

Are self-citations a normal feature of knowledge accumulation or a perversion of research evaluation?

Philippe Vincent-Lamarre
Cassidy R. Sugimoto
Vincent Larivière

vincent.lariviere@umontreal.ca
[@lariviev](https://twitter.com/lariviev)
crc.ebsi.umontreal.ca



Chaire de recherche du Canada sur
les transformations de la communication savante
École de bibliothéconomie et des sciences de l'information



Outline

- Common criticism of citation analysis
- Misconceptions around self-citations
- Two analyses
 - Self-citations vs self-references throughout the career of individuals
 - Semantic similarity between citing and cited papers
- Conclusions

Criticism of citation analysis

- Takes time for citations to accumulate (citation delays)
- Uncited influences
- Negative citations

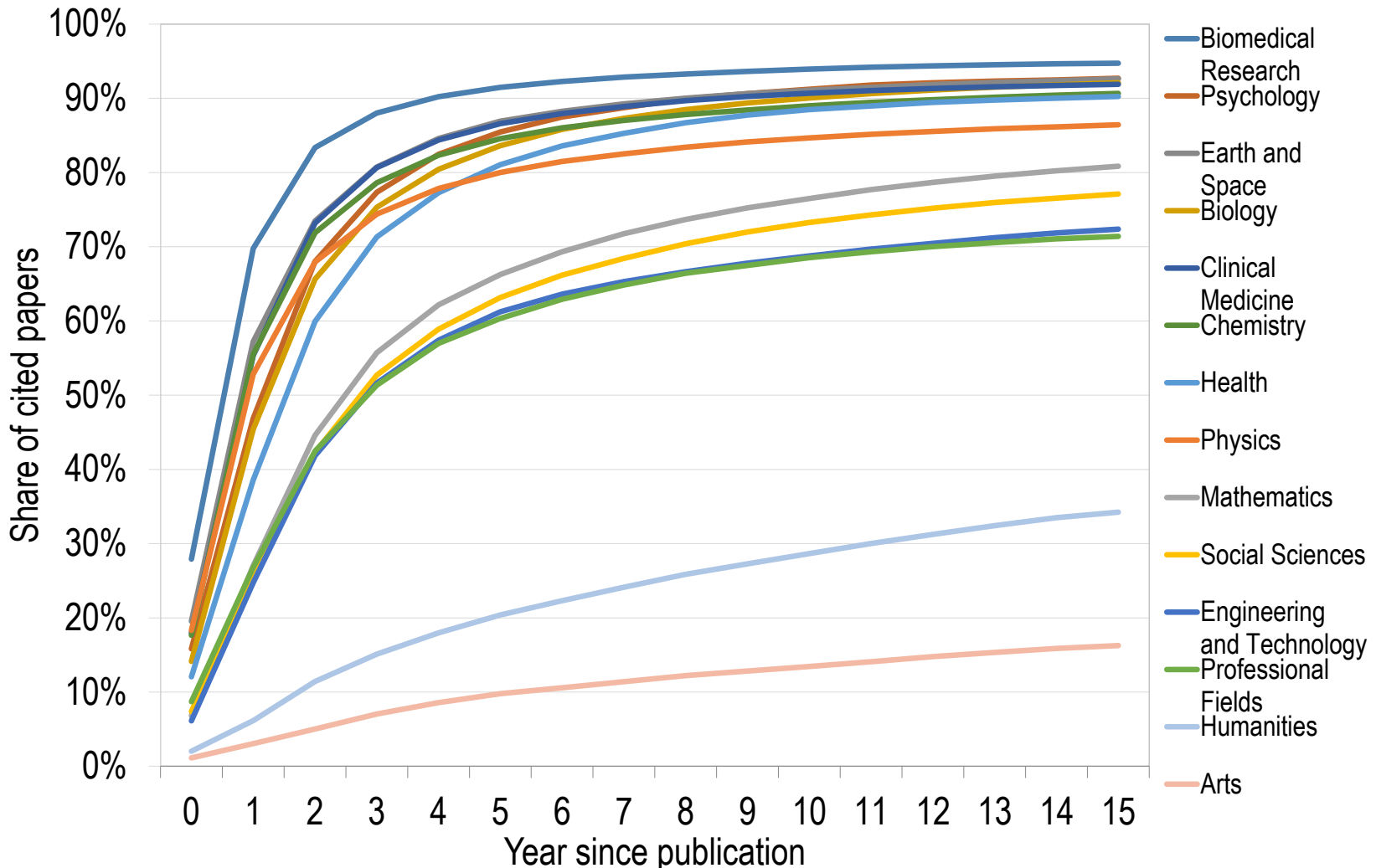
Criticism of citation analysis

- Takes time for citations to accumulate (citation delays)
- Uncited influences
- Negative citations
- Perfunctory citations

Criticism of citation analysis

- Takes time for citations to accumulate (citation delays)
- Uncited influences
- Negative citations
- Perfunctory citations
- Most papers are uncited

