Citation analysis: Indicators

Clara Calero
Centre for Science and Technology Studies, Leiden University

Course 'Bibliometrics and Scientometrics for Research Evaluation'Leiden
Leiden 4th November 2019

Outline

- Basic bibliometric indicators
- Intermezzo: h-index

Basic bibliometric indicators

Non-normalized indicators

Size-dependent indicators:

- **P**: Number of publications
- **TCS**: Total citation score

Size-independent indicators:

- **MCS**: Mean citation score

Calculation:

- Only documents classified as ‘article’ or ‘review’ (and sometimes ‘letter’)
- Author self citations are ignored

CWTS

Example

Publication	Journal	Field	Citations
P1	J1	F1	1
P2	J2	F1	8
P3	J3	F2	9

P = 3

Total Citation Score (**TCS**) = 1 + 8 + 9 = 18

Mean Citation Score (**MCS**) = TCS/P=(1 + 8 + 9) / 3 = 6

CWTS

Advanced indicators

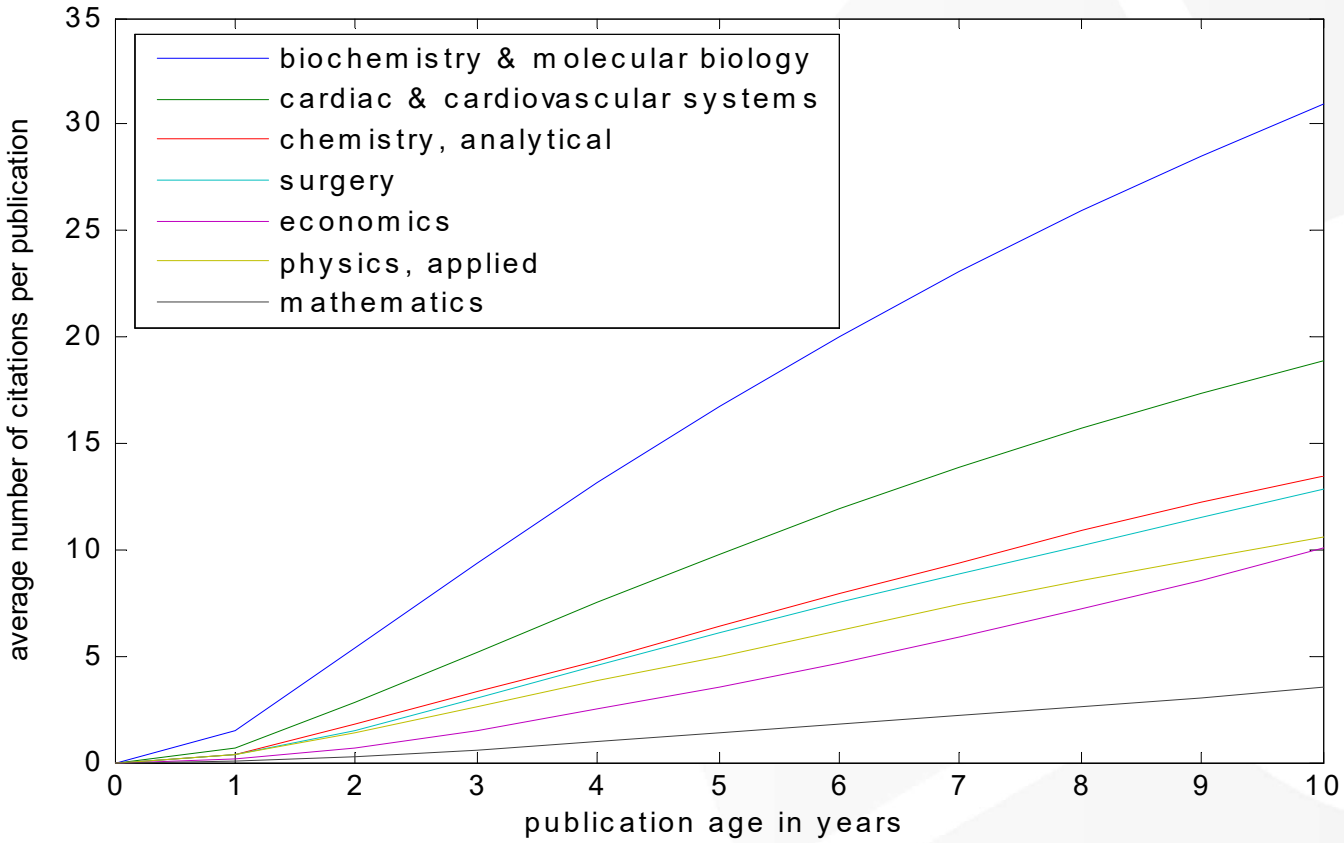
- Reference Framework

- On the level of individual papers

Expected number of citations (Field Citation Score)

- The **expected number of citations** of a publication is defined as the average number of citations of all publications published
  - in the same **scientific field** and
  - in the **same year**

