Program on Extremism
THE GEORGE WASHINGTON UNIVERSITY

UNRAVELING THE IMPACT OF SOCIAL MEDIA ON EXTREMISM: IMPLICATIONS FOR TECHNOLOGY REGULATION AND TERRORISM PREVENTION

This paper, part of the Legal Perspectives on Tech Series, was commissioned in conjunction with the Congressional Counterterrorism Caucus

ANJANA SUSARLA

SEPTEMBER 2019

About the Program on Extremism

The Program on Extremism at George Washington University provides analysis on issues related to violent and non-violent extremism. The Program spearheads innovative and thoughtful academic inquiry, producing empirical work that strengthens extremism research as a distinct field of study. The Program aims to develop pragmatic policy solutions that resonate with policymakers, civic leaders, and the general public.

About the Author

Anjana Susarla is an Associate Professor at the Eli Broad School of Business, Michigan State University.

Anjana Susarla earned an undergraduate degree in Mechanical Engineering from the Indian Institute of Technology, Chennai; a graduate degree in Business Administration from the Indian Institute of Management, Calcutta; and Ph.D. in Information Systems from the University of Texas at Austin. Before attending graduate school, she worked in Enterprise Resource Planning (ERP) consulting. Her research interests include the economics of information systems, social media analytics and the economics of artificial intelligence. Her work has appeared in several academic journals and peer-reviewed conferences such as Academy of Management Conference, IEEE Computer, Conference on Knowledge Discovery and Data Mining, Information Systems Research, International Conference in Information Systems, Journal of Management Information Systems, Management Science and MIS Quarterly. She has served on and serves on the editorial boards of Electronic Commerce Research and Applications, Information Systems Research, MIS Quarterly and the Journal of Database Management.

Anjana Susarla has been a recipient of the William S. Livingston Award for Outstanding Graduate Students at the University of Texas, a Steven Schrader Best Paper Finalist at the Academy of Management, the Association of Information Systems Best Publication Award, a Runner-Up for Information Systems Research Best Published Paper Award 2012 and the Microsoft Prize by the International Network of Social Networks Analysis Sunbelt Conference. She has worked in consulting and led experiential projects with several companies. Her op-eds and work have been quoted in several media outlets such as the Associated Press, Newsweek, The Conversation, Fast Company, Sirius XM, World Economic Forum, Quartz, Chicago Tribune, Salon, Pew Research and the Nieman Lab. She has also been a speaker at public forums such as the SXSW and the United States Institute of Peace.

The views expressed in this paper are solely those of the author, and not necessarily those of the Program on Extremism or the George Washington University.

Introduction

Social media has been remarkably effective in bringing together groups of individuals at a scale and speed unthinkable just a few years ago. While there is a positive aspect of digital activism in raising awareness and mobilizing for equitable societal outcomes, it is equally true that social media has a dark side in enabling political polarization and radicalization. This paper highlights that algorithmic bias and algorithmic manipulation accentuate these developments. We review some of the key technological aspects of social media and its impact on society, while also outlining remedies and implications for regulation. For the purpose of this paper we will define a digital platform as a technology intermediary that enables interaction between groups of users (such as Amazon or Google) and a social media platform as a digital platform for social media.

1. Background: Social Media, Online Activism and Radicalization

Social media has been effective in enabling citizen activism and connecting individuals across the world. It needs to be acknowledged, however, that a dark side to social media is that in can push individuals towards extreme content where it can end up radicalizing individuals.

