Into The Looking Glass: A Random Walk Assessment of YouTube's Radicalization Pathway

Kevin He (sjhe@sfu.ca)
Department of Mathematics, Simon Fraser University
Burnaby, BC, V5A 1S6

Abstract

YouTube is the largest and most popular video hosting platform in the world with over 30 million visitors and almost 5 billion videos watched per day. It has been hypothesized by researchers and media that a radicalization pathway exists on this platform that draws users viewing political videos towards extremist content. In this paper, we sample over 800 channels and over 2400 videos of both independent YouTube creators alongside organizations that post content. These channels are then grouped by their political ideology, where their videos and their recommended videos are then sampled from these groupings to analyze. The sampled videos were then used to develop a Markov model that paints a picture of how these groupings connect with one another. Our findings contradict the popular belief of a radicalization pipeline and suggest that YouTube has a deradicalizing nature, favouring center leaning and a few right-wing groups. We also analyzed views and channels from each ideology to gain further insight on user watch preference based on video popularity.

Keywords: Radicalization; YouTube; Recommendations Algorithm; Algorithmic Extremism

Introduction

With the COVID-19 pandemic closing many popular gathering places and forcing educational facilities and workplaces to transition towards a work-from-home model, internet usage has soared with streaming services seeing surges by 12% as well as overall internet usage surging by 70% (Beech, 2020). This transition has increased the daily use of social media as well as has provided an opportunity for conspiracy theorists and the Alt-Right to increase their presence on the internet, as they discuss COVID-19 related conspiracy theories.

These conspiracies range in severity and create ties between COVID-19 with the introduction of new technology as well as politics. The theories claim that COVID-19 is a hoax, that symptoms are related to the implementation of 5G technology, as well as claiming that Americans favouring the Democratic Party are exaggerating its effects to hurt the Republican Party's 2020 Presidential Election bid (Enders & Uscinski, 2020). None of these theories have been verified for validity, yet they have been perpetuated by popular conservative news sites and right-wing social media users on websites such as Facebook, Twitter, or YouTube (Ecarma, 2020).

YouTube-type content, where users can upload and view videos from content creators, has shown a 5% growth of users since March 2020, with at least 70% of internet users streaming videos daily. This growth amounts to 5 billion videos watched daily with an average viewing session of 40 minutes (RTE, 2020) (Rutnik, 2019). Focusing solely on the YouTube platform, YouTube allows for a wide range of content to be posted from both individual creators and organizations. The site has seen a considerable growth in channels and videos discussing political reports, opinions, and conspiracy theories as its user base grows over time.

The YouTube homepage contains an endless list of recommended videos to watch next; these videos include previously watched videos, but primarily suggests new videos that the recommendations algorithm thinks the user would like. Videos are generated dynamically, and new suggested videos are loaded as users reach the end of the screen. The homepage also includes general recommended topics (eg. Cooking, Comedy, The Sims 4) where each topic includes a finite list of previous watched and new videos that it fits under its category. This recommendation system is highly personalized and more details about the algorithm are discussed in the following section. On the individual video level, when a user is watching a video, YouTube auto generates recommended videos that have a similar theme, as well as generates user-specific recommendations titled 'Recommended for you' that are sometimes unrelated to the video. New recommendations generate as users scroll down through the web page and stops after reaching a threshold that differs per video.

The inner workings of YouTube's video recommendation algorithm is largely unknown to the public audience, but there exist previous research papers citing the different factors the algorithm considers (Covington, Adams, & Sargin, 2016).These factors including user watch time, previous videos watched, metadata, and a multitude of other elements; this algorithm plays a large role in user video consumption as over 70% of the time spent on YouTube is watching videos recommended by the algorithm (Cooper, 2019).

