

**Emergence and Impact of Twitter-Propagated
Conspiratorial Rhetoric and Misinformation Campaigns**

Wilson Chen

College of Computing, Georgia Institute of Technology

CS 4873: Computing, Society, and Professionalism

Dr. Ian Bogost

April 25, 2021

Emergence and Impact of Twitter-Propagated Conspiratorial Rhetoric and Misinformation Campaigns

January 6th, 2021 marks the date of the Capitol Hill riot, where a legion of ardent supporters of Donald J. Trump breached barricades, defaced the Capitol, and endangered the lives of countless government officials in a bid to overturn election results they believed to be “rigged” (Henderson & Alexander, 2021). Significant information exists to suggest that social media played an outsized role in organizing the takeover, with calls for violence and action circulating months in advance (Frenkel, 2021) and echoed by the President himself on Twitter in the weeks leading up to the event. The common thread among these groups was a shared belief in a conspiracy theory asserting that the presidential election was hijacked.

January 6th has shown the consequences of allowing conspiracy and misinformation to fester on social media, raising an important discussion about the obligations that social media platforms have in regulating the content hosted on their platforms. A 2020 survey by the Pew Research Center on news use among these platform estimates that around half of U.S. adults (53%) rely on social media “often” or “sometimes” to get informed about current events (Shearer & Mitchell). Of the populations using social media platforms, Twitter users are the most likely to regularly get their news from the site, with 59% of Twitter users responding that they use the platform regularly to receive news (while Facebook and Reddit users trail at 54% and 42%, respectively). Given that a significant population of U.S. citizens are now receiving news about current events from social media platforms, the danger of a widely syndicated misinformation campaign cannot be understated.

Anatomy of a Conspiracy

What mechanisms underlie the processes by which misinformation movements gain momentum? Mattie Bruton identifies a recurring theme in conspiratorial rhetoric: amorphous “elite” entities (“Big Pharma”, “Global Corporations”, “War Inc”, among others) pulling the strings to amass control over the general populace (2020). There is a certain ingenuity in this framework’s ability to easily adapt to many types of conspiratorial rhetoric. These types of conspiracies often foster a distrust in conventional sources of information credibility, with terms like the “mainstream media” becoming a derogatory buzzword to insinuate that a news source is polluted by shadowy propagandists manipulating people to believe and behave a certain way. With a distrust in traditional media, people turn to alternative sources of information and seek information that reaffirms their beliefs.

Conspiracy as a Form of Identity-Building

For some, a righteousness to their own beliefs leads them onto Twitter to become activists staking their identity on the roots of these conspiracies. For an anti-vaccination advocate, “the stakes are high. Innocent people, specifically children, are being poisoned by the forces of pure evil which operate within the medical institution and the institution of democratic government. Someone has to do something; we all have to do something.” (Bruton, 2020, p. 63). It is no longer just a belief; it is a cause to fight for. These strong attitudes pave a virulent foundation for rhetoric to take hold and spread on Twitter.

In the case of the Qanon conspiracy theory (where users follow the updates of an anonymous user known as “Q” who claims to have insider information about the corrupt American “deep-state”), hashtags and slogans are an effective way of signaling a stake in a shared belief (Bruton, 2020). #WWG1WGA (Where We Go One, We Go All) is plastered on

user biographies and within the posts of Qanon followers, providing a rallying cry to members of the in-group. Many of these accounts load their user biographies with various hashtags to communicate their positions (Bruton, 2020, p. 53):

*“Devoted wife/MomX3 #Patriot #TRUMP2020 #ExVaxxer #TheGreatAwakening
#VaccineRiskAware #Vaxxed #WWGIWGA #MAGA #Qanon #1A #2A vactivists.org.”*

*“Activist for children, Mom of vaccine injured child. #Patriot. Standing 4
#FREEDOM #Trump2020 #CDCwhistleblower VaxTruth.org #MAGA #WWGIGA #QANON”*

Particularly interesting is the conviction evident in these biographies. Phrases like “#ExVaxxer” and “Standing 4 #FREEDOM [sic]”, along with links to sites like “vactivists.org” and “VaxTruth.org”, evoke strong sentiments of fighting against a perceived injustice. To an impressionable Twitter user, these hashtags and links offer many entry points to the conspiracy pipeline.