Percentage of cited articles

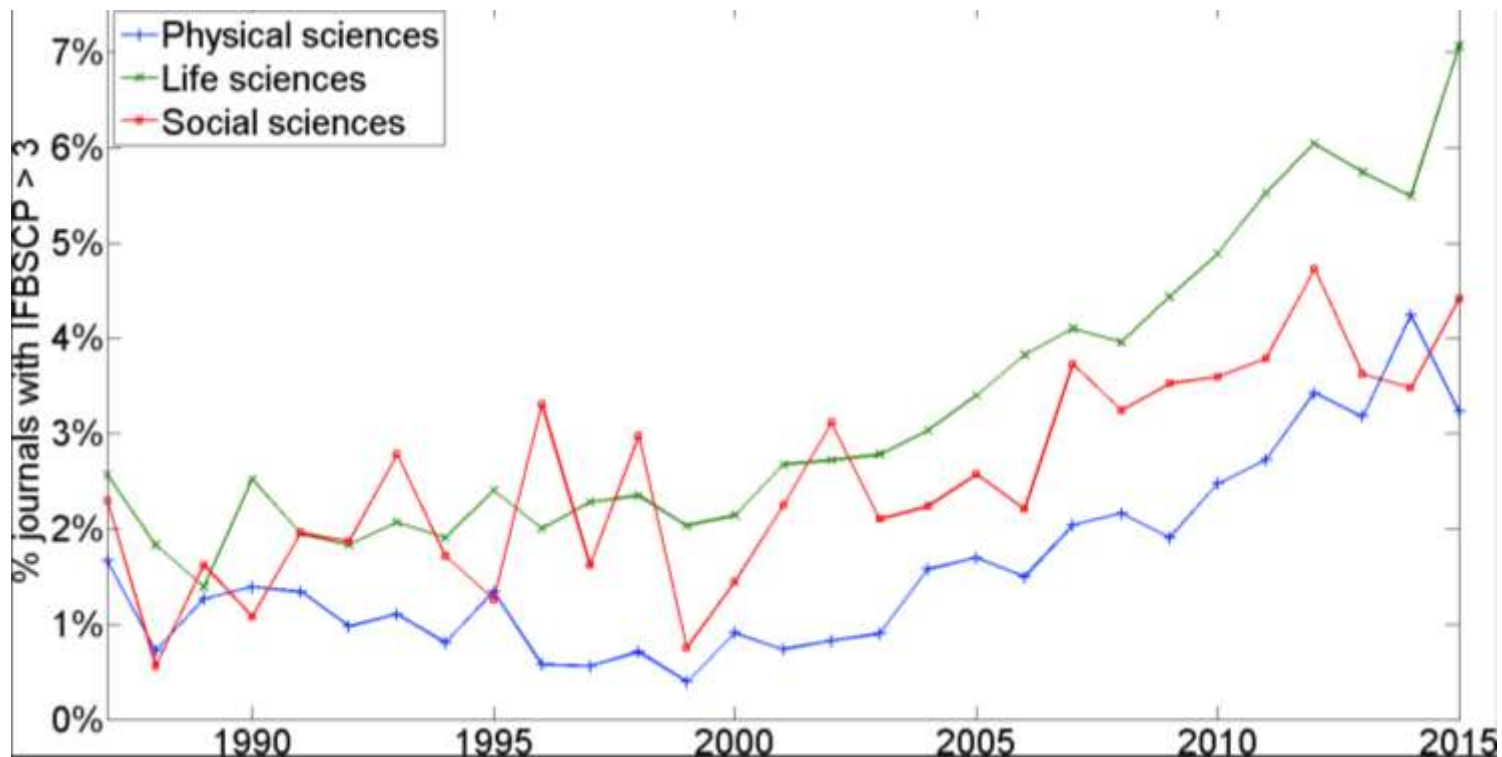


Criticism of citation analysis

- Takes time for citations to accumulate (citation delays)
- Uncited influences
- Negative citations
- Perfunctory citations
- Most papers are uncited
- Self-citations
 - Citation cartels
 - Journal
 - National
 - Paper
 - Researcher

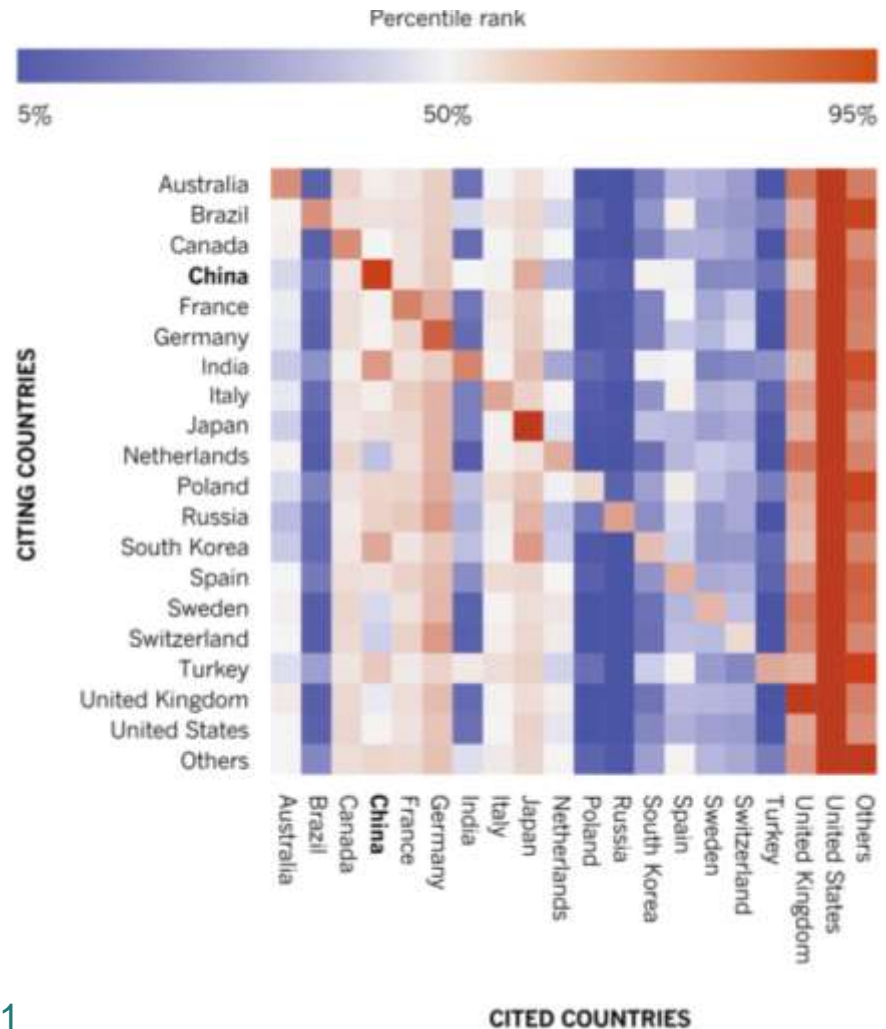
Journal self-citations

Increase in journals that self-cite their own papers more frequently in the JIF window



