Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 72, No. 1 (2007) 25–32
DOI: 10.1007/s11192-007-1712-1

Irreproducibility of the results of the Shanghai academic ranking of world universities

RĂZVAN V. FLORIAN[a,b]

[a] *Ad Astra Association of Romanian Scientists, Cluj-Napoca (Romania)*
[b] *Center for Cognitive and Neural Studies (Coneural), Cluj-Napoca (Romania)*

I discuss the difficulties that I encountered in reproducing the results of the Shanghai ranking of world universities. In the Shanghai ranking, the dependence between the score for the SCI indicator and the weighted number of considered articles obeys a power law, instead of the proportional dependence that is suggested by the official methodology of the ranking. Discrepancies from proportionality are also found in some of the scores for the N&S and Size indicators. This shows that the results of the Shanghai ranking cannot be reproduced, given raw data and the public methodology of the ranking.

Introduction

The Shanghai academic ranking of world universities[1] has been published in the last few years by the Institute of Higher Education of Shanghai Jiao Tong University, and rapidly gained international recognition. I attempted to estimate the score that Romanian universities would obtain in this ranking, as none of them entered among the top 500 universities.[2] An unexpected result of this study was that the results of the 2005 Shanghai ranking cannot be reproduced by following the public methodology of the ranking. I present here the inconsistencies that I found between the scores obtained by

Received August 23, 2006

Address for correspondence:
RĂZVAN V. FLORIAN
Ad Astra Association of Romanian Scientists
Str. Saturn nr. 24, 400504 Cluj-Napoca, Romania
E-mail: florian@ad-astra.ro

0138–9130/US $ 20.00

universities in the ranking, on one side, and data processed according the public methodology, on the other side.

The SCI indicator

The methodology of the 2005 Shanghai ranking is presented on the ranking website;[3] I will refer to it as the public methodology. According to it, one of the indicators used in the ranking is the SCI indicator, measuring the "total number of articles indexed in Science Citation Index-Expanded, Social Science Citation Index, and Arts & Humanities Citation Index in 2004. Only publications of article type are considered. When calculating the total number of articles of an institution, a special weight of two was introduced for articles indexed in Social Science Citation Index and Arts & Humanities Citation Index." It is also said that "for each indicator, the highest scoring institution is assigned a score of 100, and other institutions are calculated as a percentage of the top score. The distribution of data for each indicator is examined for any significant distorting effect; standard statistical techniques are used to adjust the indicator if necessary."

The scores obtained by the top 500 universities for each of the indicators, including the SCI one, are provided in an Excel file on the ranking website.[4] I noted with s_{ISI} the score for the SCI indicator, given in this file. The number of articles indexed in 2004 in Science Citation Index-expanded (SCIE), Social Science Citation Index (SSCI), and Arts & Humanities Citation Index (AHCI), and in either of all these three indexes (SCI), with authors from the top 500 universities, was extracted from the ISI database by a group from Institute Jozef Stefan, Slovenia, and made available on the internet.[5] I noted with n_{SCIE}, n_{SSCI}, n_{AHCI} and n_{SCI} these numbers of articles. The queries used in the Slovenian study for searching in the Thomson ISI database are made available on its website and thus the data can be easily reproduced.

I noted with n_{ISI} the total number of articles of an institution, as used in the calculation of the Shanghai score for the SCI indicator. The three indexes (SCIE, SSCI, AHCI) are not disjoint (currently, 304 journals are indexed in both SCIE and SSCI; 37 journals in both SCIE and AHCI; 53 journals in both SSCI and AHCI – data computed from the Thomson ISI journal lists available online).[6] Hence, it is not clear from the stated methodology how n_{ISI} is computed from n_{SCIE}, n_{SSCI}, n_{AHCI} and/or n_{SCI} . I considered the alternatives $n_{ISI,1}=n_{SCI}+n_{SSCI}+n_{AHCI}$ and $n_{ISI,2}=n_{SCIE}+2(n_{SSCI}+n_{AHCI})$. I did not consider the possible alternative $n_{ISI,3}=n_{SCI}+n_{SSCI,\ AHCI}$, where $n_{SSCI,\ AHCI}$ is the number of articles in either SSCI or AHCI, for lack of data, but the discrepancy between the results obtained with this option and the results expected from the Shanghai official methodology would be even greater than with the other options (as $n_{ISI,\ 3}<n_{ISI,\ 1}$).

