

DOI:10.1145/3181854

Strong correlation indicates notable research productivity of individual faculty members in turn boosts the standing of their programs.

BY SLOBODAN VUCETIC, ASHIS KUMAR CHANDA, SHANSHAN ZHANG, TIAN BAI, AND ANIRUDDHA MAITI

Peer Assessment of CS Doctoral Programs Shows Strong Correlation with Faculty Citations

RANKINGS OF UNIVERSITIES and specialized academic programs have a major influence on students deciding what university to attend, faculty deciding where to work, government bodies deciding where and how to invest education and research funding, and university leaders deciding how to grow their institutions.⁹ There is general agreement in scientometrics that the quality of a university or a program depends on many factors, and different ranking metrics might be appropriate for different types of users. However, major points of contention emerge when it comes to agreeing on ranking methodology.²⁰ Given the increasing impact of rankings, there is a need to better understand the actors

influencing rankings and come up with a justifiable, transparent formula that encourages high-quality education and research at universities.¹¹ We aim to contribute toward achieving this objective by focusing on ranking of the U.S. doctoral programs in computer science.

We broadly group quality measures into objective (such as average research funding per faculty member) and subjective (such as peer assessment). The influential *U.S. News* ranking of computer science doctoral programs^a is based purely on peer assessment in which computer science department chairs are asked to score other computer science programs on a scale of 1 to 5, with 1 being “marginal” and 5 being “outstanding,” or enter “do not know” if not sufficiently familiar with the program. The final ranking is obtained by averaging the individual scores. Due to the subjective nature of peer assessment, the factors influencing the *U.S. News* ranking of computer science programs remain hidden. Unlike the *U.S. News* ranking, the Academic Ranking of World Universities (ARWU)¹⁷ ranking of computer science programs is based on objective measures (such as

a <http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-science-schools/computer-science-rankings>

» key insights

- A correlation of 0.935 exists between the *U.S. News* peer assessment of computer science doctoral programs and the Scholar ranking we obtained by combining the average number of citations of professors in a program and the number of highly cited faculty in the same program.
- The top 62 ranked computer science doctoral programs in the U.S. per the *U.S. News* peer assessment are much more highly correlated with the Scholar ranking than are the next 57 ranked programs, indicating deficiencies of peer assessment of less-well-known programs.
- University reputation seems to positively influence peer assessment of computer science programs.



counts of papers published in computer science journals and number of highly cited faculty). The final ranking is a weighted average of these measures. The scientometrics community criticized this approach because the choice of weights is not clearly justified.^{4,6} The *U.S. News* ranking of doctoral programs in engineering^b uses a weighted average of objective measures and subjective measures. As with the ARWU, justification for the ranking formula is lacking.

Ranking of computer science doctoral programs published in 2010 by the U.S. National Research Council (NRC)² is notable for its effort to provide a justifiable ranking formula. The

NRC collected objective measures and surveyed faculty to assess peer institutions on multiple measures of perceived quality. The NRC ranking group then built a regression model that predicts subjective measures based on the objective measures. The resulting regression model was used to provide ranking order. Unfortunately, the subjective and objective data collected during the NRC ranking project had questionable quality³ and the resulting ranking did not find good reception in computer science community.^c

We find the NRC idea of calculating the ranking formula through regression modeling better justified than the alternatives. In this article, we address the data-quality issue that plagued the

NRC ranking project by collecting unbiased objective data about programs in the form of faculty-citation indices and demonstrate that regression analysis is a viable approach for ranking computer science doctoral programs. We also obtain valuable insights into the relationship between peer assessments and objective measures.

Ranking Data

Our data was collected in Fall 2016 by a team of three undergraduate students, five computer science graduate students, and one professor.

U.S. News ranking. *U.S. News* provides a well-known ranking of graduate programs in the U.S. We downloaded the most recent—the 2014 ranking of 173 U.S. computer science doctoral programs based on peer assessments administered from 2009 to 2013. Doc-

b <https://www.usnews.com/education/best-graduate-schools/articles/engineering-schools-methodology?int=9d0e08>

c <http://www.chronicle.com/article/Too-Big-to-Fail/127212/>

total programs receiving average score of at least 2.0 from their peers were ranked and their scores published. In the rest of this article, we refer to the 2014 *U.S. News* peer-assessment scores of computer science doctoral programs as “USN CS scores.” We also accessed the 2017 *U.S. News* National University Ranking^d that evaluated the quality of undergraduate programs at U.S. universities. Each university in this ranking was assigned a score between 0 and 100 based on numerous measures of quality.^e We refer to those scores as the “USN university scores.”


