

To optimize or not to optimize one's h-index...

$$N_{c,tot} = ah^2$$

by Bertil F. Dorch, Daniella Deutz, Evgenios Vlachos, Charlotte Wien, and Dorte Henriksen



Dr. Bertil



Prof. Charlotte



Dr. Evgenios



Dr. Daniella



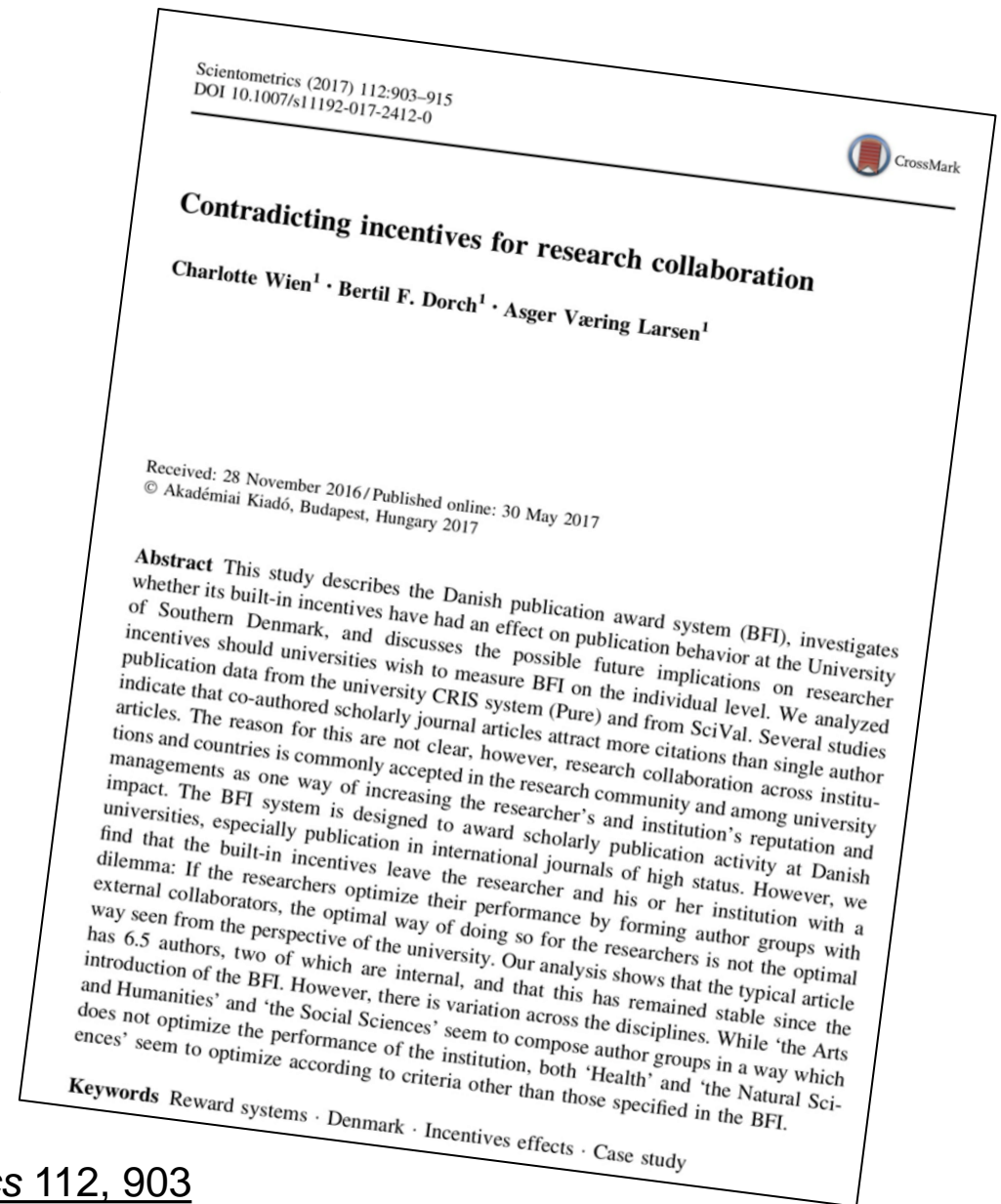
Dr. Dorte

A note on: Quantum mechanics

Incentives for "gaming"?

- General question: Does the current scholarly communication system work *for* research integrity and responsible conduct?
- Potential issues: Retractions, failure of peer review, citation cartels, ghost authorships, potential **gaming of reward mechanisms** ...
- Evidence: Studies indicate that co-authored scholarly journal articles attract more citations than single author articles: even more so for cross-institution and industrial collaborations.
- Ansatz: Non-citation based research assessment (and altmetrics) present alternative incentives to citations
- Case: The Nordic BFI model includes a reward for collaboration (co-authoring).