When plotting $n_{ISI,1}$ and $n_{ISI,2}$, computed from the data of the Slovenian group, against s_{ISI}, the score used in the Shanghai ranking (Figure 1), it is immediately clear

that the relationship between the two variables is not one of proportionality, as the official methodology suggests, but a nonlinear one. For example, the relationship between s_{ISI} and $n_{ISI,1}$ is fitted well by the power law $s_{ISI}=1.119\ n_{ISI,1}^{\ 0.470}$ (Figure 1), a function that passes through the data point corresponding to Harvard university, the institution with the highest performance for this indicator. The best power law fit is $s_{ISI}=1.698\ n_{ISI,1}^{\ 0.414}$. The difference between the forms of the distributions given by the two alternatives, $n_{ISI,1}$ and $n_{ISI,2}$, is not significant.

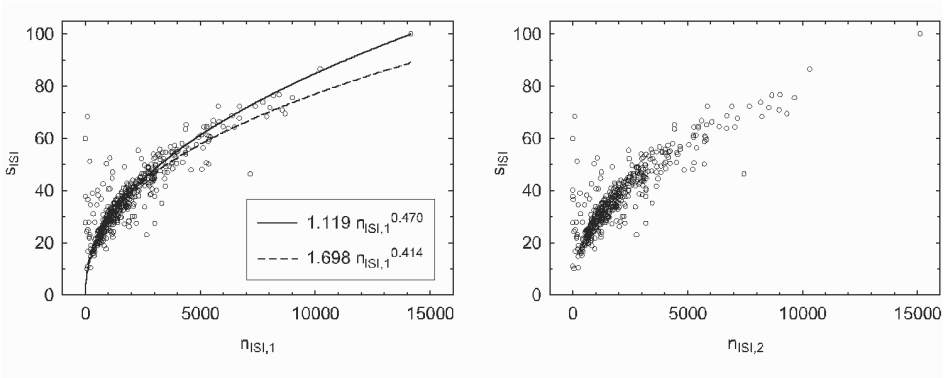

Figure 1. The dependence between the score for the SCI indicator from the Shanghai ranking and the weighted total number of articles, computed in two distinct ways. The public methodology suggests that the points should lie on a straight line passing through origin. In reality, the relationship is nonlinear, and is fitted well by a power law (the curves in the first graph; see text for details).

The fact that this power law fits so well the distribution of the data points suggests that the discrepancy with regard to proportionality is not due by errors in extracting the numbers of articles, but appears because the dependence used in the Shanghai ranking is actually nonlinear. Errors in extracting the number of articles, made either by the authors of the Slovenian study or by the authors of the Shanghai ranking, do exist, as it is shown by the spreading of some of the data points around the fitting curves. This should be expected, since defining correct search queries for extracting the output of universities from a publications database is a classical, severe problem in the analysis of academic performance, due to the variability of the names of the institutions that are indicated in scientific publications. There is an average relative error of 22% between $n_{ISI,1}$ and the corresponding values computed as $(s_{ISI}/1.119)^{\ 1/0.470}$. However, the regularity in the distribution of the data points along a power law curve suggests that in most cases there are small discrepancies between the number of articles used in the two studies (the Slovenian study and the Shanghai ranking), but the relationship between the Shanghai score for the SCI indicator and the weighted number of articles in nonlinear.

The discrepancy with regard to proportionality can be explained only by a mention in the methodology, "the distribution of data for each indicator is examined for any significant distorting effect; standard statistical techniques are used to adjust the indicator if necessary". The authors of the Shanghai ranking confirmed, indirectly, in correspondence through email, that they have used such a nonlinear dependence. It is however unclear why the authors of the Shanghai ranking considered that there was a "significant distorting effect" in the original data, and why they chose to distort the data in this particular nonlinear fashion.