National self-citations

- Researchers are more likely to draw on research from the same country



Two conceptions of self-citations

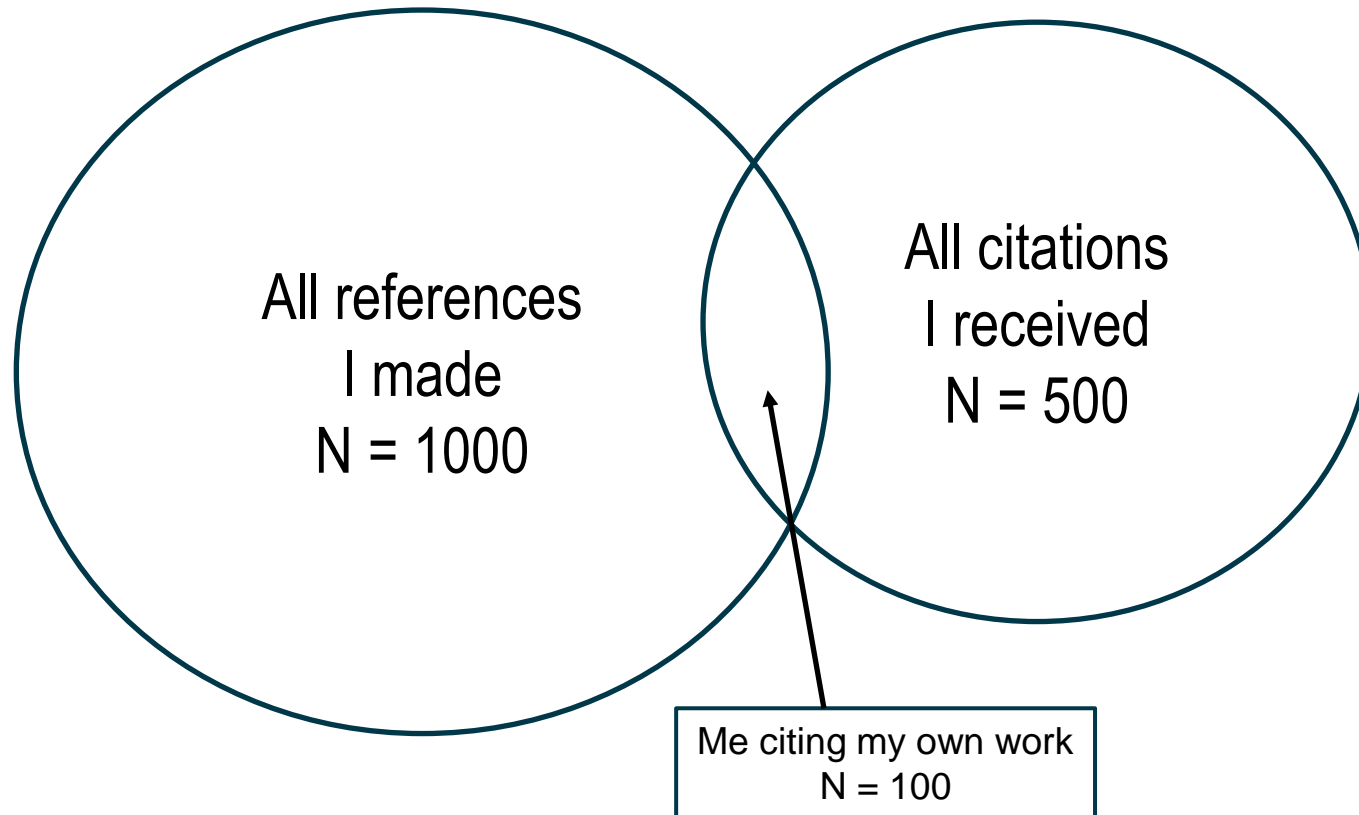
- Normal feature of research activity
 - Building on one's own previous research is natural during a research career
 - More frequent in smaller fields
 - To self-cite, one has to publish an additional article
 - Self-citations could be considered in research evaluation
- Deviant behavior
 - Having impact means influencing others
 - Self-citations are used to increase one's scholarly impact
 - Self-citations should be excluded from research evaluation

Self-citations: conceptual issues

1. Self-citations are different from self-references

- Percentage of self-citations: of all **citations received**, what is the percentage that is made by author
- Percentage of self-references: of all **references made**, what is the percentage that is made by author
- Same numerator, different denominator
 - and different interpretation!
- Self-citations and self-references obey different logics
 - Only self references are fully dependent on the author's referencing practices

Self-citations: conceptual issues



% of self-references:
 $100/1000 = \mathbf{10\%}$

% of self-citations:
 $100/500 = \mathbf{20\%}$

Self-citations at the individual researcher level?

 OPEN ACCESS

COMMUNITY PAGE

A standardized citation metrics author database annotated for scientific field

John P. A. Ioannidis  Jeroen Bass, Richard Klavans, Kevin W. Boyack

Version 2  Published: August 12, 2019 • <https://doi.org/10.1371/journal.pbio.3000384>

Article	Authors	Metrics	Comments	Media Coverage
---------	---------	---------	----------	----------------

Abstract

References

Reader Comments (7)

Media Coverage (25)

Figures

Abstract

Citation metrics are widely used and misused. We have created a publicly available database of 100,000 top scientists that provides standardized information on citations, h-index, coauthorship-adjusted hm-index, citations to papers in different authorship positions, and a composite indicator. Separate data are shown for career-long and single-year impact. Metrics with and without self-citations and ratio of citations to citing papers are given. Scientists are classified into 22 scientific fields and 176 subfields. Field- and subfield-specific percentiles are also provided for all scientists who have published at least five papers. Career-long data are updated to end of 2017 and to end of 2018 for comparison.


Figures

0 Save	31 Citation
173,518 View	7 Share

Download PDF 

Print

Share

 Check for updates

Related PLOS Articles

has **COMPANION**
Updated science-wide author databases of standardized citation indicators
[View Page](#) [PDF](#)

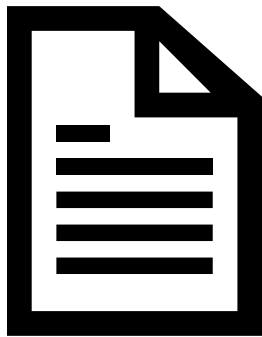
Self-citations at the individual researcher level?

2. Paper level vs. researcher level self-citations

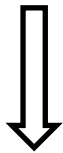
- Common issue in measuring self-citations: compiled at the paper level (as in Ioannidis et al., 2019) rather than researcher level
 - And therefore includes “self-citations” that authors’ themselves have not made
- Researcher Michael Grätzel has received 222,899 citations of which 7.44% are considered as self-citations (16,594 citations).

Author-level vs. Paper level

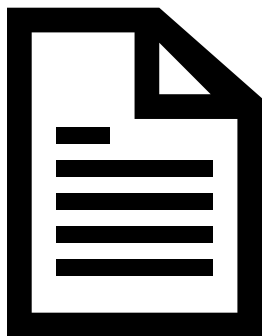
Citing paper



Larivière, V., & **Costas, R.** (2016). How many is too many? On the relationship between research productivity and impact. PLOS One, 11(9), e0162709.



Cited paper



Costas, R., van Leeuwen, T., & Bordons, M. (2010). Self-citations at the meso and individual levels: effects of different calculation methods. Scientometrics, 82(3), 517-537.

