Service and Research at the Information Retrieval Service of the CPT Section (IVS-CPT)

Robin Haunschild

12/20/2024

Service at IVS-CPT

Main services

- Bibliometric analyses (e.g., reports about persons or MPIs)
- Special scientific literature searches (e.g., exhaustive publication set regarding a topic or a compound)
- Patent searches (e.g., prior art or FTO searches)
- ivsdb
- . . .

Main databases

- Web of Science (WoS)
- Scopus
- STN (CAplus, Inspec, . . .); Derwent via WoS or Science IP
- OpenAlex (successor of Microsoft Academic Graph)
- . . .

Research at IVS-CPT

Main research activities

- Scientometrics, bibliometrics (e.g., analysis of research fields and testing/validating or proposing new indicators and methods)
- Chemical bibliometrics (e.g., counting chemical compounds or element occurrences instead of citations)
- Altmetrics (e.g., analysis of data from Twitter or Mendeley)
- . . .

Main databases

- WoS
- Scopus
- STN (CAS, Inspec, . . .)
- OpenAlex (successor of Microsoft Academic Graph)
- Overton
- Altmetric.com
- . . .

Access to WoS, Scopus, OpenAlex, CAplus/Registry, Derwent

General scientists' access

- Web interface: WoS, Scopus, OpenAlex, CAplus, Registry (SciFinder^n), Derwent (DWII in WoS)
- API: OpenAlex, CAplus/Registry (SciFinder^n)

Additional IVS-CPT access

- STN: CAplus/Registry and many other databases (some with numerical property search), e.g., Marpat, Inspec, Compendex, Reaxys, Derwent, full-text patent databases
- Raw data: WoS, Scopus, OpenAlex
  Via the German "Kompetenznetzwerk Bibliometrie" (KB) in a PostgreSQL database

Access to raw data offers the possibility of more search capabilities and calculation of advanced indicators

Access to raw data

Access to raw data

- 2013: Access via MPDL to Web of Science (Scopus was promised for later).
- 2015: Access via KB/MPDL to Web of Science and access to Scopus via KB.
- 2023: Access to Scopus raw data has still not been provided by MPDL. New indicators and other developments were blocked by MPDL over the recent years. Releases were available later in the year than usual (e.g., first release in 2024 was only at the end of September).
- 2024: Access via KB with ivsdb to Web of Science and Scopus. OpenAlex will be added next year.

KB: Kompetenznetzwerk Bibliometrie (*Competence Network Bibliometrics*)
MPDL: Max Planck Digital Library

Access to raw data via ivsdb

ivsdb

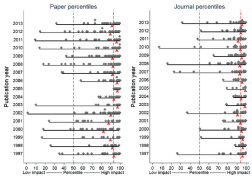
Timeline:

- Nov 2023: IVS-CPT has rented a server at MPCDF to run ivsdb
- Jan 2024: First test version of Web of Science (July 2023 data) raw data became available at ivsdb
- Jul 2024: First productive version of Web of Science (April 2024 data) raw data became available at ivsdb
- Sep 2024: First version of Scopus raw data became available at ivsdb
- Fourth quarter of 2024: Official switch from MPDL to IVS-CPT at the Administrative Headquarters of the MPG for access to bibliometric raw data

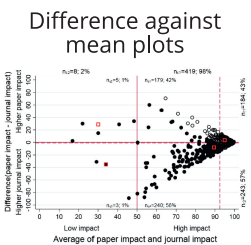
Access to raw data for:

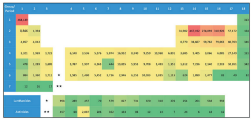
- Administrative Headquarters of the MPG
- other scientists (e.g., IVS-CPT)

IVS-CPT research impressions

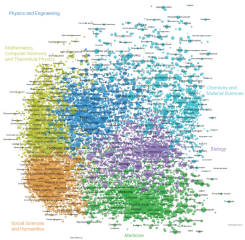


Beam plots

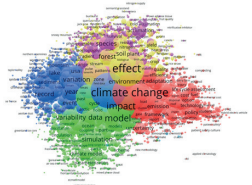




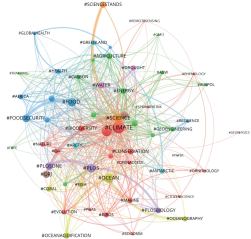
Chemical bibliometrics

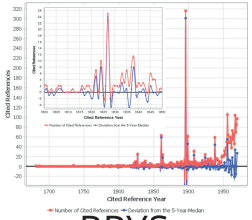


Global overlay maps



Word clouds and semantic maps



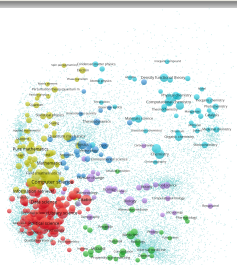


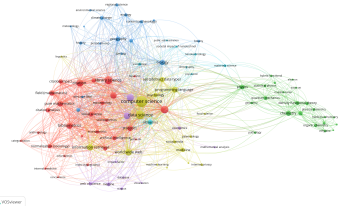
RPYS

Global maps using OpenAlex

Global vs. local

- Local maps only show the nodes of the focus dataset
- Global maps show the nodes of the focus dataset embedded in the environment
- Node positions are different for local and global maps
- Local maps are “cheaper”
- Global maps are “more expensive”





Overlay maps and base maps

What are overlay maps and base maps?