Formation and Propagation of the #FilmYourHospital Conspiracy Theory

This “hashtag activism” can be observed through the case of #FilmYourHospital, a misinformation campaign that emerged in late March of 2020 during the early months of the COVID-19 pandemic (Gruzd & Mai, 2020). Framed as a Twitter trend, it encouraged Twitter users to visit their local hospitals and take pictures of vacant hospital spaces in a collective effort to “debunk” the narrative pushed by the media that COVID-19 was a serious threat. Gruzd’s team performed social network analysis on the hashtag, identifying its origins as one simple tweet on March 28th from a small anonymous user which asked “#FilmYourHospital Can this become a thing?”. Although the original tweet only amassed limited attention at around 30 retweets, the message was amplified by notable Congressional candidate DeAnna Lorraine a day later with the following message:



Days later, the #FilmYourHospital conspiracy was further echoed by notable Trump supporter Donna Warren (@DonnaWR8) with over 100K followers, and by conservative commentator/political activist Candace Owens (@RealCandaceO) with a reach of 2.1 million followers (Gruzd & Mai, 2020). Between the two-week period of March 28th and April 9th, a total of 99,039 posts were observed by Gruzd's team with the #FilmYourHospital hashtag campaign, with contributions from 43,461 unique users.

To understand the parties that were engaging with this information, Gruzd's team used a machine learning tool to identify if the growth of the campaign was driven largely by bots. A sample of 1213 accounts that were particularly active in the #FilmYourHospital discourse had their profiles quantified through a Complete Automation Probability (CAP) score. The team discovered that 98% of the accounts spreading this information passed below the 0.25 CAP

threshold, suggesting that this activity could not be concluded as automated and that the growth of this misinformation campaign was largely sustained by real, independent users.

This is perhaps more insidious than the proposition that malicious automated systems artificially drive misinformation on Twitter. The ability for Twitter “celebrities” to instantly syndicate #FilmYourHospital to massive audiences is a powerful tool to influence an audience. When verified Twitter accounts provide credibility to conspiracy, removal of misinformation must reconcile with accusations of biases, censorship, and political tribalism. When a significant number of real users genuinely participate in online conspiratorial discourse, platforms risk antagonizing a large subset of users when they attempt to address the problem.

Consequences of COVID-19 Misinformation on Public Behaviors

#FilmYourHospital is one example out of many other COVID-related misinformation campaigns that have entrenched many people into a network of anti-vaccination rhetoric throughout 2020 and into 2021, worryingly as vaccines become available for widespread distribution. A February 2021 Gallup poll tracking the percentage of Americans willing to take a coronavirus vaccine demonstrates a large partisan divide between Democrats and Republicans at 91% and 51% willingness respectively (Brenan, 2021). The politicization of vaccine skepticism from notable conservative commentators like those who amplified the #FilmYourHospital conspiracy is likely to have played a significant role in forming these differences.

Stigma tied to non-credible COVID rhetoric has also been shown to produce an increase in hate crimes, particularly toward Asian-American communities (Choi, 2021). The origin of the virus in Wuhan, China provided certain groups with the ability to make fallacious associations between Asian-Americans and their unsubstantiated role in spreading the disease, an assessment that the FBI predicted in March of 2020 (Margolin, 2020). Disingenuous rhetoric by political

leaders, such as then-President Trump's insistence on calling it the "China virus" instead of the official terms "coronavirus" or "COVID-19" acknowledged by federal health officials, insinuate an association between the virus and an ethnicity. Many Trump supporters on Twitter adopt the problematic "China virus" vocabulary in discussing the pandemic, rather than the established terms, producing conversations that reaffirm this faulty connection. As Shinwoo Choi recounts, "It was my husband who refused to wear a mask at the supermarket, telling me, 'I don't want to be seen as the Asian who carries virus.'" (para. 5). For her husband, the risk of contracting COVID-19 was less significant than the perception that he would be seen by others as a vector of the disease himself. Choi further notes cases where "an Asian woman was attacked in New York City because she was not wearing a mask, while others were attacked because they were wearing a mask." (para. 6). Evident by Choi's experiences, stigma formed around COVID-19 from social structures that alienate significant demographics of people. Asian-Americans become a scapegoat for other citizens to express their frustrations toward in lack of a more concrete entity. With the very real consequences of false information propagation, what can be done by Twitter to mitigate these problems?

