

Bibliometrics: benefits and pitfalls

Jiří Jirát

University of Chemistry and
Technology, Prague
(VŠCHT Praha)

Approach	Complexity	Information value	Tools/data needed
All articles are equal	low	miserable	
All articles in given journal and year are equal	medium	limited, rel. simple, but hides extreme values	Metrics for journals (IF, AI in Journal Citation Reports, SNIP, SJR at JournalMetrics.com)
Articles in given journal and year have different “values” depending on citations received	high	high, but complex and not easy to decipher	Requires special tools (InCites, SciVal) or data set + non-trivial processing

here we are just now



Data sets (Czech Rep.) – comparison

	Content coverage	Identification of authors and institutions from Czech Rep.	Main disadvantages
RIV	+ besides journals and proceedings also other document types — journal list not complete, narrow selection	extremely precise	• no citation data • limited to CR: no data about foreign pubs., authors, insts. etc. • journal metrics from external source
Web of Science	• journals, proceedings (+ books) • several indices, not everybody has everything (eg. Book CI)	identification of authors, institutions, affiliations not without errors	
Scopus	• journals, proceedings (+ books) • monolithic	dtto	

Journal characteristics

(or typical article from the journal)

	Age	In use in DB	Math knowledge required	Numerical calculation
Impact Factor (IF)	60 th of 20 th century	WOS	3rd class of Primary school (division of two integers from range 0-1000)	
Eigenfactor	~ 2007	WOS	1-2 univ. semesters of linear algebra (matrix eigenvalues)	numerical methods for sparse matrices
ArticleInfluence	~ 2007	WOS		
Source Normalized Impact per Paper (SNIP/SNIP2)	~ 2010(2012)	Scopus	grammar school with focus in math/ Bc. study at technical/nat. sci. univ.	
SCImago Journal Rank (SJR)	~ 2010	Scopus	dtto	iterative algorithm

Some properties

	Publication timeframe / years	Includes self- cites?	Typical range	Includes field normalisation?
Impact Factor (IF)	2 or 5	yes (* also IF without self-cites; self-cites penalization)	> 0, up to 30 – 40	NO ex-post normalisation relies on correct field categorization
Eigenfactor	5	excluded	Caveat: journal property, not article	yes
ArticleInfluence	5	excluded	normalised – article influence: 1.0 = average, > 1.0 above avg., < 1.0 below avg.	yes
Source Normalized Impact per Paper (SNIP/SNIP2)	3	yes	0 to cca 10 - 15 (even more)	yes
SCImago Journal Rank (SJR)	3	yes, but with upper limit 33 %	min. 0,1, up to cca 10 (even more)	yes

h -index

	Occurrence	Calculation	Math knowledge required
h -index	“standard” (WOS, Scopus etc.)	order desc. by citations, then for h : cites of h th article: $\geq h$ cites of $(h+1)$ th article: $\leq h$	2 nd -3 rd class at primary school (ordering and comparison of natural numbers in range 0-100, resp. 0-1000)

Metric (quantities) types

- extensive
 - *in physics*:
 - proportional to system size
 - additive
 - depends on institution size: „size-dependent“, „power metrics“ etc.
 - examples: number of publication, number of citations, ...
- intensive
 - *in physics*:
 - Physical quantity whose magnitude is independent of the extent of the system.
 - does not depend on institution size: „size-independent“
 - examples: citations/publications; fraction of publications in best percentile, quartile, ...

Some of them are function of field and/or time.

Example 1

COMPARING INSTITUTIONS

Simplified example – comparison of institutions based on one journal



Polymer journal: in Q1 (1st quartile) in ISI JCR journal ranking in category Polymer Science

Comparing:

- ETH Zürich
- vs.
- Virtual Czech SuperUniversity
(UK + VŠCHT + VUT + UPce + MU + UTB + VŠB-TUO + ČVUT)
(Note: some Czech Univ. not included, as they hadn't any publications in Polymer in studied timeframe.)

Publication in this journal in **2005 and 2006**.

Data source: Scopus Citation Tracker

Journal Metrics	
Source Normalized Impact per Paper (SNIP):	1.618
SCImago Journal Rank (SJR):	1.469
Impact Factor:	3.766
5-Year Impact Factor:	4.224
Imprint: ELSEVIER	
ISSN: 0032-3861	

Example 1 – comparing institutions

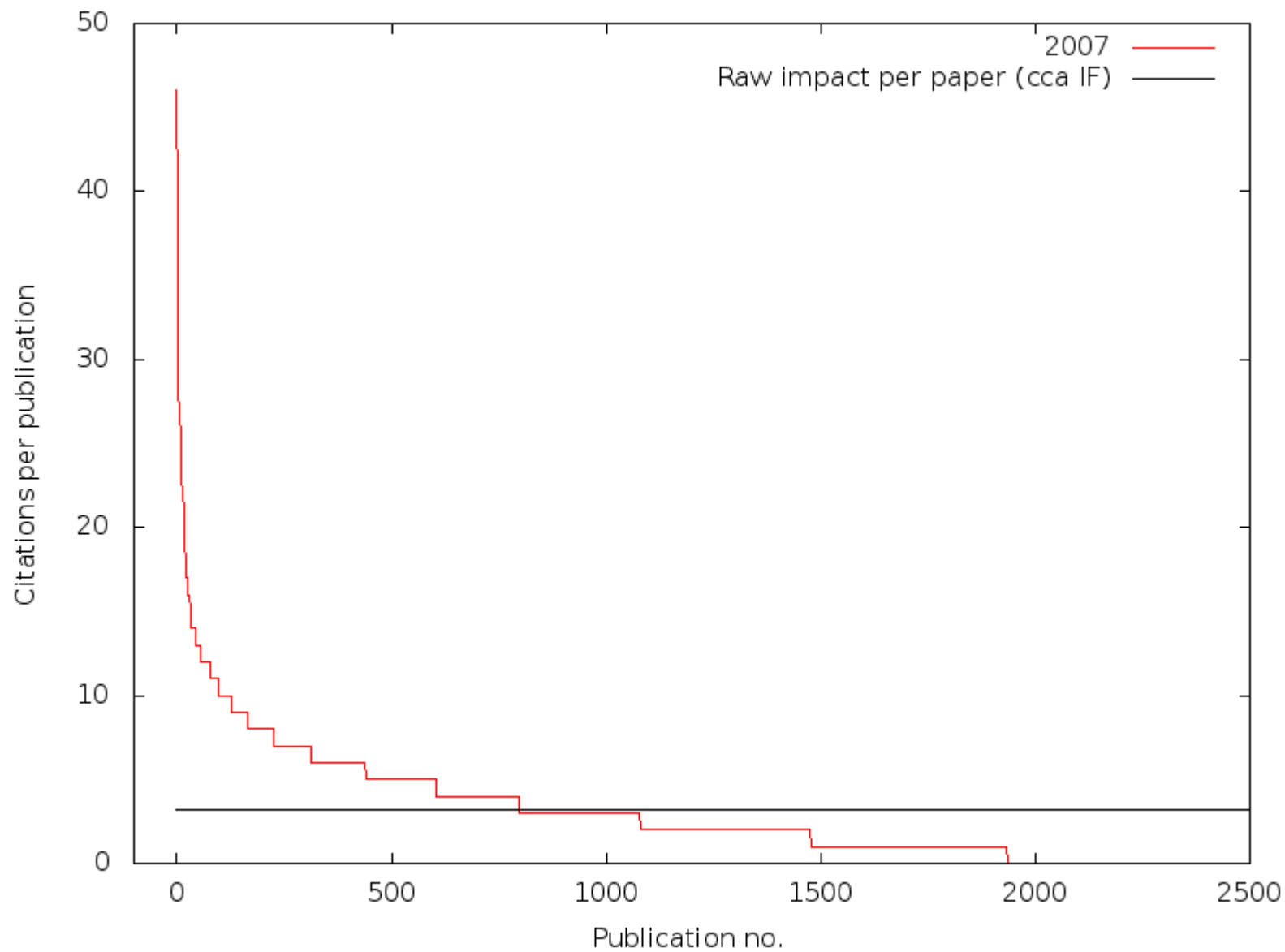
“raw Impact Factor in 2007” = $7714/2396 = 3,22$ (Scopus Citation Tracker)

for comparison: $IF(2007) = 3,065$ (2007 JCR Science Edition)

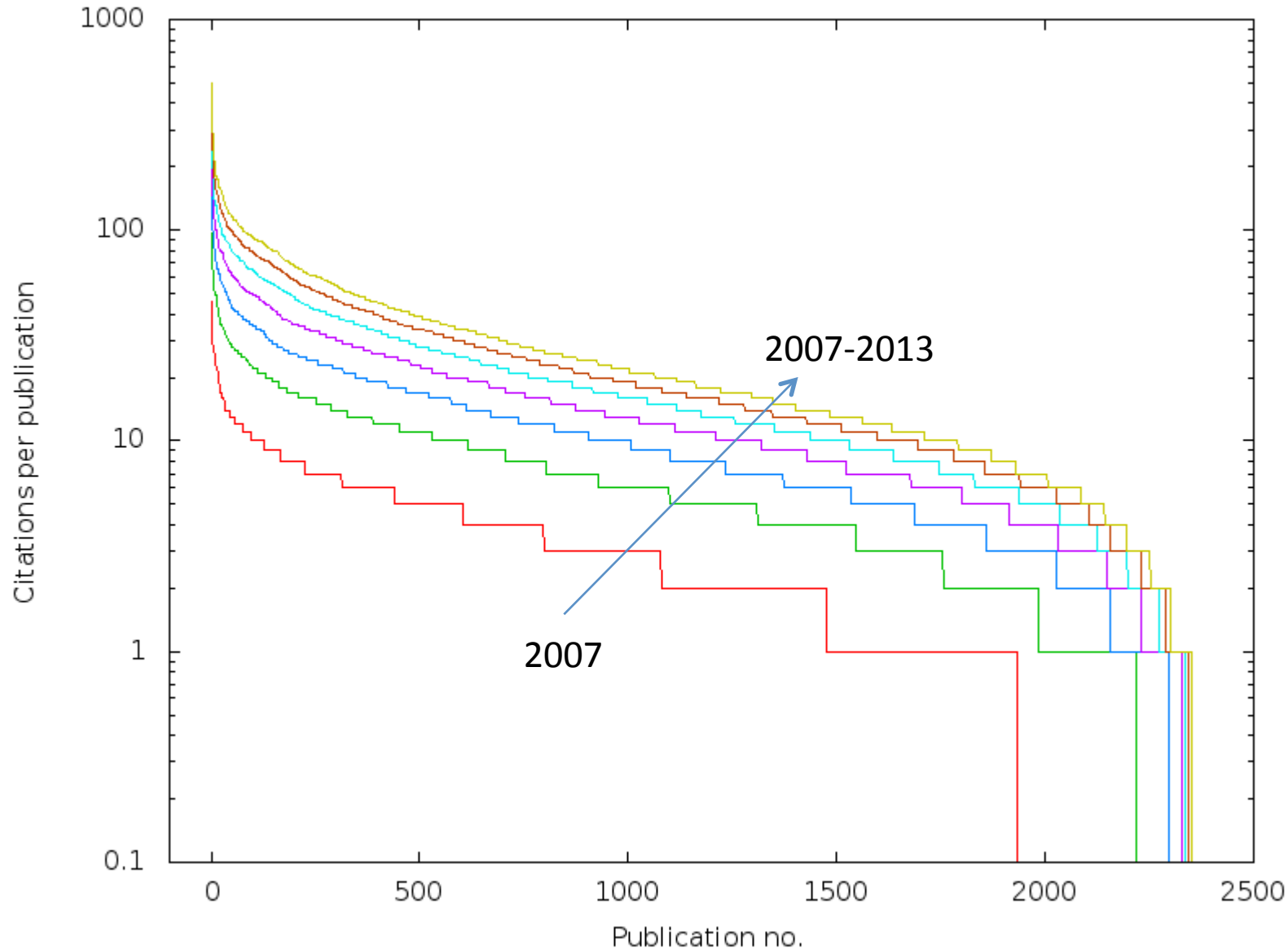
	ETH Zürich	Virtual Czech SuperUniv.
No. of publications (Polymer, 2005-2006)	13	18

What about next years?

