BIBLIOMETRIC INFORMATION FROM OPEN ALEX DATABASE

Magí Romanyà, Nico Royo, Ot Garcés, Víctor Sáez, Albert Diaz-Guilera
Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona

This application, developed at UBICS, will allow you to obtain a list of publications indexed in OpenAlex database with their bibliometric information, statistics of the journals in which they have been published, the network of collaborations between the authors, the network of concepts, the network of institutions, and interdisciplinarity information.

DATA ENTRY

The first window shows 3 alternative ways to enter the data

Search by IDs file To search from a list of IDs, upload a .dat file with the ORCIDs or the DOIs in the first column. 📄 Search by ORCIDs file 📄 Search by DOIs file	Searches from a data file that contains a column listing either ORCIDs or DOIs
Search by other parameters You can also search by affiliation name, ROR, author name, ORCID, publisher title, ISSN or keywords. 📋 Search by parameters	Search from text strings, or other alternatives, in the different fields that the OpenAlex search engine allows

And two additional features:

- TORCH Collaboration project: CHARM-EU is one of the European Universities Alliances and TORCH corresponds to its R&I dimension. Click on the left button for getting collaborations among CHARM-EU universities according to different filters. Do it on the right button for identifying researchers at any of the universities that publish according to some criteria.
- Collaboration search: choose a list of institutions and find their collaborations.

Search by other parameters

This option allows a combined search of most of the fields that appear in the publication records in the database:

Search options

You can search for papers by using the options below. Add more search terms with the [+ Add search field] button, delete them with 🗑 and use the *and*, *or*, *not* boolean modifiers to combine them. The options available are:

- **Affiliation name:** The name of the affiliation of an author.
- **ROR:** When input is introduced at the input field, a list of possible institutions is shown. Select only one institution per input field. Once an option is selected, the ROR ID is automically selected.
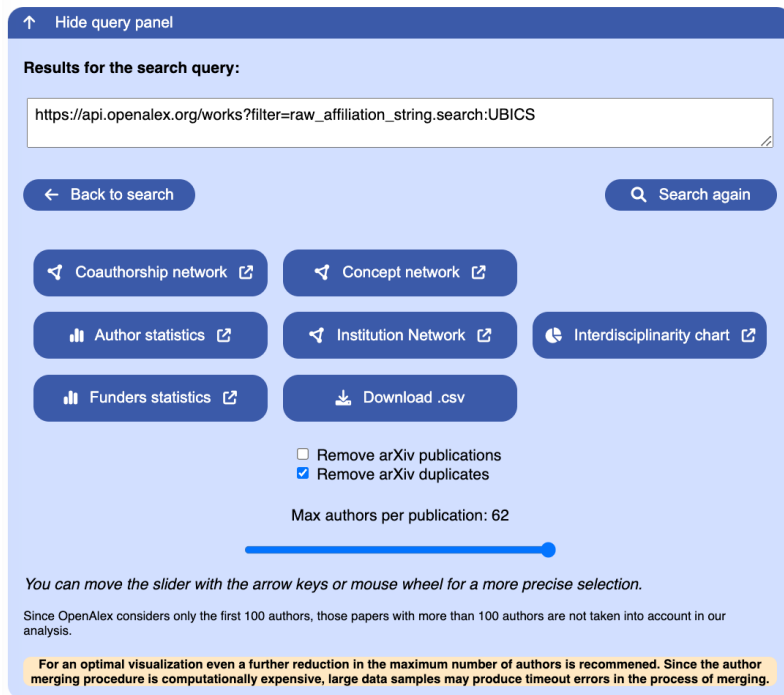 *Instructions:* For the ROR search input *Barcelona*, we get the following list of options.

  ☐ University of Barcelona (UB)
  ☐ Autonomous University of Barcelona (UAB)
  ☐ Barcelona Supercomputing Center (BSC)
  [All] [✓]

  One option must be selected and then the [✓] button pressed.

- **Author name:** The name of an author. A list of possible authors will appear, choose the one you are looking for. Instruccions are the same as for the ROR search, but multiple options can be selected.
  Example: If you are looking for *Fake Sur Name*, the author might be in the database as *Fake Sur-Name*, *F. Sur-Name*, *Fake S. N.*, etc. Select all of them.
- **ORCID:** The ORCID of an author. The format must be only the ID, not the whole link. You can find the ORCID of an author in orcid.org.
- **ISSN:** The ISSN of the journal.
- **Keyword:** As with the author name, a list will appear. This field is useful to filter by topic. Instructions are the same as for the ROR search, but multiple options can be selected.

