Auditing Radicalization Pathways on YouTube

Manoel Horta Ribeiro
manoelribeiro@dcc.ufmg.br
UFMG

Raphael Ottoni
rapha@dcc.ufmg.br
UFMG

Robert West
robert.west@epfl.ch
EPFL

Virgílio A. F. Almeida
virgilio@dcc.ufmg.br
UFMG, Berkman Klein Center

Wagner Meira Jr.
meira@dcc.ufmg.br
UFMG

ABSTRACT

Non-profits and the media claim there is a radicalization pipeline on YouTube. Its content creators would sponsor fringe ideas, and its recommender system would steer users towards edgier content. Yet, the supporting evidence for this claim is mostly anecdotal, and there are no proper measurements of the influence of YouTube's recommender system. In this work, we conduct a large scale audit of user radicalization on YouTube. We analyze 331, 849 videos of 360 channels which we broadly classify into: control, the Alt-lite, the Intellectual Dark Web (I.D.W.), and the Alt-right —channels in the I.D.W. and the Alt-lite would be gateways to fringe far-right ideology, here represented by Alt-right channels. Processing more than 79M comments, we show that the three communities increasingly share the same user base; that users consistently migrate from milder to more extreme content; and that a large percentage of users who consume Alt-right content now consumed Alt-lite and I.D.W. content in the past. We also probe YouTube's recommendation algorithm, looking at more than 2M million recommendations for videos and channels between May and July 2019. We find that Alt-lite content is easily reachable from I.D.W. channels via recommendations and that Alt-right channels may be reached from both I.D.W. and Alt-lite channels. Overall, we paint a comprehensive picture of user radicalization on YouTube and provide methods to transparently audit the platform and its recommender system.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**.

KEYWORDS

user radicalization, hate speech, extremism, alt-right

1 INTRODUCTION

On YouTube, channels that discuss social, political and cultural subjects have flourished. Among them, one may find individuals such as Jordan Peterson and Joe Rogan, associated with the so-called Intellectual Dark Web (I.D.W.) —iconoclastic thinkers, academics and media personalities [43]— but also openly declared white nationalists like Richard Spencer and Jared Taylor, which have been broadly referred to as Alt-right [5].

These individuals do not only share the same platform, but often publicly engage in debates and conversations on the website [25]. All the previously mentioned individuals, for example, are connected by joint video appearances: Jordan Peterson was interviewed by Joe Rogan [34], who interviewed YouTuber Carl Benjamin [35], who debated Richard Spencer [8], who was in a panel with Jared Taylor in an Alt-right conference [36]. According to Lewis [25], this proximity would create "radicalization pathways" for audience members and content creators. Anecdotal examples of these journeys are plenty, including *Roosh V*'s content creator trajectory, going from a Pick Up Artist to Alt-right supporter [24, 38], and Caleb Cain's testimony of his YouTube-driven radicalization [37].

The claim that there is a "radicalization pipeline" on YouTube should be considered in the context of decreasing trust in mainstream media and increasing influence of social networks. Across the globe, individuals are skeptical of traditional media vehicles and growingly consume news and opinion content on social media [22, 31]. In this setting, recent research has shown that fringe websites (like *4chan*) and subreddits (like */r/TheDonald*) have great influence over which memes [44] and news [45] are shared in large social networks, such as Twitter. YouTube is extremely popular, especially among children and teenagers [7], and, if the streaming website is actually radicalizing individuals, this can push fringe ideologies like white supremacy further into the mainstream [42].

A key issue in dealing with topics like radicalization and hate speech is the lack of agreement over what is "hateful" or "extreme" [39]. A work-around is to perform community-based analyses, rather than trying to label what is or is not hateful. For the purpose of this work, we consider three communities that have been associated with user radicalization [25, 37, 43], and that differ in the extremity of their content: the Intellectual Dark Web (I.D.W.), the Alt-lite and the Alt-right. While the I.D.W. discuss controversial subjects like race and I.Q. [43], the Alt-right sponsor fringe ideas like that of a white ethnostate [19]. Somewhere in the middle, individuals of the Alt-lite deny to embrace white supremacist ideology, although they constantly flirt with concepts associated with it (e.g., the great replacement, globalist conspiracies, etc.).

Present Work. In this paper, we audit whether users are indeed becoming radicalized on YouTube, and whether the recommendation algorithms contribute towards this radicalization. We do so by examining three proeminent communities: the Intellectual Dark Web, the Alt-lite and the Alt-right. More specifically, considering Alt-right content as a proxy for extreme content, we ask:

RQ1 How was the growth of these communities on YouTube in the last decade?

RQ2 To which extent do users systematically gravitate towards more extreme content?

RQ3 Do algorithmic recommendations steer users towards more extreme content?

We develop a data collection process where we *(i)* acquire a large pool of relevant channels from these communities; *(ii)* collect metadata and comments for each of the videos in the channels; *(iii)* annotate channels as belonging to several different communities; and *(iv)* collect YouTube video and channel recommendations. We additionally collect traditional and alternative media channels to employ as control. Notice that we use these as a sanity check to capture the growth of similar content on YouTube, rather than trying to obtain similar users in other channels. These efforts resulted in a dataset with more than 79 million comments in $331,849$ videos of 360 channels, and with more than $2M$ video and $10k$ channel recommendations. Importantly, our recommendations do not account for personalization. We analyse this large dataset extensively:

- We look at the growth of the I.D.W., the Alt-lite and the Alt-right throughout the last decade in terms of videos, likes and views, finding a steep rise in activity and engagement in the communities of interest when compared with the control channels. Moreover, comments per view seem to be particularly high in more extreme content, reaching near to 1 comment for each 5 views in Alt-right channels (Sec. 4).
- In that scenario, we inspect the intersection of commenting users within the communities, finding they increasingly share the same user base. Analyzing the overlap coefficient of the sets of commenting users, we find that more than half of users who commented on Alt-right channels in 2018 also comment on Alt-lite and on I.D.W. channels (Sec. 5).
- We also find that the intersection is not only growing due to new users but that there is significant user migration among the communities being studied. Users that initially comment only on content from the I.D.W. or the Alt-lite throughout the years consistently start to comment on Alt-right content. These users are a significant fraction of the Alt-right commenting user base. Interestingly, although control channels share, on a yearly basis, a significant number of users with Alt-right channels, we cannot observe significant user migration from them to Alt-right channels (Sec. 6).
- Lastly, we take a look on the impact of YouTube's recommendation algorithms, running simulations on recommendation graphs. Our analyses show that, particularly through the channel recommender system, Alt-lite channels are easily discovered from I.D.W. channels, and that Alt-right channels may be reached from the two other communities (Sec. 7).

This is, to our best knowledge, the first large scale quantitative audit of user radicalization on YouTube. We find strong evidence for radicalization among YouTube users, and that YouTube's recommender system enables Alt-right channels to be discovered, even in a scenario without personalization. We discuss our findings and our limitations in light of the research questions further in Sec. 8. We argue that commenting users are a good enough proxy to measure the phenomena, as more extreme content seems to beget more comments. Moreover, regardless of the degree of influence of the recommender system in the process of radicalizing users, there is significant evidence that this process is happening, and that appropriate measures should be taken.

2 BACKGROUND

Contrarian Communities. We discuss three of YouTube's prominent communities: the Intellectual Dark Web, the Alt-lite and the Alt-right. We argue that all of them are *contrarians*, in the sense that they strongly oppose mainstream views or attitudes. According to Nagle, these communities flourished in the wave of "anti-PC" culture of the 2010s, where social-political movements (e.g. the transgender rights movement, the anti-sexual assault movement) were portrayed as hysterical, and their claims, as absurd [30].

According the Anti Defamation League [5], the Alt-Right is a loose segment of the white supremacist movement consisting of individuals who reject mainstream conservatism in favor of politics that embrace racist, anti-Semitic and white supremacist ideology. The Alt-right skews younger than other far right groups, and has a big online presence, particularly on fringe web sites like 4chan, 8chan and certain corners of Reddit [1].

The term Alt-lite was created to differentiate right-wing activists who deny to embracing white supremacist ideology. Atkison argues that the Unite the Rally in Charlottesville was deeply related to this change, as participants of the rally revealed the movement's white supremacist leanings and affiliations [10]. Alt-right writer and white supremacist Greg Johnson [5] describes the difference between Alt-right and Alt-lite by the origin of it's nationalism: *"The Alt-lite is defined by civic nationalism as opposed to racial nationalism, which is a defining characteristic of the Alt-right".* This distinction was also highlighted in [28]. Yet it is important to point out that the line between the Alt-right and the Alt-lite is blurry [5], as many Alt-liters are accused of dog-whistling: attenuating their real beliefs to appeal to a more general public and to prevent getting banned [23, 26]. To address this problem, in this paper we take a very conservative approach to our labeling, naming only the most extreme content creators as Alt-right. This is explained in further detail in Sec. 3.

The Intellectual Dark Web (I.D.W.) is a term coined by Eric Ross Weinstein to refer to a particular group of academics and podcast hosts. The neologism was later popularized in a New York Times opinion article [43], where it is employed to describe: *"collection of iconoclastic thinkers, academic renegades and media personalities who are having a rolling conversation about all sorts of subjects, (...) touching on controversial issues such as abortion, biological differences between men and women, identity politics, religion, immigration, etc".*

The group described in the NYT piece includes Sam Harris, Jordan Peterson, Ben Shapiro, Dave Rubin and Joe Rogan, and also

mentions a website with an unofficial list of members. Members of the so-called I.D.W have been accused of bigotry, including Islamophobia [11], transphobia [4] and sexism [17]. Moreover, a recent report by Data & Society research institute has claimed these channels are "pathways to radicalization" [25]: they would act as an entry point to more radical channels, such as those in Alt-right. Broadly, members of this loosely defined movement see these critics as a consequence of discussing controversial subjects [43] and largely ignored/dismissed the report [41]. Similarly to what happens between Alt-right and Alt-lite, there is also blurry lines between the I.D.W. and the Alt-lite, especially for non-core members, like those listed in the website. Here, again, we take a conservative approach, considering borderline cases to belong to the Alt-lite.

