Polarization and Radicalization in a Social Media Driven Society

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Ryan Robinson

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Pedro A. P. Francisco, Department of Engineering and Society

Introduction:

In the past two decades, the internet and social media have become powerful tools that have facilitated the spread of information in many new ways. Humanity is more connected than we have ever been, and while this has had many positive effects, it also has had consequences. The majority of people are connected through a select few number of platforms, and the way these platforms drive interaction has drastic effects on how individuals may perceive their society and the world around them.

This has altered, among other things, the modern political sphere. As social media has ingrained itself more and more into the fabric of our society, we have witnessed both the emergence of more radical political groups, and a general polarization of the political sphere. Because of this, we need to know the full scope of the problem and its effects, as well as what attributes of the digital architectures are causing these trends.

It is essential to understand both the scale and methodology of these platforms' influence, as it can not only lead to a divisive political atmosphere, but also extreme radicalization and even physical acts of terror. We must dive deeper and ask ourselves what the effects are of these social media algorithms, and what drives them? In their endless pursuit of maximizing engagement, I believe they have unwittingly caused an increase in both polarization and radicalization in our society.

Research Methods:

For my research, I decided to use literary review. The problem I am discussing has been researched in different ways before, and my primary purpose with this paper is to compile a lot of these sources and provide a clear view into the issue using a variety of already conducted research. Unfortunately, because the problem is at such a large scale, it would be difficult for me

to do my own studies. For conclusive results, I would need to collect data from at a minimum a few thousand participants. Therefore, I believe it is best for me to primarily search out research from groups with more resources where studies like this have already been conducted. I can then compile the results of all the studies I've found to paint a better picture of the problem. Once I have enough sources, I will describe how all related actors are connected using Actor Network Theory. This is the best way to illustrate the problem given the complex interactions between human, non-human, and organizational entities.

Background and Significance:

From the moment you create a new account on Twitter, Instagram, or any other platform, the algorithm immediately begins tracking every aspect of your interaction with it. Very quickly, it begins to learn your likes and dislikes so it can show you exactly what you want in order to maximize the time you spend on it. This isn't inherently a bad thing. If you're into fitness and sports, and your social media of choice learns this and begins to recommend you new content related to this interest, the app is much more useful for you. However, this behavior can have negative side effects when dealing with politically charged content. If, for example, you like a politically charged tweet, the algorithms will immediately begin to feed you content with the same political swing. If you continue to engage with this content, sooner or later most recommended political content will be one of two things. Either content with the same political tilt as yours, or content from the other side that is so disagreeable and upsetting to you that you will feel obligated to interact with it. You are now in what is called an "echo chamber," where you are very unlikely to encounter rational opinions from another political viewpoint. This can create a very distorted view of the political atmosphere, making you believe that everyone with

your political views is rational and everyone else irrational (Allcott et al., 2020). This is terrible for legitimate political discourse, and makes these political environments much more hostile.

Because of the environment social media has created, people now are more apprehensive towards the other side of the aisle. Luckily, for most people, this doesn't significantly impact their everyday life. However, for a few, it can spiral into something far more sinister: "The radicalization pipeline." While for the majority of individuals, validating one's preexisting beliefs is an excellent way to improve engagement, for some, user engagement can be raised even higher by offering increasingly radical content to users who show a propensity to interact with it. The radicalization pipeline is particularly insidious because it doesn't just recommend this content to users who seek it out. Instead, it silently classifies users vulnerable to the engagement technique, often by analyzing innocent uses of the platform, and subtly pushes them towards more extreme content without their explicit knowledge or consent (Rauchfleisch & Kaiser, 2020). This increases the probability of users being drawn into extremist echo chambers where they are exposed to escalating levels of radical information. Users who have fallen deep enough into the rabbit hole may suffer deteriorating mental health, lose relationships, or even commit extreme acts (Nagourney et al., 2014).

