gesis
**Leibniz-Institut
für Sozialwissenschaften**

Altmetrics – Analyzing Academic Communications from Social Media Data

Meet the Experts! – GESIS online talks

Katrin Weller, Olga Zagovora · December 2, 2021

Leibniz
Gemeinschaft

Speakers

Dr. Katrin Weller

- Team lead Digital Society Observatory, Department Computational Social Science
- PhD in Information Science
- Social media research methods, altmetrics
- Contact: katrin.weller@gesis.org

Olga Zagovora

- Researcher in the team Digital Society Observatory, Department Computational Social Science
- MSc. in Web Science
- Scholarly communication online, altmetrics, gender biases
- Contact: olga.zagovora@gesis.org

Background - Studying scholarly communication and scientometrics:

from *bibliometrics*

to *altmetrics*

Publications and **citations** have long been at the core of scholarly communication -

they represent how information flows from one researcher to the next.

Publications and **citations** have long been at the core of scholarly communication -

they represent how information flows from one researcher to the next.

information flow
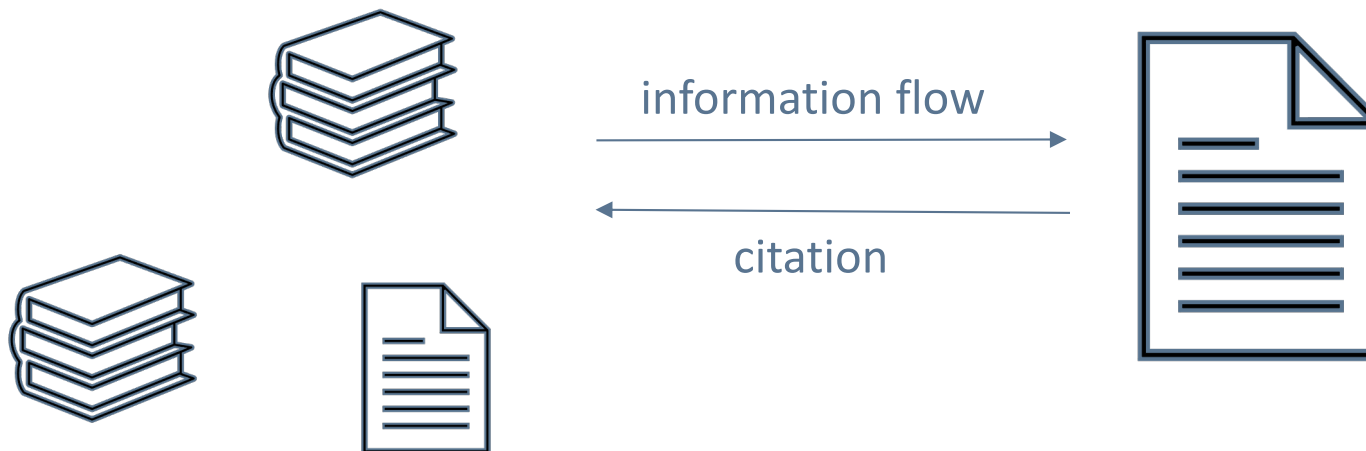
citation

Scholarly communications as a field of study

Publication and citation data enable research on scholarly communication processes, e.g.,

- sociology of science
- analyse structures of the scientific field (e.g. international or disciplinary collaborations)
- investigate patterns of scientific discovery and trends
- identify influential researchers and their networks

Scholarly communication as a field of study
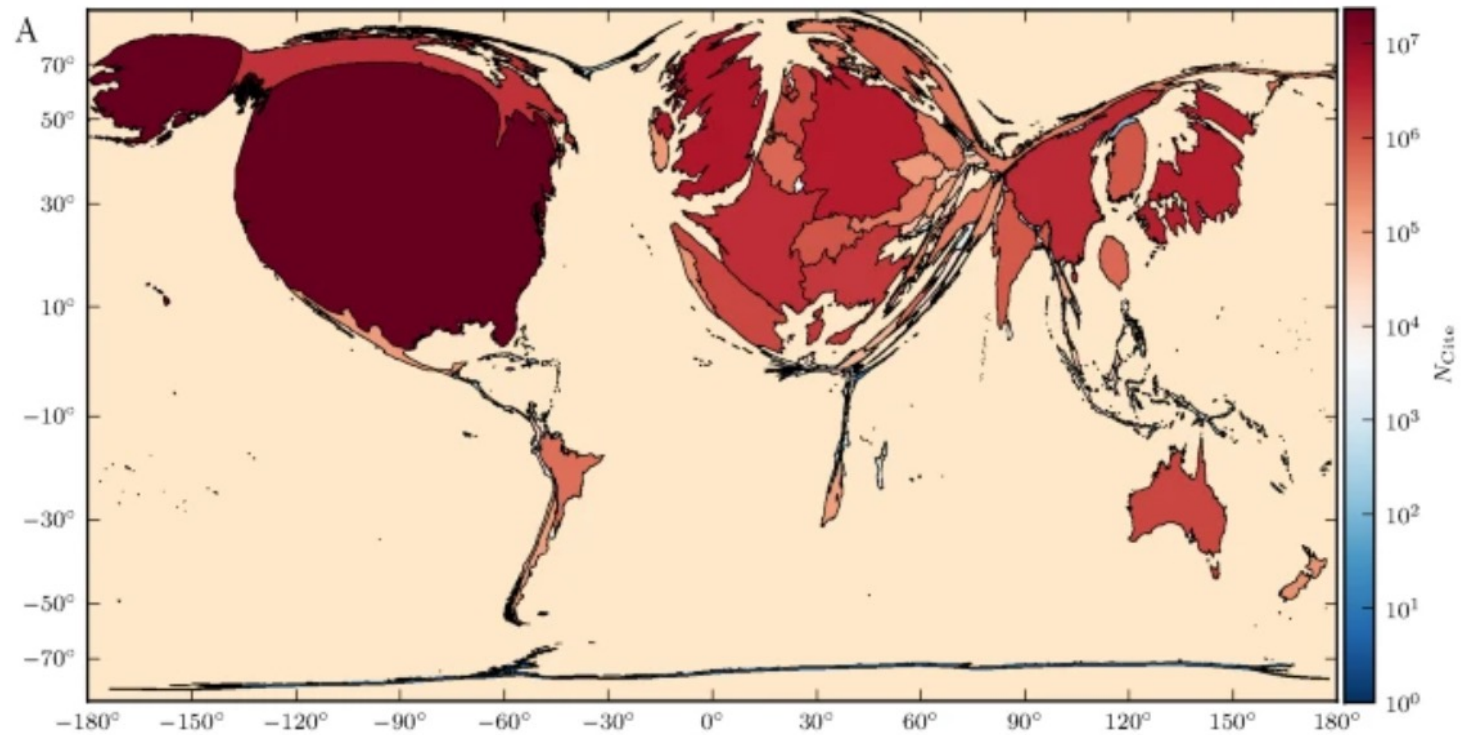
Citation map of the world where the area of each country is scaled and deformed according to the number of citations received, which is also represented by the color of each country.

