

Popularity, Quality, And Protection in Wikipedia’s Most Read Articles

Renita Murimi
University of Dallas

Abstract

Popularity, whether in online or offline environments, invites heightened scrutiny, even though popularity is not always an indicator of quality. Wikipedia articles are no exception to this phenomenon. This paper investigates the relationship between article popularity, quality, and protection using Wikipedia’s annual lists of most-read articles for over a decade. Our findings showed that article popularity is an outcome of real-world events rather than article quality, and popularity is a significant indicator of the need for protection due to heightened cultural discourse about the content of the article.

Introduction

As the world’s largest open encyclopedia, Wikipedia serves a primary role for information gathering. Although at first glance information-gathering seems episodic, there are some topics that transcend brief episodes of popularity and persist over longer periods of time. But moments of brief or sustained cultural attention are not enough to keep readers interested in a Wikipedia article. The quality of articles also plays a role in sustaining interest (Piccardi et al, 2023).

The quality of Wikipedia articles is an outcome of its open-source, volunteer-driven, editorial approach to article creation and maintenance. This also means that certain popular Wikipedia articles invite conflicting viewpoints, which can be exacerbated by local and global events. One way that Wikipedia manages conflicts in popular articles is by placing such articles under *protection*, which refers to limiting edits to only confirmed users (a registered editor account that is at least 30 days old) with an established editing history (500 or more edits). Thus, compared to non-protected articles, protected articles involve increased “frictions” due to their activity, topic area, and visibility (Ajmani et al., 2023). Still, these protections have been shown to be helpful in maintaining the article quality. A study of Wikipedia protections found that high quality articles suffer from quality degradation if left unprotected (Ruperchter, 2025).

This study examines the relationship between article quality, popularity, and protection status on Wikipedia. Specifically, the research in this study is centered on the following questions: Is the popularity of a Wikipedia article an indicator of the need for protection? Is the popularity of an

article linked to its quality? I explored these themes using Wikipedia’s annual list of its most-read articles, which has been published annually since 2015.

Methods

For the period from 2015-2025, this study used a dataset of the 20 most-read articles from the annual top-50 list. For each article in the 11-year observation window, I collected its rank, article class, number of article views, and protection status. Article class was used as a proxy for article quality, and article views were used as a proxy for popularity. In Wikipedia’s editorial system, the article class is a rating of its language and completeness, and in descending order is represented by the following labels: *Featured*, *A*, *Good*, *B*, *C*, *Start*, and *Stub*. The dataset for this study had a total of 220 Wikipedia articles with 87 class B articles (39%), 13 class C articles (5.9%), 11 *featured* articles (5%), 75 *good* articles (34%), and the rest were *unclassified*. Unclassified articles were labeled as such due to a variety of reasons such as their list-like structure (for example: article titled *Deaths in 2015*), one-off occurrences such as the *2024 AFC Asian Cup*, or due to their sub-topic classification (such as the *ChatGPT app* within the *ChatGPT* article). To determine persistence in article popularity, I compared the yearly lists to find the articles that appeared more than once. For example, some topics (e.g. United States) were observed repeatedly, while others (India, Covid-19) appeared in bursts.

To determine persistent popularity, I defined a *normalized persistence score* (N) as the ratio of the number of years an article appears in the top-20 list to the temporal span between its first and last appearance, which captures the continuity of public attention while controlling for observation window length. I also defined a *global persistence score* (G) that measures the extent to which an article maintains visibility across the entire observation period, independent of when that visibility occurs. The global persistence score (G) is defined as the proportion of years in which an article appears in the top-20 most-read list relative to the total number of years observed. Thus, if an article appears in A out of Y observed years, and S is the number of years between first and last appearance (inclusive), its global persistence score is A/Y and its normalized persistence score is A/S .