Due to the unknown nature of the recommendations algorithm, it has been the site of heavy criticism with claims that the 'autoplay' feature, where a new video plays automatically once the current video has finished, has a tendency to recommend a more extreme version of the current video. In a report by the Atlantic, videos about vegetarianism led to veganism, and videos about jogging led to running ultramarathons, suggesting that the algorithm has a bias towards radical content (Friedersdorf, 2019). This claim would also apply to the political communities of YouTube, noting a radicalization pathway on the platform, where users starting out in partisan or mainstream political content, would then be recommended content from extreme political community, thus going down into a 'rabbit hole' where the content progressively becomes more radical due to the algorithm's recommendation (Roose, 2019).

YouTube does have a strict set of rules about what is allowed to be posted on its platform, named the Community Guidelines. These rules aim at preventing violent or dangerous content from being posted, including, harassment, hate speech, and violent and/or graphical content; but videos falling under these categories are difficult for an automatic algorithm to detect (Royer, 2019). The YouTube Partner program also attempts to address this issue by incentivizing users to post content that is deemed advertiser-friendly by YouTube standards through providing monetization on videos using ads. This is a strict program where any videos that include inappropriate language, hateful content, controversial issues or sensitive events will automatically be 'demonetized', where ads are stopped from running on the specific video. This prevents users from gaining any revenue on said video until it is considered safe for monetization (YouTube, n.d.).

In this paper, we address the claims that there exists a radicalization pipeline on YouTube through a quantitative model of consumer behaviour, specifically looking at YouTube videos and their recommendations. Our data gathering procedure and code is transparent and data sets will be available to the public for examination. More specifically, we ask:

RQ1: Does the YouTube recommendation algorithm steer users towards extreme content?
RQ2: How frequently does YouTube provide recommendations outside of the original channel's ideology?
RQ3: What proportion of time is spent in each ideology?

Using a dataset labelling political ideologies for over 800 channels created by Ledwich and Zaitsev (2019), we sample videos uniformly at random on the 200 most recent videos on each channel, then we gather data about each video and collect their recommended videos. We analyze this collected data to assess video views and popularity for each ideology. As well, we aggregate the data to create a Markov model to look at long term behaviours present in the recommendations system.

Data Collection and Modelling

The classification of YouTube channels is no easy task, as YouTube content creators do not explicitly fall under an individual political ideology. There is no comprehensive list available tagging all YouTube content creators to a specific ideology, with multiple existing lists clashing with one another. We analyzed previous academic research studying radicalization on YouTube and selected the data set created by Ledwich and Zaitsev (2019). This data set collected channels from Ad Fontes Media and Media Bias Factcheck which categorizes popular political media and assesses their bias, combining these two sources covers almost 80% of all YouTube views. Ledwich and Zaitsev (2019) as well included emerging channels through following the YouTube recommendation algorithm. Specifically, including channels with over 30% of videos being political and having over 10,000 subscribers. This list of channels was then manually analyzed with the help of a volunteer labeller where they assigned soft tags for each video. When two or more labellers defined a channel by the same tag, that tag is then assigned to the video, with a maximum of 4 tags for each video. Some channels were not provided with a tag due to the political ambiguity of the channel. These tags were then aggregated together to represent an ideology rather than a collection of tags (Ledwich & Zaitsev, 2019). The result of this aggregation could be seen in Table 1.

Table 1: Ideology Groupings by Ledwich and Zaitsev (2019)

Ideologies	Examples
Alt-Light	Mark Dice, Brittany Venti
Alt-Right	AustralianRealist, ramzpaul
Anti-Theist	TMM, Amazing Atheist
Anti-White	African Diaspora News Channel, ReyRosho
Conspiracy	PowerfulJRE, David Heavener, JoyCamp
Intellectual Dark Web (IDW)	Jordan Peterson Fan Channel, Patrick
Libertarian	PragerU, ReasonTV
Men's Right Activist (MRA)	Sandman, Thinking-Ape
Partisan Left	The Nation, Jimmy Kimmel Live
Partisan Right	Newsmax TV, Fox News Insider
Religious Conservative	Girl Defined, Glenn Beck
Revolutionary Socialist	Hakim, RE-EDUCATION
Social Justice	HuffPost, Patriot Act
Socialist	Shaun, Zero Books

With this dataset, we now describe our sampling technique. We start with a clean slate of YouTube with no previous watch and search history. Since we are unable to obtain personalized user data to analyze the recommendations algorithm after learning, we look at its base level performance instead.