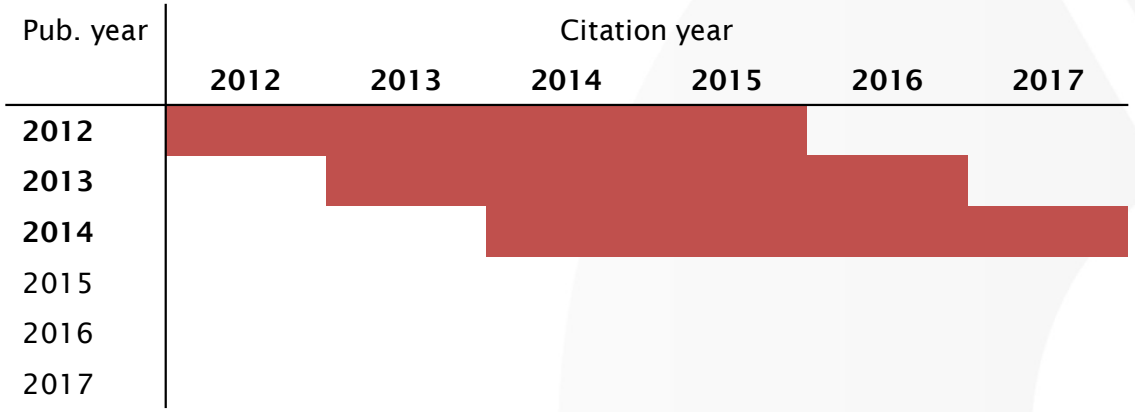
Differences among fields



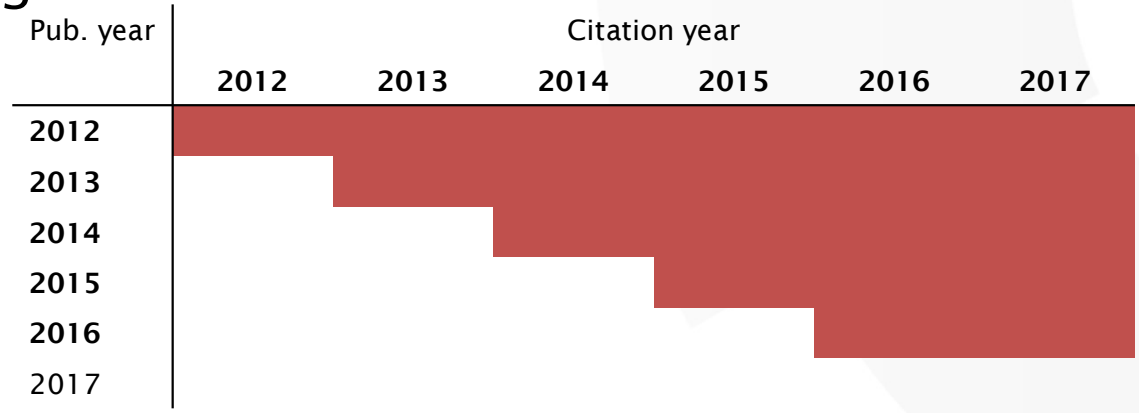
CWTS

Normalization for publication age

- Fixed-length citation window:



- Variable-length citation window:



Normalized indicators

Size-dependent indicators:

- **TNCS**: Total normalized citation score
- **$P_{top\ 10\%}$**: Number of publications in top 10%

Size-independent indicators:

- **MNCS**: Mean normalized citation score
- **$PP_{top\ 10\%}$**: Proportion of publications in top 10%

- Normalization based on
  - Field
  - Year of publication
- Calculation:
  - Only documents classified as ‘article’ or ‘review’ (and sometimes ‘letter’)
  - Author self citations are ignored
  - Citation window length must be at least **2 years** (including the publication year)
- Instead of focusing on top 10% publications, one could also consider for instance top 50%, top 5%, or top 1% publications

CWTS

Example

Average number of citations of all publications in a field (expected number of citations)

Publication	Journal	Field	Citations	Field citation score	Normalized citation score
P1	J1	F1	1	2.32	0.43
P2	J2	F1	8	2.32	3.45
P3	J3	F2	9	14.17	0.64

MNCS = (1 / 2.32 + 8 / 2.32 + 9 / 14.17) / 3

=(0.43+3.45+0.64)/3= **1.51**

CWTS

Example

Publication	Journal	Field	Citations	Field citation score	Field Threshold
P1	J1	F1	1	2.32	3
P2	J2	F1	8	2.32	3
P3	J3	F2	9	14.17	10

