Neue Entwicklungen bei der Anwendung der Bibliometrie in der Forschungsevaluation

Lutz Bornmann

Bibliometrics-based heuristics

- Definition of bibliometrics: use of publication and citation data to measure science
- The European Commission on Research and Innovation has defined bibliometrics as “a statistical or mathematical method for counting the number of academic publications, citations and authorship” (Directorate-General for Research, 2010)
- Definition is far from being satisfactory: it focusses on the used data
- Interpretation of bibliometrics in the fast-and-frugal heuristics approach
- Heuristics are decision strategies that use part of the available information and ignore the rest
- Bibliometrics-based heuristics are adaptive judgement strategies that ignore information about some performance aspects (e.g., amount of third-party funds raised or assessments of single publications by experts), thereby allowing quick (and robust) decisions in research evaluation

Bornmann, L., & Marewski, J. N. (2019). Heuristics as conceptual lens for understanding and studying the usage of bibliometrics in research evaluation. *Scientometrics, 120*(2), 419–459.

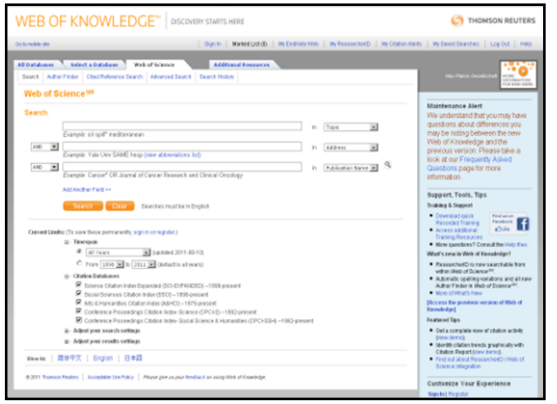
Use of bibliometric indicators in national research assessment exercises

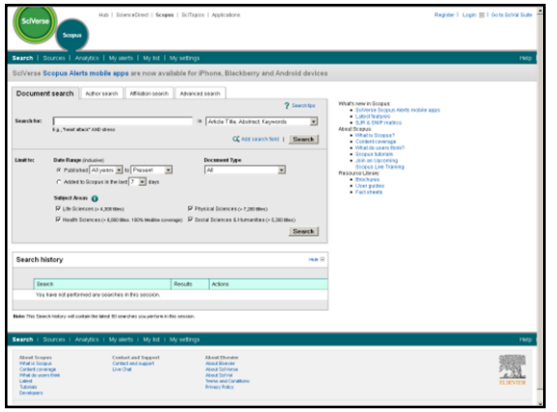
		Belgium /FL	Czech Republic	Denmark	Finland	Italy (VQR)	New Zealand	Norway	Sweden	UK
Output indicators	Academic outputs	√	√	√	√	√	√	√	√	√
	Non-academic outputs		√			√			√	√
	Innovation-related outputs (IPR)		√			√				√
External funding indicators	Competitive funding / national			√	√	√	√	√	√	√
	Competitive funding / international			√	√	√	√	√	√	√
	Contract research funding			√		√	√		√	√
	Non-competitive funding			√		√	√			√
Outcomes/ impact indicators	Academic impacts (citations)	√	√			√			√	
	Socio-economic outcomes/impacts (e.g. spin- offs)					√				√

Department for Business; Energy; Industrial Strategy. (2016). *Building on Success and Learning from Experience An Independent Review of the Research Excellence Framework*. London, UK: Department for Business, Energy & Industrial Strategy.

Databases for citation analyses

Database	Papers
Web of Science – Core Collection (Clarivate Analytics)	1900
Scopus (Elsevier)	1788
Chemical Abstracts Service (CAS) Database	1898
INSPEC database for Physics, Electronics & Computing	1897
Google Scholar Citations	???
Microsoft Academic Graph (closure later this year)	???
Dimensions	???





Explanations of databases:
QSS Volume 1 Issue 1

Necessity to have high-quality data for research evaluation

NEW METHOD FOR HIGH-ACCURACY DETERMINATION OF THE FINE-STRUCTURE CONSTANT BASED ON QUANTIZED HALL RESISTANCE

By: VONKLITZING, K (VONKLITZING, K); DORDA, G (DORDA, G); PEPPER, M (PEPPER, M)

PHYSICAL REVIEW LETTERS
Volume: 45 Issue: 6 Pages: 494-497
Published: 1980
Document Type: Article
View Journal Impact

Author Information
Reprint Address: VONKLITZING, K (reprint author)
+ UNIV WURZBURG,INST PHYS,D-8700 WURZBURG,FED REP GER.
Addresses:
+ [1] SIEMENS AG,FORSCH LAB,D-8000 MUNICH 80,FED REP GER
+ [2] UNIV CAMBRIDGE,CAVENDISH LAB,CAMBRIDGE CB3 0HE,ENGLAND

VOLUME 45, NUMBER 6 PHYSICAL REVIEW LETTERS 11 AUGUST 1980

New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance

K. v. Klitzing
Physikalisches Institut der Universität Würzburg, D-8700 Würzburg, Federal Republic of Germany, and Hochfeld-Magnetlabor des Max-Planck-Instituts für Festkörperforschung, F-38042 Grenoble, France

- Paper which leads to the Nobel Prize in physics for Klaus von Klitzing
- Research has been done in the Max Planck Society (outpost of the Max Planck Institute for Solid State Research in Grenoble)
- Klaus von Klitzing was affiliated with the Universität Würzburg

Bibliometric indicators

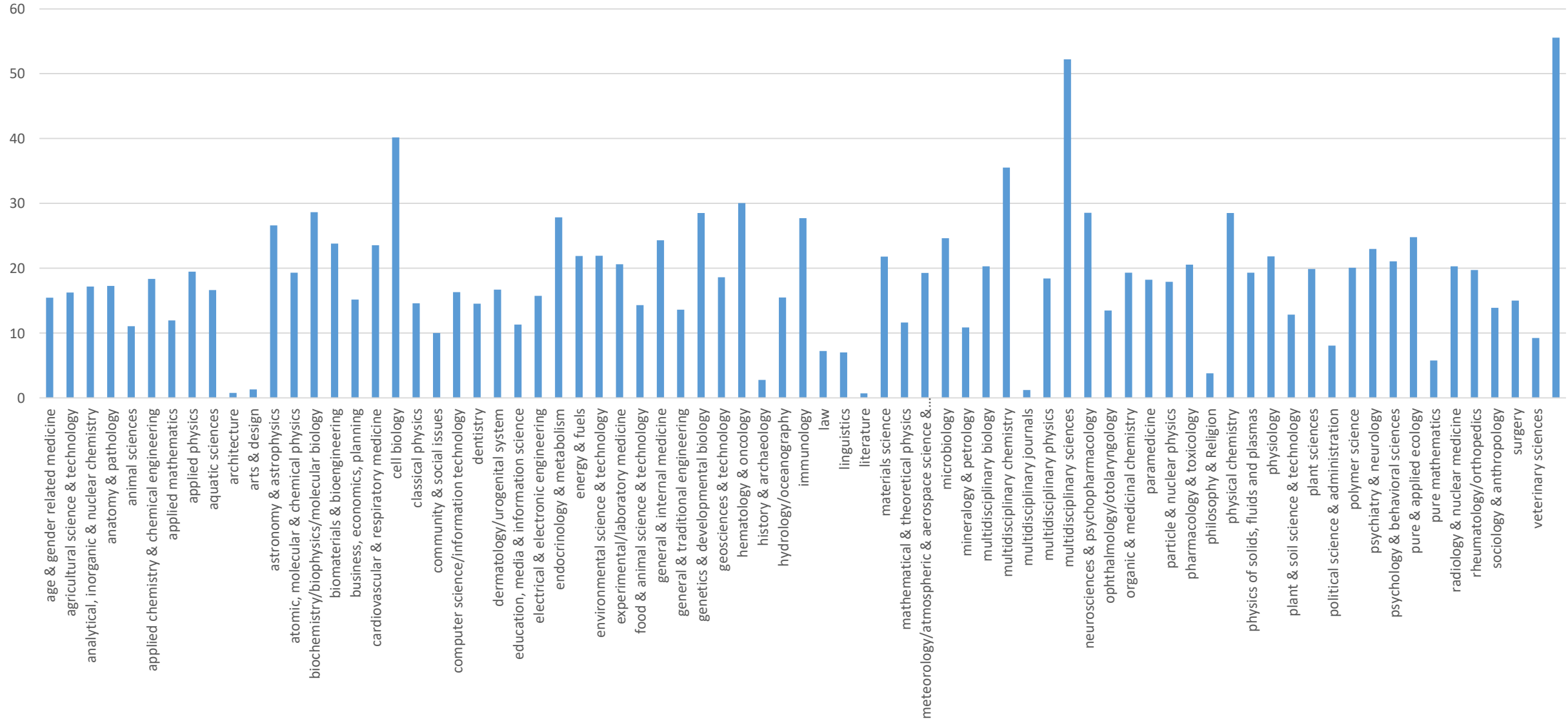
Basic indicators	Number of publications Number of citations Citations per publications (citation rate) Number of not-cited publications	Researcher Institution Country
h index-based indicators	*h* index and approximately 50 variants *m* quotient	Researcher
Normalized indicators	Field- and time-normalized indicators Cited-side and citing-side normalization	Researcher Institution Country
Technology-indicators	Number of publications cited in patents Number of patents cited in publications	Institution Country
Social indicators	Co-authorship networks	Researcher Institution Country
Journal indicators	Journal Impact Factor CiteScore	Journals
Mapping indicators	Co-citations	Institution Country

Citizen bibliometrics vs. professional bibliometrics

Citizen bibliometrics	Professional bibliometrics
• Do-it-yourself bibliometrics by researchers and research managers	• Bibliometric analyses supported by professional bibliometricians and specialized bibliometric software tools
• Journal impact factor, h-index	• Field-normalized indicators
• Google Scholar	• Web of Science, Scopus
• Mainly with small datasets (e.g. young researchers)	• Mainly with large datasets (e.g. institutions)

Leydesdorff, L., Wouters, P., & Bornmann, L. (2016). Professional and citizen bibliometrics: complementarities and ambivalences in the development and use of indicators—a state-of-the-art report. *Scientometrics*, 109(3), 2129-2150.