- "Affiliation name": authors affiliations, in a general sense
- "ROR ID": Institutional identifier (it can be found here https://ror.org/). See screenshot above.
- "Author": The name of an author. A list of possible authors will appear, choose the one you are looking for. See also screenshot above.
- "ORCID": the ORCID of an author
- "Keyword": similar to the authors name, a list will appear and just check the appropriate item
- "ISSN": journal code
- "Project ID": project ID as provided by OpenAlex
- Add dates range: the years interval can be fixed.

These searches can be combined within the same field or with others (Boolean operators follow the usual precedence rules without brackets, all "AND" are considered first).

SEARCH RESULTS

Once the "Search" is done, and depending on the number of publications it will take more or less time, this information will be returned to you:

- Search results: so that you can check what has actually been searched for, that can be edited and used for a new search through "Search again".
- "Coauthorship network": generates an interactive network of collaborations between the authors of the searched publications.
- "Author statistics": generates an interactive statistic of the publications and allows you to see the contributions of each author.
- "Funders statistics": generates a list of funding institutions and a histogram.
- "Concept network": generates an interactive network of concepts as defined in the OpenAlex database.
- "Institution network": generates an interactive network of institutions where authors of the publications are affiliated.
- "Download .csv": Allows you to download a .CSV text file without formatting, which can be immediately incorporated into an Excel or Word file, to easily generate reports, resumes, ...
- "Interdisciplinary chart": generates an interactive pie chart of disciplines as defined in CiteScore and in OpenAlex.
- arXiv duplicates can be removed (this is the default option looking for titles) or all arXiv contributions not being considered.
- "Metric info": Small window that can be shown and hidden that explains the metrics used. The indicators used are based on Citescore, but as there are articles published in 2023 and the year indexes have not yet been released, the data for these years are shown with the indicators of the last year available, in this case in 2022.[1]

And on the right panel, a list of publications with the following fields:
- TITLE: : with an immediate access to the publication in another browser window
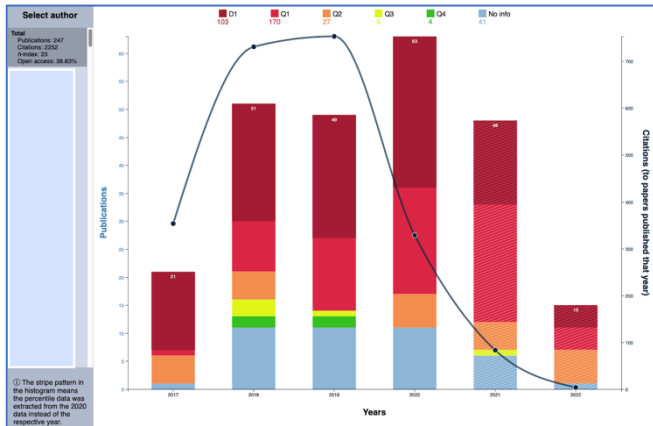- AUTHORS
- JOURNAL

---

[1] Citescore does not provide data prior to 2011.

- DOI
- Citation count

Which are the same that are downloaded in the .CSV file. In addition, we provide the bibliometric information of the journals:

- "Show journal info": detailed information on journal indicators (2023 publications currently correspond to 2022 journal indicators).

