Ranking and mapping of universities and research-focused institutions worldwide: The third release of excellencemapping.net

Lutz Bornmann¹, Moritz Stefaner², Felix de Moya Anegón³, & Rüdiger Mutz⁴

¹ bornmann@gv.mpg.de

Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Munich, Germany

² moritz@stefaner.eu

Eickendorfer Damm 35, 28865 Lilienthal, Germany

³ felix.moya@scimago.es

CSIC, Institute of Public Goods and Policies (IPP), Madrid, Spain

⁴ mutz@gess.ethz.ch

Professorship for Social Psychology and Research on Higher Education, ETH Zurich, Zurich, Switzerland

Abstract

Bornmann, Stefaner, de Moya Anegón, and Mutz (2014a) have introduced a web application (www.excellencemapping.net) which is linked to both academic ranking lists published hitherto (e.g. the Academic Ranking of World Universities) as well as spatial visualization approaches. The application visualizes the performance of institutions worldwide within specific subject areas (journal sets from Scopus) as institutional ranking lists and on custom tile-based maps. Scopus data were used which are also the base for the SCImago Institutions Ranking. The second, enhanced version of the excellencemapping.net is described in Bornmann, Stefaner, de Moya Anegón, and Mutz (2014b). In this version, the effect of covariates (such as the perceived corruption in a country in which a research institution is located) on two performance metrics (best paper rate and best journal rate) is examined and visualized. This paper describes the third version based on new data.

Introduction

In a list of the most prominent rankings Hazelkorn (2013) names 11 different international rankings. The most important source of data used for the various rankings are abstract and citation databases of peer-reviewed literature (primarily Scopus, which is provided by Elsevier and the Web of Science, WOS, from Thomson Reuters). Publication and citation data is used to make a statement about the productivity and the citation impact of institutions (Bornmann, de Moya Anegón, & Leydesdorff, 2012; Waltman et al., 2012). Recent years have seen a number of different approaches which not only put institutions in a ranking list, but also show their performance on a map (Zhang, Perra, Gonçalves, Ciulla, & Vespignani, 2013). The advantage of this visualization is that regions and countries can be explored and searched for excellent institutions (Frenken, Hardeman, & Hoekman, 2009; van Noorden, 2010).

Bornmann, et al. (2014a) have introduced a web application (www.excellencemapping.net) which is linked to both academic ranking lists published hitherto (e.g. the Academic Ranking of World Universities) as well as spatial visualization approaches. The application visualizes the performance of scientific institutions worldwide (universities or research-focused institutions) within specific subject areas (Scopus journal sets) as ranking lists and on custom tile-based maps. The second, enhanced version of the application is described in Bornmann, et al. (2014b). In this version, the effect of covariates (such as the perceived corruption in a country in which an institution is located) on two bibliometric metrics (best paper rate and best journal rate) is visualized. A covariate-adjusted ranking and mapping of the institutions

was generated in which the single covariates are held constant. This paper describes the third version which is similar to the second. The third version is based on new data.

Methods

The study is based on Scopus data which are also annually used for the SCImago Institutions Ranking (<http://www.scimagoir.com/>). In order to include in the application reliable data in terms of geo-coordinates (Bornmann, Leydesdorff, Walch-Solimena, & Ettl, 2011) and performance metrics (Waltman, et al., 2012), we only selected those institutions which have published at least 500 papers (articles, reviews and conference papers) between 2007 and 2011 in a certain subject area (Scopus journal set) in the study.¹ In other words, institutions with fewer than 500 papers in a category are not included in the application. Furthermore, only subject areas offered at least 50 institutions are considered in the application (Arts and Humanities, e.g., is not considered). This threshold was used to have no small number of institutions for a worldwide comparison. We applied the full counting method (Vinkler, 2010) to assign papers from Scopus to institutions: if an author of a paper is associated with an institution, it is assigned to this institution (with a weight of 1).

Table 1. Number of institutions included in the analyses for 17 different subject areas. The mean best paper rate/best journal rate is the mean best paper rate/best journal rate for the institutions within a subject area.

