

Wikipedia’s Polarization on Reddit

Keywords: community embeddings, Wikipedia, polarization, culture wars, Reddit

Extended Abstract

Introduction. Part of Wikipedia’s modus operandi is being a neutral source of information that presents ideas in a non-partisan, objective light. However, it has recently come under increased criticism from segments of the political right and left. The former who claim Wikipedia is composed of “leftist propaganda” and the latter who underline the potential ideological biases present in the platform. This form of negative language surrounding Wikipedia’s content has become commonplace in various communities, and has even led Wikipedia’s co-founder Larry Sanger to publicly state that “Nobody should trust Wikipedia”.¹ The current study investigates the evolution of trust in Wikipedia among partisan groups. We accomplish this by identifying pairs of communities with largely differing partisan affiliations and analyzing the changes in their respective usage of Wikipedia over time. Furthermore, we supplement this investigation by examining the kinds of Wikipedia articles that are frequently referenced by the left and right-leaning communities.

Data and methodology. We use a dataset consisting of all Reddit comments from 2006 to June 2022. We focus on the subset of Reddit comments that contain an invocation (i.e. a link) to Wikipedia articles. We then process the Wikipedia links through two steps. First we align each link with the Wikipedia article associated with it at the time of invocation. This was typically used for pages that were “renamed” or “merged”. Then we measure the partisan scores of the Wikipedia articles using the method defined by Shi et al. in [1]. Finally, we augment the Reddit communities with their vector embeddings defined by Waller and Anderson, alongside the social dimensions [2]. Social dimensions are vectors in the Reddit embedding that capture inherently social concepts like political orientation, gender, and age. These social dimensions are a consequence of the surprising properties of the Reddit embedding that permits vector algebra (r/Toronto is to r/TorontoMapleLeafs as r/Boston is to r/BostonBruins). In turn, by taking two communities that differ along partisan lines (say r/Democrats and r/conservative) we can define a partisan axis (see ref. for a clear explanation of construction and validation).

To find communities that differ along the partisan dimension, we define a set of left-leaning and right-leaning subreddits based on their projection on the partisan dimension, X_l , and X_r , respectively. Then for each left-leaning community we find its cross-partisan analogue. Explicitly, for each pair communities we measure their partisan distance as

$$d_p(\mathbf{c}_1, \mathbf{c}_2) = \cos(\mathbf{c}_1 - \mathbf{c}_2, \text{partisan}_{vec})$$

Afterwards, we run the Hungarian Algorithm to match the bipartite graph, attempting to maximize the total similarity. The final set of left-right leaning pairs is shown in Table 1. We observe that qualitatively the communities appear to differ along partisan-lines (e.g., r/climate → r/climateskeptives, r/Feminism → r/MensRights, r/AntiTrumpAlliance → r/trump).

Results. Using these pairs of communities we next explore how the right and left have relied upon Wikipedia differently. First we take a long term view by examining how the invocations

¹<https://tinyurl.com/2mnvp5nw>

Left-leaning	Right-leaning
PoliticalHumor	Conservative
EnoughLibertarianSpam	Shitstatistssay
GunsAreCool	gunpolitics
Liberal	Republican
excatholic	Catholicism
exchristian	TrueChristian
forwardsfromgrandma	TumblrInAction
Feminism	MensRights
Anarchism	Anarcho.Capitalism
racism	islam
skeptic	climateskeptics

Table 1: Matched communities used for studying how Wikipedia is invoked on Reddit. Differences in communities are mostly explained by their partisan association.

have evolved over the last 10 years. Next we connect the partisan leanings of the articles themselves with the communities that share them. In turn, we will see if the left invokes the left, and if the right invokes the right.

Temporal analysis. Figure 3 displays the percentage of links shared on Reddit that point to Wikipedia. We observe a general decline in the proportion of links to Wikipedia across all communities studied, with the largest drop occurring during Reddit’s rapid expansion in 2014. Before 2016, left and right subreddits exhibited a mirrored pattern in Wikipedia invocations. However, starting in 2016 and accelerating in 2018, we observed a clear divergence in linking behavior between the left and the right. This divide persisted through the COVID-19 pandemic, with the three months with the greatest ratio difference in invocations being September 2020 (6.7%), October 2021 (6.6%), and September 2019 (6.4%). These findings suggest that the use of Wikipedia on Reddit is becoming increasingly politicized, with potential implications for the accuracy and objectivity of information shared on the platform.