Pan, R., Kaski, K. & Fortunato, S. World citation and collaboration networks: uncovering the role of geography in science. *Sci Rep* **2,** 902 (2012). https://doi.org/10.1038/srep00902

Scholarly communication as a field of study
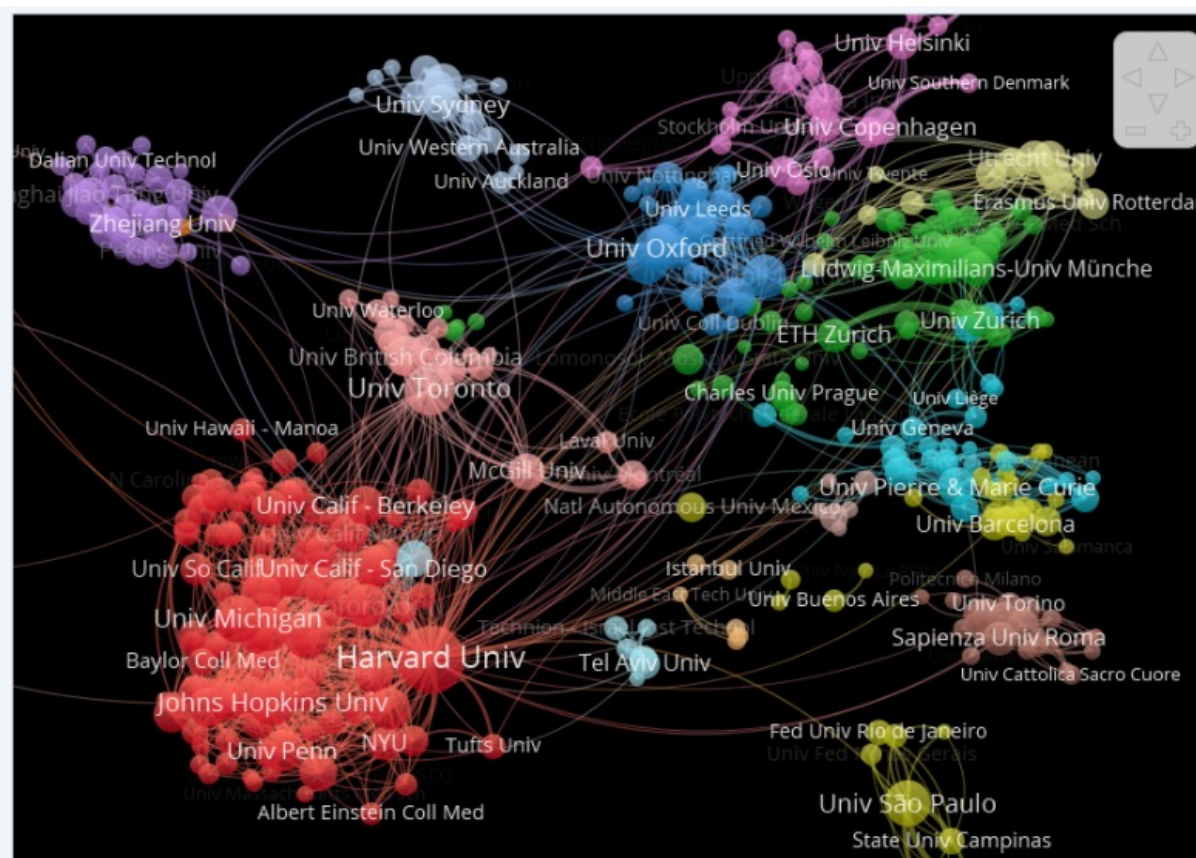
Example from VOSviewer http://www.vosviewer.com/

Over time, publications and citations have also become the key currency for measuring **scientific impact**.

Google Scholar – Author citations
https://scholar.google.com/

Scimago – Journal Rank Indicator
https://www.scimagojr.com/journalrank.php

… but these impact measures come with
several challenges and limitations.

Some limitations of traditional citation analysis

Traditional citation counts...

- ... are slow in responding to new scientific findings
- ... can often not be compared across disciplines
- ... do not differentiate different reasons to cite
- ... do only capture a specific type of impact
- ...

Some limitations of traditional citation analysis

Traditional citation counts…

- … **are slow in responding to new scientific findings**
- … can often not be compared across disciplines
- … do not differentiate different reasons to cite
- … **do only capture a specific type of impact**
- …

Potential solutions via altmetrics

Traditional citation counts…

- … do only capture a specific type of impact

Altmetrics attempt to include **alternative ways** to measure impact of scholarly work, e.g., by considering non-academic communities and general publics, journalistic ressources, teaching activities, etc.

Potential solutions via altmetrics

Traditional citation counts…

- … are slow in responding to new scientific findings

Using data from **social media** and other alternative sources enables **more timely** reflections on scientific impact.

Let us take a first look how scholarly communication is also happening in online environments such as **social media platforms**.

Alternative interactions with scholarly publications

Some examples:

- Social media users mentioning a published paper (often via links / DOIs)

- Reference lists in Wikipedia articles

- Bookmarking / networking platforms

- News reports of scientific findings

Example

Wikipedia views to measure interest
in individual researchers.

Nobel Prize in Physics 2008 Wikipedia Page Views Statistics

Before the announcement

Legend:
- Vera Rubin
- Makoto Kobayashi
- Toshihide Maskawa

Wagner, C., Zagovora, O., Sennikova, T., & Karimi, F. (2018). Collective Attention towards Scientists and Research Topics. Proceedings of the 10th ACM Conference on Web Science - WebSci '18, 11–15. https://doi.org/10.1145/3201064.3201097

