

# Mapas de la ciencia a partir de Wikipedia

# Mapas de la ciencia a partir de Wikipedia: una visión del consumo y la transferencia de conocimiento en las plataformas sociales

- Daniel Torres-Salinas (Universidad de Granada)
- Wenceslao Arroyo-Machado (Universidad de Granada)

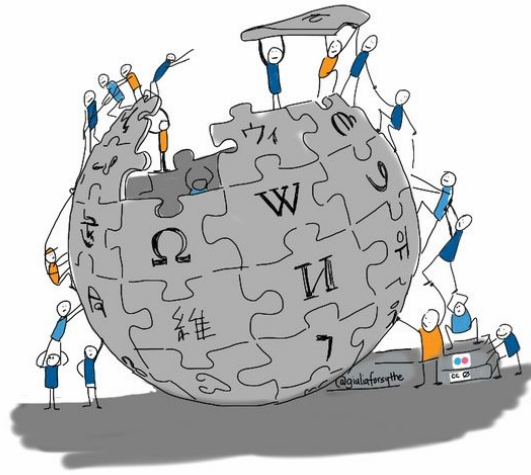
Presentado en:

EVALUACIÓN DE LA TRANSFERENCIA-EVALUACIÓN DEL IMPACTO

MÉTRICAS Y PROCEDIMIENTOS PARA LA CIENCIA Y LA INNOVACIÓN

Organizada por el **Instituto de Estudios Sociales Avanzados-CSIC**.

12–13 Diciembre 2019 - Córdoba, España



# Intro & methodology

# Bibliographic references

- Torres-Salinas, D., Romero-Frías, E., & Arroyo-Machado, W. (2019). Mapping the backbone of the Humanities through the eyes of Wikipedia. *Journal of Informetrics*, 13(3), 793-803.
- Arroyo-Machado, W., Torres-Salinas, D., Herrera-Viedma, E., Romero-Frías, E., & (2019). Science through Wikipedia: a novel representation of open knowledge through co-citation networks. *PLoS One* (under review)

# Objectives

- To apply co-citation analysis to all the articles referenced in Wikipedia in order to validate their usefulness in analyzing open knowledge platforms;
- To offer a general portrait of the use of scientific literature published in journals through the analysis of references and their obsolescence. Thus, we hope to be able to describe the consumption of scientific information by the Wikipedia community and detect possible differences between fields; and lastly, as the nuclear objective of our paper
- To discover the different visions offered by Wikipedia by using co-citation networks at different levels of aggregation: 1) journal co-citation maps 2) main field co-citation maps 3) field co-citation maps. Through these representations we intended to obtain a holistic view of how scientific articles in Wikipedia are used and consumed.

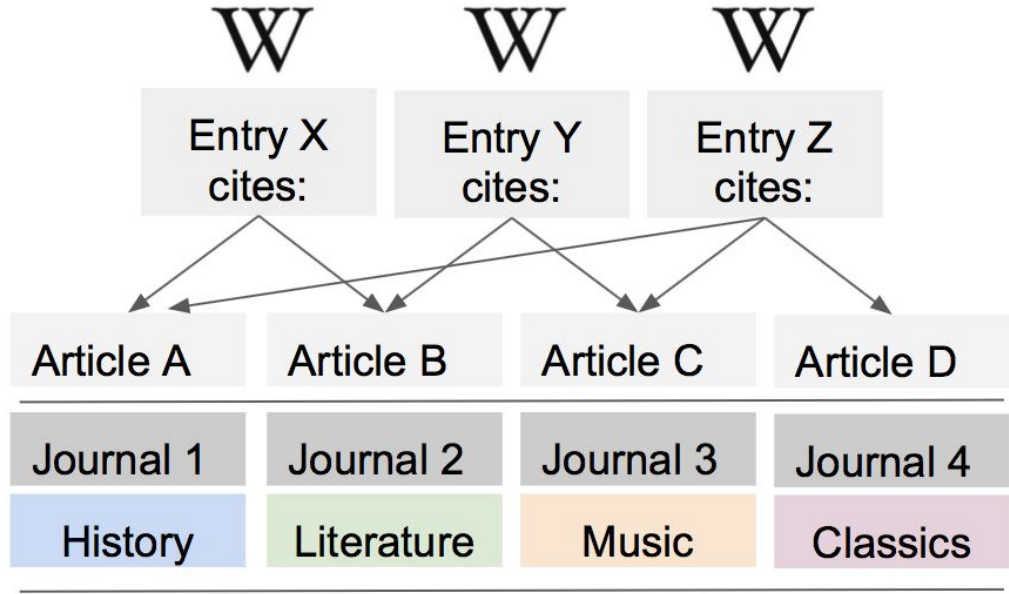


CiteScore™

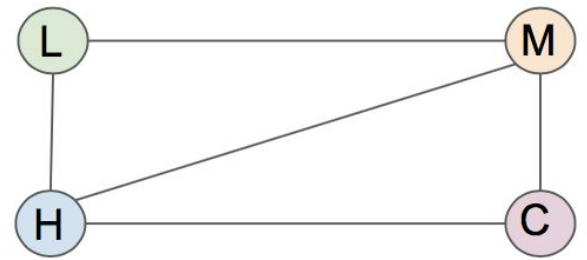
Scopus®

1	Download all resources cited in Wikipedia indexed in Altmetric.com	1 433 457 references from 321 470 Wikipedia entries to 960 017 resources
2	Clean downloaded data	1 211 904 references from 288 290 Wikipedia entries to 857 087 resources
3	Download ISSN codes of resources by using the Altmetric.com API	36 090 ISSN codes for 693 805 resources
4	Join resources with Elsevier CiteScore Metrics journals by ISSN codes	14 149 journals joined
5	Limitation to articles from joined journals and assignment them of Scopus subject area categories	847 512 references from 193 802 Wikipedia entries to 598 746 articles