Radicalization. We approach this central concept with the definition of McCauley and Moskalenko [29]: *Functionally, political radicalization is increased preparation for and commitment to intergroup conflict. Descriptively, radicalization means change in beliefs, feelings, and behaviors in directions that increasingly justify intergroup violence and demand sacrifice in defense of the ingroup.* We use the consumption of Alt-right content as a proxy for radicalization. We argue this is reasonable because the rhetoric preached by the Alt-right has been associated with multiple recent terrorist attacks (e.g. the Christchurch mass shooting [27]), and because it champions ideas associated with intergroup conflict (e.g. a white ethnostate [19]). Our conservative strategy when labeling channels is of particular importance here: Alt-right channels are closely related to these ideas, while the Alt-lite and the I.D.W. are given the benefit of the doubt.

Auditing the web. As algorithms play an ever-larger role in our lives, it is increasingly important for researchers and society at large to reverse engineer algorithms input-output relationships [15]. Previous large scale algorithmic auditing include measuring discrimination on AirBnB [16], personalization on web search [20] and price discrimination on e-commerce web sites [21]. We argue this work is an audit in the sense that is measures a troublesome phenomena (user radicalization) in a content-sharing social environment heavily influenced by algorithms (YouTube). Unfortunately, it is impossible to obtain the entire history of YouTube recommendation, so we must limit algorithmic analyses to a time slice of a constantly changing black-box. Although comments may give us insight into the past, its impossible to tease apart the influence of the algorithm in previous times. Another limitation of our auditing is that we do not account for user personalization. Despite these flaws, we argue that: *(i)* our analyses provide answers to important questions related with impactful societal processes that are allegedly happening in YouTube, and *(ii)* our framework for auditing user radicalization can be replicated through time, and expanded to handle personalization. Regardless of the extent of the contribution of YouTube's algorithm towards the process of user radicalization, understanding this process and finding ways to fight it is still a timely question.

Previous research from/on YouTube. Previous work by Google sheds light into some of the high-level technicalities of YouTube's recommender system [13, 14]. Their latest paper indicates they use feed embeddings for video searches and video histories into a dense feed-forward neural network [14]. There also exists a large body of work studying violent [18], hateful or extremist [6, 40]

and disturbing content [33] on the platform. Much of the existing work focus on creating detection algorithms for these types of content using features of the comments, the commenting users and the videos [6, 18]. Notably, Sureka et al. [40] use a seed-expanding methodology to track extremist user communities, which yielded high precision in including relevant users. This is somewhat analogous to what we do, although we use YouTube's recommender system while they use user friends, subscriptions and favourites. Ottoni et al. perform an in-depth textual analysis of 23 channels (13 broadly defined as Alt-right), finding significantly different topics across the two groups [32].

3 DATA COLLECTION

We are interested in three communities on YouTube: the I.D.W., the Alt-lite, and the Alt-right. Identifying such communities and the channels which belong to them is no easy task: the membership of channels to these communities is volatile and fuzzy, and there is disagreement between how members of these communities view themselves, and how they are considered by scholars and the media. These particularities make our challenge multi-faceted: on one hand, we want to study user radicalization, and determine, for example, if users who start watching videos by communities like the I.D.W. eventually go on to consume Alt-right content. On the other, there is often no clear agreement on who belongs to which community.

Due to these nuances, we devise a careful methodology to **(a)** collect a large pool of relevant channels; **(b)** collect data and the recommendations given by YouTube for these channels; **(c)** manually labeling these channels according to the communities of interest.

(a) For each community, we create a pool of channels as follows. We refer to channels obtained in the *i*-th step as *Type i* channels.

(1) We choose a set of *seed channels.* Seeds were extracted from the I.D.W. unofficial website [9], Anti Defamation League's report on the Alt-lite/the Alt-right [5] and Data & Society's report on YouTube Radicalization [25]. We pick popular channels that are representative of the community we are interested at. Each seed was independently annotated two times and discarded in case there was any disagreement. We further detail the annotation process later in this section.

(2) We choose a set of *keywords* related to the sub-communities. For each keyword, we use YouTube's search functionality and consider the first 200 results in English. We then add channels that broadly relate in topic to the community in question. For example, for the Alt-right, keywords included both terms associated with their narratives, such as *The Jewish Question* and *White Genocide*, as well as the names or nicknames of famous Alt-righters, such as *weev* and *Christopher Cantwell*.

(3) We iteratively search the related and featured channels collected in steps (1) and (2), adding relevant channels (as defined in 2). Note that these are two ways channel can link to each other. Featured channels may be chosen by YouTube content creators: if your friend has a channel and you want to support it, you can put it on your "Featured Channels" tab. Related channels are recommendations made by YouTube's recommender system.

(4) We repeat step (3), iteratively collecting another hop of featured/recommended channels from those obtained from 3.

Table 1: Top 16 YouTube channels with the most views per each community and for controls.

	Alt-right	Views	Alt-lite	Views	Intellectual Dark Web	Views	Control	Views
1	James Allsup	62M	StevenCrowder	727M	PowerfulJRE	1B	vox	1B
2	Black Pigeon Speaks	50M	Rebel Media	405M	JRE Clips	717M	gq magazine	1B
3	ThuleanPerspective	45M	Paul Joseph Watson	356M	PragerUniversity	635M	vice news	1B
4	Red Ice TV	42M	MarkDice	334M	SargonofAkkad100	258M	wired magazine	1B
5	The Golden One	12M	Stefan Molyneux	193M	The Daily Wire	247M	vanity fair	639M
6	AmRenVideos	9M	hOrnsticles3	145M	The Rubin Report	206M	the verge	636M
7	NeatoBurrito Productions	7M	MILO	133M	ReasonTV	138M	glamour magazine	620M
8	The Last Stand	7M	Styxhexenhammer666	132M	JordanPetersonVideos	90M	business insider	523M
9	MillennialWoes	6M	OneTruth4Life	112M	Bite-sized Philosophy	62M	huffington post (huffpost)	329M
10	Mark Collett	6M	No Bullshit	104M	Owen Benjamin	35M	today i found out	328M
11	AustralianRealist	5M	SJWCentral	90M	AgatanFoundation	33M	cbc news	324M
12	Jean-François Gariépy	5M	Computing Forever	87M	Essential Truth	32M	the guardian	300M
13	Prince of Zimbabwe	5M	The Thinkery	86M	Ben Shapiro	30M	people magazine	287M
14	The Alternative Hypothesis	5M	Bearing	81M	YAFTV	30M	big think	258M
15	Matthew North	4M	RobinHoodUKIP	64M	joerogandotnet	25M	cosmopolitan	256M
16	Faith J Goldy	4M	patcondell	64M	TheArchangel911	24M	global news	252M

The annotation process done here followed the same instructions as the one explained in detail for data collection step **(c)**. Steps (2)—(4), were done by an individual with more than 50 hours of watch-time of the communities of interest. Notice that, in steps (2)—(4), we are not labeling the channels, but creating a pool of channels to be further inspected and labeled in subsequent steps. The complete list of seeds obtained from (1) and of keywords used in (2) may be found in Appendix A. A clear distinction between featured and recommended channels may be found in Appendix B

(b) For each channel we collect the number of subscribers and views, and for their videos, all the comments and captions. Video and channel recommendations were collected separately using custom-made crawlers. We collected multiple "rounds" of recommendations, 22 for channel recommendations and 19 for video recommendations. Each "round" consists of collecting all recommended channels (on the channel web page) and all recommended video (on the video web page). To circumvent possible location bias in the data we collected we used VPNs from 7 different locations: 3 in the USA, 2 in Canada, 1 in Switzerland and 1 in Brazil. Moreover, channels were always visited in random order, to prevent any biases from arising from session-based recommendations. As we extensively discuss throughout the paper, this does not include personalization, as we do not log in into any account.

(c) Channel labeling was done in multiple steps. All channels are either seeds (*Type 1*) or obtained through YouTube's recommendation/search engine (*Types 2 and 3*). Notice that *Type 1* channels were assigned labels at the time of their collection. For the others, we had 2 researchers annotate them carefully. They both had significant experience with the communities being studied, and were given the following instructions:

> Carefully inspect each one of the channels in this table, taking a look at the most popular videos, and watching, altogether, at least 5 minutes of content from that channel. Then you should decide if the channel belongs to the Alt-right, the Alt-lite, the Intellectual Dark Web (I.D.W.), or whether you think it doesn't fit any of the communities. To get a grasp on who belongs to the I.D.W., read [43], and check out the website with some of the alleged members of the group [9]. Yet, we ask you to consider the label holistically, including channels that have content from these creators and with a similar spirit to also belong in this category. To distinguish between the Alt-right and the Alt-lite, read [5] and [28]. It is important to stress the difference between civic nationalism and racial nationalism in that case. Please consider the Alt-right label only to the most extreme content. You are encouraged to search on the internet for the name of the content creator to help you make your decision.

The annotation process lasted for 3 weeks. In case they disagreed, they had to discuss the cases individually until a conclusion was reached. Interanotator agreement was of 75.57%. We ended up with 90 I.D.W., 114 Alt-lite and 88 Alt-right channels.

Controls. Additionally we collect popular media channels as controls. These were obtained from the *mediabiasfactcheck.com* [3]. For each media source of the categories in the website (*Left, Left-Center, Center, Right-Center, Right*) we search for its name on YouTube, and consider it if there is a match in the first page of results [3]. Some of the channels were not considered because they had too many videos (15, 000+) and we were not able to retrieve them all (which is important, because our analysis are temporal). In total, we collect 68 channels in that way. Notice that although we name this group as control throughout the paper, we did not attempt to replicate the distribution of users w.r.t. some aspect in the communities of interest. We use these control channels as a sanity check to capture general trends among more mainstream YouTube channels.

We summarize the dataset collected in the Tab. 2. Data collection was performed from the 19th to the 30th of May 2019, and the collection of the recommendations between May and July 2019. We illustrate the difference between video and channel recommendations, and provide the full list of channel with raw labels, as well as seed channels and keywords used in Appendix A.

Table 2: Overview of our dataset.

