Downloaded by Indian Institute of Technology Kharagpur At 02:35 10 May 2018 (PT)

Online Information Review

Ranking and mapping of universities and research-focused institutions worldwide based on highly-cited papers: A visualisation of results from multi-level models

Lutz Bornmann, Moritz Stefaner, Felix de Moya Anegón, Rüdiger Mutz,

Article information:

To cite this document:
Lutz Bornmann, Moritz Stefaner, Felix de Moya Anegón, Rüdiger Mutz, (2014) "Ranking and mapping of universities and research-focused institutions worldwide based on highly-cited papers: A visualisation of results from multi-level models", Online Information Review, Vol. 38 Issue: 1, pp.43-58, https://doi.org/10.1108/OIR-12-2012-0214
Permanent link to this document:
https://doi.org/10.1108/OIR-12-2012-0214

Downloaded on: 10 May 2018, At: 02:35 (PT)
References: this document contains references to 49 other documents.
To copy this document: permissions@emeraldinsight.com
The fulltext of this document has been downloaded 933 times since 2014*

Users who downloaded this article also downloaded:

(2012),"Exploring the h-index at the institutional level: A practical application in world university rankings", Online Information Review, Vol. 36 Iss 4 pp. 534-547 https://doi.org/10.1108/14684521211254059

(2015),"Behind league tables and ranking systems: A critical perspective of how university quality is measured", Journal of Service Theory and Practice, Vol. 25 Iss 3 pp. 242-266 https://doi.org/10.1108/JSTP-04-2013-0059

Access to this document was granted through an Emerald subscription provided by emerald-srm:395687 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Ranking and mapping of universities and research-focused institutions worldwide based on highly-cited papers

A visualisation of results from multi-level models

Received 9 January 2013
First revision accepted
28 February 2013
Second revision
accepted March 2013

Lutz Bornmann
Division for Science and Innovation Studies,
Administrative Headquarters of the Max Planck Society, Munich, Germany
Moritz Stefaner
Independent Researcher, Lilienthal, Germany
Felix de Moya Anegón
Instituto de Políticas y Bienes Públicos,
Consejo Superior de Investigaciones Científicas, Madrid, Spain, and
Rüdiger Mutz
Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

Abstract

Purpose – The web application presented in this paper allows for an analysis to reveal centres of excellence in different fields worldwide using publication and citation data. Only specific aspects of institutional performance are taken into account and other aspects such as teaching performance or societal impact of research are not considered. The purpose of this paper is to address these issues.

Design/methodology/approach – Based on data gathered from Scopus, field-specific excellence can be identified in institutions where highly-cited papers have been frequently published.

Findings – The web application (www.excellencemapping.net) combines both a list of institutions ordered by different indicator values and a map with circles visualising indicator values for geocoded institutions.

Originality/value – Compared to the mapping and ranking approaches introduced hitherto, our underlying statistics (multi-level models) are analytically oriented by allowing the estimation of values for the number of excellent papers for an institution which are statistically more appropriate than the observed values; the calculation of confidence intervals as measures of accuracy for the institutional citation impact; the comparison of a single institution with an "average" institution in a subject area: and the direct comparison of at least two institutions.

Keywords Geography of science, Google Maps, Highly-cited papers, Scientific excellence, Spatial scientometrics, University ranking

Paper type Research paper

Online Information Review
Vol. 38 No. 1, 2014
pp. 43-58

1468-4527
DOI 10.1108/OIR-12-2012-0214

Introduction

There is a growing interest in national and international comparisons of research organisations and urban areas in terms of scientific output and impact. A sign of this trend is the continuous publication of university rankings, both inside but also outside the scientific environment (Shin *et al.*, 2011). For example the SCImago Research Group of the University of Granada in Spain (2012) publishes annually an international ranking of more than 3,000 research institutions and organisations. The reports show indicator values (e.g. publication output, relative citation rate, or excellence rate) based on publication and citation data from Scopus (Elsevier) for larger research-focused institutions (see here Scimago Reseach Group, 2011). The Leiden Ranking (www.leidenranking.com/) measures the scientific performance of 500 major universities worldwide (Leydesdorff and Bornmann, 2012b; Waltman *et al.*, 2012). In both rankings, institutions and organisations are listed one after another and can be directly compared in terms of different indicators for productivity, research impact, specialisation, and collaboration. Other academic rankings basically follow the same approach but use more or other indicators (see overviews in Kroth and Daniel, 2008; Buela-Casal *et al.*, 2007; Shin *et al.*, 2011).

According to Tijssen *et al.* (2002), Tijssen and Van Leeuwen (2006) as well as Waltman *et al.* (2012), excellent or highly-cited papers are those among the 10 per cent most-cited papers in a field (papers in or greater than the 90th percentile, referred to in the following as class 10 per cent papers). According to Waltman *et al.* (2012) the number of class 10 per cent papers is the most important impact indicator for the ranking of universities by research performance. Some different approaches have been published recently which visualise the number of excellent papers for locations of research on Google Maps instead of the presentation of the numbers in long ranking lists. Frenken *et al.* (2009) suggest grouping such approaches to mapping the geography of science under the heading "spatial scientometrics". The visualisations identify regions of excellent research and allow the comparison of excellent output in regions worldwide. Leydesdorff and Persson (2010) explore the use of Google Maps and Google Earth for generating spatial maps of science. Bornmann *et al.* (2011a) present methods to map centres of scientific excellence around the world. By colouring cities worldwide according to the output of excellent papers, their maps provide visualisations where cities with a high (or low) output of these papers can be found. Bornmann and Waltman (2011) follow their approach in general, but change the focus from mapping of single cities to a more "sliding" visualisation of broader regions. The maps generated by Bornmann *et al.* (2011a) and Bornmann and Waltman (2011) for different scientific fields point out a spatial concentration of excellent research which might possibly be explained by the fact that more competitors (here: prolific scientists) working within the same region produce better results (Bornmann *et al.*, 2011a).

The most recent stage of the developments in "spatial scientometrics" is the approach of Bornmann and Leydesdorff (2011). They consider it a disadvantage that only the output of excellent papers is used in the other approaches. If a city has a very high output in general (i.e. not only of excellent papers) one can expect a high number of excellent papers proportionally (Bornmann *et al.*, 2011b, c). Therefore the observed output should be compared with an expected output. For example if authors at a university have published 10,000 papers, one would expect for statistical reasons that approximately 1,000 (that is, 10 per cent) would also belong to the class 10 per cent papers. An observed number of 700 highly-cited papers for this university may seem a large number compared to other universities, but it turns out to be smaller than one would expect in this case. The *z* test for comparing a population proportion with a sample proportion (Sheskin, 2007) can be used for evaluating the degree to which an observed number of class 10 per cent papers differs from the expected value (Bornmann *et al.*, 2012).

