P. Ehret, A. Zucalla, and B. Gipp, "University rankings in computer science: A study and visualization of 'geo-based' impact and conference proceeding (CORE) scores," in Proceedings of ISSI 2017: 17th International Society of Scientometrics and Informetrics Conference, Wuhan, China, 2017.

University rankings in computer science: a study and visualization of 'geo-based' impact and conference proceeding (CORE) scores

Philip Ehret[1], Alesia Zuccala[2], and Bela Gipp[3]

[1] *philip.ehret@uni-konstanz.de*, [3] *bela.gipp@uni-konstanz.de*
University of Konstanz, Konstanz (Germany)

[2] *a.zuccala@hum.ku.dk*
University of Copenhagen, Copenhagen (Denmark)

Abstract
This is a research-in-progress paper concerning two types of institutional rankings, the Leiden and QS World ranking, and their relationship to a list of universities' 'geo-based' impact scores, and Computing Research and Education Conference (CORE) participation scores in the field of computer science. A 'geo-based' impact measure examines the geographical distribution of incoming citations to a particular university's journal articles for a specific period of time. It takes into account both the number of citations and the geographical variability in these citations. The CORE participation score is calculated on the basis of the number of weighted proceedings papers that a university has contributed to either an A*, A, B, or C conference as ranked by the Computing Research and Education Association of Australasia. In addition to calculating the correlations between the distinct university rankings and the separate 'geo-based' versus CORE scores, we are in the process of developing a geographical visualization tool that presents the metrics so that they may be examined in an explorative way.

Conference Topic
University policy and institutional rankings; Science communication; Mapping and visualization

Introduction
University rankings have rapidly become an influential tool in government and educational policymaking (The Guardian, 2013), and after the first Academic Ranking of World Universities (ARWU) was introduced in 2003 (also known as the 'Shanghai Ranking'), alternative rankings began to appear. These include, though in no specific order, the QS World Ranking of Universities, the Times Higher Education (THE) Ranking, SCImago ranking, and the Leiden Ranking. In past years, each ranking has been touted, examined for their disadvantages and advantages, and assessed on the bases of their similarities (e.g., Aguillo et al., 2010; Waltman et al., 2012; Bornmann et al., 2013). Certain methodological approaches have also been more heavily criticized than others (Liu & Cheng, 2005; Liu et al., 2005; VanRaan, 2005). Some critics, for example, are sceptical about approaches that rely on the amalgamation and use of weighted variables (Billaut, 2010), and others are concerned with reproducibility (Docampo, 2012). Most researchers tend to agree; however, that normalization is key to producing rankings, especially when using publication and citation data with variable field differences (López-Illescas et al., 2011).

University rankings are here to stay, and since can often be influential, there are always ample reasons to examine them further. One approach is to focus on statistical shortcomings within the ranking itself, but another is to identify and examine new variables of interest and test them to see if they show some form of positive, negative, or neutral relationship to that ranking. In this sense, correlation measures are useful, if they provide further information and insight into the communication practices of a research field, where a ranking might not present the full picture. The field that we have chosen to investigate is *information and computer science*, and the type of university rankings that we use in this study implement either a field-normalized approach – i.e., the Leiden Ranking – or a field-specific approach, such as the QS World University Ranking.

Currently, our research is 'in-progress' and it is based on two components. The first component is metric in nature, and the second involves developing a visualization tool that will allow users to explore our results geographically.

With data pertaining to: a) the Leiden University Ranking (2016), b) the QS World University Ranking in Computer Science (2016), c) university-to-university directed citation counts collected from Web of Science (WoS) journal articles (2012-2016), the first part of our study will focus on the following:

1) *What is a particular university's 'geo-based' impact (i.e., geographical reach) in the field of computer science as measured by the citations it receives from a variety of international universities?*

2) *Do high-ranking universities in the field of computer science tend to receive a broader geographical reach of citations than those that achieve a lower rank?*

Our third research question relies also on Leiden and QS World University Ranking data, but includes: a) only conference proceeding publications matched to ranked universities, and b) data from the Computing Research and Education Conference (CORE) Ranking (2014).

3) *Do computer scientists working at top ranked universities tend to participate more often in the top ranked CORE conferences than those from lower ranking universities?*

Methods

Rankings, articles and proceedings data collection:

The Leiden Ranking data, the QS World University Rankings in the field of computer science, and the CORE Computing Research and Education Conference data are publicly available on the Web. By using a standard web scraping method, we will collect data from the following sites:

1) Leiden Ranking: http://www.leidenranking.com/ranking/2016/list
2) QS World Subject 2016 - Computer Science & Information Systems: https://www.topuniversities.com/university-rankings/university-subject-rankings/2016/computer-science-information-systems
3) Computing Research and Education (CORE) Conference website: http://www.core.edu.au/conference-portal

