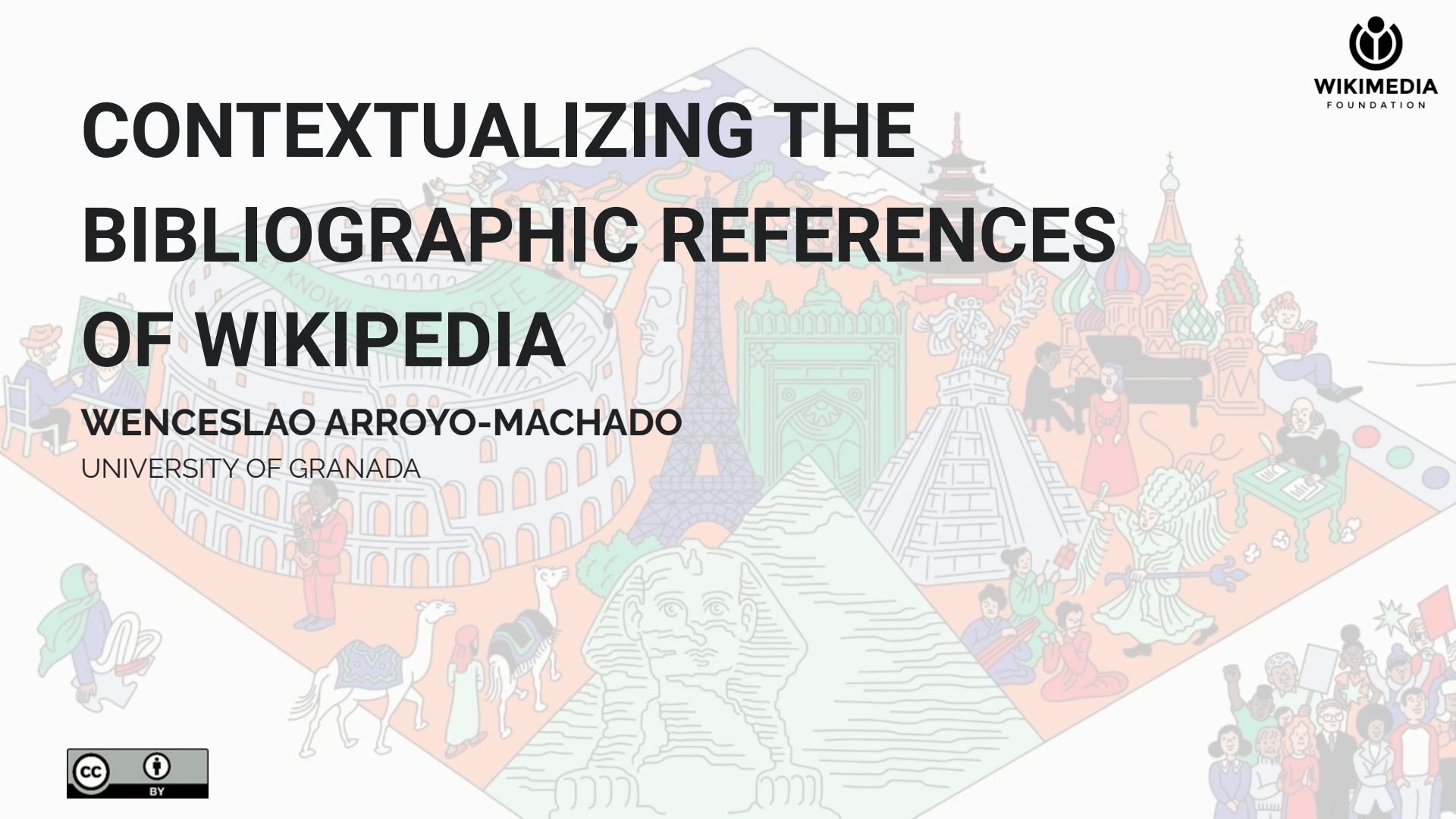
CONTEXTUALIZING THE BIBLIOGRAPHIC REFERENCES OF WIKIPEDIA

WENCESLAO ARROYO-MACHADO
UNIVERSITY OF GRANADA



PREAMBLE

Hi all,

I'm Wenceslao Arroyo-Machado, a postdoc at the University of Granada.

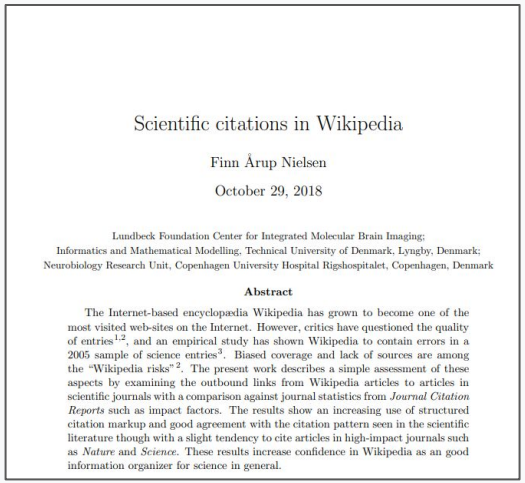
During my thesis, I combined data science with bibliometrics and social media to understand science-society relationships.

Wikipedia has been one of the main social media, and I continue to research it.

WHY WIKIPEDIA?

Since the early years of Wikipedia, the **bibliometric community** has shown significant interest in this platform for several reasons:

1. Open source
2. Social construction
3. Similarity of articles to papers



Scientific citations in Wikipedia

Finn Årup Nielsen

October 29, 2018

Lundbeck Foundation Center for Integrated Molecular Brain Imaging;
Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark;
Neurobiology Research Unit, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

Abstract

The Internet-based encyclopædia Wikipedia has grown to become one of the most visited web-sites on the Internet. However, critics have questioned the quality of entries[1,2], and an empirical study has shown Wikipedia to contain errors in a 2005 sample of science entries[3]. Biased coverage and lack of sources are among the "Wikipedia risks"[2]. The present work describes a simple assessment of these aspects by examining the outbound links from Wikipedia articles to articles in scientific journals with a comparison against journal statistics from *Journal Citation Reports* such as impact factors. The results show an increasing use of structured citation markup and good agreement with the citation pattern seen in the scientific literature though with a slight tendency to cite articles in high-impact journals such as *Nature* and *Science*. These results increase confidence in Wikipedia as an good information organizer for science in general.

