Wiki Workshop (13th edition) – March 25-26, 2026

The Five Actions of Popular Wikipedia Research

Steve Jankowski
University of Amsterdam

Abstract

Popular research about Wikipedia has existed for over 20 years. Through a bibliometric history of 50 heavily cited texts, this paper identifies five *actions* that give popular Wikipedia research its shape: modeling, optimizing, encountering, revealing, and contextualizing. By historically mapping these epistemological actions to citation clusters, I argue for reassessing how Wikipedia exists as a boundary object that shapes how the field responds to methodological concerns, such as how to study to context.

Keywords

Wikipedia research, bibliometric history, sociology of science, boundary objects, contexts

Introduction

In the past two years, there has been an increasing awareness by Wikimedia research to take "context" into account when Wikimedian data is being studied. A recent Wikimedia Foundation white paper sought to guide researchers to "[a]ppropriately contextualize any comparisons to Wikipedia content," and to "[a]ssess edits and editor actions in context" (Johnson et al., 2025). Even NPOV itself is explicitly considered to be "inherently contextual." In part, this acknowledgment highlights a tension that underpins the field's relationship to big data. On the one hand, researchers have benefited from the outcomes of treating Wikipedia as a "source of data" (Hill and Shaw, 2019). But, as danah boyd and Kate Crawford warned in 2012, "Big Data loses its meaning" when "[t]aken out of context" (danah boyd and Crawford, 2012). Far from being an emergent research problem, this focus on context is renews a longstanding, if not unresolvable, concern about what it means to study Wikipedia as a research object.

To explore this situation, I turn to sociologists of science and the concept of "boundary objects" which are "arrangement[s] that allow different groups to work together without consensus" (Star, 2010). While researchers study the same thing (like Wikipedia), the *actions* they enlist to study it provide the materiality for fundamentally different objects. I operationalize this social theory through the act of citing texts; where the accumulated significance of popular research accretes as disciplinary and interdisciplinary actions of what counts as legitimate forms of study. With this framework, I ask: How do bibliometric traces help us understand the history of Wikipedia as a boundary object that is translated through different inter/disciplinary actions? How are context and big data positioned in this history?

Methods

Building on previous systematic literature reviews (Mesgari et al., 2015) (Redi et al., 2021), especially Hill and Shaw (Hill and Shaw, 2019), I began by manually collecting data from 50 pages of Google Scholar (GS) result pages with the query "Wikipedia" while using a "clean research browser" (Rogers, 2013) p.111). After cleaning the collected pages using regex, I conducted a bibliometric analysis of 500 documents published between 2004–2025. With the goal of mapping popularly cited research, this sample was limited to texts with 100+ citations which were then coded by combining open and directed coding to discover fundamental "actions" that epistemologically connect these studies. To do this work I engaged in several rounds of iterative coding based on titles, journals, affiliations, and contents. Once these categories had at least associated 10 texts, I manually recorded each text's GS "Citations per year." Exploratory data collection was conducted in August 2025, and a new data set was collected in February 2026 with 50 texts.

Limitations: GS's data not only differs from Scopus or Web of Science, but they are subject to database errors. To address these issues I rounded all totals down to the nearest 50 when describing the data to reinforce these values as approximations. Likewise, the query of "Wikipedia" acts as a proxy for the field. Indeed, its simplicity could not collect research about other Wikimedia projects or consequential texts like *The Wealth of Networks*. This leads to the limited meaning of "popular" and "Wikipedia" research. Future bibliometric histories should involve expert and snowball sampling.

© Copyright held by the owner/author(s), published under Creative Commons CC BY 4.0 License

Results

Modeling Wikipedia (Figure 1: 20350 citations) creates new algorithms, tools, or means to assess, extract, or predict Wikipedian data. Wikipedia as an object is interpreted almost exclusively by the artificial Intelligence and natural language processing which framed it as a computational resource. The popular shape of this action shows a steady upward progression until 2014, peaks in 2022, and has quickly fallen back to 2014 levels.

Optimizing WIkipedia (Figure 2: 6100 citations) represents some of the earliest field-defining work. Here, human-computer interaction, computer-supported cooperative work, and social science researchers examined the collaboration of Wikipedians and the quality and reliability of their content, with the explicit goal of improving the encyclopedia or peer production in general. By 2014, the ten most cited works collectively dropped, and this trend has steadily continued since.

Encountering Wikipedia (Figure 3: 3400 citations) approaches Wikipedia as a set of contact points with external communities that assess the encyclopedia according to their own collective standards and commitments. This action describes and stages encounters between information gatherers and Wikipedia. It is one of the most interdisciplinary actions and followed a steep incline to 2012–2013, and then gradually receded.