Why do we need normalized indicators? Citation counts are field-dependent (papers published in 2010)



Normalized impact (NI)

- Normalized Impact (NI) = Ratio of observed citations (WoS: “times cited”) to expected citation rate
- The expected citation rate is the mean impact of the following publications:

  - published in a journal of the same subject category

  - published in the same year

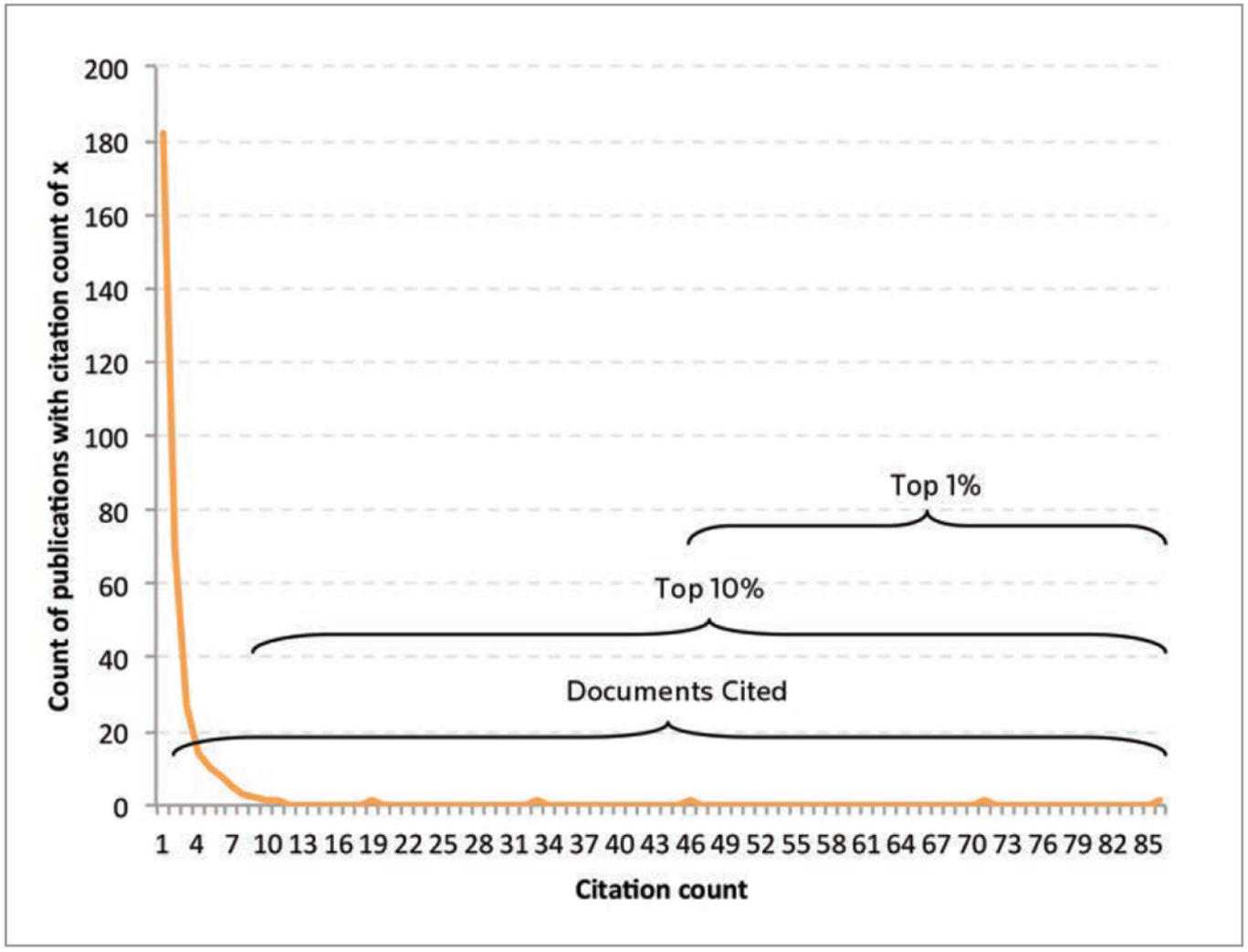
    - Suppose a publication from 2010 in an oncology journal
    - The publication has 45 citations until the end of 2015
    - On average, publications from 2010 in oncology journals have 15 citations in the same time period
    - Normalized citation score of the publication is 45 / 15 = 3

- NI values:

  NI = 1.0 : Average impact

  NI = 1.2 : 20% above average

Problem for calculation of NI: skewed distribution of citation data



A small number of highly cited papers and many papers with relatively few or no citations (source: Thomson Reuters. (2015). *InCites Indicators Handbook*. Philadelphia, PA, USA: Thomson Reuters)

The use of percentiles as an alternative to the NI

- Problem of the NI: A few highly-cited papers significantly influence the result
- Leiden Ranking 2013: University of Göttingen on ranking position two, because of only one highly-cited paper (*Göttingen effect*)
- Solution: Calculation of percentile ranks (PR)
- Definition: PR *x* is defined as the citation count at or below which *x*% (e.g., 90%) of the papers in the subject category falls
- Procedure: Calculate the cumulative percentage of papers with certain citation counts (beginning with low impact papers or papers with zero citations)
- The use of PRs avoids the problem with outliers
- PRs can be used very flexible (e.g., by calculating the top-10%)
- The use of percentiles is recommended in the Leiden Manifesto

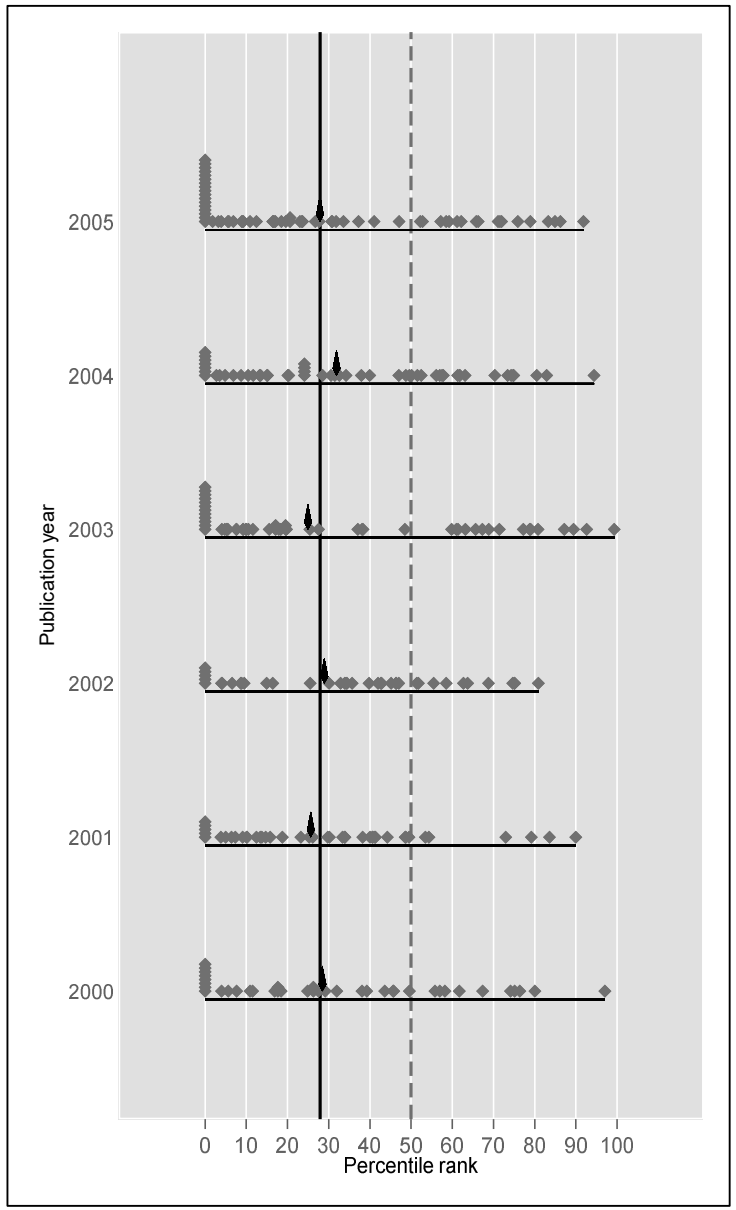
Citation count	Number of papers	Percent	Cumulative percentage (PR)
0	4	19.05	19.05
1	3	14.29	33.33
2	1	4.76	38.10
3	1	4.76	42.86
7	4	19.05	61.90
8	2	9.52	71.43
9	1	4.76	76.19
10	1	4.76	80.95
13	2	9.52	90.48
20	2	9.52	100.00
Total	21	100.00	