Paper level

- Self-citation / self-reference for each citing and cited author

Author level

- Self-reference for **R. Costas**
- Self-citation for **R. Costas**
- External reference for **V. Larivière**
- External citation for **T. van Leeuwen** and **M. Bordons**

Self-citations at the author-level

- Necessary step to understand individual researchers' citing behavior
 - In the paper-level self-citation analysis, self-citations made by co-authors are counted as one's self-citations
- Challenge: author disambiguation at a large-scale

Research questions

- What are the dynamics of self-citations and self-references?
 - How are self-citations and self-references evolving as a function of one's career?
 - How are self-citations evolving as a function of one's research output?
 - Can you become “rich” with self-citations?
 - Gender differences in self-citations?
- How similar are citing and cited papers?
 - If self-citations represent a deviant behavior, self-cited papers should be more remotely related to the citing paper than non self-cited papers.

Methods

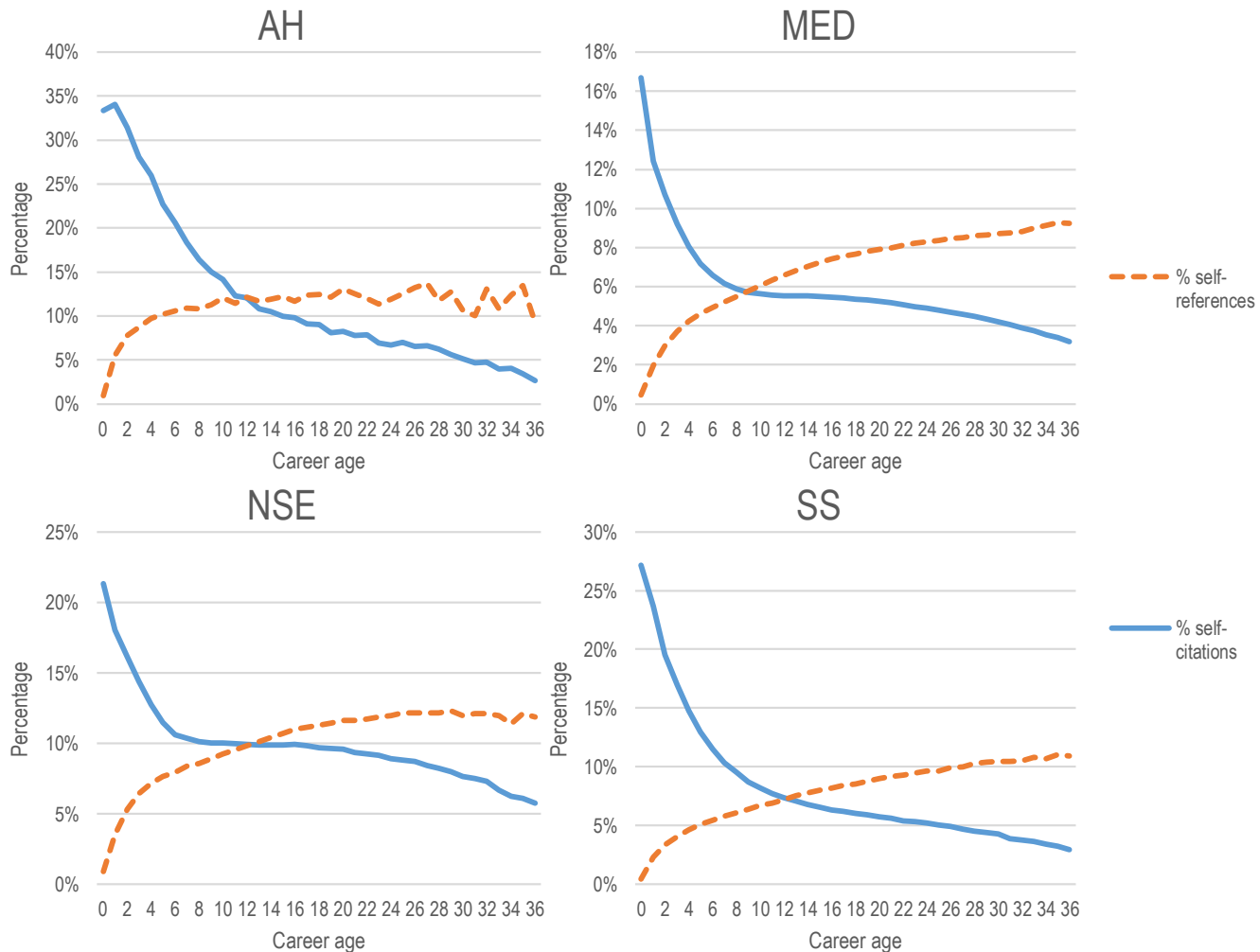
- 51,597,960 authors disambiguated individual researchers from the CWTS Web of Science database
 - Disambiguated using heuristics such as: Name and given name, institution, collaborators, cited references, topics and email addresses
 - Not perfect: likely to split researchers into two entities (rather than group two distinct researchers into one entity)
- Publication data from 1980 to 2018
 - 63,327,731 articles; 881,480,407 citations (source items only)
- Self-citation (reference) compiled at the individual researcher level
- Career age determined by year of first publication

Methods

- Four types of citations

- Direct self-citations: I directly cite a paper on which I was a co-author
- Co-authors self-citations: co-authors of the paper cite the paper
- Collaborators citations: former collaborators cite the paper
- External: other citations than the three above