One of the earliest bibliometric studies of Wikipedia by Nielsen (2007)

ARE THEY REALLY THAT SIMILAR?

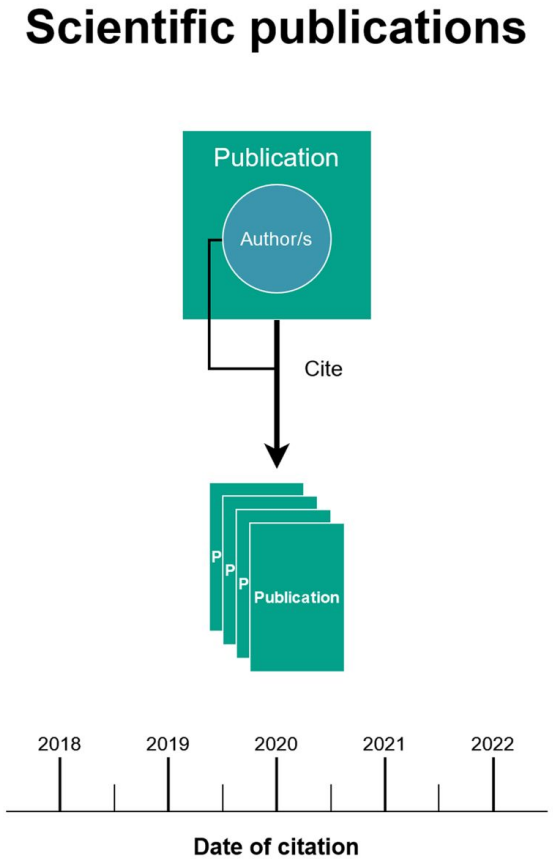
VERY SIMILAR?

Wikipedia element description		Wikipedia pages vs. Scientific publications	
		Wikipedia page	Scientific publication
State	*Document state condition*	Living	Static
ID	*Document identification number*	Page ID	DOI, ISBN, URI ...
Name	*Title of the document*	Title	Title
Type	*Document typologies*	Namespace (12 + 12 types)	Paper, proceeding, letter ...
Creation	*Date from which it is available*	First edition date	Publication date
Authorship	*Responsible for the work*	Wikipedians	Authors
Content	*Type of content*	Structured text	Structured text
Language	*Language of the resource*	Edition dependent	Document dependent
Discussion	*Comments on the contents*	Talk	Peer review
Description	*Work summary*	Short description	Abstract
Tags	*Terms describing the content*	Categories	Keywords
Media	*Audiovisual resources includable*	Images, audios, and videos	Images, audios, and videos
Internal links	*Links to the related resources*	Internal links	Citations
Format	*Standardized structure and content*	Manual of style*	Format guidelines
Bibliography	*References of cited resources*	References	References
Access	*Access model*	Open	Closed/Open
Audience	*Document target audience*	General	Specialized

Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022). Wikinformetrics: Construction and description of an open Wikipedia knowledge graph data set for informetric purposes. *Quantitative Science Studies*, 3(4), 931-952. https://doi.org/10.1162/qss_a_00226

Wikipedia combines elements of **scientific publications** with dynamic, open-editing features, making it a unique, ever-evolving knowledge resource

VERY SIMILAR?



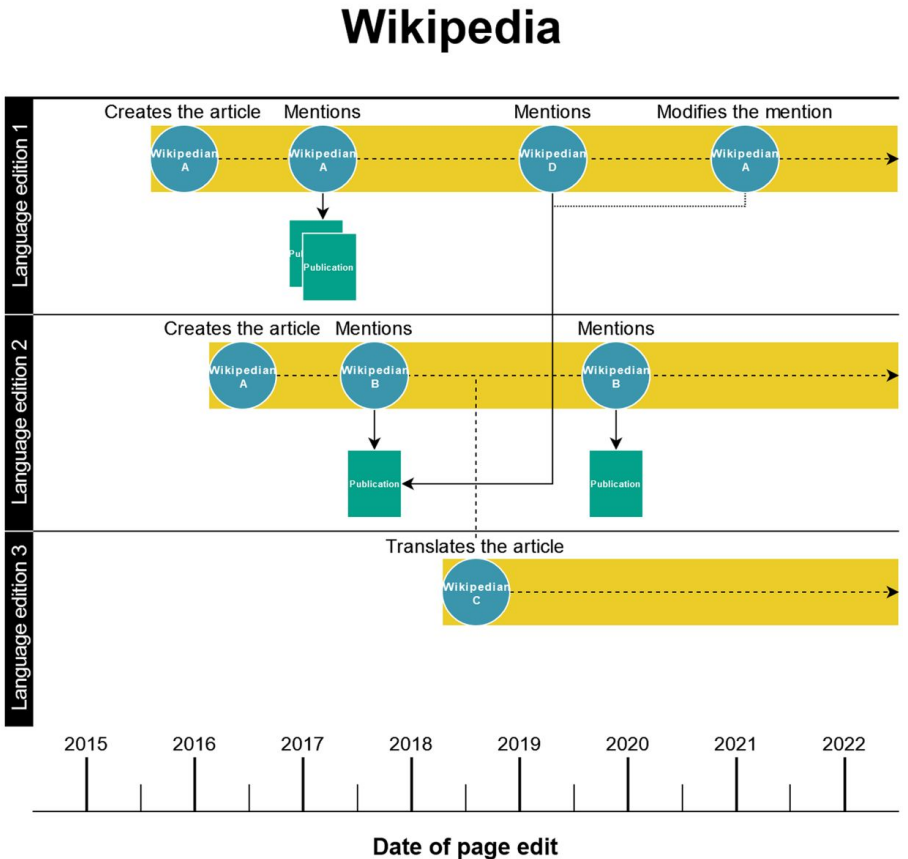
Scientific publications citations are a **static phenomenon** in which one document cites another