In this paper we introduce a new web application which is linked to both spatial visualisation approaches as well as academic ranking lists published hitherto. It tries to capture the advantages of both approaches for presenting the research performance of locations; lists and maps are visualised and intertwined. In ranking lists one can immediately see the best and worst institutions (institutions in the first and last positions). However it is difficult to directly compare institutions holding very different positions or which are located in a single region (e.g. Europe). In contrast performance indicators visualised on Google maps allow the focus on institutions in certain regions, but it is difficult to identify the best and worst locations worldwide. The web application introduced here visualises institutional performance within specific subject areas as ranking lists and on custom tile-based maps. In contrast to many other university rankings which present the results across all fields of science (e.g. the Leiden Ranking), the lists and maps shown in the web application are differentiated for subject categories.

Based on the developments of Bornmann and Leydesdorff (2011) we compare in this study the number of observed with the number of expected papers belonging to class 10 per cent within their field category for universities and research-focused institutions (referred to as institutions in the following) around the world. Bornmann and Leydesdorff (2011) conduct a single test for each city worldwide to analyse the statistical significance of the difference between observed and expected numbers (Leydesdorff and Bornmann, 2012b; Bornmann *et al.*, 2012). In this study a multi-level logistic regression analysis is calculated to analyse the differences between both numbers for all the institutions within one model. For each institution the difference between its performance and the average over all institutions in a field is tested. The model allows the calculation of shrinkage estimates and corresponding standard errors which are more precise than raw probabilities (empirical Bayes estimates) and their standard errors, especially if the information for an institution is sparse (e.g. if its publication output is low). The estimated standard errors and corresponding confidence interval of the regression model takes design effects into account which are on average higher than the corresponding standard errors and confidence intervals obtained in a sampling procedure which does not consider any clusters, i.e. institutions (Hox, 2010). Additionally multi-level models provide a very easy way to compare institutions, that is, whether they differ statistically significantly in their probabilities of having published excellent papers.

Methods

Data sets

The study is based on Scopus data (Elsevier) which has been collected for the SCImago Institutions Ranking (www.scimagoir.com/). To obtain reliable data in terms of geo-coordinates (see Bornmann *et al.*, 2011a) and the number of excellent papers (Waltman *et al.*, 2012), we considered in the study only those institutions that have published at least 500 journal papers, reviews and conference papers in the period 2005 to 2009 in a certain Scopus subject area. Institutions with fewer than 500 papers in a category were not considered. Furthermore only subject categories offered at least 50 institutions were included in the web application (e.g. arts and humanities is not included). We used this threshold to have a considerable number of institutions for a worldwide comparison. The full counting method was used (Vinkler, 2010) to attribute papers from the Scopus database to institutions: if an institution appears in the affiliation field of a paper, it is attributed to this institution (with a weight of 1). According to the results obtained by Waltman *et al.* (2012) the overall correlation between a university ranking based on the full counting and fractional counting method is very high ($r = 0.97$). The fractional counting method gives less weight (< 1) to collaborative than to non-collaborative papers ($= 1$). Table I shows the number of institutions which were considered as data sets for the 17 subject areas in this study. Out of the 27 available subject areas in Scopus, only those which include at least 50 institutions worldwide were selected for the study.

When evaluating the citation impact of publications there is the possibility of including or excluding authors' self-citations (Bornmann *et al.*, n.d., in press-a). Studies have reported different percentages of self-citations: for example in a study on researchers in Norway, Aksnes (2003) found 36 per cent; Snyder and Bonzi (1998) showed that the percentage of self-citations in natural sciences is 15 per cent, higher than in social sciences (6 per cent) and arts and humanities (6 per cent). In this study we

Table I.
Number of universities and research-focused institutions included in the statistical analyses for 17 different subject areas

Subject area	No. of universities or research-focused institutions	Mean % of highly-cited papers (class 10% papers)
Agricultural and biological science	504	0.15
Biochemistry, genetics and molecular biology	746	0.13
Chemical engineering	148	0.14
Chemistry	496	0.13
Computer science	350	0.14
Earth and planetary sciences	318	0.17
Engineering	594	0.14
Environmental science	215	0.17
Immunology and microbiology	204	0.16
Materials science	367	0.14
Mathematics	362	0.14
Medicine	1,175	0.17
Neuroscience	108	0.17
Pharmacology, toxicology and pharmaceutics	86	0.17
Physics and astronomy	650	0.14
Psychology	59	0.20
Social sciences	166	0.19

Note: The mean percentage of highly-cited papers is the mean over the percentages of class 10 percent papers for the institutions within one subject area included self-citations for two reasons. The first is that it was expected that the percentage of self-citations would not differ significantly among the different authors at the institutions. The percentage of self-citations will vary among the authors (and the publications), but in most cases it was not expected that institutions conducting research in similar areas would have very different self-citation percentages. The second reason is that following Glänzel *et al.* (2006), self-citations are usually an important feature of the science communication and publication process: "A self-citation indicates the use of own results in a new publication. Authors do this quite frequently to build on own results, to limit the length of an article by referring to already published methodology, or simply to make own background material published in 'grey' literature visible" (p. 265).

Percentile calculation

To identify the class 10 per cent papers within a subject area, the citations X_i (citation window: from publication until the end of 2011) that were received by the ith papers within n papers published in a given subject area (and publication year as well as a given document type) were gathered. Then the papers were ranked in increasing order

$$X_1 \leq X_2 \leq \ldots \leq X_n,$$

where X_1 and X_n denote the number of citations received respectively by the least and most cited paper. Where citation counts were equal, the SJR2 (Guerrero-Bote and De Moya-Anegon, 2012) of the journal which published the papers was used as a second sort key (from highest to lowest). This journal metric takes into account not only the prestige of the citing scientific publication but also its closeness to the cited journal. Finally in each field (publication year and document type), each individual publication was assigned a percentile rank based on this distribution. If, for example, a single paper within a subject area had 50 citations, and this citation count was equal to or greater than the citation counts of 90 per cent of all papers in the subject area, then the percentile rank of this paper would be 90. The paper would be in the 90th percentile and would belong to the class 10 per cent papers within the subject area. There are different approaches available for calculating percentile-based indicators (see an overview in Waltman and Schreiber, 2013; Bornmann *et al.*, 2013a). The approach used for the SCImago Institutions Ranking is comparable to the approach proposed by Rousseau (2012).