Overview of the scientific performance of three researchers

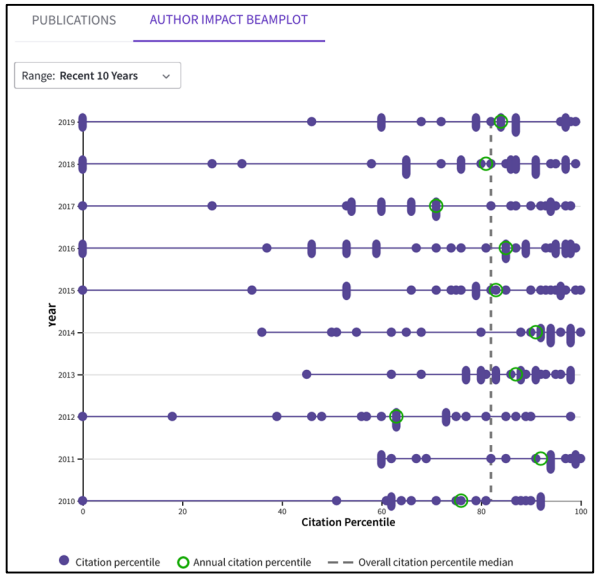
Impact	Researcher 1	Researcher 2	Researcher 3
Total citations	15,192	3,796	7,828
Number of citations per publication (arithmetic average)	83	52	89
Proportion of self-citations in total citations	3.4%	6%	5.8%
Average percentile (weighted median)	15.9	6.2	8.3
$P_{top\ 10\%}$	70	31	48
$PP_{top\ 10\%}$	39.3%	52.5%	57.8%
$P_{top\ 10\%}$ quotient	2.2	2.8	1.6
Q1 indicator	25%	46%	33%

Q1 indicator: Proportion of papers published in a journal which belongs to the 25% journals with the highest Journal Impact Factor in its field and publication year

Beamplots: measuring the performance of single researchers



- Grey diamonds: impact of single papers
- Black horizontal lines: impact range of papers in one year
- Black triangle: weighted median impact in one year
- Grey dotted line: expected average impact
- Black vertical line: weighted median impact over all years



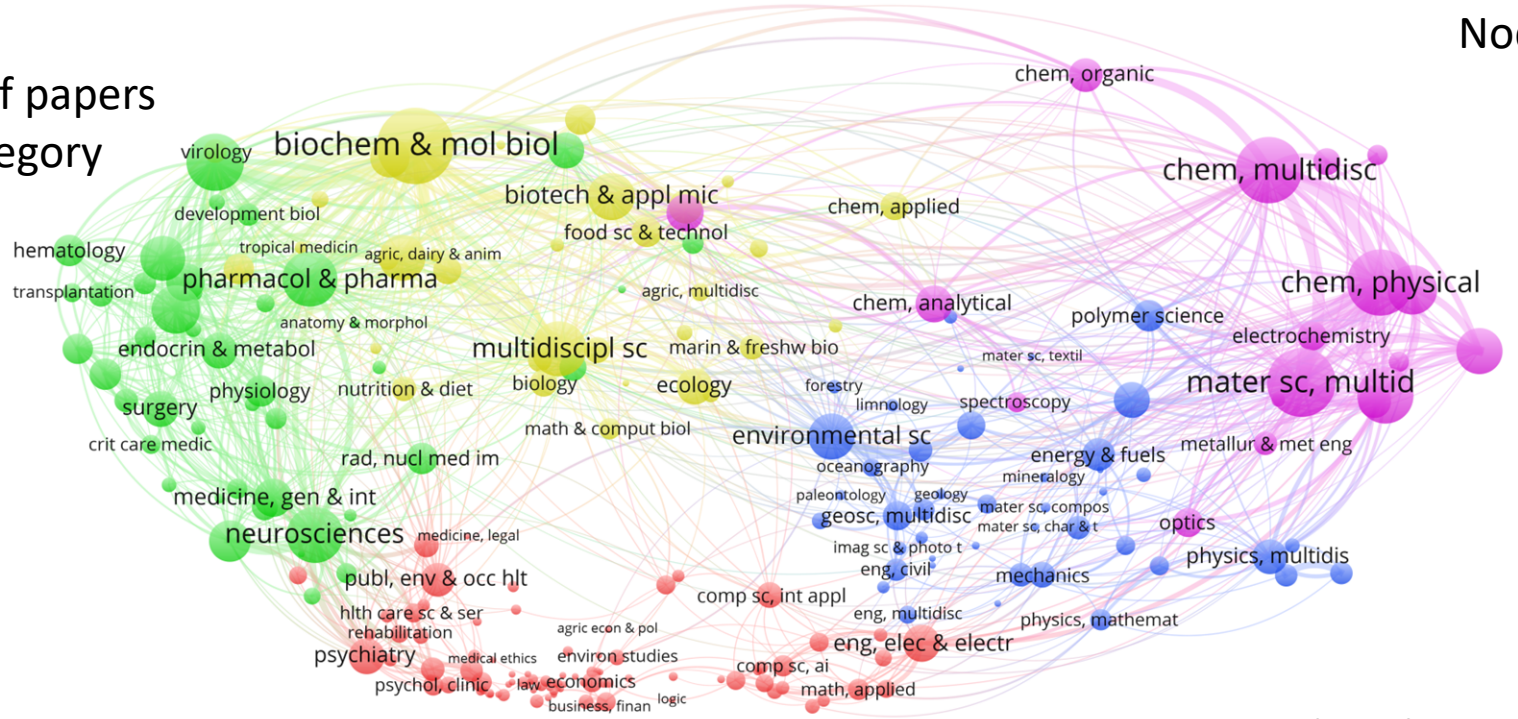
Bornmann, L., & Williams, R. (2020). An evaluation of percentile measures of citation impact, and a proposal for making them better. *Scientometrics, 124*, 1457–1478

Szomszor, M., & Pendlebury, D. A. (2021). *Interpreting the citation performance of individual researchers with beamplots*. Philadelphia, PA, USA: Clarivate Analytics.

Basic map: direct citation-relations of single subject categories (based on WoS data)

Node size: number of papers in the subject category

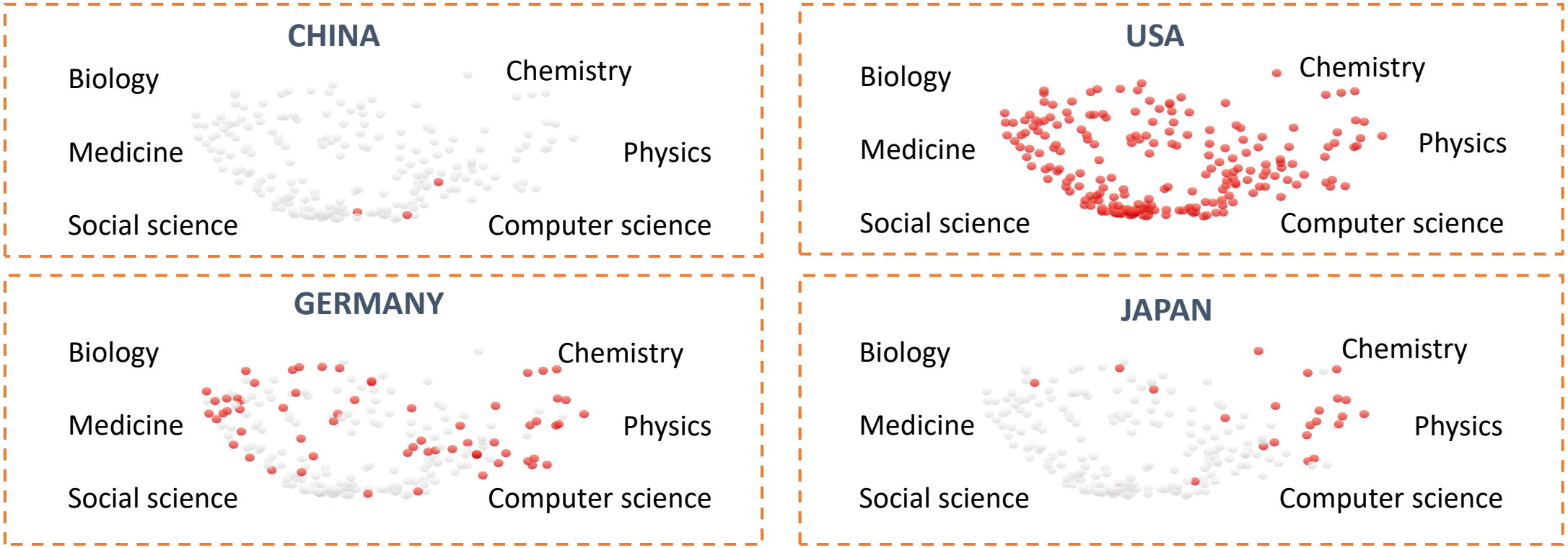
Node position: Many citation relations lead to closely positioned nodes