The following example illustrates this in the area of healthcare information. Users searching for videos on vaccine related information enter search phrases. Based on keywords, the user will then see a list of returned videos. There could be deliberately misleading content from users pushing hoaxes or from social bots, which have been proven to be more effective in spreading misinformation. The danger is that recommendation systems on such social media platforms traditionally emphasized popularity, which implies some of the top videos retrieved could be blatantly false. The motives of content creators, who could be interested in promoting their own products or

miracle cures to credulous YouTube viewers, coupled with recommendation bias towards popular videos, makes it likely for users to encounter blatantly false information about vaccines. However, for a user seeking information about vaccines, clicking on such a misleading video will only result in YouTube retrieving more of such videos, resulting potentially in even more extreme conspiracy theories or hoaxes about vaccines.[1] The video recommendations could be skewed by a tiny fraction of users who navigate from one extreme video to the other[2] leading a viewer down the rabbit hole of radicalization. Thus, one major danger from social media in the area of terrorism prevention is that of social media platforms unwittingly serving as agents of radicalization or providing exposure to inflammatory content.[3]

The second major challenge is that most information on social media, including extremist ideology and propaganda, is designed to be viral. While social media provides a sense of identity, purpose and connection, extremists also understand the virality of social media wherein disturbing content that is associated with intense emotions can be easily broadcast or even live streamed. The phenomenon of cultural products, such as videos, going viral on social media platforms illustrates that what matters to popularity is the ability of a cultural product to trigger conversations and engage viewers.[4] Viral hits become viral when, once they are posted on social media, an initial group of users engage in a number of conversations that spread on social media, resulting in a large number of responses and discussions, creating popular videos and online memes. In contrast to prior modes of broadcast media where an idea or content becomes popular after being broadcast by influential users with a large potential viewer base, it is the willingness with which a cluster of susceptible individuals engage with content transmitted on social media, as well as the imitability of content when users post response videos, that turns potential audiences from passive viewers into active commentators and sharers.[5]

The third aspect of social media that is relevant in the area of terrorism prevention is that of the opinion-making role of social media. Social media provides an opinion-making role to a few key (and connected) individuals. Most people engage with others

like them on social media,[6] so the resulting activism spirals into a large-scale online movement that would be impossible to ignore. An example will illustrate this opinion-making role. Amidst concerns about the presidential executive order on immigration, Uber's decision to turn off surge pricing spurred critics to engage in an online call for action, with the result that the hashtag #deleteuber was trending on social media, resulting in about 200,000 accounts being deleted.[7] It is important to note that the viral campaign #deleteuber campaign began with a tweet from a Chicago journalist. A study of online social interactions by Goel, Watts and Goldstein shows that relatively few of a small set of "seed" online users (with a very large degree of followers) account for a disproportionately large share of "viral" phenomenon on social media.[8] In other words, if #deleteuber is trending on Twitter, a few retweets by celebrities or news personalities would greatly increase the chance of this hashtag getting thousands if not millions of retweets.

Prior work has classified online activism as a spectrum starting from information seeking to direct action. The virality and immediacy of social media blurs the boundary between awareness, organization and action. Once such a hashtag reaches a threshold of visibility, subsequent online conversations between individuals only magnify the influence and spread of the hashtag. Every time a user deleted their Uber account, and they share the action on social media, it gives concerned individuals an immediate action they can engage in, as well as broadcast their activism to millions of other connected citizens. Prior work has showed how online conversations spiral into cascades.[9] As a hashtag starts trending, the same process is underway. It is the willingness with which groups of connected individuals engage in actions as well as advocacy, which when combined with the broadcast and reach of influential social media users, that turns an individual from passive sharers of information to active broadcasters.

The difference between what happened with the delete Uber campaign and what would have happened twenty years ago is the speed, scale and spread of digital activism.[10] In earlier times, we would not have the means to either raise awareness or mobilize for a

cause at this speed and scale, where the hashtag #deleteuber went viral and resulted in 200,000 accounts being deleted in a single day. Prior examples of successful citizen activism were for issues that involved substantial environmental damage, such as the Exxon Valdez spill or the Hinkley groundwater contamination,[11] and ones that impacted the quality of life for millions of individuals. These instances of citizen activism also involved very protracted negotiations (often for years) between companies and activists. Once again, contrast that with Uber, where activism spread virally and resulted in substantial damage to the company in less than 24 hours.