Wien, Dorch, Larsen (2017) *Scientometrics* 112, 903



Incentives for "gaming"?

- General question: Does the current scholarly communication system work *for* research integrity and responsible conduct?
- Potential issues: Retractions, failure of peer review, citation cartels, ghost authorships, potential **gaming of reward mechanisms** ...
- Evidence: Studies indicate that co-authored scholarly journal articles attract more citations than single author articles: even more so for cross-institution and industrial collaborations.
- Ansatz: Non-citation based research assessment (and altmetrics) present alternative incentives to citations
- Case: The Nordic BFI model includes a reward for collaboration (co-authoring) of 25%

Table 1 No. of BFI credits awarded the institution for a level 2 article with multiple authors from different institutions

No. of internal authors	No. of external authors									
	0	1	2	3	4	5	6	7	8	9
1	3.00	1.88	1.25	0.94	0.75	0.63	0.54	0.47	0.42	0.38
2	3.00	2.50	1.88	1.50	1.25	1.07	0.94	0.83	0.75	0.68
3	3.00	2.81	2.25	1.88	1.61	1.41	1.25	1.13	1.02	0.94
4	3.00	3.00	2.50	2.14	1.88	1.67	1.50	1.36	1.25	1.15
5	3.00	3.13	2.68	2.34	2.08	1.88	1.70	1.56	1.44	1.34
6	3.00	3.21	2.81	2.50	2.25	2.05	1.88	1.73	1.61	1.50
7	3.00	3.28	2.92	2.63	2.39	2.19	2.02	1.88	1.75	1.64
8	3.00	3.33	3.00	2.73	2.50	2.31	2.14	2.00	1.88	1.76
9	3.00	3.38	3.07	2.81	2.60	2.41	2.25	2.11	1.99	1.88
10	3.00	3.41	3.13	2.88	2.68	2.50	2.34	2.21	2.08	1.97

Table 2 No. of BFI credits awarded the individual researcher for a level 2 article with multiple authors from different institutions

No. of internal authors	No. of external authors									
	0	1	2	3	4	5	6	7	8	9
1	3.00	1.88	1.25	0.94	0.75	0.63	0.54	0.47	0.41	0.38
2	1.50	1.25	0.94	0.75	0.63	0.54	0.47	0.41	0.38	0.34
3	1.00	0.94	0.75	0.63	0.54	0.47	0.41	0.38	0.34	0.31
4	0.75	0.75	0.63	0.54	0.47	0.41	0.38	0.34	0.31	0.29
5	0.60	0.63	0.54	0.47	0.41	0.38	0.34	0.31	0.29	0.27
6	0.50	0.54	0.47	0.41	0.38	0.34	0.31	0.29	0.27	0.25
7	0.45	0.47	0.41	0.38	0.34	0.31	0.29	0.27	0.25	0.23
8	0.38	0.41	0.38	0.34	0.31	0.29	0.27	0.25	0.23	0.22
9	0.33	0.38	0.34	0.31	0.29	0.27	0.25	0.23	0.22	0.21
10	0.30	0.34	0.31	0.29	0.27	0.25	0.23	0.22	0.21	0.2