We begin by filtering the data set to each ideology label to get a clearer view of the YouTube channels clustered together. We then built a web scraper that visits every channel in each ideology label to extract 200 of the most recent videos with over 10,000 views. This provides us with a glimpse of the recent activity by the channel that has a substantial amount of views. The 10,000 views threshold was chosen to remove recently created content, which might provide a bias for each tag. For channels with less than 200 videos, all the available videos were collected.

We then chose 100 videos from the list of collected videos from each ideology to gather further data on; this process will provide us with a comprehensive view of each ideology and their recommendations. In this collection process, each video is visited iteratively, and data on the title, view count, as well

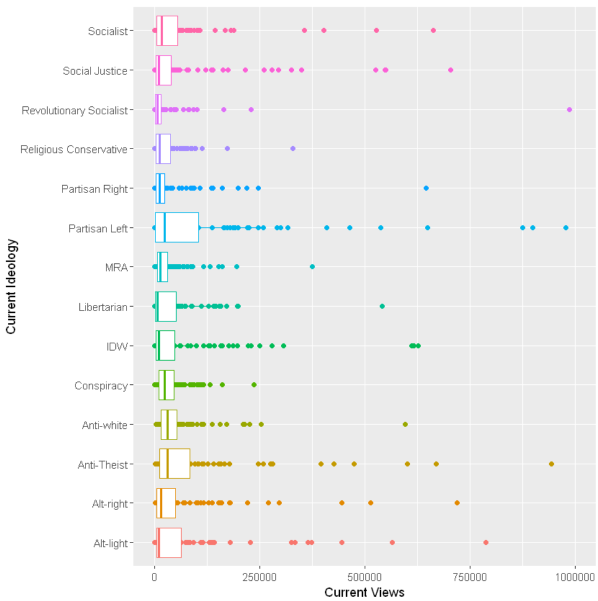

Figure 1: Views per Video in Different Ideologies

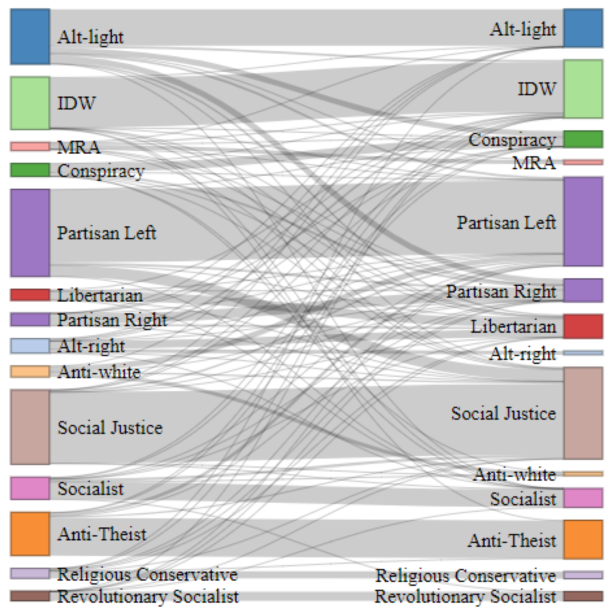

Figure 2: Flow between Ideologies

as its first 10 recommended videos titles and channels are gathered. Only the first 10 recommended videos are selected as those videos are prominently featured on the sidebar.