Who shares what. Next, we aimed to investigate the nature of shared content by determining the partisan scores of individual articles, using a method similar to the one proposed by Shi et al. [1]. Editor partisan scores, derived from their contributions to liberal or conservative articles, were mapped onto the articles. Our results, depicted in Figure 2, show a substantial surge in the number of right-edited articles shared on Reddit from 2016 onwards. Moreover, we discovered that left-leaning subreddits are more inclined to share right-edited articles. To explore this observation further, we present Figure 3, which demonstrates the difference in histograms of the number of articles shared by left- and right-leaning subreddits. The y-axis reflects the difference in the likelihood of sharing an article with a particular partisan score (x-axis). Our analysis suggests that the right tends to share articles edited by both the left and the right, while the left tends to share articles that lie on the extremes.

Discussion. In this work we illustrate the partisan differences that are emerging around Wikipedia. Through a study of how Wikipedia is invoked on Reddit we showcase a dramatic drop in its linking on the right. While this analysis is not causal, it does provide an initial account of how Wikipedia itself may be being politicized.

References

- [1] J. Becker, E. Porter, and D. Centola, “The wisdom of partisan crowds,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 22, pp. 10 717–10 722, 2019.
- [2] I. Waller and A. Anderson, “Quantifying social organization and political polarization in online platforms,” *Nature*, vol. 600, no. 7888, pp. 264–268, 2021.

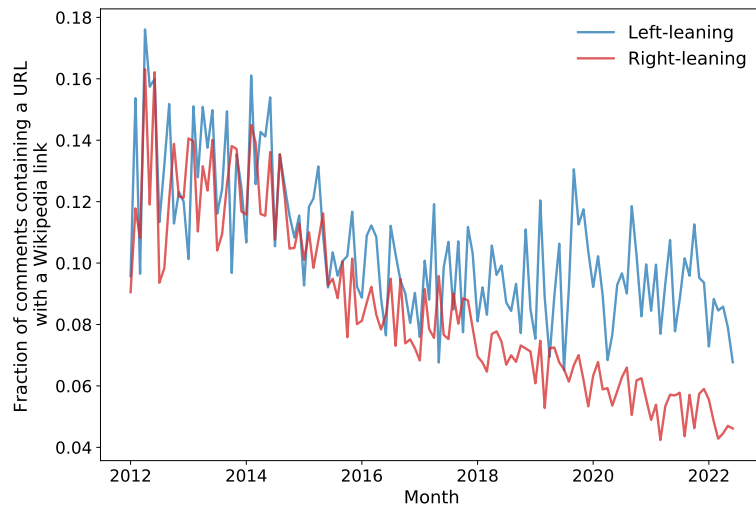


Figure 1: Proportion of URLs shared in a month that link to Wikipedia broken down into left-leaning and right-leaning subreddits.

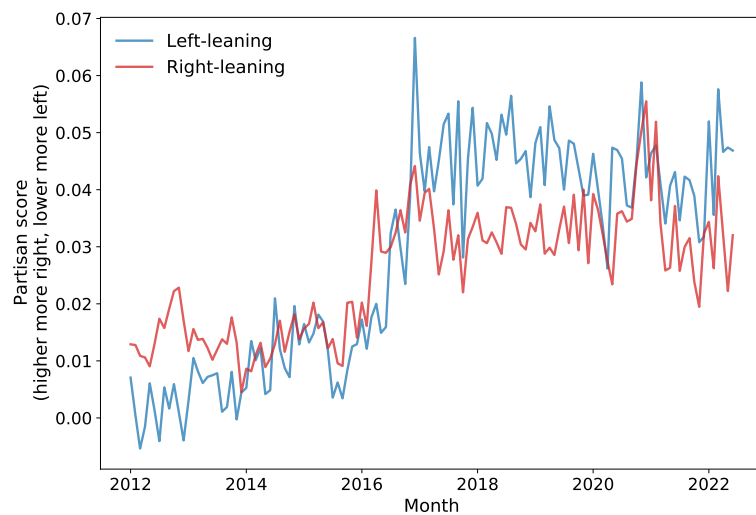


Figure 2: Average partisan score of the Wikipedia articles shared by the left and right.