Incentives: No optimization for institution

Scientometrics (2017) 112:903–915

913

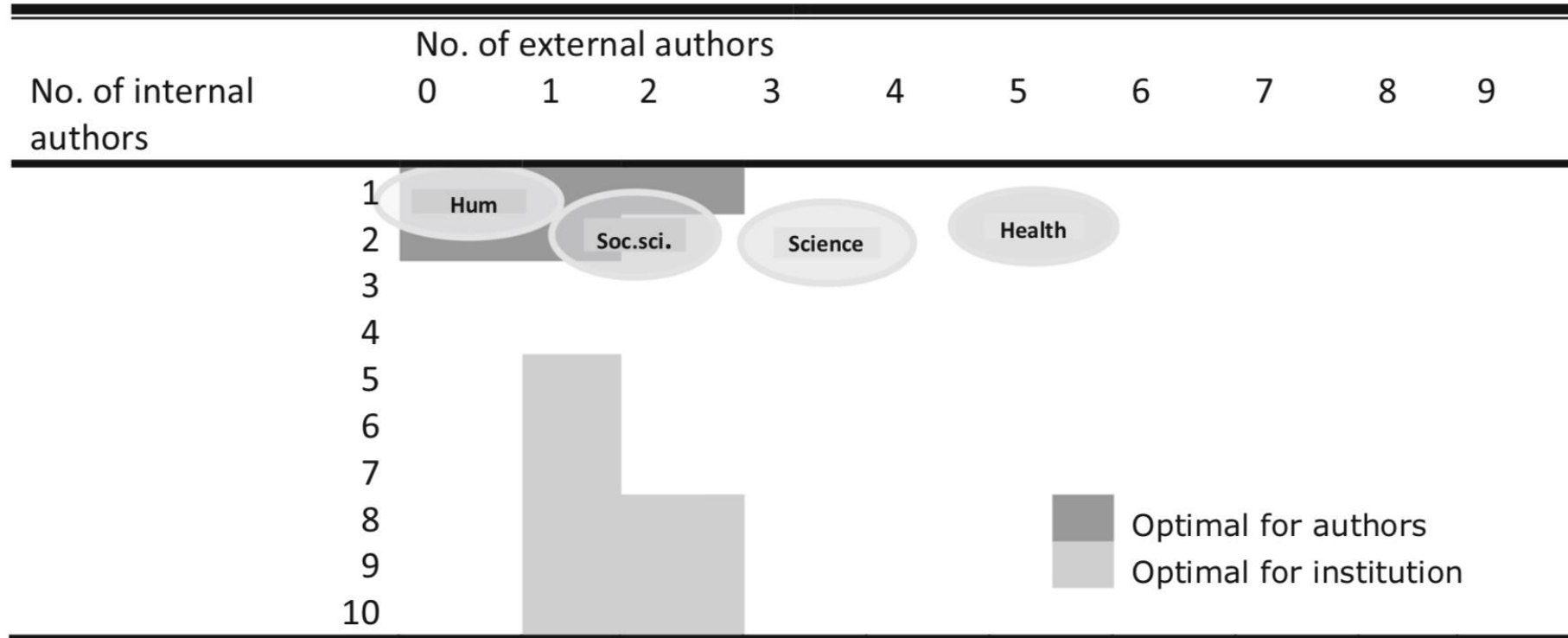


Fig. 6 Optimal and actual distribution of authors from the four main areas

Incentives: No (recent) gaming evident

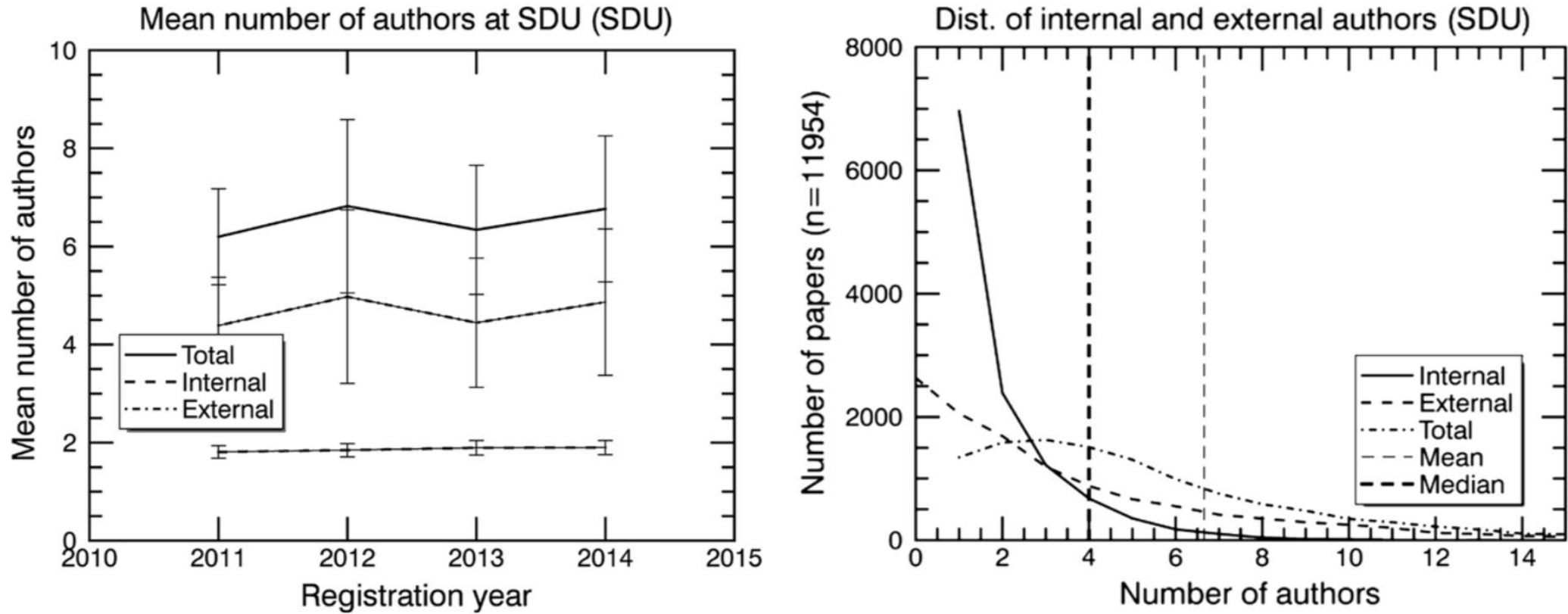


Fig. 1 Mean number of authors and distribution of internal and external authors University of Southern Denmark (2011–2014, journal articles)

Current project: What about the h-index?