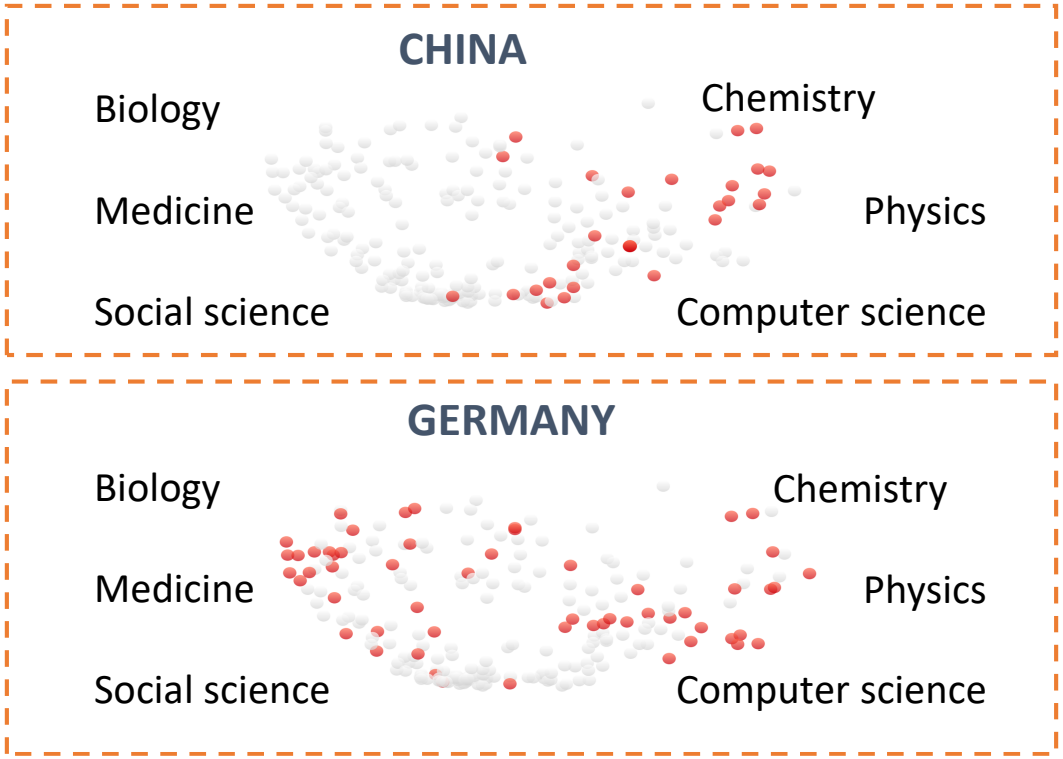
Database: Articles and reviews from 2003 to 2013 and their cited references in the same period

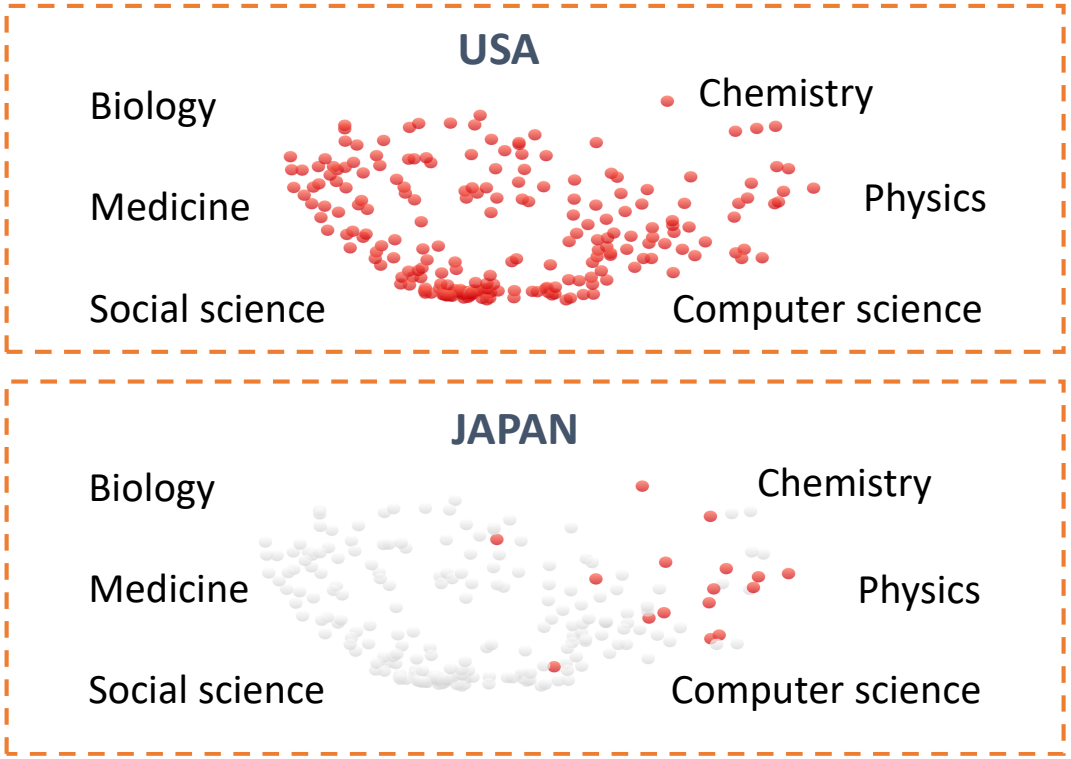
Node colour: Cluster algorithm assigns subject categories to the same colour, if they are frequently co-cited

Subject categories with more than 10% papers. Papers are counted which belong to the 10% most frequently cited papers in their subject category (2000-2002)

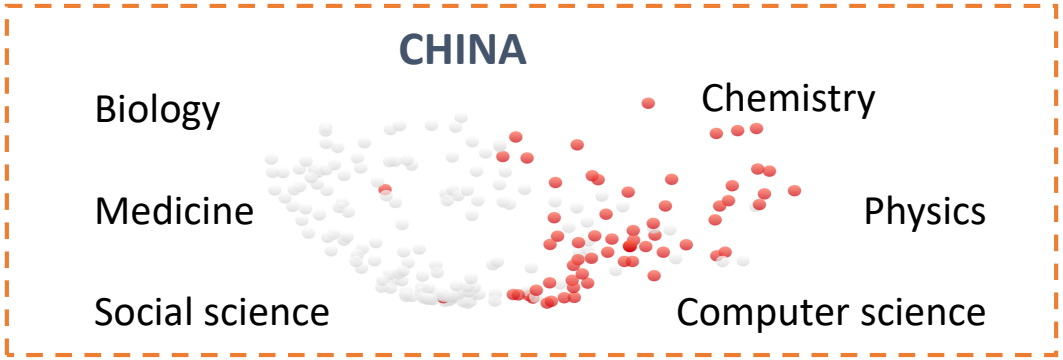


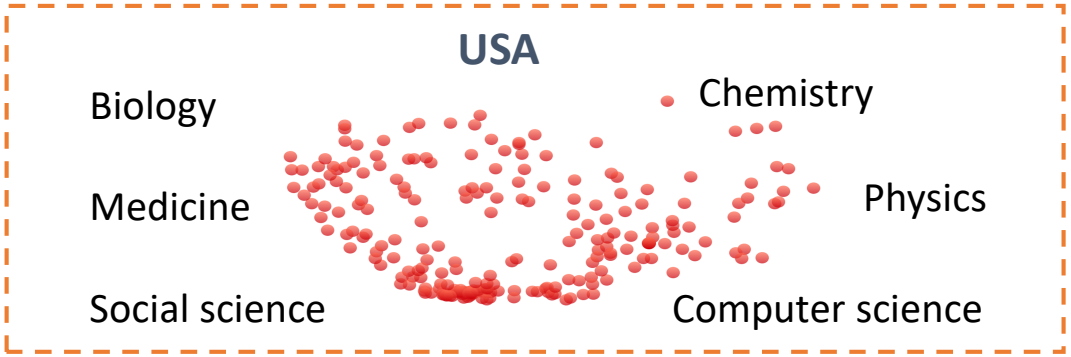
Subject categories with more than 10% papers. Papers are counted which belong to the 10% most frequently cited papers in their subject category (2003-2005)