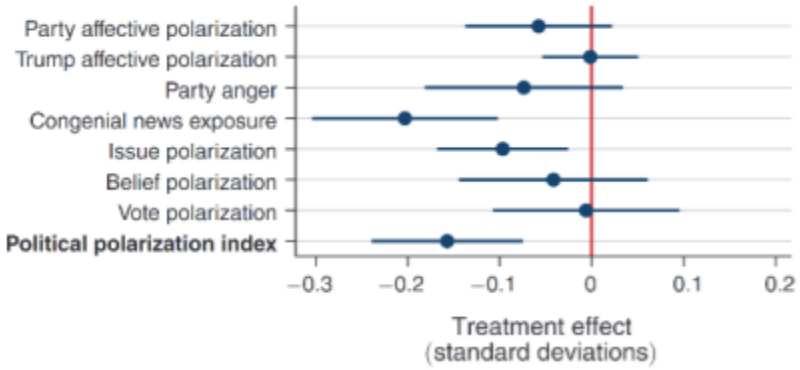
Results and Discussion:

To analyze the causes and effects of social media algorithms on political polarization and radicalization, it is imperative that we first establish with evidence that Social media does contribute to radicalization in our society. While key figures in the tech industry, including Facebook's vice president of global affairs, Nick Clegg and Facebook CEO Mark Zuckerberg, have stated that other factors, like the American media climate, are the primary drivers of polarization, they clearly have a vested interest in deflecting blame away from their business, as

seen in the transcripts from their hearing before congress (Sanford, 2021). While they are accurate in pointing out that there are clearly numerous elements that have contributed to today's political division, I believe there is enough evidence to infer that social media algorithms are a primary driver in this development.

Figure 1

Average Change in Polarization After 4 Weeks Without Social Media



We can directly observe the effects of social media on polarizing views (Allcott et al., 2020). In this study, 4 weeks prior to the 2018 election, 2743 randomized participants deactivated facebook, decreased general internet usage, and increased offline activities. After the 4 weeks, several different metrics measuring polarization were recorded for every participant. The most important recorded metric shown in Figure 1 is “issue polarization,” as it measures to what extent the participants’ opinion on certain issues lines up with the opinion of others in their party. This allows us to observe how far their opinion drifted from the original “echo chamber” they were in after 4 weeks of non-exposure. As we can see in the graphic below, issue polarization measurably decreased by -0.15 standard deviations after the 4 weeks. This data

presents strong evidence that using Facebook discourages variance of opinion and encourages polarization among differing political crowds.

These findings are backed up by another study (Levy, 2021) done in 2019 that states that our shift to primarily consuming news via social media combined with social media's reluctance to share rational sources from the other side has heavily increased polarization in America. With these two sources, we have enough evidence to conclude that Social Media has increased polarization in America and we can now place social media at the heart of our actor-network model.

Now we need to establish that polarization is in fact encouraged by the social media algorithms and why. We can start with the "why," as it is quite simple. The purpose of social media is to generate money for the companies that built them. One of the main ways they do this is through targeted advertisements delivered on their platform to their users. The more advertisements a user sees, the more money is extracted from that user. Therefore, the primary goal of any social media is to maximize the amount of time their users spend on the platform. In addition, if the users are more likely to actively engage with the content recommended to them, then they are more valuable to advertisers. The designers of these algorithms didn't expressly design them to increase polarization. Instead, they created self-learning algorithms and gave these algorithms the goal of optimizing user engagement and use time. It just so happens that recommending content that fulfills these goals also increases polarization. This is because increasing engagement between larger, like minded networks of people, can create echo chambers, as discussed in this paper (Finkel, 2020).

The evidence that the algorithms are causing this is quite damning. While the executives at Facebook may deny it, actions speak louder than words. Facebook has been known to, for

short periods of time, decrease the amount of inflammatory content it recommends during times of a possible crisis. This occurred, for example, directly after the 2020 Election (Roose et al., 2020) and during the verdict of the trial of Derek Chauvin (Meta, April 18). By doing this, they are actively admitting that during more stable times they are alright with the algorithms harnessing divisiveness to maximize ad revenue.

Another danger outside of polarization is the system's vulnerability to bad actors. Through the algorithm, it is possible for trolls and bots, sometimes controlled by a foreign power, to create discourse. Democracies, such as the US, are particularly vulnerable to these sorts of attacks. This was seen, for example, during the COVID-19 pandemic, where there was evidence that Russian bots spread misinformation regarding vaccines on American social media circles helping add to an illusion of a large popular front against the vaccines. This helped legitimize a harmful opinion to the American people.