Results

Table 1 shows the persistence scores. We see that the most persistent articles (*United States*, *Donald Trump*) are B-class

articles. We also see that featured quality does not guarantee high persistence (Taylor Swift, Elizabeth II). Regarding protection, we see that protection tracks persistence, not quality. All protected articles are high-frequency, and no low-frequency article is protected. This suggests that protection is reactive to attention, not editorial quality. For comparison of quality, I compiled a list of 20 random featured articles and their article views (to capture popularity) using Wikipedia’s *Read a Random Featured Article* tool (see Table 2). A comparison of Tables 1 and Table 2 shows that article quality does not indicate either protection status or popularity. Across the 11-year period examined, persistent top-read Wikipedia articles exhibited marked heterogeneity in both duration and temporal structure. The *United States* article showed the highest persistence, appearing in the top-20 list in 10 of 11 years, followed by *Donald Trump* (8 years) and *Elizabeth II* and *Elon Musk* (7 years each). Temporal inspection further revealed distinct persistence modes, including continuous multi-year presence (e.g., *Elizabeth II*: 2016–2022), intermittent recurrence across non-consecutive years (e.g., *India*: 2015, 2023–24), and event-driven bursts (e.g., *COVID-19*-related articles in 2020–21).

Based on the combined use of global and normalized persistence, I identify three dominant typologies of popularity of top-read Wikipedia articles as shown in Table 3. *Structural* persistence ($G \geq 0.6$) captures articles with sustained visibility across most of the observation window, reflecting enduring relevance. *Recurrent* persistence ($0.25 \leq G \leq 0.59$) characterizes articles that reappear intermittently over time, often corresponding to cyclical cultural interest or recurring events. In contrast, *bursty* persistence ($G < 0.25$ and $N \geq 0.75$) describes articles with temporally concentrated but continuous popularity within short spans, typically associated with singular events or shocks.

Discussion

The results of this study show that editorial quality classifications were only weakly aligned with persistence. In contrast, article protection status showed a clearer association with persistence, suggesting that protection is primarily a reactive response to sustained public attention rather than a function of editorial quality. While N can help identify whether the limited appearances were continuous bursts or isolated events, G measures the overall share of the observation window where an article appears. A low G indicates that an article has appeared only a few times. But those appearances may have different temporal patterns. For example, articles that appear only in consecutive years indicate short, intense attention. Others appear sporadically in distant years, indicating intermittent interest. Both of these are examples of low G . This ambiguity is resolved by N , since it measures continuity within an article’s active window. A high N implies that the article appeared consistently during the period within which it was relevant. Thus, articles with a

low G and high N indicate bursty attention. In contrast, articles with low G and low N indicate weak recurrence.

This study has several limitations. First, the analysis is based on a limited dataset comprising only the 20 most-read Wikipedia articles based on Wikipedia’s annual list of fifty most-read articles in a given year. Second, article quality classifications were treated as fixed and ordinal, although quality assessments can change over time and may lag behind shifts in readership or editorial activity. Third, protection status was analyzed as a binary indicator, which does not capture variation in the duration, intensity, or previous stints as protected articles. Finally, this study only looks at the of frequency of appearance, which does not distinguish between continuous persistence and recurrent re-emergence.

Future work could extend this analysis by incorporating higher-resolution temporal article view data to distinguish sustained attention from episodic bursts of readership. Further, an examination of changes in article quality assessments and granular protection status over time would allow for a more precise examination of how editorial interventions interact with public attention dynamics. Expanding the analysis beyond the top-20 most-read articles, or conducting cross-lingual comparisons across Wikipedia editions, would further assess the generalizability of these findings across platforms and cultural contexts.

Conclusions

Much like trending topics on various search engines and social networking sites, Wikipedia offers a reflection of cultural attention toward certain topics. Further research in this area can help determine how platform governance, public attention, and editorial quality interact over time. Additionally, the metrics in this study can be applied to other Wikipedia languages, enabling comparative studies of how collective attention, content quality, and moderation interact in diverse online knowledge systems.