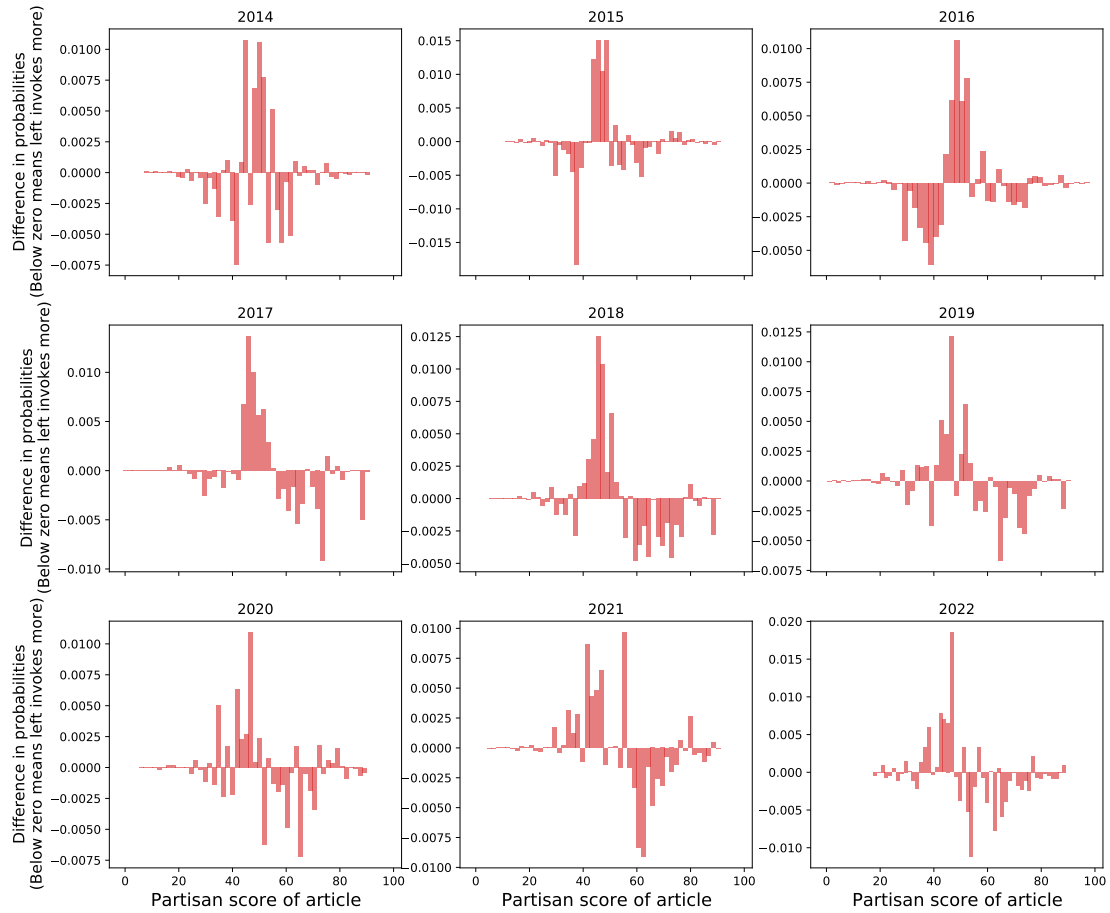


Figure 3: Difference in histograms between left and right leaning communities. The x-axis represents the partisan score of the article (higher more right, lower more left) and the y-axis showcases the difference in densities. We observe that as time progressed, the left began to invoke increasingly right-wing content.