Of the 173 doctoral programs ranked in 2014, *U.S. News* assigned scores of 2.0 or higher to 119 programs, while 54 programs had scores below 2.0 that went unpublished; 17 programs had scores of 4.0 or higher, and four programs—Carnegie Mellon University, Massachusetts Institute of Technology, Stanford University, and the University of California at Berkeley—had the maximum possible score of 5.0. The Pearson correlation between USN CS scores and 2017 USN university scores for 113 programs covered by both rankings was relatively high at 0.681.

CS faculty. We manually collected the names of 4,728 tenure-track professors of computer science from the 173 programs ranked by *U.S. News*. To be counted as a professor of computer science, tenure-track faculty had to be listed on a website of a computer science department or college.


In a number of universities, computer science faculty are part of joint departments or colleges, making it more difficult for us to identify faculty from the “people” pages. Since boundaries between computer science and related disciplines are not always clear, we decided to err on the generous side and count as computer science faculty all faculty members from such joint departments and colleges with at least some publications in computer science journals and proceedings. Another issue was dealing with affiliated faculty or faculty with secondary or joint appointments in computer sci-

^d <https://www.usnews.com/best-colleges/rankings/national-universities>

^e <https://www.usnews.com/education/best-colleges/articles/how-us-news-calculated-the-rankings>



The main contribution of this work is in showing there is a strong correlation between peer assessments and citation measures of computer science doctoral programs.



ence departments. When people pages clearly separated such faculty from primary appointments, we included only the primary appointments in our list. When the people pages did not provide discriminable information about affiliations, we included all listed tenure-track faculty. The details of faculty selection for each university are in the “CS Department Data” file we maintain on our ranking webpage.^f

Overall, we collected the names of 4,728 tenure-track faculty members, including 1,114 assistant professors, 1,271 associate professors, and 2,343 full professors. Since assistant professors are typically only starting their academic careers and publication records, we treated them differently from associate and full professors, and for the rest of this article, we refer to associate and full professors as “senior faculty.”

The distribution of program size is quite varied, with median faculty size of 22 positions, mode of 15, minimum of four, and maximum of 143 (CMU). The Pearson correlation between department size and USN CS score of the 119 programs ranked by *U.S. News* is 0.676, indicating larger departments are more likely to be higher ranked.

Faculty citations. Of the 4,728 faculty we included, 3,359 had Google Scholar profiles (71.0% coverage), and of the 3,614 senior faculty, 2,453 (67.9% coverage) had Google Scholar profiles. A Google Scholar profile includes all publications of a faculty member with citation counts for each paper and aggregate citation measures (such as the *h*-index,¹² which represents the highest integer x for which it can be claimed the author published x papers that are cited at least x times).

One option for us when we began our research was to use only citation data of faculty with Google Scholar profiles. However, we observed that less-cited faculty are less likely to have the profile and that citation data we obtained from the 3,359 profiles would be biased. To collect data with reduced bias, we introduced a new citation measure we call *t10*.

t10 index. We define the *t10* index as the number of citations of a faculty member’s 10th most-cited paper

^f http://www.dabi.temple.edu/~vucetic/CSranking/raw_data/all_departments.xlsx

and find it more convenient than the h -index because it is easier to obtain through manual search. For example, rather than having to find the 50 most-cited papers authored by a particular faculty member to establish his or her h -index is 50, $t10$ could be obtained by identifying only the faculty member's 10 most-cited papers.

We obtained $t10$ for 4,352 of the 4,728 faculty (92.0% coverage) and 3,330 of the 3,614 senior faculty (92.1% coverage) through manual search of Google Scholar. We did not collect $t10$ for 8% of the faculty whose names were too common to allow reasonably quick manual extraction. Since a faculty member's name should not influence his or her citation count, the 92% sample of faculty with known $t10$ can be treated as an unbiased sample of computer science faculty.

