

Emerging Technology From the arXiv December 7, 2015

Technology Review

Wikipedia-Mining Algorithm Reveals World's Most Influential Universities

An algorithm's list of the most influential universities contains some surprising entries.



Where are the world's most influential universities? That's a question that increasingly dominates the wawfthe public, governments, and funding agencies think about research and higher education. 01/21/2016 12:38 PM

Twaspeakammatinguage is that it's hard two reduces objective ranking of almost anything let alcom/view/544266/w... universities. Cultural, historical, and geographical factors can all influence these rankings in ways that are hard to quantify.

So an independent way of producing a ranking that avoids these controversies would be widely welcomed.

Today, we get such a ranking thanks to the work of Jose Lages at the University of Franche-Comte in France and a few pals. They've used the way universities are mentioned on Wikipedia to produce a world ranking. Their results provide a new way to think about rankings that may help to avoid some of the biases that can occur in other ranking systems.

Biases can crop up remarkably easily. For example, in the last century, English has become the *de facto* language of science, and the advantage this gives English-speaking countries is hard to quantify.

And there are other factors that are unique to university rankings. Some institutions focus more on teaching than on research – how should these factors be balanced?

The new work attempts to get around some of these problems using the Pagerank algorithm that Google famously uses to rank websites in search results. This uses the network of links between nodes on a network to determine those that are the most important.

The key insight is that the algorithm counts a node as important if other important nodes point to it. So it repeatedly works through the links, recalculating the importance of every node on each iteration, to come up with a ranking.

Exactly this process can be applied to Wikipedia articles. Each university mentioned in an article is a node in the network, and the links pointing toward it are used to determine a ranking (see also "<u>Artificial Intelligence Aims to Make Wikipedia Friendlier and Better</u>").

Lages and co apply this process to 24 different language editions of Wikipedia. This database contains some four million articles in English, 1.5 million in German and around a million in each of French, Dutch, Italian, Spanish, and Russian. It also includes Chinese, Hebrew, Hungarian, and so on. "These 24 languages cover 59% of world population and 68% of the total number of Wikipedia articles in all 287 languages," they say.

The team first determines a ranking for each language and point out that each language edition tends to favor its own universities. So the top 100 list in French includes 32 French-speaking universities, the top 100 in German includes 63 German-speaking universities, and so on.

They then combine the lists to produce a global ranking. The top 20 most influential universities ranked in this way are:

12.0f.41. University of Cambridge U.K. 2. University of Oxford U.K.

Wharvershum Reveals World's Mos...

- 4. Columbia University U.S.
- 5. Princeton University U.S.
- 6. Massachusetts Institute of Technology U.S.
- 7. University of Chicago U.S.
- 8. Stanford University U.S.
- 9. Yale University U.S.
- 10 University of California, Berkeley U.S.
- 11. Humboldt University of Berlin, Germany
- 12. Cornell University U.S.
- 13. University of Pennsylvania U.S.
- 14. University of London U.K.
- 15. Uppsala University Sweden
- 16. University of Edinburgh U.K.
- 17. Heidelberg University Germany
- 18. University of California, Los Angeles U.S.
- 19. New York University U.S.
- 20. University of Michigan U.S.

The full 100 are at: http://perso.utinam.cnrs.fr/~lages/datasets/WRWU/theta PR.php.

There are many familiar names on this list but there are also some interesting differences with conventional rankings. Perhaps the most influential of these rankings is the Academic Ranking of World Universities compiled by Shanghai Jiao Tong University since 2003.

The top 20 from this ranking (from 2013, when the Wikipedia database was compiled) are these:

- 1. Harvard University U.S.
- 2. Stanford University U.S.
- 3. University of California, Berkeley U.S.
- 4. Massachusetts Institute of Technology U.S.
- 5. University of Cambridge U.K.
- 6. California Institute of Technology U.S.
- 7, Princeton University U.S.
- 8. Columbia University U.S.
- 9. University of Chicago U.S.
- 10. University of Oxford U.K.
- 11. Yale University U.S.
- 12. University of California, Los Angeles U.S.
- 13. Cornell University U.S.
- 14. University of California, San Diego U.S.
- 15. University of Pennsylvania U.S.
- 16. University of Washington U.S.
- 17. The Johns Hopkins University U.S.
- 18. University of California, San Francisco U.S.
- 19. University of Wisconsin, Madison U.S.
- 2005 Wiss Federal Institute of Technology Zurich, Switzerland

http://www.technologyreview.com/view/544266/w...

Lagespench on make Approximer exting above wations of on a start, they point out that the Wikipedia list m/view/544266/w... tends to favor older universities that have had a greater cultural impact. For example, Humboldt University of Berlin is ranked 11 on the Wikipedia list but does not appear in the top 100 of the conventional ranking, surprising for an institution that has educated 29 Nobel Prize winners. The inclusion in the new ranking is perhaps because of the greater cultural and historical importance of this university in the arts and humanities rather than sciences.

The diversity of countries is greater in the Wikipedia list, including universities from Africa such as AI-Azhar University in Egypt, for example. Japanese and Indian universities are more prominent. Germany is the second highest ranked country after the U.S. and followed by the U.K.

By contrast, the conventional list ranks the U.S. most highly followed by the U.K. and then Australia. In general, U.S. universities are less prominent in the new ranking, accounting for 38 percent of the total. By contrast, more than half the universities in the conventional ranking are from the U.S.

The new ranking isn't perfect, of course. It lists the University of London at 14although this institution is actually comprised of several institutions such as University College London and Kings College London, which have their own separate listings.

The database does not include languages such as Ukrainian, which almost certainly introduces other biases. And Wikipedia articles are open to abuse, which may allow future rankings to be influenced by nefarious practices.

Nevertheless, the new ranking has some merit as an objective approach. While universities generally play down the significance of these kinds of rankings, these lists can have a significant influence over funding. The French strategy toward higher education and research, in particular, is thought to have been significantly influenced by the Shanghai rankings (which may also explain the interest of the authors in this topic).

Of course, the Wikipedia ranking is unlikely to replace conventional rankings – there are significant vested interests at work. However, it provides a new way to analyze the current state of affairs and should add to the debate in a useful way.

Ref: http://arxiv.org/abs/1511.09021: Wikipedia Ranking of World Universities

Tagged: Web

Reprints and Permissions | Send feedback to the editor

MIT Technology Review © 2016 v1.13.05.10