References

- Tiziano Piccardi, Martin Gerlach, Akhil Arora, and Robert West. 2023. A large-scale characterization of how readers browse Wikipedia. *ACM Transactions on the Web*, 17 (2): 1-22.
- Leah Ajmani, Nicholas Vincent, and Stevie Chancellor. 2023. Peer produced friction: how page protection on Wikipedia affects editor engagement and concentration. *Proceedings of the ACM on HCI*, 7 (2): 1-33.
- Thorsten Rupprechter, Manoel Horta Ribeiro, Robert West, and Denis Helic. 2025. Protection from evil and good: The differential effects of page protection on Wikipedia article quality. In *Proceedings of the International AAAI Conference on Web and Social Media*, 19, 1704-1718.

Article	Frequency (A)	Span (S)	Normalized persistence (N)	Global persistence (G)	# Entries in Protection Log
United States	10	11	0.909	0.909	500
Donald Trump	8	11	0.727	0.727	500
Elizabeth II	7	7	1	0.636	14
Elon Musk	7	8	0.875	0.636	17
Cristiano Ronaldo	5	8	0.625	0.455	50
Game of Thrones	4	5	0.8	0.364	14
Billie Eilish	4	6	0.667	0.364	10
Taylor Swift	3	4	0.75	0.273	25
Chadwick Boseman	3	4	0.75	0.273	11
Squid Game	3	4	0.75	0.273	11
King Charles III	3	3	1	0.273	5
Joe Biden	3	5	0.6	0.273	28
India	3	10	0.3	0.273	500
Kamala Harris	2	5	0.4	0.182	11
Stephen Hawking	2	4	0.5	0.182	20
Jeffrey Dahmer	2	4	0.5	0.182	10
Freddie Mercury	2	2	1	0.182	12
Pablo Escobar	2	2	1	0.182	6
Avengers: Infinity War	2	2	1	0.182	16
List of Marvel Cinematic Universe films	2	2	1	0.182	16
YouTube	2	2	1	0.182	17
Covid-19 pandemic	2	2	1	0.182	2
Covid-19 pandemic in the United States	2	2	1	0.182	1

Table 1: Normalized and global persistence scores for most persistently popular Wikipedia articles since 2015

Featured Article	# Views	Current Protection status	# Entries in protection log
No. 1 Squadron RAAF	955	None	1
James E Boyd	4396	None	1
Mitt Romney	1928115	None	14
American Arts Commemorative Series medallions	3453	None	1
2012 Tour De France	41716	None	1
Xenu	15324	None	17
Women's poll tax repeal movement	3202	None	1
Battle of Quebec (1775)	74105	None	1
Metallica	2548228	None	21
<i>Blood on the Floor</i> (Turnage)	35790	None	2
ForksVille Covered Bridge	723	None	1
Turf Moor	61228	None	6
Peru national football team	290284	None	4
Dance in the Dark	31268	None	10
Alice in Chains	1376083	None	4
<i>Killer Instinct Gold</i>	32994	None	1
Northern gannet	73141	None	0
<i>Oryzomys gorgasi</i>	937	None	2
1982 British Army Gazelle friendly fire incident	4933	None	2
1920–21 Gillingham F.C. season	296	None	0

Table 2: Article views and protection status for 20 random featured articles

Typology	Criteria	Articles
Structural persistence	$G \geq 0.60$	United States; Donald Trump; Elizabeth II; Elon Musk
Recurrent persistence	$0.25 \leq G \leq 0.59$	Cristiano Ronaldo; Game of Thrones; Billie Eilish; Taylor Swift; Chadwick Boseman; Joe Biden; India; Squid Game; King Charles III
Bursty persistence	$G < 0.25$ and $N \geq 0.75$	Freddie Mercury; Pablo Escobar; Avengers: Infinity War; List of Marvel Cinematic Universe films; YouTube; COVID-19 pandemic; COVID-19 pandemic in the United States
Low intermittent	$G < 0.25$ and $N < 0.75$	Kamala Harris; Stephen Hawking; Jeffrey Dahmer

Table 3: Typology of attention among top-read articles