Statistics of the publications

If you go to this section now you will find complete statistics of the journals where the entire set of authors have published or you can select the authors individually. Number of publications per year, number of citations these publications have had, and quartiles (and 1st decile) if there is information about the journal in Citescore. Years without information are marked differently.
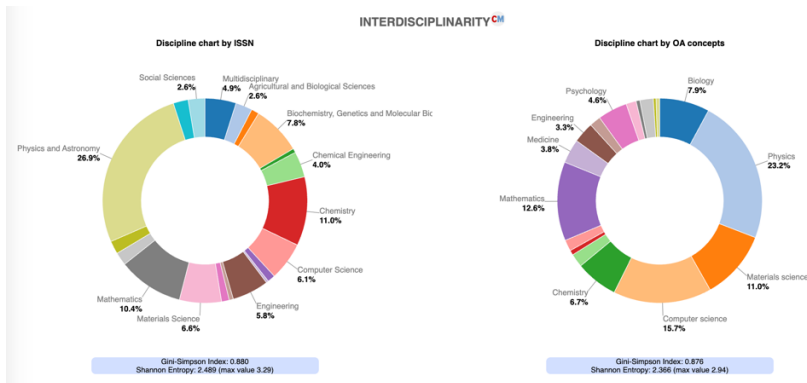
Coauthorship network
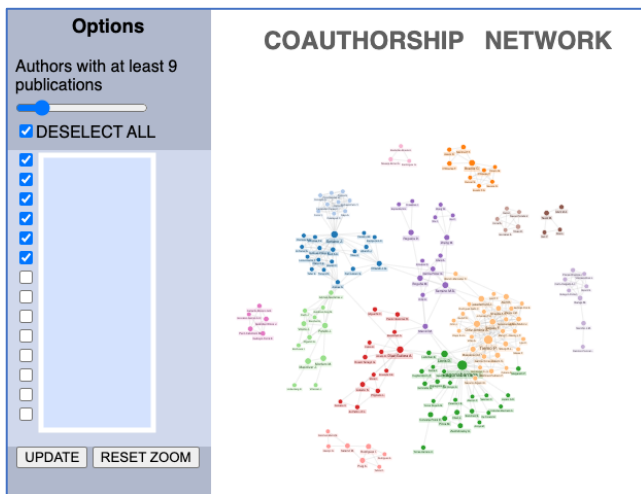
When entering this part of the application, a bar will appear that will allow you to set the minimum number of publications for an author to appear on the network, which you can change at will (the default value is 5). Once you have chosen the minimum number you can select or eliminate the authors you consider[2]. The network is interactive in determining the authors you want to appear but it is also interactive in positioning the nodes, which you can move if you want to other positions. With the Zoom you can zoom out or zoom in (the ZOOM button will allow you to restore the initial perspective).

Funders statistics

---

[2] Hint: A reasonable minimum number might be 2 or 3 publications. Once this is done, do UPDATE and it will update your network. Now you can lower the minimum number again and new authors will come up which you can remove or select one by one. Note that consortia or large collaborations tend to mask the network.

A list with the funding organizations will appear and a histogram with the contributions will show up in the right panel.

Institution network

The same than for the authors' network. Institutions can be selected and unselected by clicking the appropriate boxes. The minimal number of contributions to be represented in the network can also be chosen.

Concepts network

The same than for the authors' network. Concepts as defined in the OpenAlex database, classified according to hierarchical levels. For each publication, the database provides a list of concepts at different levels and its significance. In this case we provide two filters, one for the minimal number of appearances of a concept and the other for the maximal level of detail in the classification. Level 0 is taken as the discipline level and so on.

Interdisciplinary charts

INTERDISCIPLINARITY CM

Discipline chart by ISSN

- Social Sciences 2.6%
- Multidisciplinary 4.9%
- Agricultural and Biological Sciences 2.6%
- Biochemistry, Genetics and Molecular Bio 7.8%
- Chemical Engineering 4.0%
- Chemistry 11.0%
- Computer Science 6.1%
- Engineering 5.8%
- Materials Science 6.6%
- Mathematics 10.4%
- Physics and Astronomy 26.9%

Gini-Simpson Index: 0.880
Shannon Entropy: 2.489 (max value 3.29)

Discipline chart by OA concepts

- Psychology 4.6%
- Engineering 3.3%
- Medicine 3.8%
- Mathematics 12.6%
- Chemistry 6.7%
- Computer science 15.7%
- Materials science 11.0%
- Physics 23.2%
- Biology 7.9%

Gini-Simpson Index: 0.876
Shannon Entropy: 2.366 (max value 2.94)

Two interdisciplinarity charts. In the leftmost one, disciplines are considered in terms of the classification of the journals in CiteScore. In the rightmost chart, disciplines are extracted from the top-level concepts of the publication in OpenAlex. In both cases we provide two simple indices, the Gini-Simpson index, and the Shannon entropy, to "quantify" interdisciplinarity.[3]

---

[3] Quantifying interdisciplinarity is a complex and unsolved issue in bibliometrics. For more information about these indices and further generalizations see …….

SPECIAL QUERY: INSTITUTIONAL COLLABORATIONS

A special query to show collaborations between institutions has been added to the current release of the application. For each institution enter the desired abbreviation and select the ROR from the pop-up menu (see screenshot). A filter of either disciplines or keywords is also available with the same type of menu.

Warning about author names and identifications (OAID)

Due to some inconsistencies in the OpenAlex data that assigns different OAID to authors with similar or even identical names we have followed a merging procedure. We compute cosine similarity between names and merges those whose similarity is above some threshold, considering also the order of names and surnames and also compatibility between names and initials.