depths of wikipedia: Understanding cross-platform online attention, content creation, and success

Patrick Gildersleve
Department of Methodology, London School of Economics and Political Science

Keywords: Collective attention, Online navigation, Cross-platform, Novelty, User-generated content

Introduction

"depths of wikipedia" is a popular social media account dedicated to sharing unusual or interesting content from Wikipedia, with wide online following across Twitter, Instagram, and TikTok (760K, 1.2M, 163K followers respectively). In highlighting Wikipedia articles elsewhere on social media, depths of wikipedia can act as a significant driver of traffic to otherwise unexplored areas of the online encyclopaedia. It is thus interesting to consider how a single account, mediated by the algorithmic affordances of a large social media platform, may shape collective attention and influence content elsewhere on the Web. The nature of depths of wikipedia, in posting unusual articles unrelated to current events, makes it an appealing case study and natural experiment for studying cross-platform collective attention without the interference of external factors.

Frequently, studies of collective attention online are single platform, and/or do not consider how users navigate between different websites, or the effects of this interaction. Related works have looked at why general audiences read Wikipedia (Singer et al., 2017), modelled peaks of attention with associated editing activity in response to unexpected and planned events (Keegan et al., 2013; Kobayashi et al., 2021), studied the effects of exogenous content contributions such as editing campaigns towards unpopular articles (Zhu et al., 2020), examined different social platforms (including Wikipedia) in their combined response to exogenous events (Osborne et al., 2012), or explored wider patterns of cross-platform linking (Meier, 2022). However, work on explicit cross-platform navigation and its effects is sparing.

The aims of this study are threefold. Firstly, to understand how online content may be considered unusual and how the depths of wikipedia account is able to identify novel content. Secondly, to explore the effects of substantial inflows of traffic to typically unpopular Wikipedia pages, in terms of both page views dynamics and edits to Wikipedia content. Finally, to model how unusual content on Wikipedia may be popular elsewhere on the Internet by the response (likes/retweets) on Twitter.

Data

To study the dynamics and effects of cross-platform collective attention, we collect tweets from July 2021 to December 2022 from the depths of wikipedia account (example in figure 1), taking tweet text and image content (transcribing images where appropriate), and data on likes and retweets. The depths of wikipedia Twitter account is selected over the other social media profiles since direct links to the appropriate Wikipedia article can be and are included in each post. We also use information from Wikipedia on article content, page views, hyperlink network structure, edits, and editors for the articles that are featured. After processing, the primary dataset consists of 1027 threads, each with an associated tweet image and unique Wikipedia article. In our work, we compare these tweets and articles against a sample of all public tweets with a linked Wikipedia article (N=1000), articles tweeted by the official Wikipedia account (@Wikipedia) (N=470), and a random sample of Wikipedia articles (N=1000) from the same time period. The study is also informed by interview responses from the owner of the depths of wikipedia account.

Methods

Initially, we attempt to quantify "depth", or how unusual/unexpected a post/article is, in several ways. Firstly, using the text of the featured image from Twitter and the corresponding Wikipedia article. We take normalised term frequency scores of the image text relative to the article, the image text relative to the rest of Wikipedia, and the article compared to the rest of Wikipedia—its "textual depth". Cases where unusual terms are used more frequently, as measured by these scores, are indicative of "deeper" content. Secondly "attentional depth", using average hourly page views before the article is featured—less popular articles being "deeper". Finally by measuring an article's PageRank centrality in the article hyperlink network before being featured. In this case, we consider "structural depth" using the raw hyperlink network, as well as "navigational depth" with the clickstream network featuring how frequently links are clicked. Articles with lower centrality across these measures—less important and less integrated into the network—may be considered "deeper". We compare these depth measures

against articles tweeted by @Wikipedia, posted by other accounts on Twitter, and a random selection of articles.

To consider cross-platform attention, we turn to the dynamics of time series on page views and edits for articles featured by depths of wikipedia (visualised in Figure 2), @Wikipedia, and the rest of Twitter. We measure and model the time series peaks, accounting for any periodic variation (e.g. through anniversaries of any famous events) to quantify the peak additional views, and total excess views from the article being posted on Twitter in the following 7 days. To examine editor effects, we collect and count article revisions, also identifying logged-in editors whose first ever edit is towards an article after it is featured on Twitter. Finally, we also produce models for Tweet likes and retweets, with the aforementioned depth measures as explanatory variables. In each case we perform the analysis with negative binomial regressions.

Results

Through the analyses on depth, cross-platform attention, and success, we find that:

1. Wikipedia articles featured by depths of wikipedia are, as measured by the depth factors, more unusual than other Wikipedia articles posted on Twitter. However, they are less unusual than a typical Wikipedia article. Select further details in table 1.

2. An article being featured by depths of wikipedia typically experiences an increase in page views 25x larger than other articles posted on Twitter (7x larger than @Wikipedia), on average equivalent to 1222 views ($p < 0.001$). Total excess views increases by a factor of 9.3x (3.2x compared to @Wikipedia), equivalent to 10,993 additional views ($p < 0.001$).