Gaming the h-index?

The h-index is easy and popular

- The h -index is an author-level metric that *attempts* to measure the productivity AND citation impact of an individual.
- When proposing the h -index Hirsh stressed that it could
never give more than a rough approximation to an individual's multifaceted profile.
- Despite this the h-index is a very popular and relatively simple measure.

The screenshot shows the arXiv preprint page for the paper "An index to quantify an individual's scientific research output" by J.E. Hirsch. The page is from the Cornell University Library and is part of the physics category. It includes a search bar, a download section with options for PDF, PostScript, and other formats, and a submission history section. The paper is submitted on 3 Aug 2005 (v1) and last revised on 29 Sep 2005 (this version, v5). The abstract states: "I propose the index h , defined as the number of papers with citation number higher or equal to h , as a useful index to characterize the scientific output of a researcher." The submission history shows five versions of the paper, with the final version (v5) submitted on 29 Sep 2005. The page also includes a "Download:" section with links for PDF, PostScript, and other formats, and a "References & Citations" section with a link to NASA ADS. There are also "13 blog links" and a "Bookmark" section. The footer of the page includes a link back to the arXiv form interface and a contact link.

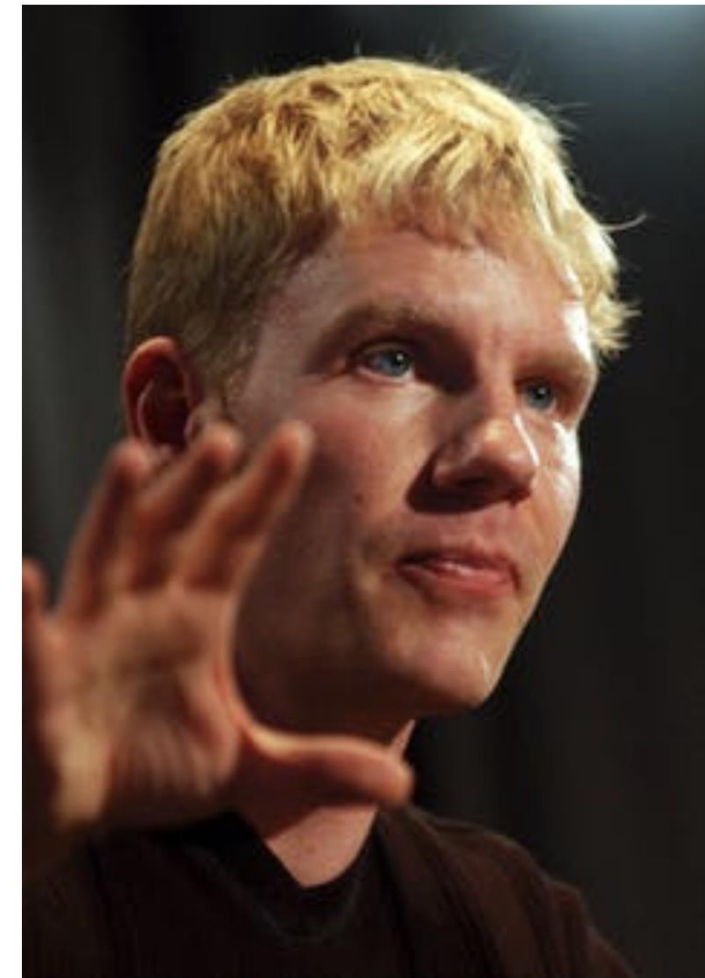
Gaming the h-index?

Ongoing study on researcher behaviour

- Assumption: Reasons for gaming the h -index are the same as for gaming citations, e.g.

boosting one's CV and increasing the chances of funding and promotion...

- Dilemma: We are critical towards h being used as an impact indicator for individuals, but we investigate how researchers (can) optimize their publishing strategies. Research can be misused.
- Question: What are the strategies of a “high h -index achiever” vs. “a low h -index achiever” ?



Bjorn Lomborg was said to have a low H-index and thus many questioned his appointment.

AAP/Alan Porritt

Spicer (2015), [The Conversation](#), May 21.

h-index basics

The Hirsch h -index is the number h of publications that have been cited h times

- In an author's complete publication list sorted by decreasing number of citations, the h -index equals the number of citations of the publication that matches that index number on the list.
- The h -index for an individual author can never exceed the number of his or her citable objects. E.g. if an author has published N_p publications, then h is at most equal to N_p .
- This means that however often an author is cited, h remains bound by N_p .
- A naive but necessary strategy is then to increase the number of publications N_p .