From this relationship, multiple analyses can be conducted

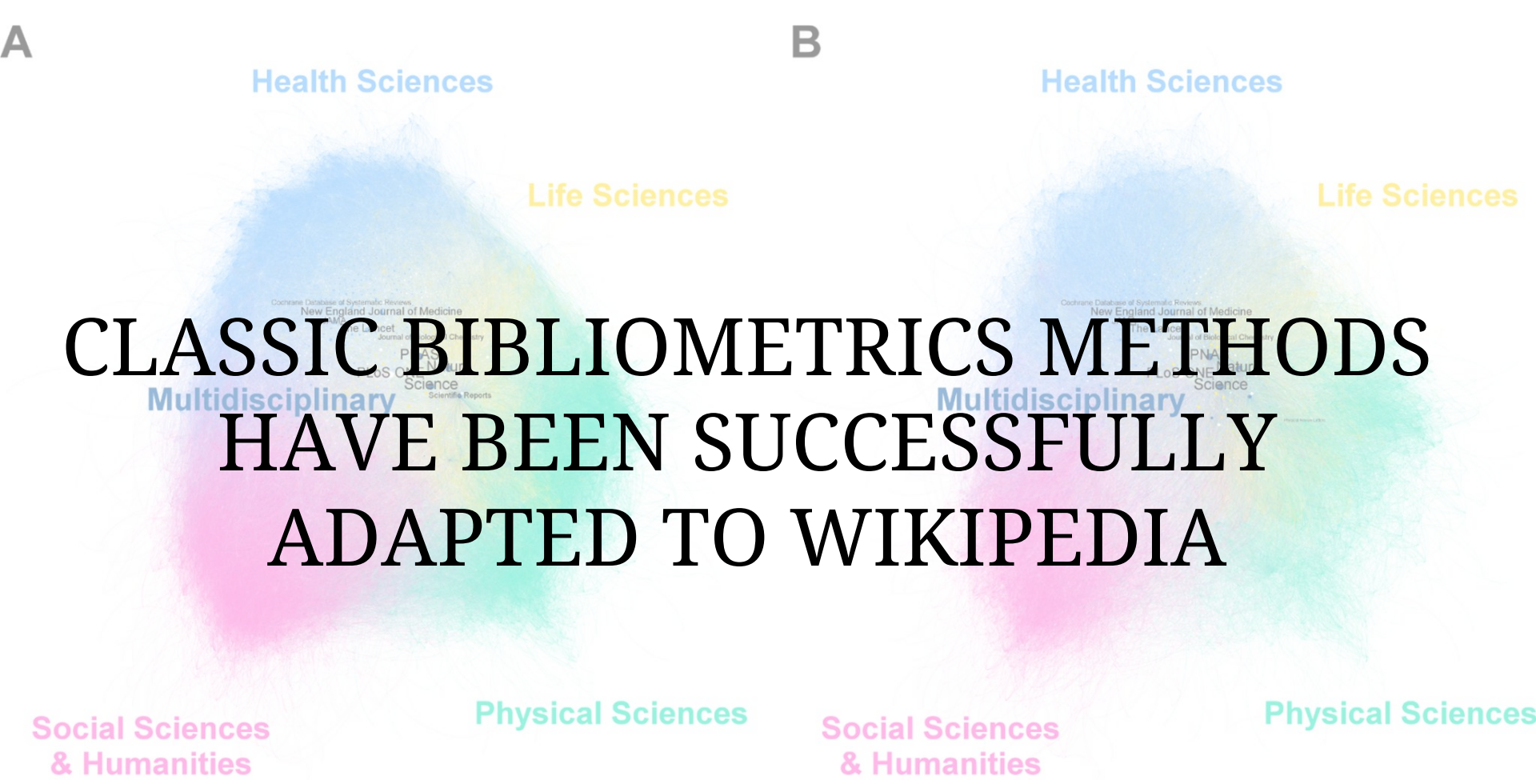
Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022). Wikinformetrics: Construction and description of an open Wikipedia knowledge graph data set for informetric purposes. *Quantitative Science Studies*, 3(4), 931-952. https://doi.org/10.1162/qss_a_00226

VERY SIMILAR?

The citations in Wikipedia articles represent a **living and complex** phenomenon



Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022). Wikinformetrics: Construction and description of an open Wikipedia knowledge graph data set for informetric purposes. *Quantitative Science Studies*, 3(4), 931-952. https://doi.org/10.1162/qss_a_00226



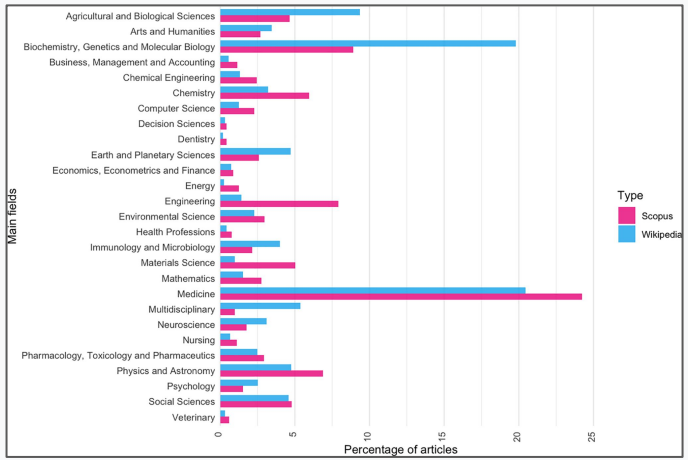
CLASSIC BIBLIOMETRICS METHODS HAVE BEEN SUCCESSFULLY ADAPTED TO WIKIPEDIA

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*, 15(2), e0228713. https://doi.org/10.1371/journal.pone.0228713

