social Sciences	648.0	29.8	589.4	706.6
Behavioral Sciences	576.6	38.7	500.6	652.5
Physics	533.8	40.5	454.3	613.2
Environmental Sciences	662.3	45.2	573.6	751.0
Computer Sciences / Engineering	545.6	17.7	510.8	580.5
Math	497.0	72.8	354.0	639.9
Geography	514.7	44.2	427.9	601.5
Medicine	544.0	67.8	410.9	677.2
Economics	541.6	26.7	489.1	594.1
Management	668.1	42.7	584.2	751.9

recommendation	report length	Std. Err.	[95% Conf.	Interval]
accepted	301.0	29.2	243.7	358.4
minor revisions	528.9	18.7	492.1	565.6
major revisions	696.7	17.6	662.0	731.4
rejected	534.8	19.4	496.6	572.9

PEERE "New Frontiers of Peer Review"
www.peere.org
peereinfo@peere.org

cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

Disciplinary composition	Number	Percentage
Mono-disciplinary referee pool	112	18.1
Multidisciplinary referee pool	506	81.8
Referees pool with two disciplinary backgrounds	362	71.5
Referees pool with three disciplinary backgrounds	141	27.8
Referees with four disciplinary backgrounds	3	0.5

Nature of the reviewing	Number	Mean citations	Std. Err.	[95%	Conf. Interval]
Mono-disciplinary	129	18.46	2.21	14.08	22.84
Multi-disciplinary	48	29.43	6.49	16.36	42.50

Yet more conclusions

Reviewing is context-specific as it is sensitive to disciplinary & status effects

Diversity premium

No reciprocity between authors/referees

A small-community effect? Cooperative tendency?

Loving

Paradiso

Yet to arrive…

PEERE

The fate of rejected papers in JASSS

To which extent does PR predict the internal/external success of a paper

Interrelate PR data …

Content of reviews, recommendations, rounds of review

… with

Journal of publication (e.g. IF) and number of citations

Thank you very much!