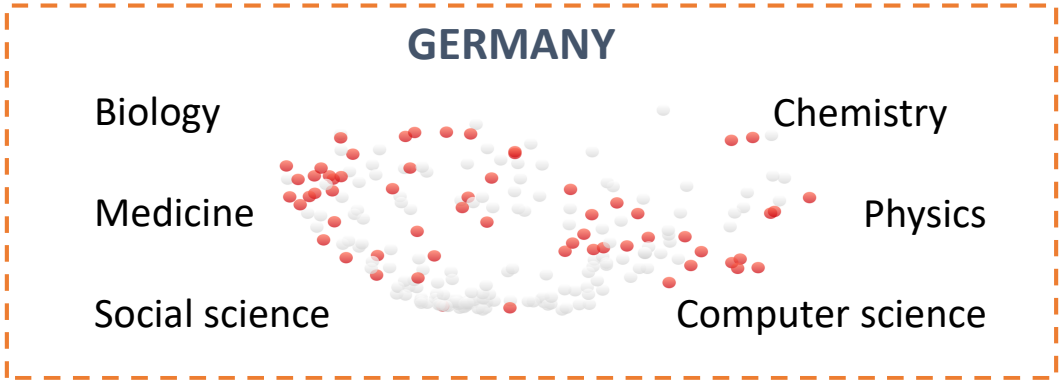


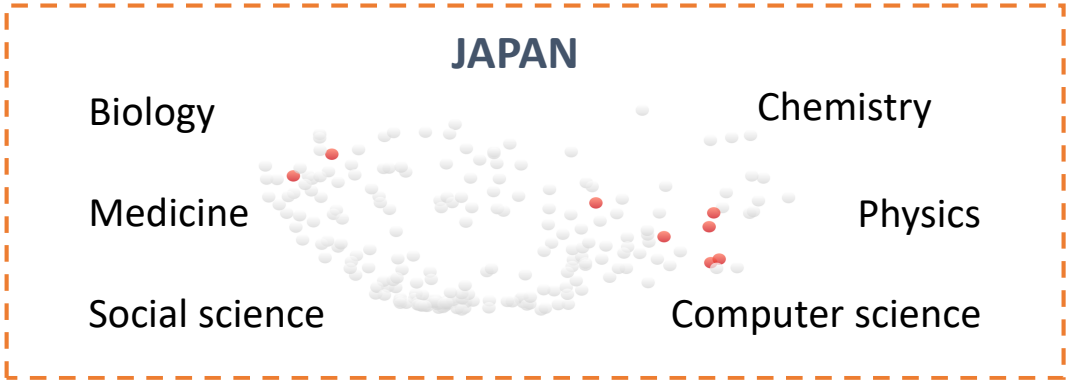


Subject categories with more than 10% papers. Papers are counted which belong to the 10% most frequently cited papers in their subject category (2006-2008)

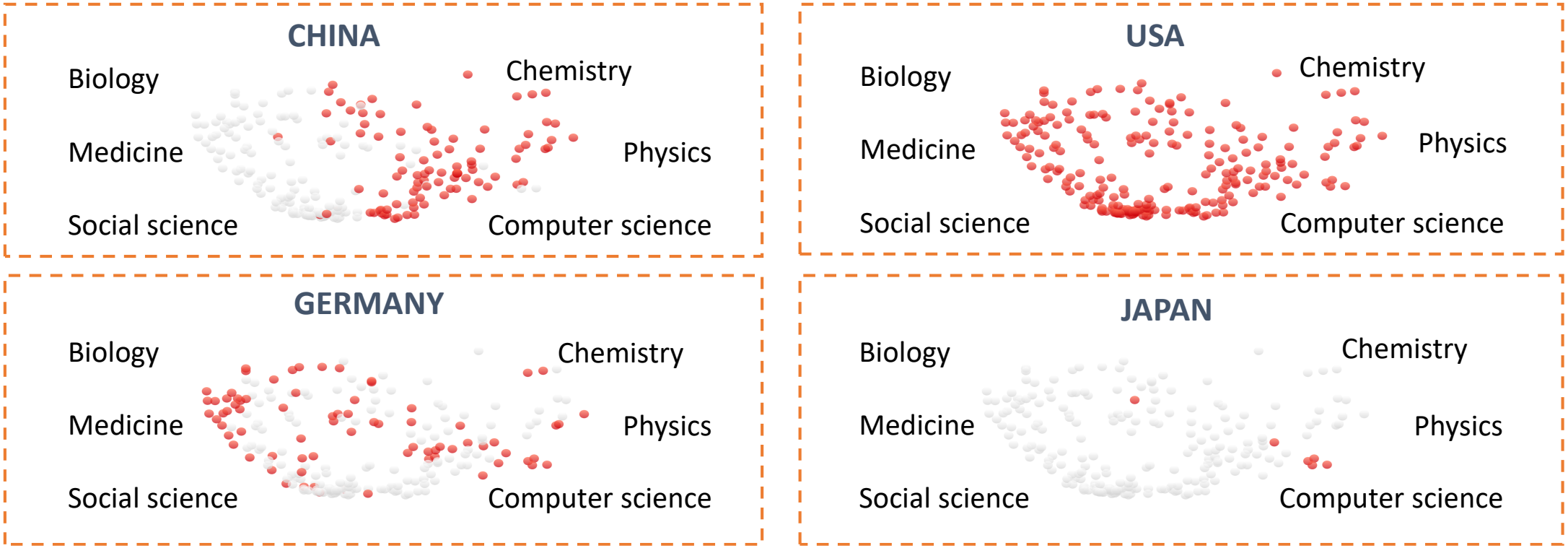




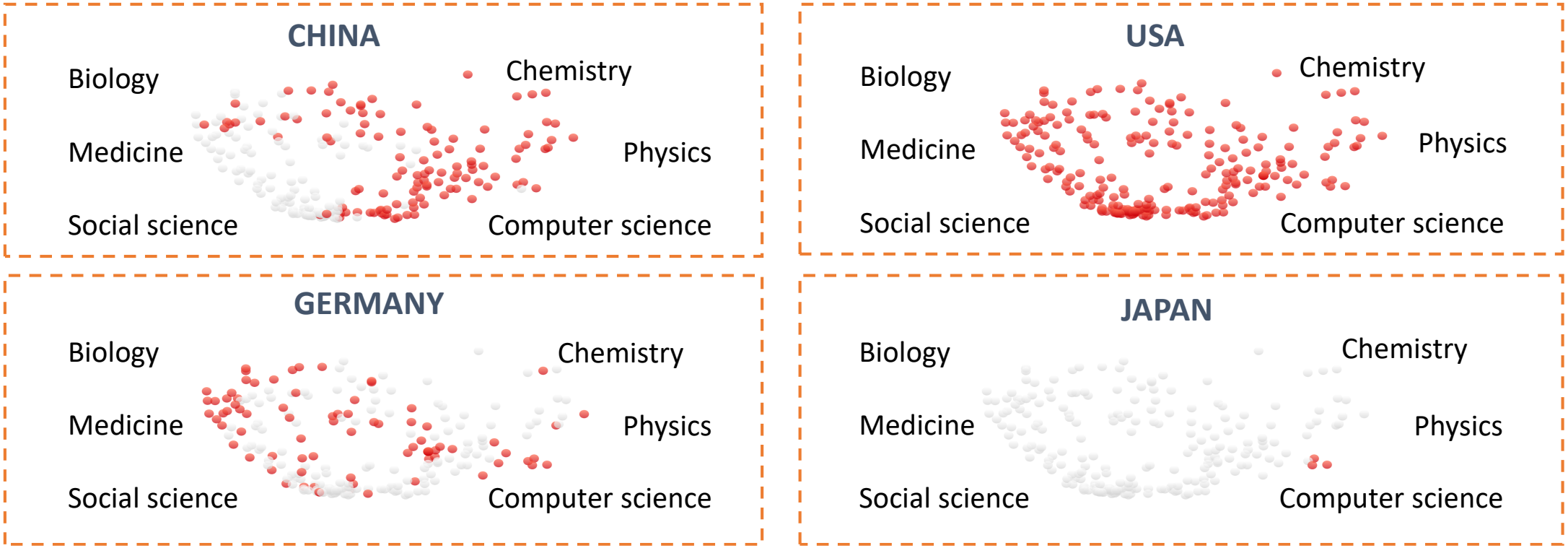




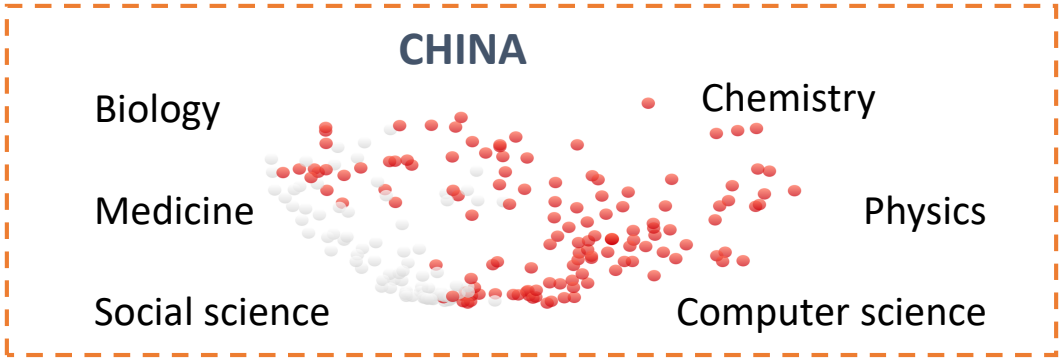
Subject categories with more than 10% papers. Papers are counted which belong to the 10% most frequently cited papers in their subject category (2009-2011)