As discussed in our background, the general public is most affected by overall political polarization and the spread of misinformation. However, certain vulnerable people can be driven to extreme radicalization by social media and by extension their algorithms. Before social media, people with extreme opinions would have a much more difficult time gaining a platform and spreading their ideas to others. Now, vulnerable people can be linked together and brough to extremism due to their propensity to engage with this content. When they create an echo chamber, it is much more dangerous than just being more intolerant of other views. We need to demonstrate that this is in fact happening and just how dangerous this can be.

First of all, we know that these extreme communities exist. Oftentimes, less extreme versions of them can be found on very common sites such as reddit. One example of this is the incel community, which is a predominantly male community defined by strong misogynistic

undertones, anger towards women, and self hatred. They believe their inability to find a partner is entirely due to genetic disadvantages and resent women because of it. While this mindset is certainly not healthy to the individual, and may pose a problem in their everyday relationships, on the surface it doesn't seem particularly dangerous to others. However, this is far from the case.

We can see this, for example, in the case of Elliot Rodger (Nagourney et al., 2014). On May 23rd, 2014 in Isla Vista California, the 22 year old man carried out a mass shooting that resulted in the deaths of six people and the injury of 14 others. Prior to the attacks and his subsequent suicide, a 141 page manifesto was posted in which he explained the motivation for his attacks. Incel ideology was the common thread throughout the manifesto, where he described a hatred towards women and a desire to punish them for their perceived rejection of him. Many have proposed that if he had not been pushed into this ideology by online incel forums, the attack may never have occurred.

Acts of violence due to online radicalization is far from isolated to the incel community. The organization of the 2017 “unite the right” Charlottesville riots, which resulted in many injuries and one death, was facilitated by online forums, including Reddit and 4Chan (Lagorio-Chafkin, 2018). While the resulting clampdown by reddit was not enough to stop other large alt right demonstrations, such as the infamous January 6th insurrection in 2021, when hundreds of alt-right Trump supporters stormed the capital building. In Western society, many acts of terror, whether it be from ISIS sympathizers, Alt-right neonazis, or any other extremism, have started with the perpetrator becoming radicalized online (Ribeiro et al., 2020).

The question we ask ourselves is: how did these seemingly normal people become involved in these radical communities that drove them to commit these violent acts? The answer, of course, goes back to the social media algorithms. Over a long period of time, users may be

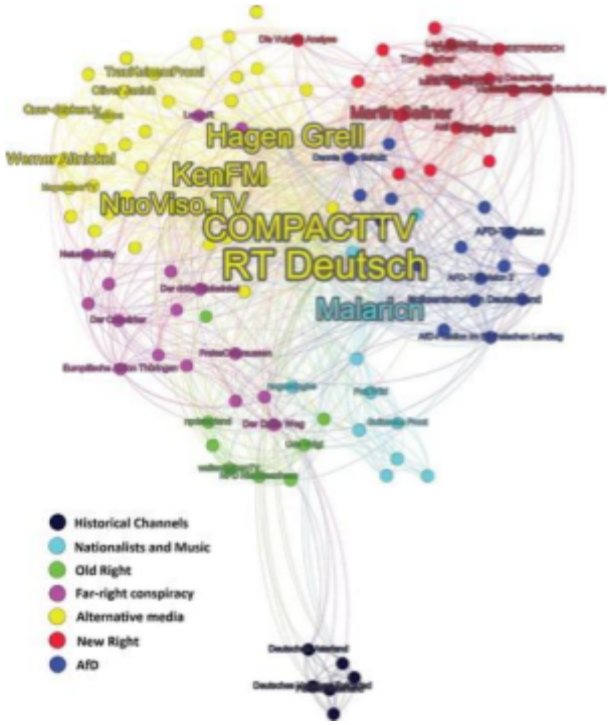
slowly drawn to extremism by being recommended progressively more and more radical content. For example, a viewer on youtube might start watching gaming content on Youtube. One day, Youtube might recommend a gaming Youtuber with edgier, more provocative humor. If the user engages with it, the algorithm may begin feeding the user videos circulating around various right wing talking points. The more the user engages with this content, the more extreme the recommended content becomes.

