The heterogeneity of social media metrics and its effects on statistics

Rodrigo Costas, Stefanie Haustein & Vincent Larivière

Université de Montréal

Canada Research Chair
on the Transformations of Scholarly Communication
École de bibliothéconomie et des sciences de l'information

Introduction

- factors affecting citation rates studied since creation of SCI:
  - characteristics of papers
  - social factors
- little knowledge about effects of document characteristics on papers' social media metrics

- research aims of this study:
  - assessing which and how document characteristics influence social media scores
  - ➢ better understanding of nature of social media indicators
  - ➢ paving the way for frameworks or theories supporting (or not) use, interpretation and validity of such indicators

Data & Methods

- Web of Science (SCIE, SSCI, AHCI)
  - five broad fields of Leiden ranking
- 1.3 million papers published in 2012, citations until the end of 2013
  - compromise between citation window, recent uptake of social media activity and Altmetric.com coverage
- bibliographic and citation characteristics
  - number of citations and self-citations
  - number of pages
  - number of authors, institutions and countries
  - number of references
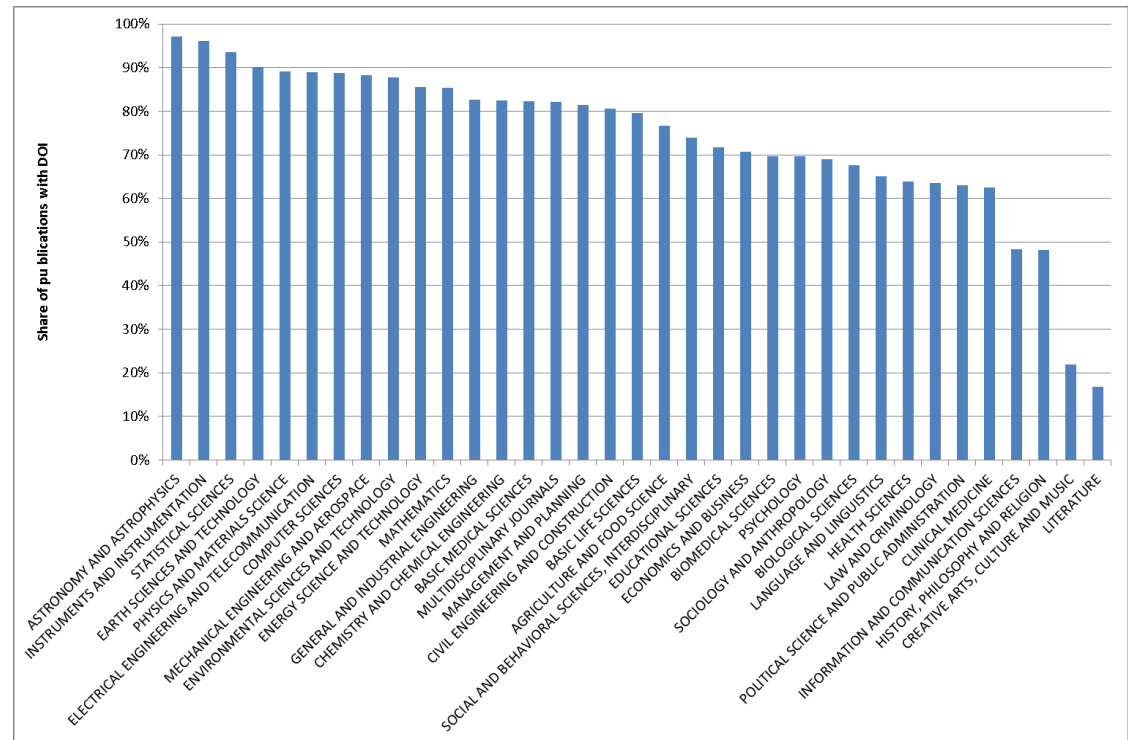 - number of characters in title

Data & Methods

- Altmetric.com as source for social media indicators
  - Twitter
  - Blogs
  - News outlets
  - Facebook
  - Google+
  - Mendeley not included due to incompleteness on Altmetric.com
- social media mentions up to 18/10/2013
  - window of more than 9 months enough to cover majority of social media activity
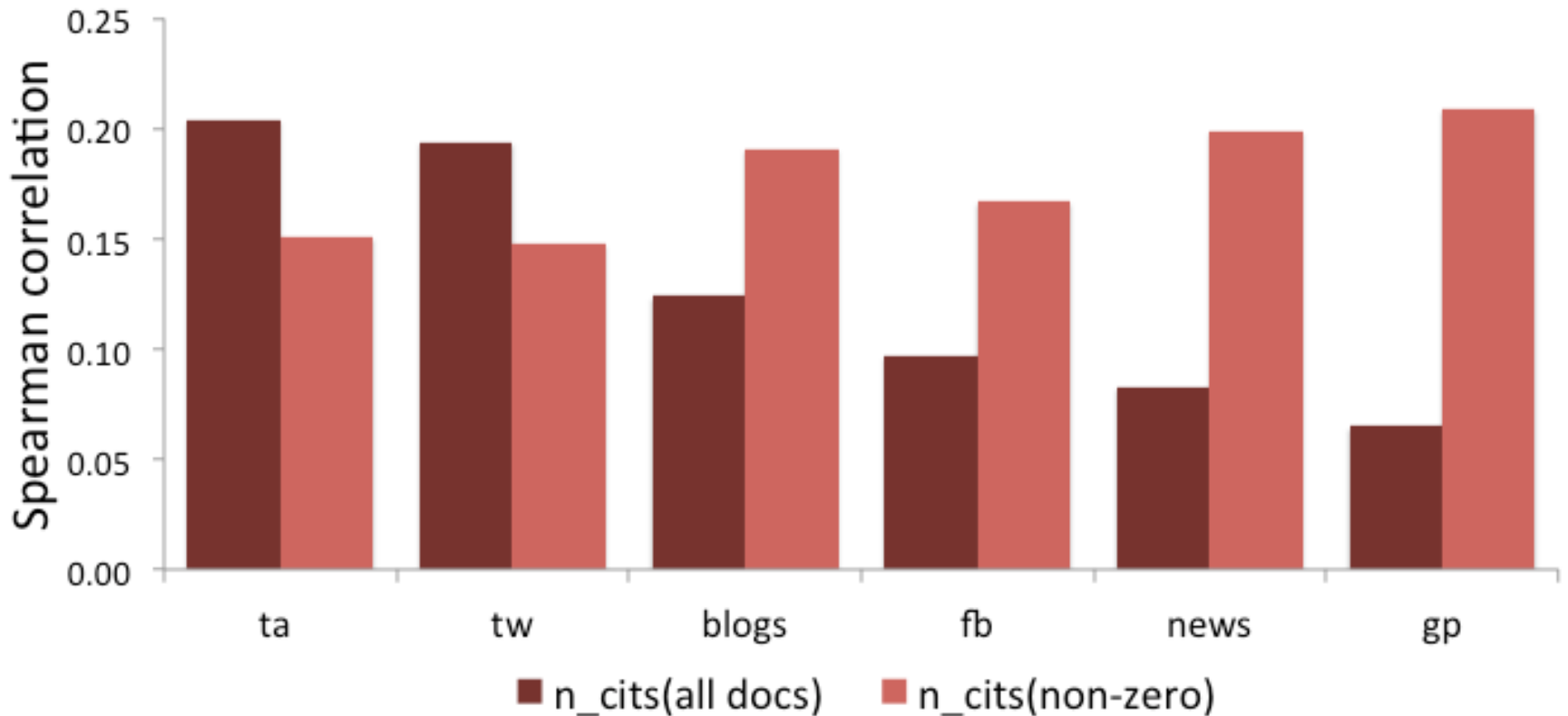
Data & Methods