<i>Subject area</i>	<i>Number of institutions</i>	<i>Mean best paper rate</i>	<i>Mean best journal rate</i>
Agricultural and Biological Science	573	0.15	0.54
Biochemistry, Genetics and Molecular Biology	808	0.14	0.49
Chemical Engineering	181	0.13	0.52
Chemistry	541	0.12	0.57
Computer Science	406	0.14	0.29
Earth and Planetary Sciences	363	0.17	0.62
Engineering	669	0.13	0.34
Environmental Science	267	0.16	0.65
Immunology and Microbiology	239	0.16	0.58
Materials Science	444	0.13	0.47
Mathematics	398	0.14	0.40
Medicine	1309	0.17	0.52
Neuroscience	133	0.17	0.54
Pharmacology, Toxicology and Pharmaceutics	102	0.18	0.60
Physics and Astronomy	702	0.16	0.56
Psychology	81	0.19	0.54
Social Sciences	235	0.17	0.47

¹ The first and second versions of the web application considered publications from 2005 to 2009 and from 2006 to 2010, respectively. Both versions of the excellence mapping tool are still accessible in the current release.

Two indicators are used to measure the performance of the institutions. The first indicator has named as “best paper rate”. It is the proportion of the institutional publications which belong to the 10% most frequently cited publications in their subject area and publication year. The second indicator – named as the best journal rate – is the proportion of publications which an institution publishes in the most influential journals worldwide. The most influential journals are those which are ranked in the first quartile (top-25%) of their subject areas as ordered by the indicator SCImago Journal Rank (SJR). While the indicator “best paper rate” informs about the long-term success of publications, the indicator “best journal rate” reflects the ability of an institution to publish its publications in reputable journals.

Table 1 shows the number of institutions which are included as datasets for the 17 Scopus subject areas in excellencemapping.net. Out of the total of 27 subject areas in Scopus, only those are selected for the web application which have at least 50 institutions worldwide. For example, 541 institutions within the subject area of chemistry were included in the analyses. The mean best paper rate for these institutions is .12 (12%) and the mean best journal rate .57 (57%).

The citation impact of the publications considered in the web application refers to the time period from publication to the beginning of 2014.

The data were analysed by using generalized linear mixed model for binomial data, which takes into account the hierarchical structure of data and properly estimates the standard errors (Bornmann, et al., 2014a; Mutz & Daniel, 2007).

The web application was implemented using modern web technologies and Open Street Map² data provided through MapBox³. It is based on the javascript frameworks backbone.js⁴, jquery⁵ and d3.js⁶.

Results

To be able to explain the performance differences among the institutions in the regression model, we included the following variables as covariates. The covariates are also utilized to create a covariate-adjusted ranking of the institutions:

- (1) Proportion of papers from one institution which were produced in an international collaboration
- (2) Corruption perception index
- (3) Number of residents in a country
- (4) Gross domestic product (GDP) per capita of a country

While the international collaboration covariate relates to individual institutions, all the other covariates apply at the level of individual countries.

Figure 1 shows a screen shot of the excellencemapping.net, which visualizes the results of the statistical analyses for 17 subject areas. There is a short description of the web application displayed in the upper right section of the excellencemapping.net with a link to click for "More information". The page with the detailed description includes the affiliations of the authors of the web application and a link to this research paper. Below the (short) description, the user can select whether to visualize the first release of the excellencemapping.net, which refers to the time period from 2005 to 2009, the second release with the time period from 2006 to 2010, or the third release with the time period from 2007 to 2011. Only in the second and third release, the user can choose two different performance metrics and several

² <http://www.openstreetmap.org>

³ <https://www.mapbox.com>

⁴ <http://backbonejs.org>

⁵ <http://jquery.org>

⁶ <http://d3js.org>

covariates. In the upper right section of excellencemapping.net, the user can select a subject area for visualization. Below the subject area selection window, there is another selection window for the covariate (for selecting the number of residents, for example).

