CWTS Meaningful metrics

Leiden Ranking Open Edition 2024

Methodology

Universiteit Leiden

Centre for Science and
Technology Studies,
Leiden University

Data

The CWTS Leiden Ranking Open Edition 2024 is based on bibliographic data from the OpenAlex database produced by OurResearch. The ranking also uses data from an organization registry created and maintained by CWTS. This registry is partly built on data from the Research Organization Registry (ROR).

Universities

The CWTS Leiden Ranking Open Edition 2024 includes 1506 universities worldwide. These are the same universities that are also included in the traditional Leiden Ranking 2024. As discussed below, a sophisticated methodology is employed to assign publications to universities.

Research Organization Registry (ROR) and affiliated institutions

A key challenge in the compilation of a university ranking is the handling of publications originating from research institutes and hospitals affiliated with universities. Among academic systems, a wide variety exists in the types of relations maintained by universities with these affiliated institutions. Usually, these relationships are shaped by local regulations and practices affecting the comparability of universities on a global scale. As there is no easy solution for this issue, it is important that producers of university rankings employ a transparent methodology in their treatment of affiliated institutions.

For the CWTS Leiden Ranking Open Edition we use the relationships between universities and their affiliated institutions included in the Research Organization Registry (ROR). For the classification of these relationships we use the same the methodology that is also used in the traditional Leiden Ranking.

CWTS distinguishes three different types of affiliated institutions:

1. Component
2. Joint research facility or organization
3. Associated organization

In the case of a *component*, the affiliated institution is actually part of or controlled by the university. Universitaire Ziekenhuizen Leuven is an example of a component, since it is part of the legal entity of Katholieke Universiteit Leuven.

A *joint research facility or organization* is the identical to a component except that it is administered by more than one organization. The Brighton & Sussex Medical School (the joint medical faculty of the University of Brighton and the University of Sussex) and Charité - Universitätsmedizin Berlin (the medical school of both the Humboldt University and the Freie Universität Berlin) are examples of this type of affiliated institution.

The third type of affiliated institution is the *associated organization*, which is more loosely connected to a university. This organization is an autonomous institution that collaborates with one or more universities based on a joint purpose but at the same time has separate missions and tasks. In many countries, hospitals that operate as teaching or university hospitals fall into this category. The Massachusetts General Hospital, one of the teaching hospitals of the Harvard Medical School, is an example of an associated organization.

A publication is counted as output of a university if at least one of the affiliations in the publication explicitly mentions either the university or one of its components or joint research facilities. In a limited number of cases, affiliations with institutions that are not controlled or owned by the university are also treated as if they were mentioning the university itself. The rationale for this is that in some cases institutions – although formally being distinct legal entities – are so tightly integrated with the university that they are commonly perceived as being a component or extension of that university. Examples of this situation include the university medical centers in the Netherlands and some of the academic health science systems in the United States and other countries. In these cases, universities have actually delegated their medical research and teaching activities to the academic hospitals and universities may even no longer act as the formal employer of the medical researchers involved. In other cases, tight integration between a university and an academic hospital may manifest itself by an extensive overlap in staff. In this situation, researchers may not always mention explicitly their affiliation with the university. An example of this tight integration is the relation between the University Hospital Zurich and the University of Zurich.

Our approach is discussed in more detail in this paper on academic hospitals.

Affiliated institutions that are not classified as a component or a joint research facility or treated as such are labeled as associated institutions. In the case of publications with affiliations from associated organizations, a distinction is made between publications from associated organizations that also mention the university and publications from associated organizations that do not include a university affiliation. In the latter case, a publication is not considered to originate from the university. On the other hand, if a publication includes an affiliation from a particular university as well as an affiliation from an associated organization, both affiliations are considered to represent that particular university. The effect of this procedure depends on the counting method that is used in the calculation of bibliometric

indicators. The procedure influences results obtained using the fractional counting method, but it has no effect on results obtained using the full counting method.

Data quality

For the assignment of publications to universities, the Leiden Ranking Open Edition relies on OpenAlex data. OpenAlex links affiliation strings to ROR identifiers. The linking of affiliation strings to ROR identifiers is a difficult task and not free of errors. It is also important to emphasize that in general the assignment of publications to universities has not been verified and approved by the universities themselves.