Limitations

- data collection of social media mentions not perfect
  - missing Mendeley
- match between WoS bibliographic data and Altmetric.com based on DOIs only

Results: correlation between variables

Indicator	n cits	nselfcits	pg	n authors	n refs	n countries	n institutes	ti length	fb	blogs	tw	gp	news
n_cits	1.000	**0.664**	0.194	**0.281**	**0.416**	**0.155**	0.188	0.142	0.097	0.124	**0.194**	0.065	0.083
n_self_cits		1.000	**0.159**	**0.225**	**0.286**	**0.166**	0.168	0.117	0.047	0.083	0.092	0.040	0.061
pg			1.000	-0.006	0.622	0.131	0.116	0.079	0.011	-0.002	0.032	0.002	-0.010
n_authors				1.000	0.155	**0.252**	**0.494**	**0.323**	0.049	0.030	0.089	0.010	0.033
n_refs					1.000	**0.146**	**0.168**	**0.165**	0.063	0.051	0.140	0.028	0.029
n_countries						1.000	0.560	0.038	0.039	0.043	0.061	0.023	0.035
n_institutes							1.000	0.135	0.062	0.045	0.102	0.026	0.037
ti_length								1.000	-0.011	-0.032	-0.004	-0.027	-0.020
fb									1.000	0.180	**0.320**	0.144	0.161
blogs										1.000	**0.194**	0.196	0.279
tw											1.000	0.142	0.137
gp												1.000	0.179
news													1.000

Results: correlation between variables

Results: statistics per document type

		all	Article	Biographical-Item	Book Review	Correction	Editorial Material	Letter	Meeting Abstract	News Item	Review
N		1,338,885	1,132,428	2,302	21,710	9,817	60,533	29,410	13,071	4,880	64,734
Citations	A	66.86%	71.6%	6.52%	1.03%	11.94%	36.28%	34.19%	3.08%	28.11%	85.08%
	B	3.17	3.24	0.10	0.01	0.22	1.05	0.80	0.04	0.72	7.48
	C	4.74	4.55	1.48	1.10	1.83	2.90	2.33	1.17	2.55	8.80
Twitter	A	21.50%	20.99%	13.03%	5.16%	10.33%	27.34%	18.16%	2.09%	42.48%	36.17%
	B	0.78	0.74	0.34	0.09	0.18	1.59	0.45	0.04	3.02	1.38
	C	3.65	3.50	2.62	1.83	1.76	5.81	2.46	1.93	7.10	3.81
Blogs	A	1.86%	1.88%	0.70%	0.14%	2.19%	2.34%	0.82%	0.03%	2.23%	2.51%
	B	0.03	0.03	0.01	0.00	0.02	0.03	0.01	0.00	0.03	0.03
	C	1.51	1.53	1.00	1.00	1.09	1.48	1.49	1.00	1.19	1.31
News	A	0.69%	0.72%	0.00%	0.00%	0.07%	0.82%	0.19%	0.01%	0.37%	0.61%
	B	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
	C	1.54	1.57	n/a	n/a	1.14	1.32	2.05	1.00	1.44	1.25
Facebook	A	4.70%	4.49%	2.74%	1.24%	1.29%	7.41%	3.36%	0.22%	7.36%	8.80%
	B	0.08	0.08	0.03	0.01	0.01	0.15	0.05	0.00	0.12	0.16
	C	1.78	1.77	1.25	1.07	1.03	1.98	1.35	1.03	1.57	1.84
Google+	A	0.75%	0.72%	0.26%	0.04%	0.19%	1.25%	0.44%	0.02%	3.85%	1.34%
	B	0.01	0.01	0.00	0.00	0.00	0.02	0.01	0.00	0.04	0.02
	C	1.66	1.70	1.17	1.00	1.16	1.64	1.49	1.00	1.11	1.39

A: Percentage of papers with at least one citation or social media mention (coverage)
B: Average number of events per paper for all papers (density)
C: Average number of events per paper for papers with at least one event (intensity)

Results: statistics per document type

		all	Article	Biographical-Item	Book Review	Correction	Editorial Material	Letter	Meeting Abstract	News Item	Review
N		1,338,885	1,132,428	2,302	21,710	9,817	60,533	29,410	13,071	4,880	64,734
Citations	A	66.86%	71.6%	6.52%	1.03%	11.94%	36.28%	34.19%	3.08%	28.11%	85.08%
	B	3.17	3.24	0.10	0.01	0.22	1.05	0.80	0.04	0.72	7.48
	C	4.74	4.55	1.48	1.10	1.83	2.90	2.33	1.17	2.55	8.80
Twitter	A	21.50%	20.99%	13.03%	5.16%	10.33%	27.34%	18.16%	2.09%	42.48%	36.17%
	B	0.78	0.74	0.34	0.09	0.18	1.59	0.45	0.04	3.02	1.38
	C	3.65	3.50	2.62	1.83	1.76	5.81	2.46	1.93	7.10	3.81
Blogs	A	1.86%	1.88%	0.70%	0.14%	2.19%	2.34%	0.82%	0.03%	2.23%	2.51%
	B	0.03	0.03	0.01	0.00	0.02	0.03	0.01	0.00	0.03	0.03
	C	1.51	1.53	1.00	1.00	1.09	1.48	1.49	1.00	1.19	1.31
News	A	0.69%	0.72%	0.00%	0.00%	0.07%	0.82%	0.19%	0.01%	0.37%	0.61%
	B	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
	C	1.54	1.57	n/a	n/a	1.14	1.32	2.05	1.00	1.44	1.25
Facebook	A	4.70%	4.49%	2.74%	1.24%	1.29%	7.41%	3.36%	0.22%	7.36%	8.80%
	B	0.08	0.08	0.03	0.01	0.01	0.15	0.05	0.00	0.12	0.16
	C	1.78	1.77	1.25	1.07	1.03	1.98	1.35	1.03	1.57	1.84
Google+	A	0.75%	0.72%	0.26%	0.04%	0.19%	1.25%	0.44%	0.02%	3.85%	1.34%
	B	0.01	0.01	0.00	0.00	0.00	0.02	0.01	0.00	0.04	0.02
	C	1.66	1.70	1.17	1.00	1.16	1.64	1.49	1.00	1.11	1.39