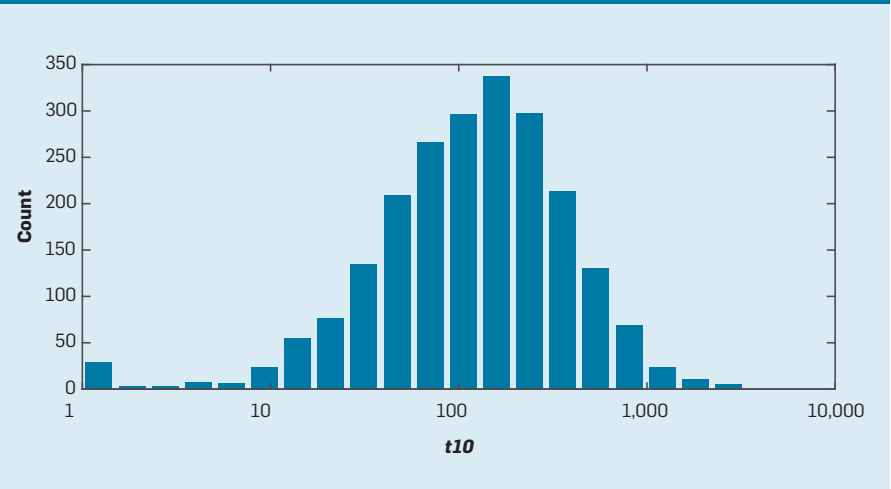
Unlike $t10$, our results show a sample of faculty with h -index is indeed biased. While $t10$ median for the 3,330 senior faculty was 89, it increased to 111 among 2,453 of faculty who also have Google Scholar profiles and dropped to 44 among 877 of those without such a profile. Among the 10% of the least-cited senior faculty, 65.3% did not have a Google Scholar profile, while among the 10% of the most-cited faculty, only 11.1% were without the profile. These results validate our effort to collect $t10$ and use it, instead of, say, h -index, in our study.

Figure 1 is a histogram of $t10$ for the 3,330 senior faculty in the study. A bump at low values represents 89 senior faculty with $t10 = 0$ with fewer than 10 cited papers listed in Google Scholar. The median $t10$ was 89, and the percentiles of $t10$ are reported in Table 1. For example, to be in the 90th percentile of all senior computer science faculty in the U.S., a faculty member must have published at least 10 papers cited at least 370 times. Pearson correlation between logarithms of h -index and $t10$ for the 2,453 senior computer science faculty having both indices was 0.937, further justifying our use of $t10$ as a replacement for the h -index.

Measuring Program Strength

We thus propose two approaches for using individual faculty citation indices to calculate citation strength of a particular university's program.

Figure 1. Histogram of $t10$ of associate and full professors of computer science.



Averaged citation measures. One way to measure program strength is to average citations of its individual faculty members.¹⁵ We explore here three different averaging schemes: The first is calculated as the median of $t10$ values of senior faculty, denoted as $m10$. The second is calculated as the geometric mean of $(1+t10)$ values of senior faculty, denoted as $g10$. The third averages $t10$ percentiles of senior faculty, denoted as $p10$. We did not count assistant professors for any of the averaged measures because their citation numbers are typically smaller and their inclusion would hurt departments with many assistant professors.

Cumulative citation measures. Another way to measure the strength of a program is to count the program's highly cited faculty. To define a highly cited faculty member, we had to decide on a $t10$ threshold. We considered all faculty above the $t10$ threshold highly cited. We introduced cN to denote the number of faculty whose $t10$ is greater than $N\%$ of all the senior faculty. For example, $c40$ counts faculty with $t10$ greater than 40% of all senior faculty in the study. To find cN for a particular university's computer science program,

we considered its faculty at all ranks, including assistant professors.

Table 2 reports the correlation between citation measures and USN CS scores. The "Original" row lists Pearson correlation between 3 averaged— $m10$, $g10$, and $p10$ —and 4 cumulative— $c20$, $c40$, $c60$, $c80$ —measures and USN CS scores. Values range from 0.794 to 0.882, indicating a strong correlation between the citation measures and peer assessments. Since the distribution of most of our citation measures was heavy-tailed, we also explored their logarithmic and square-root transformation. The correlation between the transformed measures and USN CS scores are reported in Table 2 rows "Log" and "Sqrt." The square-

Table 1. Percentiles of $t10$.