Two types of errors are possible in assigning publications to universities. On the one hand, there may be false positives, which are publications that have been assigned to a university while in fact they do not belong to the university. On the other hand, there may be false negatives, which are publications that have not been assigned to a university while in fact they do belong to the university. Both types of errors occur, but in general there are substantially more false negatives than false positives. One reason for this is that affiliation data is missing for a small share of the publications in OpenAlex. This blog post presents a comparison between the approaches for linking publications to universities used in the Leiden Ranking Open Edition and in the traditional Leiden Ranking.

Main fields

The CWTS Leiden Ranking Open Edition 2024 provides statistics not only at the level of science as a whole but also at the level of the following five main fields of science:

1. Biomedical and health sciences
2. Life and earth sciences
3. Mathematics and computer science
4. Physical sciences and engineering
5. Social sciences and humanities

As discussed below, these five main fields are defined based on large number of micro-level fields.

Algorithmically defined main fields

Each publication of a university belongs to one, or sometimes to more than one, of the above main fields. If a publication belongs to more than one main field, the publication is assigned fractionally to each of the main fields. For instance, a publication belonging to two main fields is assigned to each of the two fields with a weight of 1 / 2 = 0.5.

Publications are assigned to the five main fields using an algorithmic approach. Traditionally, fields of science are defined by sets of related journals. This approach is problematic especially in the case of multidisciplinary journals such as *Nature*, *PLOS ONE*, *PNAS*, and *Science*, which do not belong to one specific scientific field. The five main fields listed above are defined at the level of individual publications rather than at the journal level. In this way, publications in multidisciplinary journals can be properly assigned to a field.

Publications are assigned to main fields in the following three steps:

1. We start with 4521 micro-level fields of science. These fields are constructed algorithmically. Using a computer algorithm, each publication in OpenAlex is assigned to one of the 4521 fields. This is done based on a large-scale analysis of hundreds of millions of citation relations between publications.
2. We then determine for each of the 4521 micro-level fields the overlap with each of the 284 level 1 concepts defined in OpenAlex.

3. Each level 1 concept in OpenAlex has a link to one of the five main fields. Based on the link between level 1 concepts and main fields, we assign each of the 4521 micro-level fields to one or more of the five main fields. A micro-level field is assigned to a main field if at least 20% of the publications in the micro-level field cluster belong to subject categories linked to the main field.

After the above steps have been taken, each publication in OpenAlex has an assignment to a micro-level field, and each micro-level field in turn has an assignment to at least one main field. Combining these results, we obtain for each publication an assignment to one or more main fields.

More information

This blog post discusses the approach taken to construct the micro-level fields. For more information on the methodology for the algorithmic construction of the micro-level fields, we refer to a paper by Waltman and Van Eck (2012). The methodology makes use of the Leiden algorithm. This algorithm is documented in a paper by Traag et al. (2019).

Traag, V.A., Waltman, L., & Van Eck, N.J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*, 5233. doi:10.1038/s41598-019-41695-z.

Waltman, L., & Van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392. doi:10.1002/asi.22748.

Indicators

The CWTS Leiden Ranking Open Edition 2024 offers a sophisticated set of bibliometric indicators that provide statistics at the level of universities on scientific impact, collaboration, and open access publishing. The indicators are discussed in detail below.

Publications

The Leiden Ranking Open Edition is based on publications in the OpenAlex database produced by OurResearch. The most recent indicators made available in the Leiden Ranking Open Edition are based on publications in the period 2019–2022, but indicators are also provided for earlier periods.

The Leiden Ranking Open Edition takes into account only a subset of the publications in OpenAlex. We refer to these publications as core publications. Core publications are publications in international scientific journals in fields that are suitable for citation analysis. In order to be classified as a core publication, a publication must satisfy the following criteria:

- The publication has type *article* or *book chapter* and has been published in a source that has type *journal* or *book series*.
- The publication has authors, affiliations, and references.
- The publication has been written in English.
- The publication has not been retracted.
- The publication has appeared in a core journal.