A: Percentage of papers with at least one citation or social media mention (coverage)

Results: statistics per document type

		all	Article	Biographical-Item	Book Review	Correction	Editorial Material	Letter	Meeting Abstract	News Item	Review
N		*1,338,885*	*1,132,428*	*2,302*	*21,710*	*9,817*	*60,533*	*29,410*	*13,071*	*4,880*	*64,734*
Citations	A	66.86%	71.6%	6.52%	1.03%	11.94%	36.28%	34.19%	3.08%	28.11%	85.08%
	B	3.17	3.24	0.10	0.01	0.22	1.05	0.80	0.04	0.72	7.48
	C	4.74	4.55	1.48	1.10	1.83	2.90	2.33	1.17	2.55	8.80
Twitter	A	21.50%	20.99%	13.03%	5.16%	10.33%	27.34%	18.16%	2.09%	42.48%	36.17%
	B	0.78	0.74	0.34	0.09	0.18	1.59	0.45	0.04	3.02	1.38
	C	3.65	3.50	2.62	1.83	1.76	5.81	2.46	1.93	7.10	3.81
Blogs	A	1.86%	1.88%	0.70%	0.14%	2.19%	2.34%	0.82%	0.03%	2.23%	2.51%
	B	0.03	0.03	0.01	0.00	0.02	0.03	0.01	0.00	0.03	0.03
	C	1.51	1.53	1.00	1.00	1.09	1.48	1.49	1.00	1.19	1.31
News	A	0.69%	0.72%	0.00%	0.00%	0.07%	0.82%	0.19%	0.01%	0.37%	0.61%
	B	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
	C	1.54	1.57	n/a	n/a	1.14	1.32	2.05	1.00	1.44	1.25
Facebook	A	4.70%	4.49%	2.74%	1.24%	1.29%	7.41%	3.36%	0.22%	7.36%	8.80%
	B	0.08	0.08	0.03	0.01	0.01	0.15	0.05	0.00	0.12	0.16
	C	1.78	1.77	1.25	1.07	1.03	1.98	1.35	1.03	1.57	1.84
Google+	A	0.75%	0.72%	0.26%	0.04%	0.19%	1.25%	0.44%	0.02%	3.85%	1.34%
	B	0.01	0.01	0.00	0.00	0.00	0.02	0.01	0.00	0.04	0.02
	C	1.66	1.70	1.17	1.00	1.16	1.64	1.49	1.00	1.11	1.39

B: Average number of events per paper for all papers (density)

Results: statistics per document type

		all	Article	Biographical-Item	Book Review	Correction	Editorial Material	Letter	Meeting Abstract	News Item	Review
N		1,338,885	1,132,428	2,302	21,710	9,817	60,533	29,410	13,071	4,880	64,734
Citations	A	66.86%	71.6%	6.52%	1.03%	11.94%	36.28%	34.19%	3.08%	28.11%	85.08%
	B	3.17	3.24	0.10	0.01	0.22	1.05	0.80	0.04	0.72	7.48
	C	4.74	4.55	1.48	1.10	1.83	2.90	2.33	1.17	2.55	8.80
Twitter	A	21.50%	20.99%	13.03%	5.16%	10.33%	27.34%	18.16%	2.09%	42.48%	36.17%
	B	0.78	0.74	0.34	0.09	0.18	1.59	0.45	0.04	3.02	1.38
	C	3.65	3.50	2.62	1.83	1.76	5.81	2.46	1.93	7.10	3.81
Blogs	A	1.86%	1.88%	0.70%	0.14%	2.19%	2.34%	0.82%	0.03%	2.23%	2.51%
	B	0.03	0.03	0.01	0.00	0.02	0.03	0.01	0.00	0.03	0.03
	C	1.51	1.53	1.00	1.00	1.09	1.48	1.49	1.00	1.19	1.31
News	A	0.69%	0.72%	0.00%	0.00%	0.07%	0.82%	0.19%	0.01%	0.37%	0.61%
	B	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
	C	1.54	1.57	n/a	n/a	1.14	1.32	2.05	1.00	1.44	1.25
Facebook	A	4.70%	4.49%	2.74%	1.24%	1.29%	7.41%	3.36%	0.22%	7.36%	8.80%
	B	0.08	0.08	0.03	0.01	0.01	0.15	0.05	0.00	0.12	0.16
	C	1.78	1.77	1.25	1.07	1.03	1.98	1.35	1.03	1.57	1.84
Google+	A	0.75%	0.72%	0.26%	0.04%	0.19%	1.25%	0.44%	0.02%	3.85%	1.34%
	B	0.01	0.01	0.00	0.00	0.00	0.02	0.01	0.00	0.04	0.02
	C	1.66	1.70	1.17	1.00	1.16	1.64	1.49	1.00	1.11	1.39

C: Average number of events per paper for papers with at least one event (intensity)

Results: mean event rate per field

Mean rate % w/ event	cits	self cits	pg	au	refs	co	inst	title	Fb	blogs	Twitter	g+	news
All fields	3.17	0.81	9.26	5.15	36.06	1.32	2.13	96.45	0.08	0.03	0.78	0.01	0.01
	66.8%								4.7%	1.9%	21.5%	0.8%	0.7%
Biomed & Health	3.57	0.75	7.73	5.55	35.96	1.31	2.32	100.91	0.14	0.03	1.28	0.02	0.01
	70.1%								7.5%	2.1%	31.7%	1.0%	0.7%
Life & Earth	3.44	0.96	9.88	4.85	43.70	1.40	2.25	105.13	0.11	0.05	1.03	0.02	0.02
	74.3%								5.7%	2.9%	21.6%	1.0%	1.3%
Math & Computer	1.70	0.56	12.74	3.18	27.39	1.31	1.81	81.37	0.03	0.01	0.36	0.01	0.01
	54.6%								1.5%	0.8%	7.5%	0.4%	0.3%
Natural Sci & Engn	3.83	1.13	8.78	6.33	35.39	1.35	2.11	97.22	0.03	0.03	0.34	0.01	0.01
	73.7%								2.3%	1.8%	12.9%	0.6%	0.8%
Social Sci & Humanities	1.43	0.35	12.62	2.47	37.56	1.21	1.69	82.18	0.10	0.05	1.33	0.02	0.01
	45.8%								4.8%	2.7%	26.0%	1.1%	0.9%