If a covariate is selected, the probabilities of (i) publishing in reputable journals (best journal rate) or (ii) publishing most-frequently cited papers (best paper rate) is visualized adjusted (controlled) for the selected covariate. Then, the displayed performance of institutions can be interpreted as if all institutions had the same value (reference point) for the selected covariate. We z-transformed each covariate over the whole data set (with $M=0$ and $S=1$). Thus, the average probability is the value in which the selected covariate=0, i.e. equivalent to the median. The z-transformation allows the comparison of the results from models with and without the covariates.

Figure 1. Screen shot of excellencemapping.net.

One can select one of the two performance indicators (best paper rate or best journal rate) below the selection windows for the subject area and the covariates. The application shows for each indicator the residues from the regression analysis (random effects) converted to probabilities. To have values on the original scale for both performance indicators (i.e. proportion of papers in the top range or published in the top journals), the intercept was added to the residues. One can tick “Show statistically significant results only” to reduce the number of mapped institutions in a subject area to only those institutions which differ statistically significantly in their performance from the mean value.

The map on the left-hand side of the web application shows a node for each institution for a selected subject area (e.g. Physics and Astronomy). One can move the map to different regions worldwide and can zoom in (or out). Map details appear only at zoom levels of a certain depth. The node size for each institution is proportional to the number of papers in the selected subject area. For example, the Centre National de la Recherche Scientifique (CNRS) has the largest node (in Europe) in Physics and Astronomy. As many nodes overlap on larger

cities (e.g. Paris), one can select all the nodes in a certain region. These institutions are then listed on the right-hand side of the application under "Your selection". The colour of the nodes indicates the indicator values for the institutions using a colour scale from blue through grey to red: If the institutional indicator value is greater than the mean (expected) value across all institutions in a subject area, its node has a blue tint. Nodes with red colours mark institutions with indicator values lower than the mean. Grey nodes indicate values close to the expected value.

All institutions, which are taken into account in the regression model for a subject area (see section "Institutional scores"), are displayed on the right-hand side of excellencemapping.net. The name, the corresponding country, and the number of papers ("Papers") are shown for each institution. Furthermore, the best journal rate or the best paper rate is listed ("Indicator value"). The greater the confidence interval of the probability, the more unreliable it is. If the confidence interval for an institution does not overlap with the mean proportion across all institutions (the mean is the short line in the middle of "Indicator value"), the institution has published a statistically significantly higher (or lower) best paper or best journal rate than the average across all institutions ($\alpha = 0.165$). The listed institutions can be sorted (in descending or ascending order in the case of numbers). Thus, the top or worst performing institutions in a subject area can be identified by clicking on "Indicator value." Clicking on "Papers" puts the institutions with a high paper output at the top of the list (or at the end). In Biochemistry, Genetics and Molecular Biology, for example, CNRS is the institution with the highest paper output between 2007 and 2011; in terms of the best paper rate, Broad Institute of MIT and Harvard are the best-performing institution. The column farthest on the right (named as " Δ rank") in the section "Institutional Scores" shows for an institution by how many rank places it goes up (green, arrow pointing upwards) or goes down (red, arrow pointing downwards), if one selects a certain covariate. For example, the Institute for High Energy Physics (RUS) improves its position by 5 places compared to the ranking which does not take the covariate "corruption perception index" into account in the Physics and Astronomy subject area. The ranking differences in this column always relate to all the institutions included. The differences do not therefore change if one looks at only the statistically significant results.

If a covariate has been selected, one can, for example, sort the institutions by Δ rank. This puts the institutions which benefit most from the covariate at the top of the list. Using the search field at the right, one can find a specific institution in a subject area. In order to identify the institutions for a country, one can click on "Country". Then, the institutions are sorted by country and – within a country – by the indicator value. The section "Your selection" is intended for the comparison of institutions which are of interest for the user.