What needs to be understood though is that the scale, scope and speed with which stories can engage an audience occurs without the traditional filters, gatekeepers or rules regulating traditional media. The danger with such viral phenomenon is, when combined with algorithmic recommendations and echo chamber effects, ends up creating a reinforcing cycle of filter bubbles where users could be pushed into more radical views and opinions.[12] Chaslot provides experimental evidence of how social media platforms end up recommending extremist content and conspiracy theories.[13] This suggests a three-step process of radicalization since (i) groups of susceptible individuals are connected to others like them, (ii) content is being created to spread in a viral fashion (iii) content recommendations may further radicalizing individuals.

2. Radicalization and Propagation of Hate Speech: An Agenda for Social Media Regulation

Social media companies are not just platforms but publishers.[14] It is important to realize that algorithmic recommendations designed to maximize engagement can end up radicalizing or pushing viewers towards extremist content. A fundamental issue to understand is that malevolent agents are creating content that is designed to go viral and exploits the properties of online filter bubbles. It has also been shown that fake news created by bots is more effective than humans in spreading misinformation.[15] The features of these platforms of Facebook and WhatsApp[16] (owned by Facebook) that encourage participation and citizen engagement also make them vulnerable to hate

speech, fake news, and interference in the democratic process. This was, in fact, how Facebook was manipulated in the 2016 election.[17] In the recent instance, the accounts removed by Facebook were designed to inflame tensions around divisive topics like the rise of white supremacy in America and the US Immigration and Customs Enforcement. Large social media platforms also enable the transmission of fake news and opinion manipulation through micro-targeting.

Increasingly, observers point out that we trust platforms and algorithms more than our own governments and civic society. Phenomenon such as fake news and online radicalization are a consequence of our cognitive biases since they induce stronger emotions, and creators of such content are tapping into individuals' sense of activism by requesting actions such as re-sharing such content. Social media platforms are designed for content engagement, i.e., a model where users are continually engaged with the platform. There are no quality filters similar to FDA labels on our food that enable us to make an informed decision on whether or not we should consume that information.

The way digital platforms, and especially social media platforms monetize access, specifically social media, increases our vulnerability as users to disinformation. Instead of extremist videos being hidden in some darker corners of the Internet, social media platforms make it easy for anyone to stumble upon and post negative content disseminating hatred against a particular community or group, with the consequence that radicalization through exposure to hateful material.

In response to public pressure, platforms have instituted quality checks and filters that identifies (for at least highly popular content) if it is misleading. However, such an approach fundamentally misjudges cognitive biases and the process by which people make decisions. A study by Pennycook and Rand found that while tagging fake news headlines as disputed does not shift perceptions too much.[18] The viral spread of radicalism and the limitations of content filters pose challenges both to the detection, the spread of content (due to the problems of scalability of any method of detection) and finally given that objective verification or fact checking can be counterproductive.

To understand remedies and regulation, we need to acknowledge that social media platforms have become utilities. Digital platforms such as Facebook are designed for constant interaction and engaging viewers' attention, wherein the popularity bias gets reinforced. As a business model, the ad-based monetization strategies of companies such as Facebook and YouTube (owned by Google) was to understand what appeals to us, and serve us content and notifications designed to constantly keep us engaged and maximize the time we spend on those platforms. These platforms earn revenues through advertising wherein they can use the gargantuan amount of personal data that individuals have provided and sell that to advertisers for micro-targeting.[19]

In today's world, anyone can be a content creator, and there are few gatekeepers for online news, anyone can broadcast a rumor or share a news item without checking its veracity. Recommendation systems designed to maximize engagement have created filter bubbles that can be easily exploited by malevolent actors. These platforms, driven by the economics of big data, have become public utilities with monopoly power in the media industry, and wield broad societal influence. The political stakes for these platforms are enormously complicated. Trust in digital platforms is substantially different from trust in a technology company.