CLASSIC BIBLIOMETRICS

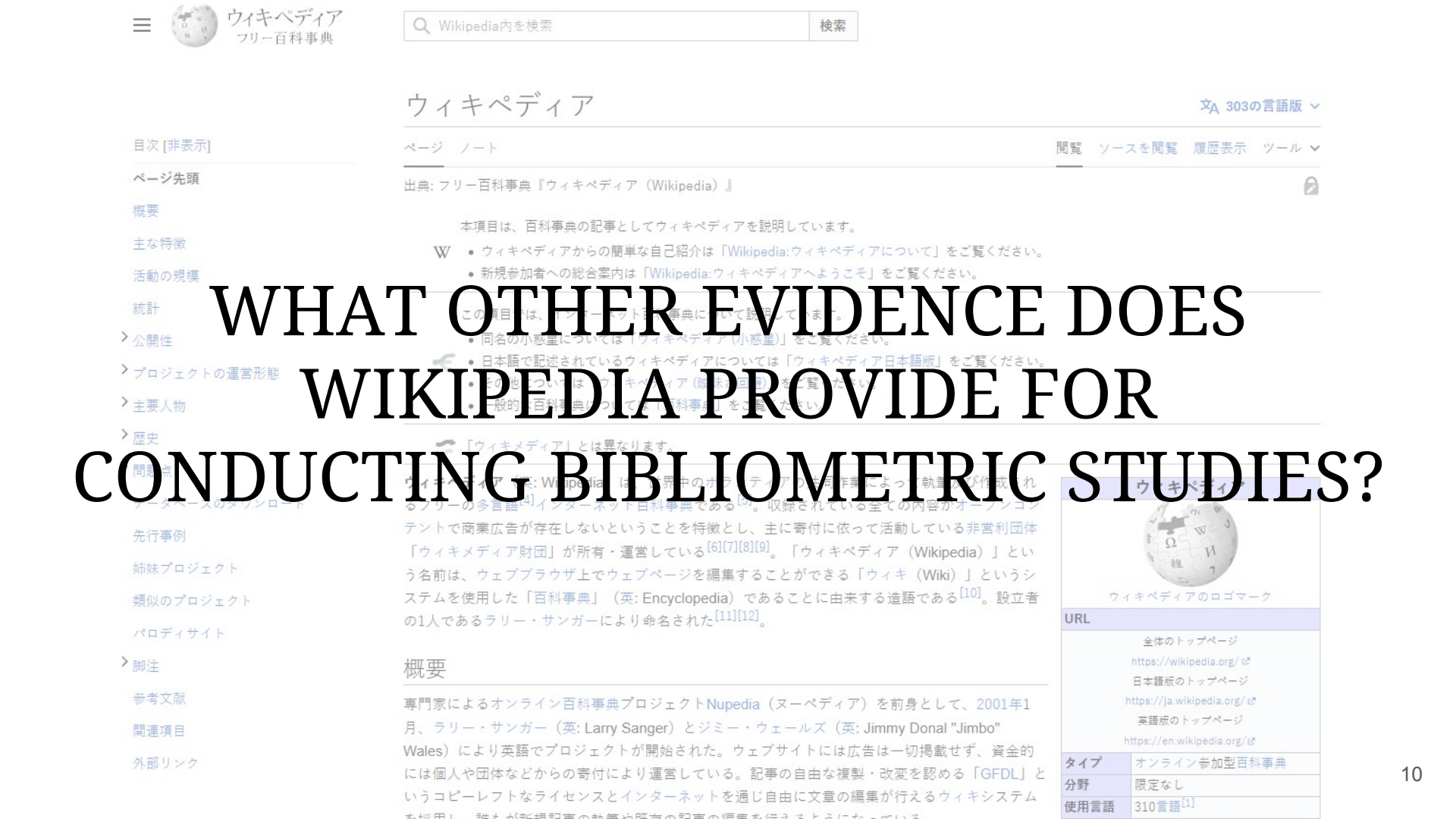
Using **classic techniques**, it has been possible to:

- map the structure of knowledge
- study the coverage
- rank publications



A topic of interest is comparing the differences between the social and academic perspectives

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*, 15(2), e0228713. https://doi.org/10.1371/journal.pone.0228713



WHAT OTHER EVIDENCE DOES WIKIPEDIA PROVIDE FOR CONDUCTING BIBLIOMETRIC STUDIES?

CLASSIC PERSPECTIVE

WIKIPEDIA
The Free Encyclopedia

Search Wikipedia | Search

COVID-19

188 languages

Article | Talk | Read | View source | View history | Tools

From Wikipedia, the free encyclopedia

For the ongoing pandemic, see COVID-19 pandemic. For other diseases caused by coronaviruses, see Coronavirus diseases.

Coronavirus disease 2019 (**COVID-19**) is a contagious disease caused by the virus SARS-CoV-2. The first known case was identified in Wuhan, China, in December 2019.[6] The disease quickly spread worldwide, resulting in the COVID-19 pandemic.

The symptoms of COVID-19 are variable but often include fever,[7] cough, headache,[8] fatigue, breathing difficulties, loss of smell, and loss of taste.[9][10][11] Symptoms may begin one to fourteen days after exposure to the virus. At least a third of people who are infected do not develop noticeable symptoms.[12] Of those who develop symptoms noticeable enough to be classified as patients, most (81%) develop mild to moderate symptoms (up to mild pneumonia), while 14% develop severe symptoms (dyspnea, hypoxia, or more than 50% lung involvement on imaging), and 5% develop critical symptoms (respiratory failure, shock, or multiorgan dysfunction).[13] Older people are at a higher risk of developing severe symptoms. Some people continue to experience a range of effects (long COVID) for years after infection, and damage to organs has been observed.[14] Multi-year studies are underway to further investigate the long-term effects of the disease.[15]

COVID-19 transmits when infectious particles are breathed in or come into contact with the eyes, nose, or mouth. The risk is highest when people are in close proximity, but small airborne particles containing the virus can remain suspended in the air and travel over longer distances, particularly indoors. Transmission can also occur when people touch their eyes, nose or mouth after touching surfaces or objects that have been contaminated by the virus. People remain contagious for up to 20 days and can spread the virus even if they do not develop symptoms.[16]

ing methods for COVID-19 to detect the virus's nucleic acid include real-time reverse transcription polymerase chain reaction (RT-PCR),[17][18] transcription-mediated amplification,[17][18][19] and reverse transcription loop-mediated isothermal amplification (RT-LAMP)[17][18] from a nasopharyngeal swab.[20]

Several COVID-19 vaccines have been approved and distributed in various countries, which have initiated mass vaccination campaigns. Other preventive measures include physical or social distancing, quarantining, ventilation of indoor spaces, use of face masks or coverings in public, covering coughs and sneezes, hand washing, and keeping unwashed hands away from the face. While work is underway to develop drugs that inhibit the virus, the primary treatment is symptomatic. Management involves the treatment of symptoms through supportive care, isolation, and experimental measures.