The N&S indicator

According to the public methodology of the Shanghai ranking, this indicator is made up by „the number of articles published in *Nature* and *Science* between 2000 and 2004. To distinguish the order of author affiliation, a weight of 100% is assigned for corresponding author affiliation, 50% for first author affiliation (second author affiliation if the first author affiliation is the same as corresponding author affiliation), 25% for the next author affiliation, and 10% for other author affiliations. Only publications of article type are considered." I computed the value of this indicator, $n_{N\&S}$, by searching in the Thomson ISI database (the source also used by the authors of the Shanghai ranking) the articles published in *Nature* and *Science* by authors from several universities with relatively low numbers of such articles. The low number of articles permitted an easier computation of the indicator. I chose the following universities: Temple, Akron, Memphis and Auburn (USA). I searched in the ISI Web of Science database by choosing the interval 2000–2004 and using the queries *og=NAME and og=univ* and (so=science or so=nature)*, where NAME was one of Temple, Akron, Memphis or Auburn. I then manually processed the results, selected only the records with authors from the targeted universities (for example, the queries also returned articles from Auburn University in Scotland, etc.), records of article type, and counted the value of the indicators according to the position of the target universities' addresses in the affiliations list. The results are listed in Table 1, together with the total number of retrieved publications of article type with authors from the target universities, the value of the actual score for the N&S indicator in the Shanghai ranking for the target universities, noted with $s_{N\&S}$, and the ratio $s_{N\&S}/n_{N\&S}$. I also included in the table the total number of publications of article type in *Nature* and *Science* for Harvard and Stanford universities, from 2000–2004, retrieved with the queries *og=NAME univ* and (so=science or so=nature)*, where NAME was one of Harvard or Stanford, and then processed with a computer program to count only the publications of article type.

Table 1. Dependence between some computed values of the N&S indicator and the corresponding scores in the Shanghai ranking

University	*Nature* and *Science* publications of article type	$n_{N\&S}$	$s_{N\&S}$	$s_{N\&S}/n_{N\&S}$
Temple	2	0.6	3.8	6.33
Akron	2	1.1	4.9	4.45
Memphis	4	2.25	6.5	2.89
Auburn (USA)	4	0.95	5.2	5.47
Harvard	623		100.0	
Stanford	300		70.9	

It can be seen that the ratio $s_{N\&S}/n_{N\&S}$ is not a constant. Its average is 4.79, which seems much too high to be the value used in the Shanghai ranking. Given this ratio, and the value of 100 for the score of Harvard University at this indicator, it results that the value of the $n_{N\&S}$ indicator for Harvard is 20.88. Harvard has 623 publications of article type in *Nature* and *Science* in the given time interval, so its value for the $n_{N\&S}$ indicator should be at least 62.3, given that the minimum weigth is 10%. This suggests another discrepancy from proportionality in the computation of Shanghai ranking scores.

The size indicator

This indicator is given by "the weighted scores of the above five indicators [Alumni, Award, HiCi, N&S, SCI] divided by the number of full-time equivalent academic staff. If the number of academic staff for institutions of a country cannot be obtained, the weighted scores of the above five indicators is used. For ranking 2005, the numbers of full-time equivalent academic staff are obtained for institutions in USA, Japan, China, Italy, Australia, Netherlands, Sweden, Switzerland, Belgium, Slovenia, etc." I assumed that the weights used for computing this indicator, which are not specified, are the same that were used in computing the final score of the ranking. I thus computed the weighted sum of the five indicators as

$$s_p = (0.1\, s_{Alumni} + 0.2\, s_{Award} + 0.2\, s_{HiCi} + 0.2\, s_{N\&S} + 0.2\, s_{ISI})/0.9,$$

using the data for the top 500 universities from the Shanghai ranking. I then computed the size r of the universities relative to the size of Harvard University that was used in the Shanghai ranking, using the formula

$$r = 0.724\, s_p/s_{Size},$$

where s_{Size} is the score for the Size indicator. This is because Harvard University has s_p=100 and s_{Size}=72.4, and because s_{Size} is proportional to s_p / n_{Size}, according to the methodology, where n_{Size} is the number of full-time equivalent (FTE) academic staff. I

then tried to obtain the coefficient of proportionality between r and n_{Size}, using data for the FTE academic staff obtained from the internet, for a few universities.

Reproducing this indicator is difficult, as the definition of „full time academic staff" is ambiguous – does this include researchers, visitors, etc. or not? Moreover, available data is scarce and inconsistent – many universities report the headcount of faculty, but few report the FTE; and when reporting the academic staff, the included categories of personnel vary from case to case. However, there is at least a source that permits verifying the data of the Shanghai ranking in a consistent manner. The University of California (UC) publishes on the internet detailed data on the FTE academic staff of its composing Universities,[7] of which many are included in the Shanghai top 500. I present below data corresponding to the situation of UC personnel in October 2004. I presumed that the authors of the Shanghai ranking used the same data, if they made the ranking during the first months of 2005, as the next available data is for April 2005. The results obtained with the two data sets are, however, similar.