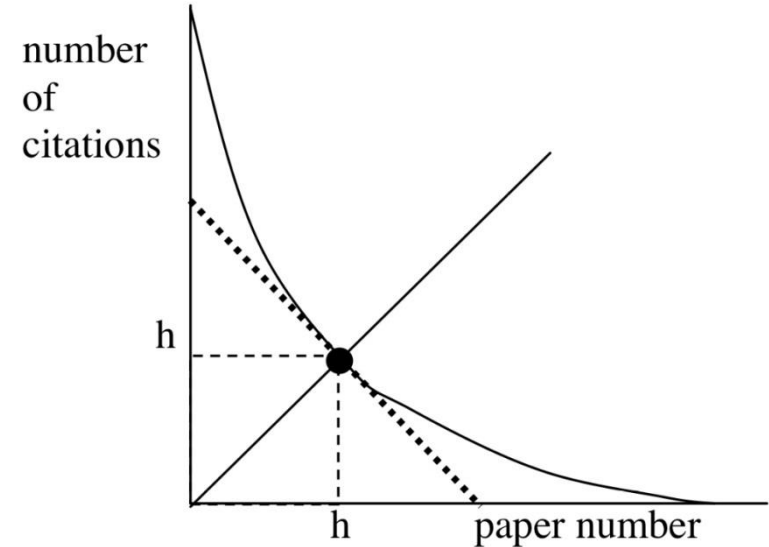
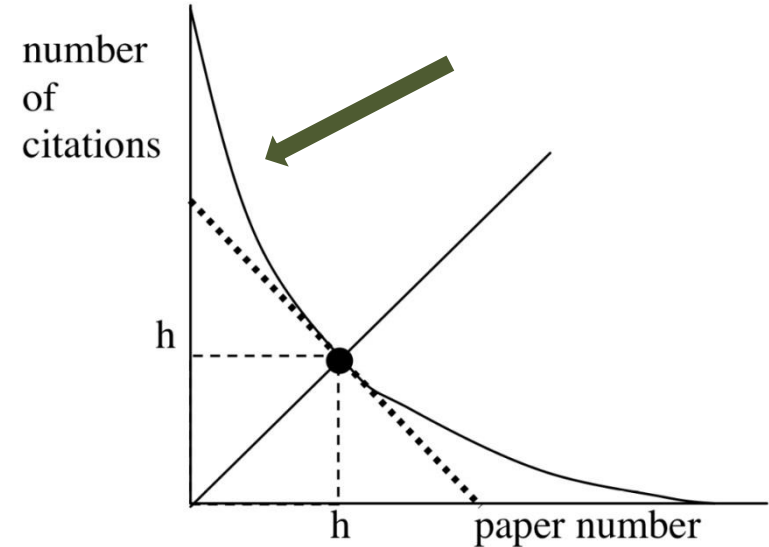


FIG. 1: The intersection of the 45 degree line with the curve giving the number of citations versus the paper number gives h . The total number of citations is the area under the curve. Assuming the second derivative is non-negative everywhere, the minimum area is given by the distribution indicated by the dotted line, yielding $a=2$ in Eq. 1.

Hirsch (2005), arXiv:physics/0508025 .

Efficiency \equiv a bounds

- The h -index is limited by the total number of citations $N_{c,tot}$ through Hirsch's first equation, where $\mathbf{a} \geq 1$.
- The squared h cannot be greater than N_c / \mathbf{a} .
- $\mathbf{a} = 1$ corresponds to the most "efficient" distribution, i.e. a step function resulting from all cited papers being cited h times, while the rest of the papers are not cited at all. $\mathbf{a} = 2$ is a straight line.
- Hirsch empirally finds \mathbf{a} is = 3 – 5.
- The lower your \mathbf{a} , the higher your h for a given citations record – or a given citation impact.



$$N_{c,tot} = ah^2$$

Focusing on a

- Quantitative data:
Citations and publications for 75 researchers from the Dept. of Clin. Res. (SDU).
 - Visually representation:
Scatter plots of various parameters distinguish high achievers from low achievers.
- ✓ Most researchers have an **a** between 3–5, as noted by Hirsch 2005 (for physicists).
- ✓ The value of the h-index may not be comparable from one author to another.