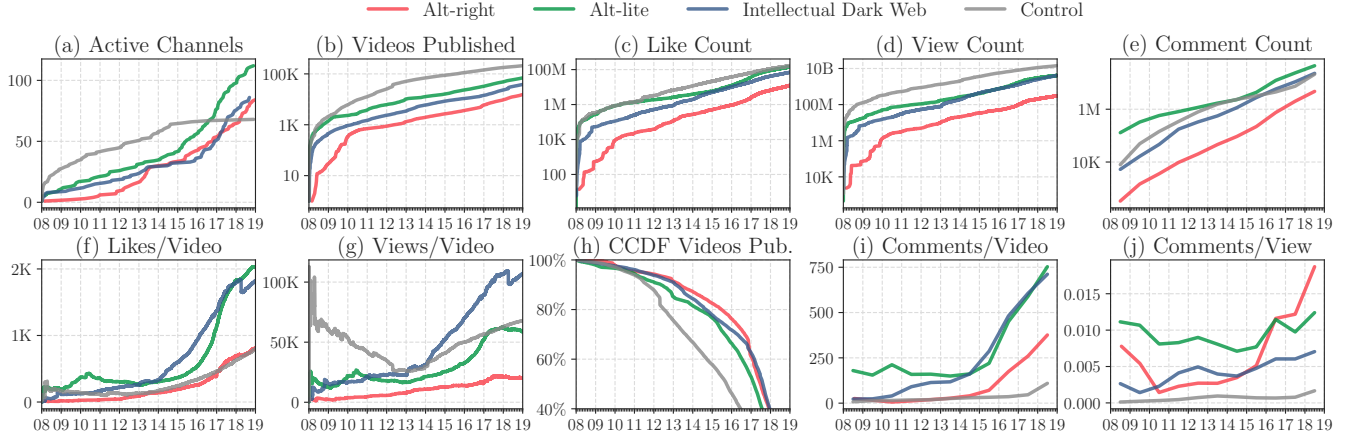|			
Channels	360	Video Recs rounds	19
Videos	331, 849	Video Recs	2,474,044
Comments	79,180,534	Channel Recs Rounds	22
Commenting users	6,308,759	Channel Recs	14,283

Figure 1: On the top row (a)–(e), for each community and controls, we have the cumulative number of active channels (that posted at least one video), of videos published, of likes, views and of comments. In the bottom row, we have CDFs for engagement metrics, and the CCDF of videos published, zoomed in in the range [40%, 100%] on the y-axis. Notice that for comments, we know only the year when they were published, and thus the CDFs granularity is coarser (years rather than seconds). The raw numbers of views, likes, videos published and more is shown in App. C.

4 THE RISE OF CONTRARIANS

We present an overview of the channels in the communities of interest, and show results about their growth in the last years, setting the stage to more in-depth analyses in later sections. Tab. 1 shows the 16 most viewed YouTubers for each of the communities and for the controls, and Figure 1 shows information on the number of videos published, channels created, likes, views and comments per year, as well as several engagement metrics.

Recent rise in activity. Figs. 1(a)–(e) show the rise in channel creation, video publishing, likes, views and comments in the last decade. The four latter are growing exponentially for all the communities of interest and for the control channels. Noticeably, the rise in the number of active channels is much more recent for the communities of interest than for controls, as shown in Fig. 1(a). In mid 2015, for example, while more than 75% of the channels in the control group were active (posted their first video), only slightly more than 50% of Alt-lite and less than 50% of Alt-right and I.D.W. channels had done so. This growth in the communities of interest during 2015 may also be noted in Fig. 1(i), which shows the CDF of number comments per videos, and can also be seen between early 2014 and late 2016 in Figs. 1(f)–(g), which show the number of likes and views per video, respectively. Notice that the number of likes and views is obtained during data collection, and thus, it might be that older videos from those channels became popular later. Altogether, our data corroborates with the narrative that these communities gained traction in (and fortified) Donald Trump's campaign during the 2016 presidential elections [2, 12].

Engagement. A key difference between the communities of interest and the control channels is the level of engagement with the videos, as portrayed by the number of likes per video, comments per video and comments per view, shown in Figs. 1 (f), (i), and (j), respectively. For all these metrics, the communities of interest have more engagement than the control channels: Although control

channels have more views per video, as shown in Figs. 1(g), these views are less often converted into likes and comments. Notably, Alt-right channels have, since 2017, become the ones with the highest number of comments per view, with nearly 1 comment per 5 views by 2018. The alignment of the high rate of comments/view with the extremity of communities (Alt-right > Alt-lite > I.D.W. > Control) is particularly helpful as we do have access to the entire trajectories of comments by users.

Dormant Alt-right Channels. Although by 2013, approximately the same number of channels of all three communities had become active (∼ 30), as it can be seen in Fig. 1(a), the number of videos they published by the Alt-right was very low before 2016. This can be seen in the CCDF in Fig. 1(h): while control and Alt-lite channels had published nearly 40% of their content, the Alt-right had published a bit more than 20%. This is not because the most popular channels did not yet exist: 4 out of the 5 current top Alt-right channels (accumulating approximately 150M views) had already been created by 2013. Moreover, it is noteworthy that many of the channels now dedicated to Alt-right content have initial videos related to other subjects. Take for example the channel "The Golden One", number 5 in Tab. 1. Most of the initial videos in the channel are about working out or video-games, with politics related videos becoming increasingly occurring. The growth in engagement metrics such as likes per video and comments per video of the Alt-right succeeds that of the I.D.W. and of the Alt-lite, resonating with the narrative that the rise of Alt-Lite and I.D.W. channels created fertile grounds for individuals with fringe ideas to prosper [25, 30].

Although our data-driven analysis sheds light into existing narratives on the communities of interest, it is still impossible to determine, from these simple CDFs, whether there is a radicalization pipeline. To do so, in the following two sections, we dig deeper into the relationship between these communities looking closely at the users who commented on them.
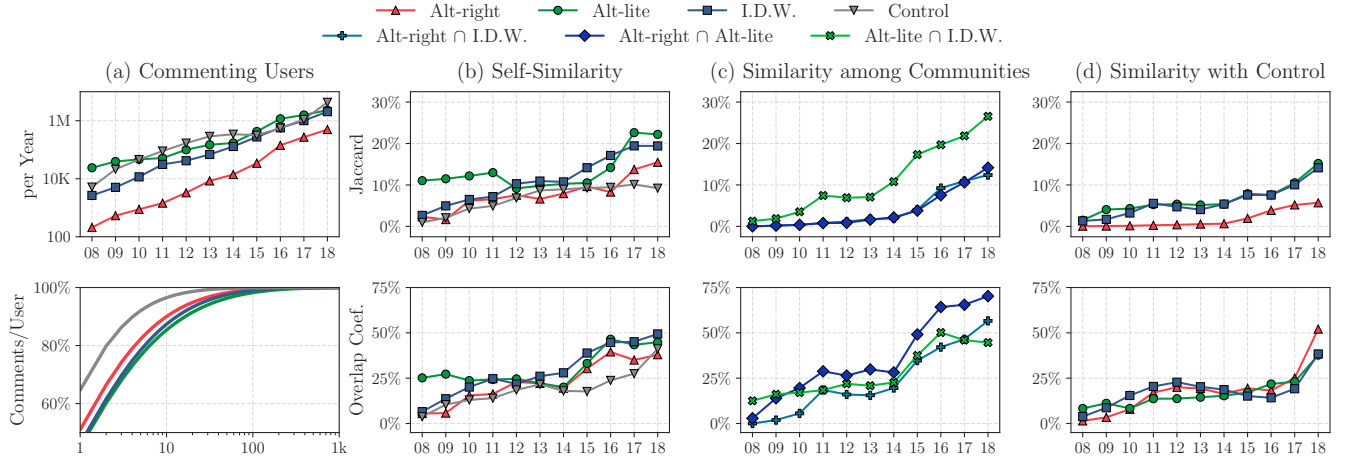
Figure 2: In *(a)*, the number of unique commenting users per year in the top plot and the CDF of comments per user for each one of the communities in the bottom plot. In *(b)—(d)* we show two similarity metrics (Jaccard and Overlap Coefficient) for different pairs of sets of commenting users across the years. In *(b)* these pairs are the sets of users of each community in subsequent years. In *(c)* these pairs are the sets of users of each one of the communities of interest. In *(d)* these pairs are the sets of users of the communities compared with the users who commented in control channels.

5 USER INTERSECTION

We begin our in-depth analysis of users who commented in the channels of interest by analysing the *intersection* between the users in different channels and communities. In that context, we use two set similarity metrics: the Jaccard Similarity $\frac{|A \cap B|}{|A \cup B|}$; and the Overlap Coefficient $\frac{|A \cap B|}{\min(|A|,|B|)}$. Notice that the overlap coefficient is particularly useful to compare communities of different sizes. For example a small subset of a large set may yield low Jaccard Similarity, but will necessarily yield an Overlap Coefficient of 1.

Column (a) of Fig. 2 characterizes commenting users. The top plot shows the absolute number of commenting users per year, while the bottom one shows the CDF of the number of comments per users per community. It is interesting to compare these plots with that of Fig. 1(e), as we can see that the communities of interest have many more highly active commenters. This supports the hypothesis that users who consume content in the communities of interest are more "engaged" than those who consume the content from the control channels. Notice also that, although the Alt-right commenters have, in average, fewer comments than those in Alt-lite or the I.D.W., the community is much younger (as discussed in Sec. 4), and thus it is hard to tell whether their users are less engaged.

In columns (b)—(d) of Fig. 2 we consider the intersection between the commenting users of the I.D.W., the Alt-lite, the Alt-right and control channels. The top figure for each column shows the Jaccard Similarity and the bottom one shows the Overlap Coefficient.

Column (b) in Fig. 2 shows the similarity measures for a community with itself a year before (which here we name self-similarity). We find that the retention of users among the three communities is growing with time for both metrics. However, for control channels we find that the Jaccard similarity is actually decreasing since 2015, and that the overlap coefficient only recently started to grow, perhaps due to the sharp increase in commenting users since 2015.

Commenting users from the communities of interest seem to go back more often than those in control.

Column (c) in Fig. 2 shows the pairwise similarity between the three communities. Notably, the Jaccard Similarity between the Alt-lite and the I.D.W. is higher than the self-similarity of these two communities, reaching almost 30% over time. Moreover, the Overlap Coefficient of the Alt-right with the Alt-lite and the Alt-right is high: approximately 55% for the I.D.W. and 70% for the Alt-lite in 2018. This means more than half of the users who commented in Alt-right channels commented in the other communities.