Percentile	$t10$	Percentile	$t10$
10%	16	70%	166
20%	31	80%	231
30%	47	90%	370
40%	65	95%	543
50%	90	98%	780
60%	123	99%	1,078

Table 2. Correlation between averaged program measures and U.S. News computer science scores.

Transform	$m10$	$g10$	$p10$	$c20$	$c40$	$c60$	$c80$
Original	0.880	0.875	0.882	0.794	0.844	0.874	0.842
Log	0.865	0.856	0.870	0.807	0.840	0.875	0.904
Sqrt	0.890	0.887	0.861	0.825	0.877	0.909	0.906

root transformation of $m10$ has the greatest correlation (0.890) among averaged citation measures, and the square root of $c60$ has the greatest correlation (0.909) among cumulative citation measures. This result supports our original hypothesis that peer assessment is closely tied to research productivity of individual faculty members.

Regression Analysis

Our preliminary results found that combining one averaged and one cumulative citation measure increases correlation with the USN CS scores. They also found that linear regression with two measures is nearly as successful as linear regression with more than two measures or nonlinear regression with two or more measures. We thus used linear regression models of type $s_i = \beta_0 + \beta_1 a_i + \beta_2 c_i$, where s_i is the predicted USN CS score, a_i is an aggregated citation measure, and c_i is a cumulative citation measure of the i^{th} program. The regression parameters are β_0, β_1 , and β_2 . Instead of learning the intercept parameter β_0 , we set it to $\beta_0 = 1$ by default. A justification is that a given university computer science program with $a_i = 0$ and $s_i = 0$ does not have research-active faculty and, based on the peer-assessment instructions by *U.S. News*, such a

program would be scored as 1 (“Marginal”).

By combining one of the averaged citation measures ($\sqrt{m10}, \sqrt{g10}, p10$) and one of the cumulative citation measures ($\sqrt{c40}, \sqrt{c60}, \sqrt{c80}$) we trained nine different regression models. For that training, we used 119 computer science doctoral programs ranked by *U.S. News*. The correlation of all nine models with USN CS scores ranged from 0.920 to 0.934, which is greater than for any of the individual citation measures in Table 2. The best four models combined either $\sqrt{m10}$ or $\sqrt{g10}$ averaged measure and $\sqrt{c40}$ or $\sqrt{c60}$ cumulative measure; their parameters are reported in Table 3. The best overall model, which achieved $R^2 = 0.869$ and Pearson correlation 0.934, combines $\sqrt{m10}$ and $\sqrt{c60}$ and measures

$$s = 1 + 0.130\sqrt{m10} + 0.218\sqrt{c60}$$

If the median faculty members in a given computer science program has $t10 = 100$ and there are nine faculty over the 60th percentile (with $t10 \geq 123$) based on $t10$ index, the calculated score of that program would be 2.95.

By averaging the output of four regression models in Table 3, we obtained a joint model (last row in Table 3), with $R^2 = 0.874$ and Pearson corre-

lation 0.935, making it more accurate than any of the individual models. Figure 2 is a scatter plot of the USN CS scores and the joint model scores for the 173 programs we studied. For the 55 computer science doctoral programs not ranked by *U.S. News*, we set their default USN CS score to 1.5 in Figure 2. The result illustrates a strong correlation between the peer-assessed USN CS scores and the objectively measured joint-model scores.

A closer look at the scatterplot reveals two groups of computer science programs can be distinguished with respect to the correlation between joint model scores and USN CS scores. The first group includes 62 programs scored 2.7 and higher by *U.S. News*. The correlation between the USN CS scores and joint-model scores in this group was 0.911. The second group included 57 programs with USN CS scores between 2.0 and 2.6. The correlation between the USN CS scores and joint-model scores in this group was only 0.360. Our hypothesis is that the programs with USN CS scores between 2.0 and 2.6 might not be sufficiently well known among their peers at the national level to allow objective and reliable peer assessments.

The issue of reliability raises the question as to whether USN CS scores of the programs scored below 2.7 might be too noisy for us to include in our regression. To explore this, we trained another joint model using data from the 62 programs that scored above 2.6 by *U.S. News* ranking.

Scholar Model

When evaluated on the top 62 computer science programs, the new updated joint model produced $R^2 = 0.830$ and correlation 0.913. When measured on all 119 ranked programs, the new updated joint model produced $R^2 = 0.872$

Figure 2. Aggregated model scores and USN CS scores of 173 computer science graduate programs compared; for 55 programs not ranked by *U.S. News*, we set the default score at 1.5 out of 5.