# Analysis of co-citation networks



**Map of categories based on co-citation**

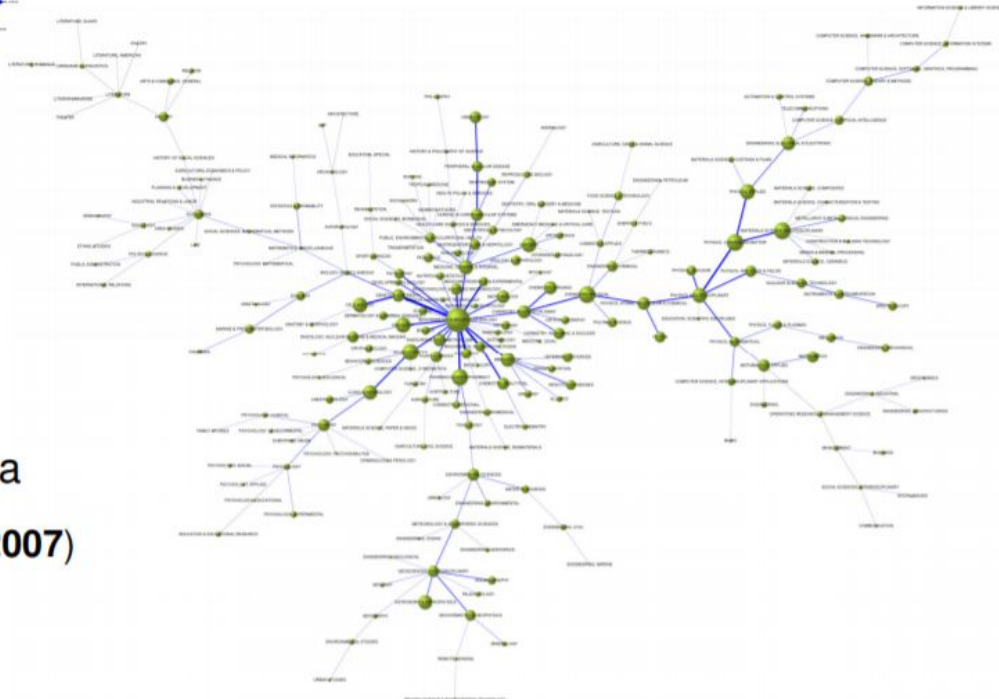
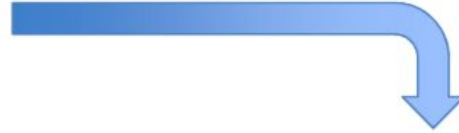
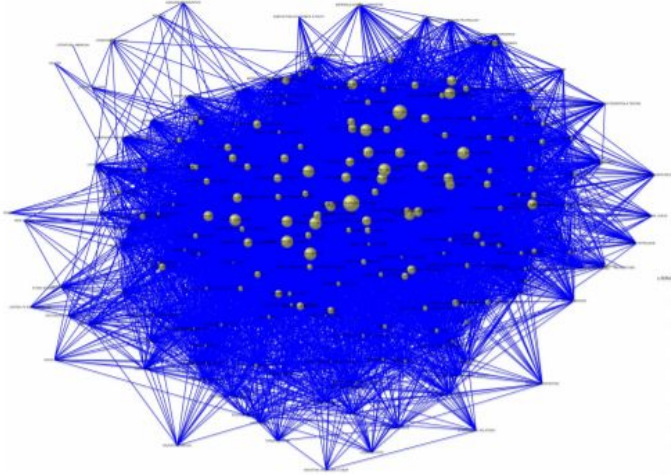


# Analysis of co-citation networks

In the co-citation networks of journals, fields and main fields, the Pathfinder algorithm has been applied as a pruning method following a common configuration ( $r=\infty$ ,  $q=n-1$ ) that reduces the networks to a minimum covering tree. This algorithm—successfully applied in the field of Library and Information Sciences—keeps only the strongest co-citation links between all pairs of nodes and offers a diaphanous view of large networks. Since it is applied to values in relation to distances, the inverse value of the co-cites has been used in our analysis. Local measures of proximity, betweenness and eigenvector centrality have also been calculated.



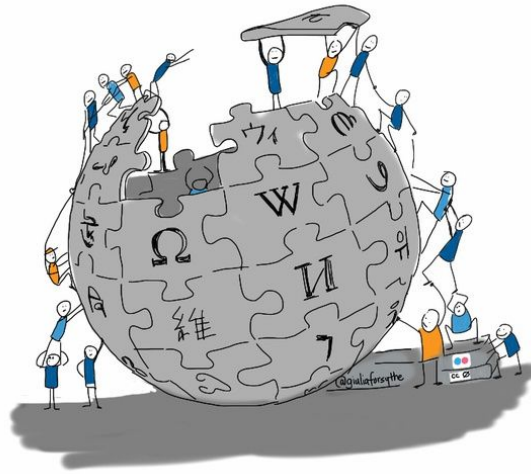
# Analysis of co-citation networks



Ej: redes de cocitación científica

(Vargas-Quesada y Moya-Anegón, 2007)

[Source Oscar Cordón](#)



**Results: main data**

**A****Health Sciences****Life Sciences****Multidisciplinary****Social Sciences  
& Humanities****Physical Sciences**

Cochrane Database of Systematic Reviews  
 New England Journal of Medicine  
 JAMA  
 The Lancet  
 Journal of Biological Chemistry  
 PNAS  
 Nature  
 PLoS ONE  
 Science  
 Scientific Reports

**B****Health Sciences****Life Sciences****Multidisciplinary****Social Sciences  
& Humanities****Physical Sciences**

Cochrane Database of Systematic Reviews  
 New England Journal of Medicine  
 JAMA  
 The Lancet  
 Journal of Biological Chemistry  
 PNAS  
 Nature  
 PLoS ONE  
 Science  
 World Health Library