Lastly, column (d) in Fig. 2 shows the similarity of the three communities with the control channels. We have that the Jaccard similarity between the I.D.W. and the Alt-lite and the control channels is not so different from the similarity between these communities and the Alt-right. This is a subtle finding. On one hand, it means that individuals in these communities make up a significant portion of the massive media channels we use for control, which gather billions of views. These communities do not exist in a vacuum, but are part of the existing online information environment. On the other, it shows that the Alt-right, a group of channels with order of magnitudes fewer views, subscribers and comments, are actually *on par* with these large channels. Inspecting the Overlap Coefficient, however, we get a different view: there we have that the communities overlap more with themselves than with the control channels. Interestingly, for both similarity measurements we can see a sharp growth in the similarity with control after 2016. This may be explained due to the sharp increase in popularity of these communities since 2015, as discussed in Sec. 4.

These analyses take us one step further in understanding the communities being studied. We again see that their users are more engaged, and, notably, find that the I.D.W, the Alt-lite, and the Alt-right increasingly share the same commenting user base.
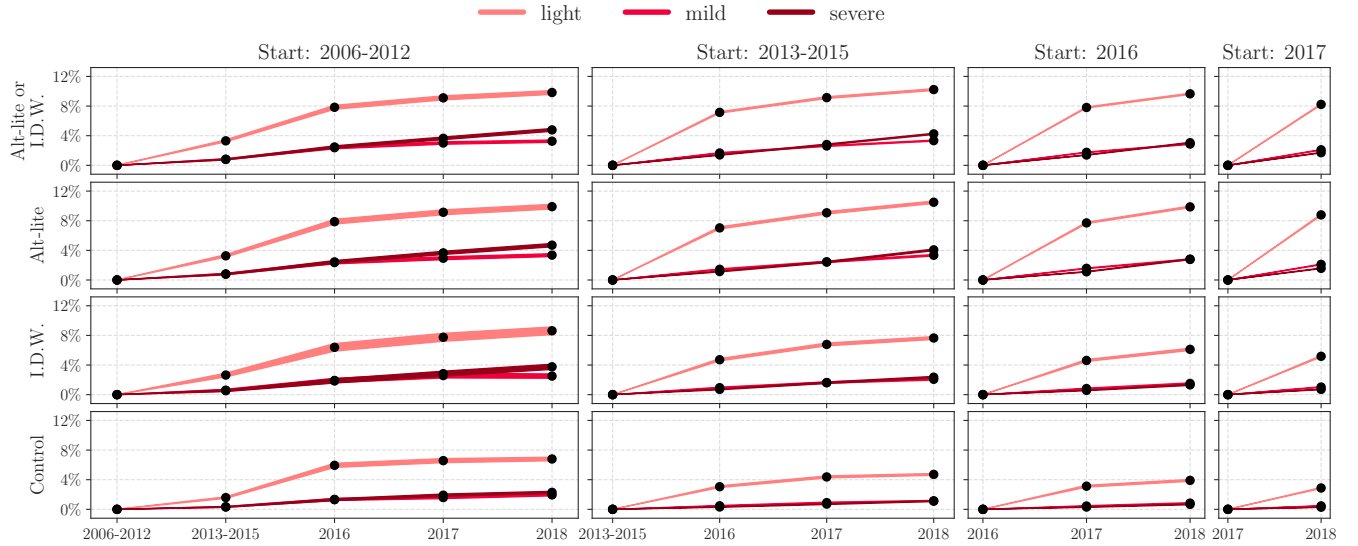
Figure 3: We show how users "migrate" towards Alt-right content. For users who consumed only videos in the communities indicated by the labels in the rows (Alt-lite or I.D.W., only Alt-lite, only I.D.W., and Control), we show the probability of them consuming Alt-right content. We consider three levels of "infection": light (commented in 1 to 2 Alt-right videos), mild (3 to 5) and severe (6+). Each column tracks users in a different starting date. Initially, their infection rates are 0 (as they did not consume any Alt-right content). As time passes, we show the infection rates in the y-axis, for each of the years, in the x-axis.

6 USER MIGRATION

In the previous section we showed that the commenting user bases among the I.D.W., the Alt-lite, and the Alt-right are increasingly similar —and the effect is stronger than for control channels. This indicates that there is a growing percentage of users consuming extreme (Alt-right) content on YouTube *while also* consuming content from other milder communities (Alt-lite/I.D.W.). Yet, it does not, *per se*, indicate that there is a radicalization pipeline in the website. It could be, for example, that new users who join the website go on to consume content from all three communities. To better address this question, we find users who *did not comment* in Alt-right content in a given year and track their subsequent activity. Notice that we do not have the user's entire activity history, and thus, we track their activity only in the channels whose videos we collected.

For four time brackets [(2006−2012), (2013−2015), (2016), (2017)] we track four sets of users: those who only commented in videos of the Alt-lite or the I.D.W, those who did so only for videos in the Alt-lite, those who did so only for videos in the I.D.W., and those who commented only in videos of the control channels. Then, for subsequent years, we track the same users. Notice that when users are tracked for one year they are not eligible for selection in upcoming years. We consider these users to be "infected" if they commented in 1-2 (light), 3-5 (mild) or 6+ (severe) Alt-right videos.

The results for this analysis are shown in Fig. 3. We show the percentage of users we managed to track that became infected. The number of users tracked and infected at each step may be found in Appendix C. Consider, for example, users who in 2006 − 2012 commented only on I.D.W. or Alt-lite content (227, 945 users), as shown in the subplot in the first column and the first row. By 2018,

around 10% were lightly infected, and roughly 4% severely or mildly so —which amounts to more than 9k users in total. From the ones who in 2017 commented only in Alt-lite or I.D.W. videos (1, 253, 751 users), as shown in the last column of the first row, approximately 12% of them became infected —more than 60k users altogether.

We also find that control channels present lower infection rates, as can be seen comparing the first and last row of the figure. The difference is particularly significant for the last three time brackets. Less than 1% of users in control became mildly or severely infected, while 3% to 4% of Alt-lite or I.D.W. users did so. Moreover, while in control around 4% of users eventually became lightly infected, more than 8% of Alt-lite or I.D.W users did so.

When teasing apart users that commented only on Alt-lite or only on I.D.W. content, we find that, not only users who commented only on content from the I.D.W. get less infected, but increasingly less so, as with the control channels. For example, the radicalization rates of users who watched only Alt-lite (second row) or only I.D.W. (third row) content are much more similar for those tracked in 2006 − 2012 (first column) than for those tracked in 2017 (last column). For users who were tracked in 2006 − 2012, around 15% became infected for both scenarios, while for those tracked in 2017, this difference grew farther apart (∼ 12% Alt-lite vs. ∼ 6% I.D.W.).

The previous study suggests that the pipeline effect does exist, and that indeed, users systematically go from milder communities to the Alt-right. However, it does not give insight into how expressive the effect is in terms of what part of the Alt-right user base has gone through it. We address this question by tracking users exactly as we did before, and then analyzing what percentage of "infected" users at each year can be traced back to users who initially watched content from other communities. In other terms, for each year we
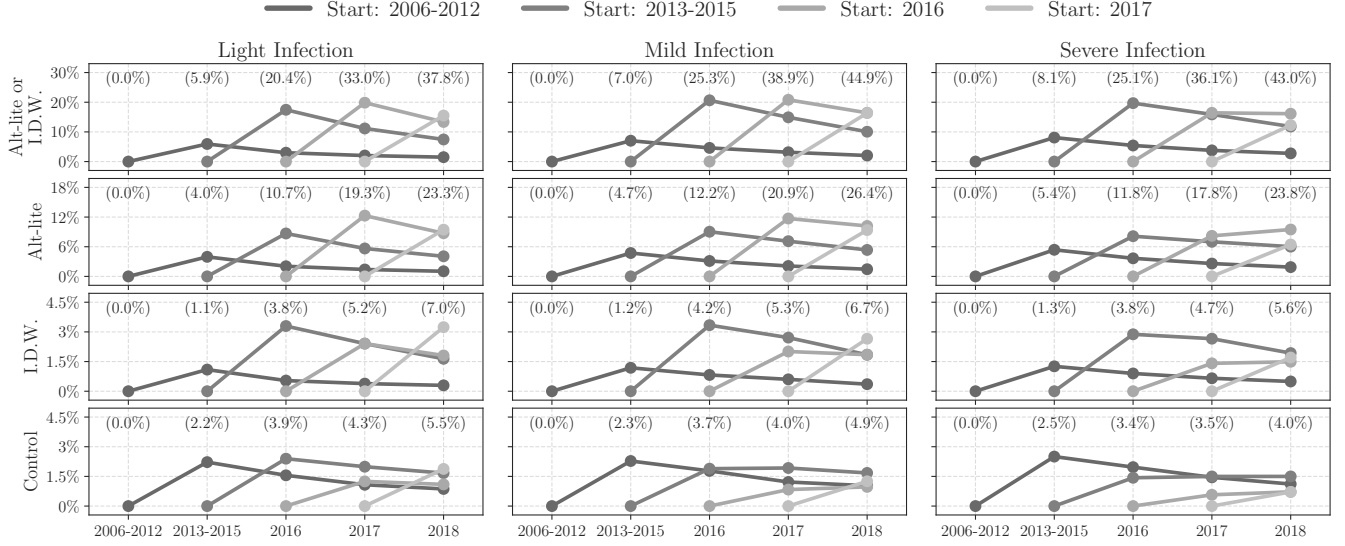
Figure 4: We show how expressive the tracked users are in terms of the Alt-right user base. Each row shows a the a different condition for tracking users and each column shows a different level of infection. Each line corresponds to users tracked at a different starting date (in the x-axis), and the y-axis shows the percentage of the total Alt-right commenting users they went to become (notice that all lines begin at 0, because initially they did not consume any Alt-right content).

calculate, of the users who are infected (i.e. who watched Alt-right videos), which percentage belongs to each one of the sets of tracked users we just described.

The results for this analysis are shown in Fig. 4. We find that these users are a considerable fraction of the Alt-right commenting audience. In 2018, for all kinds of infections, roughly 40% of commenting users can be traced back from cohorts of users that commented only in Alt-lite or I.D.W. videos in the past. This can be seen in the first row of the plot. Moreover, we can observe that, consistently, users who consumed Alt-lite or I.D.W. content in a given year, go on to become a significant fraction of the Alt-right user base in the following year. This number is much more expressive than the number of users which came from control channels —in the last row— which never surpasses 6% for all levels of infection.