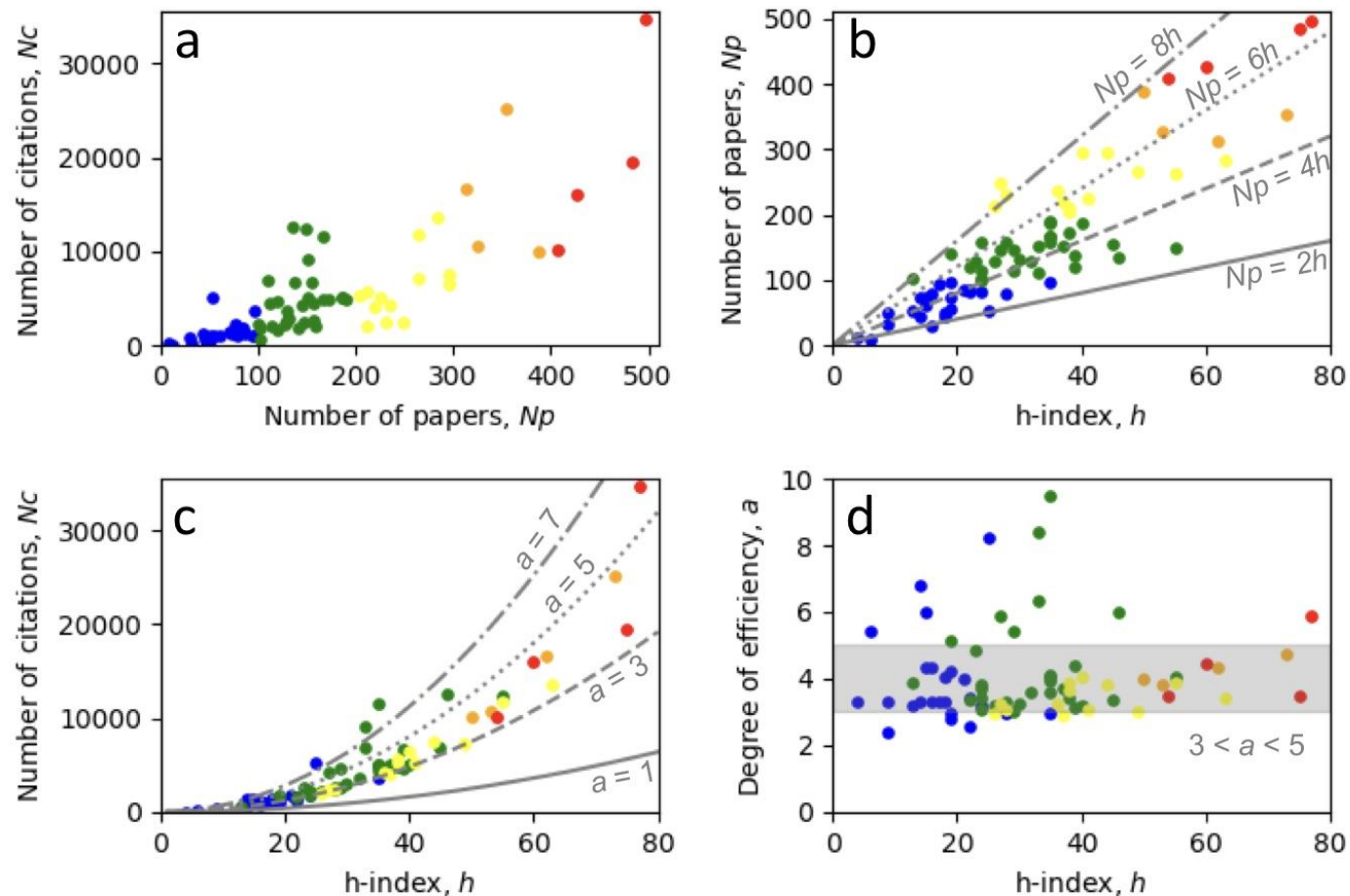


Figure 1. Publication data of 75 researchers affiliated with the Department of Clinical Research at SDU. (a) Number of citations, N_c , versus the number of papers, N_p , of each researcher. (b) N_p plotted against the h-index, h , of each researcher. The line represents the N_p equal to h . (c) N_c versus h . Each line represents the curve for the degree of efficiency, a . (d) The derived 'a' value of each researcher versus h . The highlighted grey square represents the expected spread of a from 3 to 5, as described in (Hirsch 2005).

Data: Low and high h -index "performers"