Nobel Prize in Physics 2008 Wikipedia Page Views Statistics

Time around the announcement

Legend: Vera Rubin, Makoto Kobayashi, Toshihide Maskawa
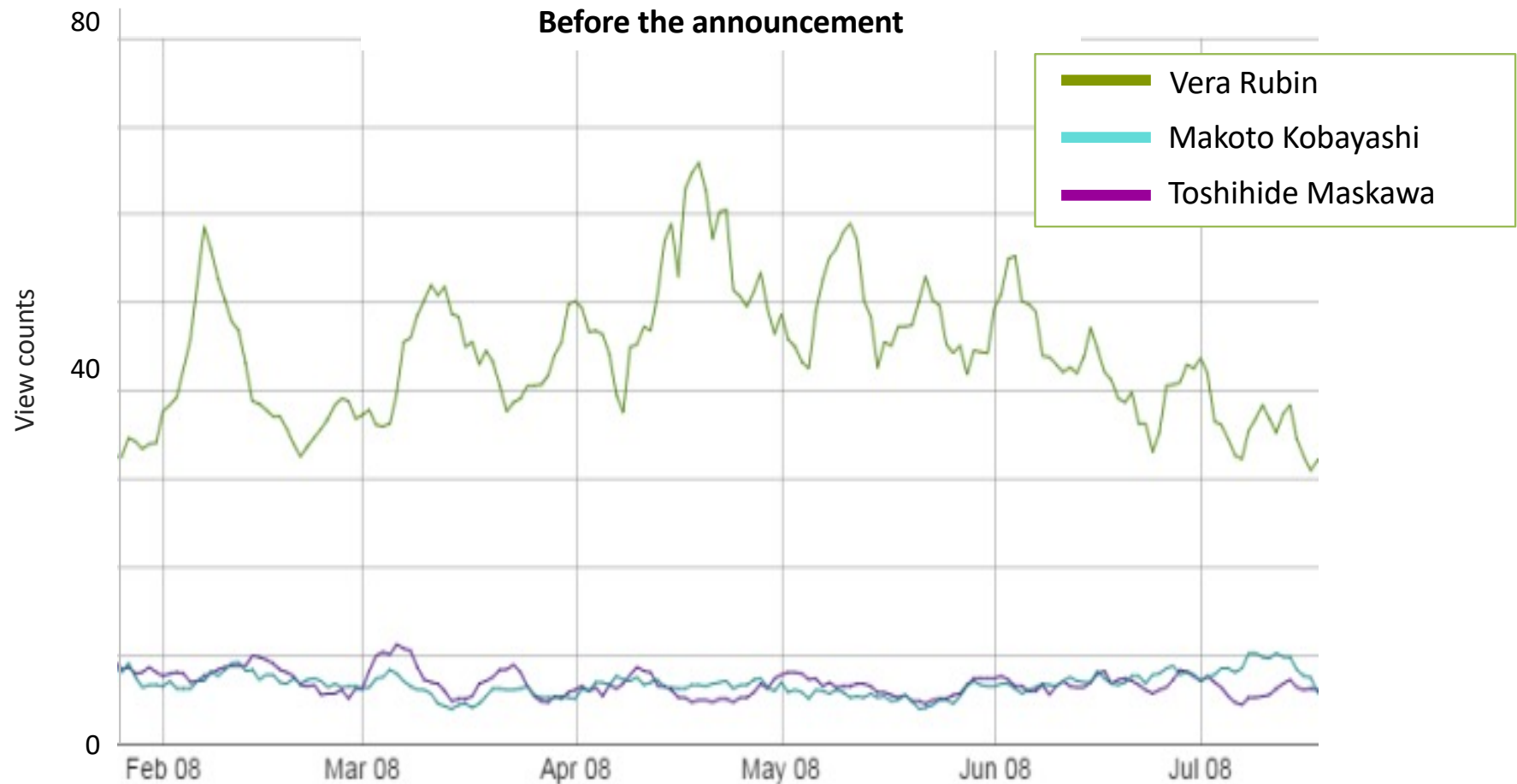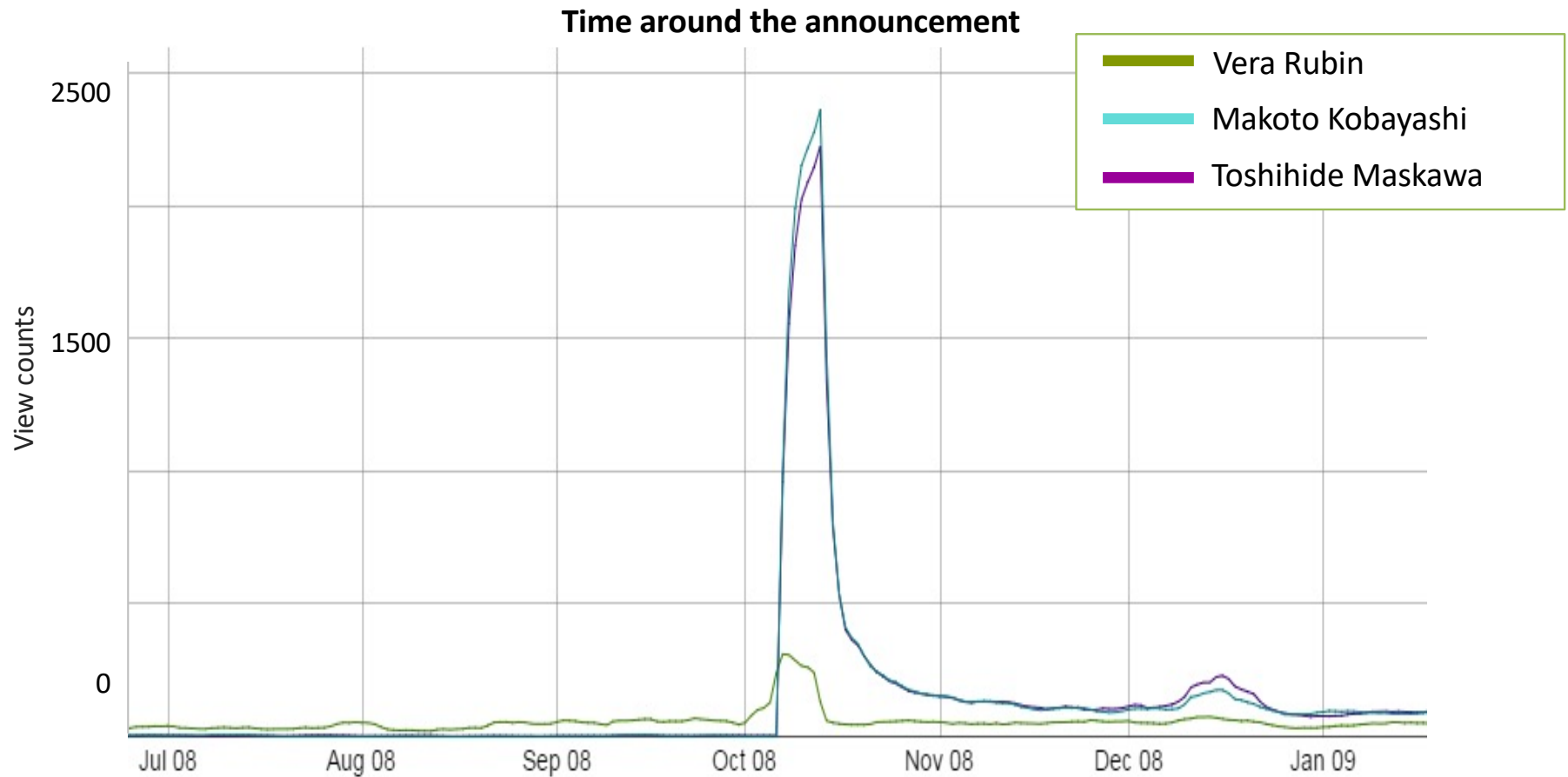
Practical examples:

Cited papers on YouTube
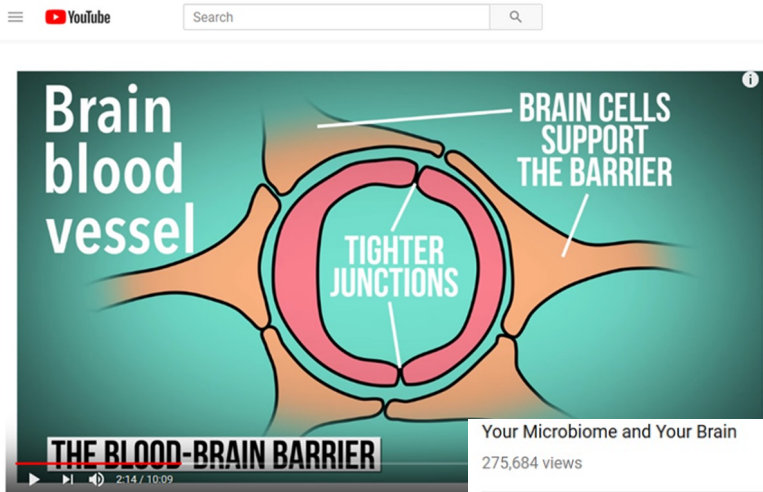
YouTube video
with textual video
description

that includes a link to a

published paper

Research subjects (based on Scopus) of papers that are mentioned in YouTube video descriptions.