In Table I the mean percentage of highly-cited papers for the institutions included in this study is the mean over the percentages of class 10 per cent papers for the single institutions within one subject area. For example physics and astronomy consists of 650 different institutions with a mean proportion of excellent papers of 0.14. Three reasons can be given for the fact that the mean average for physics and astronomy (as well as all other subject areas in the table) is higher than 10 per cent: ties in citation data lead to a higher number of class 10 per cent papers (Leydesdorff *et al.*, 2011; Waltman and Schreiber, 2013); the highly-selected set of institutions considered here (institutions with at least 500 publications) has published more class 10 per cent papers than institutions not considered; and "First, collaborative publications are counted multiple times in the full counting method, and second, collaborative publications tend to be cited more frequently than non-collaborative publications" (Waltman *et al.*, 2012, p. 2427).

Statistical model

The choice of the statistical procedure to analyse the data depends strongly on the scale of the dependent variable (e.g. ordinal, continuous). In our case the dependent variable is dichotomous: a paper published by an author located in an institution belongs to the class 10 per cent papers or not. The relative frequency of the papers in the class 10 per cent for an institution is an estimate of its probability of class 10 per cent papers. The simplest way for the statistical analysis is to report these probabilities of class 10 per cent papers for each institution (see Waltman *et al.*, 2012). However this procedure is not statistically appropriate and leads to incorrect solutions, because the hierarchical structure of the data (papers, level 1, are nested within institutions, level 2) is not taken into account. We can assume that papers published by authors within one institution are somewhat more homogeneous regarding their probability of being class 10 per cent papers than papers published by authors located in different institutions. The homogeneity reduces the effective sample size and increases the standard errors.

We prefer a multi-level logistic regression intercept-only model for binary outcomes, which properly estimates the standard errors. Not only the standard errors of the regression parameter, but also the size of the standard error of the estimated class 10 per cent probabilities (might) differ from those of a one-level model with consequences for the statistical comparison of institutions. Another great advantage of multi-level modelling is that the statistical results can be summarised with a small set of parameters. For instance one parameter allows us to test statistically whether the institutional performances vary only randomly (i.e. as random samples of the same population) or systematically. Only in the case of systematic differences between institutions are comparisons reasonable.

In multi-level logistic regression, papers are clustered within universities, whereas j ($j = 1, \ldots, N$) denotes the level-2 units ("institutions") and i ($i = 1, \ldots, n_j$) the level-1 units ("papers"). Due to the fact that the dependent variable x_{ji} is dichotomous ($1 =$ paper i belongs to the class 10 per cent publications, $0 =$ paper i does not belong to the class 10 per cent publications), ordinary multi-level models for continuous data are not appropriate. Therefore generalised linear mixed models are favoured, especially the multi-level logistic model for binary data, which comprises three components (Hox, 2010):

(1) The probability distribution for p_{ji} ($= \Pr(x_{ji} = 1)$) is a Bernoulli distribution $(1, \mu)$ with mean μ.

(2) A linear multi-level regression part with a latent (unobserved) predictor η_{ji} of the binary outcome x_{ji}: $\eta_{ji} = \beta_0 + u_{0j}$, where u_{0j} is a normally distributed random effect $u_{0j} \sim N(0, \sigma^2_{u0})$ with the variance σ^2_{u0},

(3) A link function connects the expected value of the dependent variable x with the latent predictor η, which is here the logit function: $\eta = \text{logit}(\mu) = \log(\mu/(1-\mu))$. Probabilities which range between 0 and 1 are transformed by the logit link function to logits, which continuously vary between $-\infty$ and $+\infty$ with a variance of $\pi^2/3 = 3.29$.

The multi-level logistic model for the observed proportions p_j of papers which belong to the class 10 per cent publications can be formulated as follows (Snijders and Bosker, 2004):

$$p_j = \text{logistic}\left(\beta_0 + u_{0j}\right) u_{0j} \sim N\left(0, \sigma_{u0}^2\right), \qquad (1)$$

where “logistic” means the logistic transformation of p_j ($\text{logistic}(x) = e^x/(1 + e^x)$), which is the inverse logit link function. There is a intra-class correlation between papers within institutions with $\rho = \sigma_{u0}^2/(3.29 + \sigma_{u0}^2)$ which reflects the homogeneity of papers within an institution. The Wald test allows us to test whether σ_{u0}^2 deviates from 0 (the null hypothesis). If the Wald test is statistically significant, the institutions systematically vary with respect to their number of class 10 per cent papers. Then a ranking or comparison of institutions is reasonable. Covariates can be included in the model in order to control, for instance, for socio-economic differences between countries (Bornmann *et al.*, 2013b).

Most importantly the multi-level model allows the calculation of so-called Empirical Bayes (EB) or shrinkage estimates which are more precise than their empirical counterparts, the raw probabilities. The following information is considered in the calculation of EB. First if there is no further information for an institution, the mean value (i.e. mean probability) of class 10 per cent papers across all institutions is the best estimate. Second the more reliable the information for an institution (i.e. the greater the variance between institutions σ_u^2 and the higher the total number of papers for the institution under consideration), the more the raw probability of class 10 per cent papers is the best estimate for this institution. The EB, therefore, vary between the mean value of class 10 per cent papers across all institutions and the raw probability of class 10 per cent papers for a certain institution. If the sample size (number of papers) for an institution is low, the EB is shrunken towards the mean value. The estimated standard errors take design effects into account (SAS Institute Inc., 2008). They are different from the corresponding standard errors and confidence intervals obtained in a sampling procedure which does not consider any clusters, especially where the sample size of level-2 units is small.

The EB and the confidence intervals can be transformed back to probabilities to facilitate the interpretation of the results. The multiplication of standard errors by 1.39 instead of 1.96 results in Goldstein-adjusted confidence intervals (Goldstein and Healy, 1994) with the property that if the confidence intervals of two institutions do not overlap, they differ statistically significantly ($\alpha = 5$ per cent) in their estimates (i.e. class 10 per cent papers’ probabilities). If the 95 per cent confidence interval does not include the mean proportion of class 10 per cent papers across all institutions, the authors located at this institution have published a statistically significantly greater or smaller number of class 10 per cent papers than the average across all institutions. The Goldstein-adjusted confidence interval test can only be done on the 16.3 per cent probability level, rather than on the usual 5 per cent level.