3. Daily edits increase by a factor of 2.26x (typically 1 edit) ($p < 0.001$) when an article is posted by depths of wikipedia. Moreover, depths of wikipedia has attracted an estimated 493 new account-holding editors who have made a further 15,241 edits.

4. Article features are weakly associated with that article proving popular when posted to Twitter.

Discussion

Overall, the consequences of these findings suggest that social media accounts like depths of wikipedia can have a significant impact on the visibility and development of online content. The depths of wikipedia account successfully identifies unusual articles for social media, though there are evidently many more candidates available in the online encyclopaedia. Alternative avenues, such as meme accounts, can be significant drivers of attention, in the form of page views and edits / (new) editors, to Wikipedia. They can also be particularly effective in targeting less popular pages as compared to typical sources such as official channels, current events, or popular entertainment. Though one must also be conscious of the risks of attracting this attention, and perhaps low quality or bad faith edits that could follow. Finally, whilst initial results on modelling article popularity on social media are mixed, being able to predict this popularity of content from Wikipedia may help more people discover relevant informative articles, and can assist in finding new editing communities for neglected articles.

Conclusion

Results from this work contribute to understanding how Internet users navigate between popular platforms, as well as the effects of sudden targeted influxes of users to particular sites. The work also offers wider insights for online content creation in computationally identifying interesting content and optimising clickthrough rates. Furthermore, there are clear practical applications in helping to understand how Wikipedia attracts users and edits for the future promotion and development of the site.

References

[Keegan et al.2013] Brian Keegan, Darren Gergle, and Noshir Contractor. 2013. Hot off the wiki: Structures and dynamics of Wikipedia's coverage of breaking news events. *American Behavioral Scientist*, 57(5):595–622.

[Kobayashi et al.2021] Ryota Kobayashi, Patrick Gildersleve, Takeaki Uno, and Renaud Lambiotte. 2021. Modeling collective anticipation and response on Wikipedia. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 315–326.

[Meier2022] Florian Meier. 2022. TWikiL–the Twitter Wikipedia link dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1292–1301.

[Osborne et al.2012] Miles Osborne, Saša Petrovic, Richard McCreadie, Craig Macdonald, and Iadh Ounis. 2012. Bieber no more: First story detection using Twitter and Wikipedia. In *SIGIR 2012 Workshop on Time-aware Information Access*, pages 16–76. Citeseer.

[Singer et al.2017] Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. 2017. Why we read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1591–1600.

[Zhu et al.2020] Kai Zhu, Dylan Walker, and Lev Muchnik. 2020. Content growth and attention contagion in information networks: Addressing information poverty on wikipedia. *Information Systems Research*, 31(2):491–509.

Figure 1: An example of a tweet from depths of wikipedia with linked article.

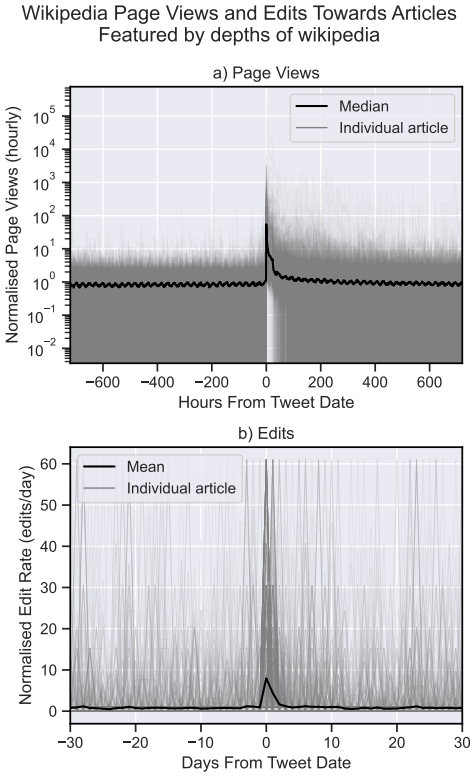

Figure 2: Spikes in both page views and editing activity are observed when an article is featured by depths of wikipedia. Page views and edits are normalised to average activity in the 30 days before featuring.

Source	Textual depth (higher is "deeper")	Attentional depth (lower is "deeper")	Structural depth (lower is "deeper")	Navigational depth (lower is "deeper")
@depthsofwiki	**0.017**	**20.5**	$\mathbf{1.1 \times 10^{-6}}$	$\mathbf{2.3 \times 10^{-6}}$
@Wikipedia	0.009^{*}	83.3^{***}	$2.4 \times 10^{-6***}$	$6.4 \times 10^{-6***}$
All Tweets	0.015	29.2^{*}	$1.7 \times 10^{-6***}$	$3.4 \times 10^{-6***}$
Random Articles	0.069^{***}	0.8^{***}	$3.2 \times 10^{-7***}$	$2.9 \times 10^{-7***}$

Table 1: Mean "depth" score across different measures. Values for @Wikipedia, All Tweets, and Random Articles are compared against @depthsofwiki with Welch's t-tests. Wikipedia articles posted by depths of wikipedia are more unusual than other articles posted on Twitter on all measures except textual depth. However, randomly selected Wikipedia articles are more unusual than those posted by depths of wikipedia. * $p < 0.05$; ** $p < 0.01$; ***$p < 0.001$.