Citations per publication: Polymer (2005+2006) in 2007

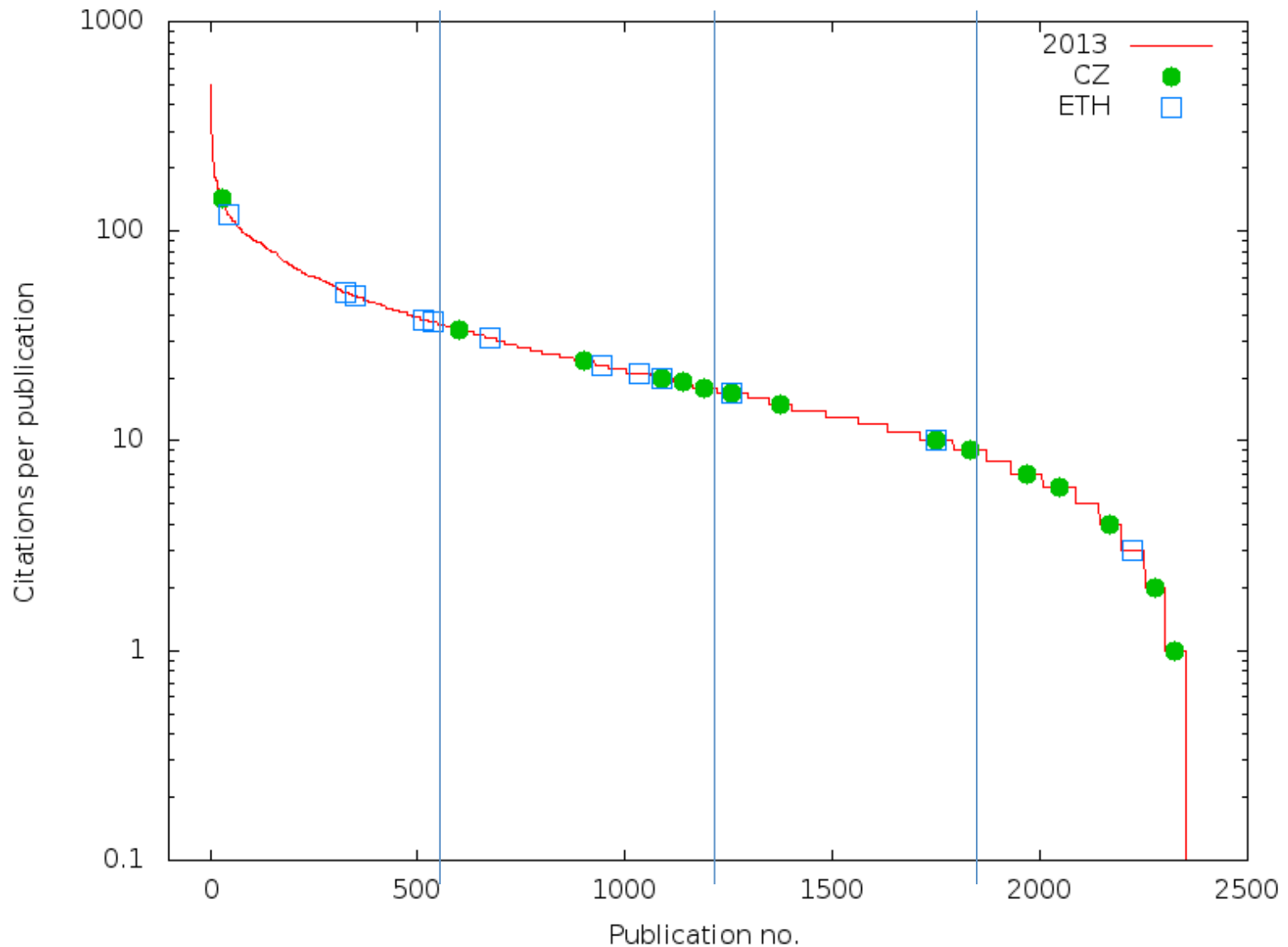


Development of cumulative citation count – in 2007-2013



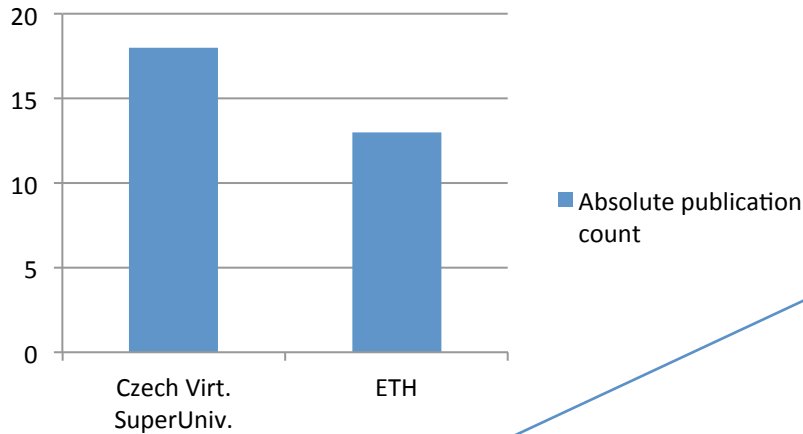
Note:
publications'
order
changes
from year to
year!

2013 – comparison of univs. (*Polymer* journal, 2005+2006)



Use percentiles as characteristics

Absolute publication count



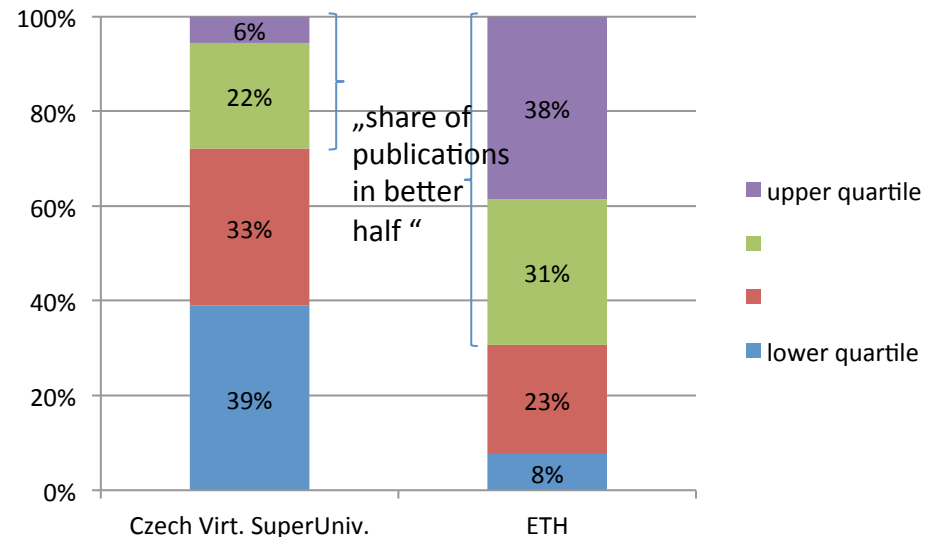
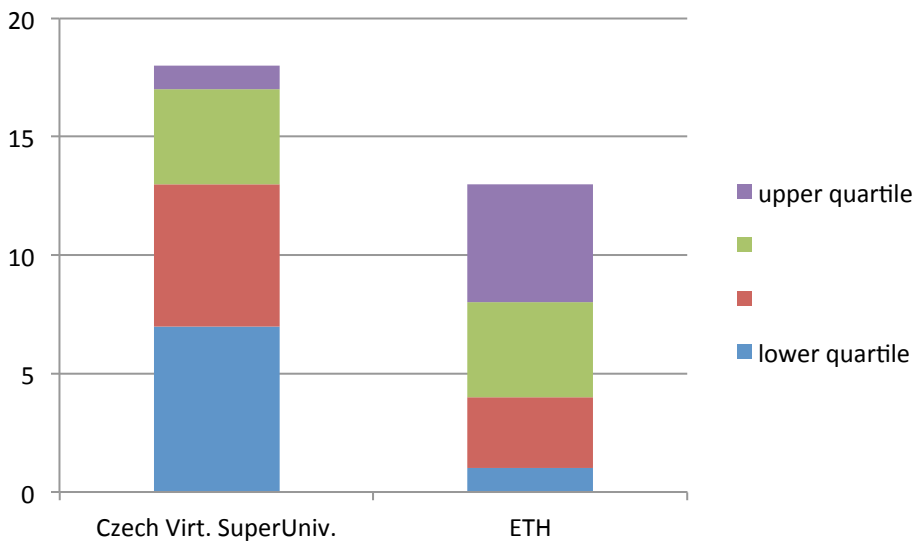
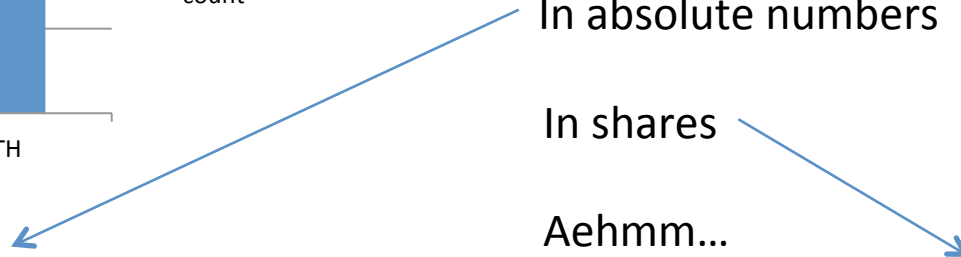
We are better!

What about quartile distribution by citations?