The power of statistical tests is defined as the probability of rejecting the null hypothesis in the case that is actually true. Based on simulation studies Moineddin *et al.* (2007) recommended at least 100 groups for a multi-level logistic regression with a group size of 50 for an acceptable power. Except for the subject areas psychology ($n = 59$) as well as pharmacology, toxicology and pharmaceutics ($n = 86$) all subject areas in this study significantly exceed this threshold for the number of groups (here: institutions). With respect to group size (here: publications) all subject areas exceed the threshold of 50. We did not perform a power analysis here, because we reanalysed with the publication and citation data observed data. Retrospective or *post hoc* power analyses do not provide useful and valid estimators of the true power (Levine and Ensom, 2001; Yuan and Maxwell, 2005). Simulation studies have shown that "a low power does not always indicate that the test is unpowered" (Sun *et al.*, 2011, p. 81). Against this backdrop we followed Levine and Ensom (2001) who pointed out that "confidence intervals better inform readers about the possibility of inadequate sample size than do *post hoc* power calculations" (p. 405).

The analyses for this study were calculated using the proc glimmix procedure implemented in the SAS statistical software (SAS Institute Inc., 2008).

Programming of the visualisation

We have tried "to make a visualisation that is attractive and informative, and yet conveys its own contingency and limitations" (Spiegelhalter *et al.*, 2011, p. 1400). The following data and results of 17 multi-level regression models (one model for each subject area) have been used as input for the visualisation: number of papers published by authors located in an institution; number of papers for an institution belonging to the class 10 per cent papers in a subject area; the "true" proportion of class 10 per cent papers for an institution as the result of the multi-level model; the proportion's confidence interval (lower and upper limits); and whether an institution's proportion differs statistically significantly from the mean over all institutions in a subject area (the expected value). To rank the institutions within a subject area, the logarithmised quotient of the "true" proportion of class 10 per cent papers for an institution and the "true" proportion across all institutions was calculated for each institution. The rationale for applying a logarithm is to provide comparable scales for values above and below the expected value; in other words, on our rank scale, an institution producing twice as many papers as expected is as far from the point for the expected value as an institution producing half as many as expected.

The maps used in the visualisation are custom-styled map tiles generated with TileMill (http://mapbox.com/tilemill/) based on Open Street Map (http://openstreetmap.org) data. For developing the data overlays, we used the polymaps library (http://polymaps.org/); the dynamic tables were compiled with the help of the DataTables jquery plugin (http://datatables.net/).

Results

Figure 1 shows a screen shot of the web application (www.excellencemapping.net) visualising the results of multi-level analyses for 17 different subject areas. The web application is password-protected: it can be used for research purposes only. The password can be received from the authors of this paper.

For a selected subject category (e.g. physics and astronomy), the map on the left-hand side of the screen shows a circle for each institution with a paper output greater than or equal to 500. Users can move the map to different regions by using the mouse (click and drag) and zoom in (or out) by using the mouse wheel. Country labels and map details appear only at zoom levels of a certain depth, primarily in order to facilitate perception of the data markers. Both moving and zooming can also be done by using the control buttons at the top left of the screen. The circle area for each institution on the map is proportional to the number of published papers in the respective subject area. For example, the Centre National de la Recherche Scientifique (CNRS) has the largest circle (in Europe) on the physics and astronomy map,

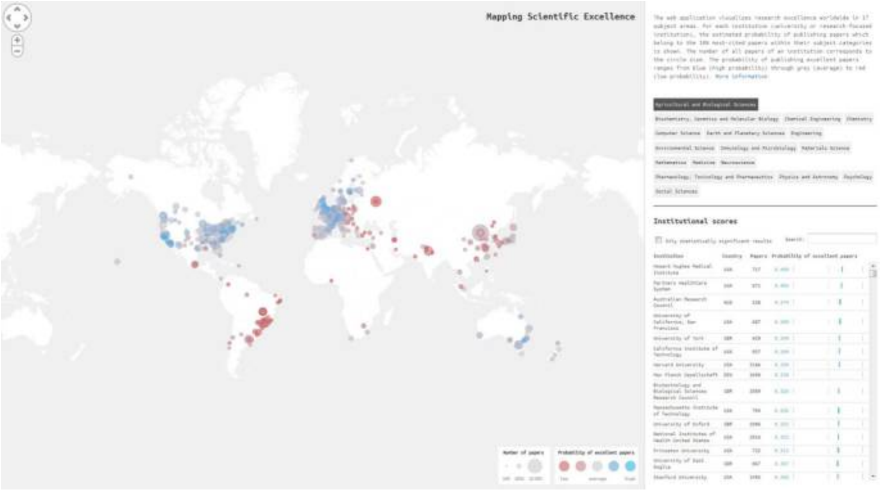

Figure 1.
Screen shot of the web application

highlighting the high output of papers in this subject area (see Figure 2). The circle colour indicates the proportion of class 10 per cent papers for the respective institution using a diverging colour scale, from blue through grey to red (without any reference to statistical testing). If the proportion of class 10 per cent papers for an institution is greater than the mean (expected) value across all institutions, its circle has a blue tint. Red circles mark institutions with proportions of class 10 per cent papers lower than the mean. Grey circles indicate a value close to the expected value.

On the right-hand side of the web application, all those institutions are listed which are considered in the multi-level model for a subject area (section "Institutional scores"). For each institution the name, the respective country, the number of all the papers published ("Papers"), and the EBs with confidence intervals are visualised ("Probability of excellent papers"). The greater the confidence interval, the more