Results: number of authors
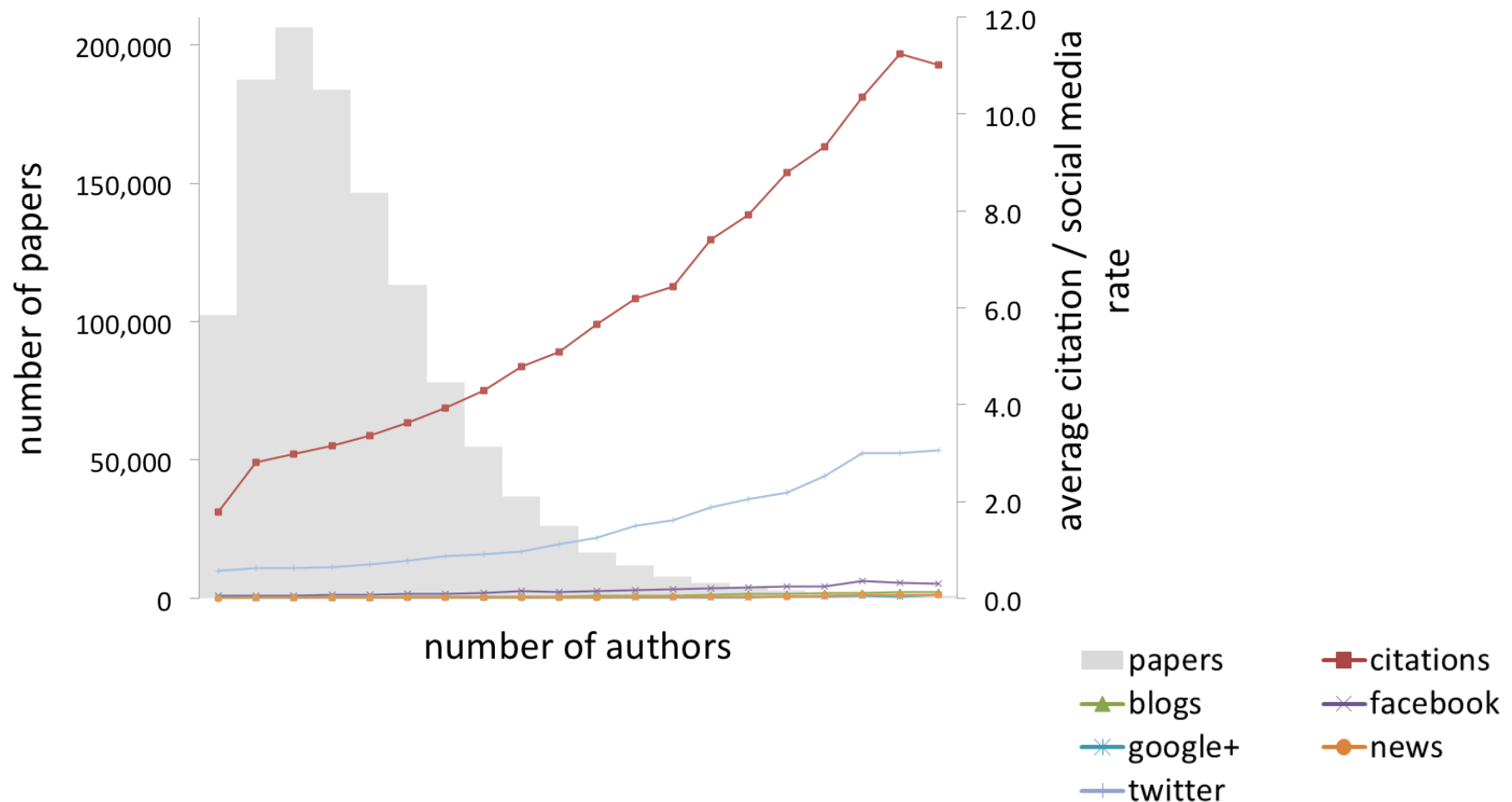
Results: number of countries
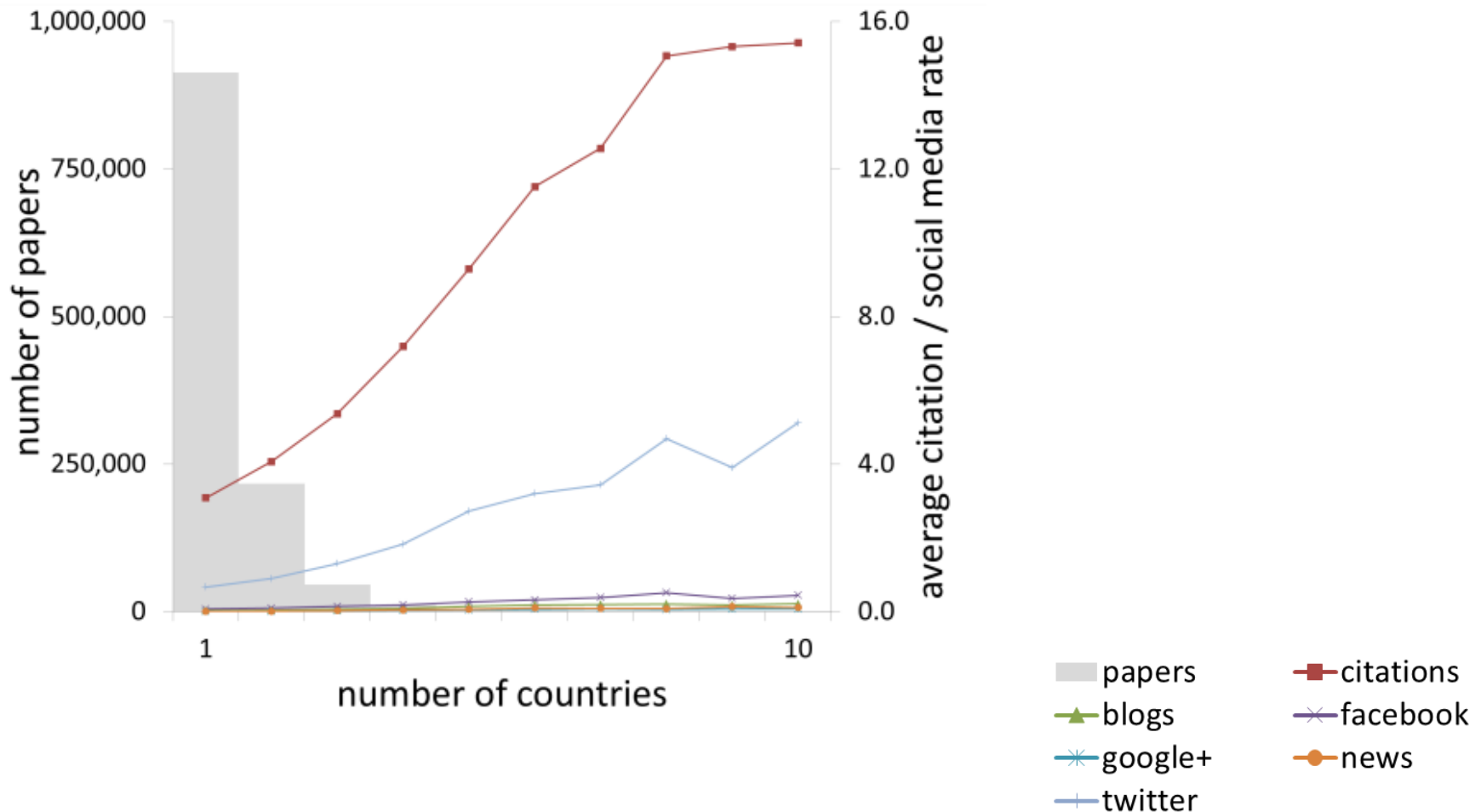
Results: number of pages