Zagovora, O., & Weller, K. (2018). YouTube videos as a place for public discussion of science? The 2018 Altmetrics Workshop, 1–16. http://altmetrics.org/wp-content/uploads/2018/04/altmetrics18_paper_8_Zagovora.pdf
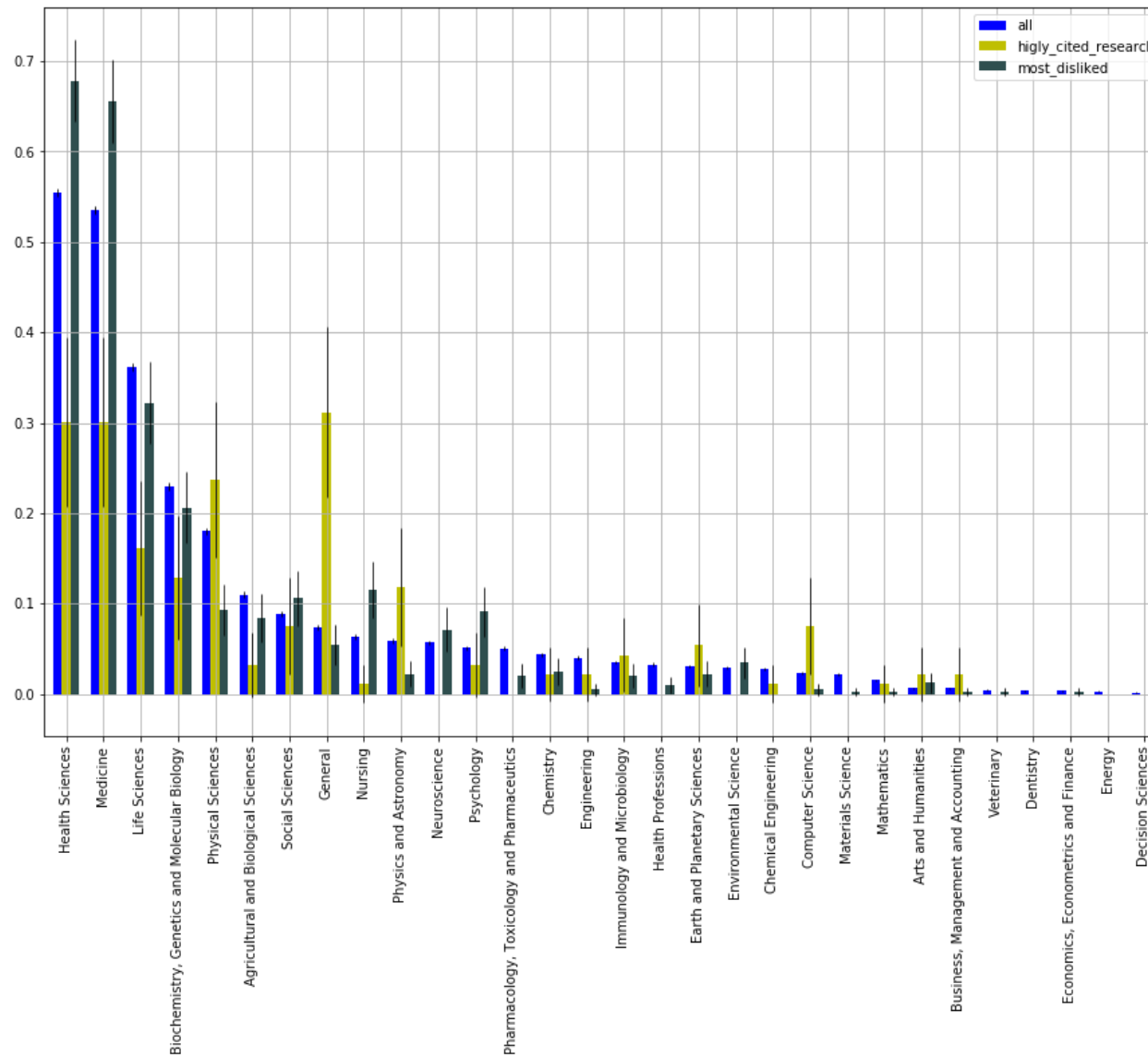
Practical examples:
publishers implementing altmetrics