In absolute numbers

In shares

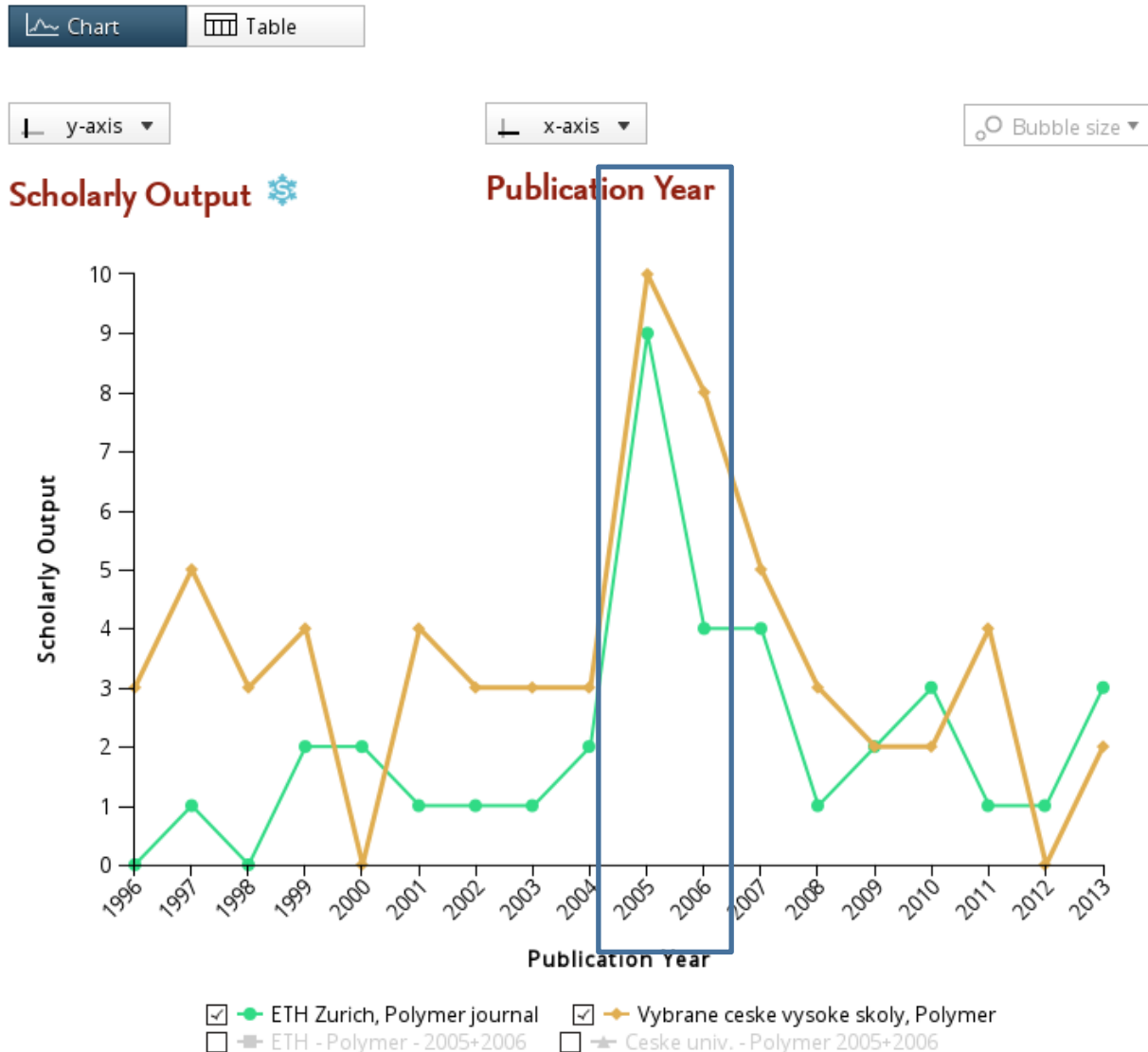
Aehmm...



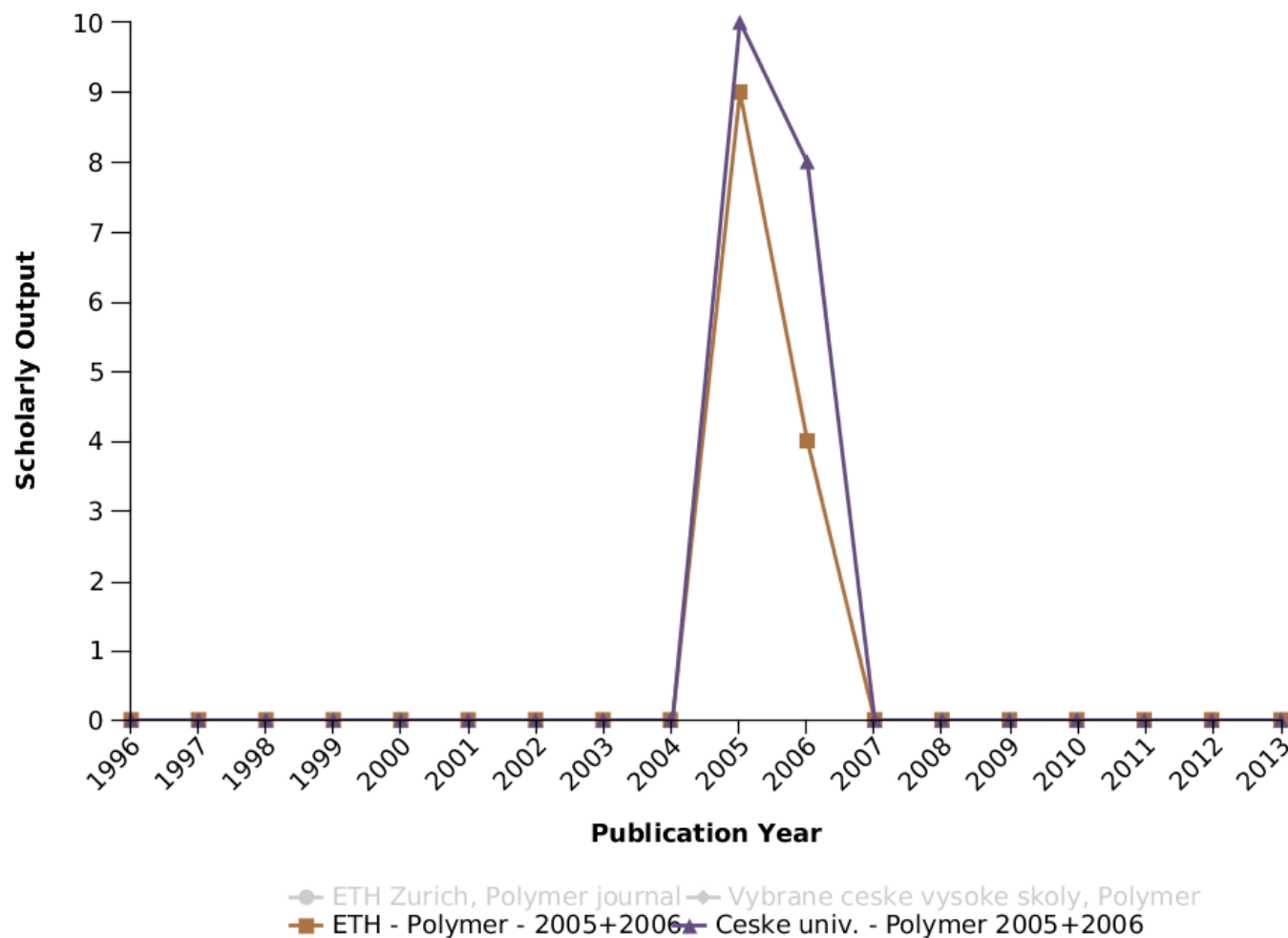
Comparison – bibliometric tools

- Comparison using one journal – limited information value
- Comparison within country, Europe, world?
 - without special tools extremely labourious/
impossible
- Comparison using SciVal:
 - two data sets created (DOI upload)
 - comparison both by “quantity” and “quality “