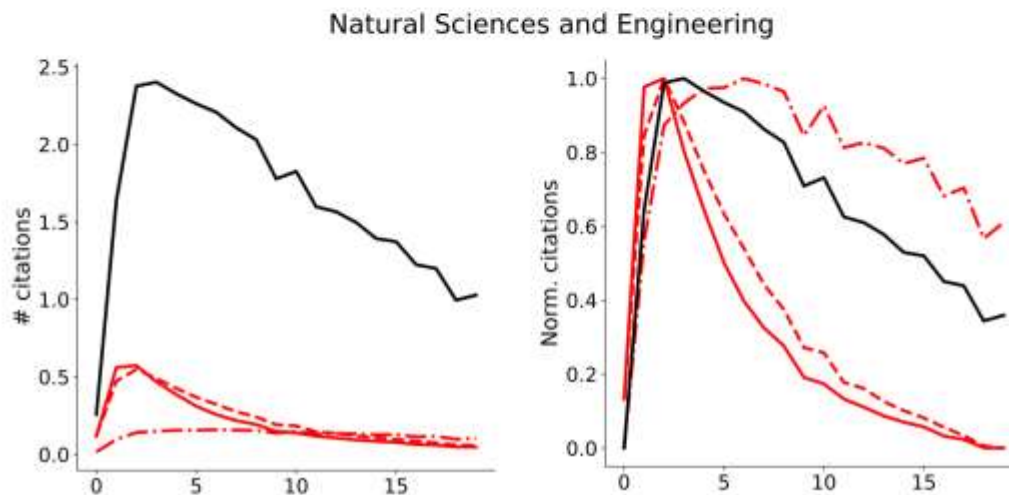
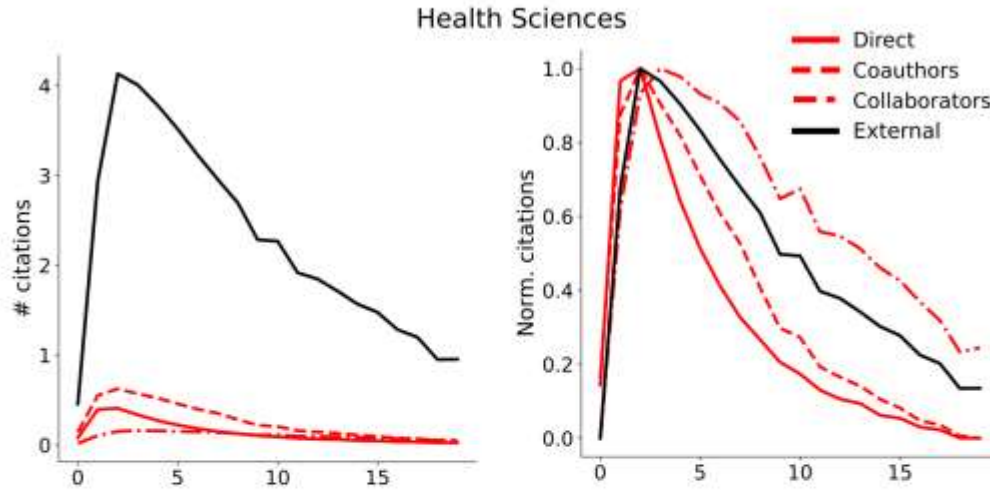
Self-citations and self-references as a function of career age