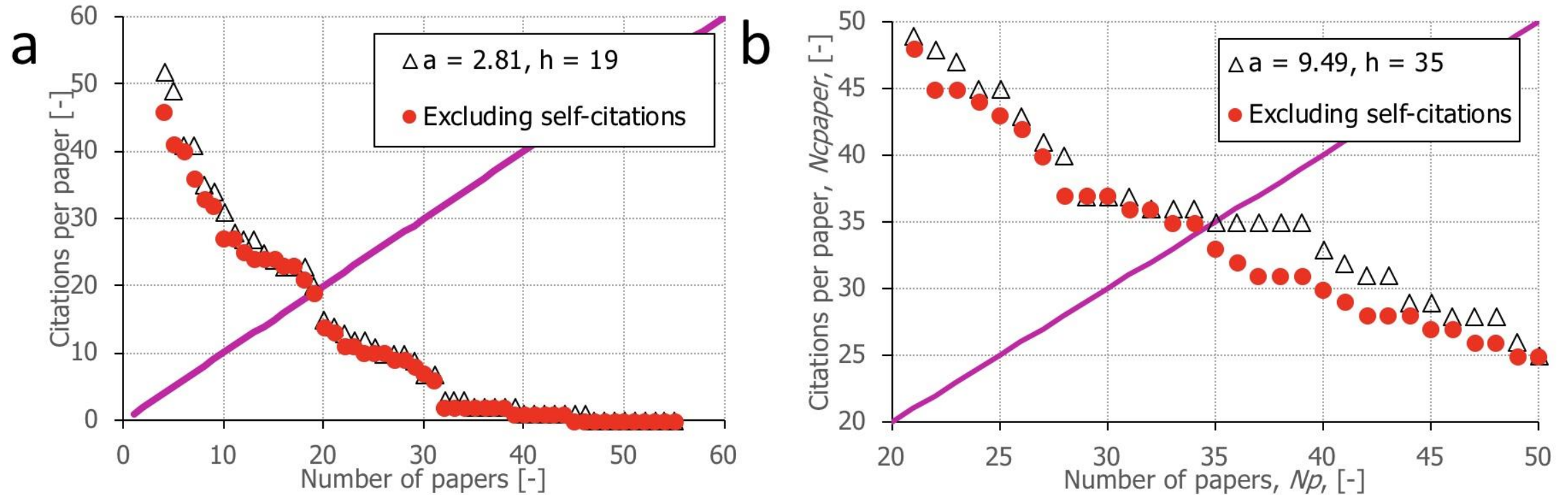


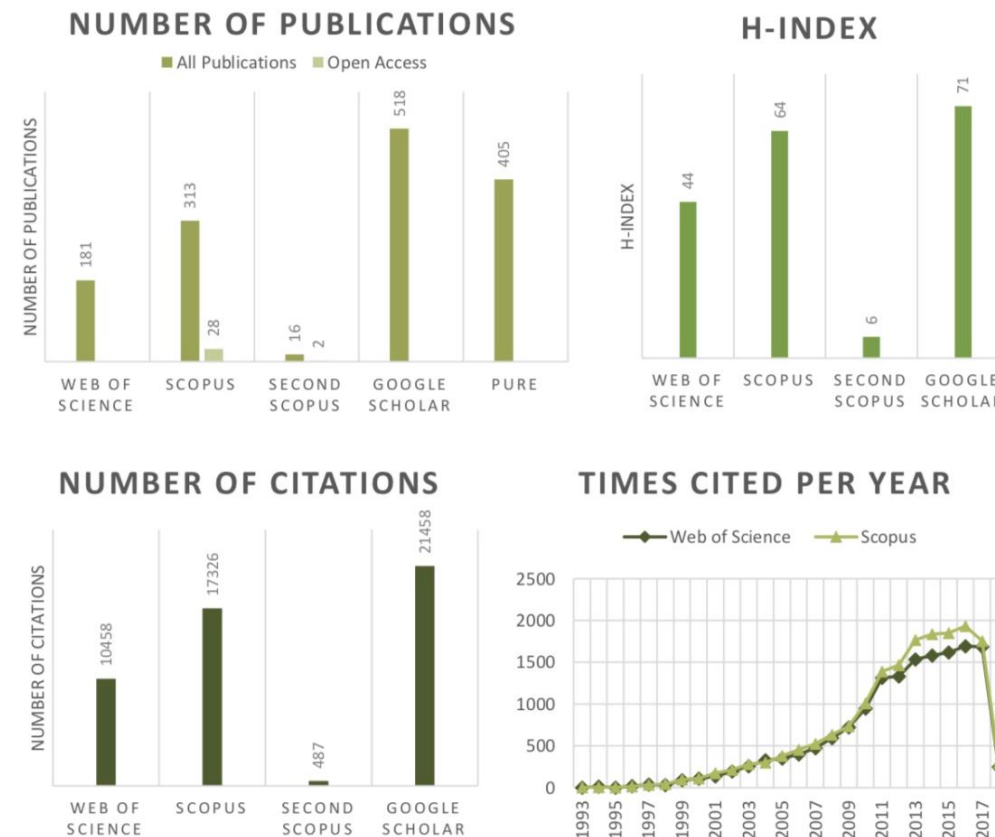
Figure 2. h -index curves for two outlier researchers. (a) Low a , low h . (b) High a , med h .