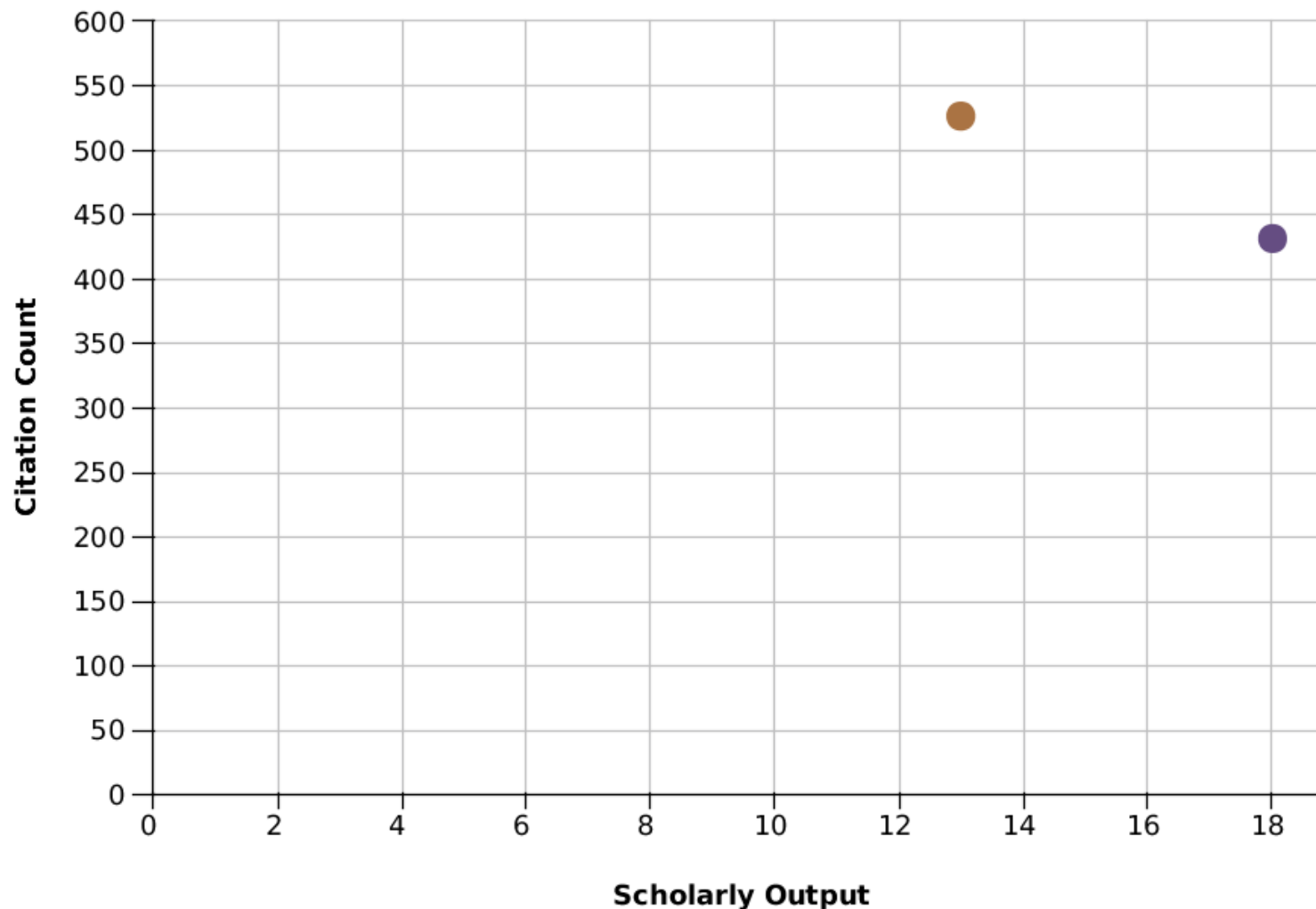
Institutions' publications in *Polymer*



Focus on *Polymer*, 2005+2006



Quantity vs quantity: Citation count vs article count

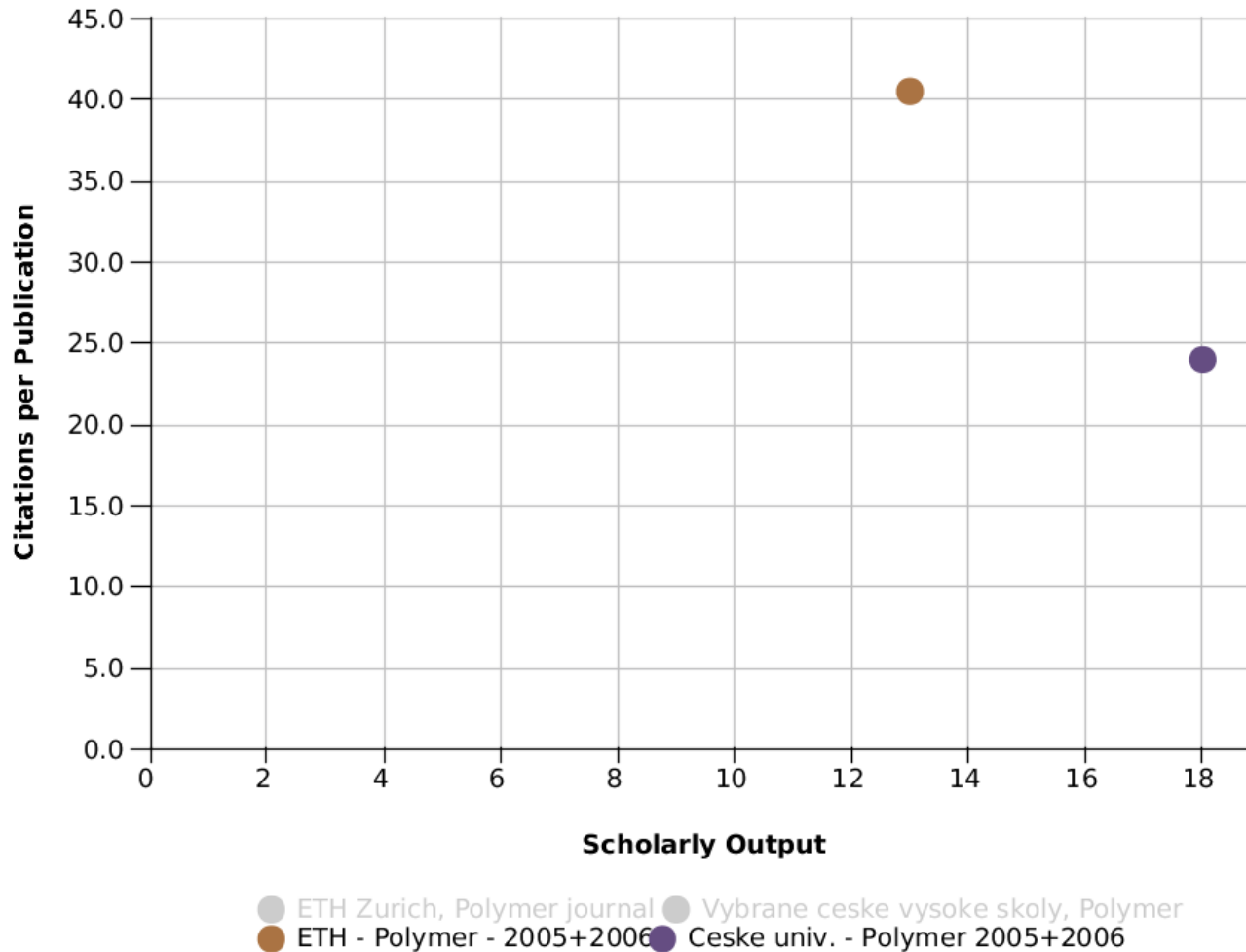


ETH Zurich, Polymer journal Vybrane ceske vysoke skoly, Polymer
ETH - Polymer - 2005+2006 Ceske univ. - Polymer 2005+2006

Zdroj: SciVal, Elsevier

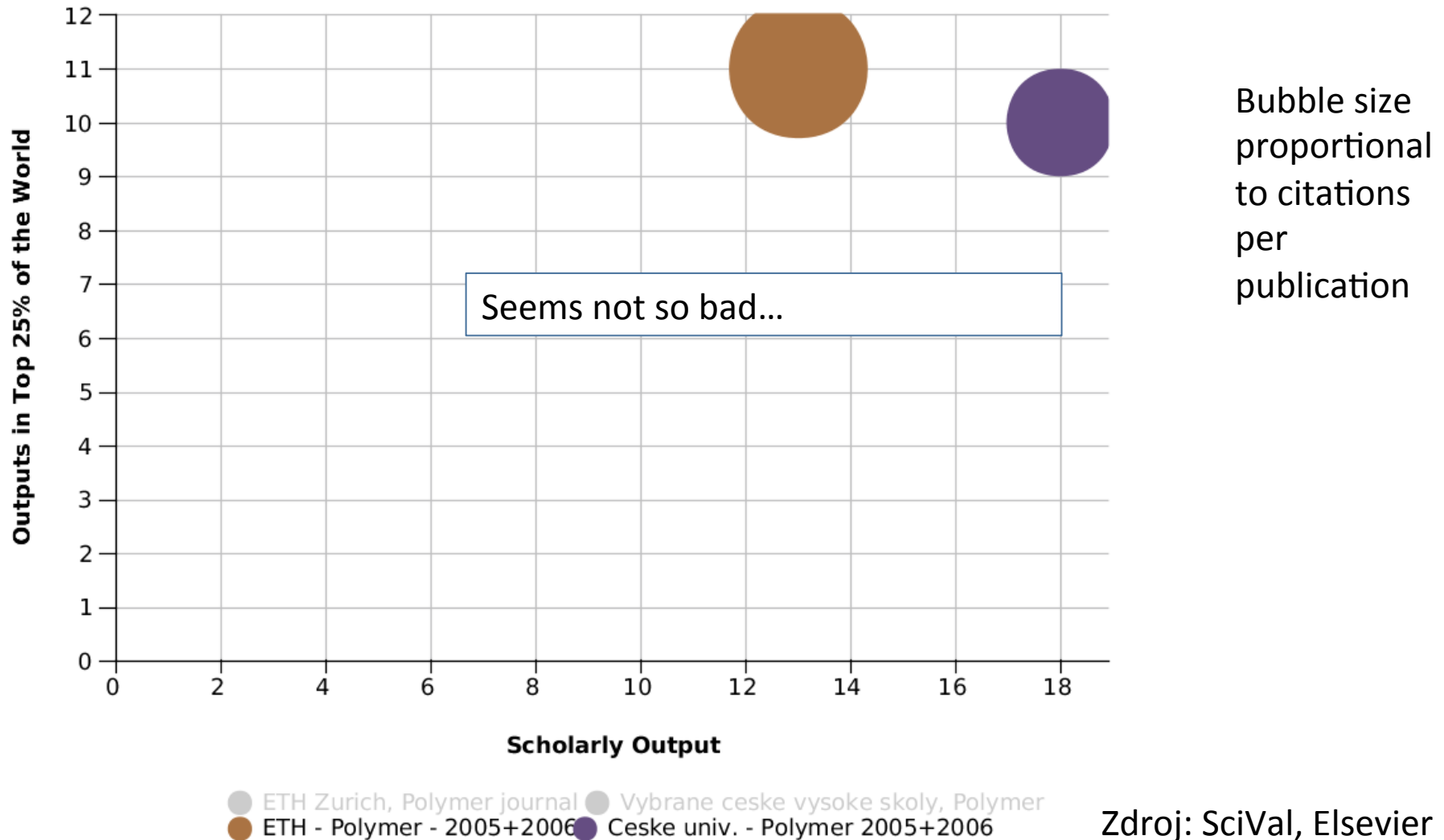
Intensity vs quantity:

Citations per publication vs article count



Quantity (quality selection) vs quantity vs intensity:

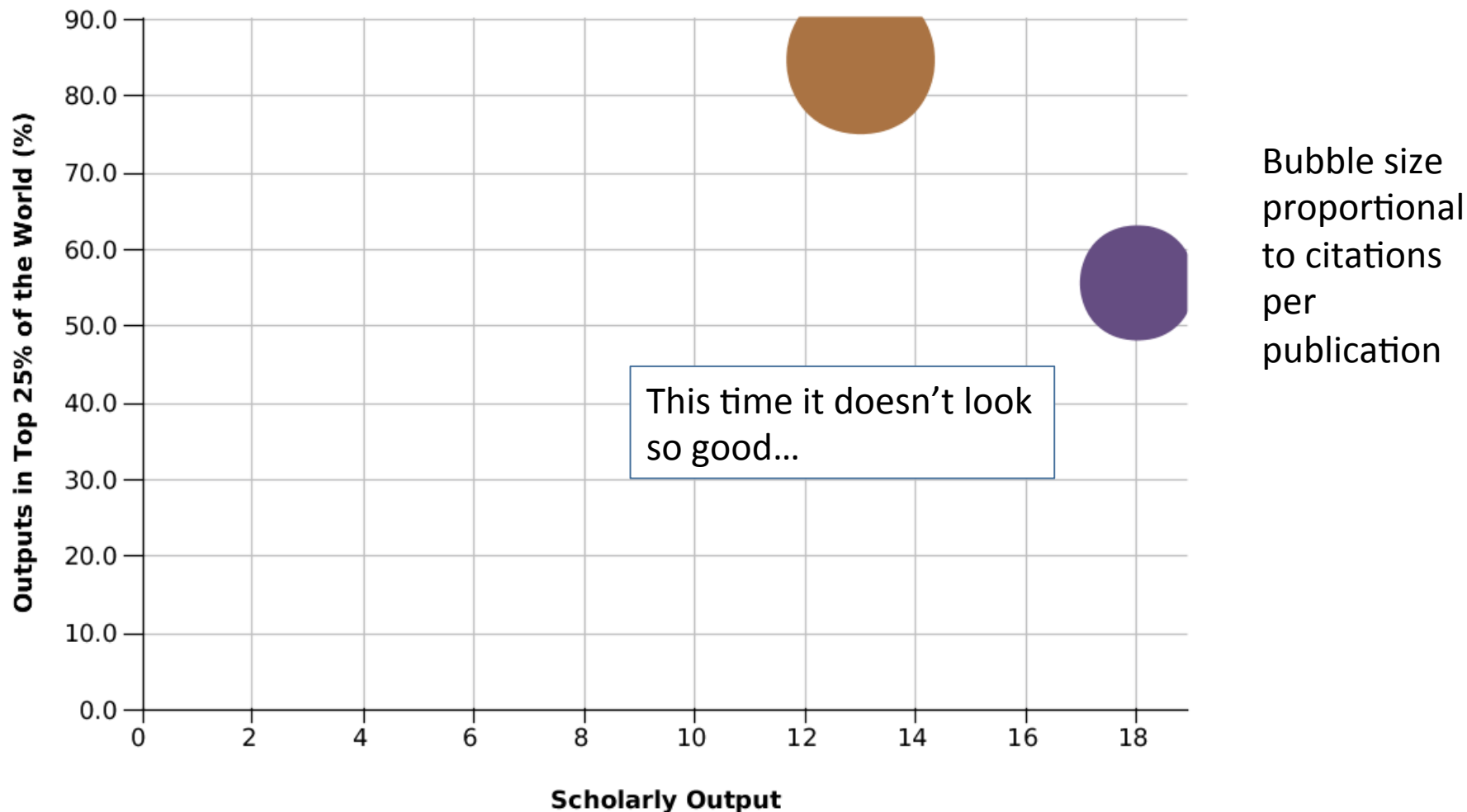
Article count in 25 % top articles worldwide vs article count vs article count per publication