Minimum number of citations to be in top 10%

PP(top 10%) = (0 + 1 + 0) / 3 = **0.33**

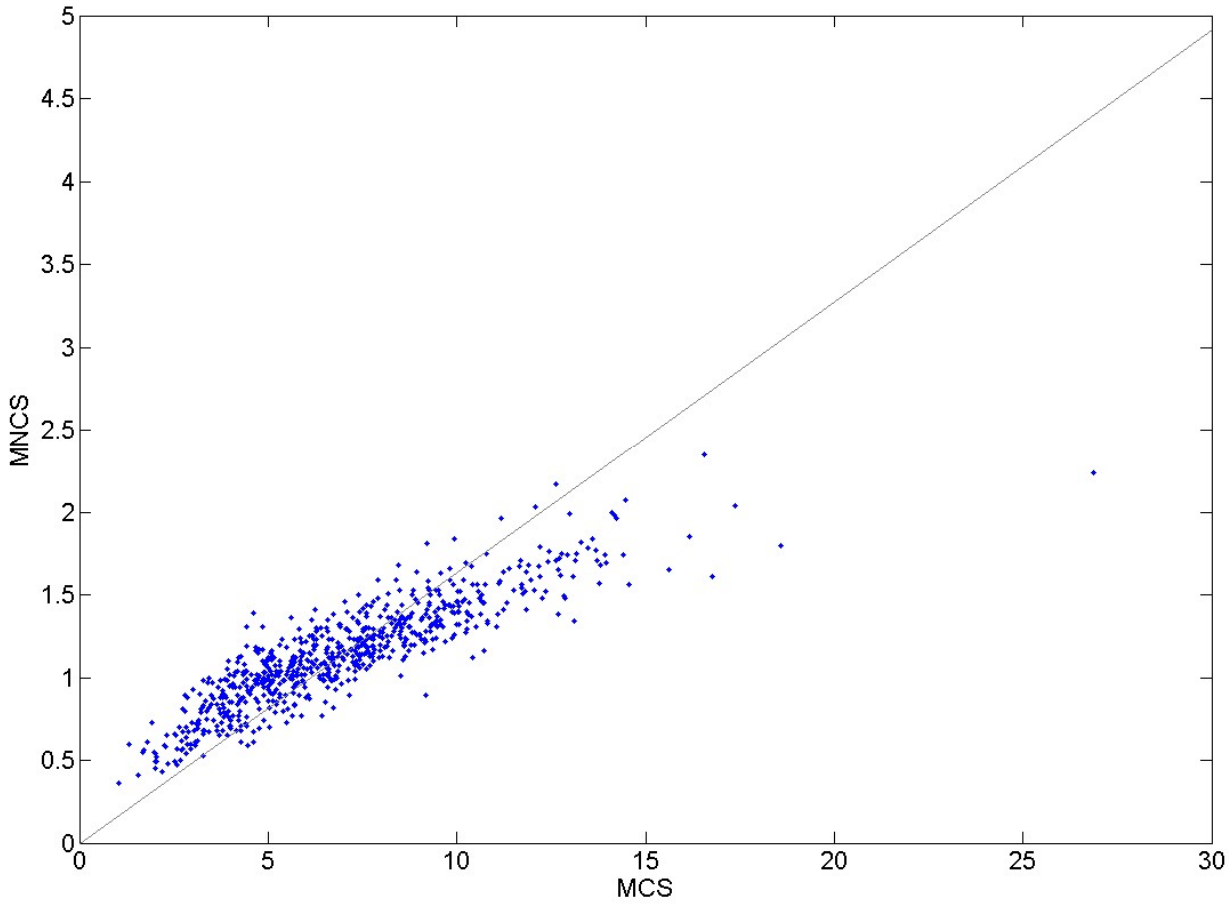
CWTS

Benchmark values

- **MNCS** of all publications in a database (e.g., Web of Science or Scopus) equals exactly **1**
- **PP(top 10%)** of all publications in a database equals exactly **10%** (or 0.10)
- These benchmark values are sometimes referred to as '*expected values*', '*world averages*', or '*database averages*'