With the Leiden ranked list we will limit the universities to those associated with the field of 'mathematics and computer science' and use the basic P indicator (i.e., total number of publications), and percentile-based indicators: $P_{top\text{-}10\%,}$ and $PP_{top\text{-}10\%.}$ A percentile-based indicator is that which values publications based on their position within the citation distribution of their field (Waltman & Schreiber, 2012). The $P_{top\text{-}10\%}$ is more precisely defined as: “the number of publications which belong to the top 10% most frequently cited publications; a publication belongs to the top 10% most frequently cited if it is cited more than 90% of the publications published in the same subject area and in the same year” (Bornmann & Williams, 2013)

The indicator we use from the QS 2016 World University Rankings in Computer Science & Information Systems is computed as an “overall score” (a score of 0 to 100) for each university, and it is comprised of four components: 1) academic reputation (i.e., a global survey of academics), 2) employer reputation (i.e., a global employer survey), 3) research citations per paper, and 4) the h-index (see https://www.topuniversities.com/subject-rankings/methodology)

The Web of Science data for a set of computer science journal articles (doctype = Article,) for the period of 2007 to 2011 (publication years) has been provided to us from Clarivate Analytics (i.e. formerly Thomson Reuters). Our indicator for ‘geo-based’ impact is calculated in terms of citation variability, that is, the variability in the origin of citations received by the universities from other universities in countries worldwide (i.e., this is specifically for the articles published in 2007 to 2011, which have been cited during the period of 2012 to 2016). This approach, shown in the formula below, is adapted from what has been done earlier by Gao et al. (2013):

$$gi(I) = \left|\bigcup_{i=1}^{P(I)} \bigcup_{j=1}^{C(i)} O(j)\right|$$

I: A certain institution

$P(I)$: Number of papers published by researchers affiliated with institution I

$C(i)$: Number of incoming citations for paper i

$O(j)$: Country of origin for citation j

And finally, a set of conference proceedings articles recorded in WoS for the year 2016 (doctype = Article,Proceedings Paper), will be matched to both their university of origin as well as the rank of the actual conference, as listed at the (CORE) Conference website. According to the CORE webpage: “conference rankings are determined by a mix of indicators, including citation rates, paper submission and acceptance rates, and the visibility and research track record of the key people hosting the conference and managing its technical program.” This includes a more detailed statement concerning the categorization ranks, which are labelled as A*, A, B, and C (see CORE, 2016).

For each university we determine an overall CORE score based on the weighted proportion of its proceedings articles that match with one of the CORE rankings: A*, A, B, and C. Thus for each Core rank we will give an ‘A*’ a weight of 3, an ‘A’ a weight of 2, a ‘B’ a weight of 1 and a ‘C’ a weight of 0.5. If, for example, a university produces 5 articles each associated with the following CORE ranks (3=A*, 1=A, 1=B) we would obtain a CORE score = [3(3/5) + 2(1/5) + 1(1/5)] = [3(0.6) + 2(0.2) + 1(0.2)] = (1.8 + 0.4 + 0.2) = 2.4.

Visualization:

The visualization component of this study is under development at the Department of Computer Science, University of Konstanz (see Figure 1, below). The aim is to provide users with an interactive tool that can support immediate geographical comparisons for a set of ranked universities from the field of computer science. A drop-down menu (left of screen) presents the university's most recent rank, as per the Leiden ranking method for 'mathematics and computer science', the QS World Academic ranking in computer science, and associated Computing Research and Education Conference CORE score and geo-based impact.

The circular nodes on the map identify each university within its specific country. Red lines, or trajectories connect the university to other universities worldwide and illustrate the degree to which it is *receiving* citations. For instance, a user can click on a specific university node (University of Konstanz) in our geographical visualization tool, then hover over the trajectory and find a pop-up text at the right of the screen, which indicates total citations received for a specific time period. For example, (s)he might see a count of 30 incoming citations to University of Konstanz (Germany) from University of Toronto (Canada) for the period 2011-2016.

The countries in which the universities are situated will also be coloured as per their average geo-based impact. Again, the user can hover over a country, and at the right side of the screen a pop-up text will appear with the country name, its flag, and average geo-based impact, as calculated from the geo-based impacts of all of its regional universities'. A country with an above average geo-based impact will be coloured a darker shade of blue, and a country with a below average geo-based impact will be coloured a lighter shade of blue.