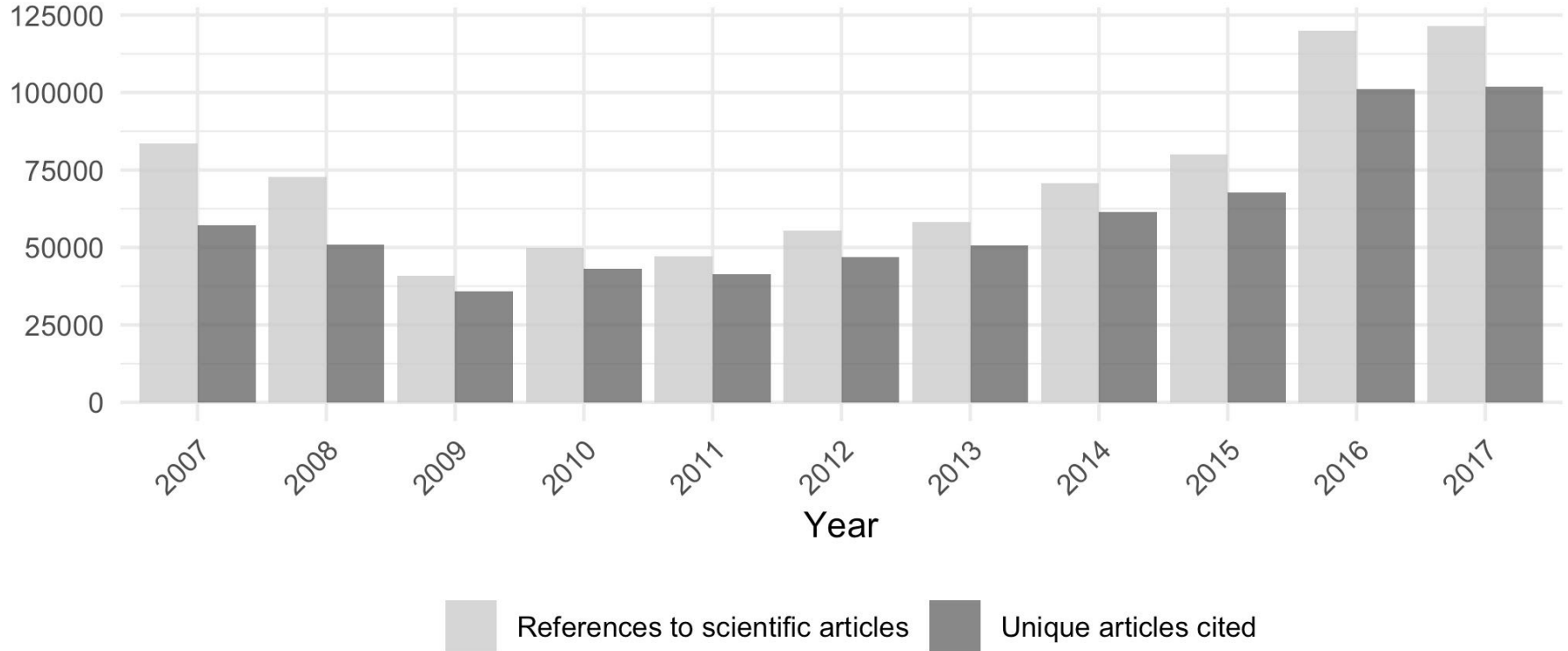
# General view

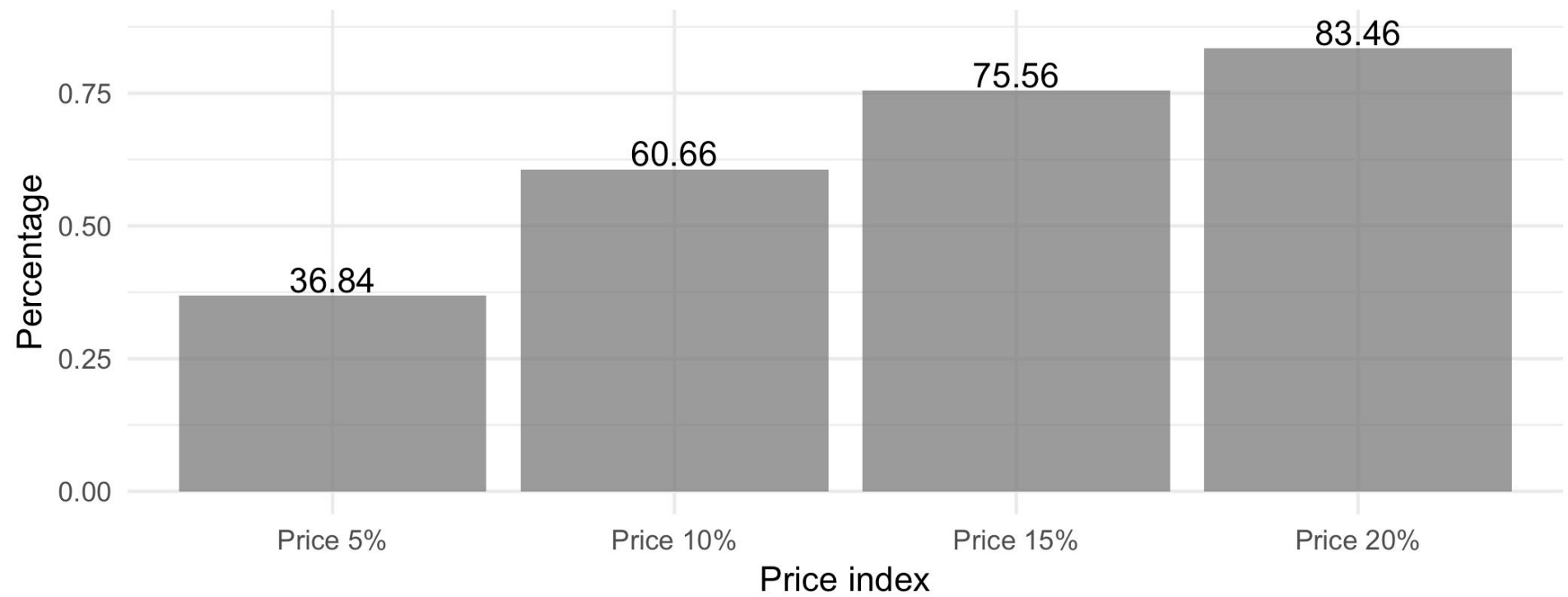


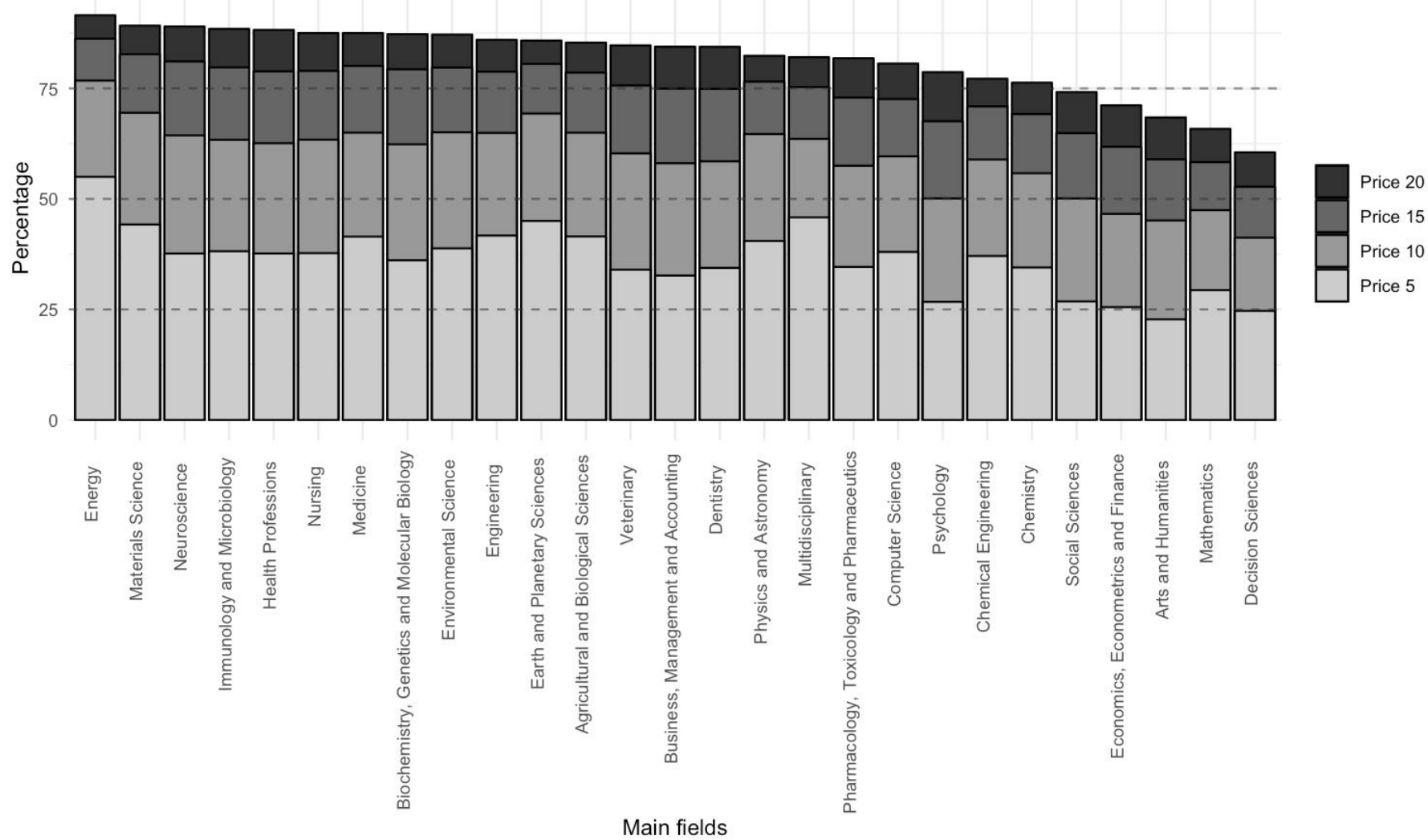
- 847512 references to scientific articles distributed across 193 802 Wikipedia entries.
- 598 746 scientific articles published in 14 149 journals are cited.
- Each Wikipedia entry includes 4.373 ( $\pm 8.351$ ) references to scientific articles
- Articles receive a mean of 1.415 ( $\pm 10.15$ ) citations.
- 81.71% of the total number of scientific articles (489 235) receive only one citation and this corresponds to 57.73% of all references in the study sample.

	Descriptive statistics of references made by Wikipedia entries to articles by areas					Descriptive statistics of citations that articles receive from Wikipedia entries by areas				
Area	Wikipedia entries	References (% total)	Mean ( $\pm$ standard deviation)	Median	Range	Articles	References (% total)	Mean ( $\pm$ standard deviation)	Median	Range
Health Sciences	73 402	282 750 (33.36%)	3.85 ( $\pm$ 6.89)	2	221	212 292	282 750 (33.36%)	1.33 ( $\pm$ 9.35)	1	3591
Life Sciences	102 933	414 400 (48.90%)	4.03 ( $\pm$ 6.61)	2	265	287 674	414 400 (48.90%)	1.44 ( $\pm$ 9.70)	1	3591
Multidisciplinary	33 943	72 346 (8.54%)	2.13 ( $\pm$ 2.72)	1	99	38 422	72 346 (8.54%)	1.88 ( $\pm$ 26.40)	1	4997
Physical Sciences	73 453	195 678 (23.09%)	2.66 ( $\pm$ 4.65)	1	301	137 492	195 678 (23.09%)	1.42 ( $\pm$ 7.54)	1	1483
Social Sciences & Humanities	49 688	115 947 (13.68%)	2.33 ( $\pm$ 3.96)	1	127	91 617	115 947 (13.68%)	1.27 ( $\pm$ 1.01)	1	107