The last criterion is very important. A journal is considered a core journal if it meets the following conditions:

- The journal has an international scope, as reflected by the countries in which researchers publishing in the journal and citing to the journal are located.
- The journal has a sufficiently large number of references to other core journals, indicating that the journal is situated in a field that is suitable for citation analysis. Many journals in the arts and humanities do not meet this condition. The same applies to trade journals and popular magazines.

In the calculation of indicators, only core publications are taken into account. In this way, the Leiden Ranking Open Edition aims to resemble the traditional Leiden Ranking as closely as possible.

Size-dependent vs. size-independent indicators

Indicators included in the Leiden Ranking Open Edition have two variants: A size-dependent and a size-independent variant. In general, size-dependent indicators are obtained by counting the absolute number of publications of a university that have a certain property, while size-independent indicators are obtained by calculating the proportion of the publications of a university with a certain property. For instance, the number of highly cited publications of a university and the number of publications of a university co-authored with other organizations are size-dependent indicators. The proportion of the publications of a university that are highly cited and the proportion of a university's publications co-authored with other organizations are size-independent indicators. In the case of size-dependent indicators, universities with a larger publication output tend to perform better than universities with a smaller publication output. Size-independent indicators have been corrected for the size of the publication output of a university. Hence, when size-independent indicators are used, both larger and smaller universities may perform well.

Scientific impact indicators

The Leiden Ranking Open Edition provides the following indicators of scientific impact:

- *P*. Total number of publications of a university.
- *P(top 1%) and PP(top 1%)*. The number and the proportion of a university's publications that, compared with other publications in the same field and in the same year, belong to the top 1% most frequently cited.
- *P(top 5%) and PP(top 5%)*. The number and the proportion of a university's publications that, compared with other publications in the same field and in the same year, belong to the top 5% most frequently cited.
- *P(top 10%) and PP(top 10%)*. The number and the proportion of a university's publications that, compared with other publications in the same field and in the same year, belong to the top 10% most frequently cited.

- *P(top 50%) and PP(top 50%)*. The number and the proportion of a university's publications that, compared with other publications in the same field and in the same year, belong to the top 50% most frequently cited.
- *TCS and MCS*. The total and the average number of citations of the publications of a university.
- *TNCS and MNCS*. The total and the average number of citations of the publications of a university, normalized for field and publication year. An MNCS value of two for instance means that the publications of a university have been cited twice above the average of their field and publication year.

Citations are counted until the end of 2023 in the calculation of the above indicators. Author self-citations are excluded. All indicators except for TCS and MCS are normalized for differences in citation patterns between scientific fields. For the purpose of this field normalization, about 4500 fields are distinguished. These fields are defined at the level of individual publications. Using a computer algorithm, each publication in OpenAlex is assigned to a field based on its citation relations with other publications. More information is provided in this blog post.

The TCS, MCS, TNCS, and MNCS indicators are not available on the main ranking page. These indicators can be accessed by clicking on the name of a university. An overview of all bibliometric statistics available for the university will then be presented. This overview also includes the TCS, MCS, TNCS, and MNCS indicators.

Collaboration indicators

The Leiden Ranking Open Edition provides the following indicators of collaboration:

- *P*. Total number of publications of a university.
- *P(collab) and PP(collab)*. The number and the proportion of a university's publications that have been co-authored with other organizations.
- *P(int collab) and PP(int collab)*. The number and the proportion of a university's publications that have been co-authored by multiple countries.
- *P(industry) and PP(industry)*. The number and the proportion of a university's publications that have been co-authored with organizations classified as industry in OpenAlex.
- *P(<100 km) and pp(<100 km)*. The number and the proportion of a university's publications with a geographical collaboration distance of less

than 100 km. The geographical collaboration distance of a publication equals the largest geographical distance between two addresses mentioned in the publication's address list.

- *P(>5000 km) and PP(>5000 km)*. The number and the proportion of a university's publications with a geographical collaboration distance of more than 5000 km.