- Overlay maps use a base map to project data/information onto it
- Usually, the node size is scaled with the number of papers
- Node colors can be associated with impact, cluster assignment, field assignment, or other pieces of information
- Overlay maps can be constructed from local or global base maps
- Base maps provide the coordinates and cluster assignments of the nodes – all other data are provided by the overlay data

Why OpenAlex?

Why is OpenAlex suitable for global maps?

- OpenAlex is free
  $\rightarrow$ Everyone can obtain and use the data
- OpenAlex has a very large coverage of the literature
  $\rightarrow$ More of the scientific environment is included

Earlier approaches

Earlier approaches of constructing global maps

Clustering publications by

- direct citation relations
- bibliographic coupling relations
- co-citation relations
- semantic similarities based on title, abstract, and/or keywords
- using commercial (WoS or Scopus) or field-specific databases (PubMed)

Problems:

- Labeling of the nodes
- Interpretation, usage, and sharing of the maps

New approach

Method – node selection

- Concepts from OpenAlex are nodes on the map.
- Levels 0-2 out of 0-5 are used.
- 19 concepts on level 0, 284 concepts on level 1, and 21,460 concepts on level 2

Method – positioning of the nodes

- Direct citation relations are used for the calculation of the node positions
- Usage of a five-year citation (i.e., referencing) window (plus publication year) as default
- One map was created using a thirty-year citation window (plus publication year)

→ R. Haunschild & L. Bornmann (2024). The use of OpenAlex to produce meaningful bibliometric global overlay maps of science on the individual, institutional, and national levels. *PLOS ONE* 19(12): e0308041 DOI: 10.1371/journal.pone.0308041

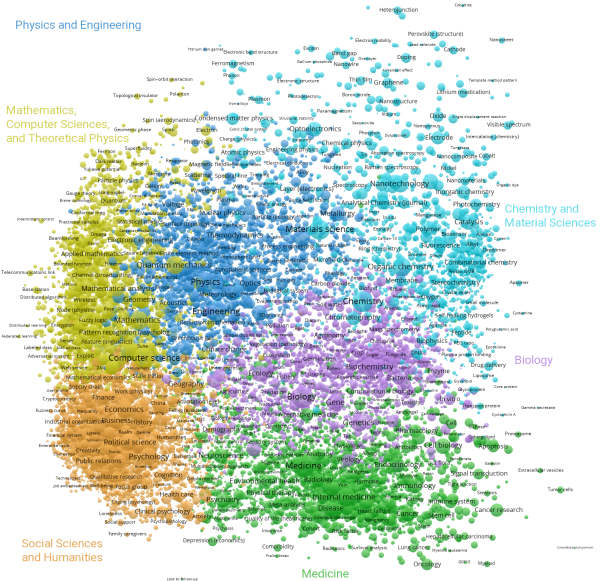
Dataset and method

Dataset for the base maps

OpenAlex snapshot from August 2023 with 243,053,925 documents in the “Kompetenznetzwerk Bibliometrie” (KB)

- 1800-2022: 237,876,541 (237,830,057) documents
- **2008-2022: 134,092,007 (134,054,634) documents**
- 2013-2022: 95,438,459 (95,406,638) documents
- 2018-2022: 47,665,990 (47,641,330) documents
- 2022: 8,496,167 (8,478,584) documents
- No document type restrictions for obtaining global maps

2008-2022 with a five-year citation window



Cluster colors were assigned according to cluster size: orange (social sciences and humanities), green (medicine), blue (physics and engineering), yellow (mathematics, computer sciences, and theoretical physics), pink (biology), and light-blue (chemistry and material sciences)

Examples for overlay maps

Different overlay maps on the base map 2008-2022

- A) Papers that are assigned to the OpenAlex author ID of Robin Haunschild
- B) Papers that are assigned to the OpenAlex source ID of the journal *Science*