# Annual values of total references made by Wikipedia and single articles cited

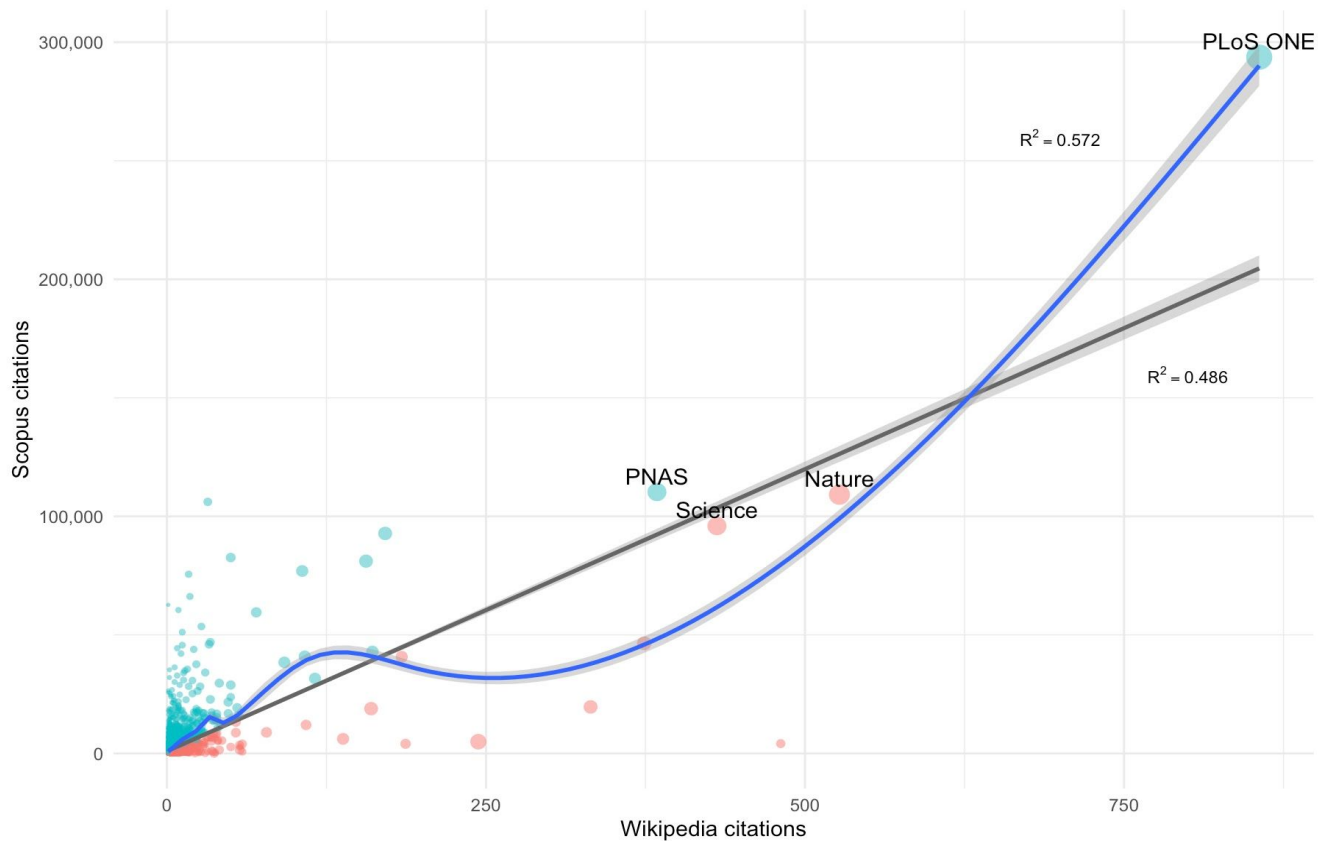









# Scatter plot of journals by citation collected in Scopus and Wikipedia in 2016 (citescore formula)



	<b>Journal</b>	<b>Articles</b>	<b>Citations</b>	<b>References/ articles</b>	<b>Open Access</b>	<b>Area</b>
1	Nature	13 312	26 434	1.986	No	Multidisciplinary
2	PNAS	12 234	24 104	1.970	No	Multidisciplinary
3	Journal of Biological Chemistry	16 611	21 921	1.320	No	Life Sciences
4	Science	10 191	17 853	1.752	No	Multidisciplinary
5	PLoS ONE	5793	9603	1.658	Yes	Health Sciences/Life Sciences
6	Genome Research	613	8839	1.419	No	Health Sciences/Life Sciences
7	Astrophysical Journal	3722	8552	2.298	No	Physical Sciences
8	Cell	3673	7223	1.967	No	Life Sciences
9	Astronomy and Astrophysics	1584	5546	3.501	No	Physical Sciences
10	Journal   American Chemical Society 	4549	5529	1.215	No	Life Sciences/Physical Sciences

# Wikipedia singularities

Gerhard, D. S., Wagner, L., Feingold, E. A., Shenmen, C. M., Grouse, L. H., Schuler, G., ... & Guyer, M. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome research*, 14(10B), 2121-2127.

## 3408 Wikipedia Citations

SUMMARY	News	Twitter	Patents	Wikipedia
---------	------	---------	---------	-----------

 So far, Altmetric has seen **3408** citations on **3408** Wikipedia pages.



### LOC101928193

Cited by user **Boghog** on **08 May 2019**

LOC101928193 is a protein found in Homo sapiens that is en



### SLITRK1 (gene)

Cited by user **Artoria2e5** on **08 Apr 2019**

SLITRK1 ("SLIT and NTRK-like family, member 1") is a human  $\zeta$  the SLITRK gene family, which is...