Open access indicators

The Leiden Ranking Open Edition provides the following indicators of open access publishing:

- *P*. Total number of publications of a university.
- *P(OA) and PP(OA)*. The number and the proportion of open access publications of a university.
- *P(gold OA) and PP(gold OA)*. The number and the proportion of gold open access publications of a university. Gold open access publications are publications in an open access journal.
- *P(hybrid OA) and PP(hybrid OA)*. The number and the proportion of hybrid open access publications of a university. Hybrid open access publications are publications in a subscription journal that are open access with a license that allows the publication to be reused.
- *P(bronze OA) and PP(bronze OA)*. The number and the proportion of bronze open access publications of a university. Bronze open access publications are publications in a subscription journal that are open access without a license that allows the publication to be reused.
- *P(green OA) and PP(green OA)*. The number and the proportion of green open access publications of a university. Green open access publications are publications in a subscription journal that are open access not in the journal itself but in a repository.

In the calculation of the P(OA) and PP(OA) indicators, a publication is considered open access if it is gold, hybrid, bronze, or green open access.

Counting method

The scientific impact indicators in the Leiden Ranking Open Edition can be calculated using either a full counting or a fractional counting method. The full counting method gives a full weight of one to each publication of a university. The fractional counting method gives less weight to collaborative publications than to non-collaborative ones. For instance, if a publication has been co-authored by five researchers and two of these researchers are affiliated with a particular university, the publication has a weight of 2 / 5 = 0.4 in the calculation of the scientific impact indicators for this university. The fractional counting method leads to a more proper field normalization of scientific impact indicators and therefore to fairer comparisons between universities active in different fields. For this reason, fractional counting is the preferred counting method for the scientific impact indicators in the Leiden Ranking Open Edition.

Collaboration and open access indicators are always calculated using the full counting method.

Trend analysis

To facilitate trend analyses, the Leiden Ranking Open Edition provides statistics not only based on publications from the period 2019–2022, but also based on publications from earlier periods: 2006–2009, 2007–2010, …, 2018–2021. The statistics for the different periods are calculated in a fully consistent way. For each period, citations are counted until the end of the first year after the period has ended. For instance, in the case of the period 2006–2009 citations are counted until the end of 2010, while in the case of the period 2019–2022 citations are counted until the end of 2023.

Stability intervals

Stability intervals provide some insight into the uncertainty in bibliometric statistics. A stability interval indicates a range of values of an indicator that are likely to be observed when the underlying set of publications changes. For instance, the PP(top 10%) indicator may be equal to 15.3% for a particular university, with a stability interval ranging from 14.1% to 16.5%. This means that the PP(top 10%) indicator equals 15.3% for this university, but that changes in the set of publications of the university may relatively easily lead to PP(top 10%) values in the range from 14.1% to

16.5%. The Leiden Ranking Open Edition employs 95% stability intervals constructed using a statistical technique known as bootstrapping.

More information

More information on the indicators available in the Leiden Ranking can be found in a number of articles published by CWTS researchers. Field normalization of scientific impact indicators based on algorithmically defined fields is studied by Ruiz-Castillo and Waltman (2014). The calculation of percentile-based indicators of scientific impact is discussed by Waltman and Schreiber (2013). The methodology adopted in the Leiden Ranking for identifying core publications and core journals is outlined by Waltman and Van Eck (2013a, 2013b). The application of this methodology in the OpenAlex database is described by Van Eck and Waltman (2024). Finally, the importance of using fractional rather than full counting in the calculation of field-normalized scientific impact indicators is explained by Waltman and Van Eck (2015).

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, *9*(1), 102–117. doi:10.1016/j.joi.2014.11.010.

Van Eck, N.J., & Waltman, L. (2024). A methodology for identifying core sources and core publications in OpenAlex. *Zenodo*. doi:10.5281/zenodo.13879947.

Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. *Journal of the American Society for iInformation Science and Technology*, *64*(2), 372–379. doi:10.1002/asi.22775.

Waltman, L., & Van Eck, N.J. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716. doi:10.1007/s11192-012-0913-4.

Waltman, L., & Van Eck, N.J. (2013b). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics*, *7*(4), 833–849. doi:10.1016/j.joi.2013.08.002.

Waltman, L., & Van Eck, N.J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, *9*(4), 872–894. doi:10.1016/j.joi.2015.08.001.