Ranking of World Universities from 2017 Wikipedia Network

Célestin Coquidé[1], José Lages[1], and Dima L. Shepelyansky[2]

[1] Institut UTINAM, Observatoire des Sciences de l'Univers THETA, CNRS, Université de Bourgogne-Franche-Comté, Besançon 25030, France
http://perso.utinam.cnrs.fr/~lages/

[2] Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, 31062 Toulouse, France
http://www.quantware.ups-tlse.fr/dima/

1 Introduction

The efficiency of academic education is nowadays a matter of political, economical and societal importance. University rankings, reviewed e.g. in many details in [4], are among the most important tools to measure academic efficiency. The available ranking approaches are based on human selection rules which can not be exhaustive or can favor certain cultural choices and preferences. Thus it would be highly desirable to construct an independent mathematical statistical method which ranks universities independently of any human rules. In order to fill this gap, we have recently proposed the Wikipedia Ranking of World Universities (WRWU) [5,7] based on the statistical analysis of networks of Wikipedia articles. Wikipedia now supersedes Encyclopedia Britannica in size and even in accuracy of articles devoted to many scientific domains [3]. Presently, Wikipedia contains more than 280 language editions representing different and complementary cultural views on human knowledge. This huge amount of encyclopedic data encodes also hidden information about how different cultures and societies are entangled. For all these reasons probing Wikipedia is relevant to build rankings for various aspects of human activities, one of these being higher education.

2 Methods

The mathematical grounds of this approach are based on Markov chain theory and, in particular, on the Google matrix analysis initially introduced in 1998 by Google's co-founders, Brin and Page [1], for hypertext analysis of the World Wide Web. Let's consider the network of the N articles of a given Wikipedia edition. The network adjacency matrix element A_{ij} is equal to 1 if article j points towards article i and equal to zero otherwise. The Google matrix element $G_{ij} = S_{ij} + (1-\alpha)/N$ gives the transition probability that a random reader jump from article j to article i. The stochastic matrix element is $S_{ij} = A_{ij}/\sum_{i=1}^{N} A_{ij}$ if article j is not a dangling node, otherwise $S_{ij} = 1/N$. The dumping factor $\alpha = 0.85$ allows the random reader to escape from dangling subnetworks. The right eigenvector $\mathbf{P}$ corresponding to the $\lambda = 1$ Google matrix eigenvalue is the PageRank vector. The vector element P_i is proportional to the number of times

the random reader reads article i. The CheiRank vector $\mathbf{P}^*$ is the $\lambda = 1$ right eigenvector of the Google matrix constructed with the inverted network using A_{ji} instead of A_{ij}. PageRank measures the relative influence of nodes. Recursively, the more a node is pointed by influent nodes, the more it is influent. CheiRank measures the relative communicative ability of nodes. Recursively, the more a node points toward important communicative nodes, the more it is communicative. The ranking of the most influent (communicative) universities is obtained by extraction from PageRank (CheiRank) the articles devoted to universities.

3 Results

Table 1 (leftmost column) gives for the 2017 English edition of Wikipedia, the top10 of the most influent universities using PageRank algorithm. As a comparison, the top10 of the 2017 Academic Ranking of World Universities (ARWU) is shown in Table 1 (rightmost column). The two top10s (top100s; not shown) have 9 (61) universities in common confirming the fact that Wikipedia ranking is indeed able to measure academic excellence. Comparing each of these two rankings with the corresponding ones in 2013, we see that Wikipedia ranking is more robust since 9 universities are in common and keep their positions [7] and for ARWU 10 universities are in common but only 4 keep their positions. Fig. 1 (left panel) shows the geographical distribution of top100 universities from the 2017 English Wikipedia network PageRank analysis. As in ARWU, Anglo-Saxon universities dominate in number. Fig. 1 (right panel) gives the distribution of the density of 2017 English Wikipedia articles in the plane of PageRank index K *vs.* CheiRank index K^*. We clearly see that the most influent universities (low PageRank index K) are also among the top100 of the most communicative universities (CheiRank K^*); PageRanking and CheiRanking share 41% universities in common. We observe that top influent universities have a PageRank as low as $\sim 10^2$ indicating the very importance of the corresponding articles in the 2017 English Wikipedia network ($\sim 5 \times 10^6$ articles). Also, most of the ARWU top100 universities which are not present in Wikipedia top100 PageRank are also not present in the top100 CheiRank, indicating their lack of communicative ability via Wikipedia.