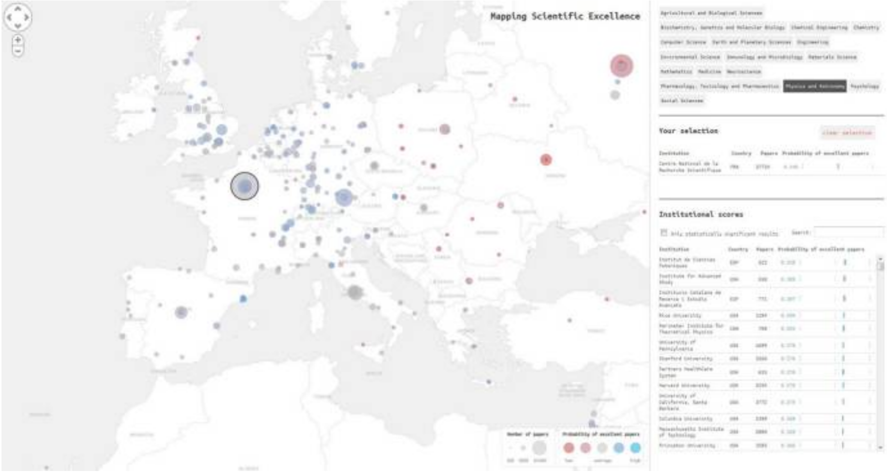

Figure 2.
Physics and astronomy map with the CNRS selected

unreliable the probability for an institution is. If the confidence interval does not overlap with the mean proportion of class 10 per cent papers across all institutions (the mean is visualised by the short line in the middle of "Probability of excellent papers"), the authors located at this institution have published a statistically significantly higher (or lower) number of class 10 per cent papers than the average across all the institutions ($\alpha = 0.165$). The institutions in the list can be sorted (in decreasing or increasing order in case of numbers) by clicking on the relevant heading. Thus the top or worst performers in a field can be identified by clicking on "Probability of excellent papers". Institutions with high productivity in terms of paper numbers appear at the top of the list (or at the end) by clicking on "Papers". In biochemistry, genetics and molecular biology, for example, the institution with the highest productivity between 2005 and 2009 is the CNRS; in terms of probabilities of class 10 per cent papers, the best-performing institution is the Broad Institute of MIT and Harvard (see Figure 3). To reduce the set of visualised institutions in a field to only those which differ statistically significantly in their performance from the mean value, the corresponding tick mark can be set by the user.

Using the search field at the top right, the user can find a specific institution. By clicking on a heading (e.g. "Papers") the reduced list is ordered accordingly. To identify the institutions for a specific country, click on "Country". Then the institutions are first sorted by country and second by the probability of excellent papers (in increasing or decreasing order). "Your selection" is intended to be the section for the user to compare institutions of interest directly. If the confidence intervals of two institutions do not overlap, they differ statistically significantly on the 5 per cent level in the probability of class 10 per cent papers' output. For example in physics and astronomy Stanford University and the Fraunhofer Gesellschaft are visualised without overlap (see Figure 4). The selected institutions in "Your selection" can be sorted by each heading in different orders. These institutions are also marked on the map with a black border. Thus both institutional lists and institutional maps are linked by the section "Your selection". For the comparison of different institutions, it is not only possible to select

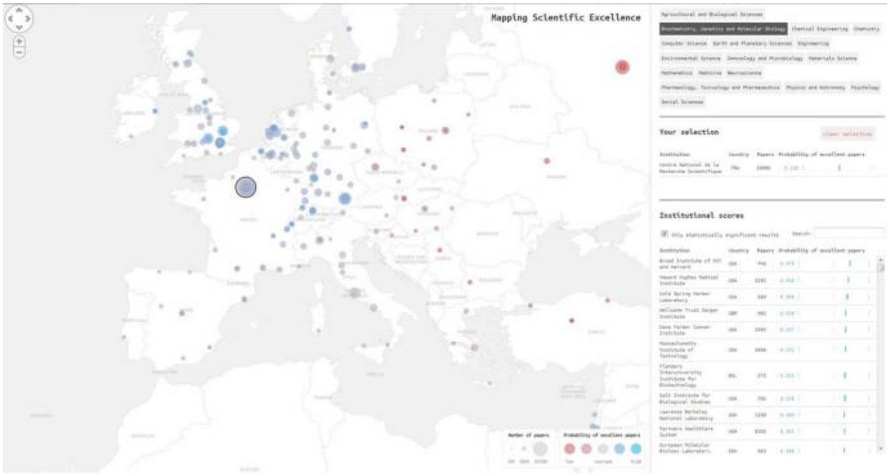

Figure 3.
Biochemistry, genetics and molecular biology map with the CNRS selected

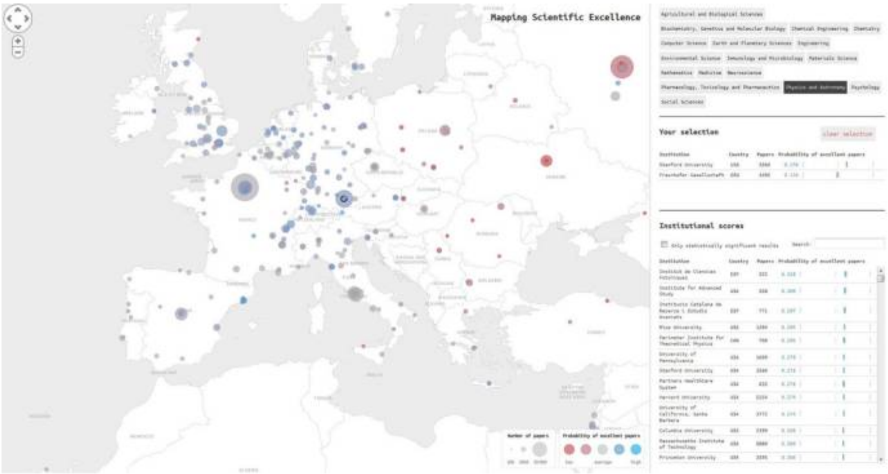

Figure 4.
Physics and astronomy map with Stanford University and Fraunhofer Gesellschaft selected

them in the list but also on the map with a mouse click. A new comparison of institutions can be started by clicking on “Clear selection”.

If the user has selected some institutions or has sorted them in a certain order, the selection and sort order are retained if the subject area is changed. This feature makes it possible to compare the results for certain institutions across different subject areas directly.

Discussion

The web application presented in this paper allows for an analysis which reveals centres of excellence in different fields worldwide using publication and citation data. Similar to the Leiden Ranking, only specific aspects of institutional performance are taken into account and other aspects such as teaching performance or the societal impact of research (Bornmann, 2012, 2013) are not considered (see Waltman *et al.*, 2012). Based on data gathered from Scopus, field-specific excellence can be identified in institutions where highly-cited papers have been published frequently. The web application combines both a list of institutions ordered by different indicator values and a map with circles visualising indicator values for geocoded institutions. Compared to the mapping and ranking approaches introduced hitherto, our underlying statistic (multi-level models) are analytically oriented by allowing the estimation of statistically more appropriate values for the number of excellent papers for an institution than the observed values; the calculation of confidence intervals as reliability measures for the institutional citation impact; the comparison of a single institution with an “average” institution in a subject area; and the direct comparison of at least two institutions. With these features our approach can not only identify the top performers in (excellent) output but the “true jewels” in different disciplines. These are institutions that have published statistically significantly more class 10 per cent papers than an “average” institution in a subject area. Against the backdrop of these advantages, our web application can be used by scientists for exploring excellence centres and regions worldwide in a specific subject area, students and parents to get helpful hints about the comparative merits of different universities when making choices, and governments and policymakers to compare the performance of institutions within a specific country to those outside (see Hazelkorn, 2011).