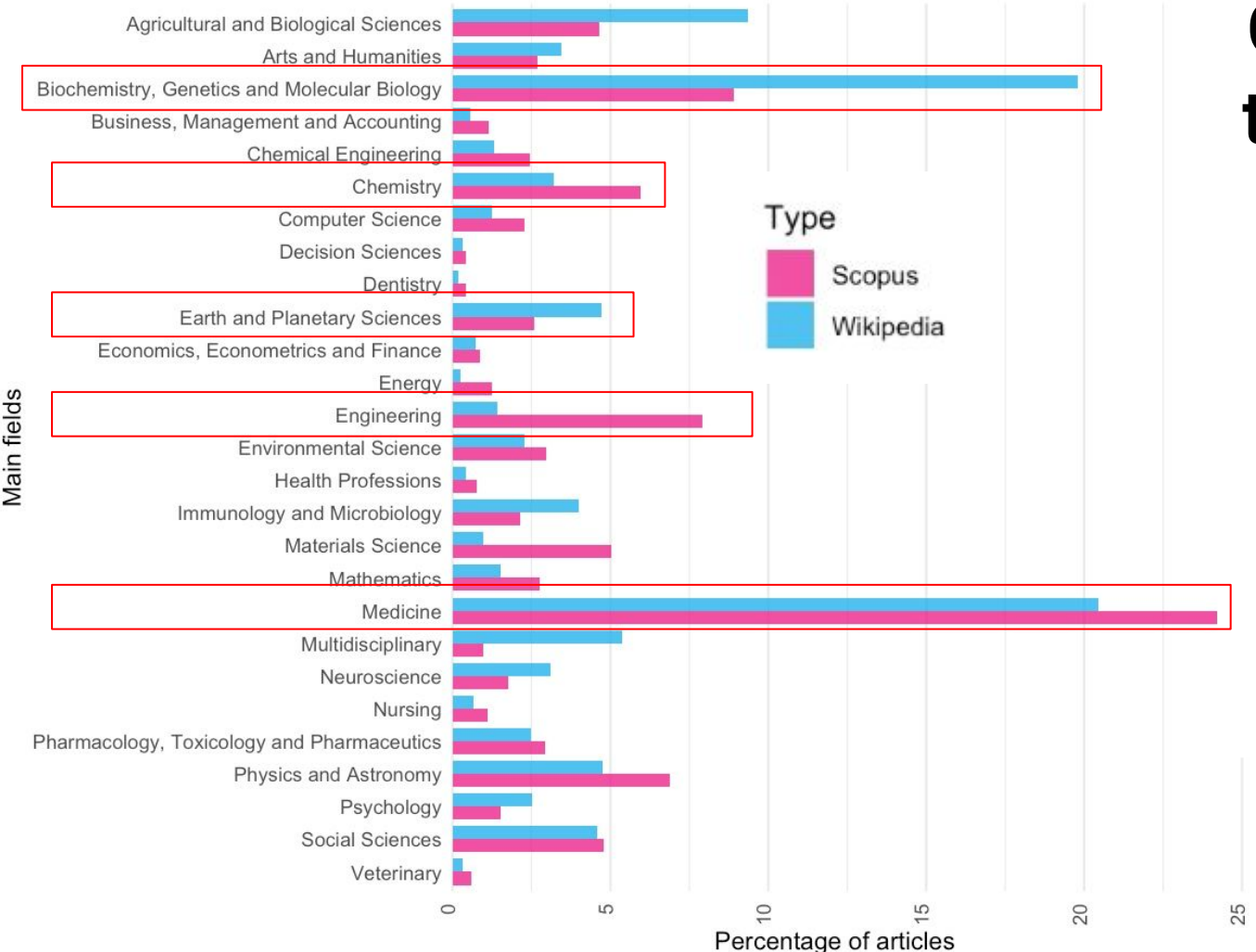


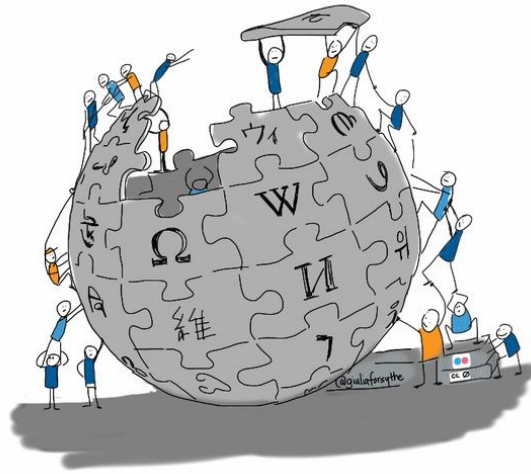
### LRRC8C

Cited by user **Angiebarrett** on **26 Jul 2017**

Leucine-rich repeat-containing protein 8C is a protein that in

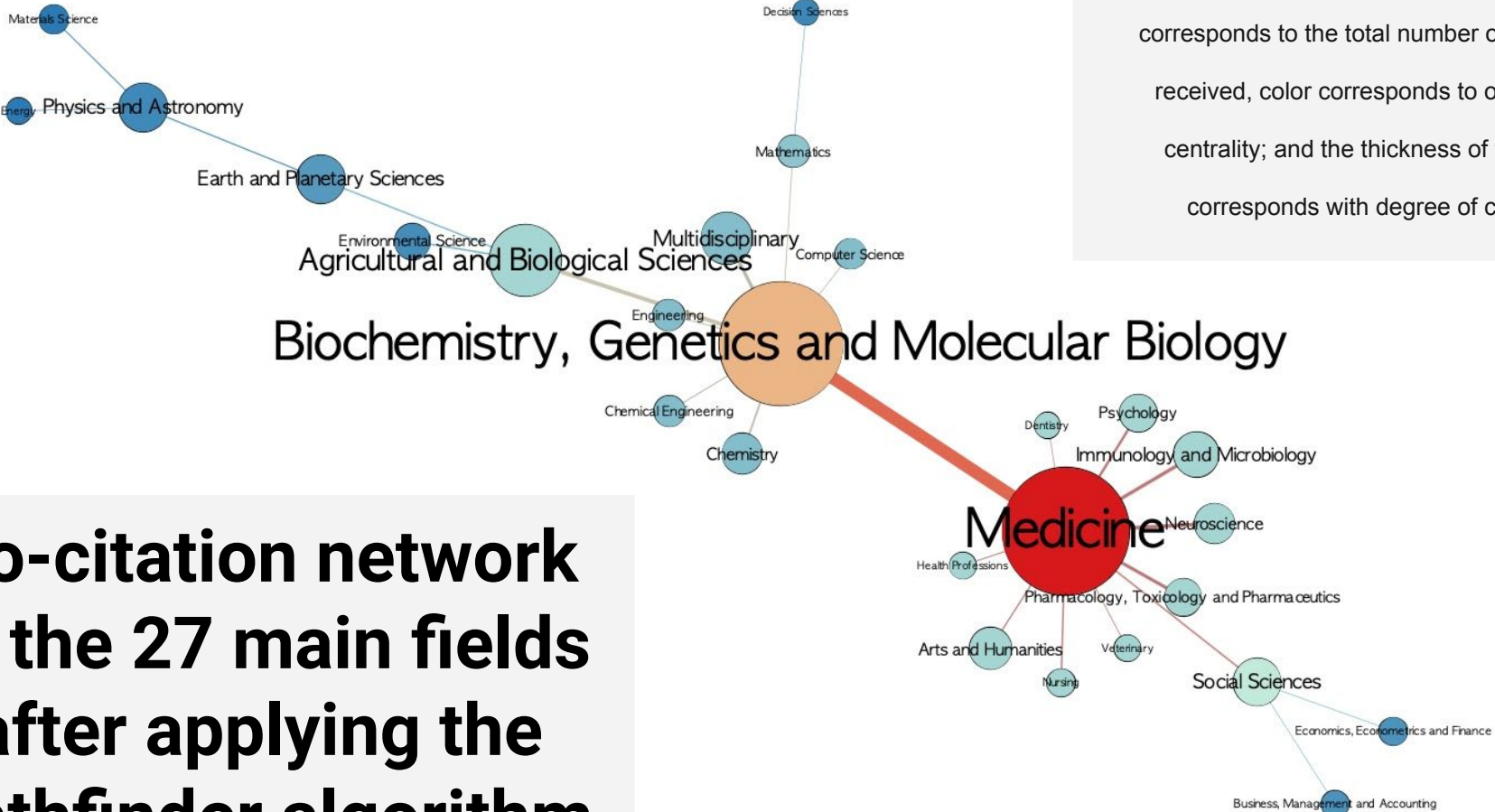
# Comparison of the percentage of articles by main field in Scopus and Wikipedia





**Results: maps**

The nodes represent each main field; node size corresponds to the total number of citations received, color corresponds to own vector centrality; and the thickness of the edges corresponds with degree of co-citation.