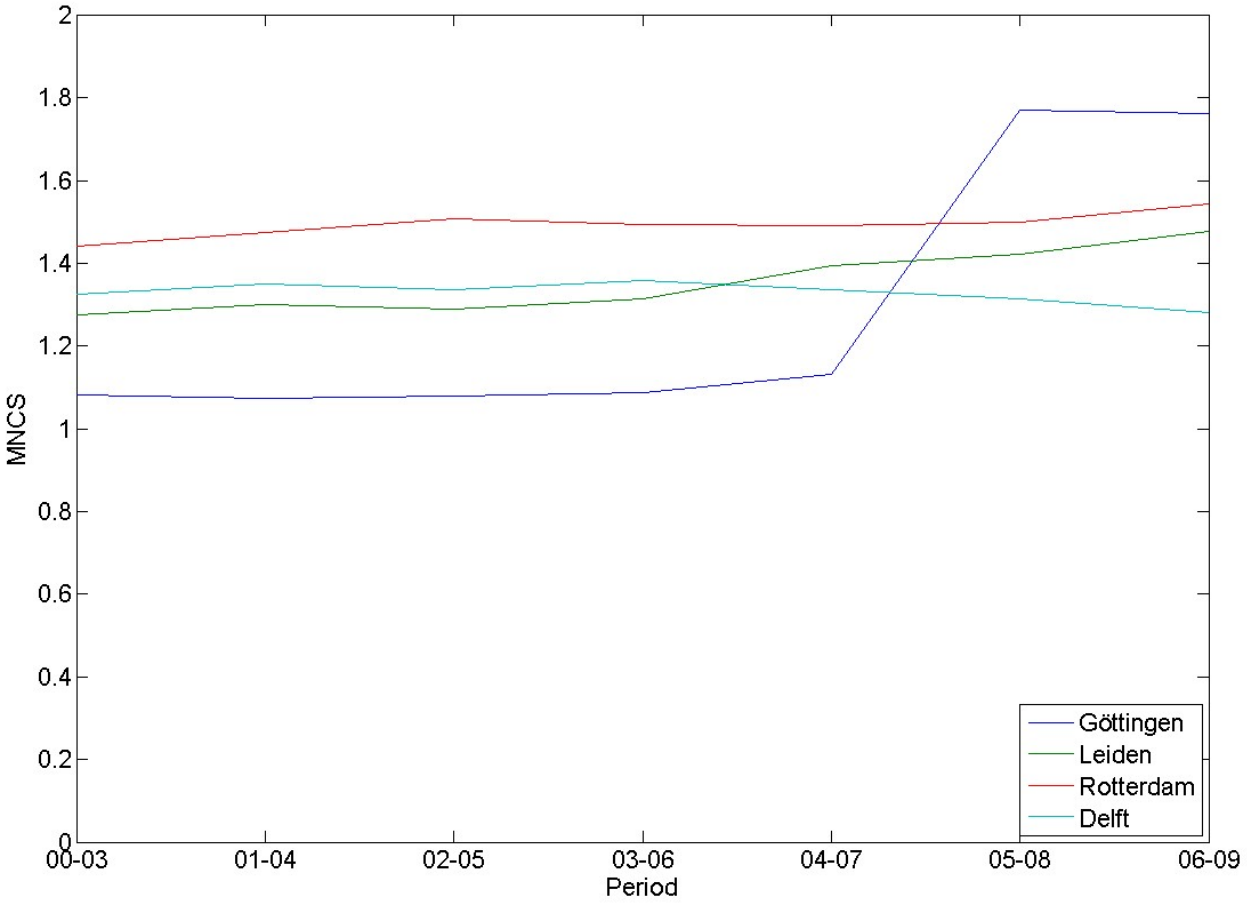
Assignment field normalization

MCS vs. MNCS

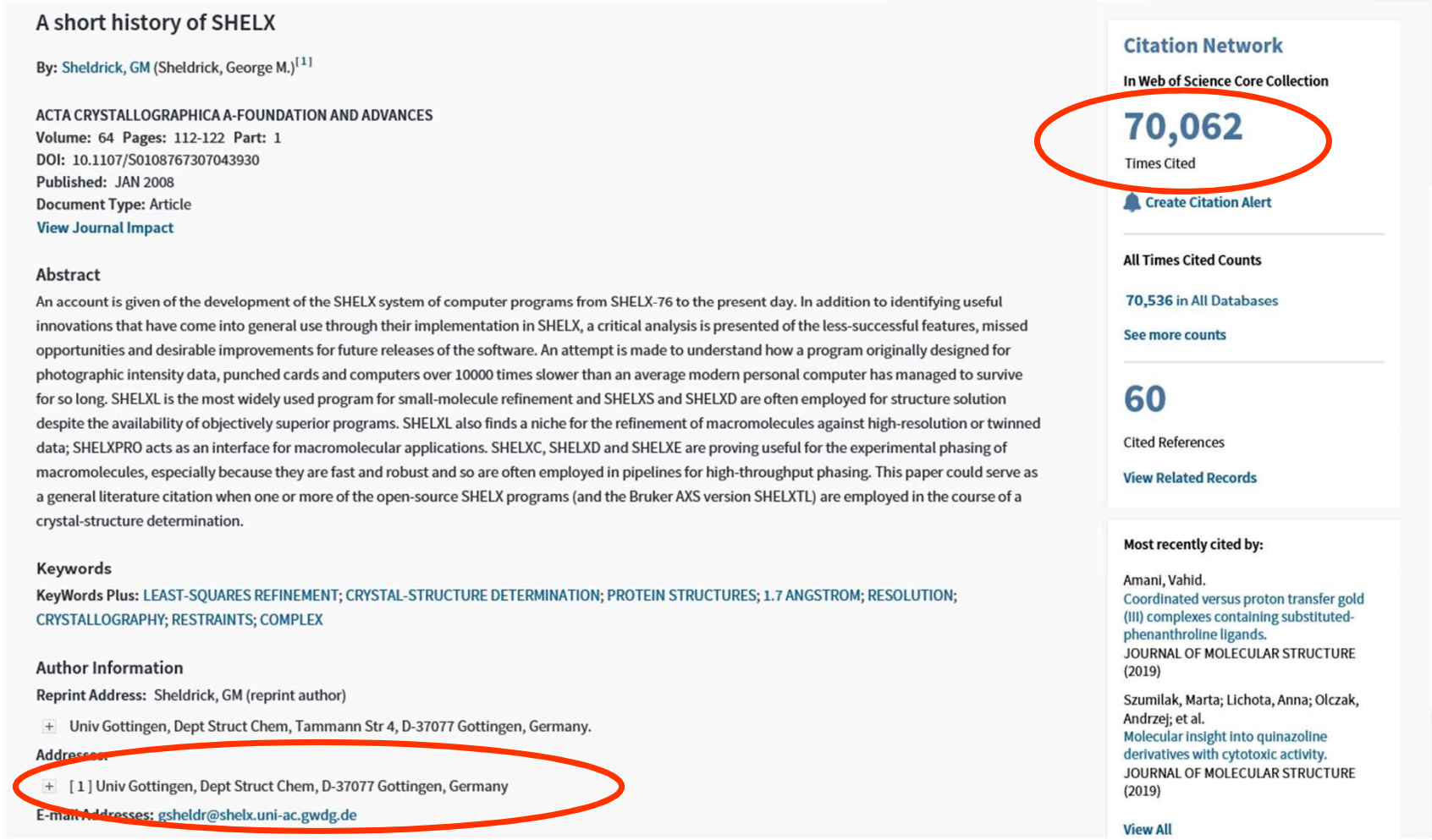


Averages and the effect of outliers

Sensitivity of indicators to ‘outliers’ (1)



Sensitivity of indicators to ‘outliers’ (2)



A short history of SHELX

By: Sheldrick, GM (Sheldrick, George M.)[1]

ACTA CRYSTALLOGRAPHICA A-FOUNDATION AND ADVANCES
Volume: 64 Pages: 112-122 Part: 1
DOI: 10.1107/S0108767307043930
Published: JAN 2008
Document Type: Article
View Journal Impact

Abstract

An account is given of the development of the SHELX system of computer programs from SHELX-76 to the present day. In addition to identifying useful innovations that have come into general use through their implementation in SHELX, a critical analysis is presented of the less-successful features, missed opportunities and desirable improvements for future releases of the software. An attempt is made to understand how a program originally designed for photographic intensity data, punched cards and computers over 10000 times slower than an average modern personal computer has managed to survive for so long. SHELXL is the most widely used program for small-molecule refinement and SHELXS and SHELXD are often employed for structure solution despite the availability of objectively superior programs. SHELXL also finds a niche for the refinement of macromolecules against high-resolution or twinned data; SHELXPRO acts as an interface for macromolecular applications. SHELXC, SHELXD and SHELXE are proving useful for the experimental phasing of macromolecules, especially because they are fast and robust and so are often employed in pipelines for high-throughput phasing. This paper could serve as a general literature citation when one or more of the open-source SHELX programs (and the Bruker AXS version SHELXTL) are employed in the course of a crystal-structure determination.

Keywords

KeyWords Plus: LEAST-SQUARES REFINEMENT; CRYSTAL-STRUCTURE DETERMINATION; PROTEIN STRUCTURES; 1.7 ANGSTROM; RESOLUTION; CRYSTALLOGRAPHY; RESTRAINTS; COMPLEX

Author Information

Reprint Address: Sheldrick, GM (reprint author)

+ Univ Gottingen, Dept Struct Chem, Tammann Str 4, D-37077 Gottingen, Germany.