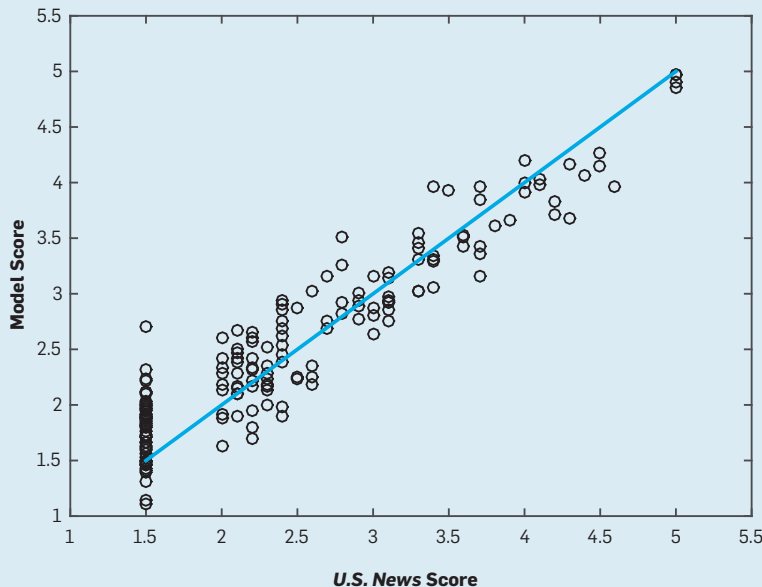


Table 3. Parameters of the four best individual ranking models and the Scholar model.

	$\sqrt{m10}$	$\sqrt{g10}$	$\sqrt{c40}$	$\sqrt{c60}$
Model 1	0.111		0.223	
Model 2	0.130			0.218
Model 3		0.113	0.225	
Model 4		0.133		0.220
Joint Model	0.060	0.062	0.112	0.109

and correlation 0.935, virtually identical to the joint model in Table 3. We thus concluded that USN CS scores of programs ranked from 2.0 to 2.6 are indeed too noisy to be helpful. As a result, we endorse a joint model trained on the top 62 programs as the best model for ranking computer science doctoral programs, calling it the “Scholar model” and its outputs the “Scholar scores.”^g The Scholar scores are calculated as

$$s = 1 + 0.058\sqrt{m10} + 0.059\sqrt{g10} + 0.121\sqrt{c40} + 0.127\sqrt{c60}$$

Impact of Reputation

We were also interested in the effect of university reputation on the ranking of computer science programs. We thus trained regression models of type $s_i = 1 + \beta_1 a_i + \beta_2 c_i + \beta_3 u s_i$, where $u s_i$ is USN university score of the i^{th} university. Note the maximum USN university score was 100 (Princeton) and the lowest score was 20. For universities that did not have a published score, we set that score at 20 by default. By averaging four regression models that use one of the averaged measures— $\sqrt{m10}$ or $\sqrt{g10}$ —one of the cumulative measures— $\sqrt{c20}$ or $\sqrt{c60}$ —and 2017 USN university score, all trained on 119 universities with a USN CS score 2.0 and greater, the resulting updated joint model had $R^2 = 0.888$ and correlation 0.942, which is an increase in accuracy compared to the joint model in Table 3. This result indicates university reputation might have an impact on peer assessments of computer science doctoral programs.

To help explain it, consider the most accurate individual model whose $R^2 = 0.884$ and correlation 0.941

$$s = 1 + 0.090\sqrt{m10} + 0.238\sqrt{c60} + 0.0061 \cdot us$$

This model thus adds 0.61 to the score of the doctoral program at Princeton and only 0.12 to programs from universities not ranked by the 2017 *U.S. News National University* ranking. As a result, if the citation measures of two computer science programs are identical, their scores assigned by this model

^g For Scholar scores of the 173 computer science doctoral programs, see <http://www.dabi.temple.edu/~vucetic/CSranking/>

Beyond measures of publication quality, other measures have also been proposed, including faculty recognition, student placement, student selectivity, research funding, resources, and diversity.

could differ by as much as 0.49 assessment points, depending on their university score.