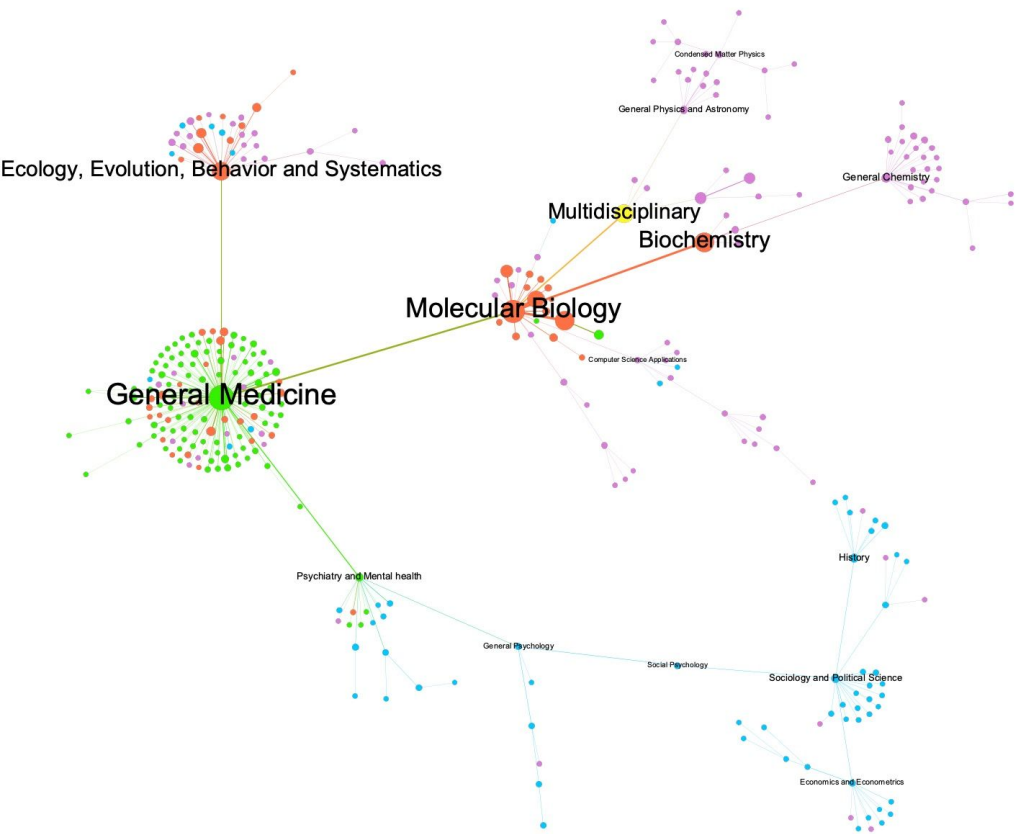


Biochemistry, Genetics and Molecular Biology

Medicine

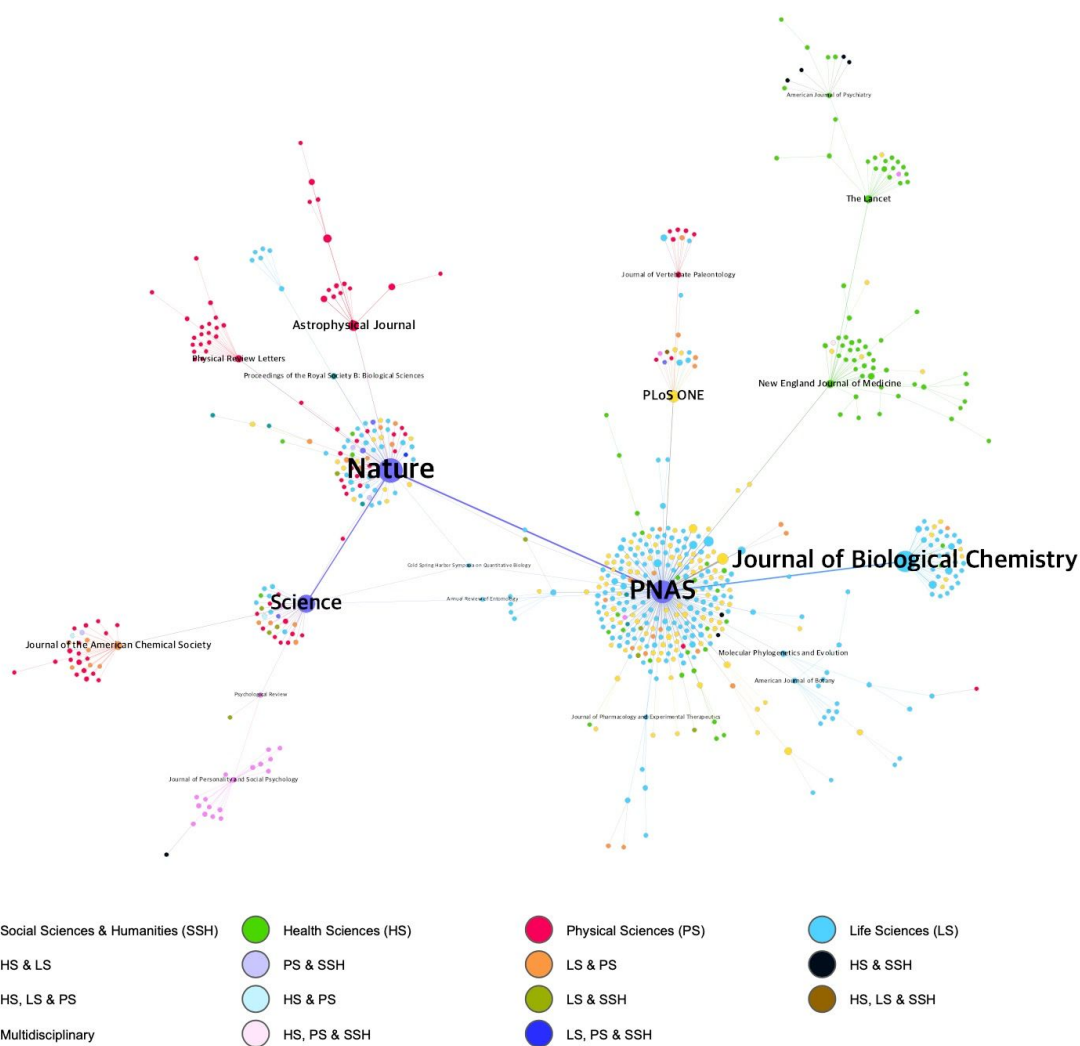
**Co-citation network  
of the 27 main fields  
after applying the  
Pathfinder algorithm**

# Co-citation network of the 330 fields after applying the Pathfinder algorithm



The nodes represent each field, indicating size, total number of citations received, color, thematic area or areas, and the thickness of the edges indicates the degree of co-citation. Field titles are given for the 15 fields with the highest levels of intermediation.