Addresses:

+ [1] Univ Gottingen, Dept Struct Chem, D-37077 Gottingen, Germany

E-mail Addresses: gsheldr@shelx.uni-ac.gwdg.de

Citation Network

In Web of Science Core Collection

70,062

Times Cited

Create Citation Alert

All Times Cited Counts

70,536 in All Databases

See more counts

60

Cited References

View Related Records

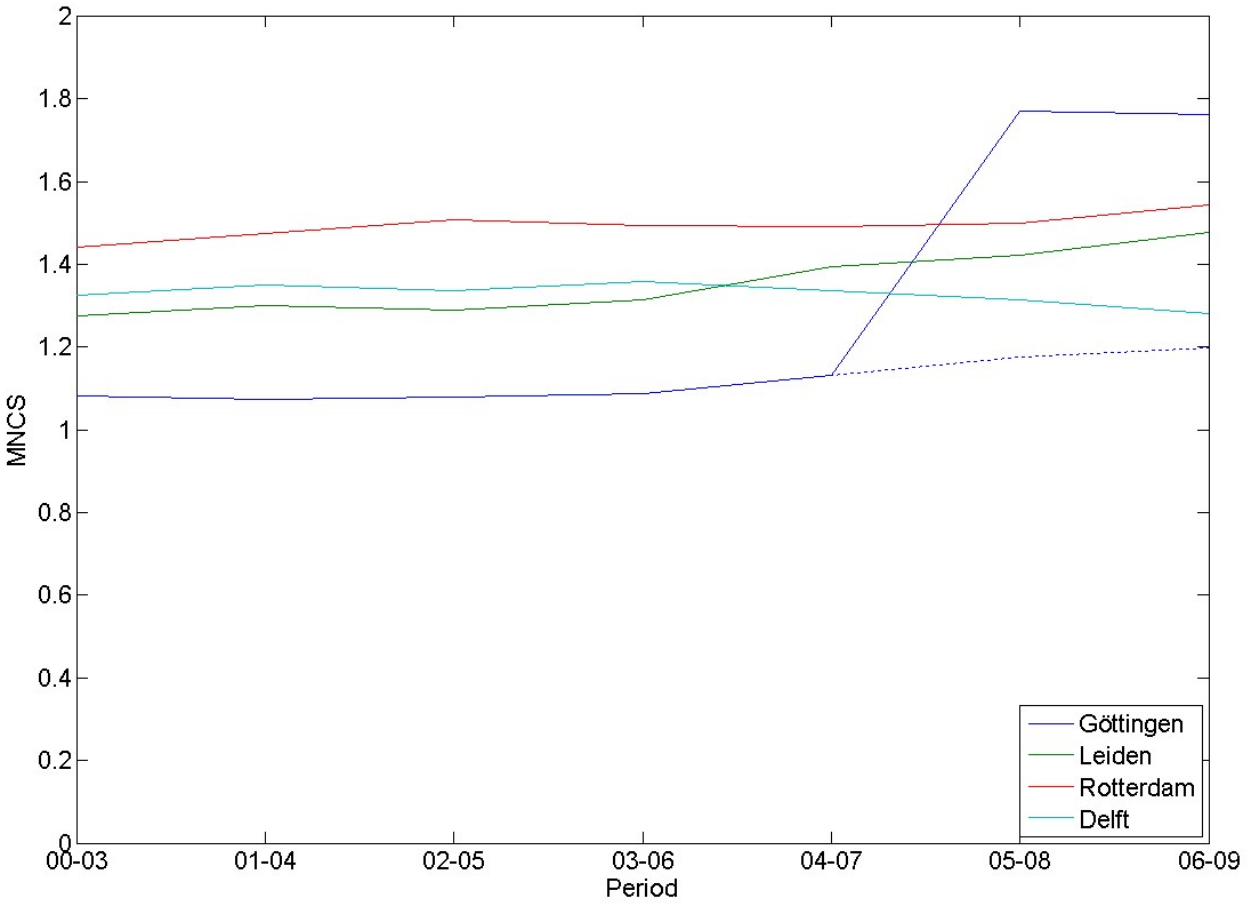
Most recently cited by:

Amani, Vahid.
Coordinated versus proton transfer gold (III) complexes containing substituted-phenanthroline ligands.
JOURNAL OF MOLECULAR STRUCTURE (2019)

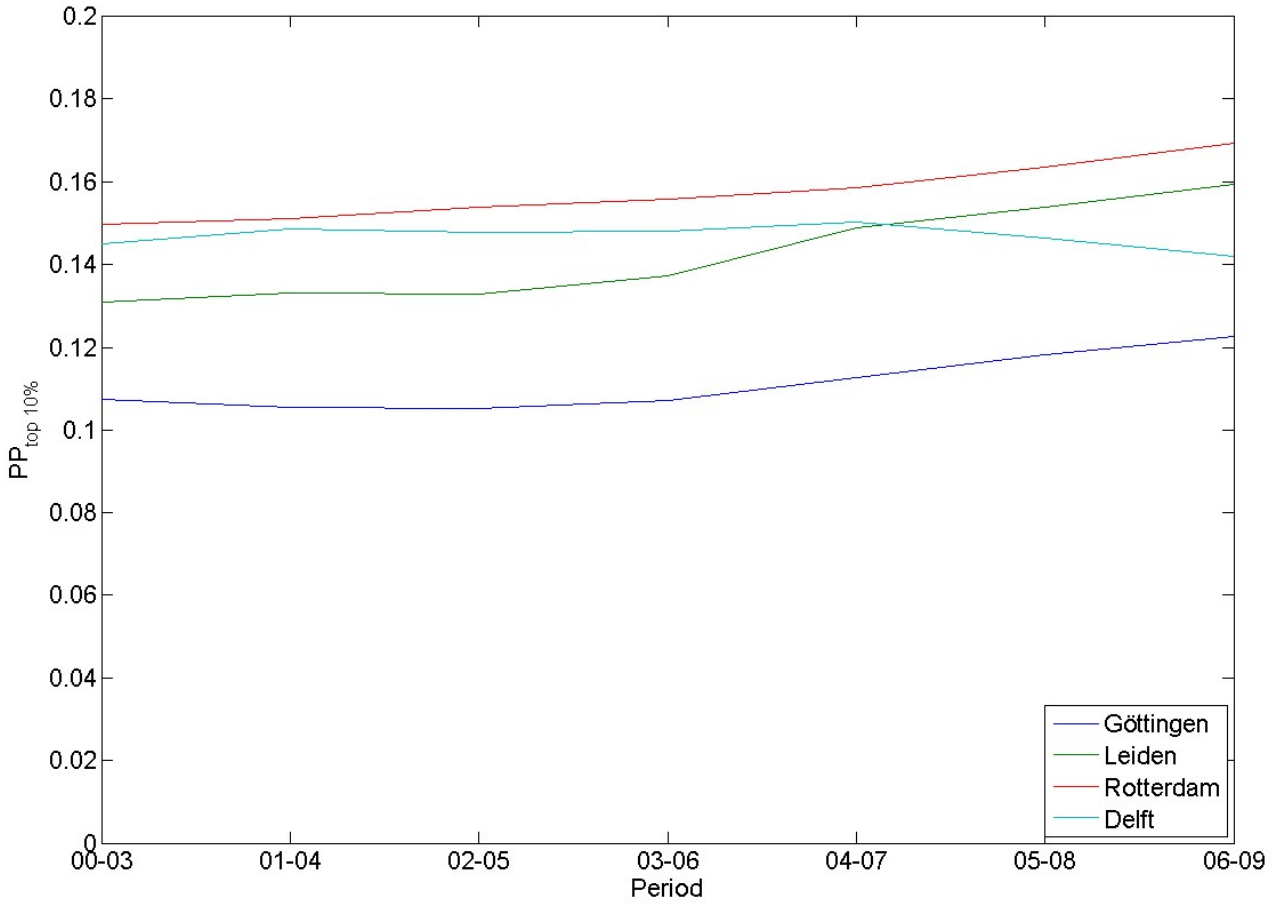
Szumilak, Marta; Lichota, Anna; Olczak, Andrzej; et al.
Molecular insight into quinazoline derivatives with cytotoxic activity.
JOURNAL OF MOLECULAR STRUCTURE (2019)

