CrossMark

Applying the CSS method to bibliometric indicators used in (university) rankings

Lutz Bornmann[1] · Wolfgang Glänzel[2,3]

Abstract This Letter to the Editor proposes to use the CSS method for classifying ranking results (e.g. from university rankings) into meaningful groups.

The ranking of institutions and other units (e.g. countries and single researchers) characterizes the modern science. For example, institutions and countries are ranked by the Nature Index (http://www.natureindex.com) which counts the number of publications in reputable journals (Bornmann and Haunschild 2016). The Leiden Ranking (http://www.leidenranking.com) lists universities by using different bibliometric indicators which focus on the publication output and citation impact of publications. The problem with these and other rankings is that the user is especially interested in ranking positions rather than meaningful differences between units. The Leiden Ranking, for example, publishes for universities the share of papers which belongs to the 10% most frequently cited papers in corresponding subject categories and publication years ($P_{top10\%}$). There are 63 ranking positions between Univ Hull (position 200) and Univ Bern (position 137), although the difference between both universities is only 1% point. In order to avoid the interpretation

✉ Lutz Bornmann
bornmann@gv.mpg.de

Wolfgang Glänzel
wolfgang.glanzel@kuleuven.be; glanzw@iif.hu

1   Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany

2   ECOOM and Department of MSI, KU Leuven, Naamsestraat 61, 3000 Louvain, Belgium

3   Department of Science Policy and Scientometrics, Library of the Hungarian Academy of Sciences, Arany János u. 1, Budapest 1051, Hungary

of ranking positions without considering the differences in indicator values, it has been suggested to group the evaluated units meaningfully.

The CHE Ranking (http://www.che-ranking.de) uses several indicators to rank German universities. In order to avoid the focus of the users on ranking positions, it applies two methods to group the universities. (1) The indicator values are used to identify the 25% of the institutions with the best values (the first quartile) as well as the next 50% (second and third quartiles) and the bottom 25% (fourth quartile). (2) Confidence intervals of institutional indicator values are calculated. If the mean value across all institutions is outside the lower bound of an institution, the institution is assigned to the top group. If the mean is outside the upper bound, the institution is assigned to the bottom group. All other institutions, for which the mean is positioned within the confidence interval, are part of a middle group. Both methods which are used in the CHE Ranking allow meaningful separations of institutional groups based on established statistical methods.

In this Letter to the Editor a third method of group formation is proposed which is based on the Characteristics Scores and Scales (CSS) method—"a parameter-free solution for the assessment of outstanding performance" (Glänzel et al. 2016). This method was introduced by Glänzel and Schubert (1988) to assign single publications in a field and publication year to meaningful impact groups (as a rule four groups). Characteristic scores are obtained from iteratively truncating samples at their mean value and recalculating the mean of the truncated sample until the procedure is stopped or no new scores are generated: In the first step the mean citation impact is calculated for all papers in a field and publication year. All papers with citations counts below the mean are classified as "poorly cited". The papers with citation counts above the mean are used for further calculations in the second step. For these papers the mean citation impact is calculated again and the papers with citation counts below the mean are assigned to the category "fairly cited". In the third step the procedure of mean calculation and separation of two groups is repeated which results in two further impact groups labelled as "remarkably cited" and "outstandingly cited". If all papers in a database (e.g. Web of Science) are classified in this way based on their different assignment to fields and publication years, the share of papers published by units (e.g. researchers or institutions) in these impact categories can be determined. Large shares of papers in the categories "remarkably cited" and "outstandingly cited" point to successful research activities measured in terms of citations. It is an advantage of CSS that "it is self-adjusting, that is, no predefined thresholds, such as given numbers of citation or percentiles, are necessary" (Glänzel et al. 2016).

In principle, the CSS method can not only be used for the classification of single papers, but also to assign institutions in rankings to meaningful groups. For that, the user of the ranking selects the indicator of interest and calculates the necessary mean values for the group formation—provided the complete data are freely available. For example, the underlying data for the Leiden Ranking can be downloaded from the Centre for Science and Technology Studies (CWTS) web site (http://www.leidenranking.com/downloads). Using the $P_{top10\%}$ indicator (the number of top-cited papers), one receives the following thresholds for the four impact groups (considering fractionally counted papers from all sciences in 2011–2014): "poorly cited" $< P_{top10\%} = 386.3774 \geq$ "fairly cited" $< P_{top10\%} = 888.7805 \geq$ "remarkably cited" $< P_{top10\%} = 1544.48 <$ "outstandingly cited". Twenty-six universities are outstandingly cited (3%), 56 remarkably cited (7%), 172 fairly cited (20%) and 588 poorly cited (70%). This distribution of percentages (3–7–20–70%) is similar to distributions which are reported by Glänzel and Schubert (1988) and Albarrán and Ruiz-Castillo (2011) based on the impact of single publications. Thus, the distribution seems to follow a general pattern which is independent of publication year and citation window. Similar distribution properties

have recently been reported by Chi and Glänzel (2016) in the context of Thomson Reuters' "usage count" statistics provided by the Web of Science Core Collection database.

The application of the CSS method could also be of interest for the classification of highly-cited researchers (HCR, http://highlycited.com/). Clarivate Analytics, formerly the IP and Science business of Thomson Reuters, publishes annually those researchers which have published the most highly-cited papers in their discipline. In order to reduce the set of researchers with at least one highly-cited paper to researchers with a substantial number of highly-cited papers, the square root of all researchers with at least one highly-cited paper is used as threshold. If the CSS method would be used instead, more than two HCR groups could be built within each discipline. In order to show only researchers with a substantial number of highly-cited papers the HCR list on www.highlycited.com could show all researchers above the "poorly cited" group (here: researchers with only a few highly-cited papers). The listed HCR could then be separated into three groups: "fairly number of highly-cited papers", "remarkably number of highly-cited papers" and "outstandingly number of highly-cited papers".

References

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology, 62*(1), 40–49. doi:10.1002/asi.21448.

Bornmann, L., & Haunschild, R. (2016). An empirical look at the nature index. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23682.

Chi, P. S., & Glänzel, W. (2016). Do usage and scientific collaboration associate with citation impact? In I. Rafols, J. Molas-Gallart, E. Castro-Martínez, & R. Woolley (Eds.), *Proceedings of the 21st international conference on science and technology indicators—Peripheries, frontiers and beyond* (pp. 1223–1228). Valencia, Spain.

Glänzel, W., Debackere, K., & Thijs, B. (2016). *Citation classes: A novel indicator base to classify scientific output*. Retrieved October 21, 2016, from http://www.oecd.org/sti/051%20-%20Blue%20Sky%20Biblio%20Submitted.pdf.

Glänzel, W., & Schubert, A. (1988). Characteristic Scores and Scales in assessing citation impact. *Journal of Information Science, 14*(2), 123–127.