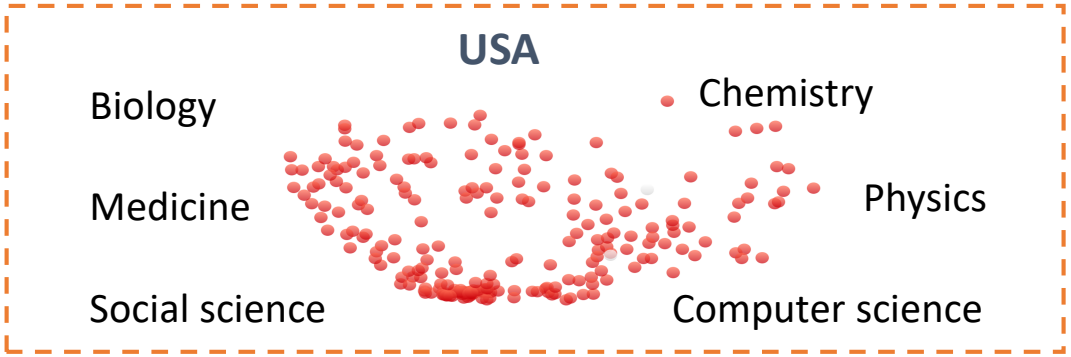


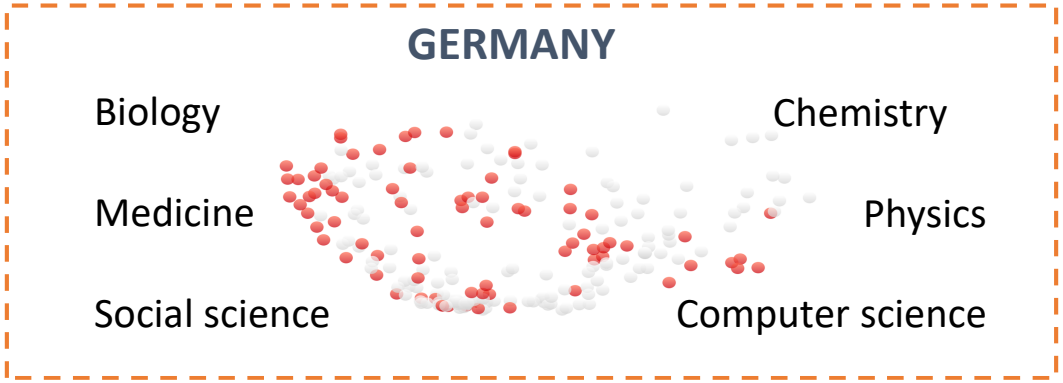
Subject categories with more than 10% papers. Papers are counted which belong to the 10% most frequently cited papers in their subject category (2012-2014)

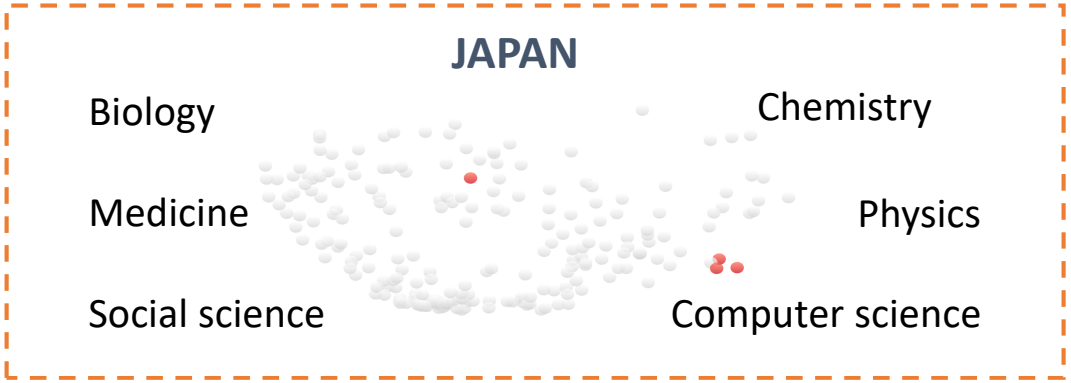


Subject categories with more than 10% papers. Papers are counted which belong to the 10% most frequently cited papers in their subject category (2015-2017)

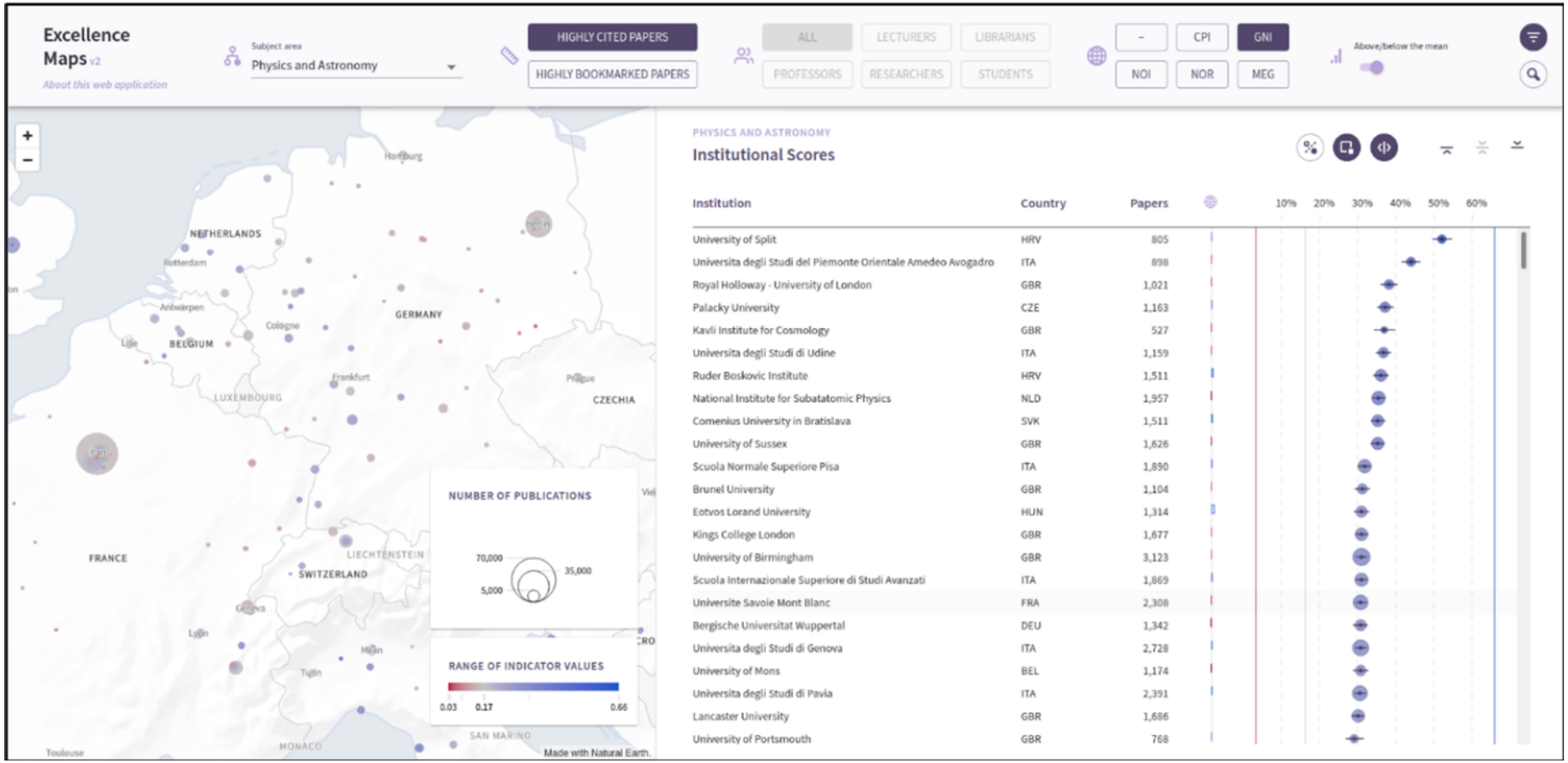








Mapping scientific excellence: application which visualizes research excellence worldwide in several subject areas (www.excellencemapping.net)