View All

Sensitivity of indicators to 'outliers' (3)

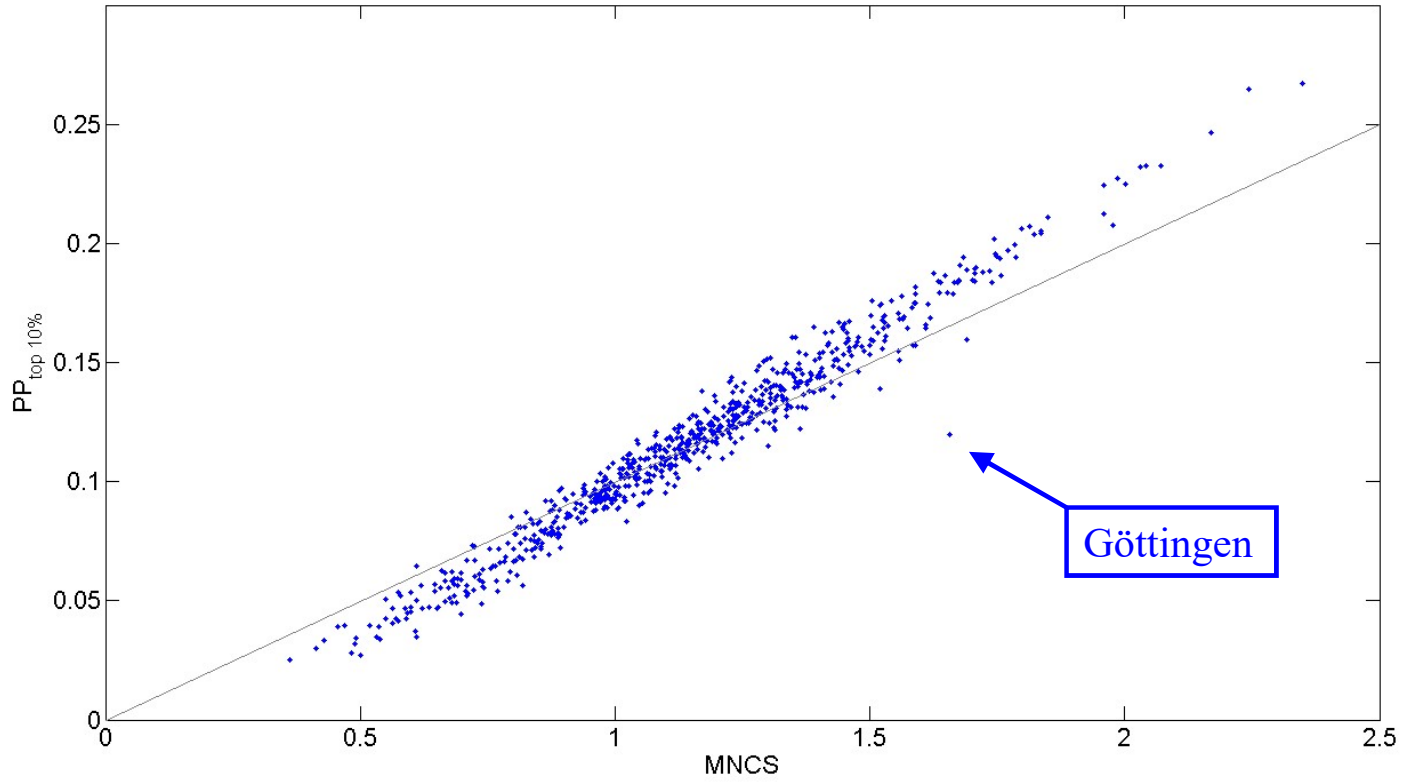


Sensitivity of indicators to 'outliers' (4)