What Can Twitter Do? A Timeline of Twitter's Evolving Moderation Efforts

Twitter's position as a prominent medium of information dissemination makes it impossible to separate the platform from the discourse within it. It is evident that there are notable figures on Twitter that hold a degree of responsibility for circulating conspiracy and misinformation. Given that these accounts amass hundreds of thousands to millions of followers, targeting these accounts appears to be a logical vector for Twitter to act through.

Examining the Twitter blog's history regarding its stance on political figures and world leaders, their *World Leaders on Twitter* 2018 post indicates that Twitter has an incentive to give

these prominent figures a platform in the interest of advancing global, public conversation. With regards to controversy, Twitter justifies that removing the figure from Twitter would “hide important information people should be able to see and debate”, and “would not silence that leader, but it would certainly hamper necessary discussion around their words and actions” (2018, para. 3).

In the summer of 2019, Twitter announced a public interest framework in blog post *Defining public interest on Twitter* that would add a new notice on tweets made by public officials that violated the Twitter rules. These notices would clarify that a tweet violated Twitter Rules on abusive behavior but would also state that it would be in the public’s interest to keep the tweet available. Furthermore, all interactions (except for quote-retweets) would be disabled with the intention of reducing the reach of the tweet. Notable instances of this framework’s application are with various tweets that former President Trump has composed. Among those: repeated claims of a “rigged election”, downplaying the threat of the coronavirus, encouraging supporters to vote twice in North Carolina, and the characterization of Black Lives Matter protesters as “thugs” with the infamous threat “when the looting starts, the shooting starts” (Cohen, 2021). In one way or another, some notice was appended by Twitter onto each of these Tweets, but the messages themselves were not removed.

Returning to the events of the January 6th Capitol riot, Twitter – along with many other social media platforms – took unprecedented action and permanently suspended the account of Donald Trump. For some critics, this was an action that should have taken place far earlier. For others, the ability to deplatform world leaders was seen as an overreach of power and alienated millions of Twitter users.

Nonetheless, there is evidence to suggest that while controversial, the suspension of Trump has proved to be an effective way to combat misinformation. In the week following Trump's suspension from Twitter, conversations of election fraud dropped from 2.5 million mentions to just 688,000 mentions across several social media sites, representing a significant 73% decline (Timberg & Dwoskin, 2021). It should be noted that confounding factors may correlate with this decline, as Twitter also reported purging over 70,000 Qanon-affiliated accounts following the Capitol riot as well (Dwoskin & Romm, 2021). Regardless, these measures taken together seem to indicate that a more liberal approach to suspending accounts that spread misinformation and conspiracy could be an effective strategy.

A caveat to the aforementioned analysis is that removing these users does not address the fact that they still hold these convictions. Research on conspiracy theories following ostracism finds that individuals who are ostracized are more likely to endorse conspiracy beliefs related to politics (Poon, Chen, & Wong, 2020). Alternative platforms such as Gab and Parler have gained momentum as a response to perceived mainstream biases, suggesting that users who find themselves ostracized from mainstream social media platforms will seek networks to turn to that will accept their beliefs, with worrying implications for the future of political polarization and discourse in the United States.

Final Considerations

Twitter's ramping enforcement of its Rules and Guidelines in response to conspiracy and misinformation on its platform are a welcome change when considering its relatively lax moderation policies in years past. However, given the amount of people who have already entrenched themselves in conspiratorial belief, avenues of deprogramming these individuals must be a topic of conversation. A proactive, rather than reactive, approach could be formulated

to minimize the chances of misinformation manifesting into violence, as seen in the Anti-Asian microaggressions associated with the “China virus” phrase and the chaos of the January 6th Capitol riots. As a common theme in conspiratorial rhetoric is a distrust in conventional media (Bruton, 2020), there is a fine line for Twitter between enforcing its rules and alienating users deeper down the conspiracy pipeline. Areas for possible further research may involve the efficacy of rehabilitative (as opposed to punitive) approaches to moderation on Twitter on reducing conspiracy beliefs, as a practice typically defined within the scope of reducing recidivism within the justice system. The use of algorithms as a means of moderating at scale may hold promise in enhancing the arsenal of moderation tools, although this approach must be cautious in addressing challenges of enforcement transparency and bias (Gorwa, Binns, & Katzenbach, 2020). Lastly, another target for ongoing research may lie in studying the communities of “alternative” social networks (Gab, Parler) and recurring themes of rhetoric that develop within them.