The societal impact of firms such as Facebook has crossed over from the corporate domain into broader societal issues such as their ability to provide transparency and public awareness and their role in shaping public opinion. They are no longer corporate entities responsible for their shareholders alone, but their ability to mold private interactions and sway public opinion affects the strength of the participative process and institutions of democracy. While the social media platforms have introduced some type of news validation, it is important to note that only content provided from top news sites and validated sources are carefully vetted. There is not much screening on these platforms for videos linked to dubious sources such as sub-reddits, 8chan etc.

One solution is for sites to establish a digital watermark to show that content is validated, especially for controversial topics. Platforms should also change their recommendation algorithms to make such hateful content less likely to be discoverable in related videos. The problem is with the algorithmic governance is the sheer size of these social media platforms. For instance, in the case of the most popular video sharing platform, YouTube, it was estimated that YouTube's video-recommendation algorithm influences about 700,000,000 hours of watch time per day.[20]

3. Some Challenges for Regulation

Platforms need regulatory pressure to establish that the cost of misleading information is higher than the revenue to platforms from their model of economics of continuous attention. While social media platforms such as Facebook and YouTube introduced fact checks, these approaches still rely on algorithmic governance. What needs to be understood is that we need a crowdsourced and citizen-centered, Wikipedia-type mechanism to address terrorism prevention and misinformation. Wikipedia has used this model effectively to regulate the quality of its content, though obviously for large-scale digital platforms this has to occur at a much higher scale and scope, as well as much faster pace. Increasingly, these platforms will face pressures from politicians, regulators, researchers, and social commentators to not only filter out fake news and in its role of combating online radicalization but also cognizant of its role in a democratic system.

While some observers have proposed top-down regulation, including breaking up technology companies, it needs to be recognized that economic agents are complex entities, and the economy (or any subsystem of the economy, such as a firm or an industry or a market) is a system made up of many sub agents. Complexity theory would suggest that in any complex adaptive system, the agents typically serve as "building blocks," and the overall organization exhibits tangled interactions across agents. The challenge with regulation is that the range of scenarios wherein agents interact through complex dynamics results in outcomes that are difficult to anticipate or be

systematically identified. The challenge in regulation is how government regulators from a single country could identify and impose a standardized, across-the-board solution to a problem like deep fakes, artificially-generated photos and videos that portray real people saying or doing something fictitious.[21] Due to the complexity of technology, unintended effects and reward hacking engendered by algorithms, it would be very difficult to develop a regulatory model that accounts for the effects of the regulatory regime on the system being regulated. Regulation is also challenging due to regulatory capture in high technology settings wherein the opaqueness or difficulty of interpreting the decision-making by the platform is opaque to the platform users. Technology companies then could be likely to lobby for legislation that actually enhances their ability to seek rents from users.

4. Solutions for Platform Governance and Regulation

The two issues to address are (i) the ownership of data, and (ii) the platforms' ability to curate content based on detailed information from the users.

The first issue is that while social media data is voluntarily provided, the process of data collection is not transparent to the end user. Users may not realize that Facebook has the ability to make inferences about political preferences, attitudes toward issues such as race and society. In other words, users are completely in the dark about the extent to which platforms can use the data collected about them to monetize user interactions, and in turn, how such nudges and recommendations from platforms impact aggregate behavior. As part of the recently approved GDPR legislation by the European Union (EU), there are articles that mandate that there be 'meaningful information about the logic involved' in automated decisions, which roughly suggest a right to explanation of what are the criteria that are used by algorithms in their decision-making.[22] Rules like the GDPR will require that companies elicit consent before the collection of data, and make the information they collect accessible to the consumer. It may be beyond the scope of most users to comprehend the extent to which their data is used and for the purposes their data is used for, unless the most digitally sophisticated users who can

monitor every piece of data they share with digital platforms. Simple data use agreements with end users is not feasible since platforms combine user-supplied data with data from third parties.