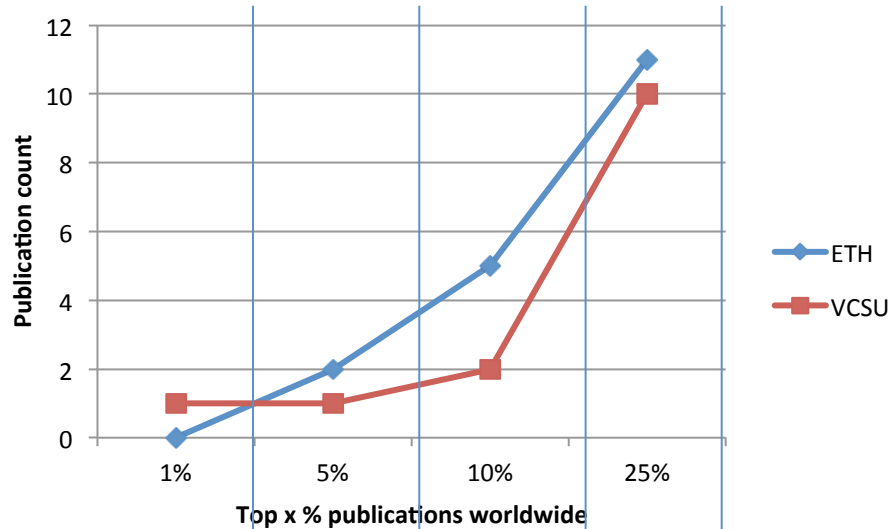
Zdroj: SciVal, Elsevier

Quality vs quantity vs intensity:

Fraction of articles in 25 % top articles worldwide vs article count
vs article count per publication



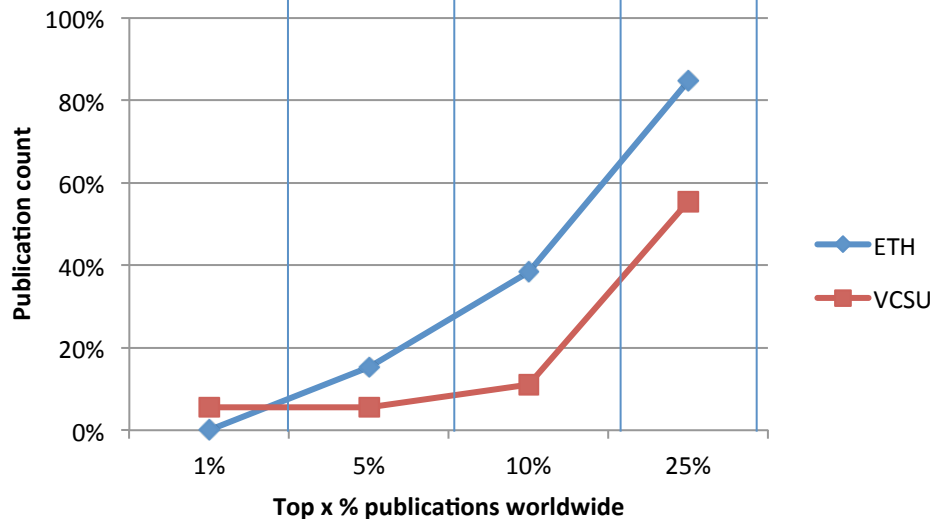
Looking for excellence – final decision



Where to put the delimiter to find “top” results?

If at 1 %:
VCSU wins

Else at 5 %, 10 % or 25 %:
ETH wins



Example 2

QUALITY OF RESEARCH GROUP

Traditional evaluation

Kde se v ČR dělá nejlepší výzkum

9.8. 2012 (DOPLNĚNO 20.12.2012)

ŠTĚPÁN JURAIDA a DANIEL MÜNICH:

(publications put to tertiles based on journal IF,
years 2006-2010,
but includes shares on publication from RIV etc.)

CA-Anorganická chemie (45)

Inorganic chemistry

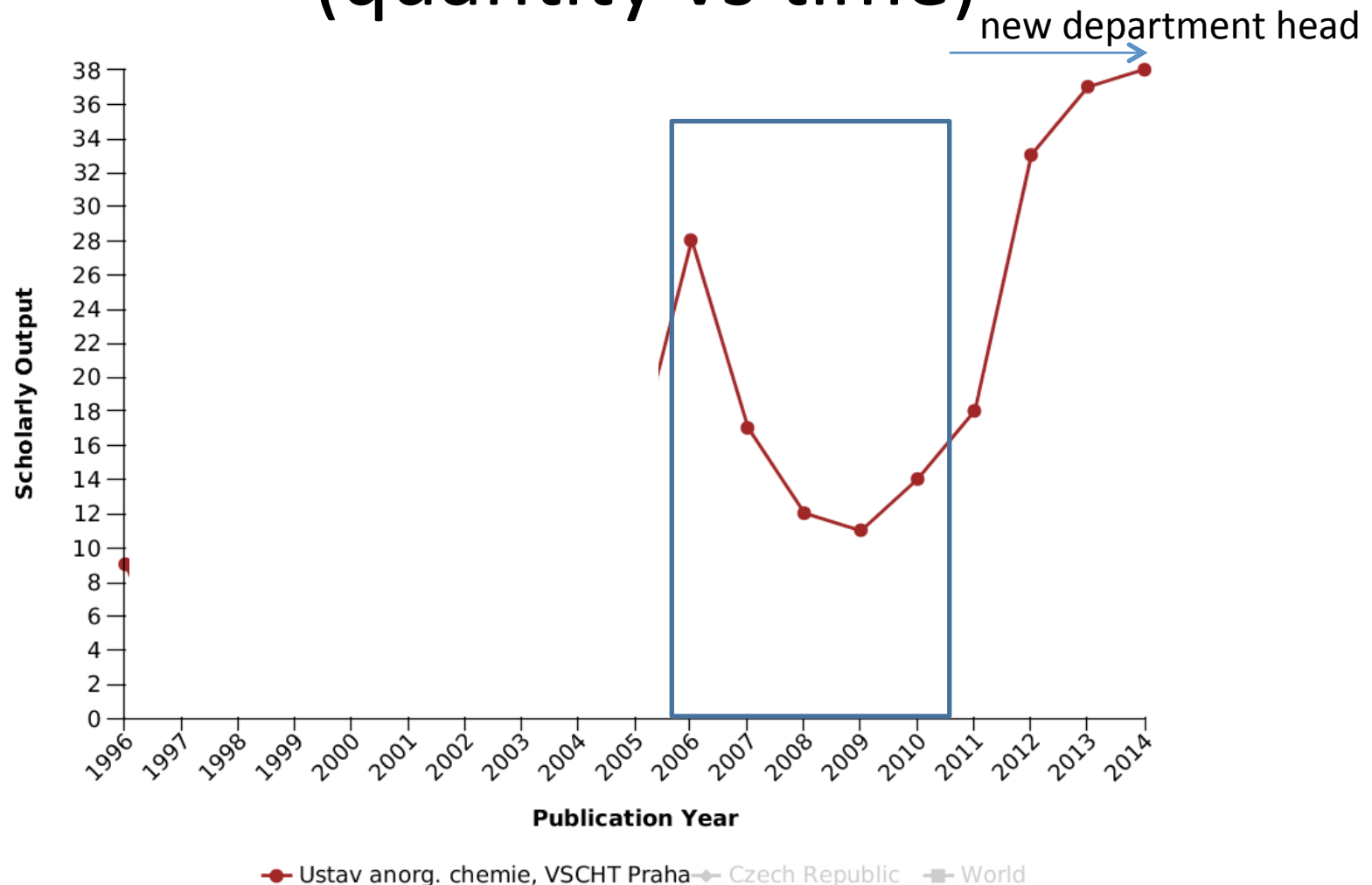
PRACOVISŤE	H	S	D	Celkem	Pořadí	Alt.pořadí	Oborů
AVCR-Ustav anorganické chemie ..	112	59	27	198	1	1	5
UPa-Fakulta chemicko-technolog..	86	77	26	189	2	2	17
UK-Prirodovědecká fakulta	78	44	15	137	3	3	34
AVCR-Ustav makromolekulární ch..	41	30	7	78	4	4	11
UP-Prirodovědecká fakulta	27	24	19	70	5	5	20
MU-Prirodovědecká fakulta	25	30	10	65	6	6	37
VŠCHT-Fakulta chemické technol..	19	20	6	45	7	7	18
UK-Farmaceutická fakulta v Hra..	5	3	2	10	8	8	7
MU-Fakulta informatiky	3	1	0	4	9	9	5
AVCR-Ustav experimentální bota..	2	2	0	4	10	10	12



Is Department of Inorganic Chemistry (under Faculty of Chemical Technology)
of UCT Prague really so bad?

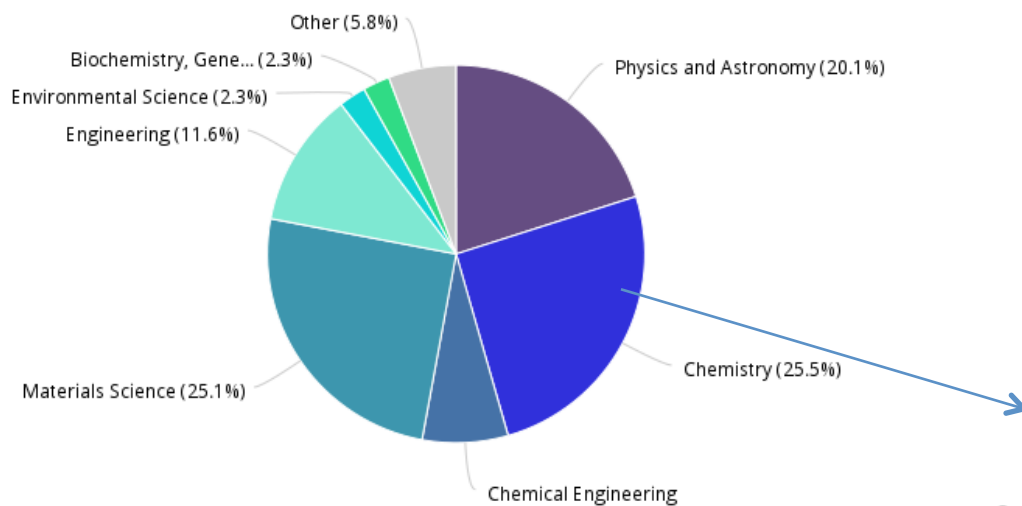
Quick check using bibliometric analysis tool (SciVal)...