Looking at the second and third row of Fig. 4, we find a substantial difference between the I.D.W. and the Alt-lite. Whereas in Sec. 5 we find that the intersection between them both and the Alt-right are very similar, here we see that users who initially commented *only* on I.D.W. channels constitute a much less significant percentage of the Alt-right consumer base in upcoming years. For all levels of infection, at all times, the number of infected users that can be traced back to commenting exclusively on I.D.W. channels is around 3 times lower. So, while in 2018, 23.3% of users who were lightly infected can be traced back to users who commented in Alt-lite channels in previous years, only 7.6% can be traced back to I.D.W. channels. Overall, in both analyses, users who consumed only I.D.W. channel seem to behave more similarly to the users in the control group. Yet, as we see in Sec. 5, the intersection between the Alt-lite and the I.D.W. is increasing with time, which means this population is becoming less significant.

The experiments performed show that, not only the commenting user bases are becoming increasingly similar (as shown in Sec. 5), but that, systematically, users who commented only in I.D.W. or Alt-lite content go on to comment on Alt-right channels. This phenomenon is significant both in terms of the percentage of the users tracked —as in Fig. 3— and in terms of the total Alt-right commenting user base —as in Fig. 4.

7 THE RECOMMENDATION ALGORITHM

In this section, we inspect the impact of YouTube's recommendation algorithm. Unfortunately, we have only a snapshot of the recommender system which does not take into account personalization. Thus, it is hard to reach significant conclusions on what was the role of the recommender system in the radicalization process we depicted in Sec. 6. Yet, we argue that analyzing these data is relevant, for it is a blueprint of how the influence of the recommender system may be measured, and because it allows us to understand how the recommender system is behaving for our scenario.

Table 3: Percentage of edges in-between communities in the recommendation graphs (normalized per weight). Video recommendations are in bold. Rows indicate the source of edges columns indicate their destination.

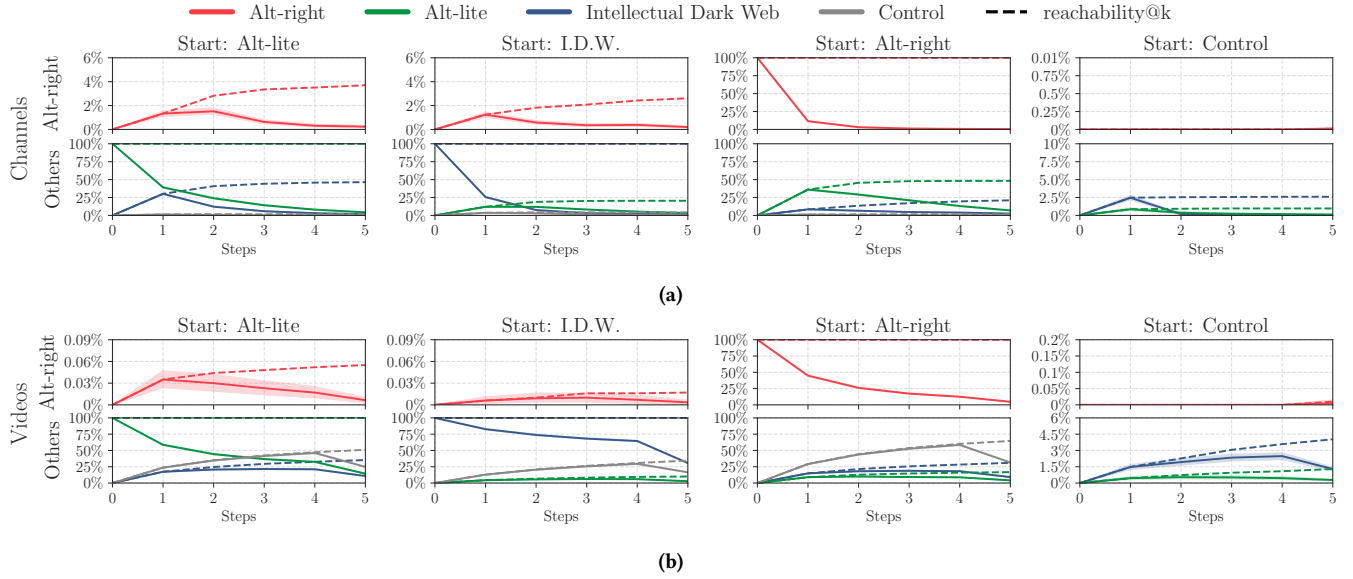| Src|Dst	I.D.W.	Alt-lite	Alt-right	Control	Other
I.D.W.	53.74/**19.07**	19.72/**1.59**	0.37/**0.03**	2.99/**3.02**	23.18/**76.29**
Alt-L	20.68/**2.43**	47.93/**12.52**	3.08/**0.13**	2.84/**3.27**	25.47/**81.65**
Alt-R	13.01/**1.89**	39.03/**1.15**	25.83/**8.55**	1.36/**3.38**	20.78/**85.03**
Control	5.00/**0.31**	4.41/**0.08**	0/**0**	29.12/**14.84**	61.47/**84.77**

(a)

(b)

Figure 5: We show the results for the simulation of random walks for channels (a) and videos (b). We show two metrics, as described in the text, the probability of the walker being in a given community at each step (solid line) and the reachability at each step for a given community (dashed line). The different columns portray different starting rules for the initial node in the simulations. Error bands are 95% confidence intervals.

We perform our analysis in a recommendation graph, built using the data collected. The graph is built as follows: for each channel, we join together all recommendations obtained in all rounds of data collection. Each channel is a node, and edges between nodes indicate recommendations from a channel to another (for both video and channel recommendations). Notice that, in case there was a recommendation towards a channel or a video we are not aware of, we add an edge to a special sink node we name "Other". Each edge is weighted proportionally to the number of times that recommendation appeared in the data collection, and weights are normalized so that outcoming edges of each node sum up to 1.

The percentage of edges flowing from each community to another (normalized by their weight) is shown in Tab. 3 for channel and video recommendations. For channel recommendations, we have that control channels are recommended scarcely by the communities of interest. In fact, there are more edges flowing *out* of control channels towards Alt-lite and I.D.W. channels than the other way around. Alt-lite and I.D.W. channels recommend channels from the same community around 50% of the time, and recommend each other around 20% of the time. Alt-right channels are not often recommended, but significantly more so than by the Alt-lite (3.08%), than by the I.D.W. (0.37%). For video recommendations, there is a high prevalence of recommendation to videos we were not able to track (more than 75% of outgoing edges from all communities pointed towards the "Other" node). We also find that control channels are more often recommended in this setting (around 3% for all other communities), and that the Alt-lite and the I.D.W. recommend each other roughly 2% of the time. Lastly, it is noteworthy that Alt-right videos are not significantly recommended here.

Given these graphs, we perform experiments considering a random walker. The random walker begins in a random node, chosen

with chance proportional to the number of subscribers in each channel. Then, the random walker randomly navigates the graph for 5 steps, choosing edges at random with probabilities proportional to their weights. We store the random walks and calculate two metrics. The first is the probability of it being in a channel from each of the communities, that is, the probability that there is channel of a given community in the k-th step. The second is the reachability of each of the communities at step k. That is, at step k, the percentage of times that the random walker has found a node of a given community. We run the simulation $10,000$ times for four starting scenarios. In each scenario, the initial node is restricted to one of the three communities or the control channels.

Importantly, we consider a small difference in the experimental set-up for each of the graphs. In the channel recommendation graph, we allow the random walker to choose the "Other" node. When this happens the walk stops, thus at each step there is a probability this walk is interrupted by this —or by the fact that there are no recommended channels. In the channel recommendation graph, as the number of edges to the "Other" node is too high, we do not allow the random walker to go towards it. Notice that the scenario for the channels is more realistic, and we give more weight to the conclusions drawn there. The two aforementioned metrics, at each step, given different starting conditions, are shown on Fig. 5, for channel and video recommendations.

For channel recommendations, we have that the reachability@5 of Alt-right channels is of approximately 4% for the simulations starting from Alt-lite and I.D.W. channels. Moreover, starting from an I.D.W. channel, users have approximately 10% of chance of being in an Alt-lite channel at the next step, and in 5 steps, there is 25% of chance that the user has found at least one Alt-lite channel. Starting from the Control channels, reachability@5 of I.D.W. channels is of

2.5%, and of slightly less than 1% for Alt-lite channels. These can be seen on the bottom row of Fig. 5 (a).

For video recommendations, reaching Alt-right channels from other communities is less likely. From the Alt-lite, reachability@5 is of around 0.06%. Going from the I.D.W. to the Alt-lite is more difficult: the reachability@5 is of roughly 10%. More relevant, though, starting from control channels, the reachability@5 of I.D.W. and Alt-lite channels is of around 4.5% and 1.5% respectively. It is worth recalling that this experiment is less realistic than the former, as here we ignore the possibility of the random walker being in a video we are not aware of.

Overall, we find that, particularly in the channel recommender system, it is easy to navigate from the I.D.W. to the Alt-lite (and vice-versa), and it is possible to find Alt-right channels. If we follow the recommender system 5 times, approximately 1 out of each 25 times we will have spotted an Alt-right channel. In the video recommender system, Alt-right channels are less recommended, but finding Alt-lite channels from the I.D.W. and I.D.W. channels from the large media channels in the control group is also feasible. Considering the sheer amount of views the channels in the Alt-lite, the I.D.W. and the Alt-lite, these percentages, although low, may result in a very significant number of views towards fringe content. This process may also be amplified when taking personalization into account. Notice that we depict the two graphs in which we performed our experiments in Appendix E.

8 DISCUSSION

We performed a throughout analysis of three YouTube communities —the I.D.W., the Alt-lite, and the Alt-right— inspecting a large dataset with millions of comments and recommendations from thousands of videos. In this section, we discuss how the insights of our analyses shed light into our research questions. We also talk about the limitations and potential implications of this work.

RQ1. How was the growth of these communities on YouTube in the last decade? The three communities studied sky-rocketed in terms of views, likes, videos published and comments, particularly, since 2015, coinciding with the turbulent presidential election of that year, as shown in Sec. 4. However, this seems to be the case not for these communities, but also for the larger channels in the control group. A difference between the communities and control channels lays in the engagement of their users. The number of comments per view seems to be particularly high for extreme content (Sec. 4), and users in all three communities are more assiduous commentators than in the control group (Sec. 5).