Findings and Discussion

With the collected data, we first analyze the viewership within each political community to assess popularity for each ideology. Figure 1 provides a scatterplot and a boxplot of views for each video within their respective labels providing a range from 0 to 1 million views, specifically with videos over 10,000 views in the data set. This boxplot suggests that Partisan Left, Anti-Theist, and the Alt-Light communities have a wider spread of views in the videos sampled. As well, for each community, the distribution of views is right skewed, which aligns with video behaviour on YouTube, as few videos becomes 'viral', which allows that video to grow rapidly in views for a short period, but often tapers off over time. Most communities hold a median view count around the 30,000 mark.

We then analyze the recommended videos from each ideology to see the connections between the ideology labels by using a modified impressions system. The impressions system provides insight on the number of times a video is recommended in a particular YouTube video, no matter if the recommended video was viewed or not (Zarzycki, 2018). This allows us to formally analyze the YouTube recommendation algorithm by providing a rough estimate of the types of recommended videos that appear in each ideology group. We make slight modifications to this system by dividing the number of impressions by the number of recommended videos we used for each video in our model. The result permits us to assign a proportion of the users who watched that video to different channels in the recommended videos list by aggregating together the different recommended videos belonging to the same channel.As recommendations are generated dynamically for each video, and may differ for different users, we assume that collecting recommendations from all videos belonging to a certain Ideology, would provide a generalization of recommended videos and channels.

Specifically, in this study, we assume that all users that viewed the original video goes on to choose one of the recommended videos and that they are equally likely to view any of recommended videos; we focus our findings on the number of times a specific channel is recommended, rather than the number of times a specific video appears. The modified impressions system naturally scales with the viewership a specific video has and the number of times a channel appears in its recommendations list, thus videos with higher views or channels that appears multiple times can respectively send and receive more impressions. From this, our impressions calculation becomes the number of views the current video has multiplied by the number of times a specific channel appears in the first ten recommendations divided by ten.

Using this method, looking at the first ten recommendations, if a YouTube video suggests n different videos belonging to one channel, that channel will then receive $10n\%$ of the users that watched the original video.This calculation is performed on each video sampled, and then summed together to determine the flow of users between ideologies.

Figure 2 provides a visualization for the result of this process through a Sankey diagram that displays this flow of users between the various ideologies. The categories on the left side is scaled in proportion with the total sum of views for each video within that community, as well the right side is scaled in proportion to the number viewers received. Here, **RQ2** is answered as this diagram displays that ideologies pri-

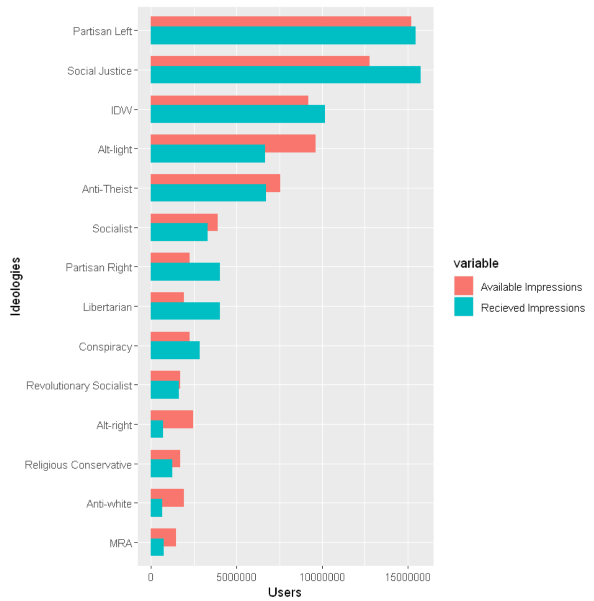

Figure 3: Users Gained and Lost from recommended videos

marily send users back to itself, with the rest of the audience being sent to a wide range of ideologies. The Alt-Light, IDW, Partisan Left, Social Justice, and Anti-Theist ideologies have a higher viewer count when compared with other ideologies, as well they dominate majority of the political sphere within YouTube in terms of viewership.