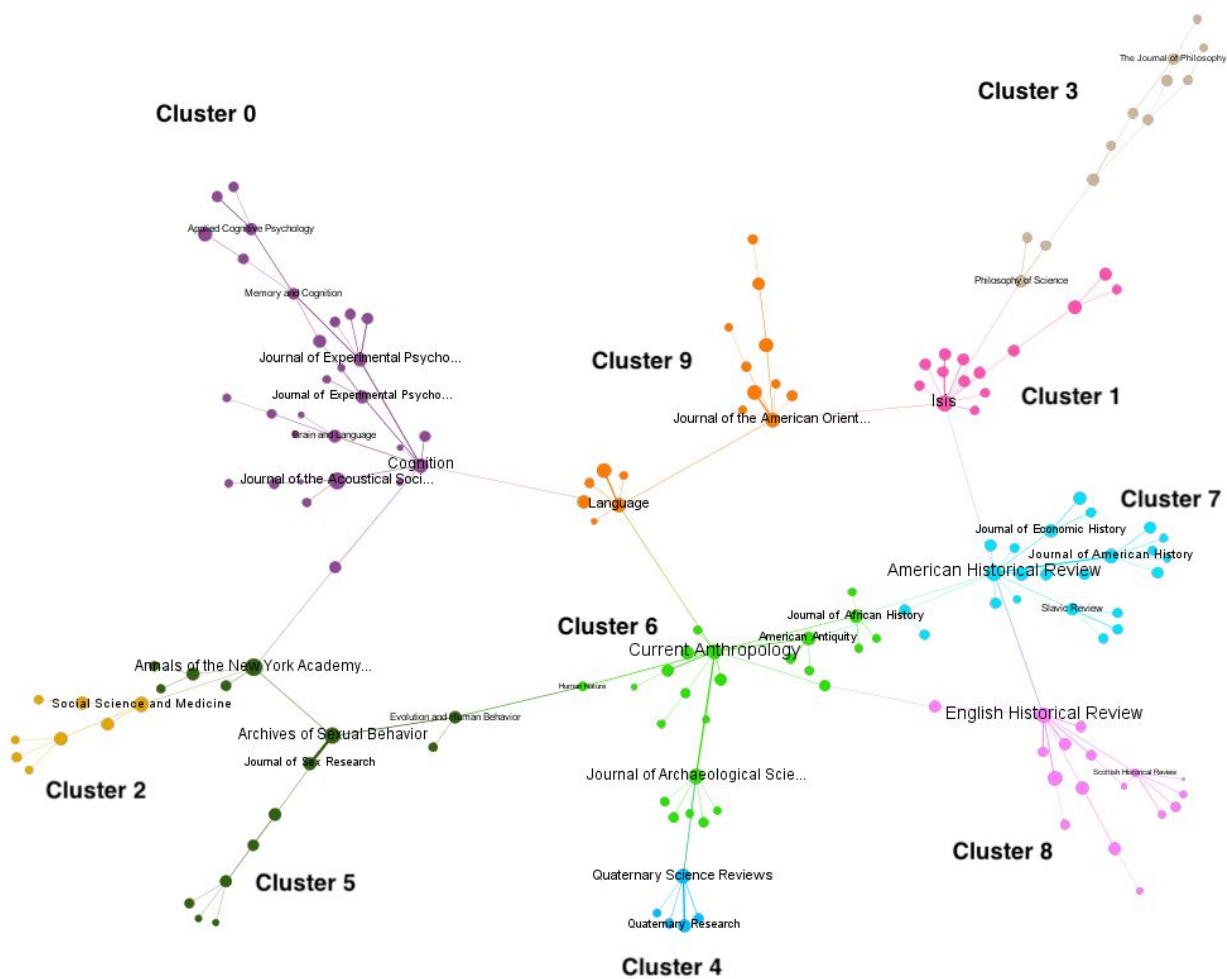
# Co-citation network of journals based on Wikipedia article references

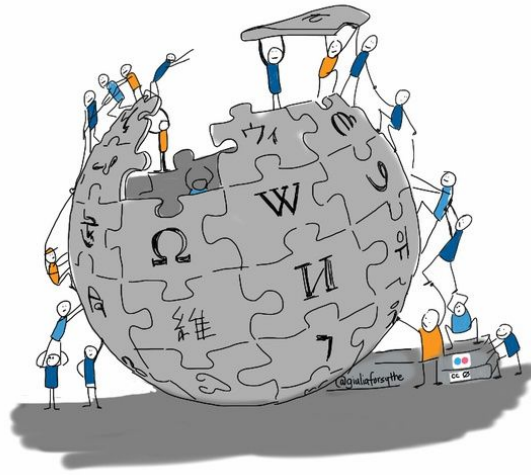


This network is produced by applying the Pathfinder algorithm—based on a minimum of 50 co-cites—and shows a total of 629 relationships. Each node represents one journal and node size corresponds to the total number of citations received; color corresponds to the area or combination of subject areas to which it belongs; and the thickness of the edges corresponds to the degree of co-citation between the two. The titles of the 20 journals with the highest intermediation value have been included.



# Map of co-citation in Wikipedia of scientific journals in the Humanities grouped according to similarity clusters





**Final comments**



## **BoB. Bibliometría o Barbarie. Open**

50 members, 7 online

Compartimos enlaces, artículos y todos de tipo de recursos, siempre de manera informal y divertida, sobre el mundo de las métricas y la evaluación científica

[VIEW IN TELEGRAM](#)