PLOS ONE

HTML Views as Citation Count Predictors
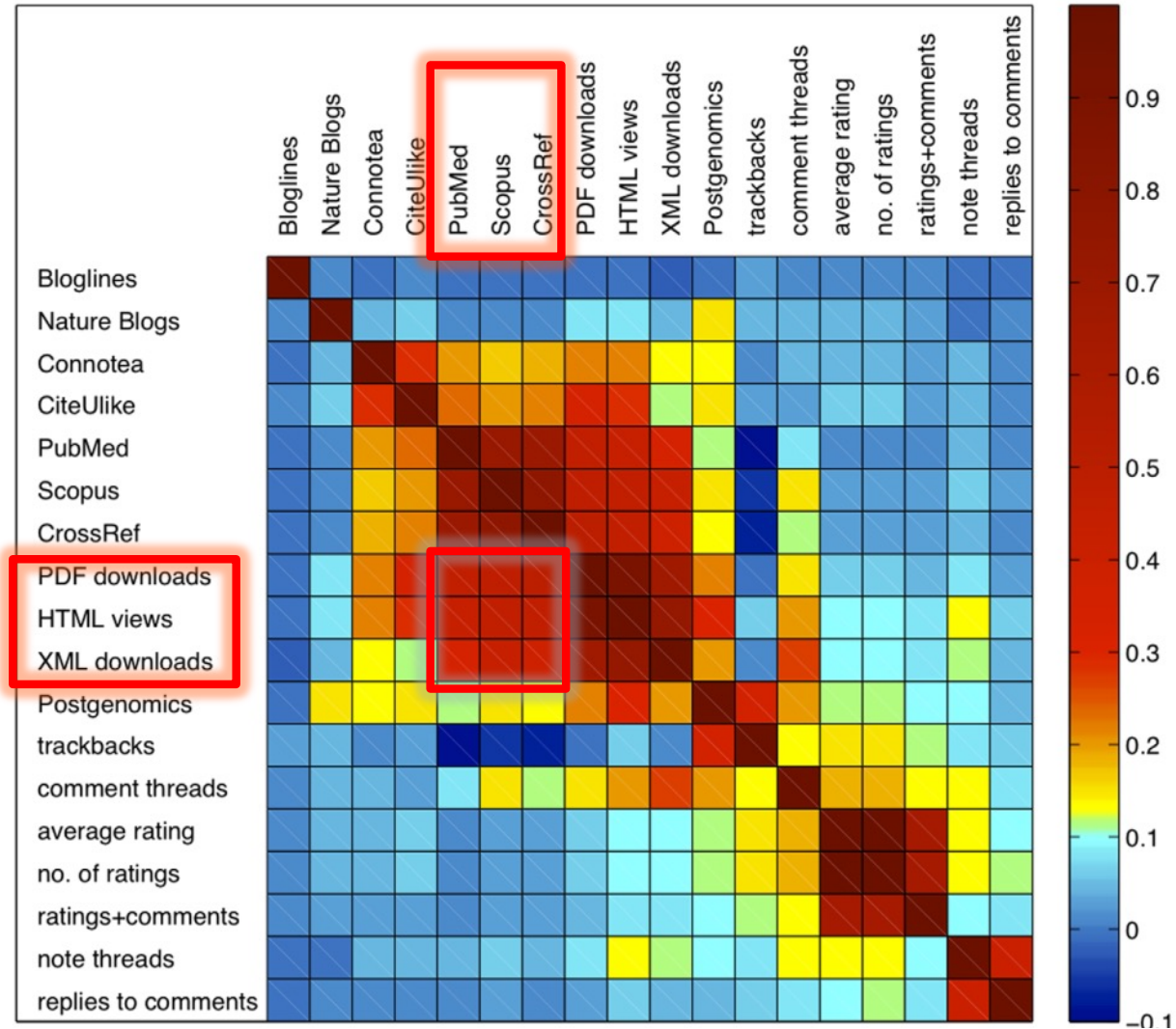
Yan K-K, Gerstein M (2011) The Spread of Scientific Information: Insights from the Web Usage Statistics in PLoS Article-Level Metrics. PLoS ONE 6(5): e19917. https://doi.org/10.1371/journal.pone.0019917

So how can you practically work
with altmetrics data? Some options:

Data providers
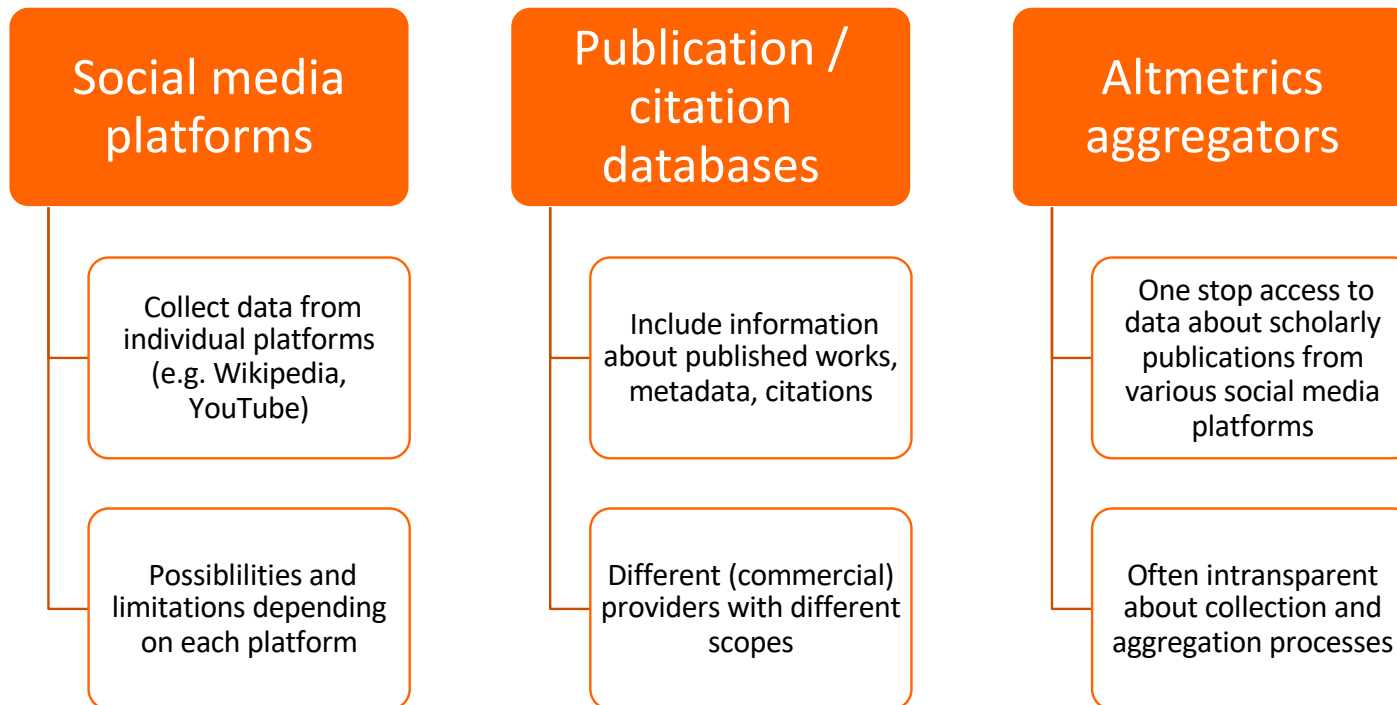
Social media platforms

Collect data from individual platforms (e.g. Wikipedia, YouTube)

Possiblilities and limitations depending on each platform

Publication / citation databases

Include information about published works, metadata, citations

Different (commercial) providers with different scopes

Altmetrics aggregators

One stop access to data about scholarly publications from various social media platforms

Often intransparent about collection and aggregation processes

Altmetrics aggregators' mechanism

Publications
Scholarly products
Publication databases

Online Platforms
Social media
News and other sources

Information flow?

"citation"

Altmetrics aggregators' mechanism

Publications
Scholarly products
Publication databases

Online Platforms
Social media
News and other sources

Information flow?