Qualitative Analysis of Scholar Scores

Based on R^2 of the Scholar model, we see that measures derived from the faculty citations we collected in Fall 2016 can explain 87.4% of the variance of peer-assessed USN CS scores. For qualitative analysis, we looked at the programs with the greatest discrepancies between USN CS scores and Scholar scores.

Among programs with Scholar scores significantly greater, one significant group included those with fewer than 25 faculty members, including the University of California, Santa Cruz (+0.7), Colorado State University (+0.6), and Lehigh University (+0.6). We hypothesize it is more likely that a surveyed peer does not know any faculty members in smaller programs and might thus lead to conservative ratings. Another dominant group included programs that have recently experienced significant growth, including New York University (+0.6), the University of California, Riverside (+0.5), and Northeastern University (+0.5). This might be explained by the lag between *U.S. News* peer assessments (collected in 2009 and 2013) and citation measures (collected in Fall 2016).

Among programs ranked significantly lower by the Scholar model, many are hosted at universities with strong non-computer science departments (such as electrical engineering and computer engineering) in which a number of faculty members publish in computer science journals and proceedings but we did not select for our list of computer science faculty based on our inclusion criteria. It is possible that inclusion of such faculty would increase Scholar scores of the related programs.

Discussion

The main contribution of this work is in showing there is a strong correlation between peer assessments and citation measures of computer science doctoral programs. This result is remarkable considering the subjective nature of peer assessments, demon-

strating that committees of imperfect raters are able to produce good decisions, as has been observed in many other settings.^{5,19}

An open question is: Can the correlation between peer assessment and citation measures be further improved? It would certainly help reduce the time gap between collecting peer assessments and citation measures. Further improvements could be achieved by addressing several concerns about peer assessments and objective measures.

On the peer-assessment side, we observed the quality of smaller or less-known programs might be underestimated. We also observed that peers might overestimate computer science doctoral programs at highly reputable universities. The root of both issues might be difficulty by the peers to obtain relevant information about a large number of programs from a survey. A remedy might involve collecting and publishing unbiased and objective measures about computer science doctoral programs.

The citation measures we collected have several complications. One is related to the definition of a computer science doctoral program; for example, does it refer to an administrative unit (such as a computer science department) and its primary faculty or to all computer science-related faculty in a given university? While we used the former definition, the latter might be equally valid. Another is the quality of the Google Scholar data we used. Although automated Web crawling^{7,13} used by Google Scholar is imperfect, its advantage is its broad coverage of journal and conference papers, both important in computer science. Our proposed *t10* index also has its limitations; for example, while it includes self-citations, it is blind to location in the author list and number of co-authors.¹⁶ To calculate the citation measures of a doctoral program, we relied on aggregating citation indices of its individual faculty members. Such aggregation is blind to the research field of the faculty members, possibly hurting sub-disciplines with smaller communities or requiring more work to publish a paper.¹⁸

We used only Google Scholar citation data to create objective measures

of program quality. It is likely that additional measures, if collected in an unbiased manner and with sufficient quality, might further improve the explained variance of regression models;^{1,17} for example, a recently created ranking of computer science departments^h is based on counts of faculty publications in selected conferences, ensuring the less-represented sub-areas of computer science are given greater weight in ranking and down-weights papers with many co-authors.


Future Assessment

Beyond measures of publication quality, other measures have also been proposed, including faculty recognition, student placement, student selectivity, research funding, resources, and diversity. Moving forward, it might be helpful for the computer science community to create a public repository of objective program measures. For such a resource to be useful, it would also have to contain raw data, a detailed description of data-collection process, and any potential issues with data quality, as well as all the relevant code. Such a resource would be very useful for peer assessment by providing peers objective, unbiased information about the assessed programs.

One key caveat when ranking universities and programs based on objective measures is the potential risk of it being susceptible to gaming.^{8,10,14} To be truly useful, any ranking should thus be examined for potential negative incentives for change or for the presence of shortcuts that could artificially improve rankings.

The resulting rankings, raw data, and codes used in this study are publicly available at <http://www.dabi.temple.edu/~vucetic/CSranking/>.

Acknowledgments

We thank Sharayu Deshmukh, Charis Yoo, Mary Margaret Okonski, and Taha Shamshudin from Temple University for their help with data collection. 