Table 2. Dependence between the size of UC universities relative to Harvard, as used in the Shanghai ranking, and some actual size data

University	Size relative to Harvard (according to Shanghai ranking)	Regular teaching faculty (ladder ranks, acting ranks), lecturers, other teaching faculty	Ratio	Regular teaching faculty (ladder ranks, acting ranks), lecturers, other teaching faculty, researchers	Ratio	Regular teaching faculty (ladder ranks, acting ranks), lecturers, other teaching faculty, researchers, student assistants	Ratio	All academic staff	Ratio
UC Berkeley	1.00	1630	1630.77	2786	2787.32	4697	4699.23	5041	5043.39
UC San Diego	0.77	1614	2083.71	2691	3474.14	4355	5622.40	4525	5841.88
UC Los Angeles	1.07	3029	2821.74	4194	3907.02	6811	6344.95	7121	6633.74
UC San Francisco	0.75	1789	2394.47	3063	4099.65	4177	5590.68	4243	5679.01
UC Santa Barbara	0.73	976	1345.56	1433	1975.60	2265	3122.64	2361	3254.99
UC Irvine	0.72	1532	2128.95	2182	3032.23	3635	5051.40	3741	5198.70
UC Davis	0.78	2027	2589.26	3583	4576.87	5636	7199.35	5793	7399.90
Relative standard deviation			24.30%		25.91%		24.03%		23.53%

Table 2 lists the size r of several UC branches relative to the size of Harvard University, as it was used in the Shanghai ranking, together with several values for its FTE academic staff (obtained by including various categories of staff included in the academic staff, as listed on the cited website presenting the data on UC personnel), and

the ratios of these values to *r*. If the Shanghai ranking used the same data for counting the FTE academic staff for UC branches (and it should have done so as this information can be easily found on the internet), and the values of the Size indicator were computed according to the official methodology that I also followed, these ratios should have been constant. However, it can be seen from the table below that the ratios vary considerably, with a relative standard deviation of about 24%.

Conclusions

The data presented here suggests that the results of the Shanghai ranking are irreproducible. At least the data concerning the SCI indicator suggests that the authors of the Shanghai ranking deviated from the official published methodology when computing the scores of the universities. It is understandable that the values of some of the indicators used are hard to reproduce, as in the case of the Size indicator, where data about universities' personnel are hard to obtain and inconsistent, and sometimes requires using educated guesses, or as in the case of the N&S indicator, where the necessity of weigthing as a function of author importance may require an error-prone counting method. It is less acceptable that the values for an objective indicator such as SCI cannot be reproduced using the public methodology. These findings undermine the relevance of the Shanghai ranking, and adds to other critiques of its methodology and results.[8]

In correspondence through email with the authors of the Shanghai ranking, they insist (in response to a draft version of this paper) that the results found here do not mean that the ranking results are irreproducible, since "any two institutions having the same amount of SCI articles, they should get the same scores in our ranking". I believe that reproducibility means that, given correct raw data for one university inside top 500, one should be able to compute its score according to the published methodology, and this score should be equal to the score used in Shanghai ranking. This has practical importance, since, for example, it also permits computing scores for universities that are not in top 500, and permits devising strategies that these universities may use for entering the top 500. Since I have shown that this is not possible, and one has instead to get the distribution of data for most of the universities in top 500 in order to infer the dependencies used in the ranking, which are not the ones in the published methodology, I consider that the results are irreproducible, given the currently published methodology. The inconsistencies found here may be clarified if the authors of the Shanghai ranking will publish also the un-normalized values that they used for the indicators, and the actual transformations they performed on data. It would be useful if they would insure the transparency of their results, or at least implement on their website the possibility that the public compute automatically the score of any university, given raw data.

References

1. http://ed.sjtu.edu.cn/rank/2005/ARWU2005Main.htm
2. R. FLORIAN, *Universităţile din România şi clasamentul Shanghai* [Romanian universities and the Shanghai ranking; in Romanian], Ad Astra 5, 2006, http://www.ad-astra.ro/journal/8/florian_shanghai_romania.pdf
3. http://ed.sjtu.edu.cn/rank/2005/ARWU2005Methodology.htm; another document presenting a slightly different methodology is also linked from the ranking homepage and is available at http://ed.sjtu.edu.cn/rank/file/ARWU-M&P.pdf, but this latter one presents the methodology of the 2004 ranking, not of concern to us.
4. http://ed.sjtu.edu.cn/rank/2005/ARWU2005.xls
5. http://ai.ijs.si/mezi/iassatena/shanghai-relative.html
6. http://scientific.thomson.com/mjl/
7. http://www.ucop.edu/ucophome/uwnews/stat/
8. VAN RAAN, A. F. J., Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62 (1) (2005) 133–143.