"citation"
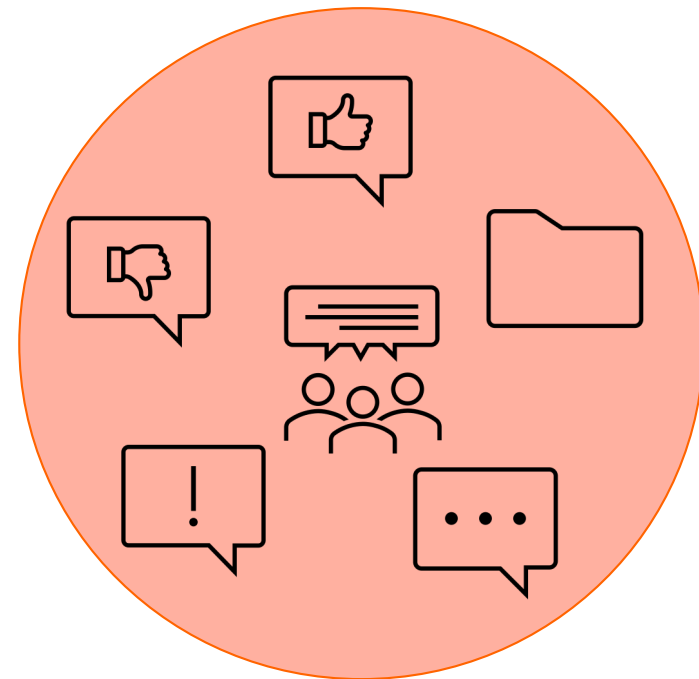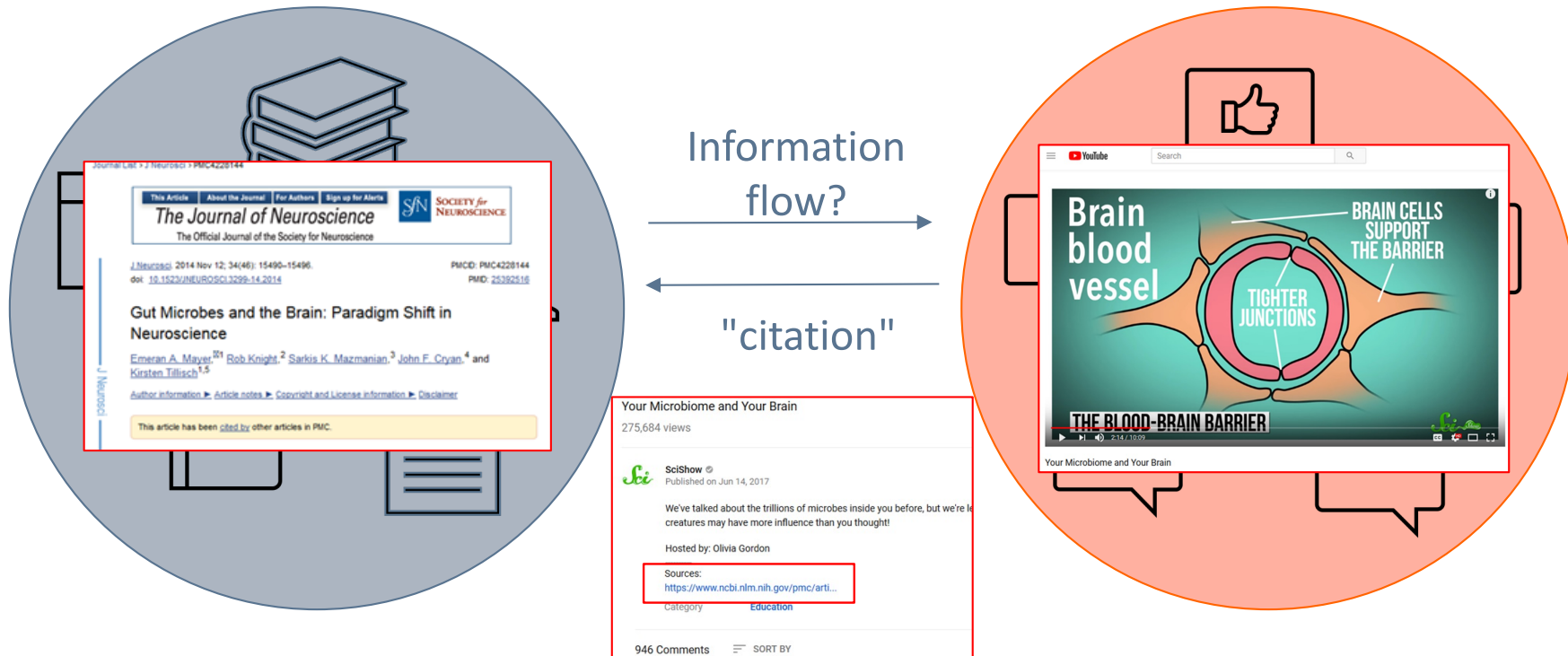
Altmetrics Aggregators

Name	API and access	Metrics
Altmetric.com	free API for researchers	Donuts/badges/bars, attention score, summary counts (Tweets, Facebook posts, etc.; see next slides), demographics (Twitter+Mendeley), rankings&percentiles (among all/same source/same age/same source + age), cited (Scopus).
PlumX	[Elsevier]; API; tracks also data stored in institutional repo	Clicks, downloads, views, library holdings, video plays, bookmarks, code forks, favorites, readers, watchers , blog posts, comments, reviews, Wikipedia links, likes, shares, tweets, citation indexes, patent citations, clinical citations
CrossRef Event Data	free API	Tweets, Newsfeed (incl. blogs), F1000, Hypothes.is, Reddit, StackExchange, The Lens (patent documents), Wikipedia, other webpages.
ImpactStory	API (deprecated), pages can be viewed in json	Author: percentiles & Achievements (open access, hot Steak, etc), Timeline mentions, Publication: saved & shared
Lagotto, ALM (reports)	[PLOS]; free API, Python client, R package; open-source API	Viewed (PLOS+PubMed Central+figshare), cited (Scopus, CrossRef, WoS, PubMed, PMC Europe, etc), Saved (Mendeley+CiteULike), Share (Faceboof+Tweeter), Discussed (Blogs, social media, news, Wikipedia, etc.) , Recommended (F1000)
ResearchGate	no API	ResearchGate Score, Reach (direct & indirect), percentile of the Score, Reads/Citations/Views, Followers
Mendeley	[Elsevier]; API	Citations, h-index, publications, views, readers, demographics, followers, following

Altmetric.com Sources (1)
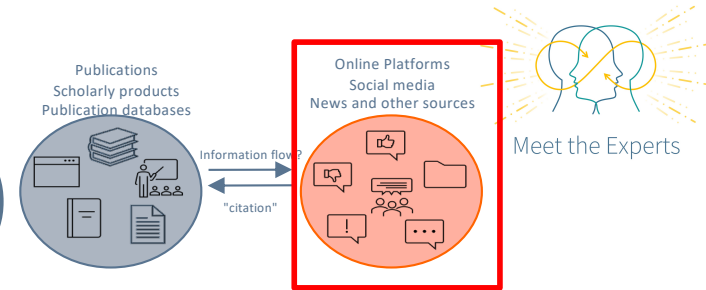
Source	Metrics	Mining methods	Limitations
Mainstream media: [2,000 outlets and magazines](#)	# news stories, # outlets	1. Link recognition from RSS feeds or APIs 2. Text mining	1. ~~Text mining is used for English only~~ 2. Text mining mostly for articles with DOIs
Wikipedia	# citations, # pages	Pages that used any of the {{cite journal}}, {{cite doi}}, {{cite pmid}}, or {{DOI}} tags	English Wikipedia only
Facebook	# public posts, # users	Direct link to a scholarly output	Mentions on public pages only, likes are not tracked
Twitter	# public tweets, # users, upper bound of followers, demographics of users	Direct link to a scholarly output	Started from the summer 2011; no favorites
Reddit	# original posts	Direct link to a scholarly output	Comments are not tracked