2017 English Wikipedia PageRanking	Rank	2017 ARWU
Harvard University	1	Harvard University
University of Oxford	2	Stanford University
University of Cambridge	3	University of Cambridge
Columbia University	4	MIT
Yale University	5	University of California, Berkeley
Stanford University	6	Princeton University
MIT	7	University of Oxford
University of California, Berkeley	8	Columbia University
Princeton University	9	California Institute of Technology
University of Chicago	10	University of Chicago

Table 1. Comparison between the ranking of world universities obtained from the PageRank of the 2017 English Wikipedia network (leftmost column) and the 2017 academic ranking of world universities provided by the Shanghai Jiao Tong University (rightmost column).

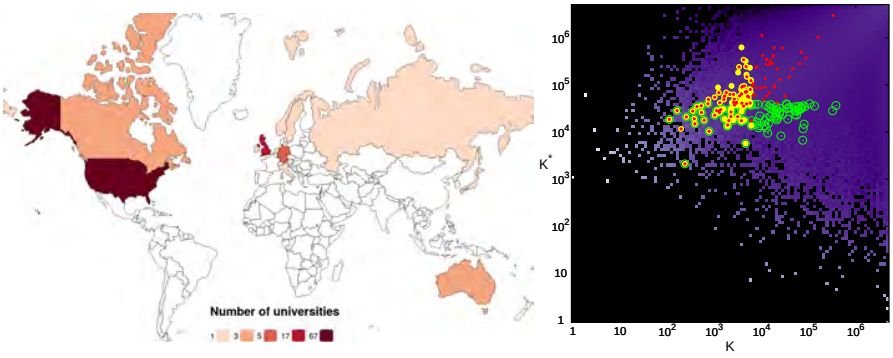

Fig. 1. Left panel: geographical distribution of the top100 universities obtained from the PageRanking of the 2017 English Wikipedia network. Right panel: density distribution $dN/dKdK^*$ of 2017 English Wikipedia articles in the plane of PageRank and CheiRank indexes (K, K^*) shown by color with dark violet for minimum and white for maximum (black for zero). Yellow disks (green circles) indicate the top100 universities using PageRank (CheiRank) algorithm. Red points indicate 2017 ARWU top100.

Fig. 1 shows results for 2017 English edition of Wikipedia. At the Complex Networks 2017 conference, we will present an exhaustive study of 24 different language editions of Wikipedia ($\sim 17 \times 10^6$ articles) representing about 60% of the total articles in Wikipedia and corresponding to about 60% of the total world population. Consequently, we will construct a network of culture comparing the different cultural point of views encoded in these language editions. Aggregating rankings for the 24 Wikipedia editions, we will provide the 2017 global Wikipedia Ranking of World Universities. Also, using the recently developed reduced Google matrix method [2,6], we will present hidden links existing between the most influent universities.

References

1. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. 30(1-7), 107–117 (Apr 1998), http://dx.doi.org/10.1016/S0169-7552(98)00110-X
2. Frahm, K.M., Jaffrès-Runser, K., Shepelyansky, D.L.: Wikipedia mining of hidden links between politicalleaders. The European Physical Journal B 89(12), 269 (Dec 2016), https://doi.org/10.1140/epjb/e2016-70526-3
3. Giles, J.: Internet encyclopaedias go head to head. Nature 438, 900–901 (Dec 2005)
4. Hazelkorn, E.: Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence. Palgrave Macmillan (2015)
5. Lages, J., Patt, A., Shepelyansky, D.L.: Wikipedia ranking of world universities. The European Physical Journal B 89(3), 69 (Mar 2016), https://doi.org/10.1140/epjb/e2016-60922-0
6. Lages, J., Shepelyansky, D., Zinovyev, A.: Inferring hidden causal relations between pathway members using reduced google matrix of directed biological networks. bioRxiv (2017), https://www.biorxiv.org/content/early/2017/02/06/096362
7. Lages, J., Shepelyansky, D.: WRWU website. http://perso.utinam.cnrs.fr/~lages/datasets/WRWU/ ([Online; accessed 06-October-2017])