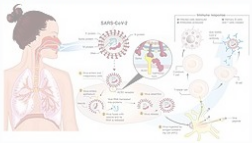
Contents [hide]

- (Top)
- Nomenclature
- Symptoms and signs
- Cause
- Pathophysiology
- Diagnosis
- Prevention
- Treatment
- Prognosis and risk factors
- Mortality
- History
- Misinformation
- Other species
- Research
- See also
- References
- Further reading
- External links

Cited references

Coronavirus disease 2019 (COVID-19)

Other names	COVID, (the) coronavirus



Transmission and life-cycle of SARS-CoV-2, which causes COVID-19

Pronunciation	/kə roʊnəvaɪrəs/ / kouvɪdneɪn ˈtiːn, kovɪd-[1]
Specialty	Infectious disease
Symptoms	Fever, cough, fatigue, shortness of breath, vomiting, loss of taste or smell; some cases asymptomatic[2][3]
Complications	Pneumonia, sepsis, ARDS, kidney failure, respiratory failure, pulmonary fibrosis, CKS, MIS-C, long COVID
Usual onset	2–14 days (typically 5) after infection
Duration	5 days to chronic
Causes	SARS-CoV-2
Diagnostic method	RT-PCR testing, CT scan, rapid antigen test
Prevention	Vaccination, face coverings, quarantine, social distancing, ventilation, hand washing

NOVEL PERSPECTIVE

This article is about a health issue

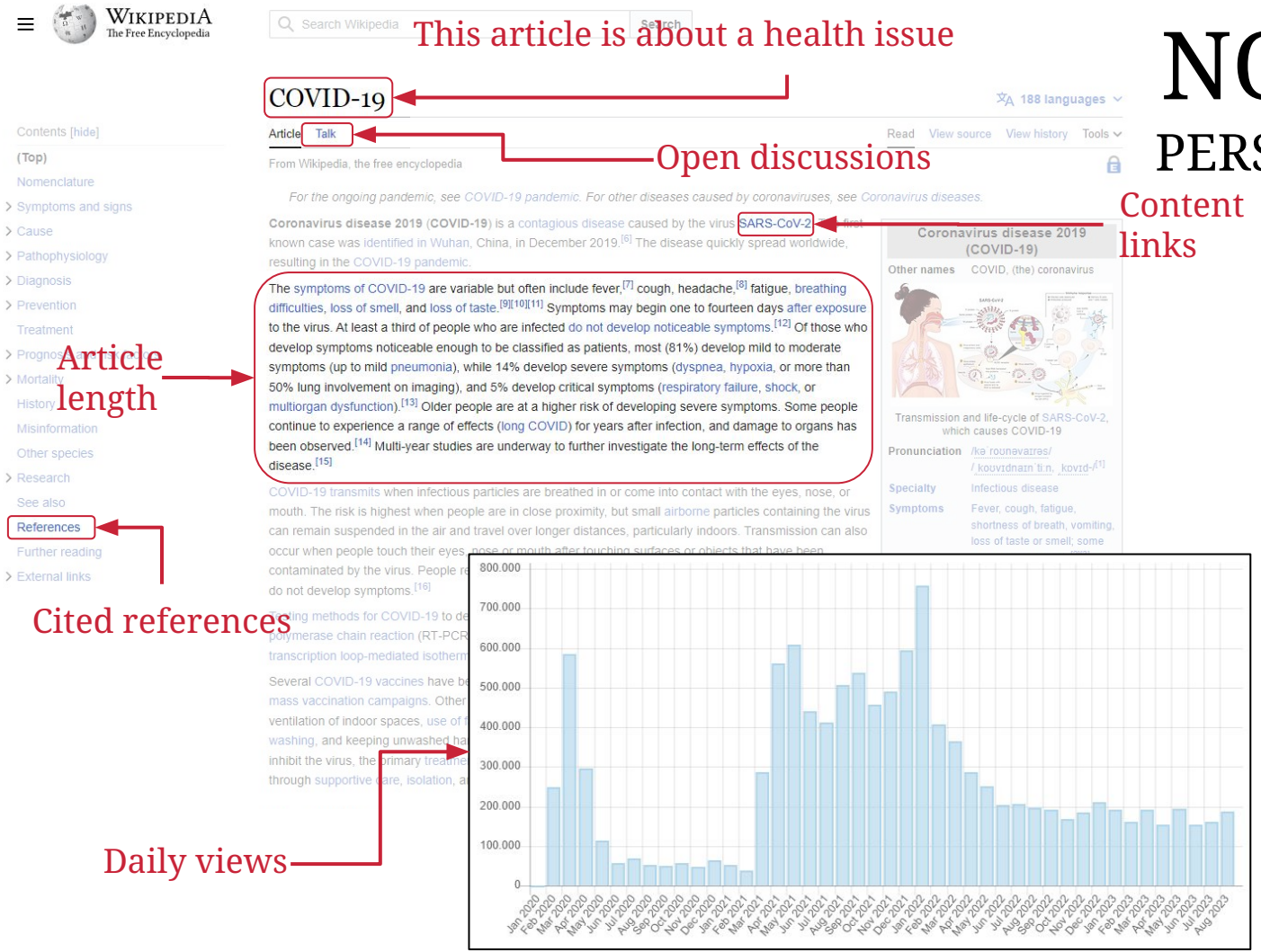
Open discussions

Content links

Article length

Cited references

Daily views



TOWARDS NEW HORIZONS

Wikipedia has many features that help characterize articles and are useful for contextualizing citations

Metric	Analytical dimension	Description
Editors	Activity	Number of unique editors that have edited a Wikipedia article
Edits	Activity	Number of total edits that have a Wikipedia article
Linked	Connectivity	Number of Wikipedia articles in which the article is linked to
Links	Connectivity	Number of internal links that include a Wikipedia article to others
Age	Description	Years that have passed since the creation of the page to the date of data collection
Length	Description	Length in bytes of the page
Talkers	Discussion	Number of unique editors that have edited a Wikipedia article's talk page
Talks	Discussion	Number of total edits that the talk page of a Wikipedia article has
Views	Outreach	Number of daily views of a Wikipedia page
References	Support	Number of elements listed in the references
Pub. referenced	Support	Number of publications referenced
URLs	Support	Number of external links that include a Wikipedia article

Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022). Wikinformetrics: Construction and description of an open Wikipedia knowledge graph data set for informetric purposes. *Quantitative Science Studies*, 3(4), 931-952. https://doi.org/10.1162/qss_a_00226