Altmetric.com Sources (2)

Publications
Scholarly products
Publication databases

Online Platforms
Social media
News and other sources

Meet the Experts

Information flow?
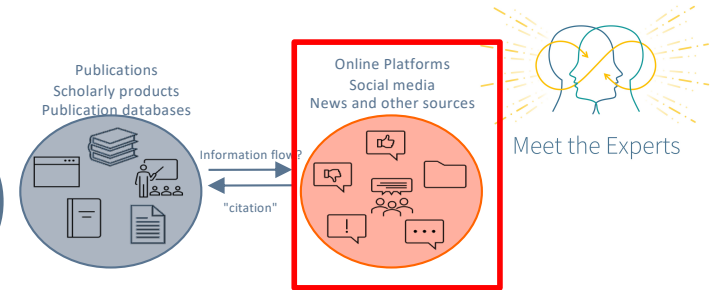
"citation"

Source	Metrics	Mining methods	Limitations
Faculty Opinions (formerly F1000)	# recomendations	„Through XML feed"	F1000 is for Biology and Medicine only
Blogs: over 11,000 academic and non-academic blogs	# posts, # blogs	Direct links to research outputs [from RSS feed]	
Pubpeer, Publons	# reviews		
Mendeley	# users who saved the article to their library, demographics	„Directly from the API"	
CiteULike	# bookmarks		
LinkedIn, Sina Weibo, Pinterest, Google+	# posts		Historical data only

Other sources:
- YouTube
- StackOverflow, StackExchange
- Public policy documents, Patents
- Open Syllabus Project: about 4,000 institutions

34

Altmetric.com Identifiers & Outputs tracked

Publications
Scholarly products
Publication databases

Online Platforms
Social media
News and other sources

Information flow?
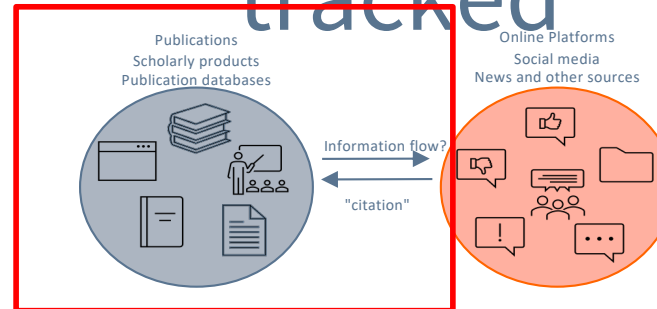
"citation"

Research outputs:

- ▶ Books
- ▶ Book Chapters
- ▶ Journal Articles
- ▶ Presentations
- ▶ Reports
- ▶ Data Sets
- ▶ Policy Documents
- ▶ Syllabi
- ▶ White papers

Identifier attached to the output:

- ▪ DOI
- ▪ ISBN
- ▪ URN
- ▪ arXiv ID
- ▪ PubMedID
- ▪ ADS ID
- ▪ SSRN ID
- ▪ RePEC ID
- ▪ Handle.net identifiers
- ▪ System handles disambiguation of the same research output

Donuts

- The colors of the donut represent a source of attention

● Policy documents		● Google+	
● News		● LinkedIn	
● Blogs		● Reddit	
● Twitter		● Faculty1000	
● Post-publication peer-reviews		● Q&A (stack overflow)	
● Facebook		● Youtube	
● Sina Weibo		● Pinterest	
● Wikipedia			

- The amount of each color identifies the ratio of attention sources

252   37   115

  - Number is the attention score

Altmetric.com. Access

- How one can get access to data?
  - ▶ Institutions. [Details here](#)
  - ▶ Researchers in scientometrics. [Details here](#)
- Types of access
  - ▶ Altmetric Explorer
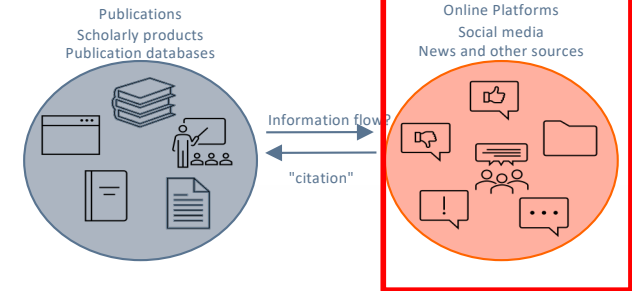 - ▶ Altmetric Details Page API
  - ▶ Full Access

PlumX

Effects of fructose vs glucose on regional cerebral blood flow in brain regions involved with appetite and reward pathways.
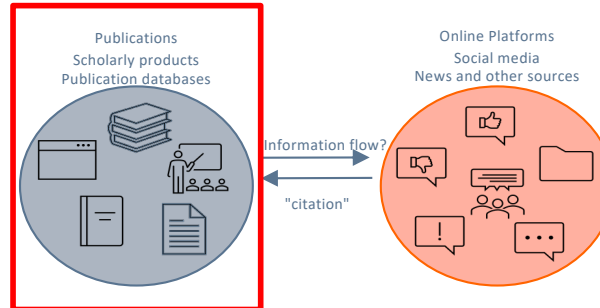
Citation data: JAMA, ISSN: 1538-3598, Vol: 309, Issue: 1, Page: 63-70
Publication Year: 2013

USAGE ^	954	CAPTURES ^	311	MENTIONS ^	32	SOCIAL MEDIA ^	1346	CITATIONS ^	96
Abstract Views ⊙	825	Readers ⊙	251	Comments ⊙	26	Shares, Likes & Comments ⊙	755	Citation Indexes ⊙	96
EBSCO	825	Mendeley	251	Reddit	13	Facebook	755	Scopus	96
				Reddit	12			CrossRef	91
Link-outs ⊙	76	Exports-Saves ⊙	60	Reddit	1	Scores ⊙	349	PubMed Central	53
EBSCO	76	EBSCO	60			Reddit	177		
				Blog Mentions ⊙	4	Reddit	171		
Clicks ⊙	52			Blog	3	Reddit	1		
Bitly	35			Blog	1				
Bitly	17					Tweets ⊙	193		
				Economics Blog Mentions ⊙	1	jama.jamanetwork.com	175		
HTML Views ⊙	1			Blog	1	www.ncbi.nlm.nih.gov	13		
EBSCO	1					www.ncbi.nlm.nih.gov	2		
				News Mentions ⊙	1	www.ncbi.nlm.nih.gov	2		
				News	1	www.ncbi.nlm.nih.gov	1		
						+1s ⊙	49		
						Google+	49		

| ARTICLE SUMMARY | TWEETS | BLOG MENTIONS | ECONOMICS BLOG MENTIONS | NEWS MENTIONS |

Publications
Scholarly products
Publication databases

Online Platforms
Social media
News and other sources

Information flow?
"citation"