How much they publish (article count) (quantity vs time)

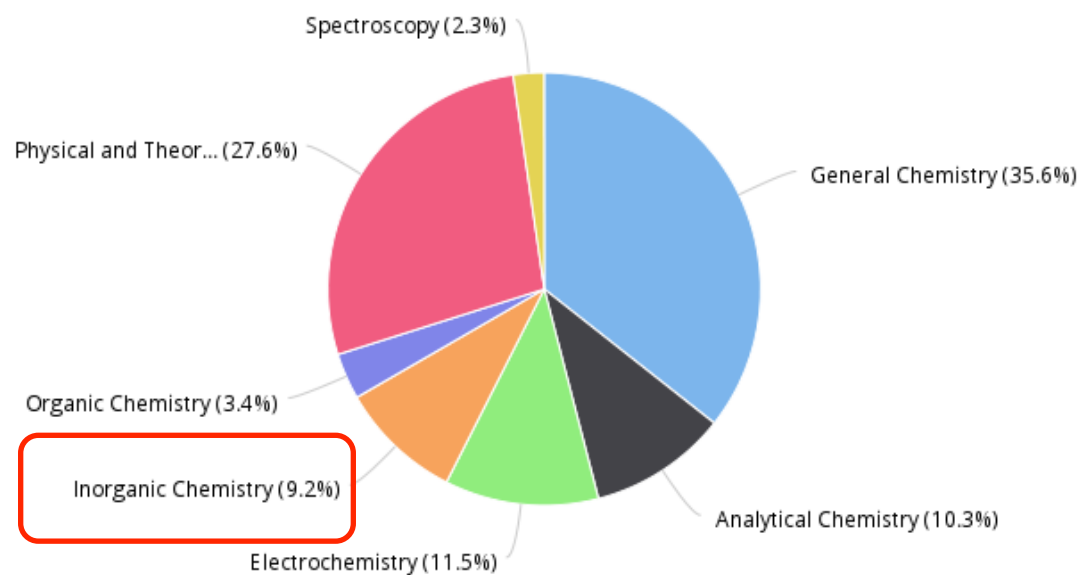


And where?

Main fields



Chemistry subfields



Overview for 2011-2013, field Chemistry

Overall research performance

[Export ▼](#)[Shortcuts ▼](#)

Publications

66 ▲

Researchers

8

Citations

402

Field-Weighted Citation Impact

2.39

[View list of publications](#)

Performance by Journal Category

[Export ▼](#)

Pie chart



Bar chart

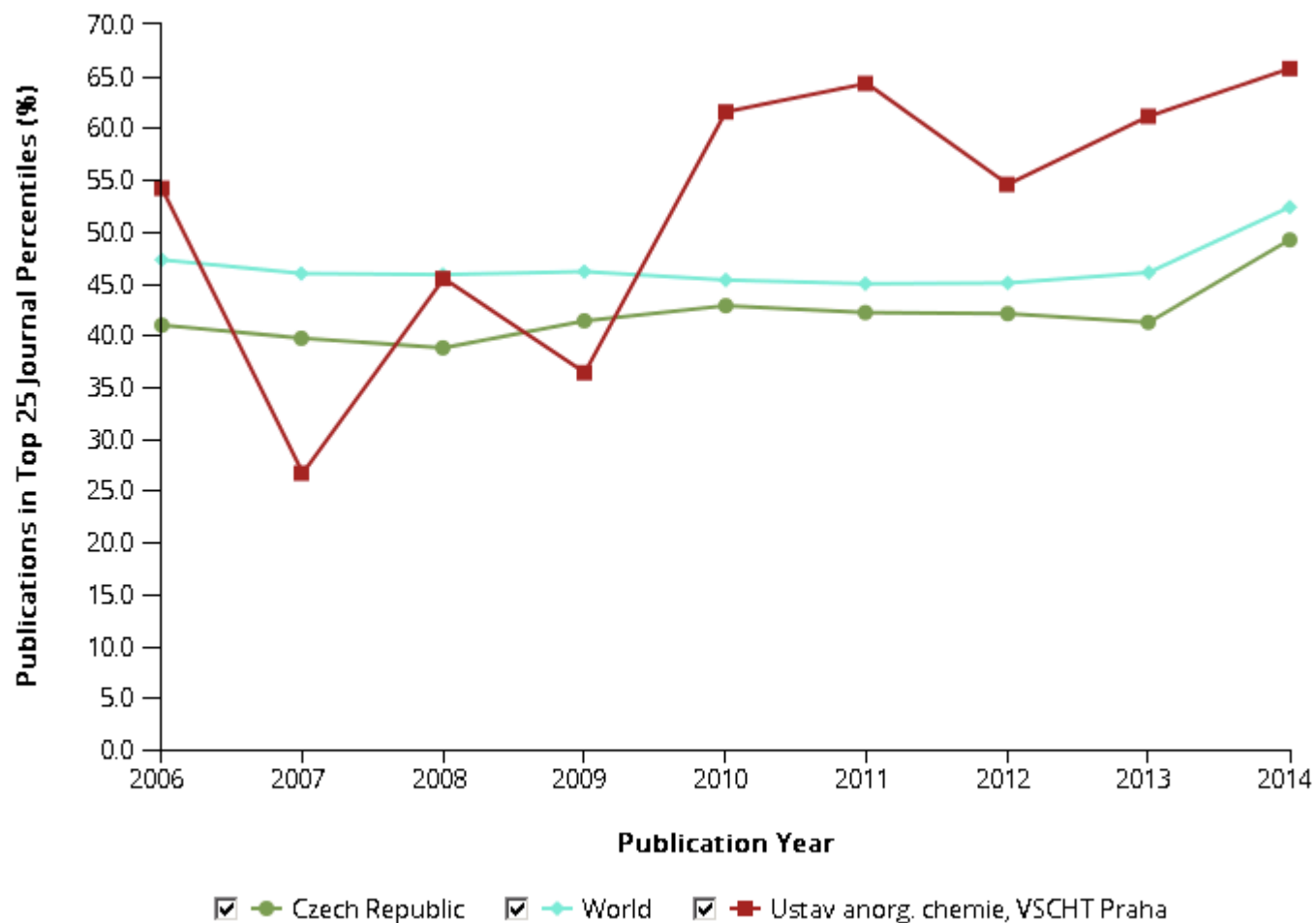


Table

Journal Category	Publications	Citations	Researchers	Field-Weighted Citation Impact
Ustav anorg. chemie, VSCHT Praha	126 ▲	633	10	2.07
Chemistry	66 ▲	402	8	2.39
General Chemistry	31 ▲	254	6	3.15
Physical and Theoretical Chemistry	24 ▲	74	6	1.71
Electrochemistry	10	71	3	1.88
Analytical Chemistry	9 ▲	28	2	1.03
Inorganic Chemistry	8 ▼	25	4	0.71
Organic Chemistry	3 ▲	10	2	0.93
Spectroscopy	2	15	1	1.15

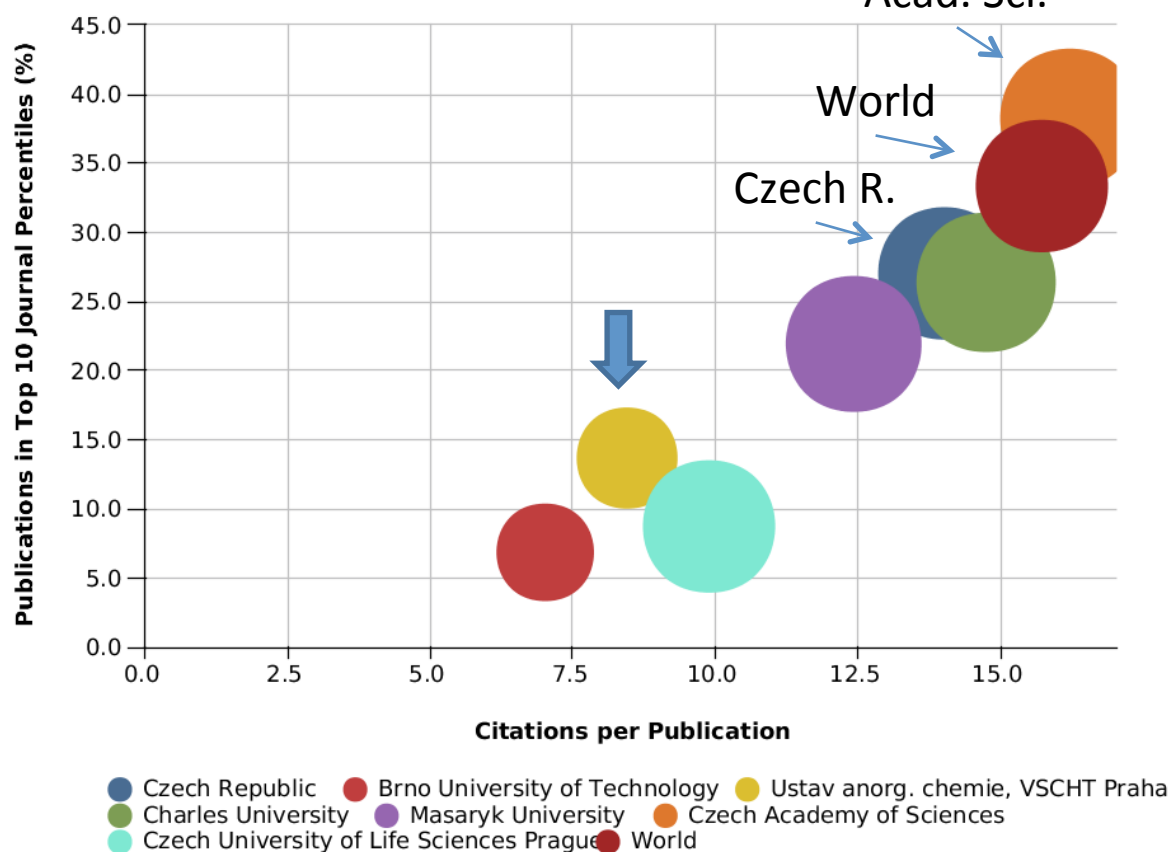
FWCI – field-weighted citation impact, value > 1: better than world average (normalized for given field, year, ...)

Comparison with world and Czech Rep. : fraction of articles in top 25 % journals (by SJR)

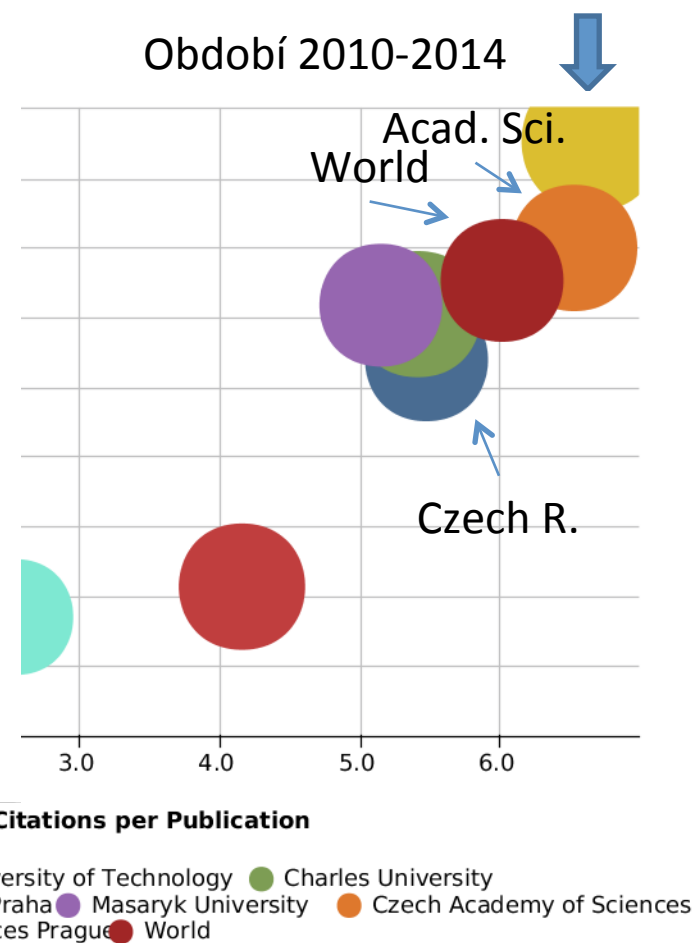


Development! – field Chemistry, institutions in Czech Rep. fraction of articles in top 10 % journals (SJR) vs citations per publication vs fraction in 25 % top articles (bubble)

