CWTS Leiden Ranking 2014

Methodology

Data collection

The CWTS Leiden Ranking 2014 ranks the 750 universities in the world with the largest contribution in international scientific journals in the period of 2009–2012. The ranking is based on data from the Web of Science bibliographic database produced by Thomson Reuters.

Below, a summary of the data collection methodology of the CWTS Leiden Ranking 2014 is provided. It should be emphasized that, in general, universities did not verify and approve the publication data of their institution and that publications have been assigned to universities on the basis of the institutional affiliations mentioned by the authors of the publications. However, the assignment of publications from these affiliations is by no means a straightforward task. A university may be referred to by many different (non-English) name variants and abbreviations. In addition, the definition and delimitation of universities as separate entities is not always obvious.

Identification of universities

The criteria that have been adopted to define universities for the Leiden Ranking are not very formal. Typically, a university is characterized by a combination of education and research tasks in conjunction with a doctorate-granting authority. However, these characteristics do not mean that the universities are particularly homogeneous entities that allow for international comparison on every aspect. The focus of the Leiden Ranking on scientific research certifies that the institutions included in the Leiden Ranking have a high degree of research intensity in common. Nevertheless, the ranking scores for each institution should be evaluated in the context of its particular mission and responsibilities. These missions and responsibilities in turn are strongly linked to the national and regional academic systems in which universities operate. Academic systems - and the role of universities therein - differ substantially from one another and are constantly changing. Inevitably, the outcomes of the Leiden Ranking reflect these differences and changes.

The international variety in the organization of academic systems also poses difficulties in terms of identifying the proper unit of analysis. In many countries, there are collegiate universities, university systems, or federal universities. Again, instead of applying formal criteria when possible we followed common practice based on the way these institutions are perceived locally. Consequently, we treated

the University of Cambridge and the University of Oxford as entities but in the case of the University of London, we distinguished between the constituent colleges. For the United States, university systems (e.g. University of California) were split up into separate universities. The higher education sector in France, like in many other countries, has gone through many reorganizations in recent years. Many French institutions of higher education have been grouped together in *Pôles de Recherche et d'Enseignement Supérieur* (PRES), or in consortia. In most cases, the Leiden Ranking still distinguishes between the different constituent institutions but in particular cases of very tight integration, consortia were treated as if they were a single university (e.g. Grenoble INP).

Publications are assigned to universities based on their most recent configuration. Changes in the organizational structures of universities up to 2013 have been taken into account. For example, in the Leiden Ranking 2014, the University of Lisbon which merged with the Technical University of Lisbon in 2013 encompasses all publications assigned to the old University of Lisbon as well as the publications previously assigned to the Technical University of Lisbon.

Affiliated institutions

A key challenge in the compilation of a university ranking is the handling of publications originating from research institutes and hospitals associated 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 and affect 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.

CWTS distinguishes three different types of affiliated institutions:

1. component
2. joint research facility or organization
3. associated organization

In the case of *components* the affiliated institution is actually part of the university or so tightly integrated with it or with one of its faculties that the two can be considered as a single entity. The University Medical Centres in the Netherlands are examples of components. All teaching and research tasks in the field of medicine that were traditionally the responsibility of the universities have been delegated to these

separate organizations that combine the medical faculties and the university hospitals.

*Joint research facilities or organization*s are the same as components except for the fact that they are 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, the medical school for both the Humboldt University and Freie Universität Berlin are both examples of this type of affiliated institution.

The third type of affiliated institution is the *associated organization* which is more loosely connected to the 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. Massachusetts General Hospital, one of the teaching hospitals of Harvard Medical School, is an example of an associated organization.

The treatment of university hospitals in particular is of substantial consequence as medical research has a strong presence in the Web of Science. The importance of associated organizations is growing as universities present themselves more and more frequently as network organizations. As a result, researchers formally employed by the university but working at associated organizations may not always mention the university in publications. On the other hand, as universities become increasingly aware of the significance of their visibility in research publications, they actively exert pressure on researchers to mention their affiliation with the university in their publications.

In the Leiden Ranking 2014, publications from affiliated institutions of the first two types are considered as output from the university. A different procedure has been followed for publications 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 contain such a university affiliation. In the latter case, publications are not counted as publications originating from the university. In the event that a publication contains affiliations from a particular university as well as affiliations from its associated organization(s), both type of affiliations are credited to the contribution of that particular university to the publication in the fractional counting method.

The 750 universities: selection and counting method

The 750 universities that appear in the Leiden Ranking have been selected based on their contribution to articles and review articles published in international scientific journals in the period of 2009–2012. The contribution of a university to an article is calculated based on the number of affiliations mentioned in the article. If an article mentions three different affiliations of which two belong to a particular university, then the contribution of that university to the article is counted as two thirds. Only publications in core journals are included. The equivalent of more than 1,000 papers was required for a university to be ranked among the 750 universities with the largest scientific output.

Data quality

It is important to highlight that the assignment of publications to universities is not free of errors. There are generally two types of errors: 'false positives', which are publications that have been assigned to a university when they do not in fact belong to that university, and 'false negatives', which are publications that have not been assigned to a university when they should in fact have been. Considerably more false negatives than false positives should be expected, especially since the 5% least frequently occurring addresses in the database may not have been manually checked. This can be considered a reasonable upper bound for errors, since the majority of these addresses are probably non-university addresses.

Main fields

The CWTS Leiden Ranking 2014 provides statistics not only at the level of science as a whole but also at the level of the following seven broad fields of science:

1. Cognitive and health sciences
2. Earth and environmental sciences
3. Life sciences
4. Mathematics, computer science, and engineering
5. Medical sciences
6. Natural sciences
7. Social sciences

The above fields have been defined using a unique bottom-up approach. Traditionally, fields are defined as sets of closely 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 particular field. The seven broad fields of science listed above have been defined at the level of individual publications rather than at the journal level. Using a computer algorithm, each publication in the Web of Science database has been assigned to one of these seven fields. This has been done based on a large-scale analysis of hundreds of millions of citation relations between publications.