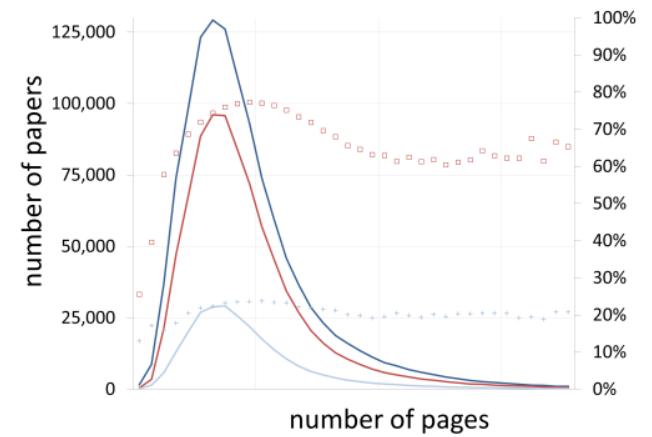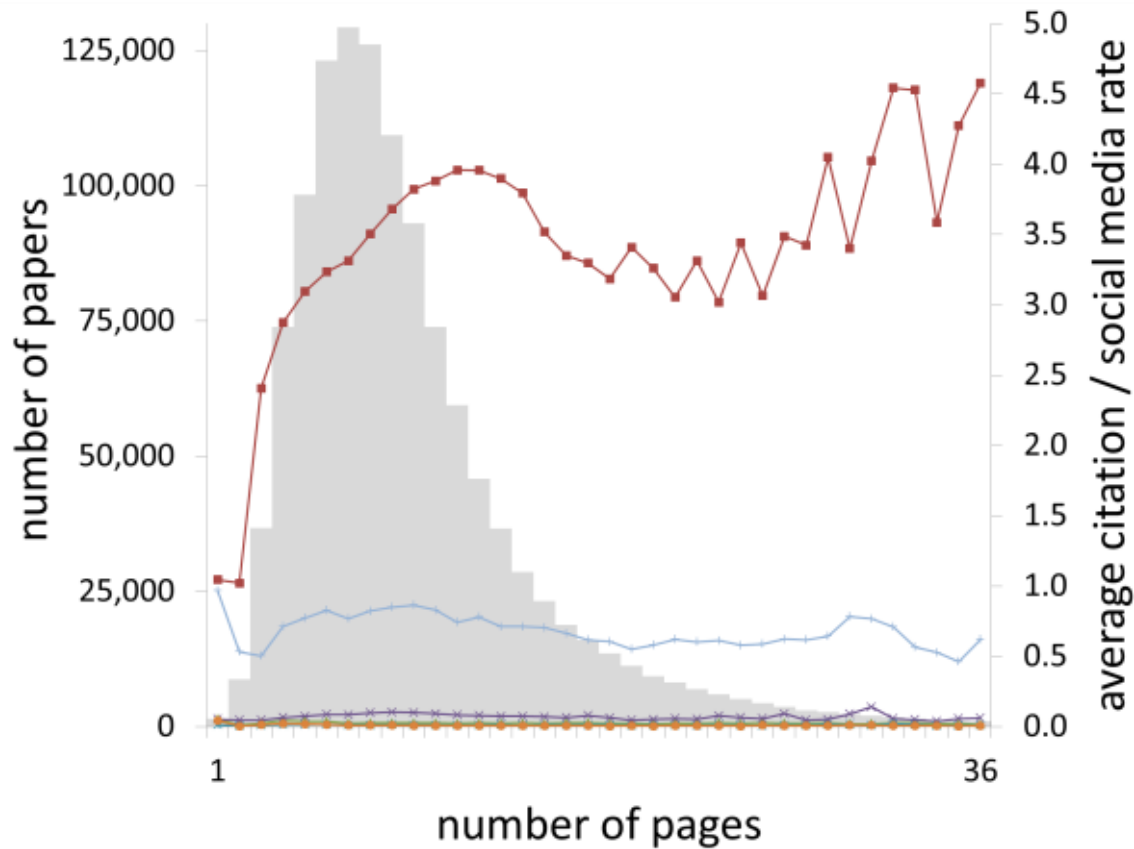
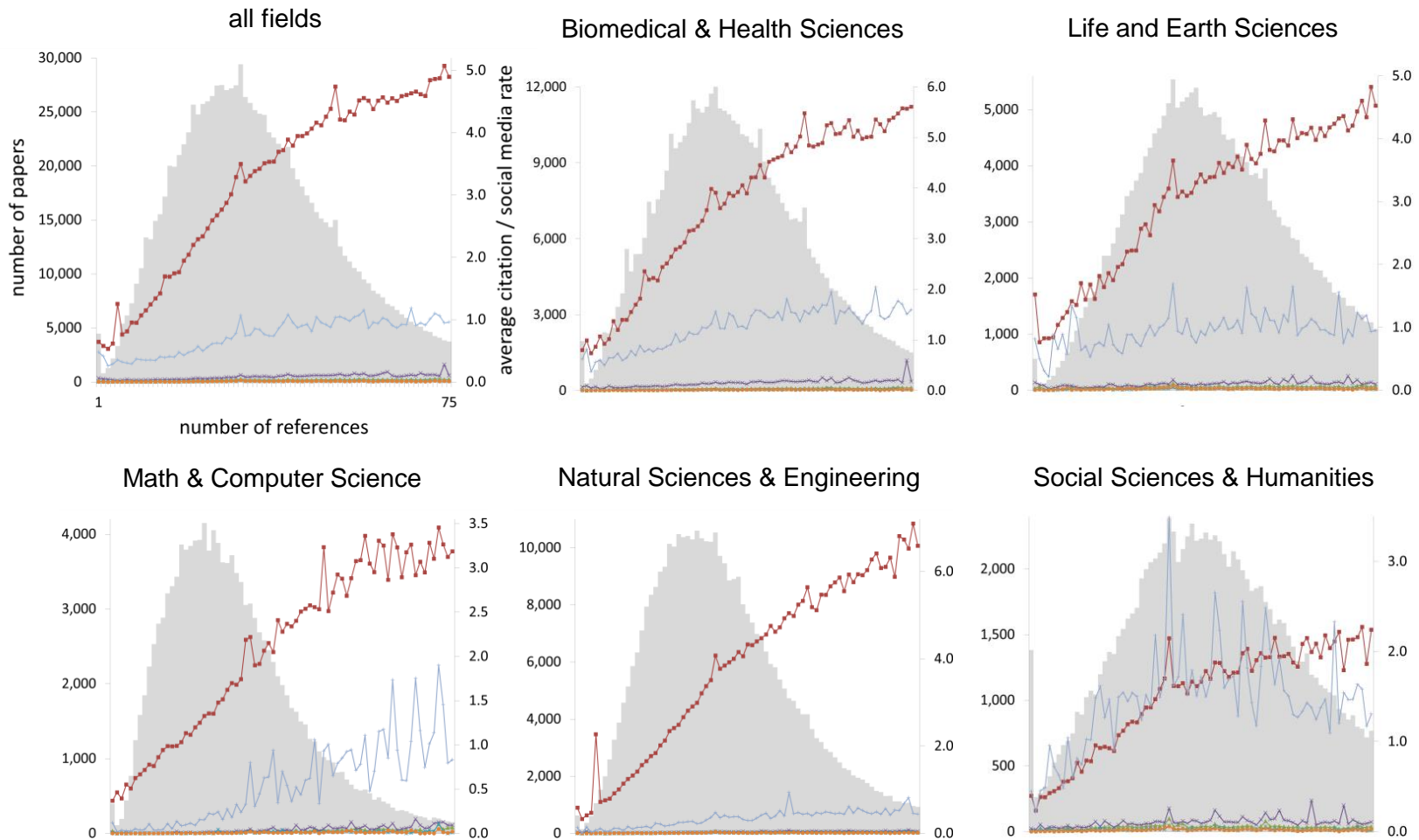# **Results**: number of pages