Despite the advantages of the web application for mapping excellence in science, we recognise the limitations inherent in bibliometric data. First papers are only one among several types of scientific activities. Research is a multi-dimensional endeavour which cannot be captured with only a single indicator. Second it is not guaranteed that the addresses listed on the publication reflect the locations where the reported research was conducted. There might be several addresses on a publication but the research was mainly conducted at one location. Third no standard technique exists for the subject classification of journal papers (Bornmann and Daniel, 2008; Bornmann *et al.*, 2008; Leydesdorff and Rafols, 2009). For the web application we have used the standard technique based on journal classification schemes. Although this approach has been frequently criticised and other solutions have been proposed (e.g. the categorisation of papers based on index terms from field-specific databases), it has not been possible to establish any other proposals as a (new) standard for interdisciplinary studies up to now.

Besides the limitations inherent in bibliometric data, there are several problems inherent in mapping approaches such as that proposed here. A detailed list of these problems is published in Bornmann *et al.* (2011a). The user should always be aware of these limitations when the web application is used. For example there may be circles for institutions on the maps that are not in the right position. In the various routines we tried to avoid these misallocations, but they could not be completely resolved. The misallocations do have different sources: address errors in the Scopus data or erroneous coordinates provided by the geocoding process. Furthermore high numbers of excellent papers visualised on the map for a single institution might be due to the following two effects: many scientists located in that institution produced at least one excellent paper each or only a few scientists located in this institution produced many influential papers. Assuming institutions as units of analysis, one is not able to distinguish between these two interpretations.

The web application described here allows future developments in several directions. First further data sets can be uploaded, e.g. for the visualisation of patent data (Leydesdorff and Bornmann, 2012a). Second with multi-level models it is possible to consider data for more than one subject area. Third this data can be used to categorise institutions in different groups such as universal performers that are successful in all subject areas and specific performers which are successful only in some areas (Bornmann *et al.*, n.d., in press-b). Covariates might be included in the regression models to control class 10 per cent probabilities for certain factors (e.g. institutional size, economic force) (Bornmann *et al.*, 2013b). Fourth the evaluation of spatial performance could be extended by network approaches proposed by Hennemann (2012) "in which co-authorships of scientific papers represent a definable relationship between knowledge-producing players in a system" (p. 2402) and by Calero Valdez *et al.* (2012) who proposed a performance metric of interdisciplinarity. According to Mazloumian *et al.* (2013) "given the large relevance of network theory in many scientific areas, we believe that classical, node-based indices must be complemented by network-based indices."

References

Aksnes, D.W. (2003), "A macro study of self-citation", *Scientometrics*, Vol. 56 No. 2, pp. 235-246.

Bornmann, L. (2012), "Measuring the societal impact of research", *EMBO Reports*, Vol. 13 No. 8, pp. 673-676.

Bornmann, L. (2013), "What is societal impact of research and how can it be assessed? A literature survey", *Journal of the American Society of Information Science and Technology*, Vol. 64 No. 2, pp. 217-233.

Bornmann, L. and Daniel, H.-D. (2008), "What do citation counts measure? A review of studies on citing behavior", *Journal of Documentation*, Vol. 64 No. 1, pp. 45-80.

Bornmann, L. and 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*, Vol. 62 No. 10, pp. 1954-1962.

Bornmann, L. and Waltman, L. (2011), "The detection of 'hot regions' in the geography of science: a visualization approach by using density maps", *Journal of Informetrics*, Vol. 5 No. 4, pp. 547-553.

Bornmann, L., De Moya Anegón, F. and Leydesdorff, L. (2012), "The new Excellence Indicator in the World Report of the SCImago Institutions Rankings 2011", *Journal of Informetrics*, Vol. 6 No. 2, pp. 333-335.

Bornmann, L., De Moya Anegón, F. and Mutz, R. (n.d.), "Do universities or research institutions with a specific subject profile have an advantage or a disadvantage in institutional rankings? A latent class analysis with data from the SCImago ranking", *Journal of the American Society for Information Science and Technology* (in press-b).

Bornmann, L., Leydesdorff, L. and Mutz, R. (2013a), "The use of percentiles and percentile rank classes in the analysis of bibliometric data: opportunities and limits", *Journal of Informetrics*, Vol. 7 No. 1, pp. 158-165.

Bornmann, L., Mutz, R. and Daniel, H.-D. (2013b), "A multilevel-statistical reformulation of citation-based university rankings: the Leiden Ranking 2011/2012", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 8, pp. 1649-1658.

Bornmann, L., Leydesdorff, L., Walch-Solimena, C. and Ettl, C. (2011a), "Mapping excellence in the geography of science: an approach based on Scopus data", *Journal of Informetrics*, Vol. 5 No. 4, pp. 537-546.

Bornmann, L., Mutz, R., Neuhaus, C. and Daniel, H.-D. (2008), "Use of citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results", *Ethics in Science and Environmental Politics*, Vol. 8 No. 1, pp. 93-102.

Bornmann, L., Schier, H., Marx, W. and Daniel, H.-D. (2011c), "Is interactive open access publishing able to identify high-impact submissions? A study on the predictive validity of Atmospheric Chemistry and Physics by using percentile rank classes", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 1, pp. 61-71.

Bornmann, L., Mutz, R., Marx, W., Schier, H. and Daniel, H.-D. (2011b), "A multilevel modelling approach to investigating the predictive validity of editorial decisions: do the editors of a high-profile journal select manuscripts that are highly cited after publication?", *Journal of the Royal Statistical Society – Series A (Statistics in Society)*, Vol. 174 No. 4, pp. 857-879.

Bornmann, L., Bowman, B.F., Bauer, J., Marx, W., Schier, H. and Palzenberger, M. (n.d.), "Standards for using bibliometrics in the evaluation of research institutes", in Cronin, B. and Sugimoto, C. (Eds), *Next Generation Metrics*, MIT Press, Cambridge, MA (in press-a).

Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M. and Vadillo-Muñoz, O. (2007), “Comparative study of international academic rankings of universities”, *Scientometrics*, Vol. 71 No. 3, pp. 349-365.

Calero Valdez, A., Schaar, A., Ziefle, M., Holzinger, A., Jeschke, S. and Brecher, C. (2012), “Using mixed node publication network graphs for analyzing success in interdisciplinary teams”, in Huang, R., Ghorbani, A., Pasi, G., Yamaguchi, T., Yen, N. and Jin, B. (Eds), *Active Media Technology*, Springer, Berlin, pp. 606-617.