Effect in ranking: Leiden Ranking

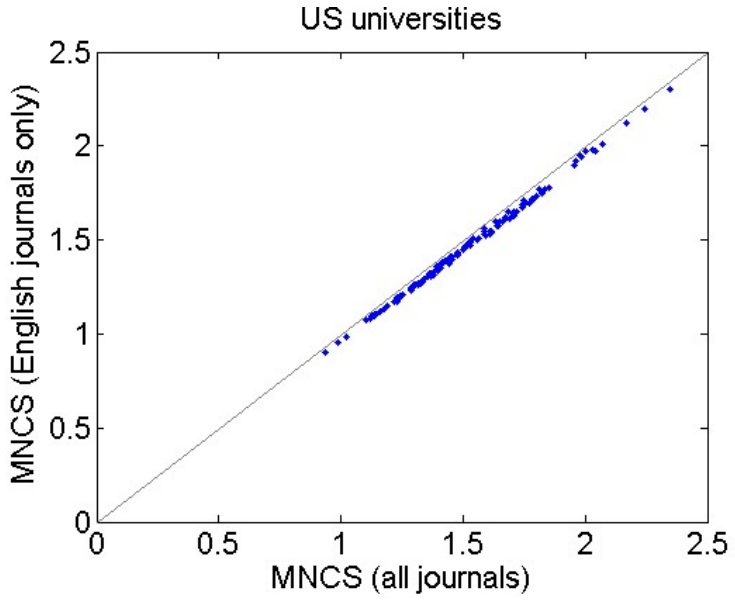
MNCS vs $PP_{top\ 10\%}$

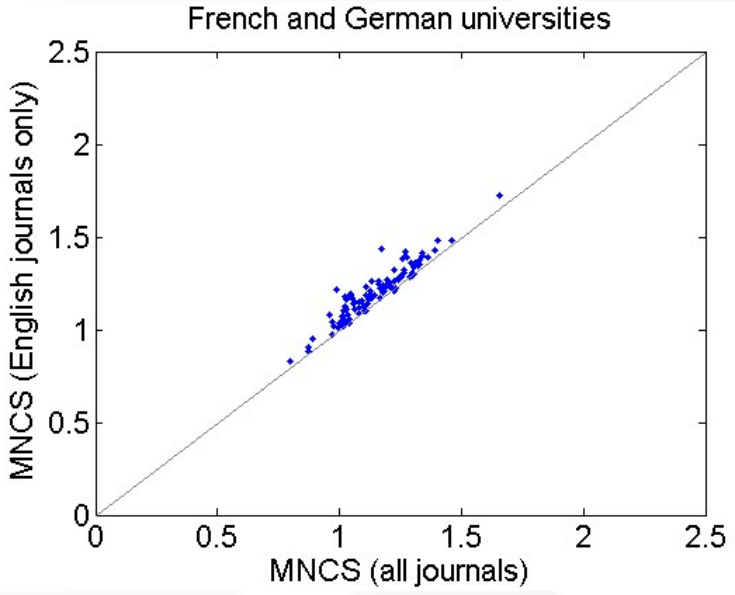


Other issues regarding measuring impact

Dependence on database coverage

Effect of excluding non-English journals from WoS





CWTS

Productivity

- Two research units of equal size (FTE or money)
- Unit 1:
  - 100 publications with 20 citations each
  - Mean citation score: (100 × 20) / 100 = 20
- Unit 2:
  - 100 publications with 20 citations each and 50 publications with 10 citations each
  - Mean citation score: (100 × 20 + 50 × 10) / (100 + 50) = 16.67
- Unit 2 has a lower mean citation score, even though this unit seems to have performed better.

CWTS

Impact indicators and tools

Similar indicators in InCites and SciVal

Indicator	CWTS	InCites	SciVal
Number of publications	P	Web of Science documents	Scholarly output
Total number of citations	TCS	Times cited	Citation count
Average number of citations per publication	MCS	Citation impact	Citations per publication
h-index	-	h-index	h-index
Total normalized number of citations	TNCS	-	-
Average normalized number of citations per publication	MNCS	Normalized citation impact	Field-weighted citation impact
Number of publications in top 10% of their field	P(top 10%)	# documents in top 10%	Field-weighted outputs in top 10%
Proportion of publications in top 10% of their field	PP(top 10%)	% documents in top 10%	% Field-weighted outputs in top 10%

CWTS

Intermezzo: h-index

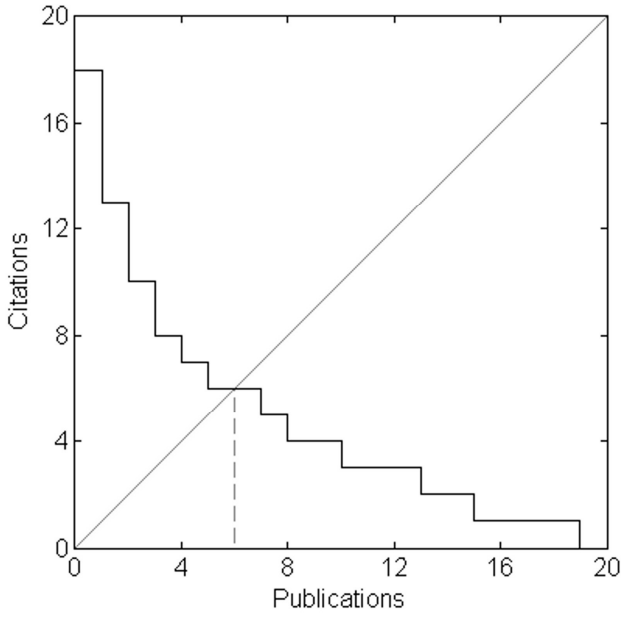
h-index

- Introduced in 2005 by physicist Jorge E. Hirsch
- Originally intended for the evaluation of individual researchers
- Received a lot of attention and quickly became popular
- Lots of *h*-index variants have been proposed, such as the *g*-index

CWTS

Definition of the *h*-index

A scientist has index *h* if *h* of his papers have at least *h* citations each and the other papers have at most *h* citations each



CWTS

What do you see as the advantages and disadvantages of the *h*-index?