Meet the Experts

PlumX. What is tracked

Source: https://plumanalytics.com/learn/about-artifacts/
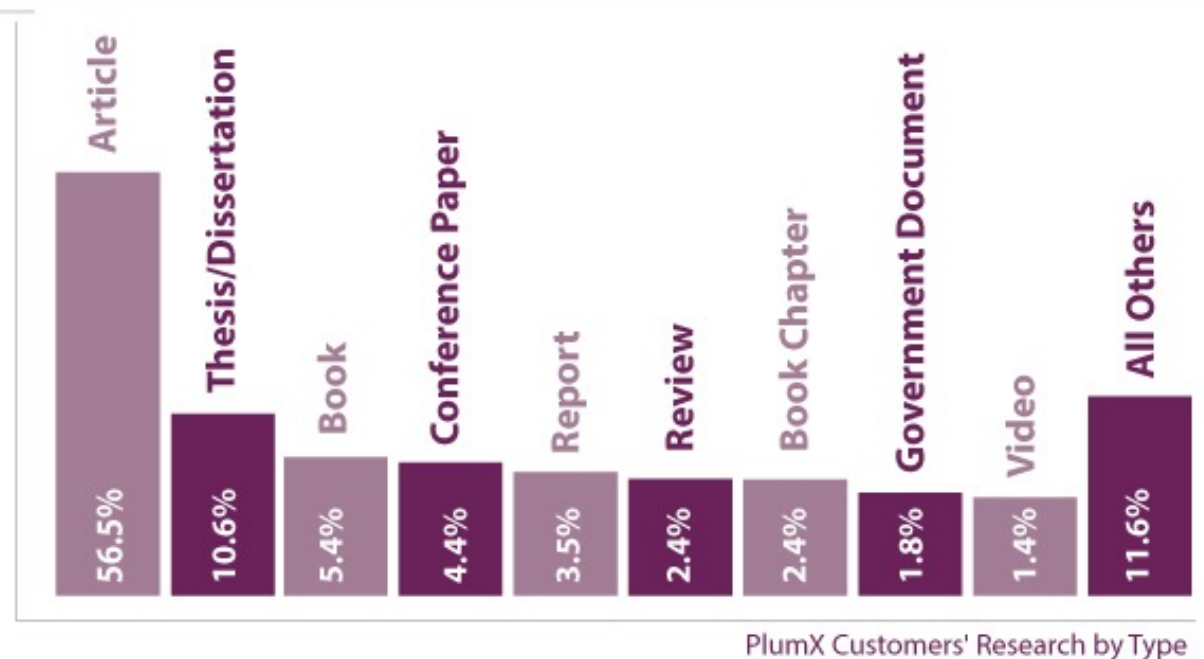
PlumX. Data access

Access:

- Artifact-level PlumX pages are free and publicly accessible; they provide access to all our article-level metrics
- API access: link

Challenges

- Intransparency of data mining
- Inconsistencies and variations across aggregators

readerships	Mendeley.com (n = 30,124)				Altmetric.com (n = 19,073)				Lagotto (n = 30,117)			
	overlapped	equal	>	<	overlapped	equal	>	<	overlapped	equal	>	<
Altmetric.Com (n = 19,073)	19,015	18,613	153	249								
%		97.9%	0.8%	1.3%								
Lagotto (n = 30,117)	30,117	14,416	13,974	1,727	19,012	7,977	9,823	1,212				
%		47.9%	46.4%	5.7%		42.0%	51.7%	6.4%				
Plum Analytics (n = 30,389)	30,089	9,027	10,531	10,531	19,057	5,120	6,974	6,963	30,086	7,676	7,815	14,595
%		30.0%	35.0%	35.0%		26.9%	36.6%	36.5%		25.5%	26.0%	48.5%

(Dis)agreement among aggregators in Mendeley readership counts (Zahedi & Costas, 2018)

- Altmetrics are dependent on aggregation time
  - Changes
  - Deleted content

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. PLOS ONE, 13(5), e0197326. https://doi.org/10.1371/journal.pone.0197326

Where to see my own altmetrics?

Dimensions - [Try it out!](#)

Conclusions

- Altmetrics offer an alternative view to scientific impact, including the general public and their interactions with scholarly publications.

- Altmetrics data enable new forms of Computational Social Science research, but they come with their own challenges when it comes to data collection and quality.

- Altmetrics may be relevant for individual researchers' day to day activities.

Thank you !

gesis Leibniz-Institut
für Sozialwissenschaften

Leibniz
Gemeinschaft

Expert Contact & GESIS Consulting

Contact: you can reach the speakers via e-mail:
katrin.weller@gesis.org   olga.zagovora@gesis.org

GESIS Consulting: GESIS offers individual consulting in a number of areas – including survey design & methodology, data archiving, digital behavioral data & computational social science – and across the research data cycle.

Please visit our website www.gesis.org for more detailed information on available services and terms.

More Services from GESIS

- Get materials for capacity building in computational social science and take advantage of our expanding expertise and resources in digital behavioral data.

- Use GESIS data services for finding data for secondary analysis and sharing your own data.

- Check out the GESIS blog "Growing Knowledge in the Social Sciences" for topics, methods and discussions from the GESIS cosmos – and beyond.

- Keep up with GESIS activities and subscribe to the monthly newsletter.

- Search | Search GESIS ▼   for publications, tools & services.

More from CSS Experts in the Series

June 24   Katrin Weller: **A Short Introduction to Computational Social Science and Digital Behavioral Data**

July 01   Fabian Flöck, Indira Sen: **Digital Traces of Human Behavior from Online Platforms – Research Designs and Error Sources**

July 08   Sebastian Stier, Johannes Breuer: **Combining Survey Data and Digital Behavioral Data**

Sept 16   Katrin Weller, Oliver Watteler: **Ethics and Data Protection in Social Media Research**

Sept 30   Roberto Ulloa: **Introduction to Online Data Acquisition**

Oct 07   Roberto Ulloa: **Auditing Algorithms: How Platform Technologies Shape our Digital Environment**

Oct 14   Marius Sältzer, Sebastian Stier: **The German Federal Election: Social Media Data for Scientific (Re-)Use**

Nov 04   Arnim Bleier: **Introduction to Text Mining**

Nov 11   Haiko Lietz: **Social Network Analysis with Digital Behavioral Data**

Dec 02   Olga Zagovora, Katrin Weller: **Altmetrics: Analyzing Academic Communications from Social Media Data**

Dec 16   Andreas Schmitz: **Online Dating: Data Types and Analytical Approaches**

Jan 13   Gizem Bacaksizlar: **Political Behavior and Influence in Online Networks**

Jan 27   David Brodesser: **SocioHub – A Collaboration Platform for the Social Sciences**

Feb 03   Regina Pfeifenberger, Wolfgang Otto: **Pollux – Literature and Research Tools for Political Scientists**