



Researchers at the University of Granada map the Sciences according to Wikipedia

20/03/2020

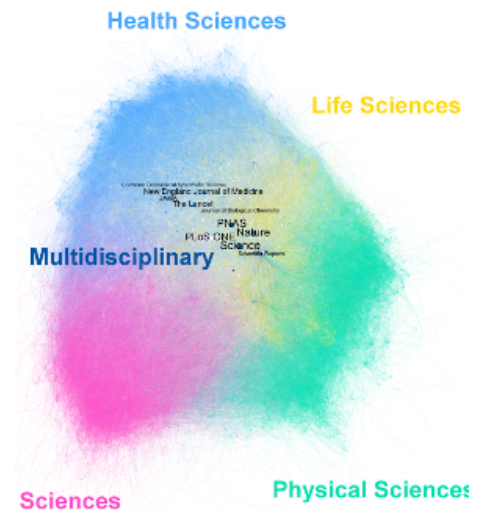
Research news

A study carried out by scientists from the University of Granada (UGR) and published in the journal PLOS ONE has successfully produced a representation of the Sciences according to Wikipedia, thereby showing how this social platform has constructed the knowledge-base via its open and collaborative dissemination model. The study was based on the bibliographical references included in the Wikipedia entries, as its editors rely on these to ensure the rigour of all contributions to the platform.

The researchers had previously carried out a similar study focusing only on the Social Sciences. This revealed a predominant presence of articles dealing with History, as well as significant differences between Wikipedia and other academic environments. However, the Social Sciences represent just 5% of the articles cited in Wikipedia.

In this latest study, the researchers analysed Wikipedia from a more global perspective. According to their findings, the citation approach studied here differs from that of Scopus and other social media, with the popular digital encyclopaedia offering a different view of the Sciences. Articles dealing with Medicine and Biochemistry are the main focus, while those related to the Social Sciences and Humanities have a lesser presence.

To carry out the study, the researchers retrieved the bibliographical references of the English version of Wikipedia, which is the most extensive, covering a period from the very beginning of the platform in 2004 up to early 2018. Having identified the references, they linked each scientific article to its source, using the Scopus database, and assigned each one to the most appropriate thematic category. The



main output of the study was a series of maps showing how articles and journals are cited from Wikipedia, based on an adaptation of co-citation theory to this particular context. In total, the sample comprised 847,512 references made by 193,802 Wikipedia entries to 598,746 articles from 14,149 journals.

Differences between Wikipedia and Scopus

On the one hand, the results show differences between Wikipedia and Scopus, both in the coverage of the articles and in the citations they use, albeit no causality could be established. Despite the open approach of Wikipedia, Open Access journals represent just 13% of the total number of journals cited. Among this total, high-impact journals are among the most commonly referenced, albeit with some differences compared to other social media.

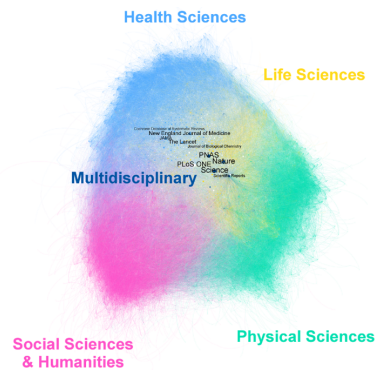
On the other hand, the researchers were able, by means of the mapping process, to illustrate the relationships that the platform's publishers establish between scientific articles and journals when these are cited together within the same entry. Thanks to this feature, it is possible to see how journals are grouped together under four main areas (Physics, Health Sciences, Social Sciences, and Life Sciences) and also their interdisciplinarity. Those journals of a multidisciplinary nature occupy a central position—in particular Science, Nature, PNAS, PLOS ONE, and The Lancet. Meanwhile, the disciplines are centred around Medicine and Biochemistry—a phenomenon that can also be observed at the level of specialties.

This study was conducted by researchers Wenceslao Arroyo Machado, Daniel Torres Salinas, Enrique Herrera Viedma, and Esteban Romero Frías. It was made possible thanks to research funding from 'Fundación BBVA a Equipos de Investigación Científica 2016'.

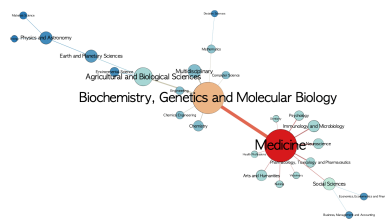
Bibliography:

Arroyo-Machado W., Torres-Salinas D., Herrera-Viedma E. & Romero-Frías, E. (2020) 'Science through Wikipedia: A novel representation of open knowledge through co-citation networks'. PLOS ONE. <https://doi.org/10.1371/journal.pone.0228713>

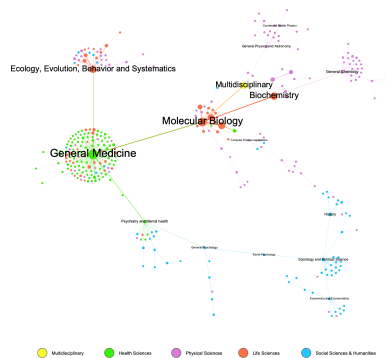
Image captions:



Journal co-citation network based on Wikipedia article references. Main component of the network. The colour corresponds to the thematic area, but those with more than one area are shown in white; the thickness of the edges corresponds to the degree of co-citation between the two. The titles of the 10 journals with the highest levels of intermediation were included.



Co-citation network of the 27 major fields after applying the Pathfinder algorithm. The nodes represent each main field; the size of the nodes corresponds to the total number of citations received; the colour to eigenvector centrality; and the thickness of the edges to the degree of co-citation.



Co-citation network of the 330 subfields after applying the Pathfinder algorithm. The nodes represent each of the subfields, their size indicating the total number of citations received, the colour according to the area or thematic areas, and the thickness of the edges according to the degree of

co-citation. The names of the 15 subfields with the highest levels of intermediation were included.

Media enquiries:

Daniel Torres-Salinas

Department of Information and Communication Sciences, University of Granada

Tel.: +34 958 244128

Email: torressalinas@go.ugr.es