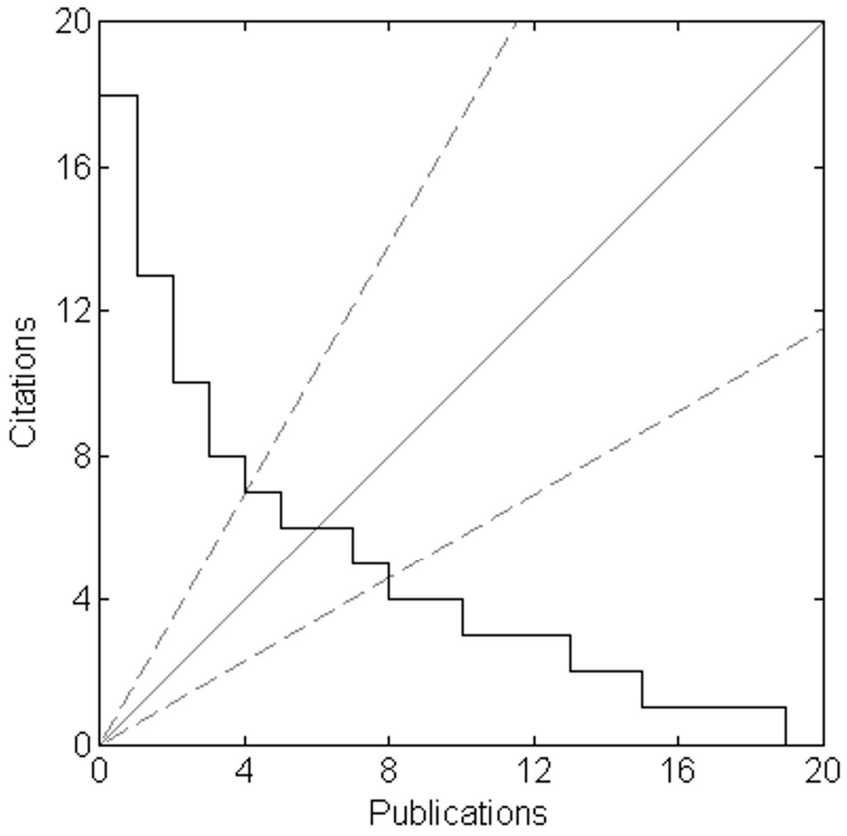
What are the main differences with the CWTS indicators?

Differences with CWTS indicators

- No document type restrictions;
- No customized publication and citation windows;
- No exclusion of author self citations;
- No normalization for field differences.

(Other) issues related to the h-index

Arbitrariness of the *h*-index



CWTS

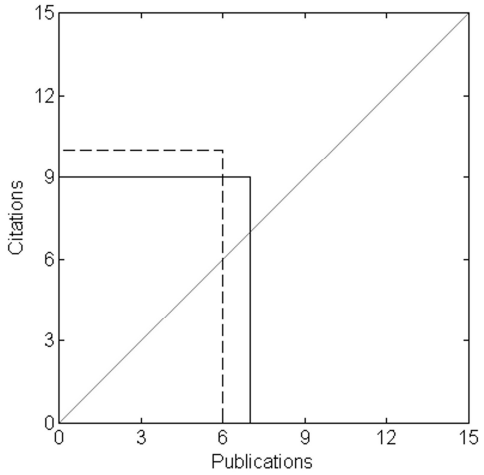
Consistency requirement

If two scientists achieve the same performance improvement, their ranking relative to each other should remain unchanged

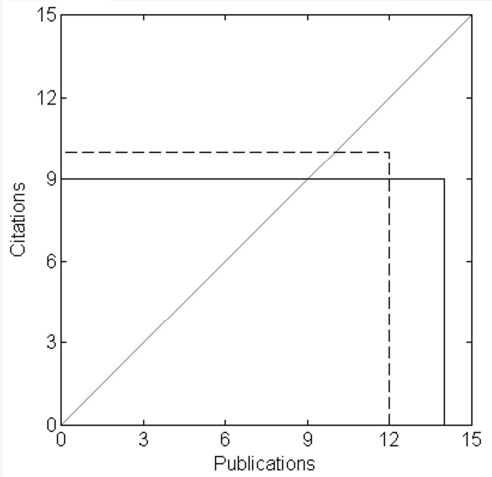
Inconsistency of the *h*-index

Researcher A

Researcher B

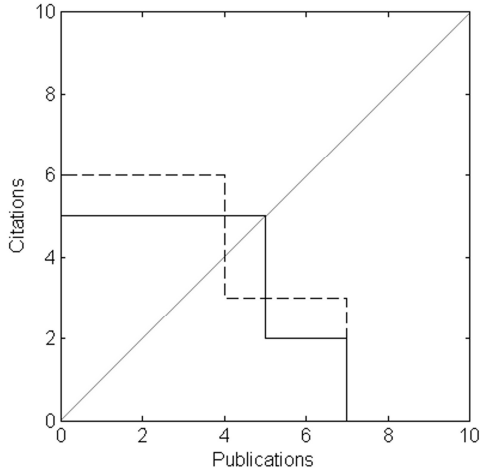


Double publications

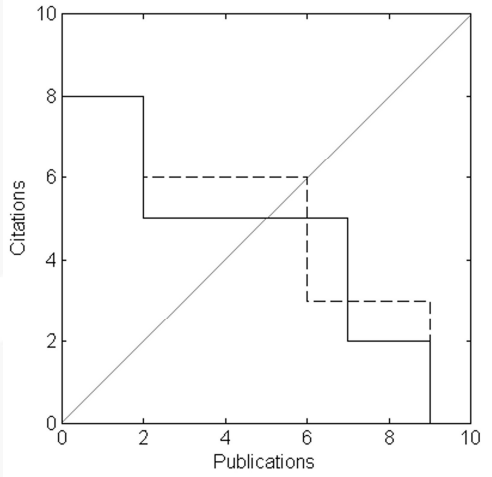


Researcher C

Researcher D



Double citations



CWTS

Thank you for your attention!