The second is that through microtargeting, curation of content, and recommendations, these platforms also have the ability to direct users' attention to more extreme content. Our online interactions on social media platforms are controlled by algorithms. Tufecki provides an example where a book highlighting an outlandish conspiracy theory started rising in the sales rankings.[23] This was because Amazon started listing this book as a related book recommendation for shoppers not looking for this particular title. Such a recommendation increases visibility of a book, which then increases the likelihood that it will be subsequently recommended to another shopper.

Recommendations work by discovering association - what products consumers buy in tandem, for example. For a large range of categories, however, social media platforms may not have enough data, or the viewership statistics may be sometimes based on only a fraction of viewers who could be extreme in their views. This can also end up giving a wider audience to conspiracy theorists. Fundamentally, the process by which a reader or viewer interacts with books, videos or other content on digital platforms is different than how we read a newspaper. When we read a physical book or newspaper, we could discuss it with a handful of acquaintances. However, the role of algorithms in curation and ranking systems implies that digital platforms have the power to nudge our decisions by highlighting collective influence, both through the power of social influence in that we observe what others are doing and in the platforms' ability to selectively highlight which content we see and which we do not. The digital neighborhood is widened, but it also widens our access to the most vitriolic, extreme or provocative content. While it is known that algorithmic recommendations lead to echo chamber effects, these problems are accentuated when content creators understand and game the process of algorithmic recommendations.

Below we highlight different approaches adopted in various models wherein the data ownership as well as content recommendations are shared between the users and the platform. Each of these has strengths but could be vulnerable to different types of regulatory and technological challenges.

Table 1. Different Approaches to Data Ownership and Content Curation

Ownership of Data	Content Recommendations	Centralization vs. Decentralization tradeoff	Examples
Centralized	Centralized	Centralized policy, but could be vulnerable to bots and misinformation	Facebook
Centralized	Content moderators	Self-correcting but cannot address the speed of content propagation of the scale of social media platforms	Wikipedia
Centralized	Decentralized	Policies need to be agreed upon by participants, issues with scaling	Sub-reddits
Decentralized	Decentralized	Offer control to the end user, but may not be as scalable as the user base grows	Bitcoin based social media platforms

Conclusion

Digital platforms can be vulnerable to interference from authoritarian governments, and be used to spread false news, influence electoral processes, and in short threaten the functioning of democratic institutions. Rather than top-down regulation, the best way to engage in terrorism prevention may be a combination of external pressure from governments, user driven monitoring, including data access rights, and adoption of community norms.

Social media platforms use our data and our interactions to get to know us and then monetize our personal information for advertisers to micro target us based on the data we revealed through our digital interactions. In a sense, we are all stakeholders of these platforms, but we do not get rewarded for providing access to our data. Some states have suggested that digital platforms need to provide data ownership rights to users. Financial institutions have collateral posted on bilateral transactions that protects us against specific default risk. Digital platforms should likewise consider a user bill of rights for their usage and safeguarding of data to keep them within terms.[24]

References

[1] Liu X., B Zhang, A Susarla, R Padman, "Go To YouTube and Call me in the Morning: Go To YouTube and Call Me in the Morning: Use of Social Media for Chronic Conditions," MIS Quarterly, forthcoming.

[2] Chaslot, G. 2017. How YouTube's A.I. boosts alternative facts. Retrieved from https://medium.com/@guillaumechaslot/how-youtubes-a-i-boosts-alternative-facts-3cc276f47cf7.

[3] Nicas, J. 2018. How YouTube Drives People to the Internet's Darkest Corners. Wall Street Journal, Retrieved Online from https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478.

[4] Susarla A., J. Oh, Y. Tan, "Influentials, Imitables or Susceptibles? Virality and Word of Mouth Conversations in Online Social Networks," Journal of Management Information Systems, June 2016, 33(1): 139-170.