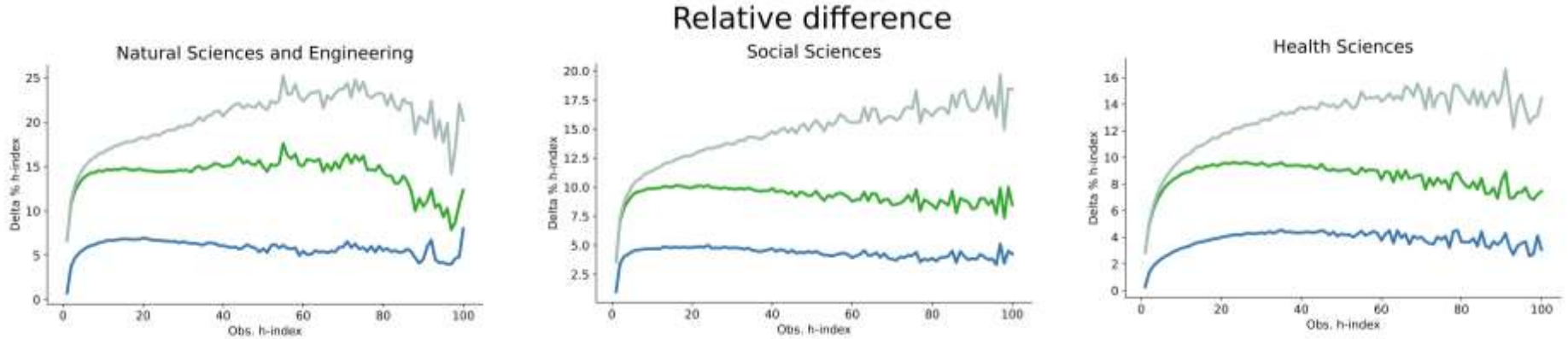
Self-citations by age of cited document

Absolute

Normalized

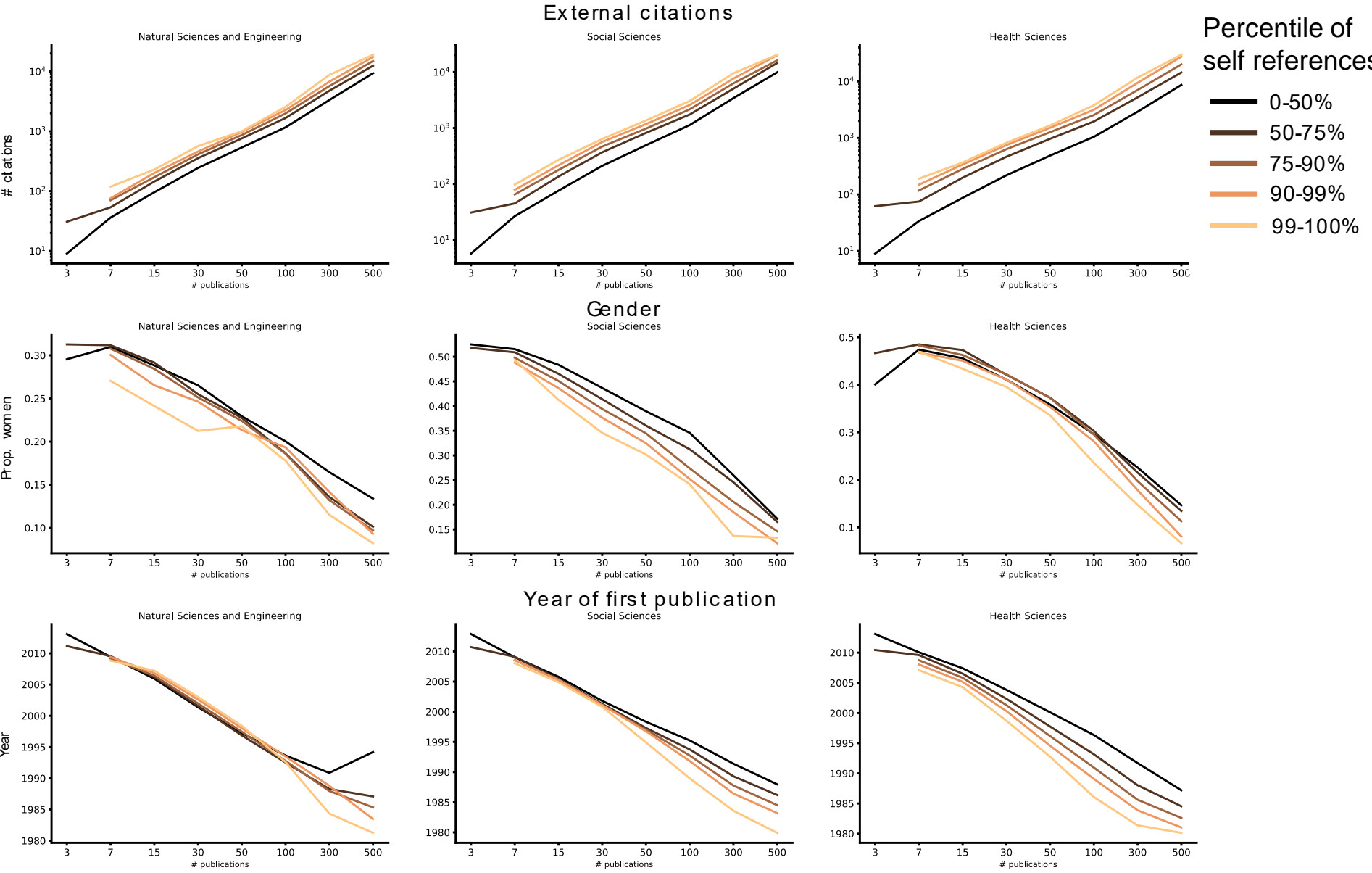


Effect on the H-Index

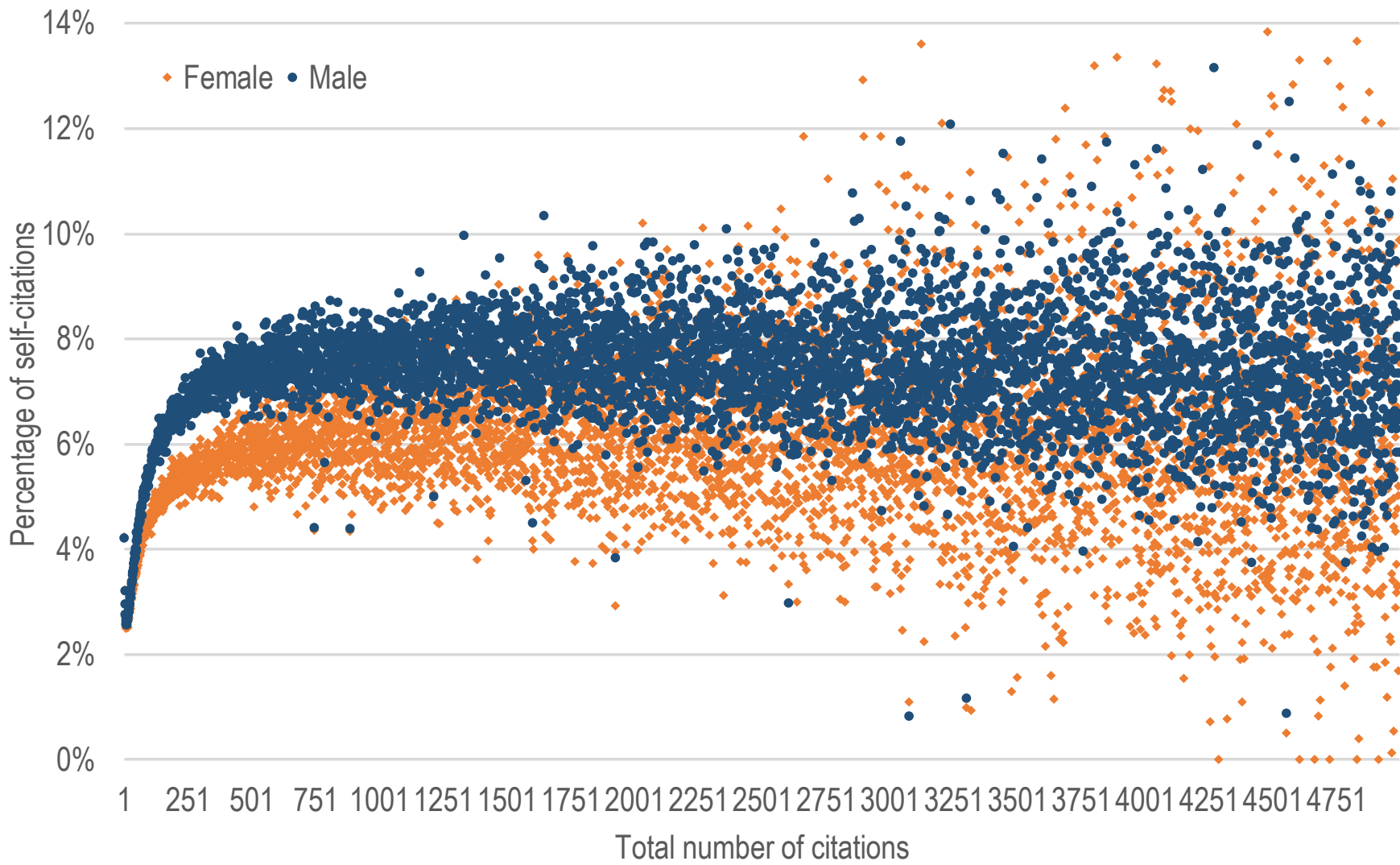


- Obs. - Direct
- Obs. - Direct - Co.
- Obs. - Direct - Co. - Coll.

Correlates of self-references



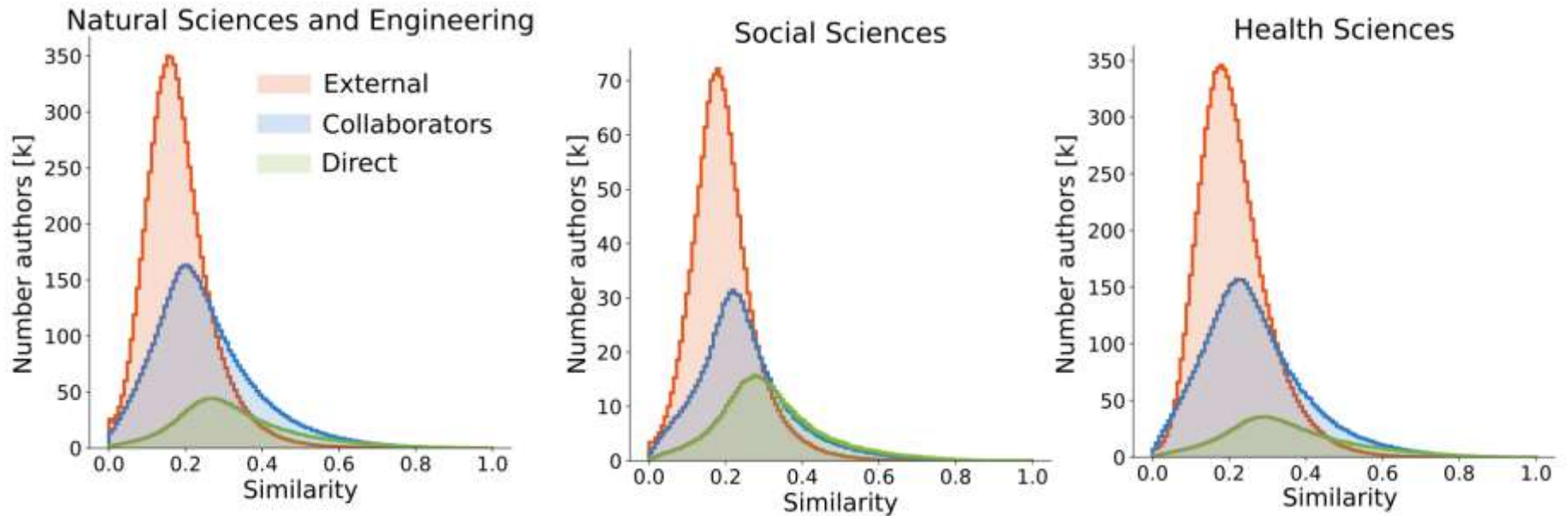
Percentage of self-citations by gender and total citations