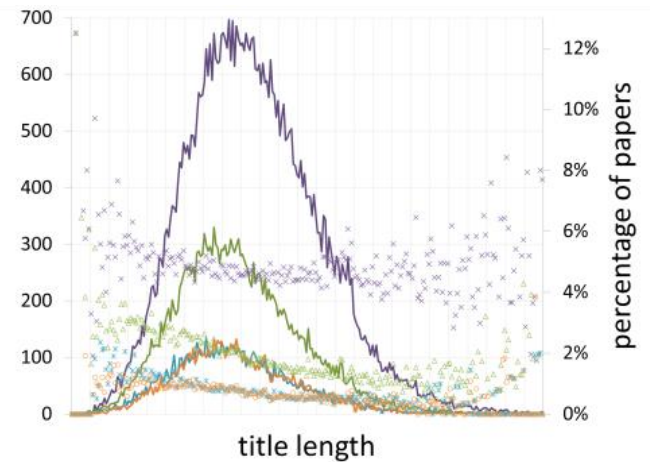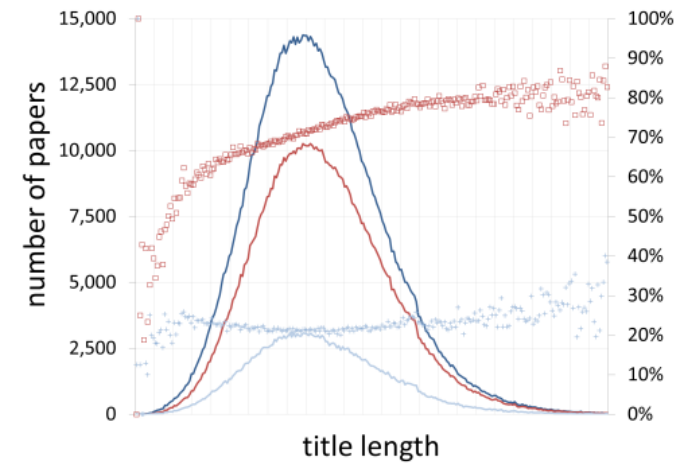
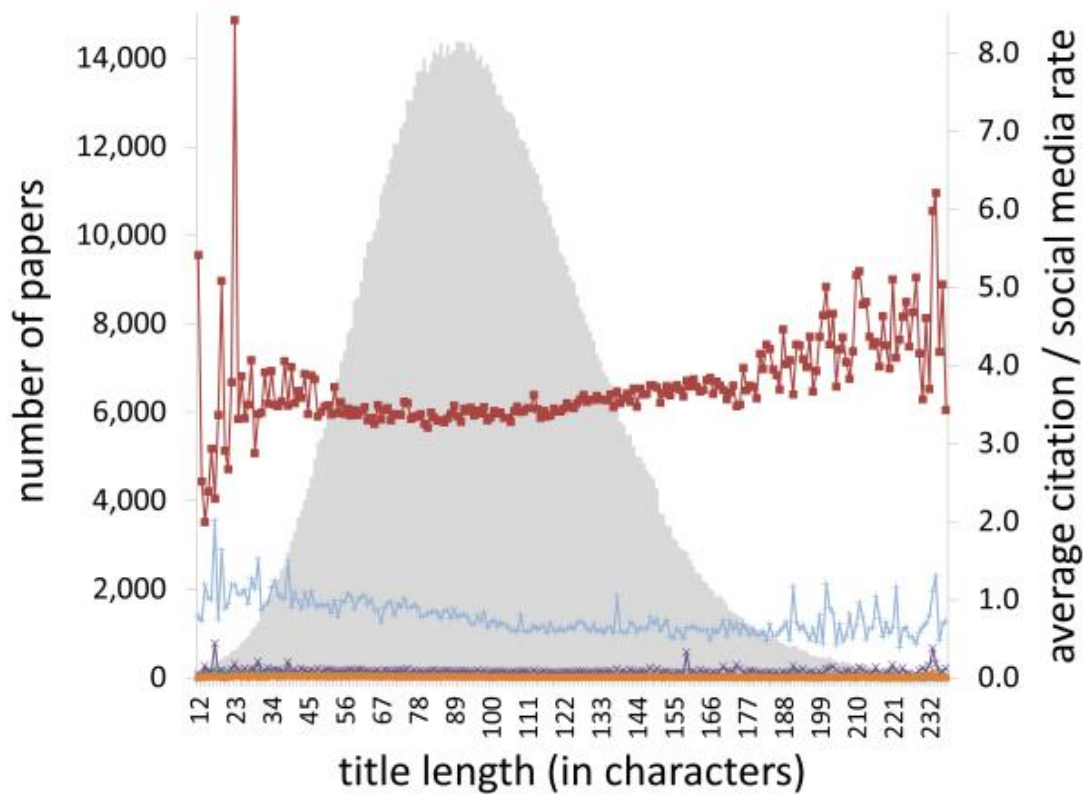Life and Earth Sciences

