IMT 575: User Roles in Electoral Disinformation on Twitter

Group 2: Stephen Prochaska, Chris Fu, Leo Moley, Nayan Kaushal

1. Introduction

During the 2020 U.S. Presidential Election there were numerous false and misleading claims surrounding the election and its processes. Previous work has demonstrated that at least some of these claims were part of a coordinated disinformation campaign (Benkler et al, 2020) meant to erode trust in electoral systems and ultimately the outcome of the election in the form of the #StopTheSteal campaign and the January 6 insurrection (Election Integrity Partnership, 2021). There were hundreds of such claims over the course of 2020 and although the claims worked together to erode trust, it is educational to examine specific claims in detail to better understand how they may have functioned to delegitimize the election as a whole. The current study examines Tweets related to a narrative consisting of both misinformation and disinformation: the Dominion conspiracy theory.

At its core, the Dominion conspiracy theory centers around false/misleading claims that Dominion Voting Systems, a company that sells electronic voting software and hardware, was connected to alleged voting irregularities in multiple states. The conspiracy theory was supported and spread by Trump and his allies after election day, when Trump began to tweet about fraud that he claimed was due to Dominion based on articles by One American News Network (Collins, 2020). It wasn't just Trump who promoted the theory at high levels, both Rudy Guiliani and Sidney Powell (among others, including Fox News anchors and Newsmax) actively spread conspiracies related to Dominion and QAnon, also drawing debunked connections between elections software company Smartmatic and Dominion (Swenson, 2020). Additionally, Ron Watkins, former administrator of fringe imageboard website 8kun (previously 8chan) where he was suspected by some to have been posting as Q, from QAnon notably tweeted multiple times claiming to have knowledge of how the Dominion software could have been compromised and offered to help Giuliani and Powell investigate the potential fraud (Collins, 2020).

As the case of the January 6, 2021 insurrection demonstrates, the effects of mis/disinformation are not constrained only to digital spaces. Tactics from participatory activism and social movements are evident in both disinformation campaigns and extremist movements, but whether and how the two different styles of movement differ in their participatory structure is unknown. The current study aims to better understand the relationship between the online community structures present in disinformation campaigns and extremist movements on different platforms. To that end, we answer the following question:

- Do the roles identified by Phadke & Mitra (2021) in extremist movements on Facebook (Educators, Solicitors, Flamers, Motivators, Sympathizers) translate to disinformation on Twitter?

2. Related Work

The following section introduces the concept of participatory disinformation, of which the Dominion conspiracy theory is a strong example, and moves into a discussion of the taxonomy developed by Phadke & Mitra (2021).

2.1 Participatory Disinformation, Social Movements, and Repeat Spreaders

Benkler et al. (2020) present a conception of the disinformation present in 2020 as primarily driven by traditional media platforms such as cable news (e.g. Fox News) that is then amplified in online spaces. Although more traditional media platforms undoubtedly played an influential role in the disinformation surrounding the 2020 election, there is an increasingly large body of literature that describes the participatory relationships that develop between disseminators of disinformation and supporters who unwittingly spread and even generate disinformation (Starbird et al., 2019; see also Ong & Cabanes, 2018; Nemer, 2021). From this perspective, members of the general public participate in the production and propagation of disinformation without realizing that they are participating in a larger information operation. Many of these unwitting agents are true believers of the disinformation that they help spread, which makes it difficult to tell where a disinformation campaign ends and a genuine social movement begins. Critically, the unwitting agents are not merely peripheral actors in the proliferation of claims making up a campaign. Indeed, historical descriptions of disinformation describe the often significant role that unwitting agents play in the success of disinformation campaigns by actively participating in conversation and action (Rid, 2020). In the context of the U.S. in 2020, it is made even more difficult to tell the difference between a genuine social movement and a disinformation campaign by the fact that, as Benkler et al. (2018) rightfully point out, the current iteration of disinformation that the U.S. is experiencing is the result of decades of rhetoric from traditional media sources that have continuously been attempting to influence public perceptions.

Previous work on modern iterations of digital social movements has focused on how pro-democratic movements such as Occupy Wall Street, the Arab Spring, and the Indignados Movement have mobilized resources to further their causes (e.g. Postill, 2014;

Howard & Hussain, 2013; Theocharis et al. 2015; Garrett, 2006). According to these conceptions of the interaction between social movements and internet communication technologies (ICTs), the internet is a democratizing force that allows for widespread participation in movements that would otherwise be impossible to mobilize around.

Unfortunately, one of the primary weaknesses of this conception, as Benkler et al. (2018) point out, is that digitally distributed social movements suffer from an inability to easily create and maintain organizational structure. This lack of organizing structure is something that a deliberately organized disinformation campaign (or extremist group) does not lack, and therefore disinformation campaigns have been successful at weaponizing the affordances of online platforms and communities using the exact same strategies that previous movements had used in a liberating way (see Starbird et al., 2019; Ong & Cabanes, 2018; Nemer, 2021; Benkler et al. 2018). In order to better understand the roles present in this structure, we turn next to a taxonomy of participation in extremist movements.