Revealing Wikipedia (Figure 4: 3350 citations) involves describing and naming underlying norms, roles, routines, and decision-making mechanisms. In doing so, this action renders elusive processes and structures as observable and transformable. It also relies on the same disciplines as optimization but leans more towards media studies' critical tradition. The action itself peaked in 2012 and, after dropping the following year, has maintained a relatively consistent level.

Contextualizing Wikipedia (Figure 5: 3000 citations) studies the encyclopedia and its community through ethnography, history, and critical theory. While some of these studies focus on Wikipedia itself, it is more common to situate it within the fabric of society, culture, and politics. Unlike other actions, contextual research was dependent largely on a single text until 2011 and then quickly peaked in 2015 where it plateaued for 7 years, before dropping in 2022.

Conclusions

Viewing the actions together (Figure 6), the first 8 years saw overlapping citation values and represents a interdisciplinary period where Wikipedia was materialized as an object of peer production. But between 2009–2012, modeling began to outpace optimization, and by the end of 2013, modeling was the only action that retained its upward trajectory which resulted in the unambiguous bifurcation of popular citations. This development happened during the creation of Wikidata in 2012, a project designed to transform peer produced content into easily computable semantics (Iliadis, 2022). A decade later, modeling precipitously fell after 2022, an event that remarkably corresponds with the public release of ChatGPT. In light of these two events, the modeling action of computer science materialized Wikipedia as a disciplinary semantic extraction object.

In view of these bibliometrics, I argue that the history of popular Wikipedia research has been ordered by the inter/disciplinary actions that materialized two research objects. I suggest that this history provides us with a map to reflect on the course of popular Wikipedia research and the citational declines of both objects suggest that it is time for researchers to experiment with alternative interpretations of Wikipedia.

References

[danah boyd and Crawford2012] danah boyd and Kate Crawford. 2012. Critical questions for big data. *Information, Communication & Society*, 15(5):662–679.

[Hill and Shaw2019] Benjamin Mako Hill and Aaron Shaw. 2019. The most important laboratory for social scientific and computing research in history. *Wikipedia @ 20*, 9. https://wikipedia20.pubpub.org/pub/laboratory.

[Iliadis2022] Andrew Iliadis. 2022. *Semantic media: Mapping Meaning on the Internet*. Polity Press, Oxford, England, November.

[Johnson et al.2025] Isaac Johnson, Yu-Ming Liou, Jacob Rogers, Aaron Shaw, and Leila Zia. 2025. Recommended practices for npov research on wikipedia. *arXiv*.

[Mesgari et al.2015] Mostafa Mesgari, Chitu Okoli, Mohamad Mehdi, Finn Årup Nielsen, and Arto Lanamäki. 2015. "the sum of all human knowledge": A systematic review of scholarly research on the content of w ikipedia. *Journal of the Association for Information Science and Technology*, 66(2):219–245.

[Redi et al.2021] Miriam Redi, Martin Gerlach, Isaac Johnson, Jonathan Morgan, and Leila Zia. 2021. A taxonomy of knowledge gaps for wikimedia projects (second draft). page https://arxiv.org/abs/2008.12314.

[Rogers2013] Richard Rogers, 2013. *Digital methods*, chapter Search as Research, page 111. MIT Press.

[Star2010] Susan Leigh Star. 2010. This is not a boundary object: Reflections on the origin of a concept. *Science, technology, & human values*, 35(5):601–617.

© Copyright held by the owner/author(s), published under Creative Commons CC BY 4.0 License

Modeling Wikipedia

Yearly Citations (2005–2025)

These top ten most cited documents of the "modeling" research action are grounded in Natural Language Processing and Artificial Intelligence; as well as Social Science.

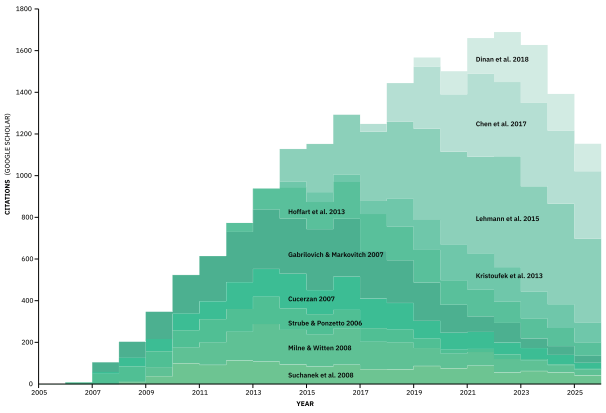

Figure 1

Optimizing Wikipedia

Yearly Citations (2005–2025)

These top ten cited documents of the "optimization" research action are grounded in Human-computer interaction, Computer-Supported Cooperative Work, Social Science and Information Science.