RQ2. To which extent do users systematically gravitate towards more extreme content? We find that the commenting user bases for the three communities are increasingly similar (Sec. 5), and, considering Alt-right channels as a proxy for extreme content, that a significant amount of commenting users systematically migrates from commenting exclusively on milder content to commenting on more extreme content (Sec. 7). We argue that this finding comprises significant evidence that there has been, and there continues to be, user radicalization on YouTube, and our analyses of the activity of these communities (Sec. 4) is consistent with the theory that more extreme content "piggybacked" the surge in popularity of

I.D.W. and Alt-lite content [30]. We show this migration phenomenon is not only consistent throughout the years, but also that it is significant in its absolute quantity. Noticeably, the findings related to this research question make the implicit assumption that commenting users are a good enough proxy for radicalization, and that comments in YouTube channels are supportive of the videos they are associated with. We find these to be minor issues. First, the sheer amount of comments and high prevalence of comments per views in Alt-right videos suggest that commenting users are a population worth studying, specially when in Sec. 4 we found that more Alt-right channels had a very high percentage of comments per view. Secondly, during the three week annotation period, it was noted that the number of opposing comments is rather small. We went as far as to manually check 900 randomly selected comments (300 for each community of interest), and found that only 5 could be interpreted as criticisms to the videos they were associated with. As identifying stance in short comments is a difficult research question on its own, we considered that commenting was a good enough proxy for our work.

RQ3. Do algorithmic recommendations steer users towards more extreme content? Our simulations suggest that YouTube's recommendation algorithms frequently suggest Alt-lite and I.D.W. content. From these two communities, it is possible to find Alt-right content from recommended channels, but not from recommended videos. Noticeably, our analysis has several shortcomings which do not allow us to make bold claims about this research question. Firstly, we are able to look only at a tiny fraction of actual recommendations —it could very well be that this content was being promoted in the past. Secondly, our analysis does not take into account personalization, which could reveal a completely different picture. Still, even without personalization, and after continuous backlash against YouTube in the last years, we were still able to find a path in which users could find extreme content from large media channels.

Implications. Our work resonates with the narrative that there is a radicalization pipeline [37, 42]. Indeed, we manage to measure traces of this phenomenon using commenting users. Yet, given that it is not possible to know how YouTube's recommender system worked in the past, and given that we did not take into personalization into account, our work is not able to determine the extent to which the recommender system played a role in this process. Regardless, we argue that *(i)* our analysis of the recommender system does show that this is indeed feasible, even in our simplistic scenario, and *(ii)* regardless of the role of the recommender system, user radicalization on YouTube seems to be significant.

Future Work. In this paper, we focused exclusively on basic statistics (likes, views and comments) and on the trajectory of users, be they inferred through comments or simulated in the recommendation graphs. Another interesting direction would be to trace the evolution of the *speech* of content creators and commenting users throughout the years: what are the narratives that arose, how did their tone change. Moreover, we intend to extend the existing framework to audit radicalization to take into account user personalization —this is not trivial, as knowing the trajectories of radicalized users through YouTube content is hard.

REFERENCES

[1] [n. d.]. Alt Right. https://www.adl.org/resources/glossary-terms/alt-right

[2] [n. d.]. How 2015 Fueled The Rise Of The Freewheeling, White Nationalist Alt-Movement. https://www.buzzfeednews.com/article/rosiegray/how-2015-fueled-the-rise-of-the-freewheeling-white-nationali

[3] [n. d.]. Media Bias/Fact Check - Search and Learn the Bias of News Media. https://mediabiasfactcheck.com/

[4] 2017. Jordan Peterson and the transgender wars. https://life.spectator.co.uk/2017/09/jordan-peterson-and-the-transgender-wars/

[5] 2019. From Alt Right to Alt Lite: Naming the Hate. https://web.archive.org/web/20190422202936/https://www.adl.org/resources/backgrounders/from-alt-right-to-alt-lite-naming-the-hate

[6] Swati Agarwal and Ashish Sureka. 2014. A Focused Crawler for Mining Hate and Extremism Promoting Videos on YouTube.. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media (HT '14)*. ACM, New York, NY, USA, 294–296. https://doi.org/10.1145/2631775.2631776 event-place: Santiago, Chile.

[7] Monica Anderson and Jingjing Jiang. 2018. *Teens, Social Media & Technology 2018*. Technical Report. Pew Research Center. https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/

[8] Andywarski. [n. d.]. Richard Spencer, Styx and Sargon Have a Chat - Andy and JF moderate. https://web.archive.org/web/20190616134137/https://www.youtube.com/watch?v=UiUH-tWHbr8

[9] Anonymous. [n. d.]. The Intellectual Dark Web. https://web.archive.org/web/20190407170300/http://intellectualdark.website/

[10] David C. Atkinson. 2018. Charlottesville and the alt-right: a turning point? *Politics, Groups, and Identities* 6, 2 (April 2018), 309–315. https://doi.org/10.1080/21565503.2018.1454330

[11] Khaled A. Beydoun. 2018. US liberal Islamophobia is rising – and more insidious than rightwing bigotry | Khaled A Beydoun. *The Guardian* (May 2018). https://www.theguardian.com/commentisfree/2018/may/26/us-liberal-islamophobia-rising-more-insidious

[12] in Campaigns, Elections, Parties, Citizen Action, Interest Groups, Papers, Politics, Government, Research, and Technology. [n. d.]. "Alt-Lite" Bloggers and the Conservative Ecosystem. https://shorensteincenter.org/alt-lite-bloggers-conservative-ecosystem/

[13] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 191–198. https://doi.org/10.1145/2959100.2959190 event-place: Boston, Massachusetts, USA.

[14] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. 2010. The YouTube Video Recommendation System. In *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10)*. ACM, New York, NY, USA, 293–296. https://doi.org/10.1145/1864708.1864770 event-place: Barcelona, Spain.

[15] Nicholas Diakopoulos. 2014. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. (2014). https://doi.org/10.7916/D8ZK5TW2

[16] Benjamin G. Edelman and Michael Luca. 2014. *Digital Discrimination: The Case of Airbnb.com*. SSRN Scholarly Paper ID 2377353. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=2377353

[17] Lisa W. Foderaro. 2018. Alexandria Ocasio-Cortez Likens $10,000 Debate Offer by Conservative Columnist to Catcalling. *The New York Times* (Nov. 2018). https://www.nytimes.com/2018/08/10/nyregion/alexandria-ocasio-cortez-debate-catcalling-ben-shapiro.html

[18] T. Giannakopoulos, A. Pikrakis, and S. Theodoridis. 2010. A Multimodal Approach to Violence Detection in Video Sharing Sites. In *2010 20th International Conference on Pattern Recognition*. 3244–3247. https://doi.org/10.1109/ICPR.2010.793

[19] Keegan Hankes and Alex Amend. [n. d.]. The Alt-Right is Killing People. https://www.splcenter.org/20180205/alt-right-killing-people

[20] Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring Personalization of Web Search. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 527–538. https://doi.org/10.1145/2488388.2488435 event-place: Rio de Janeiro, Brazil.

[21] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring Price Discrimination and Steering on E-commerce Web Sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, New York, NY, USA, 305–318. https://doi.org/10.1145/2663716.2663744 event-place: Vancouver, BC, Canada.

[22] Mathew Ingram. 2018. Most Americans say they have lost trust in the media. *Columbia Journalism Review* (Sept. 2018). https://www.cjr.org/the_media_today/trust-in-media-down.php

[23] Brendan Joel Kelly. 2017. Lauren Southern: The alt-right's Canadian dog whistler. https://www.splcenter.org/hatewatch/2017/11/07/lauren-southern-alt-right%E2%80%99s-canadian-dog-whistler

[24] Max Kutner. 2016. Roosh V's journey from pickup artist to right-wing provocateur. *Newsweek* (Oct. 2016). https://www.newsweek.com/2016/10/21/roosh-v-pickup-artist-right-wing-provocateur-509319.html

[25] Rebecca Lewis. 2018. *Alternative influence: Broadcasting the reactionary right on YouTube*. Technical Report. Data and Society.

[26] Cristina Lopez G. 2019. Stefan Molyneux is MAGA Twitter's favorite white nationalist. https://www.mediamatters.org/people/stefan-molyneux

[27] Alex Mann, Kevin Nguyen, and Katherine Gregory. 2019. 'Emperor Cottrell': Accused Christchurch shooter had celebrated rise of the Australian far-right. https://www.abc.net.au/news/2019-03-23/christchurch-shooting-accused-praised-blair-cottrell/10930632

[28] Andrew Marantz. 2017. The Alt-Right Branding War Has Torn the Movement in Two. (July 2017). https://www.newyorker.com/news/news-desk/the-alt-right-branding-war-has-torn-the-movement-in-two

[29] Clark McCauley and Sophia Moskalenko. 2008. Mechanisms of Political Radicalization: Pathways Toward Terrorism. *Terrorism and Political Violence* 20, 3 (July 2008), 415–433. https://doi.org/10.1080/09546550802073367

[30] Angela Nagle. 2017. *Kill All Normies: Online Culture Wars From 4Chan And Tumblr To Trump And The Alt-Right*. John Hunt Publishing.

[31] Newman Nic, Richard Fletcher, Antonis Kalogeropoulos, David AL Levy, and Rasmus Kleis Nielsen. 2018. *Reuters Institute Digital News Report 2018*. Technical Report. Reuters Institute for the Study of Journalism.

[32] Raphael Ottoni, Evandro Cunha, Gabriel Magno, Pedro Bernardina, Wagner Meira Jr., and Virgílio Almeida. 2018. Analyzing Right-wing YouTube Channels: Hate, Violence and Discrimination. In *Proceedings of the 10th ACM Conference on Web Science (WebSci '18)*. ACM, New York, NY, USA, 323–332. https://doi.org/10.1145/3201064.3201081 event-place: Amsterdam, Netherlands.

[33] Kostantinos Papadamou, Antonis Papasavva, Savvas Zannettou, Jeremy Blackburn, Nicolas Kourtellis, Ilias Leontiadis, Gianluca Stringhini, and Michael Sirivianos. 2019. Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children. *arXiv:1901.07046 [cs]* (Jan. 2019). http://arxiv.org/abs/1901.07046 arXiv: 1901.07046.