We are also able to analyze how the different ideologies are connected through the recommendations system. Outside of sending impressions back into itself, the remainder of viewers are sent towards other similar categories as the Partisan Left sends viewers towards Social Justice, Libertarian and Partisan Right ideologies, as well, there does exist recommendations towards more fringe communities.

In order for an ideology to be at an advantage using the modified impressions system, the ideology must receive more views than it originally began with. This could be easily assessed by looking at Figure 2 and comparing the sizes of categories before and after. A proper comparison of viewers gained and lost is displayed in Figure 3, which shows the difference in users recommended to a category and the users on the original category. Here, we are able to see that Partisan Right and Libertarian almost doubled the size of their community, while the Alt-Right, Anti-White and MRA ideologies have shrunk by half of their original viewers or more from this model. Generally, further right-wing groups had fewer recommendations, but there is a growth in views with IDW and Conspiracy channels on the platform suggesting that these communities are favoured by the recommendations algorithm in comparison to other right-wing ideologies. From this data **RQ1** is able to be answered, as through this model majority of central and leftist ideology communities have shown growth through the YouTube video recommendations.

Markov Model of Consumption

To gain a better understanding of the flow of users within the YouTube political sphere we look at a probabilistic view through a Markov model.

The key assumption that we are making is that the past videos viewed do not affect the user's choice of recommended video. By using a Markov model we're able to look at the long term behaviour of the Markov chain as users consume more content.

Similar to our above work with the modified impression system, for this model we tweak this system and calculate impressions for each individual video by multiplying the number of views the current video has by the number of times a specific channel appears in the recommended list. As well, we only look at the top 10 recommendations from each video. The recommended channels are then linked to an ideology, and then summed together to find the total number of impressions sent.

To finalize this model we normalized the data by the total impressions sent by each ideology to gain the transition probabilities for the Markov model. Similarly, videos with higher views as well as channels that appear multiple times in the recommended list are scaled appropriately to denote their significance. Videos with fewer views are put at a disadvantage since that would suggest lower viewership within that community.

Using this model, users take a random walk, and the proportion of time they spend consuming content from each ideology is collected by solving for the limiting distribution the Markov model, this data could be found in Table 3. Fringe political communities including Religious Conservative, Revolutionary Socialist, MRA, Anti-White, and Alt-Right are shown to be at a disadvantage in this model and as well stays consistent with **RQ1** suggesting that the recommendations algorithm does not steer users towards extreme content.

Table 3: Markov Model Steady Probability

Ideologies	Proportion
Alt-Light	0.01
Alt-Right	0
Anti-Theist	0.01
Anti-White	0
Conspiracy	0.01
IDW	0.14
Libertarian	0.05
MRA	0
Partisan Left	0.13
Partisan Right	0.05
Religious Conservative	0
Revolutionary Socialist	0
Social Justice	0.55
Socialist	0.04

This model answers **RQ3** as Social Justice channels are heavily favoured by the model as 55% of the time in the

YouTube political sphere are spent in the Social Justice ideology. This finding is consistent with the work with the modified impressions system we previously worked with as Social Justice has the second-largest impressions pool, and it receives more impressions than it had to send by a statistically significant margin. In a further analysis analyzing the channels belonging to the Social Justice ideology, the ideology contains popular channels such as 'VICE','Last Week Tonight', as well as 'Vox'. These channels primarily cater to the general YouTube audience while promoting Social Justice issues, as well as bolstering large amounts of subscribers and video views. Despite these channels not being strictly political, they still effectively claim the majority of the time spent in the political communities of YouTube.

Outside of Social Justice, the IDW and Partisan Left communities hold 14% and 13% of the proportion of time in the YouTube political sphere respectively. Similarly, this finding holds with the modified impressions system as both categories had a substantial number of available impressions and both shown an increase in the impressions received. The notion that the IDW provides a gateway towards radicalization is proven to be false with our findings, as further fringe communities are shown to be at a disadvantage by the recommendations algorithm by the findings above. Instead, there exists a notion of growth within the IDW bubble despite its recent inception.