Indicators

The CWTS Leiden Ranking 2014 is based on publications in Thomson Reuters' Web of Science database (Science Citation Index Expanded, Social Sciences Citation Index, and Arts & Humanities Citation Index) in the period 2009–2012. Book publications, publications in conference proceedings, and publications in journals not indexed in the Web of Science database are not included. Within the Web of Science database, only publications in international scientific journals are included. In addition, only publications of the Web of Science document types *article* and *review* are considered.

Impact indicators

The Leiden Ranking offers the following indicators of the scientific impact of a university:

- *MCS (mean citation score)*. The average number of citations of the publications of a university.
- *MNCS (mean normalized citation score)*. The average number of citations of the publications of a university, normalized for field differences and publication year. An MNCS value of two for instance means that the publications of a university have been cited twice above world average.
- *PP(top 10%) (proportion of top 10% publications)*. The proportion of the publications of a university that, compared with other publications in the same field and in the same year, belong to the top 10% most frequently cited.

Citations are counted until the end of 2013 in the above indicators. Author self citations are excluded. Both the MNCS indicator and the PP(top 10%) indicator correct for differences in citation practices between scientific fields. 828 fields are distinguished. These fields are defined at the level of individual publications. Using a computer algorithm, each publication in the Web of Science database has been assigned to a field based on its citation relations with other publications. Because the PP(top 10%) indicator is more stable than the MNCS indicator, the PP(top 10%) indicator is regarded as the most important impact indicator of the Leiden Ranking.

Collaboration indicators

The following indicators of scientific collaboration are provided in the Leiden Ranking:

- *PP(collab) (proportion of interinstitutional collaborative publications).* The proportion of the publications of a university that have been co-authored with one or more other organizations.
- *PP(int collab) (proportion of international collaborative publications).* The proportion of the publications of a university that have been co-authored by two or more countries.
- *PP(UI collab) (proportion of collaborative publications with industry).* The proportion of the publications of a university that have been co-authored with one or more industrial partners.
- *PP(<100 km) (proportion of short distance collaborative publications).* The proportion of the publications of a university with a geographical collaboration distance of less than 100 km, where the geographical collaboration distance of a publication equals the largest geographical distance between two addresses mentioned in the publication's address list.
- *PP(>1000 km) (proportion of long distance collaborative publications).* The proportion of the publications of a university with a geographical collaboration distance of more than 1000 km.

Core journals

A journal is considered a core journal if it meets the following two conditions:

1. The journal publishes in English and has an international scope, as reflected by the countries in which researchers publishing in the journal and citing to the journal are located.
2. The journal has a sufficiently large number of references to other core journals in the Web of Science database, indicating that in terms of citation traffic the journal is well-connected to these other journals. Many journals in the humanities do not meet this condition. The same applies to trade journals and popular magazines.

In the calculation of the Leiden Ranking indicators, only publications in core journals are included. The MNCS and PP(top 10%) indicators become significantly more accurate by excluding publications in non-core journals. About 16% of the publications in the Web of Science database are excluded because they have appeared in non-core journals.

Size-dependent vs. size-independent indicators

The Leiden Ranking by default reports size-independent indicators. These indicators provide average statistics per publication, such as a university's average number of citations per publication. The advantage of size-independent indicators is that they enable comparisons between smaller and larger universities. As an alternative to size-independent indicators, the Leiden Ranking can also report size-dependent indicators, which provide overall statistics of the publications of a university. An example is the total (rather than the average) number of citations of the publications of a university. Size-dependent indicators are strongly influenced by the size of a university (i.e., a university's total publication output) and therefore tend to be less useful for comparison purposes.

Counting method

The impact indicators included in the Leiden Ranking can be calculated using either a full counting method or a fractional counting method. The full counting method gives equal weight to all publications of a university. The fractional counting method gives less weight to collaborative publications than to non-collaborative ones. For instance, if the address list of a publication contains five addresses and two of these addresses belong to a particular university, then the publication has a weight of 2 / 5 = 0.4 in the calculation of the indicators for this university. The fractional counting method leads to a more proper field normalization of impact indicators and to fairer comparisons between universities active in different fields. Fractional counting is therefore regarded as the preferred counting method in the Leiden Ranking. Collaboration indicators are always calculated using the full counting method.

Stability intervals

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 MNCS indicator may be equal to 1.50 for a particular university, with a stability interval from 1.40 to 1.65. This means that the true value of the MNCS indicator equals 1.50 for this university, but that changes in the set of publications of the university may relatively easily lead to MNCS values in the range from 1.40 to 1.65. The Leiden Ranking employs 95% stability intervals constructed using a statistical technique known as bootstrapping.

More information

More information on the Leiden Ranking methodology can be found in a number of publications by CWTS researchers. An extensive discussion of the Leiden Ranking is offered by Waltman et al. (2012). This publication relates to the 2011/2012 edition of the Leiden Ranking. Although not entirely up to date anymore, the publication still provides a lot of relevant information on the Leiden Ranking. The bottom-up approach taken in the Leiden Ranking to define scientific fields is described in detail by Waltman and Van Eck (2012). The methodology adopted in the Leiden Ranking for identifying core journals is outlined by Waltman and Van Eck (2013a, 2013b).

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., Van Eck, N.J., Van Leeuwen, T.N., Van Raan, A.F.J., Visser, M.S., & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, *63*(12), 2419-2432. DOI: 10.1002/asi.22708.

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.

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.