[34] PowerfulJRE. [n. d.]. Jordan Peterson - Joe Rogan Experience #1208. https://web.archive.org/web/20190401093845/https://www.youtube.com/watch?v=vIeFt88Hm8s

[35] PowerfulJRE. [n. d.]. Sargon of Akkad - Joe Rogan Experience #979. https://web.archive.org/web/20190616134332/https://www.youtube.com/watch?v=xrBCsLsSD2E

[36] RedIceTV. [n. d.]. NPI 2016 Panel Discussion: Jared Taylor, Peter Brimelow, Kevin MacDonald & Millennial Woes. https://web.archive.org/web/20190616134429/https://www.youtube.com/watch?v=aWr2NUvbcVI

[37] Kevin Roose. 2019. The Making of a YouTube Radical. *The New York Times* (June 2019). https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html,https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

[38] Roosh V. 2016. I Do Not Disavow Richard Spencer. https://web.archive.org/web/20190331073037/https://www.rooshv.com/i-do-not-disavow-richard-spencer

[39] Andrew Sellars. 2016. *Defining Hate Speech*. SSRN Scholarly Paper ID 2882244. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=2882244

[40] Ashish Sureka, Ponnurangam Kumaraguru, Atul Goyal, and Sidharth Chhabra. 2010. Mining YouTube to Discover Extremist Videos, Users and Hidden Communities. In *Information Retrieval Technology (Lecture Notes in Computer Science)*, Pu-Jen Cheng, Min-Yen Kan, Wai Lam, and Preslav Nakov (Eds.). Springer Berlin Heidelberg, 13–24.

[41] The Rubin Report. 2018. Eric Weinstein: The Future of The Intellectual Dark Web. https://www.youtube.com/watch?v=tUl7-SvntQ4

[42] Zeynep Tufekci. 2018. Opinion | YouTube, the Great Radicalizer. *The New York Times* (June 2018). https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

[43] Bari Weiss and Damon Winter. 2018. Opinion | Meet the Renegades of the Intellectual Dark Web. *The New York Times* (May 2018). https://www.nytimes.com/2018/05/08/opinion/intellectual-dark-web.html

[44] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. 2018. On the Origins of Memes by Means of Fringe Web Communities. In *Proceedings of the Internet Measurement Conference 2018 (IMC '18)*. ACM, New York, NY, USA, 188–202. https://doi.org/10.1145/3278532.3278550 event-place: Boston, MA, USA.

[45] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtelris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2017. The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources. In *Proceedings of the 2017 Internet Measurement Conferen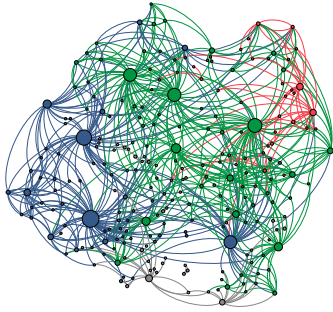ce (IMC '17)*. ACM, New York, NY, USA, 405–417. https://doi.org/10.1145/3131365.3131390 event-place: London, United Kingdom.

Figure 6: Recommendation graph of YouTube channels.

Figure 8: Example of YouTube channel with featured channel on the side.
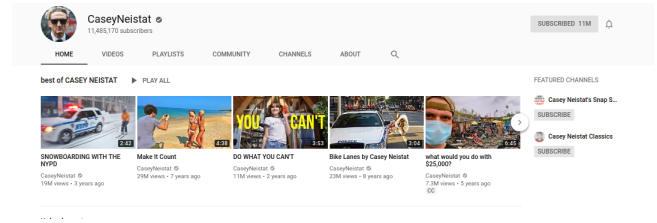
For the I.D.W. *Stephen Hicks, Camille Paglia, Carl Benjamin, Elon Musk, Akira the Don, Nicholas Christakis, Claire Lehmann, Matt Christiansen, Steven Pinker, Rebel Wisdom, Tim Pool, Quillette, Jonathan Haidt, Peter Thiel, Lindsay Shepherd, James Damore*

For the Alt-lite *Brittany Pettibone, Jack Posobiec, Gavin McInnes, Kyle Chapman, Kyle Prescott, Lucian Wintrich, Mike Cernovich, Milo Yiannopoulos, Stefan Molyneaux, Vee, Blonde in the Belly of the Beast, Paul Joseph Watson, Styxhenxenhammer666, Rebel Media, Lauren Chen, Computing Forever, Andy Warski, Owen Benjamin, Steven Crowder*

For the Alt-right *Evola, Evropa, The Jewish Question, White Genocide, Mass immigration, Andrew Anglin, weev, Andy Nowicki, Augustus Invictus, Christopher Cantwell, Collin Liddell, Daniel J. Kleve, Daniel Friberg, Dillon Irizarry, Greg Johnson, Jared Taylor, Jason Kessler, Jason Reza Jorjani, Johnny Monoxide, Lana Lokteff, Matt Forney, Matthew Heimbach, Matthew Parrott, Mike Enoch, Nathan Damigo, Pax Dickinson, Richard Spencer, Tara McCarthy, Vox Day, Baked Alaska*

B FEATURED VS RECOMMENDED

We illustrate the difference between featured and recommended channel. In Fig. 8 you may see an example of featured channels, these are chosen by the channel owner. In Fig. 9, letter (e) shows related channels, these are recommendations made by YouTube.

C LIKES, VIDEOS, VIEWS, COMMENTS

Tab. 4 shows, for the three communities, the number of likes, views, videos and commenting users accross the years.

D USER TRAJECTORIES

Tab. 5 shows the absolute numbers of users tracked and infected (at all levels, as mentioned in Sec. 6. It also shows what percentage of the total number of users who watched Alt-right the number of users infected was.

E RECOMMENDATION GRAPHS

In Figs. 6 and 7 we show the recommendation graphs used for the experiment in Section 7.

Figure 7: Recommendation graph of YouTube videos. Colors for communities are the same as those in the paper.

A DATA COLLECTION

We give some details in the data collection process. Tab. 6 and Tab. 7 show for channels labeled as Alt-right, Alt-lite and I.D.W., their communities and data collection steps. Tab 8 shows all control channels we obtained. Fig. 9 highlights what was collected on YouTube. Below, we enumerate the keywords employed to search for channels of each of the communities:

Table 4: For all categories, we list the number of likes, views, videos and commenting users across the years.

Category	Year	Like Count	View Count	Video Count	Comment Count
Alt-lite	2008	272639	18145720	1392	129130
	2009	585060	32993863	929	197934
	2010	503744	30519109	1498	248432
	2011	527221	28400257	2344	236961
	2012	805166	61779929	4142	360688
	2013	1227089	100673417	2468	529538
	2014	1973844	137877955	3375	608113
	2015	6561782	352751031	7132	2041333
	2016	24384665	931704328	9324	7982815
	2017	51764529	1330158802	15927	11502964
	2018	52323713	1051654691	20784	20513250
Alt-right	2008	559	73159	29	332
	2009	8389	1236895	313	1135
	2010	14124	1897892	363	2136
	2011	15992	1778120	174	6076
	2012	75092	4925868	434	10452
	2013	160494	11219639	654	25825
	2014	233381	12718956	725	47032
	2015	434925	17148672	958	127319
	2016	1411778	44177307	2334	529821
	2017	4253888	108482909	3548	1262549
	2018	5773031	106455102	5843	2725573
Control	2008	348137	128986765	2115	7932
	2009	511468	196992273	3939	41492
	2010	573299	203399250	6531	94379
	2011	1824078	350120542	12748	200385
	2012	3432239	454969357	25716	447302
	2013	5238196	716009326	18135	756691
	2014	9217725	1538251895	18836	814124
	2015	16569182	2015671151	24168	830655
	2016	27807514	2481994316	30119	1317648
	2017	46467022	3102590498	35678	2931209
	2018	54106314	2997876294	30951	13667470
I.D.W.	2008	54185	7034287	447	5263
	2009	61340	8661426	488	11249
	2010	135205	15457288	549	29932
	2011	269021	22797055	884	129453
	2012	365241	23918023	1067	154322
	2013	1095128	109814034	2528	229144
	2014	3072296	256213378	2504	588068
	2015	4975362	310760188	2324	1610366
	2016	9515078	518317670	3497	3062565
	2017	22103578	1201736706	8954	5133976
	2018	28079111	1620909482	14981	11904865

Table 5: We show absolute numbers for users infected and tracked in Sec. 6, as well as what percentage of the total number of users who watched Alt-right the number of users infected was.

Category	Start	Year	Number of Users Infected	Number of Users Tracked	Percentage of Users Alt-right
Alt-lite	2006-2012	2006-2012	0	170301	0.0%
		2013-2015	2133	43895	14.07%
		2016	3426	27052	8.83%
		2017	4557	28960	6.11%
		2018	6186	34447	4.37%
	2013-2015	2013-2015	0	417113	0.0%
		2016	12319	128093	25.84%
		2017	16402	117689	19.76%
		2018	22819	127464	15.43%
	2016	2016	0	719232	0.0%
		2017	31308	301559	32.2%
		2018	45017	291136	28.41%
	2017	2017	0	779316	0.0%
		2018	44117	354039	25.27%
Alt-lite or I.D.W.	2006-2012	2006-2012	0	227945	0.0%
		2013-2015	3194	64913	21.05%
		2016	5016	39539	13.02%
		2017	6644	42158	8.94%
		2018	8896	49766	6.29%
	2013-2015	2013-2015	0	696899	0.0%
		2016	25881	253479	57.8%
		2017	33586	230700	41.96%
		2018	42691	239787	29.32%
	2016	2016	0	1041550	0.0%
		2017	52621	480584	57.1%
		2018	70906	455166	45.91%
	2017	2017	0	1253751	0.0%
		2018	74622	620581	44.14%
Control	2006-2012	2006-2012	0	248214	0.0%
		2013-2015	1136	50753	7.0%
		2016	2331	27175	5.3%
		2017	3122	31001	3.75%
		2018	4628	41941	3.01%
	2013-2015	2013-2015	0	637009	0.0%
		2016	3129	81357	5.72%
		2017	5150	86027	5.41%
		2018	8110	116368	4.84%
	2016	2016	0	365592	0.0%
		2017	2928	75539	2.65%
		2018	4998	92312	2.79%
	2017	2017	0	696089	0.0%
		2018	7792	214516	3.85%
I.D.W.	2006-2012	2006-2012	0	47914	0.0%
		2013-2015	566	14961	3.54%
		2016	889	8750	2.26%
		2017	1240	9426	1.64%
		2018	1687	11329	1.15%
	2013-2015	2013-2015	0	211977	0.0%
		2016	4621	72519	9.51%
		2017	6729	67145	7.78%
		2018	8529	70630	5.41%
	2016	2016	0	232102	0.0%
		2017	5937	98623	5.83%
		2018	8702	97275	5.15%
	2017	2017	0	419833	0.0%
		2018	14241	205895	7.6%

Table 6: For the three communities, we list all the websites analysed in this paper (part 1).