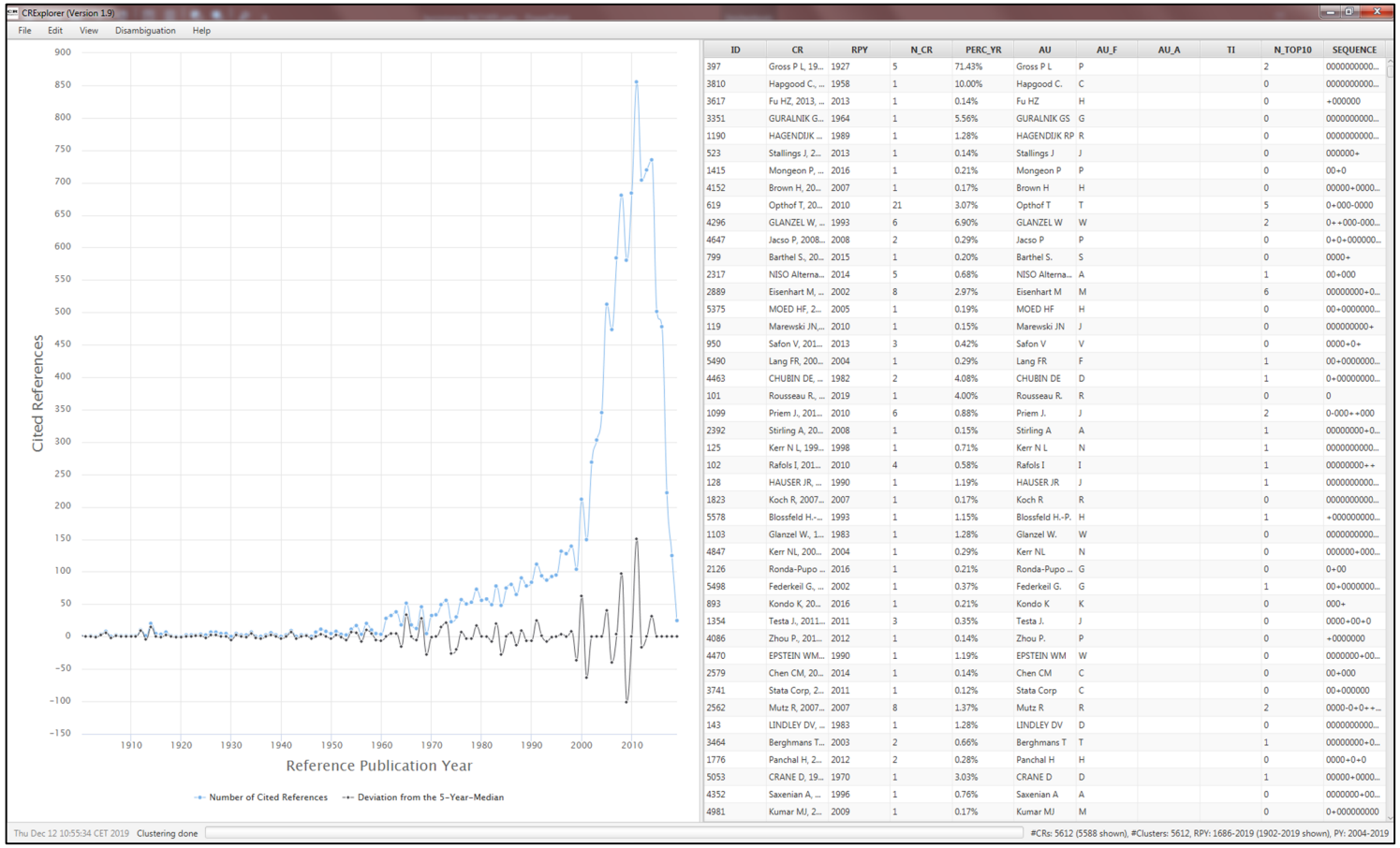
Lutz Bornmann, Rüdiger Mutz, Robin Haunschild, Felix de Moya-Anegon, Mirko de Almeida Madeira Clemente, Moritz Stefaner (2021). Mapping the impact of papers on various status groups: A new excellence mapping tool based on citation and reader scores. https://arxiv.org/abs/2103.10225

Excellence networks: application which visualizes how successful institutions collaborate (www.excellence-networks.net)

Bornmann, L., Stefaner, M., de Moya Anegón, F., & Mutz, R. (2016). Excellence networks in science: A Web-based application based on Bayesian multilevel logistic regression (BMLR) for the identification of institutions collaborating successfully. *Journal of Informetrics, 10*(1), 312-327.

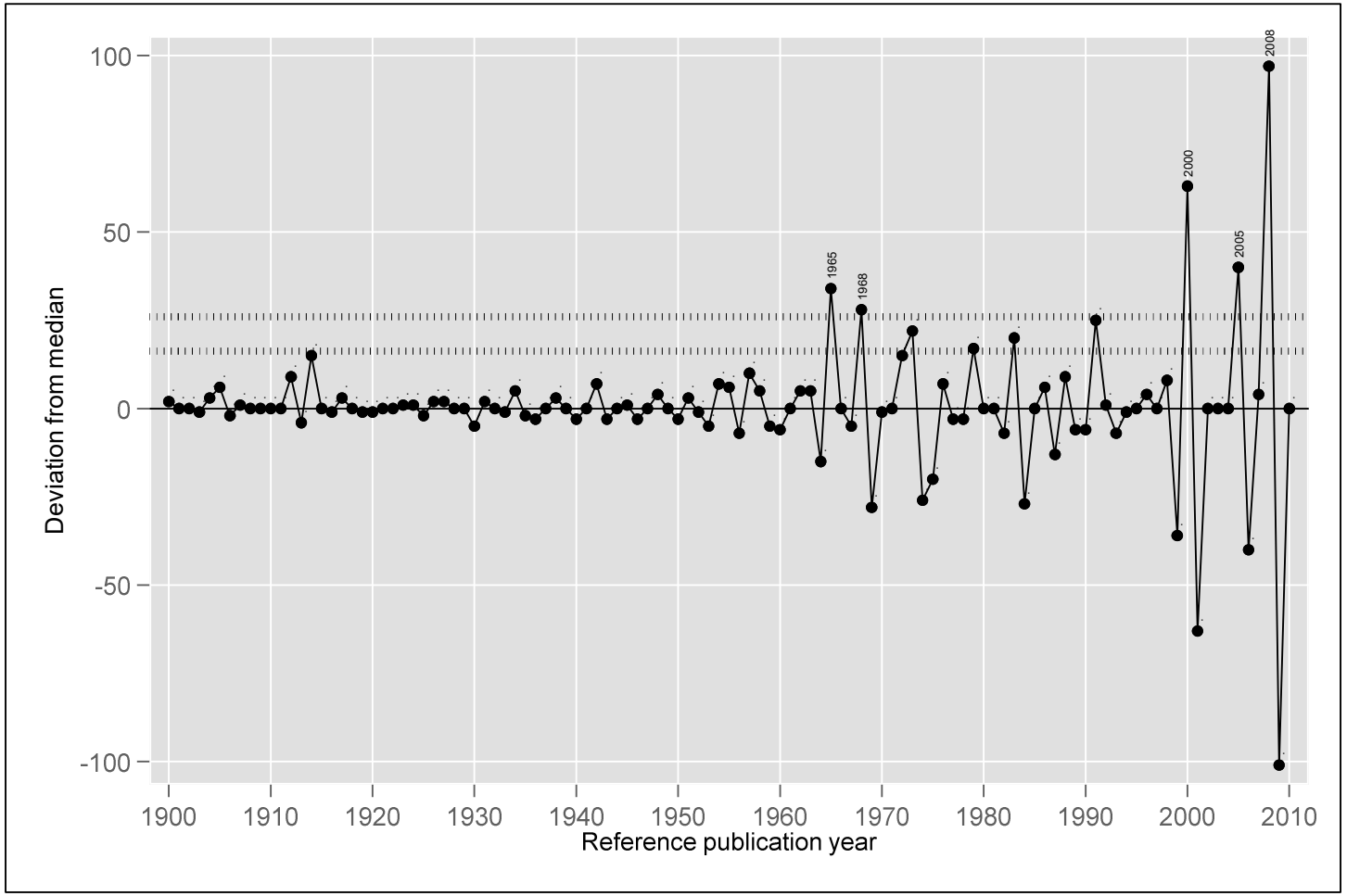
New developments:
Cited references analysis instead of times cited analysis

- Identification of historical roots of research fields, topics, researchers etc.
- Identifying landmark papers which have been influential over many years
- CRExplorer (www.crexplorer.net)
- First step: selection of publication set on which citation impact should be measured
- Second Step: exporting publication data including cited references (e.g., Web of Science)
- Third step: analyzing peaks (reference publication years)



New developments: Cited references analysis instead of times cited analysis

- Identification of my historical roots using CRExplorer
- Publication set: 324 papers (Web of Science)
- Several peaks (1965, 1968, 2000, 2005, 2008)
- Price, D. J. d. S. (1965). Networks of scientific papers. Science, 149(3683), 510-515 (23%; identification of research fronts)
- Merton, R. K. (1968). The Matthew effect in science. Science, 159(3810), 56-63 (41%)



New developments: Indicators measuring disruptiveness

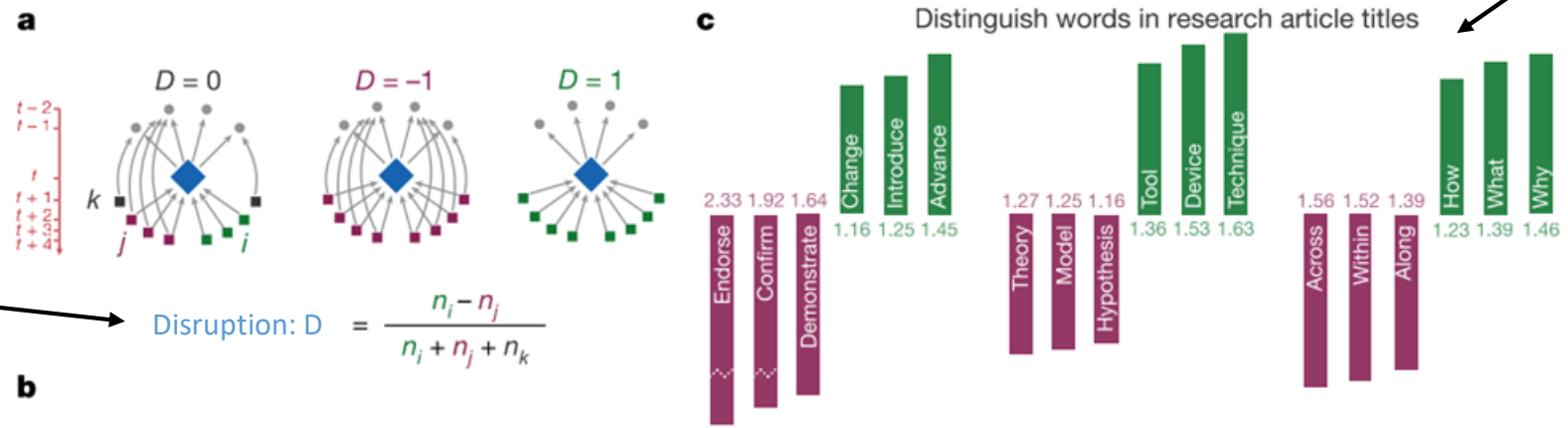
- One-dimensional perspective: simple citation counting
- Multi-dimensional perspective: cited references data of focal papers (FPs) and citing papers are considered
- Disruption indicators combine cited references of citing papers with cited references data of FPs
- The disruptiveness of a FP is measured based on the extent to which the cited references of the papers citing the FP also refer to the cited references of the FP
- Many citing papers not referring to the FP's cited references indicate disruptiveness
- In this case, the FP is the basis for new work which does not depend on the context of the FP, i.e. the FP gives rise to new research.