This effect has been mapped out in several studies. For example, this study (Champion, A. R. 2021) mapped out a network of incel-related community clusters by searching for typical incel keywords to identify the correct content on Youtube. She aimed to find the central and most influential related videos on Youtube, and if any more innocent content bridged the gap between incel and non-incel communities. They found there was indeed evidence that innocuous youtube channels are linked to more radicalized ideology.

A more in depth study was conducted in Germany (Rauchfleisch & Kaiser, 2020), with the results shown in Figure 2 below. They started by downloading 2.2 million comments from 200k distinct users on both alt-right and non-alt-right YouTube videos. They then created a network of user interaction, where they determined which users were commenting on the innocuous videos, when they commented, and what percentage of them were also commenting on more radical videos at a later date. Using the network, they were able to determine where users who commented on radical videos had originally migrated from and draw conclusions about where the algorithm was leading its users. One of the more interesting results is the connection to historical channels. While being interested in historical content is certainly not synonymous with being alt right, the evidence here points to the algorithm funneling these people towards that type of content.

Figure 2

Youtube User Interaction Over Right Wing and Right Wing Adjacent Channels In Germany



Overall, even large studies like this fail to capture the scope of the problem. While it documented 2 million comments, there were only 116 unique channels shown in the graph. The sheer scale of the data required to network out all of one social media platforms' algorithmic behaviors would require a significant amount of investment. For better visibility into these platforms, it may be smart for an agency with a larger amount of funding to map out larger parts of common social media in a similar way to give us a better view into where these algorithms are actually guiding us.

Framework Analysis

The core actor in this Actor Network is, of course, the Social Media algorithms. They are designed to maximize user engagement and time spent on their platforms. In doing so, they

inadvertently contribute to polarization and radicalization by creating echo chambers and recommending increasingly extreme content. Every human actor in the network is linked to social media through the social media algorithms.

There are a variety of different human actors that interact with these algorithms. The most common is the average social media user. Their interactions with the content drive the algorithms to recommend more polarizing content, which can further entrench their existing beliefs and contribute to a distorted perception of reality.

The other human actor in the "user" category is the vulnerable social media user, who is more susceptible to the influence of extreme content. As they interact with social media algorithms, their engagement with increasingly radical content may lead them into extremist communities. This user's experiences and interactions within these online spaces contribute to their further radicalization, with potential consequences for both their personal well-being and broader society.

Another important human actor is the content creators, who produce content for the platforms. They may create inflammatory or extreme content to attract more attention and engagement, which in turn can be amplified by the algorithms.

Finally, the most important human actors are the policy makers and regulators. They are the only bodies that have the power to implement policy changes or regulations that can address the negative consequences of social media algorithms, such as requiring greater transparency or limiting the extent to which algorithms can amplify extreme content.

Conclusion:

Despite the problems that come with how social media algorithms are influencing the current political space, it is very difficult for a governing body to step in and directly force the

companies to change their algorithms. While the effects are documented, it is very difficult to quantify the technical specifications for limiting them, especially when many members of congress are older and don't understand even basic technology. One interesting legislative proposal gets around this all together by forcing companies to allow users the ability to opt out of content recommended by algorithms (Person, 2021). A law like this does have precedence. In 2021, Apple released a privacy update that forced all apps to prompt users to opt in to sharing private data for the purposes of improving ads (Apple, 2023). This was a huge blow to companies such as Facebook that made a significant profit from targeted advertising and had a large effect on the industry. A similar feature could be a good step forward towards addressing the concerns in this paper.

If a law like this was instated, it would be very important to make the dangers of algorithmic content well known and force companies to be as transparent as possible. Users should know the consequences of opting into content recommended from the algorithms. In addition, more awareness could force advertisers and by extension the social media companies to change their strategies. If, for example, public opinion turned against advertisers with ads running alongside divisive content, it's possible these companies would pull their ads from these platforms and request change from the companies. Perhaps if the government helped to fund studies that mapped out large sections of interactions of the internet, it would be more easy to show to the general populace how these algorithms are affecting their daily lives and increase overall awareness of the issue. The more people who are taught to recognize the dangers of social media, the less will fall into this terrible trap.