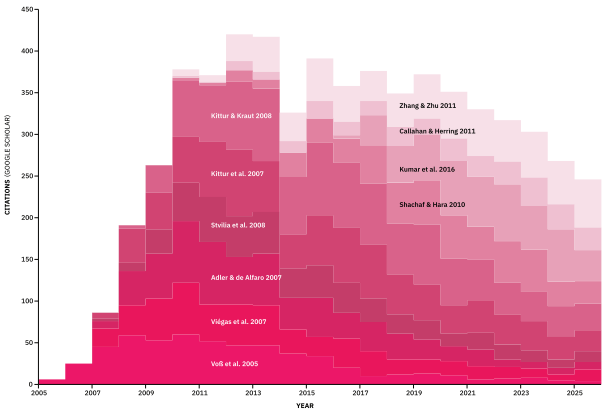

Figure 2

Encountering Wikipedia

Yearly Citations (2005–2025)

These top ten most cited documents of the "encountering" research action are grounded in Information Science, Media Studies, Medicine, Computer Science, Economics and Education.

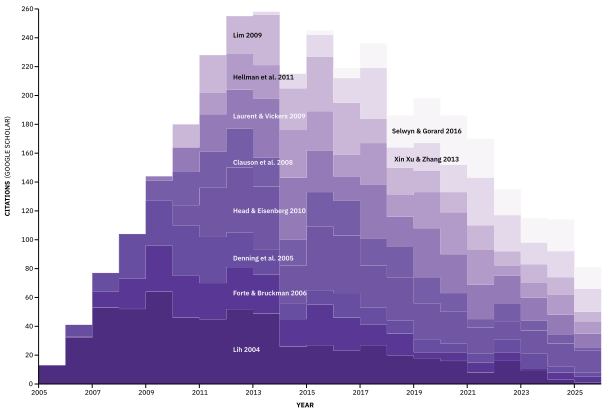

Figure 3

Revealing Wikipedia

Yearly Citations (2005–2025)

These top ten most cited documents of the "revealing" research action are grounded in Social Science, Human-computer Interaction, Computer-Supported Cooperative Work; and Media Studies.

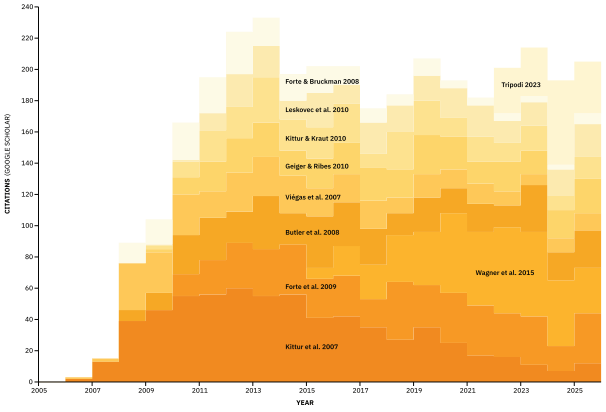

Figure 4

Contextualizing Wikipedia

Yearly Citations (2005–2025)

These top ten most cited documents of the "contextualizing" research action are grounded in Media Studies, History, Social Science, Information Science and Education.

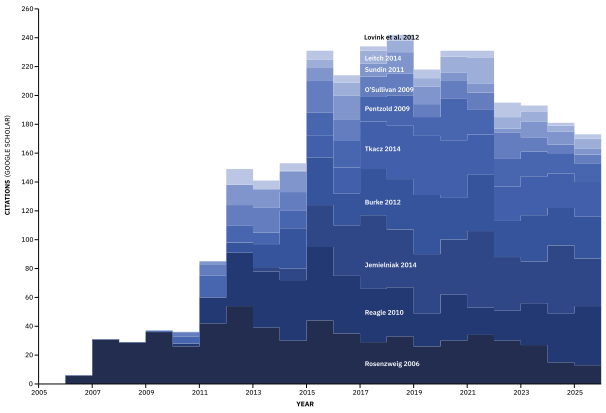

Figure 5

Popular Wikipedia Research

Yearly citations for five major research actions

Aggregrated Google Scholar citations for the top ten most cited documents per research action. 2009–2013 represents the split of a common ascent which materializes the peer production and semantic extraction boundary objects.

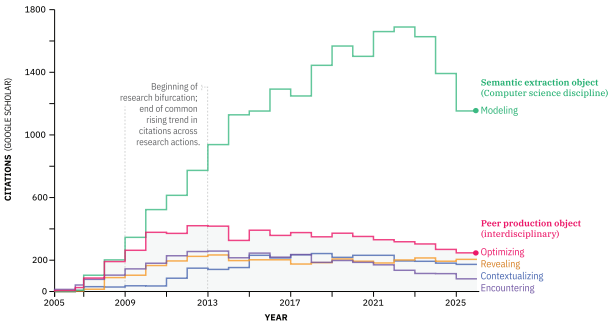

Figure 6

© Copyright held by the owner/author(s), published under Creative Commons CC BY 4.0 License