TOWARDS NEW HORIZONS

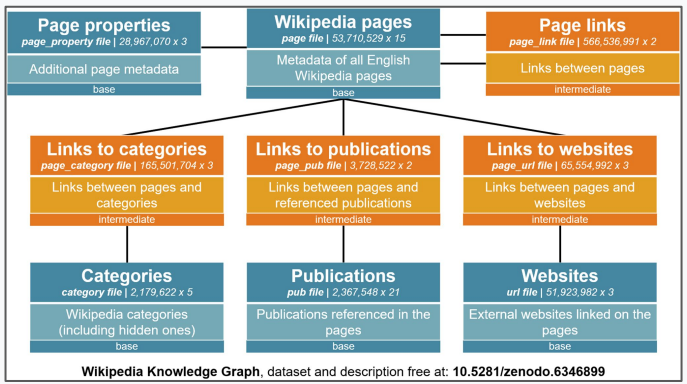
Moreover, there are classification systems for the level of development of articles

Class	Description
Featured article	The best possible content on Wikipedia, no need for improvement
Featured list	The best possible list on Wikipedia, no need for improvement
A	Fully addresses the subject and requires only minor improvements
Good article	It satisfies Wikipedia's main criteria and is close to a professional article
B	The content is almost complete and has no major problems
C	The content is considerable, but has significant problems
Start	It includes significant content, but is still in development
Stub	The content is very short and requires substantial work
List	Content displayed in a list linking to Wikipedia articles on a specific topic

Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022). Wikinformetrics: Construction and description of an open Wikipedia knowledge graph data set for informetric purposes. *Quantitative Science Studies*, 3(4), 931-952. https://doi.org/10.1162/qss_a_00226

TOWARDS NEW HORIZONS

After reviewing all these possibilities, we constructed a **knowledge graph** of the English Wikipedia, linking the characteristics and relationships of Wikipedia articles with the cited references



Schema of the Wikipedia Knowledge Graph

Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022). Wikinformetrics: Construction and description of an open Wikipedia knowledge graph data set for informetric purposes. *Quantitative Science Studies*, 3(4), 931-952. https://doi.org/10.1162/qss_a_00226

TOWARDS NEW HORIZONS

Analyzing all this data, we found that **references** play a relevant role within Wikipedia and closely approximate the quality of Wikipedia articles

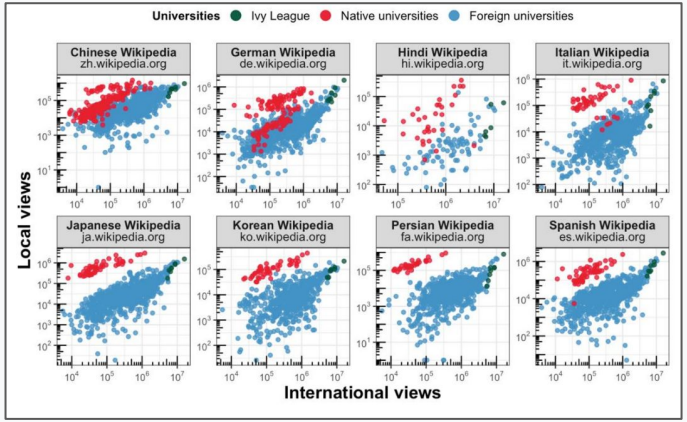
N. of articles → / Wiki Metrics ↓	All articles	Featured articles	Featured lists	A	Good	B	C	List	Start	Stub
	6,328,134	5945	3816	958	34,004	109,019	394,065	253,066	1,818,356	3,079,778
Editors	48.38	516.93	179.13	176.80	275.71	297.62	165.36	56.27	63.13	22.85
Edits	101.92	1491.35	593.61	564.91	724.13	705.41	369.89	159.80	129.52	40.23
Linked	80.53	725.25	175.84	202.01	330.18	417.00	234.08	107.34	93.03	55.70
Links	87.77	329.68	270.16	236.56	224.88	233.87	164.23	174.78	101.28	69.90
Age	9.59	14.33	11.52	12.74	12.06	12.47	10.92	9.13	10.45	9.20
Length	7844.68	61,248	51,549	43,329	39,444	35,009	21,676	18,202	10,033	3748
Talkers	5.38	66.17	16.62	27.90	29.64	28.16	15.03	4.98	6.56	3.64
Talks	9.19	258.40	42.36	92.21	88.56	88.35	35.32	9.07	9.69	4.32
Views	3345.07	64,801	26,685	16,011	29,229	30,359	15,829	3777	4094	710
References	4.6	53.95	55.49	31.76	38.87	26.51	15.40	9.20	5.79	1.84
Pub. Ref.	0.59	14.27	2.34	8.51	5.83	4.77	2.37	0.53	0.69	0.22
URLs	10.33	58.03	67.32	33.32	46.10	40.31	25.95	22.82	12.90	6.09

Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022). Wikinformetrics: Construction and description of an open Wikipedia knowledge graph data set for informetric purposes. *Quantitative Science Studies*, 3(4), 931-952. https://doi.org/10.1162/qss_a_00226

TOWARDS NEW HORIZONS

Thanks to this perspective, it is possible to formulate new research questions and conduct much **broader bibliometric studies**

We are no longer limited to replicating methods of bibliometrics, but can take advantage of the wide possibilities of Wikipedia



Another advantage is the opportunity to compare the perspectives of the different Wikipedias

Arroyo-Machado, W., Díaz-Faes, A. A., Herrera-Viedma, E., & Costas, R. (2023). From academic to media capital: To what extent does the scientific reputation of universities translate into Wikipedia attention? https://doi.org/10.48550/ARXIV.2307.05366