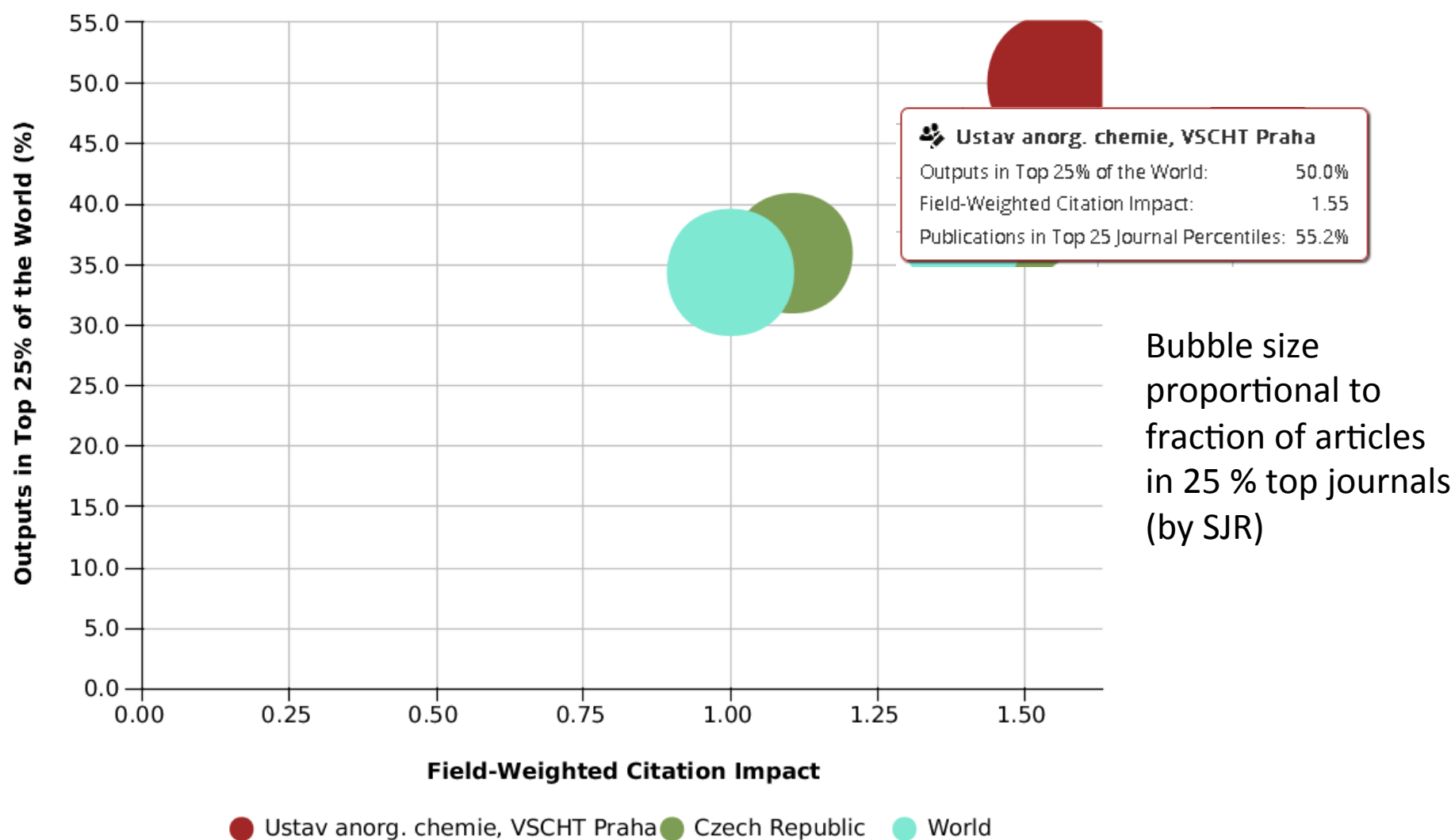
Období 2006-2010



Období 2010-2014



Comparing quality:



List of journals

Year range: 2006 to 2010

View the Scholarly Output of the selected entities, by journal:

Journal	SJR	 Ustav a
Chemistry of Materials	3.915	
Physical Review B - Condensed Matter and Materials Physics	2.143	
Langmuir	2.005	
Physical Chemistry Chemical Physics	1.811	
IEEE Journal of Quantum Electronics	1.594	
Electrochimica Acta	1.529	
Journal of Chemical Thermodynamics	1.292	
Journal of Alloys and Compounds	1.181	
Calphad: Computer Coupling of Phase Diagrams and Thermochemistry	1.063	
Journal of Crystal Growth	0.962	

Check: journal titles

List of journals

Year range: 2010 to 2013

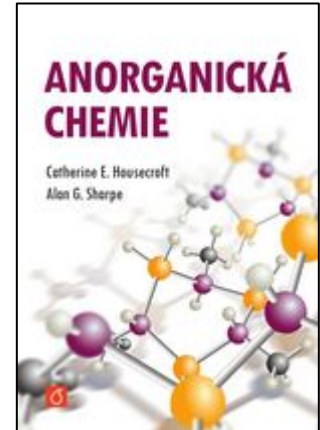
View the Scholarly Output of the selected entities, by jour

Export ▼

Shortcuts ▼

Journal	SJR ▼	Ustav anorg. chemie, VSCHT Praha
Proceedings of the National Academy of Sciences of the United States	7.048	1
ACS Nano	7.016	3
Angewandte Chemie - International Edition	6.016	1
Chemistry - A European Journal	2.744	6
Scripta Materialia	2.706	1
Nanoscale	2.704	4
Journal of Physical Chemistry C	2.265	1
Langmuir	2.005	1
Electrochemistry Communications	1.937	2
Inorganic Chemistry	1.926	1

Field definition



- What is „inorganic chemistry“? (informal discussion at the launch of the book „Anorganická chemie“ with org. chemists):
„Inorganic chemistry is everything what is not organic chemistry (inorg. chemists can take carbon dioxide too, about carbon monoxide we will negotiate later).“
- Could we use such exact scientific definitions for all normalizations, medians, rankings in categories etc. ...?

Example 3

LOOKING FOR EXCELLENT SCIENTIST

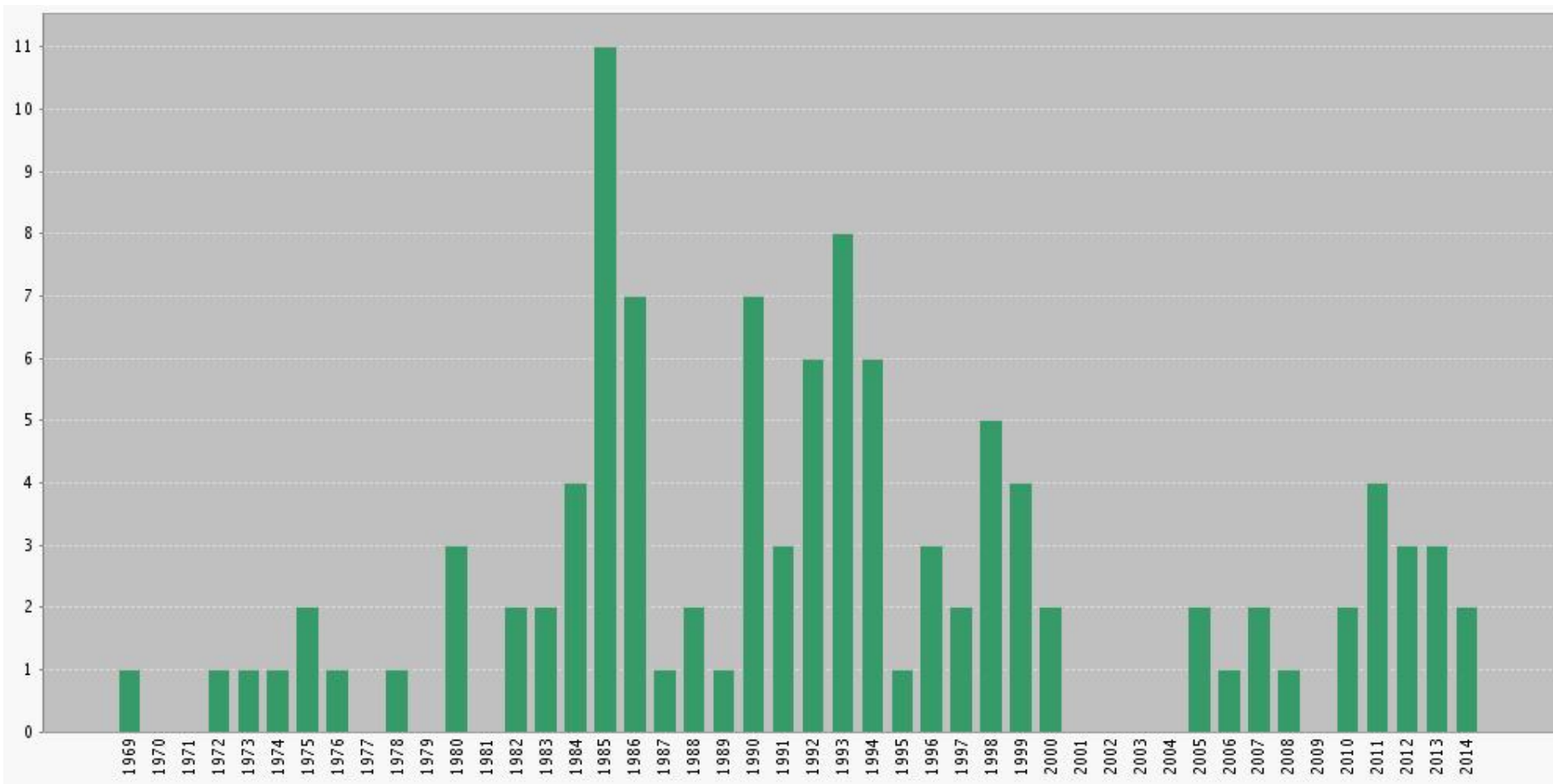
Comparing candidates – 1st round

	field (Research Area WOS)	<i>h</i> -index	article count	article count in recent 15 years
	Physics, Science technology other topics, Materials Science	29	148	140
	Materials Science, Metallurgy Metalurgical eng.	22	108	26
	Chemistry, Biochemistry molecular biology	54	754	307
	Physics	10	22	2

1st screening – removing low values

	field (Research Area WOS)	<i>h</i> -index	article count	article count in recent 15 years
	Physics, Science technology other topics, Materials Science	29	148	140
	Materials Science, Metallurgy Metalurgical eng.	22	108	26
	Chemistry, Biochemistry molecular biology	54	754	307
	Physics	10	22	2
very low article count, <i>h</i> -index too low				