Social Sciences & Humanities
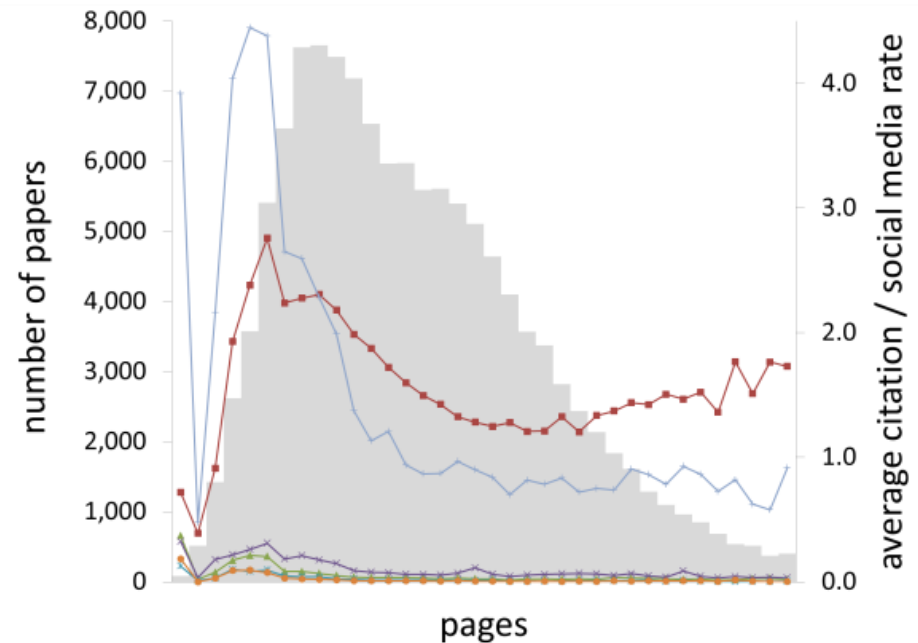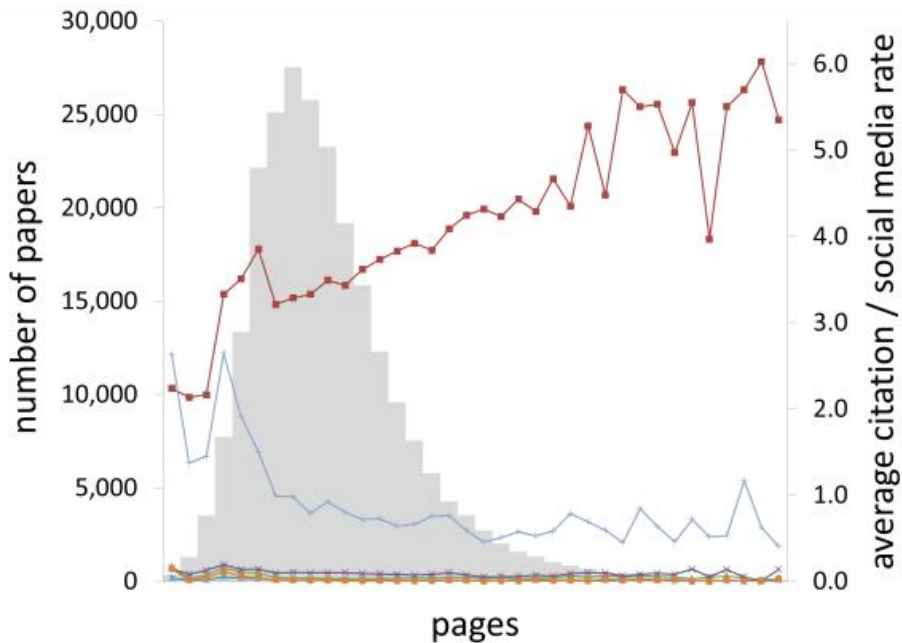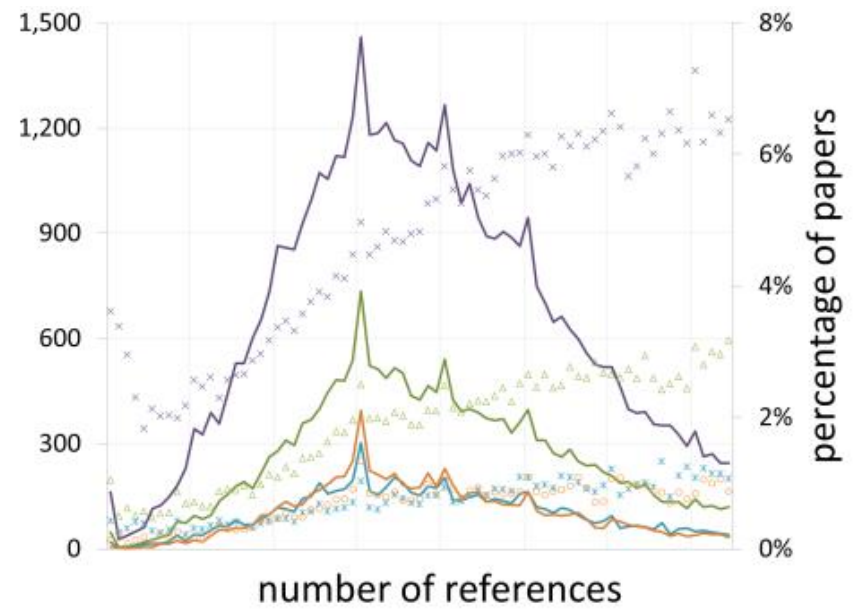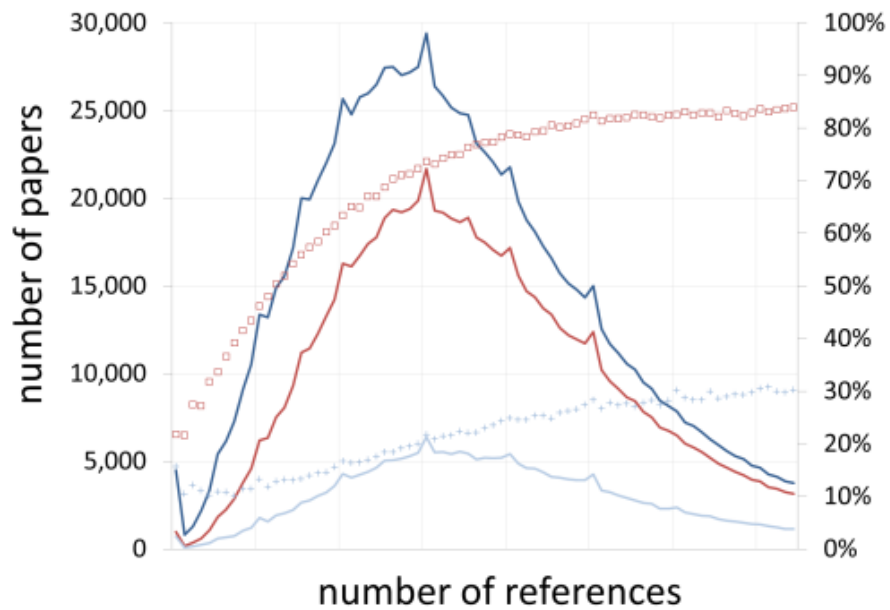
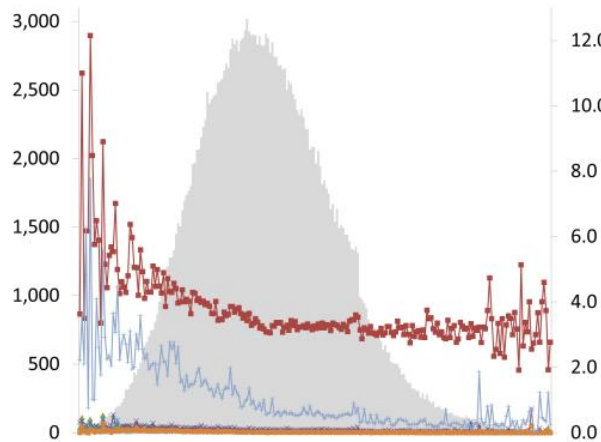# **Results**: number of references

