9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Implicit Polarization on YouTube

Keywords: social media, polarization, extremism, word embeddings, timeseries analysis

Extended Abstract

How information spreads on social media and what individual characteristics or behaviors drive migration to radical echo chambers is a topic of much debate in political science. As the most widely used platform in the US with the most representative user base [1], and a host for embedded content across the web, the wide ranging impact of YouTube makes it a prime platform for understanding this phenomenon. Though researchers have addressed questions of polarization and extremism through a combination of surveys and interventions, these as well as large scale explorations of behavior over time focus on desktop data and access to a set of pre-labelled channels. This paper uses novel longitudinal data on individuals' YouTube consumption from mobile data provider mFour to consider how patterns of user viewing behavior, the networks of YouTube channels they watch, and their demographics influence viewing over time. In particular, we consider whether an individual's history of watching YouTube videos with implicit messages regarding polarizing topics, as found using content analysis of video transcripts, increases the likelihood that they will watch more videos of this type in the future.

YouTube is a key player on the web due to its use as both an on-site content provider and storage provider for videos embedded elsewhere. Similar to Wikipedia [2], this connectivity throughout other social networking sites enables it to play a vital role online. In acknowledgement of this influence, a variety of research has considered the role YouTube plays in the spread of pernicious ideologies including extremism and election misinformation. Longitudinal studies of user behavior have shown mixed results in understanding individual's trajectories towards more extreme content. In a carefully outlined study, [3] isolate the role of the YouTube algorithm from the role of individual preference, and find that individuals more skeptical of the 2020 election were more likely to be recommended content discussing narratives about election legitimacy. Similarly, a study by ADL found that extremist and white supremacist content remains prevalent on YouTube despite policy changes, and individuals who engaged with this content were more likely to get recommended and then choose to watch similar videos [4]. One of the original papers to create a list of categorized radicalization channels, [5], finds that Intellectual Dark Web content easily leads to Alt-lite content. Conversely, a longitudinal study of 300,000 Americans between 2016 and 2019 found that content from left and mainstream channels was more prevalent than far right or anti-woke content on YouTube, and a link between watching anti-woke content and extreme content remained unclear [6].

We seek to expand upon these findings using mobile data and content analysis. While in the acquisition process for the mFour data, which runs from fall 2020 to early 2023, we have run preliminary analysis on a dataset from Qrious, a different data provider using the same data collection mechanism. This smaller panel of 8,856 participants includes more than 3 million YouTube entries, over 1 million of which are videos and not ads. Individuals in this panel have watched 314,079 unique videos during the data collection period. Comparing the videos in our dataset with the labels from [5], we note that 2,418 of the videos are from labelled channels and about half of these are unique. Channel labels include Alt-right, Alt-lite, Intellectual Dark Web, Men Going Their Own Way (MGTOW), Men's Rights Activist (MRA), Pickup Artist (PUA), and media websites. A breakdown of watches within these categories can be seen in Table

1. The relatively low prevalence of labelled channels motivates our use of content analysis. We collect transcripts from the videos using the YouTube API and selenium when YouTube provides closed captioning and whisper [7] when these captions are not available. We tune a pre-trained word embeddings model to find instances of polarizing or hateful speech in the videos, and create a time series for each individual to note how much of this content they are watching over time and predict the viewing trajectories of other participants.

Our work makes three key contributions. One, we focus on mobile browsing using a two year panel from data provider mFour which also includes information about Twitter, Instagram, and Facebook use. Mobile behavior is especially critical as access to social network sites has become primarily mobile based [8]. Two, we perform content analysis on the videos watched by each individual to determine polarization at a video rather than a channel level. Not all content on a single channel may be aligned, and many channels contain messaging that is implicitly misogynistic, racist, or otherwise intolerant in nature without the channel outwardly holding such a stance. Three, we plan to add video analysis in order to pick up on gestures lost in the transcripts. Our longitudinal analysis of user behavior and content prevalence, though not causal in nature, provides implications for policy and future research.

References

[1] Brooke Auxier and Monica Anderson. Social media use in 2021. Washington, D.C., 2021. Pew Research Center.

[2] Connor McMahon, Isaac Johnson, and Brent Hecht. The substantial interdependence of wikipedia and google: A case study on the relationship between peer production communities and information technologies. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1):142–151, May 2017.

[3] James Bisbee, Megan Brown, Angela Lai, Richard Bonneau, Joshua A Tucker, and Jonathan Nagler. Election fraud, youtube, and public perception of the legitimacy of president biden. *Journal of Online Trust and Safety*, 1(3), Aug. 2022.

[4] Annie Y. Chen, Brendan Nyhan, Jason Reifler, Ronald E. Robertson, and Christo Wilson. Exposure to alternative extremist content on youtube. Washington, D.C., May 2022. Anti-Defamation League.

[5] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. Auditing radicalization pathways on youtube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 131–141, New York, NY, USA, 2020. Association for Computing Machinery.

[6] Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, Markus Mobius, David M. Rothschild, and Duncan J. Watts. Examining the consumption of radical content on youtube. *Proceedings of the National Academy of Sciences*, 118(32):e2101967118, 2021.

[7] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision, 2022.

[8] S. Dixon. Mobile share of social media visits in the united states from 4th quarter 2017 to 4th quarter 2019. Statista, 2022.

Channel Type	**Unique Videos**
Incel	1
Intellectual Dark Web	491
MGTOW	6
NONE	18
PUA	37
Center	104
Left	87
Left-Center	358
Right	9
Right-Center	27
Alt-Lite	130

Table 1: The number of videos in the Qrious dataset which fall into each category specified by [5]. As much research has noted, Center and Left-Center content is highly represented, though Intellectual Dark Web and Alt-Lite is still relatively common. Efforts to remove Alt-Right and Incel content are clear both through the lack of videos and the fact that many of these channels have been taken down.