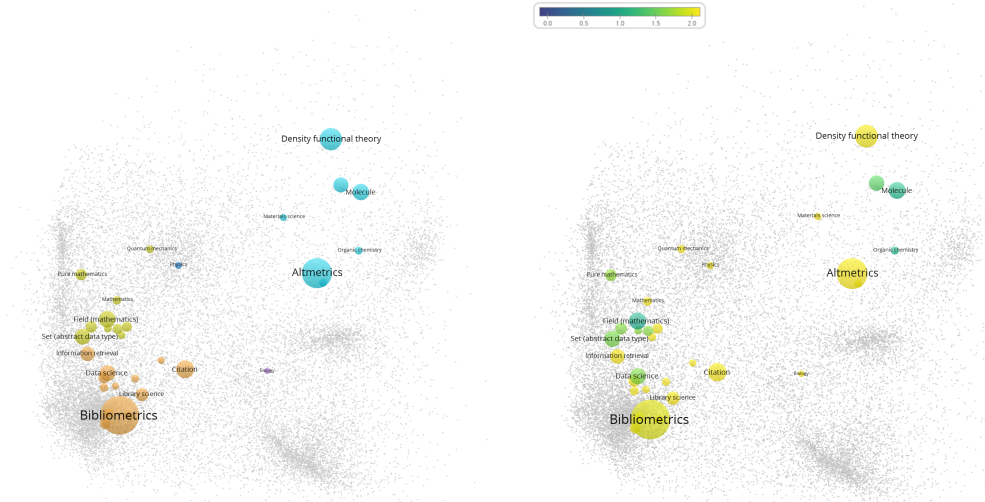
Node colors

- Cluster assignment
- Mean normalized citation score (MNCS)

Dataset

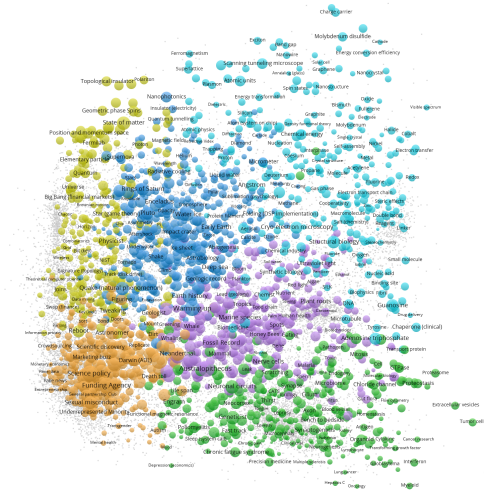
- Only documents of type article
- Only documents with impact (NCS) measurement
- Time period 2008-2020
- Only nodes with at least ten documents are shown

Example: global overlay maps of Robin Haunschild

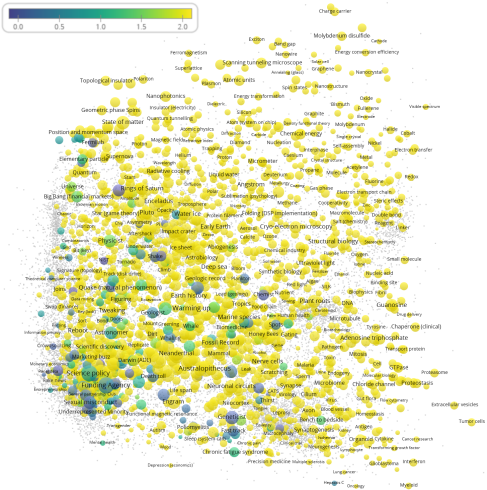


(with cluster coloring) (with impact (MNCS) coloring)

Example: global overlay maps of the journal *Science*

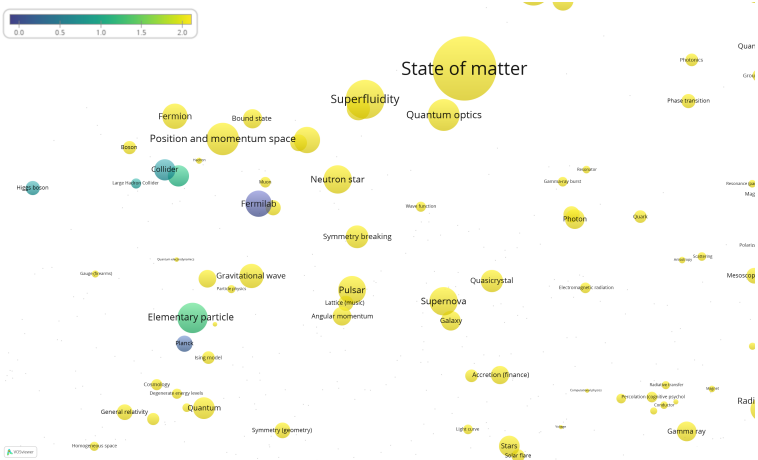


(with cluster coloring)



(with impact (MNCS) coloring)

Example: global overlay maps of the journal *Science* – Zoom



(with impact (MNCS) coloring)

Summary

Take-home messages

- IVS-CPT offers a broad range of services and performs research with the aim of improving its services
- IVS-CPT is active in bibliometric research
- In collaboration with the administrative headquarters, we have set up our own database server for bibliometric raw data to serve ourselves and the administrative headquarters
- Open global base maps of science for different time periods were provided
- Overlay maps for different bibliometric data were presented
- Data for overlay maps can be retrieved via the GUI, API, or from a snapshot of OpenAlex
- Other data (reader impact, average publication year, proportion of SDG-relevant papers, . . .) can be used as overlay color

Scientific staff at IVS-CPT

Dr. Thomas Scheidsteger (2M9, 1284)
T.Scheidsteger@fkf.mpg.de

Dr. Robin Haunschild (2M11, 1285)
R.Haunschild@fkf.mpg.de

Email: R.Haunschild@fkf.mpg.de Twitter: @rhaunschild
Mastodon: @robinhaunschild@academiccloud.social
https://www.fkf.mpg.de/ivs