[5] Susarla A., J. Oh, Y. Tan, "Influentials, Imitables or Susceptibles? Virality and Word of Mouth Conversations in Online Social Networks," Journal of Management Information Systems, June 2016, 33(1): 139-170.

[6] This is called homophily, or a preference for like-minded others.

[7] Huston, C. 2017. Consumers lash out at Uber and turn to Lyft after Uber's immigration response. Market Watch. Retrieved Online from http://www.marketwatch.com/story/consumers-lash-out-at-uber-and-turn-to-lyft-after-ubers-immigration-response-2017-01-29.

[8] Goel, S., D. Watts, D.G. Goldstein, 2012. The structure of online diffusion networks. Proceedings of the 13th ACM Conference on Electronic Commerce, 623-638. Valencia, Spain.

[9] Gonzalez-Bailon, S., Borge-Holthoefer, J., A. Rivero and Y. Moreno, 2011. The Dynamics of Protest Recruitment through an Online Network. Nature Scientific Reports, 1 (197)

Greene. D., We Don't Need New Laws for Faked Videos, We Already Have Them, Electronic Frontier Foundation. (Feb. 13, 2018), https://www.eff.org/ deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them.

[10] Susarla, A. 2017. How social media turned United's PR flub into a firestorm. The Conversation. Retrieved Online from https://theconversation.com/how-social-media-turned-uniteds-pr-flub-into-a-firestorm-76210.

[11]Where Pacific Gas and Electricity was held accountable.

[12] Bessi A, Zollo F, Del Vicario M, Puliga M, Scala A, Caldarelli G, et al. (2016) Users Polarization on Facebook and YouTube. PLoS ONE 11(8): e0159641.

[13]Chaslot, G. 2017. How YouTube's A.I. boosts alternative facts. Retrieved from https://medium.com/@guillaumechaslot/how-youtubes-a-i-boosts-alternative-facts-3cc276f47cf7.

[14] Susarla, A. 2018. Facebook begins to shift from being a free and open platform into a responsible public utility. The Conversation. Retrieved Online From https://theconversation.com/facebook-begins-to-shift-from-being-a-free-and-open-platform-into-a-responsible-public-utility-101577.

[15] Lazer, D.M.J., M. A. Baum, Y. Benkler, A J. Berinsky, K M. Greenhill, F Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, J. L. Zittrain. 2018. The science of fake news. Science. 359(6380): 1094-1096.
[16] "Understanding Media and Information Quality in an Age of Artificial Intelligence, Automation, Algorithms and Machine Learning," Berkman Klein Center, July 13, 2018, https://cyber.harvard.edu/story/2018-07/understanding-media-and-information-quality-age-artificial-intelligence-automation.
[17] "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach | News | The Guardian," accessed June 20, 2019, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.
[18] Pennycook, G. & Rand, D. G. 2019. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality, Forthcoming.
[19] Susarla, A. 2018. Facebook begins to shift from being a free and open platform into a responsible public utility. The Conversation. Retrieved Online From https://theconversation.com/facebook-begins-to-shift-from-being-a-free-and-open-platform-into-a-responsible-public-utility-101577.

[20] "Press - YouTube," accessed June 20, 2019, https://www.youtube.com/yt/about/press/.
[21] Greene. D., We Don't Need New Laws for Faked Videos, We Already Have Them, Electronic Frontier Foundation. (Feb. 13, 2018), https://www.eff.org/ deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them.
[22] Selbst, A.D., and J. Powles. 2017. Meaningful information and the right to explanation. International Data Privacy Law, 7(4): Pages 233–242.
[23] Tufekci, Z. 2019. How Recommendation Algorithms Run the World. Wired. Retrieved Online from https://www.wired.com/story/how-recommendation-algorithms-run-the-world/.
[24] Hosanagar, K. 2019. A Human's Guide to Machine Intelligence: How Algorithms Are Shaping Our Lives and How We Can Stay in Control. Penguin: NY.