Interviews, ongoing

- Qualitative data:**
 18 researchers with high h-index from Clinical Research (SDU) invited to a recorded interview on their publication strategy (10 researchers with low-a and 8 with high-a).
- Data so far:**
 Conducted 9 interviews (5 with low a, and 4 with high a researchers).
- Method:**
 Each semi-structured interview lasts approx. 10 minutes and are conducted at the office of the interviewees. Interviewees gave oral consent for being recorded.
- Interviews take outset in our concept Research Footprint, generated individually for each researcher.

RESEARCH FOOTPRINT

RESEARCHER: KIRSTEN O. KYVIK



ORCID



0000-0003-2981-0245, Coupled to PURE?

RESEARCHERID

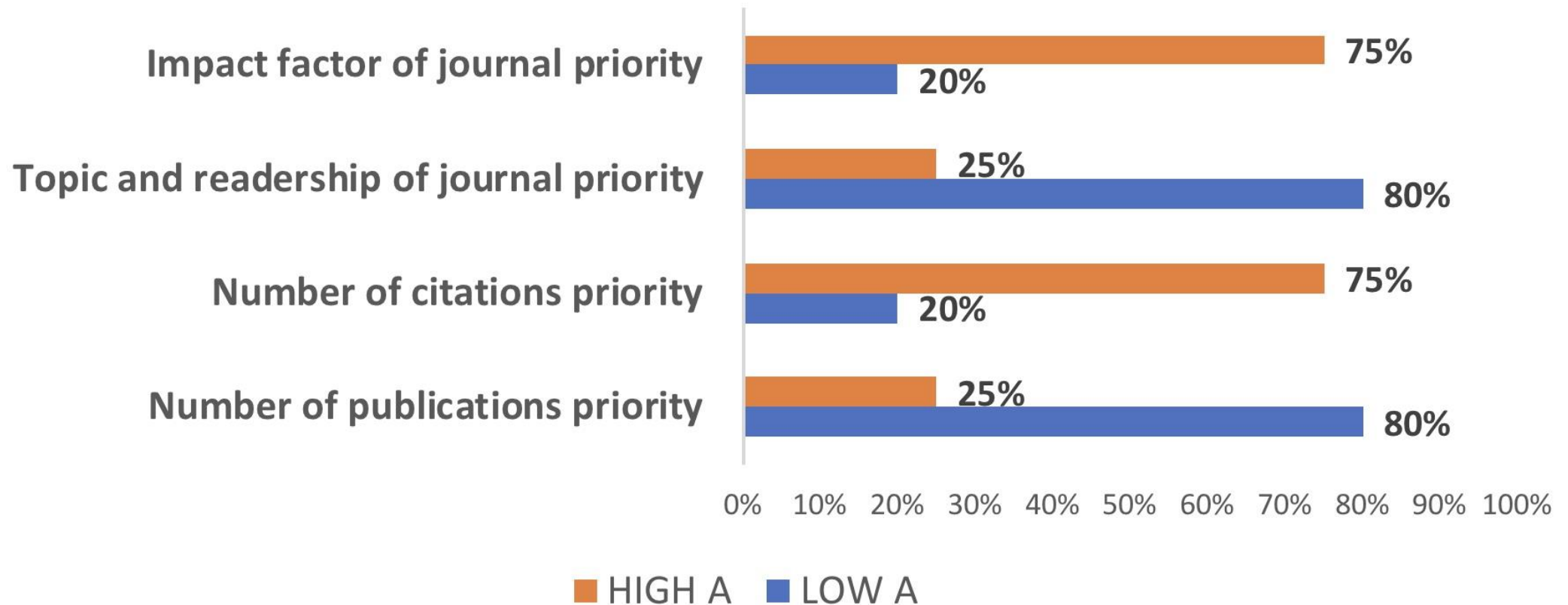


K-5680-2016



Preliminary results summarized

FACTORS AFFECTING EFFECTIVENESS



Discussion & The End

Questions

1. Academic: Could we derive a practice for optimizing the h-index by considering the corresponding publication behavior in more detail?
 2. Ansatz: Is there a practice of focusing at the $h+1$ 'th paper?
 3. Conflicting incentives: If one tries to optimize one's h-index, does this conflict with possible incentives resulting from other kinds of optimizing – in other words, are there opposing strategies at work?
- Remains to be answered cf. forthcoming ...

Note on biases

- On the one hand, our proposed measure of efficiency focus on achieving as high an h as possible given the total number of citations, i.e. it indirectly favors lower total citation counts.
- Also, it does not favor high citation counts to individual papers.
- On the other hand, the efficiency measure is blind towards other non-citation based bibliometric indicators such as impact factors, *altmetrics* and research evaluation exercises.
- When the focus is on the individual paper, i.e. how to increase its citation impact, it is a process that must be very subject specific and content oriented.