COMPARISON BETWEEN ACADEMIC PERFORMANCE AND ENGAGEMENT IN WIKIPEDIA

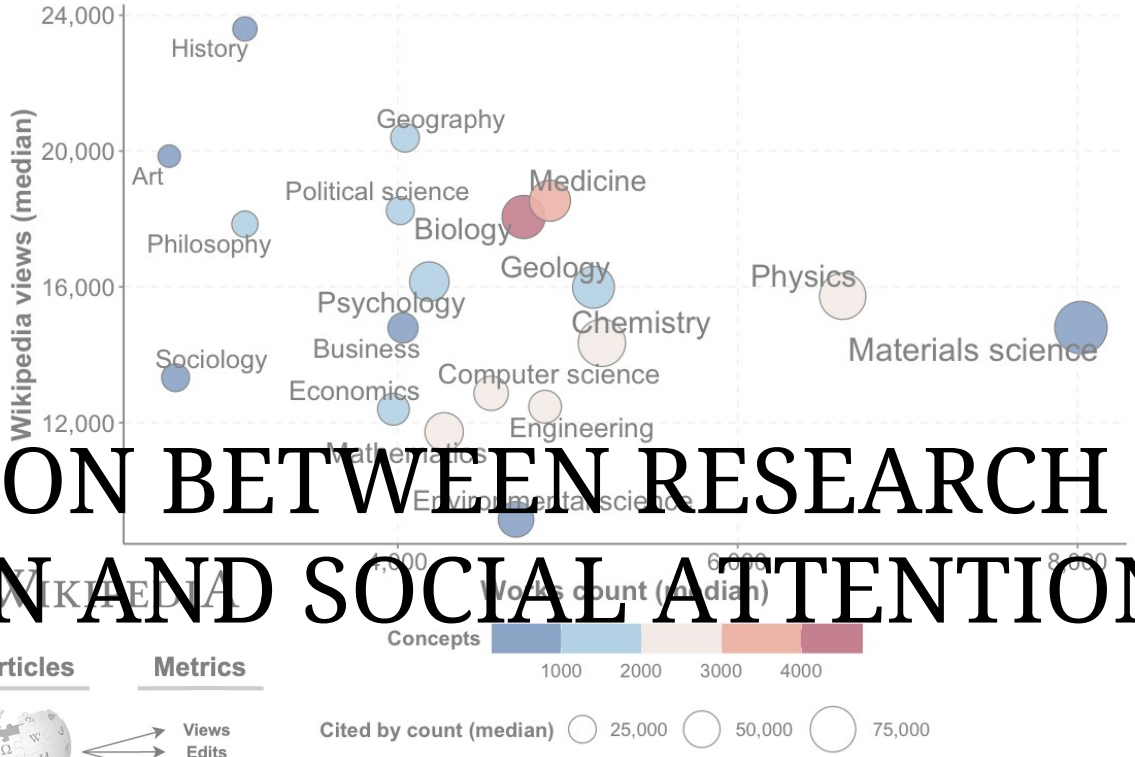
	University		Research performance			Wikipedia			
Country	No.	Age	P	P int collab	P top 10%	Language links	Local views	International views	Edits
China	217	89.4	8015.9	2254.1	917	8.2	95825.9	93236.2	799.6
United States	200	147.3	10694.9	4621.6	1796.4	31.2	1979702.4	1979702.4	4906.3
United Kingdom	61	161.8	9448.3	6180.3	1679.3	37.5	1244018.8	1242904.7	4350.6
Germany	54	232	8322.2	4908.8	1217.3	36.5	320047.9	291413.9	2932.4
Japan	54	104.6	5765.9	1974.7	494.6	16.4	756578.5	130824.8	1633.9
South Korea	46	86.6	6492.7	1956.2	532	12.6	121928	248131.2	1549.7
Italy	42	323.1	7596.4	4058.7	980.5	25.5	262138.7	178701.6	1714.4
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
India	[illegible]	74.1	3349.7	[illegible]	278.2	[illegible]	788956.2	[illegible]	1622.2
Iran	36	61.5	3797.6	1066.3	333.1	8.1	242611.4	60166.7	985.5
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Turkey	32	78.8	2702.6	819.3	185.8	14.5	201000.8	114498.3	1219.9
Brazil	31	71.4	6191.3	[illegible]	[illegible]	[illegible]	181464.9	48964.7	1165
Poland	31	116.7	3139.3	[illegible]	[illegible]	[illegible]	171856.4	80605.7	1125.6
Canada	30	120.9	10716.4	6024.1	1538.9	30.9	1079240.8	995113.6	3521.5
France	30	272.9	10101.3	6278.6	1442.6	27.3	147720.5	166818.3	1291.4
Taiwan	21	78.7	5128.4	1754.3	408.2	13.1	570572.1	99502.3	2151.9
The Netherlands	13	194.9	15213.8	9608.5	2604.1	39.3	131776.4	447745.3	2715.2
Austria	12	284.9	4813.4	3416.4	685.8	24.9	141606.1	190930.5	1467.7
Sweden	12	163.7	11142.8	7558.3	1651.8	38.6	132864.3	330566.3	2410.5
Russia	10	138.2	5009.7	3017.4	400.8	35.5	772638.6	306399.3	2995.3

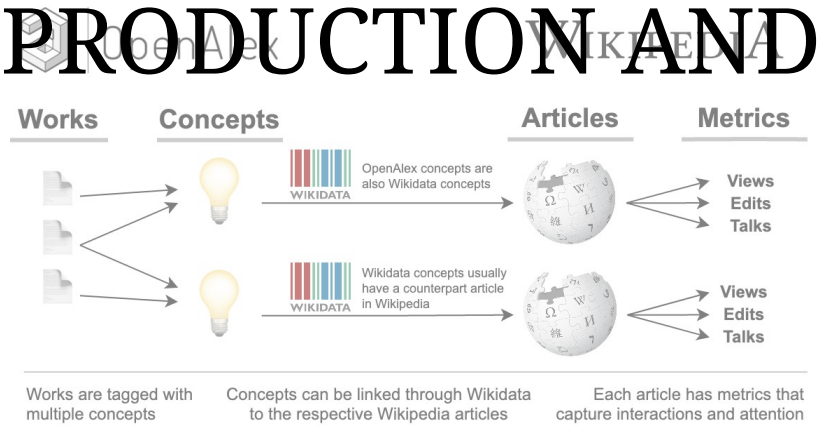
Arroyo-Machado, W., Díaz-Faes, A. A., Herrera-Viedma, E., & Costas, R. (2023). From academic to media capital: To what extent does the scientific reputation of universities translate into Wikipedia attention? https://doi.org/10.48550/ARXIV.2307.05366