Limitations

The primary limiting factor to this study involves personalization of the recommendations algorithm through previous watch history and search history. As this personalization process is done on an individual level we are unable to see the effects of how this process would affect the recommended videos displayed and how the users would interact with those videos specifically. As well, we would need to analyze user interactions with the recommended videos to gain a clearer understanding of how users select which video to watch next. In order to address this, we would require the history of personalized recommendations over a longer period of time for each user, which is something only YouTube as access to.

There also exist numerous issues with categorizing channels into different ideology categories. YouTube content creators are not limited to post videos about a specific topic only as they have full control of what content they want to post. They might choose to cater their channel towards a specific topic, but frequently there are videos that cover other content. This makes categorizing channels extremely difficult as it oversees the other contents that the channel may cover in addition to its primary videos. Due to this, there is no definitive list of YouTube channels and their political ideology, and current data sets often conflict with each other about their categorizations.

From this issue in the study, we only focused on the channels that were categorized by an ideology, thus ignoring other channels that were not categorized. When analyzing video recommendations approximately 50% of the channels listed were not categorized, and this significant portion of videos was ignored. Further categorization is required to gain a complete understanding of the recommendation algorithm's behaviour.

References

Beech, M. (2020, March). *Covid-19 pushes up internet use 70% and streaming more than 12%.* `https://www.forbes.com/sites/markbeech/2020/03/25/covid-19-pushes-up-internet-use-70-streaming-more-than-12-first-figures-reveal/#b84d3153104e`. (Accessed 07-17-2020)

Cooper, P. (2019, November). *How does the youtube algorithm work? a guide to getting more views0.* `https://blog.hootsuite.com/how-the-youtube-algorithm-works/`. (Accessed 07-17-2020)

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th acm conference on recommender systems.* New York, NY, USA. (Accessed 08-05-2020)

Ecarma, C. (2020, April). *Covid-19 conspiracies are turobocharged through conservative media.* `https://www.vanityfair.com/news/2020/04/coronavirus-conspiracies-charged-conservative-media-fox-news`. (Accessed 07-22-2020)

Enders, A., & Uscinski, J. (2020, May). *Conspiracy theories run rampant when people feel helpless. like now.* `https://www.washingtonpost.com/outlook/2020/05/05/coronavirus-conspiracy-theories-pandemic/`. (Accessed 07-17-2020)

Friedersdorf, C. (2019, June). *Youtube extremism and the long tail.* `https://www.theatlantic.com/politics/archive/2018/03/youtube-extremism-and-the-long-tail/555350/`. (Accessed 07-17-2020)

Ledwich, M., & Zaitsev, A. (2019, 12). *Algorithmic extremism: Examining youtube's rabbit hole of radicalization.*

Roose, K. (2019, June). *The making of a youtube radical.* `https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html`. (Accessed 07-17-2020)

Royer, A. (2019, July). *Why youtube's decision to remove far-right content is not enough.* `https://www.themantle.com/arts-and-culture/why-youtubes-decision-remove-far-right-content-not-enough`. (Accessed 07-17-2020)

RTE. (2020, June). *Internet usage increased during march covid-19 lockdown.* `https://www.rte.ie/news/business/2020/0619/1148430-cso-internet-usage/`. (Accessed 07-17-2020)

Rutnik, M. (2019, August). *Youtube in numbers: Monthly views, most popular video, and more fun stats!* `https://www.androidauthority.com/youtube-stats-1016070/`. (Accessed 07-17-2020)

YouTube. (n.d.). *Advertiser-friendly content guidelines - youtube help.* https://support.google.com/youtube/answer/6162278?hl=en. (Accessed 07-17-2020)

Zarzycki, N. (2018, June). *Reach vs. impressions: What's the difference (and what should you track)?* https://blog.hootsuite.com/reach-vs-impressions/. (Accessed 07-17-2020)