Frenken, K., Hardeman, S. and Hoekman, J. (2009), “Spatial scientometrics: towards a cumulative research program”, *Journal of Informetrics*, Vol. 3 No. 3, pp. 222-232.

Glänzel, W., Debackere, K., Thijs, B. and Schubert, A. (2006), “A concise review on the role of author self-citations in information science, bibliometrics and science policy”, *Scientometrics*, Vol. 67 No. 2, pp. 263-277.

Goldstein, H. and Healy, M.J.R. (1994), “The graphical representation of a collection of means”, *Journal of the Royal Statistical Society Series a-Statistics in Society*, Vol. 158 No. 1, pp. 175-177.

Guerrero-Bote, V.P. and De Moya-Anegon, F. (2012), “A further step forward in measuring journals’ scientific prestige: the SJR2 indicator”, available at: http://arxiv.org/abs/1201.4639 (accessed 3 July 2012).

Hazelkorn, E. (2011), *Rankings and the Reshaping of Higher Education. The Battle for World-class Excellence*, Palgrave Macmillan, New York, NY.

Hennemann, S. (2012), “Evaluating the performance of geographical locations within scientific networks using an aggregation-randomization-re-sampling approach (ARR)”, *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 12, pp. 2393-2404.

Hox, J.J. (2010), *Multilevel Analysis: Techniques and Applications*, Routledge, New York, NY.

Kroth, A. and Daniel, H.D. (2008), “International university rankings – a critical review of the methodology”, *Zeitschrift für Erziehungswissenschaft*, Vol. 11 No. 4, pp. 542-558.

Levine, M. and Ensom, M.H.H. (2001), “Post hoc power analysis: an idea whose time has passed?”, *Pharmacotherapy*, Vol. 21 No. 4, pp. 405-409.

Leydesdorff, L. and Bornmann, L. (2012a), “Mapping (USPTO) patent data using overlays to Google maps”, *Journal of the American Society of Information Science and Technology*, Vol. 63 No. 7, pp. 1442-1458.

Leydesdorff, L. and Bornmann, L. (2012b), “Testing differences statistically with the Leiden ranking”, *Scientometrics*, Vol. 92 No. 3, pp. 781-783.

Leydesdorff, L. and Persson, O. (2010), “Mapping the geography of science: distribution patterns and networks of relations among cities and institutes”, *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 8, pp. 1622-1634.

Leydesdorff, L. and Rafols, I. (2009), “A global map of science based on the ISI subject categories”, *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 2, pp. 348-362.

Leydesdorff, L., Bornmann, L., Mutz, R. and Opthof, T. (2011), “Turning the tables in citation analysis one more time: principles for comparing sets of documents”, *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 7, pp. 1370-1381.

Mazloumian, A., Helbing, D., Lozano, S., Light, R.P. and Börner, K. (2013), “Global multi-level analysis of the ‘Scientific Food Web’”, *Scientific Reports*, Vol. 3, p. 1167.

Moineddin, R., Matheson, F.I. and Glazier, R.H. (2007), “A simulation study of sample size for multilevel logistic regression models”, *BMC Medical Research Methodology*, Vol. 7 No. 1.

Rousseau, R. (2012), "Basic properties of both percentile rank scores and the I3 indicator", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 2, pp. 416-420.

SAS Institute Inc. (2008), *SAS/STAT® 9.2 User's Guide*, SAS Institute, Cary, NC.

Scimago Research Group (2011), *SIR World Report 2011*, University of Granada, Granada.

Scimago Research Group (2012), *SIR World Report 2012*, University of Granada, Granada.

Sheskin, D. (2007), *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, FL.

Shin, J.C., Toutkoushian, R.K. and Teichler, U. (2011), *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education*, Springer, Dordrecht.

Snijders, T.A.B. and Bosker, R.J. (2004), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Sage Publications, London.

Snyder, H. and Bonzi, S. (1998), "Patterns of self-citation across disciplines (1980-1989)", *Journal of Information Science*, Vol. 24 No. 6, pp. 431-435.

Spiegelhalter, D., Pearson, M. and Short, I. (2011), "Visualizing uncertainty about the future", *Science*, Vol. 333 No. 6048, pp. 1393-1400.

Sun, S.Y., Pan, W. and Wang, L.L. (2011), "Rethinking observed power concept, practice, and implications", *Methodology – European Journal of Research Methods for the Behavioral and Social Sciences*, Vol. 7 No. 3, pp. 81-87.

Tijssen, R. and Van Leeuwen, T. (2006), "Centres of research excellence and science indicators. Can 'excellence' be captured in numbers?", in Glänzel, W. (Ed.), *Ninth International Conference on Science and Technology Indicators*, Katholieke Universiteit Leuven, Leuven, pp. 146-147.

Tijssen, R., Visser, M. and Van Leeuwen, T. (2002), "Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference?", *Scientometrics*, Vol. 54 No. 3, pp. 381-397.

Vinkler, P. (2010), *The Evaluation of Research by Scientometric Indicators*, Chandos Publishing, Oxford.

Waltman, L. and Schreiber, M. (2013), "On the calculation of percentile-based bibliometric indicators", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 2, pp. 372-379.

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. and Wouters, P. (2012), "The Leiden Ranking 2011/2012: data collection, indicators, and interpretation", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 12, pp. 2419-2432.

Yuan, K.H. and Maxwell, S. (2005), "On the post hoc power in testing mean differences", *Journal of Educational and Behavioral Statistics*, Vol. 30 No. 2, pp. 141-167.

About the authors

Lutz Bornmann works as a Sociologist of Science at the Division for Science and Innovation Studies in the Administrative Headquarters of the Max Planck Society in Munich. His current research interests include research evaluation, peer review and bibliometric indicators. He is a member of the editorial boards of the *Journal of Informetrics* (Elsevier), *PLOS ONE* and *Scientometrics* (Springer). He is also senior associate editor for the *International Journal of Biomedical Science Editing* (Kowsar) and advisory editorial board member of *EMBO Reports* (Nature Publishing). Dr Lutz Bornmann is the corresponding author and can be contacted at: lutz.bornmann@gv.mpg.de

Moritz Stefaner is a self-employed "Truth and Beauty Operator" working at the crossroads of data visualisation, information aesthetics and user interface design. With a background in cognitive science (BSc with distinction, University of Osnabrück) and interface design (MA, University of Applied Sciences, Potsdam), his work balances analytical and aesthetic aspects in mapping abstract and complex phenomena. He is especially interested in the visualisation of large-scale human activity. His personal portfolio may be seen at http://moritz.stefaner.eu. He also publishes the Data Stories podcast together with Enrico Bertini.