^h <http://csranks.org/>

References

1. Aguilto, I.F., Bar-Ilan, J., Levene, M., and Ortega, J.L. Comparing university rankings. *Scientometrics* 85, 1 (Feb. 2010), 243–256.
2. Altbach, P.G. The dilemmas of ranking. *International Higher Education* 25, 42 (Mar. 2015); <https://ejournals.bc.edu/ojs/index.php/ihe/article/view/7878>

3. Bernat, A. and Grimson, E. Doctoral program rankings for U.S. computing programs: The national research council strikes out. *Commun. ACM* 54, 12 (Dec. 2011), 41–43.
4. Billaut, J.C., Bouyssou, D., and Vincke, P. Should you believe in the Shanghai ranking? *Scientometrics* 84, 1 (July 2010), 237–263.
5. Black, D. On the rationale of group decision-making. *Journal of Political Economy* 56, 1 (Feb. 1948), 23–34.
6. Docampo, D. Reproducibility of the Shanghai Academic Ranking of World Universities results. *Scientometrics* 94, 2 (Feb. 2013), 567–587.
7. Delgado Lopez-Cozar, E., Robinson-Garcia, N., and Torres-Salinas, D. The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology* 65, 3 (Mar. 2014), 446–454.
8. Gertler, E., Mackin, E., Magdon-Ismael, M., Xia, L., and Yi, Y. Computing manipulations of ranking systems. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems* (Istanbul, Turkey, May 4–8). ACM Press, New York, 2015, 685–693.
9. Hazelkorn, E. Learning to live with league tables and ranking: The experience of institutional leaders. *Higher Education Policy* 21, 2 (June 2008), 193–215.
10. Hazelkorn, E. How rankings are reshaping higher education. Chapter in *Los Rankings Universitarios, Mitos y Realidades*, V. Climent, F. Michavila, and M. Ripolles, Eds. Tecnos, Dublin Institute of Technology, Ireland, 2013, 1–9.
11. Hicks, D., Wouters, P., Waltman L., De Rijcke, S., and Rafols, I. The Leiden Manifesto for research metrics. *Nature* 520, 7548 (Apr. 23, 2015), 429–431.
12. Hirsch, J.E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102, 46 (Nov. 15, 2005), 16569–16572.
13. Jacsoo, P. Deflated, inflated and phantom citation counts. *Online Information Review* 30, 3 (May 2006), 297–309.
14. Kehm, B.M. and Erkkila, T. Editorial: The ranking game. *European Journal of Education* 49, 1 (Mar. 2014), 3–11.
15. Lazaridis, I. Ranking university departments using the mean *h*-index. *Scientometrics* 82, 2 (Feb. 2010), 211–216.
16. Lin, C.S., Huang, M.H., and Chen, D.Z. The influences of counting methods on university rankings based on paper count and citation count. *Journal of Informetrics* 7, 3 (July 2013), 611–621.
17. Liu, N.C. and Cheng, Y. The Academic Ranking of World Universities. *Higher Education in Europe* 30, 2 (July 2005), 127–136.
18. Radiocchi, F. and Castellano, C. Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics* 6, 1 (Jan. 2012), 121–130.
19. Raykar, V.C., Yu, S., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., and Moy, L. Learning from crowds. *Journal of Machine Learning Research* 11 (Apr. 2010), 1297–1322.
20. Saisana, M., d'Hombres, B., and Sattelli, A. Rickety numbers: Volatility of university rankings and policy implications. *Research Policy* 40, 1 (Feb. 2011), 165–177.

Stobodan Vucetic (vucetic@temple.edu) is a professor in the Department of Computer and Information Sciences at Temple University, Philadelphia, PA, USA.

Ashis Kumar Chanda (ashis@temple.edu) is a doctoral student in the Department of Computer and Information Sciences at Temple University, Philadelphia, PA, USA.

Shanshan Zhang (zhang.shanshan@temple.edu) is a doctoral student in the Department of Computer and Information Sciences at Temple University, Philadelphia, PA, USA.

Tian Bai (tianbai@temple.edu) is a doctoral student in the Department of Computer and Information Sciences at Temple University, Philadelphia, PA, USA.

Aniruddha Maiti (tuf86648@temple.edu) is a doctoral student in the Department of Computer and Information Sciences at Temple University, Philadelphia, PA, USA.