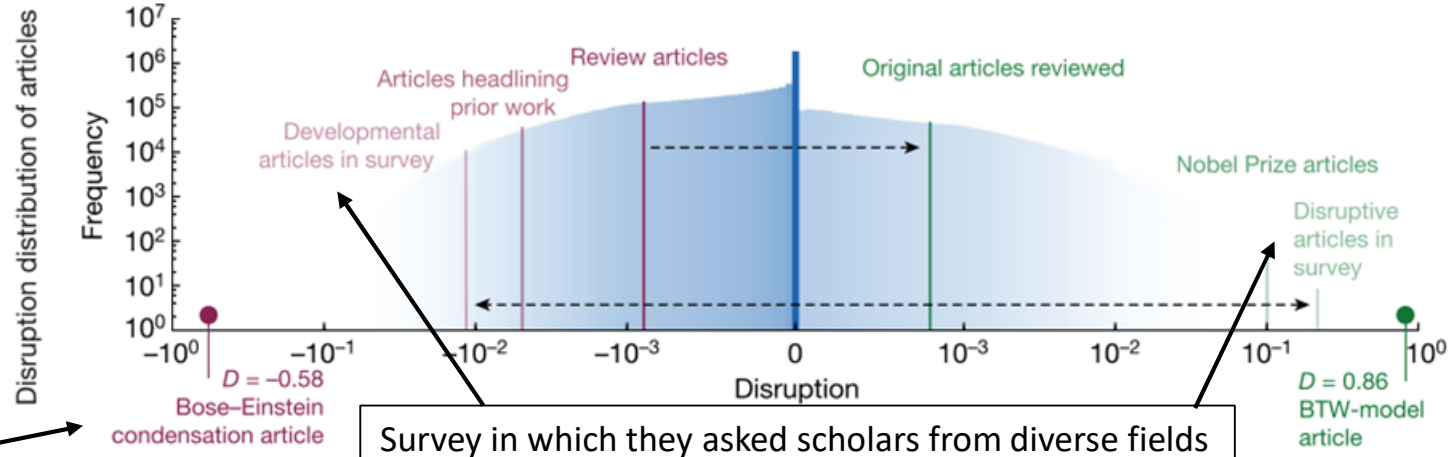
New developments: Indicators measuring disruptiveness

n_i : papers that cite the FP without citing any of its cited references

n_j : papers that cite both the FP and at least one of its cited references

n_k : papers that cite at least one of the FP's cited reference without citing the paper itself





Differences in the content and writing style between disrupting versus developing articles in terms of verbs, nouns, and adverbs and prepositions. Ratio (r) of word frequency in disrupting versus developing articles (title words). r is in green if $r > 1$, and $1/r$ in red otherwise

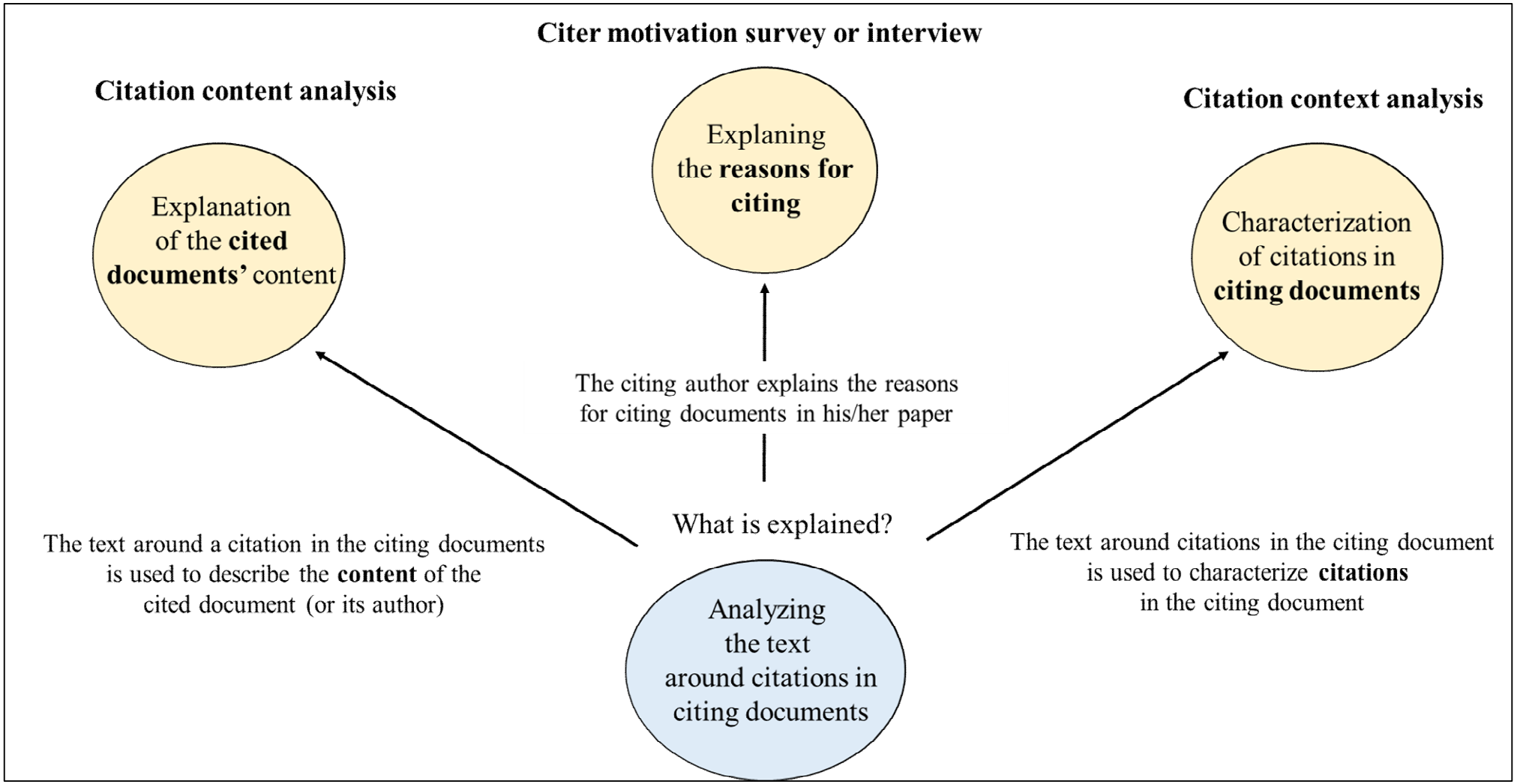
Bak, P., Tang, C. & Wiesenfeld, K. (1987). Self-organized criticality: an explanation of the 1/f noise. *Phys. Rev. Lett.* 59, 381-384 (1987)

Davis, K. B. et al. (1995) Bose–Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* 75, 3969-3973.

Survey in which they asked scholars from diverse fields to propose disruptive and developmental articles

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, 566, 378–382.

New developments:
More detailed citation analysis instead citation counting



Tahamtan, I., & Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics, 121*(3), 1635–1684

Citation concept analysis of *The structure of scientific revolutions* by Thomas Kuhn. How frequently have concepts been cited?

Concept	Frequency	Percent
Paradigm	2294	40.44
Normal science	321	5.66
Scientific revolution	278	4.9
Anomaly	148	2.61
Incommensurability	110	1.94
Structure	85	1.5
Crisis	78	1.37
Total	3314	58.42
Cases	5673	100

Bornmann, L., Wray, K. B., & Haunschild, R. (2019). Citation concept analysis (CCA): A new form of citation analysis revealing the usefulness of concepts for other researchers illustrated by two exemplary case studies including classic books by Thomas S. Kuhn and Karl R. Popper. *Scientometrics, 122*(1), 1051-1074.

- Which concepts introduced in a landmark publication have the most impact on science?
- Identification of the concepts in the publication and generating corresponding search terms
- Search for the concepts in citances – sentences in citing publications around citations of the landmark publication
- Counting the occurrences of concept mentions in citances

Web pages of tools

- Web application which visualizes research excellence worldwide in several subject areas:

  www.excellencemapping.net

- Web application which visualizes how successful universities or research-focused institutions collaborate:

  www.excellence-networks.net

- CRExplorer: A program for identifying citation classics and landmark papers of fields:

  www.crexplorer.net