If the confidence intervals of two institutions do not overlap, they differ statistically significantly on the 5% level (in the best paper or best journal rate, respectively). For example, in Physics and Astronomy, Stanford University and the Helmholtz Association are shown without overlap (publication years 2007 to 2011).

The institutions in "Your selection" can be sorted by each heading. These institutions are also marked on the map on the left-hand side with a black border. Thus, the section "Your selection" links both, institutional lists and institutional maps. For the institutional comparison, one can select them in the list or on the map. By clicking on "Clear", one can start a new comparison of institutions.

Discussion

Following Bornmann and Leydesdorff (2011), our underlying statistics (multi-level regression models) are analytically oriented. They enable (1) the estimation of statistically more appropriate values for the performance indicators than the observed values; (2) the calculation

of confidence intervals which can be used as measures of reliability for the institutional performance; (3) the comparison of an institution with an “average” institution in a specific subject area and (4) the direct comparison of two or more institutions. (5) By taking covariates into account when mapping and ranking institutions an adjusted view of institutional research performance is possible. For example, with our application it is possible to look at the performance of institutions worldwide in countries with the same financial background (that is, the same GDP). This highlights institutions showing a relatively high performance despite a bad financial situation in the country (Bornmann, 2014).

Conclusions

The second, substantially enhanced version of the web application visualizes institutional performance as ranking lists and on custom tile-based maps. The effect of covariates (such as the perceived corruption in a country) on two performance metrics (best paper rate and best journal rate) is visualized. The web application can be used, e.g., by students, (postdoctoral) researchers and people working in the area of science policy to identify scientifically interesting institutions worldwide.

References

- Bornmann, L. (2014). Ranking institutions by the handicap principle. *Scientometrics*, 100(2), 603-604.
- Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2012). The new Excellence Indicator in the World Report of the SCImago Institutions Rankings 2011. *Journal of Informetrics*, 6(2), 333-335. doi: 10.1016/j.joi.2011.11.006.
- Bornmann, L., & Leydesdorff, L. (2011). Which cities produce more excellent papers than can be expected? A new mapping approach—using Google Maps—based on statistical significance testing. *Journal of the American Society of Information Science and Technology*, 62(10), 1954-1962.
- Bornmann, L., Leydesdorff, L., Walch-Solimena, C., & Ettl, C. (2011). Mapping excellence in the geography of science: an approach based on Scopus data. *Journal of Informetrics*, 5(4), 537-546.
- Bornmann, L., Stefaner, M., de Moya Anegón, F., & Mutz, R. (2014a). Ranking and mapping of universities and research-focused institutions worldwide based on highly-cited papers: A visualization of results from multi-level models. *Online Information Review*, 38(1), 43-58.
- Bornmann, L., Stefaner, M., de Moya Anegón, F., & Mutz, R. (2014b). What is the effect of country-specific characteristics on the research performance of scientific institutions? Using multi-level statistical models to rank and map universities and research-focused institutions worldwide. *Journal of Informetrics*, 8(3), 581-593.
- Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: towards a cumulative research program. *Journal of Informetrics*, 3(3), 222-232. doi: 10.1016/j.joi.2009.03.005.
- Hazelkorn, E. (2013). Reflections on a decade of global rankings: what we've learned and outstanding issues. *Beiträge zur Hochschulforschung*, 35(2), 8-33.
- Mutz, R., & Daniel, H. D. (2007). Development of a ranking procedure by mixed Rasch model and multilevel analysis - psychology as an example. *Diagnostica*, 53(1), 3-16. doi: 10.1026/0012-1924.53.1.3.
- van Noorden, R. (2010). Cities: building the best cities for science. Which urban regions produce the best research - and can their success be replicated? *Nature*, 467, 906-908. doi: 10.1038/467906a.
- Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Oxford, UK: Chandos Publishing.
- Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., . . . 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.

Zhang, Q., Perra, N., Gonçalves, B., Ciulla, F., & Vespignani, A. (2013). Characterizing scientific production and consumption in Physics. *Sci. Rep.*, 3. doi: 10.1038/srep01640.