Felix de Moya Anegón (PhD, University of Granada) is Professor of Investigation at the Instituto de Políticas y Bienes Públicos, where his research interests include advanced techniques for information retrieval in the development of information systems in real time, techniques for the visualisation of scientific information in interface design, and the analysis and evaluation of scientific domains; information system design in scientometrics, and in aiding decision-making processes. Professor Anegón's publications have appeared in such journals as *JASIST*, *Chemistry Today*, *Aslib Proceedings* and *Scientometrics*.

Rüdiger Mutz is a Senior Researcher at the Swiss Federal Institute of Technology Zurich. Dr Mutz holds degrees in psychology and economics. His research interests include research evaluation, quality assurance in higher education, rankings and advanced statistical procedures (e.g. multi-level modelling, meta-analysis).

This article has been cited by:

1. Ole Ellegaard. 2018. The application of bibliometric analysis: disciplinary and user aspects. *Scientometrics* **465**. . [Crossref]
2. Xiaomei Bai. 2018. Predicting the Number of Publications for Scholarly. *IEEE Access* 1-1. [Crossref]
3. Matjaž Mikoš. 2017. Landslides: a top international journal in geological engineering and engineering geology?. *Landslides* **9**. . [Crossref]
4. Gangan Prathap. 2017. A three-dimensional bibliometric evaluation of recent research in India. *Scientometrics* **110**:3, 1085-1097. [Crossref]
5. Lutz Bornmann, Alexander I. Pudovkin. 2017. The Journal Impact Factor Should Not Be Discarded. *Journal of Korean Medical Science* **32**:2, 180. [Crossref]
6. Xiaomei Bai, Ivan Lee, Zhaolong Ning, Amr Tolba, Feng Xia. 2017. The Role of Positive and Negative Citations in Scientific Evaluation. *IEEE Access* **5**, 17607-17617. [Crossref]
7. Xiaomei Bai, Fuli Zhang, Jie Hou, Feng Xia, Amr Tolba, Elsayed Elashkar. 2017. Implicit Multi-Feature Learning for Dynamic Time Series Prediction of the Impact of Institutions. *IEEE Access* **5**, 16372-16382. [Crossref]
8. Marco Frittelli, Loriano Mancini, Ilaria Peri. 2016. Scientific research measures. *Journal of the Association for Information Science and Technology* **67**:12, 3051-3063. [Crossref]
9. Zaida Chinchilla-Rodríguez, Grisel Zacca-González, Benjamín Vargas-Quesada, Félix de Moya-Anegón. 2016. Benchmarking scientific performance by decomposing leadership of Cuban and Latin American institutions in Public Health. *Scientometrics* **106**:3, 1239-1264. [Crossref]
10. Ashraf Uddin, Jaideep Bhoosreddy, Marisha Tiwari, Vivek Kumar Singh. 2016. A Sciento-text framework to characterize research strength of institutions at fine-grained thematic area level. *Scientometrics* **106**:3, 1135-1150. [Crossref]
11. Lutz Bornmann, Moritz Stefaner, Felix de Moya Anegón, Rüdiger Mutz. 2016. Excellence networks in science: A Web-based application based on Bayesian multilevel logistic regression (BMLR) for the identification of institutions collaborating successfully. *Journal of Informetrics* **10**:1, 312-327. [Crossref]
12. Marina Dobrota, Milica Bulajic, Lutz Bornmann, Veljko Jeremic. 2016. A new approach to the QS university ranking using the composite I-distance indicator: Uncertainty and sensitivity analyses. *Journal of the Association for Information Science and Technology* **67**:1, 200-211. [Crossref]
13. Giovanni Abramo, Ciriaco Andrea D’Angelo. 2015. A methodology to compute the territorial productivity of scientists: The case of Italy. *Journal of Informetrics* **9**:4, 675-685. [Crossref]
14. Ivan Dario Ramos Vacca, Victor Andres Bucheli Guerrero. Automatic geolocation of the scientific knowledge: Geolocarti 416-424. [Crossref]
15. Wei-Chao Lin, Chih-Fong Tsai, Shih-Wen Ke. 2015. Correlation analysis for comparison of the citation impact of journals, magazines, and conferences in computer science. *Online Information Review* **39**:3, 310-325. [Abstract] [Full Text] [PDF]
16. Giovanni Abramo, Ciriaco Andrea D'Angelo, Flavia Di Costa. 2015. A new approach to measure the scientific strengths of territories. *Journal of the Association for Information Science and Technology* **66**:6, 1167-1177. [Crossref]

17. Guo Chen, Lu Xiao, Chang-ping Hu, Xue-qin Zhao. 2015. Identifying the research focus of Library and Information Science institutions in China with institution-specific keywords. *Scientometrics* **103**:2, 707-724. [Crossref]
18. Lutz Bornmann, Moritz Stefaner, Felix de Moya Anegón, Rüdiger Mutz. 2015. Ranking and mappping of universities and research-focused institutions worldwide: The third release of excellencemapping.net. *Collnet Journal of Scientometrics and Information Management* **9**:1, 65-72. [Crossref]
19. Christophe Vandeviver. 2014. Applying Google Maps and Google Street View in criminological research. *Crime Science* **3**:1. . [Crossref]
20. Shih-Wen Ke, Wei-Chao Lin, Chih-Fong Tsai, Ya-Han Hu. 2014. Citation impact analysis of research papers that appear in oral and poster sessions. *Online Information Review* **38**:6, 738-745. [Abstract] [Full Text] [PDF]
21. Lutz Bornmann, Moritz Stefaner, Felix de Moya Anegón, Rüdiger Mutz. 2014. 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. [Crossref]
22. Edmilson J.T. Manganote, Mariana S. Araujo, Peter A. Schulz. 2014. Visualization of ranking data: Geographical signatures in international collaboration, leadership and research impact. *Journal of Informetrics* **8**:3, 642-649. [Crossref]
23. L. Bornmann. 2014. How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation* **23**:2, 166-173. [Crossref]
24. Nicolás Robinson-García, Clara Calero-Medina. 2014. What do university rankings by fields rank? Exploring discrepancies between the organizational structure of universities and bibliometric classifications. *Scientometrics* **98**:3, 1955-1970. [Crossref]
25. Nicolas Robinson-Garcia, Evaristo Jiménez-Contreras. Analyzing the Disciplinary Focus of Universities 161-185. [Crossref]