Results: number of references

all fields

Biomedical & Health Sciences

Life and Earth Sciences

Math & Computer Science

Natural Sciences & Engineering

Social Sciences & Humanities

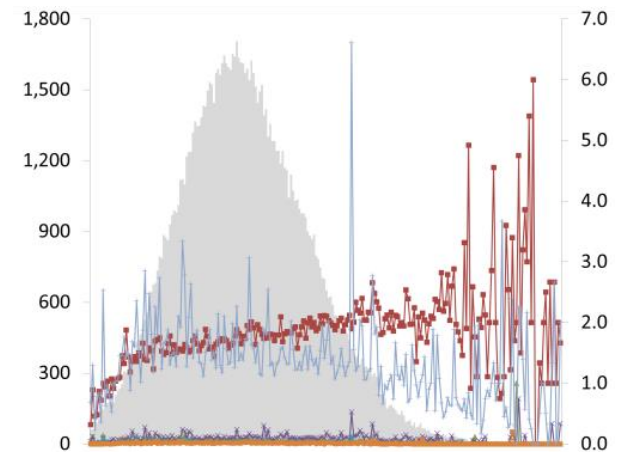# **Results**: title length

Results: title length

Life and Earth Sciences

Natural Sciences & Engineering
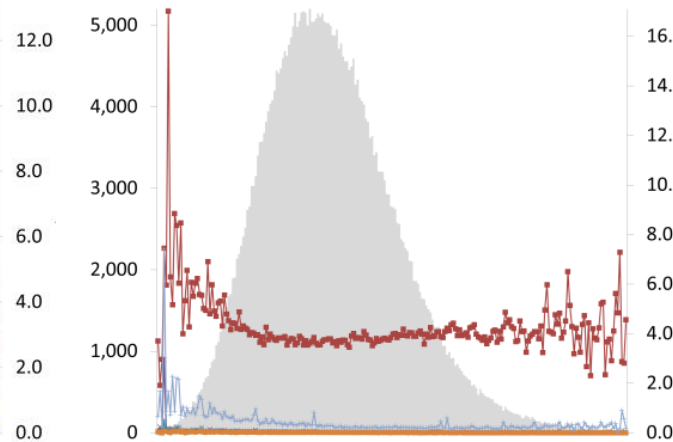
Social Sciences & Humanities

Conclusions

- social media metrics are heterogeneous
    - they vary among each other
    - they differ from citation patterns
- social media metrics are more skewed than citations
- except for Twitter, signals are so low that they do not allow for the computation of meaningful statistics
    - means and correlations are influenced by the inflation of zeros

➤ further research
    - use of catchy words
    - social factors
    - is research with known societal impact more tweeted?

Thanks to Euan Adie and ⬤ Altmetric for access to their data!

Please submit to *Aslib Journal of Information Management*
Special Issue Social Media Metrics in Scholarly Communication: exploring tweets, blogs, likes and other altmetrics
Submission Deadline 15.11.2014
http://www.emeraldgrouppublishing.com/products/journals/call_for_papers.htm?id=5754

Thank you for your attention!
Questions?

Vincent Larivière   vincent.lariviere@umontreal.ca

Université de Montréal

Canada Research Chair
on the Transformations of Scholarly Communication
École de bibliothéconomie et des sciences de l'information

ALFRED P. SLOAN FOUNDATION