	University		Research performance			Wikipedia			
Country	No.	Age	P	P int collab	P top 10%	Language links	Local views	International views	Edits
China	217	89.4	8015.9	2254.1	917	8.2	95825.9	93236.2	799.6
United States	200	147.3	10694.9	4621.6	1796.4	31.2	1979702.4	1979702.4	4906.3
United Kingdom	61	161.8	9448.3	6180.3	1679.3	37.5	1244018.8	1242904.7	4350.6
Germany	54	232	8322.2	4908.8	1217.3	36.5	320047.9	291413.9	2932.4
Japan	54	104.6	5765.9	1974.7	494.6	16.4	756578.5	130824.8	1633.9
South Korea	46	86.6	6492.7	1956.2	532	12.6	121928	248131.2	1549.7
Italy	42	323.1	7596.4	4058.7	980.5	25.5	262138.7	178701.6	1714.4
Spain	42	219.3	5305.1	2921.9	608.9	21.7	245805.1	116310.5	1346.1
India	38	74.1	3349.7	910.6	278.2	15.1	788956.2	745249.5	1822.2
Iran	36	61.5	3797.6	1066.3	333.1	8.1	242611.4	60166.7	985.5
Australia	32	73.5	10643.6	6275.8	1631.8	24.7	563584.3	563584.3	2430.8
Turkey	32	78.8	2702.6	819.3	185.8	14.5	201000.8	114498.3	1219.9
Brazil	31	71.4	6191.3	2573.9	501.1	17.8	181464.9	48964.7	1165
Poland	31	116.7	3139.3	1205.1	244.1	18.3	171856.4	80605.7	1125.6
Canada	30	120.9	10716.4	6024.1	1538.9	30.9	1079240.8	995113.6	3521.5
France	30	272.9	10101.3	6278.6	1442.6	27.3	147720.5	166818.3	1291.4
Taiwan	21	78.7	5128.4	1754.3	408.2	13.1	570572.1	99502.3	2151.9
The Netherlands	13	194.9	15213.8	9608.5	2604.1	39.3	131776.4	447745.3	2715.2
Austria	12	284.9	4813.4	3416.4	685.8	24.9	141606.1	190930.5	1467.7
Sweden	12	163.7	11142.8	7558.3	1651.8	38.6	132864.3	330566.3	2410.5
Russia	10	138.2	5009.7	3017.4	400.8	35.5	772638.6	306399.3	2995.3

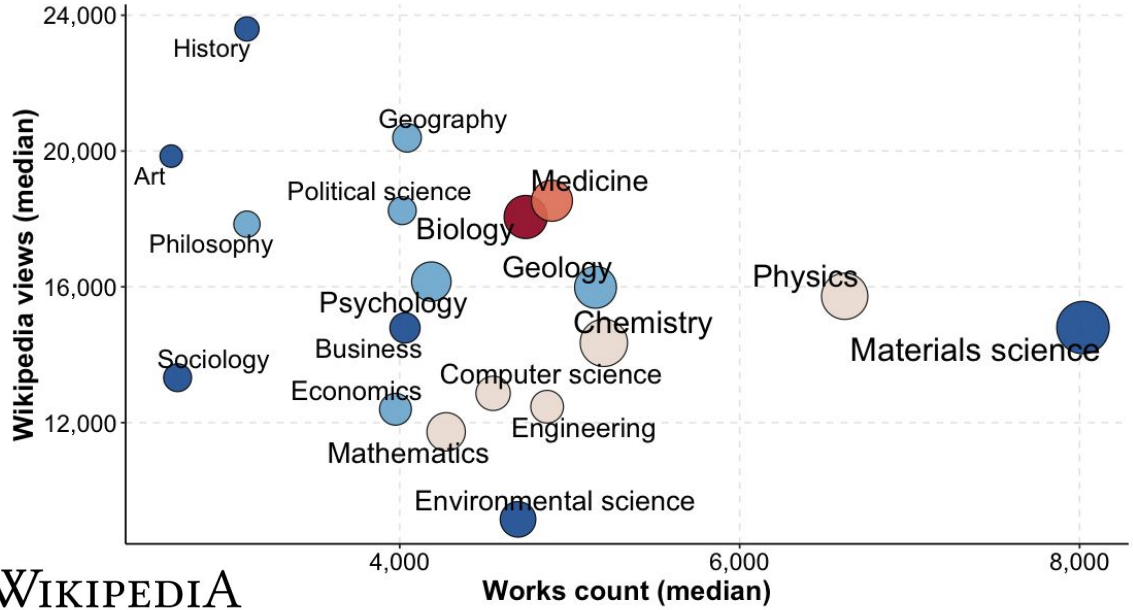
Arroyo-Machado, W., Díaz-Faes, A. A., Herrera-Viedma, E., & Costas, R. (2023). From academic to media capital: To what extent does the scientific reputation of universities translate into Wikipedia attention? https://doi.org/10.48550/ARXIV.2307.05366

COMPARISON BETWEEN RESEARCH PRODUCTION AND SOCIAL ATTENTION



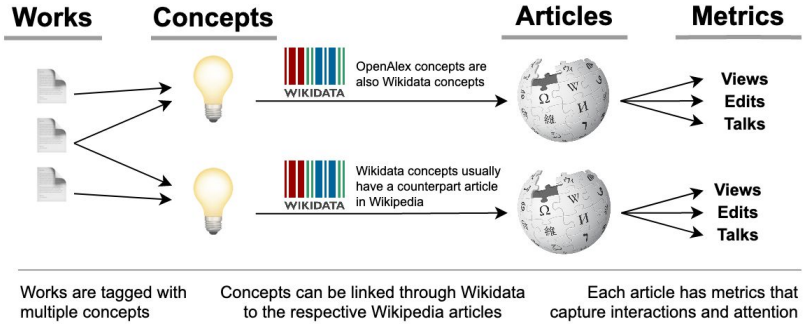


Arroyo-Machado, W., & Costas, R. (2023). Do popular research topics attract the most social attention? A first proposal based on OpenAlex and Wikipedia. *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. 27th International Conference on Science, Technology and Innovation Indicators (STI 2023). https://doi.org/10.55835/6442bb04903ef57acd6dab9e



OpenAlex

WIKIPEDIA



Arroyo-Machado, W., & Costas, R. (2023). Do popular research topics attract the most social attention? A first proposal based on OpenAlex and Wikipedia. *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. 27th International Conference on Science, Technology and Innovation Indicators (STI 2023). https://doi.org/10.55835/6442bb04903ef57acd6dab9e

MAIN CHALLENGES

These methods are very useful and Wikipedia offers all its data openly, but there are two main challenges:

1. Extensive data processing required
2. Lack of standardization of references
3. There are no rules!

WIKIMEDIA FOUNDATION

Thanks!

Any questions?