References:

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020, March). The welfare effects of social media. American Economic Review. Retrieved April 13, 2023, from https://www.aeaweb.org/articles?id=10.1257%2Faer.20190658

Apple. (2023, February 13). *If an app asks to track your activity*. Apple Support. Retrieved April 13, 2023, from https://support.apple.com/en-us/HT212025

Barrett, P., Hendrix, J., & Sims, G. (2022, March 9). *How tech platforms fuel U.S. political polarization and what government can do about it*. Brookings. Retrieved April 13, 2023, from https://www.brookings.edu/blog/techtank/2021/09/27/how-tech-platforms-fuel-u-s-political-polarization-and-what-government-can-do-about-it/

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, *108*(10), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

Champion, A. R. (2021). Exploring the radicalization pipeline on youtube. *The Journal of Intelligence, Conflict, and Warfare*, *4*(2), 122–126. https://doi.org/10.21810/jicw.v4i2.3754

Faris, R., Hal, R., Bruce, E., Nikki, B., Ethan, Z., & Yochai, B. (2019, December 17). *Partisanship, propaganda, and disinformation: Online media and the 2016 u. S.*

Presidential election | berkman klein center. https://cyber.harvard.edu/publications/2017/08/mediacloud

Finkel, E. J. (2020, October 30). *Political sectarianism in america | science*. Retrieved April 13, 2023, from https://www.science.org/doi/10.1126/science.abe1715

Lagorio-Chafkin, C. (2018, September 23). *How charlottesville forced reddit to clean up its Act*. The Guardian. Retrieved April 13, 2023, from https://www.theguardian.com/technology/2018/sep/23/reddit-charlottesville-we-are-the-nerds-book-extract-christine-lagorio-chafkin

Levy, R. (2021). *Social Media, news consumption, and polarization: Evidence from a field ...* Retrieved April 13, 2023, from https://pubs.aeaweb.org/doi/pdf/10.1257/aer.20191777

Matthews, J. (n.d.). *Radicalization pipelines: How targeted advertising on social media drives people to extremes*. The Conversation. Retrieved March 24, 2022, from http://theconversation.com/radicalization-pipelines-how-targeted-advertising-on-social-media-drives-people-to-extremes-173568

Nagourney, A., Cieply, M., Feuer, A., & Lovett, I. (2014, June 2). Before brief, deadly spree, trouble since age 8. The New York Times. Retrieved April 13, 2023, from https://www.nytimes.com/2014/06/02/us/elliot-rodger-killings-in-california-followed-years-of-withdrawal.html

Person. (2021, November 9). *Social media users could disable algorithms in new U.S. proposal*. Reuters. Retrieved April 13, 2023, from

https://www.reuters.com/technology/social-media-users-could-disable-algorithms-new-us-proposal-2021-11-09/

Preparing for a verdict in the trial of Derek Chauvin. Meta. (2021, April 18). Retrieved April 13, 2023, from https://about.fb.com/news/2021/04/preparing-for-a-verdict-in-the-trial-of-derek-chauvin/

Rauchfleisch, A., & Kaiser, J. (2020). The German Far-right on YouTube: An Analysis of User Overlap and User Comments. *Journal of Broadcasting & Electronic Media*, *64*(3), 373–396. https://doi.org/10.1080/08838151.2020.1799690

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on YouTube. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 131–141. https://doi.org/10.1145/3351095.3372879

Roose, K., Isaac, M., & Frenkel, S. (2020, November 24). *Facebook struggles to balance civility and growth*. The New York Times. Retrieved April 13, 2023, from https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html

Sanford, C. (2021, March 25). Mark Zuckerberg opening statement transcript: House hearing on Misinformation. Rev. Retrieved April 13, 2023, from https://www.rev.com/blog/transcripts/mark-zuckerberg-opening-statement-transcript-house-hearing-on-misinformation

Von Behr, I., Reding, A., Charlie, E., & Gribbon, L. (n.d.). *Radicalization in the digital era*. Retrieved March 24, 2022, from https://www.rand.org/randeurope/research/projects/internet-and-radicalisation.html