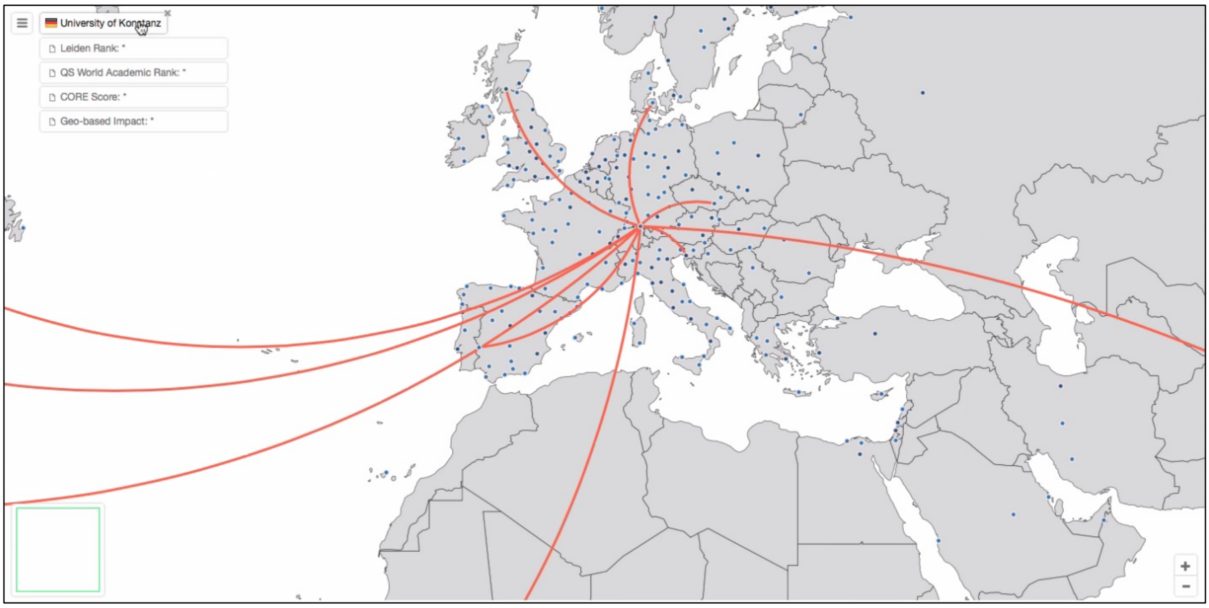

Figure 1. Prototype of the geographical visualization tool for university rankings, CORE scores, and geo-based impacts in computer science.

Expected Results

The results of this study aim to give scholars the opportunity to monitor their university's current rank within a specific research field and to identify additional field-related measures associated with these rankings. With the metric part of our study we expect to find that high-ranking universities in the field of computer science tend to receive a broader geographical reach of citations than those that achieve a lower rank. We further expect to find that computer scientists working at top ranked universities tend to participate more often in the top ranked CORE conferences, than those working at lower ranked universities. This will be shown on our geographical visualization tool in terms of multiple trajectories (red lines) corresponding to received international citations. It will also be shown on the basis of a high 'geo-based' impact measure and a high CORE score for each university in a pop-up menu beneath their corresponding Leiden and QS rankings.

References

Aguillo, I.F., Bar-Illan, J., Levene, M. & Ortega J L. (2010). Comparing university rankings. *Scientometrics, 85*, 243–256.

Billaut, J-C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics, 84*, 237-263.

Bornmann, L., de Moya Anegón, F. and Mutz, R. (2013), Do universities or research institutions with a specific subject profile have an advantage or a disadvantage in institutional rankings? *Journal of the American Society for Information Science and Technology, 64*, 2310–2316.

Bornmann, L. & Williams, R. (2013). How to calculate the practical significance of citation impact differences? An empirical example from evaluative institutional bibliometrics using adjusted predictions and marginal effects. *Journal of Informetrics, 7*, 562–574.

CORE. (2017). Computer science conference ranking descriptions. Retrieved April 10, 2017 from https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnx3d3djb3JlZWR1fGd4OjJjNjkxOWE1NWQ4ZGY5MjU.

Docampo, D. (2012). Reproducibility of the Shanghai academic ranking of world universities results. *Scientometrics, 94*, 567–587.

Gao, S., Hu, Y., Janowicz, K. & McKenzie, G. (2013). A spatiotemporal scientometrics framework for exploring the citation impact of publications and scientists. *In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL'13)* (pp. 204-213) ACM, New York: USA. DOI=http://dx.doi.org/10.1145/2525314.2525368

Liu, N.C. & Cheng, Y. (2005). Academic ranking of world universities: methodologies and problems. *Higher Education in Europe*, 30, 127–136.

Liu, NC, Cheng, Y & Liu, L. (2005). Academic ranking of world universities using scientometrics: A comment to the "fatal attraction". *Scientometrics, 64*,101–109

López-Illescas, C., de Moya-Anegón, F., Moed, H. F. (2011) A ranking of universities should account for differences in their disciplinary specialization. *Scientometrics, 88*, 563-574.

The Guardian. (2013). *World university rankings: how much influence do they really have?* Retrieved April 9, 2017 from https://www.theguardian.com/higher-education-network/blog/2013/sep/10/university-rankings-influence-government-policy

VanRaan, A. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientimetrics, 62,* 133-143.

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, 63*, 2419–2432.

Waltman, L. and Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. *Journal of the American Society for Information Science and Technology, 64*, 372–379.