Certainly, moderating a platform with billions of users with varied opinions is a monumental challenge. Nonetheless, this should not discount Twitter’s obligation to limit the propagation of verifiably false content. As a networking service with worldwide reach, the actions users make online influence real-world behaviors and consequences. In serving its users as a platform for public conversation, Twitter must continue to evolve its response to misinformation to ensure platform integrity and minimize harm.

References

- Brenan, M. (2021, April 6). *Two-Thirds of Americans Not Satisfied With Vaccine Rollout*. Gallup.com. <https://news.gallup.com/poll/329552/two-thirds-americans-not-satisfied-vaccine-rollout.aspx>.
- Bruton, M. (2020). *Morality, Epistemology, and Activism: How Anti-vaccination Advocates on Twitter Construct a Rhetoric of Alternative Immunity* (Doctoral dissertation, Purdue University Graduate School). https://hammer.figshare.com/articles/thesis/Morality_Epistemology_and_Activism_How_Anti-vaccination_Advocates_on_Twitter_Construct_a_Rhetoric_of_Alternative_Immunity/12762080/files/24157430.pdf
- Choi, S. (2021). "People look at me like I AM the virus": Fear, stigma, and discrimination during the COVID-19 pandemic. *Qualitative Social Work*, 20(1-2), 233-239. <https://journals.sagepub.com/doi/full/10.1177/1473325020973333>
- Cohen, L. (2021, January 26). *A running list of Trump tweets that have been flagged by Twitter*. The Daily Dot. <https://www.dailydot.com/debug/trump-twitter/>.
- Dwoskin, E., & Romm, T. (2021, January 12). *Twitter purged more than 70,000 accounts affiliated with QAnon following Capitol riot*. The Washington Post. <https://www.washingtonpost.com/technology/2021/01/11/trump-twitter-ban/>.
- Frenkel, S. (2021, January 6). *The storming of Capitol Hill was organized on social media*. The New York Times. <https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html>.
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*. <https://doi.org/10.1177/2053951719897945>
- Gruzd, A., & Mai, P. (2020). Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society*, 7(2), 2053951720938405. <https://journals.sagepub.com/doi/full/10.1177/2053951720938405>
- Henderson, C., & Alexander, B. (2021, January 7). *President Trump supporters violently storm Capitol Hill: Here's everything we know*. USA Today. <https://www.usatoday.com/story/news/nation/2021/01/06/capitol-hill-riot-heres-everything-we-know/6573033002/>.
- Margolin, J. (2020, March 27). *FBI warns of potential surge in hate crimes against Asian Americans amid coronavirus*. ABC News. <https://abcnews.go.com/US/fbi-warns-potential-surge-hate-crimes-asian-americans/story?id=69831920>.

- Poon, K. T., Chen, Z., & Wong, W. Y. (2020). Beliefs in conspiracy theories following ostracism. *Personality and Social Psychology Bulletin*, 46(8), 1234-1246.
<https://journals.sagepub.com/doi/full/10.1177/0146167219898944>
- Shearer, E., & Mitchell, A. (2021, January 12). *News Use Across Social Media Platforms in 2020*. Pew Research Center's Journalism Project.
<https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/>.
- Timberg, C., & Dwoskin, E. (2021, January 17). *Misinformation dropped dramatically the week after Twitter banned Trump and some allies*. The Washington Post.
<https://www.washingtonpost.com/technology/2021/01/16/misinformation-trump-twitter/>.
- Twitter. (2018, January 5). *World Leaders on Twitter*. Twitter.
https://blog.twitter.com/en_us/topics/company/2018/world-leaders-and-twitter.html.
- Twitter. (2019, June 27). *Defining public interest on Twitter*. Twitter.
https://blog.twitter.com/en_us/topics/company/2019/publicinterest.html.
- Twitter. (2021, January 8). *Permanent suspension of @realDonaldTrump*. Twitter.
https://blog.twitter.com/en_us/topics/company/2020/suspension.html.