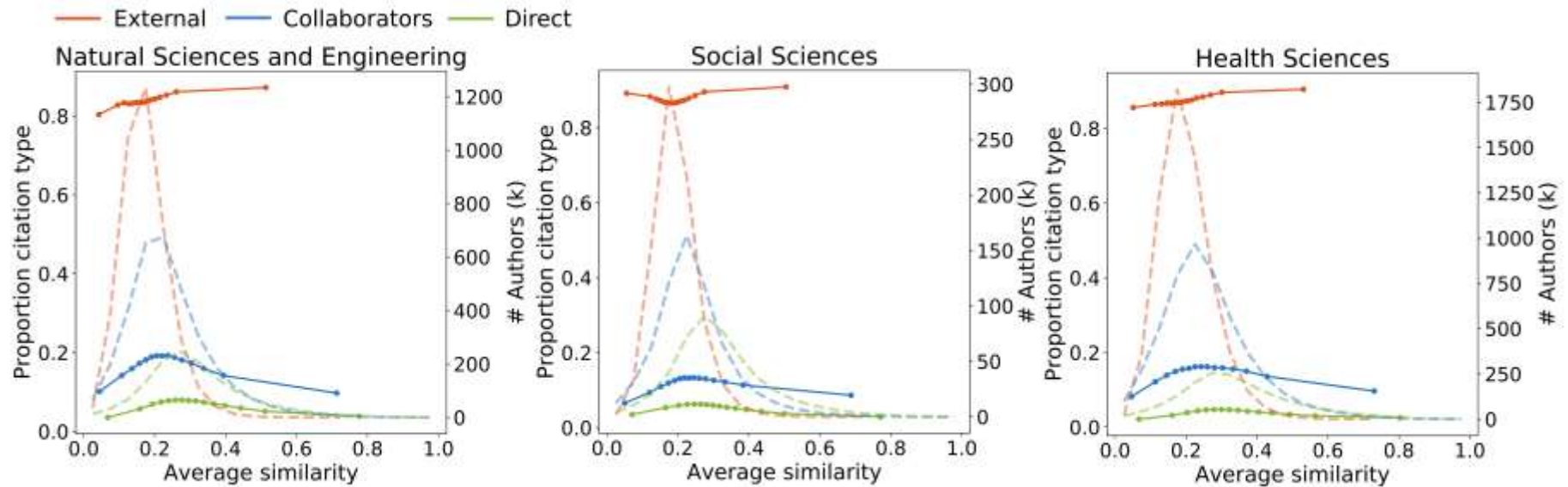
Similarity analysis

- Of the 63,327,731 articles, 26,329,066 had an abstract and a disambiguated author.
 - 15,749,808 pairs of citing-cited articles.
 - Since a given citing-cited pair can involve more than one author this leads to 1,689,855,285 citing-cited comparisons with unique authors.
- The text of each abstract was preprocessed to remove stop words and we took the stem (root) of every word. We then used a tf-idf representation of every abstract, and obtained the cosine similarity between each pair of citing-cited articles.
- We computed an average similarity score for each citation type of each author.