	Alt-right channels	Step	Alt-lite channels	Step	I.D.W. channels	Step
0	AltRight.com	1	America First with Nicholas J Fuentes	1	Ben Shapiro	1
1	AmRen Podcasts	1	Andy Warski	1	Bret Weinstein	1
2	AmRenVideos	1	Blonde in the Belly of the Beast	1	Gad Saad	1
3	Ayla Stewart Wife With A Purpose	1	Brittany Pettibone	1	JRE Clips	1
4	Baked Alaska 2	1	Computing Forever	1	Jordan B Peterson Clips	1
5	Black Pigeon Speaks	1	Gavin McInnes	1	JordanPetersonVideos	1
6	Bre Faucheux	1	Laura Loomer	1	Lindsay Shepherd	1
7	CounterCurrentsTV	1	Lauren Chen	1	Matt Christiansen	1
8	Darkstream	1	Lauren Southern	1	Owen Benjamin	1
9	Faith J Goldy	1	MILO	1	Owen Benjamin Clips	1
10	James Allsup	1	Mike Cernovich	1	PowerfulJRE	1
11	Jason Kessler	1	Nick Fuentes Clips	1	Rebel Wisdom	1
12	Jean-François Gariépy	1	No Bullshit	1	Sam Harris	1
13	Johnny Monoxide	1	No Bullshit 2	1	SargonofAkkad100	1
14	MW Live	1	Paul Joseph Watson	1	The Rubin Report	1
15	Matt Forney	1	Rebel Canada	1	joerogandotnet	1
16	MillennialWoes	1	Rebel Edge	1	1791	2
17	NPI / Radix	1	Rebel Media	1	American Justice	2
18	Red Ice TV	1	Stefan Molyneux	1	Atheist Foundation of Australia Inc	2
19	Staying Woke	1	StevenCrowder	1	AynRandInstitute	2
20	The Golden One	1	Styxhexenhammer666	1	Ben Shapiro Thug Life	2
21	The Reality Calls Show	1	The Thinkery	1	Benjamin A Boyce	2
22	Traditionalist Worker Party	1	Vee	1	Brother Nathanael	2
23	Voxiversity	1	6oodfella	2	CISAus	2
24	augustussolinvictus	1	A1Cvenom	2	Clash of Ideas	2
25	iambakedalaska	1	AIU-Resurrection	2	Conversations with Bill Kristol	2
26	Alt Right	2	AltRight Truth	2	Crysta	2
27	Alt-Right Tankie	2	AustralianNeoCon1	2	Desi-Rae Thinking	2
28	American Pride	2	BlazeTV	2	Douglas Murray Archive	2
29	American Pride 2	2	Brave New World	2	Enlightainment	2
30	ArktosOnline	2	Bull Brand	2	Essential Truth	2
31	Augustus Invictus for United States Senate	2	Carpe Donktum	2	Freedom Speaks	2
32	AustralianRealist	2	Christopher Anderson	2	Glenn Beck	2
33	Be Open MInded	2	Daily Caller	2	Gravitahn	2
34	BigCatKayla Livestreams	2	DailyCallerVideo	2	Informative	2
35	Charles Zeiger	2	DailyKenn	2	Jordan Peterson Fan Channel	2
36	Corpus Mentis	2	Dinesh D'Souza	2	Liberty us	2
37	Dismantle The Matrix	2	DoctorRandomercam	2	MG	2
38	Dissident View	2	Domination Station	2	Maximilien Robespierre	2
39	Engländer	2	Harrison Hill Smith	2	MeaningofLife.tv	2
40	Jan Kerkoff	2	Jacob Wohl	2	Mike Nayna	2
41	Mark Collett	2	Kelly Day	2	Motte & Bailey	2
42	Matthew North	2	Leo Stratton	2	MrAndsn	2
43	Nacionalista Blanco del SoCal	2	Liberty Machine News	2	Notes For Space Cadets	2
44	Nationalist Media Network	2	Luke Ford	2	Pangburn	2
45	No White Guilt	2	Luke Ford Livestreams	2	PhilosophyInsights	2
46	Patrick Slattery	2	Make Cringe Great Again	2	Pragmatic Entertainment	2
47	Real McGoy	2	News2Share	2	ReasonTV	2
48	Revcon Media	2	On The Offensive	2	Savage Facts	2
49	Stand Up Europe	2	Oppressed Media	2	The Daily Truth	2
50	Steve Trueblue	2	Revenge Of The Cis	2	The Free Speech Club	2
51	The Alternative Hypothesis	2	RobinHoodUKIP	2	The Heritage Foundation	2
52	The Great Dolemite	2	SJW CRINGE MACHINE	2	The New Criterion	2
53	The James Delingpole Channel	2	SJWCentral	2	The Pondering Primate	2
54	The Last Stand	2	Semiogogue	2	The Unplugged Observer	2
55	The Rational Rise	2	Social Justice Fails	2	TheArchangel911	2
56	TheArmenianNation	2	The Fallen State	2	TheAtlasSociety	2
57	This is Europa	2	The Glass Blind Spot	2	Transliminal	2

Table 7: For the three communities, we list all the websites analysed in this paper (part 2).

	Alt-right channels	Step	Alt-lite channels	Step	I.D.W. channels	Step
58	ThuleanPerspective	2	The Hateful Gaels	2	Trigger Happy Media	2
59	Traditionalist Youth Network	2	The Iconoclast	2	VikNand	2
60	Truth Against The World	2	TheSchillingShow	2	Washington Watch	2
61	WhiteRabbitRadioTV	2	Tipping Point With Liz Wheeler on OAN	2	WisdomTalks	2
62	andy nowicki	2	Tommy Robinson	2	YAFTV	2
63	eliharman	2	Tree Of Logic	2	ZIEeICoZ	2
64	jackburton2009	2	UNITE AMERICA FIRST	2	ZeroFox Given	2
65	nightmarefuel	2	Western Man	2	battleofideas	2
66	14 Sacred Words	3	Zach Hing	2	bloggingheads.tv	2
67	Awakened Saxon	3	grapjas60	2	bmdavll	2
68	Borzoi Boskovic	3	hOrnsticles3	2	successcouncil	2
69	Danny 1488	3	ramzpaul	2	tmcleanful	2
70	InvincibleNumanist	3	theovonk	2	wikileaksplus	2
71	LaughingMan0X	3	theturningpointusa	2	xUnlimitedMagz	2
72	Laura Towler	3	thkelly67	2	ybrook	2
73	LibertarianRealist2	3	Actual Justice Warrior	3	AgatanFoundation	3
74	Little Revolution	3	AllNationsParty	3	Bite-sized Philosophy	3
75	Marie Cachet	3	Alt Hype Streams	3	CoolHardLogic	3
76	Morrakiu	3	Aydin Paladin	3	Davie Addison	3
77	NeatoBurrito Productions	3	Beacom Of Light	3	Dose of Truth	3
78	NewEuropeANP	3	Bearing	3	DronetekPolitics	3
79	OnlineWipe	3	Count Dankula	3	Galactic Bubble Productions	3
80	Oswald Spengler	3	Dangerfield	3	HowTheWorldWorks	3
81	Prince of Zimbabwe	3	Demirep	3	ManOfAllCreation	3
82	Serp Kerp	3	Dr. Steve Turley	3	PragerUniversity	3
83	TRS Radio	3	IRmep Stream	3	Rekt Idiots	3
84	The Leftovers	3	Jericho Green	3	Sinatra_Says	3
85	The Lion	3	John Ward	3	Sorting Myself Out	3
86	The Revolutionary Conservative	3	JustInformed Talk	3	The Andrew Klavan Show	3
87	VertigoPolitix	3	Liberty Hangout	3	The Daily Wire	3
88			MR. OBVIOUS	3	The Heartland Institute	3
89			MarkDice	3	The Propertarian Institute	3
90			MichelleRempel	3	Timcast	3
91			Mister Metokur	3		
92			NateTalksToYou	3		
93			OneTruth4Life	3		
94			ProductiehuisEU	3		
95			Reverend Simon Sideways	3		
96			Sanity 4 Sweden	3		
97			Sargon of Akkad Live	3		
98			SkidRowRadio	3		
99			Slightly Offens*ve	3		
100			Tea Clips	3		
101			The Amazing Lucas	3		
102			The Weekly Sweat	3		
103			TheBechtloff	3		
104			TheIncredibleSaltMine	3		
105			Toad McKinley	3		
106			TokenLibertarianGirl	3		
107			Undoomed	3		
108			Vincent James of The Red Elephants	3		
109			ataxin	3		
110			brianoflondon	3		
111			jaydyer	3		
112			libertydollshouse	3		
113			patcondell	3		

Table 8: Control communities

	Left	Center	Left-Center	Right-Center	Right
0	cosmopolitan	big think	(the)atlantic	forbes	american enterprise institute
1	democracy now	c-span	business insider	gulf news	judicial watch
2	elite daily	consumer reports	cbc news	learn liberty	national rifle association (nra)
3	good magazine	financial times	engadget	new york post	pj media
4	gq magazine	harvard business review	feminist frequency	ntd.tv (new tang dynasty)	project veritas
5	huffington post (huffpost)	investopedia	glamour magazine	russia insider	ron paul liberty report
6	mashable	makeuseof	global citizen		
7	merry jane	mental floss	global news		
8	new york magazine	military.com	hollywood reporter		
9	new yorker	recode	la times		
10	people magazine	relevant magazine	lifehacker		
11	slate	the economist	new york daily news		
12	uproxx	the indian express	rolling stone		
13	upworthy	today i found out	san francisco globe		
14	vanity fair	vocativ	scoopwhoop		
15	vox	world economic forum	scroll.in		
16			sky news		
17			techcrunch		
18			the guardian		
19			the verge		
20			vice news		
21			washington post		
22			wired magazine		
23			yahoo news		

Figure 9: Overview of the elements we collected: (a) video captions, when available, (b) video recommendations, (c) video description and metadata, (d) comments, (e) channel recommendations, and (f) video metadata.