Kandidát č. 2



Zdroj dat: Web of Science, Thomson-Reuters

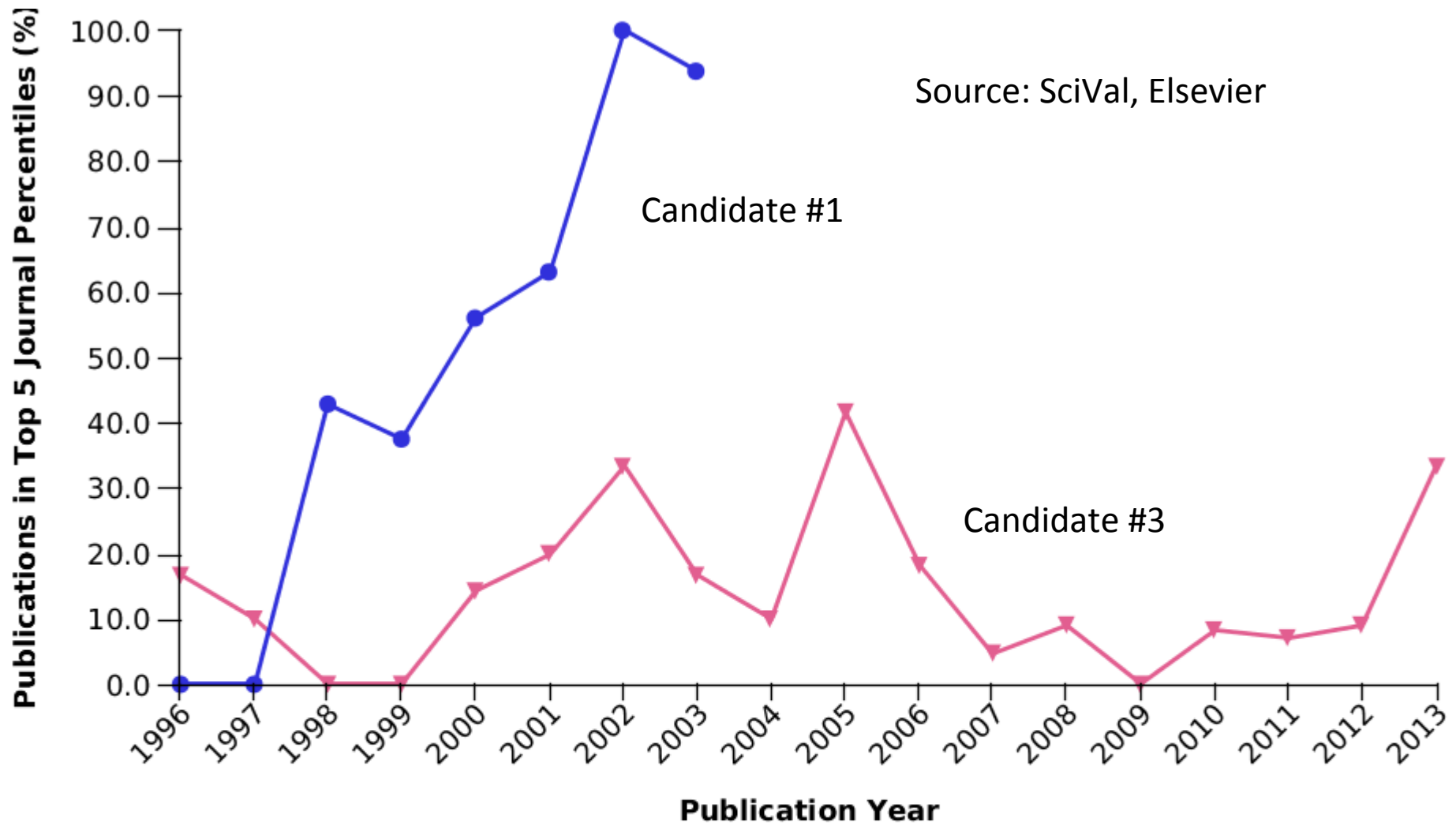
Další filtr

field (Research Area WOS)	<i>h</i> -index	article count	citations / citations without self-cites
Physics, Science technology other topics, Materials sci.	29	148	6544/6466
Materials Science, Metallurgy Metalurgical eng.	22	108	3492/3181
Chemistry, Biochemistry molecular biology	54	754	15451/10984

h-index nothing extra, rel. low citation count



Comparing finalists – articles in top 5 % best journals (SJR)



List of journals



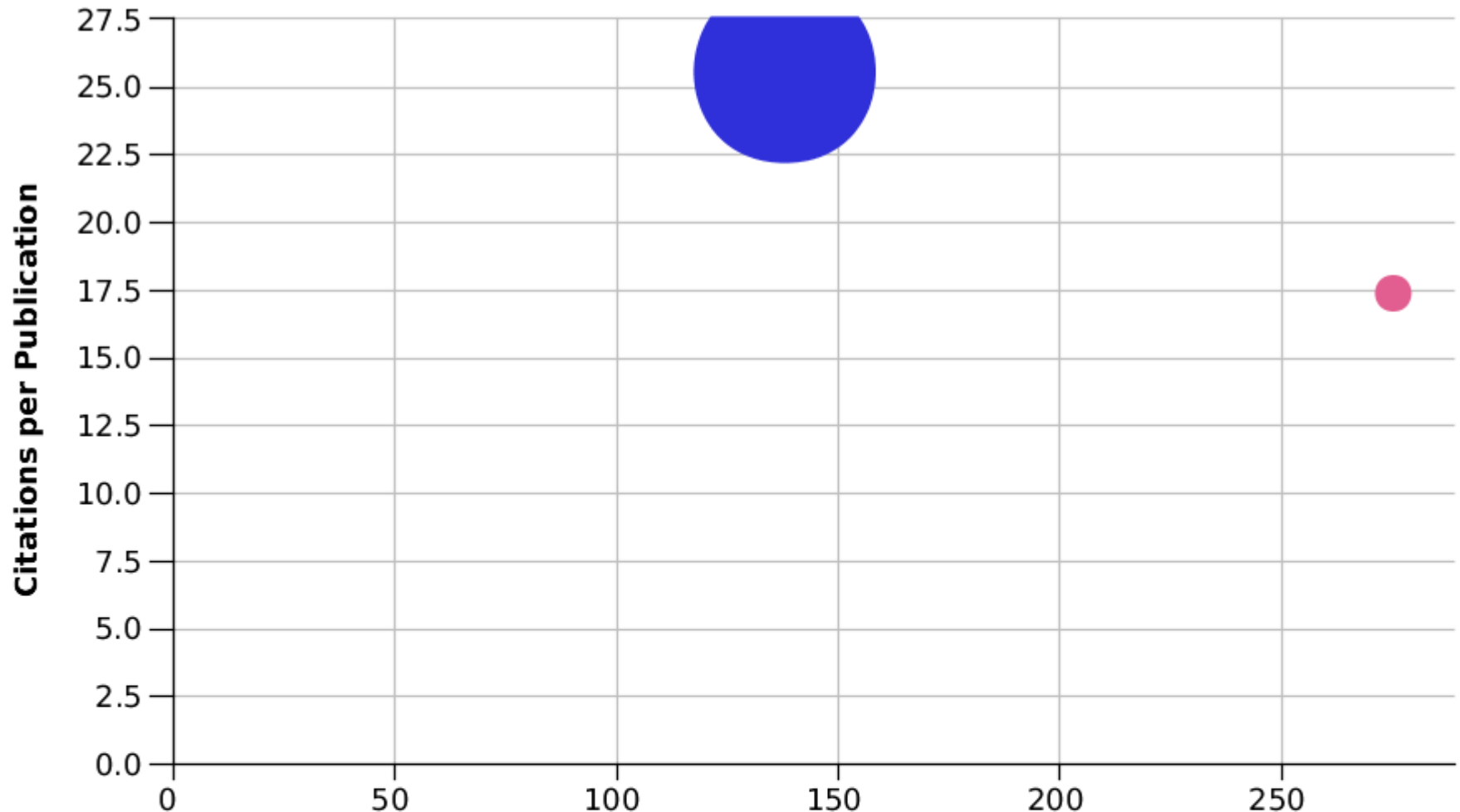
Year range: 1996 to 2013

View the Scholarly Output of the

Journal	SJR	Candidate #3	Candidate #1
Nature	21.323	1	17
Science	12.465	0	21
Advanced Materials	7.857	0	3
Nature Reviews Drug Discovery	7.139	1	0
Angewandte Chemie - International Edition	6.016	1	1
PLoS Pathogens	5.017	1	0
Physical Review Letters	4.099	0	3
Journal of Virology	3.492	1	0
Journal of Biological Chemistry	3.391	1	0
Journal of Molecular Biology	3.158	1	0
Chemistry and Biology	3.054	1	0
Chemical Communications	2.918	1	0
Chemistry - A European Journal	2.744	2	0
Journal of Controlled Release	2.494	1	0
Molecular Pharmacology	2.491	1	0
Journal of Antimicrobial Chemotherapy	2.420	1	0

Source: SciVal, Elsevier

Citations per publication vs article count (bubble size – article fraction in top 5 % journals) (1996-2013)



Scholarly Output

Source: SciVal, Elsevier

Porovnání lidí (WOS)?



obor (Res. Area WOS)	<i>h</i> -index	počet publ.	počet citací celkem / bez autocitací	průměrný počet citací na publ.	
Physics, Science technology other topics, Materials sci.	29	148	6544/6466	23.59	
Materials Science, Metallurgy Metalurgical eng.	22	108	3492/3181	60.59	
Chemistry, Biochemistry molecular biology	54	754	15451/10984	20.49	
Physics	10	22	4588/4572	208.55	

Porovnání lidí (WOS)?



obor (Res. Area WOS)	<i>h</i> -index	počet publ.	počet citací celkem / bez autocitací	průměrný počet citací na publ.	Odhalení
Physics, Science technology other topics, Materials sci.	29	148	6544/6466	23.59	
Materials Science, Metallurgy Metalurgical eng.	22	108	3492/3181	60.59	Prof. Dan Shechtman , Nobelova cena za chemii 2011
Chemistry, Biochemistry molecular biology	54	754	15451/10984	20.49	
Physics	10	22	4588/4572	208.55	Prof. P.W.

Na 4. místě...



Potato. [Photography]. Retrieved from Encyclopædia Britannica ImageQuest.
http://quest.eb.com/#/search/156_2410939/1/156_2410939/cite

Na 3. místě...



London 2012 Medals Being Produced by Royal Mint. [Photography]. Retrieved from Encyclopædia Britannica ImageQuest.
http://quest.eb.com/#/search/115_3907467/1/115_3907467/cite

Na 2. místě...



London 2012 Medals Being Produced by Royal Mint. [Photography]. Retrieved from Encyclopædia Britannica ImageQuest.
http://quest.eb.com/#/search/115_3907467/1/115_3907467/cite

Vítězem se stává...



London 2012 Medals Being Produced by Royal Mint. [Photography]. Retrieved from Encyclopædia Britannica ImageQuest.
http://quest.eb.com/#/search/115_3907467/1/115_3907467/cite

And the result is...



Conclusions

- Bibliometric tools
 - benefits:
 - possibility to identify quality and changes in institution's research
 - deeper insight into institution's research
- Be aware:
 - mechanical application can yield misleading results
 - evaluator should (in ideal world) have detailed knowledge of institution **and** science field