Distribution of citations, by similarity between citing and cited paper

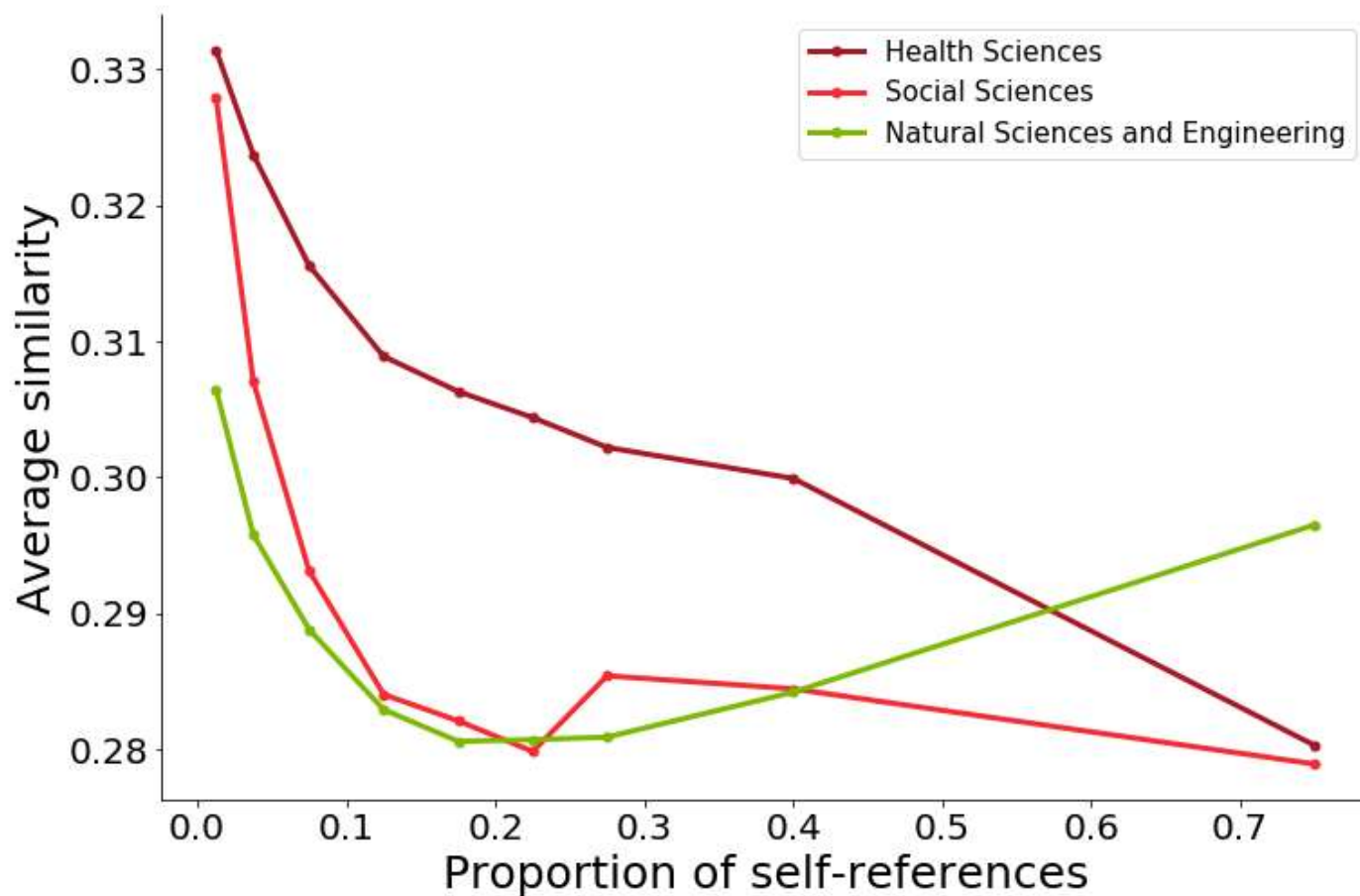


Proportion of citation type as a function of average similarity

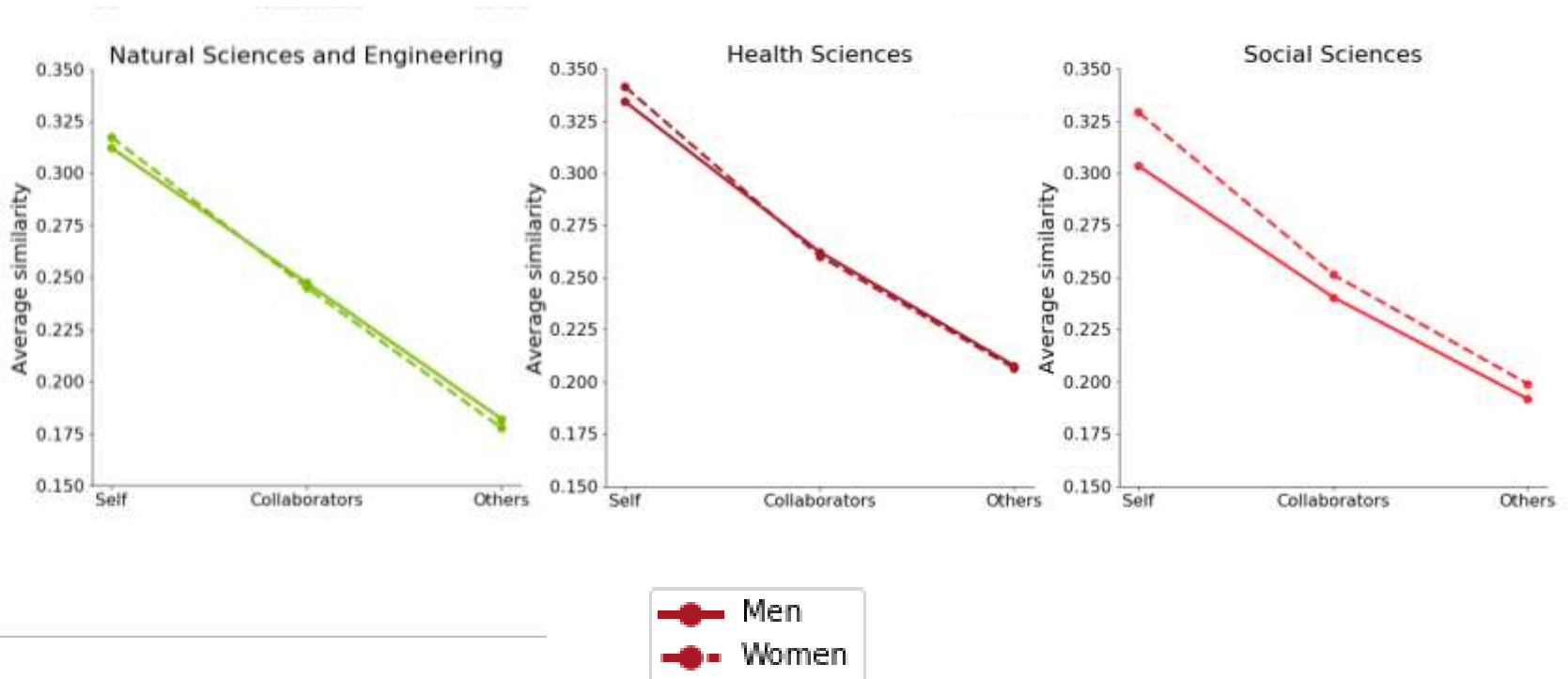


- Authors who self-cite the most and the least are the ones where the similarity between citing and cited papers is the lowest
 - Suggests a deviant behavior at extremes

Similarity as a function of the proportion of self-references



Average similarity of self, collaborators and others' citations, by gender



Conclusions

- Self-citations and self-references follow a different path over the course of a career.
 - Self-citations decrease over time
 - Self-references are stable / increase (especially for highly productive researchers)
- Self-references are more common for younger documents
- External citations are correlated with self-citations
- Direct self-citations have little effect on the h-index once a threshold is reached
 - But citations from collaborators have an important effect
- Women are self-referencing less
 - Men are benefiting from their self-citations
 - Converges with previous research at the paper level

Conclusions (2)

There are “mechanical” aspects to self-citations and self-references

- Young scholars cannot self-cite much, but are likely to self-reference
- Researchers of different cohorts cannot be held to the same “standard”
- Contrary to our hypothesis, self-references are more “relevant” to the citing paper than non-self-citations
 - Especially for self-references made by woman
- Our results suggest that self-citations are, in most cases, a normal feature of knowledge accumulation.
 - Except for the subset of highly self-referencing authors

Thank you!

Philippe Vincent-Lamarre
Cassidy R. Sugimoto
Vincent Larivière

vincent.lariviere@umontreal.ca
[@lariviev](https://twitter.com/lariviev)
crc.ebsi.umontreal.ca



Chaire de recherche du Canada sur
les transformations de la communication savante
École de bibliothéconomie et des sciences de l'information