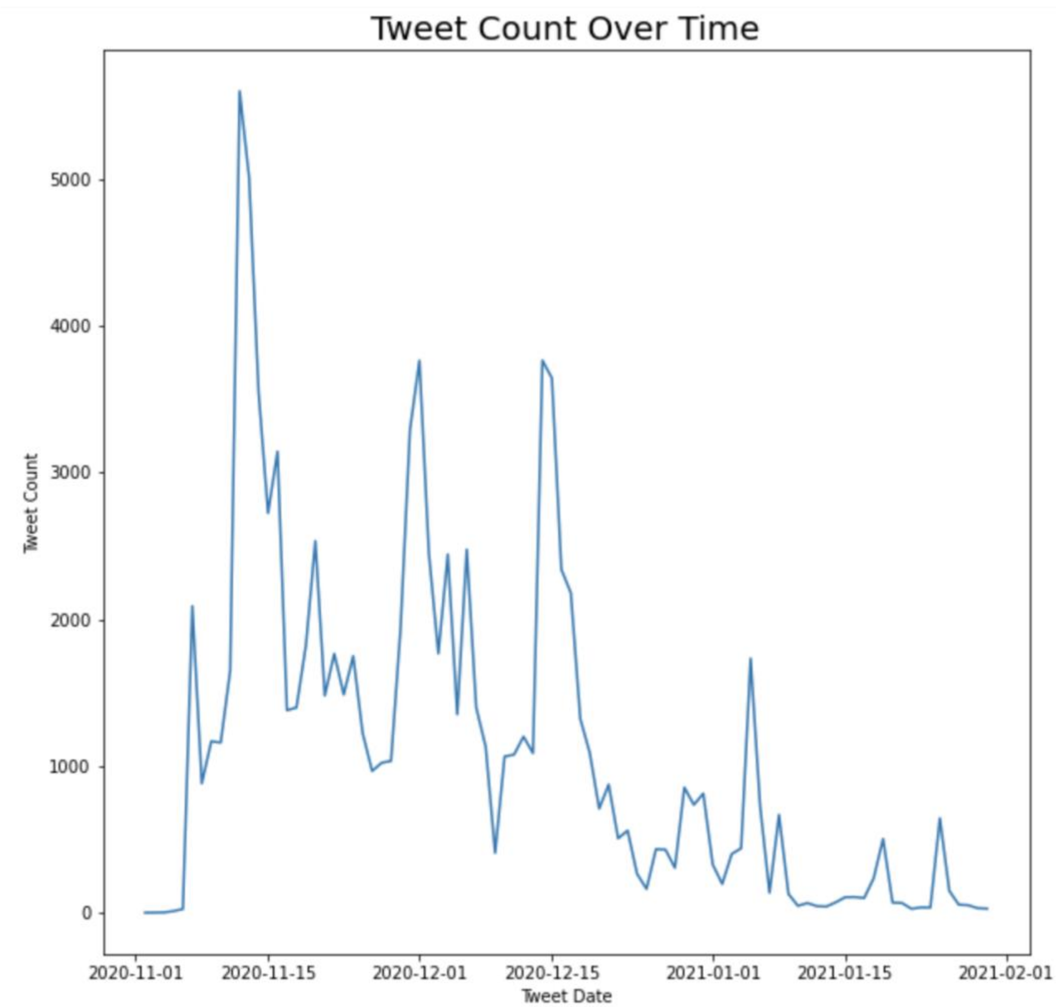
2.2 Role Taxonomy: Educators, Solicitors, Flamers, Motivators, Sympathizers

Phadke and Mitra (2021) observed a role structure in the context of online extremist social movements. In their study of Facebook pages and groups that shared content from websites of entities designated as extremist groups by the Southern Poverty Law Center (SPLC), Phadke and Mitra classified each page and group into one of five categories – — educators, solicitors, flamers, motivators and sympathizers – according to the social role they served within the broader extremist ecosystem. Although this study focused on U.S. domestic extremist content specifically, there are similarities between how information flows within the context of extremism and political disinformation, and indeed, the two often meld together in the context of a campaign. They developed their taxonomy by generating 13 different features that they used to perform k-means clustering for accounts on Facebook involved in extremist movements. Each feature was generated based on the different types of participation present in social movements as discussed in theories based on participatory activism and social psychology. These theories and resulting features were grouped into three main categories: drives for participation, engagement in the movement, and strategies for mobilization. Phadke and Mitra then operationalized each feature for extremism on Facebook and clustered based on the resulting features, resulting in the qualitatively validated role taxonomy. Due to their success identifying roles on Facebook, we adapted the same process and features for use on Twitter within disinformation, discussed next.

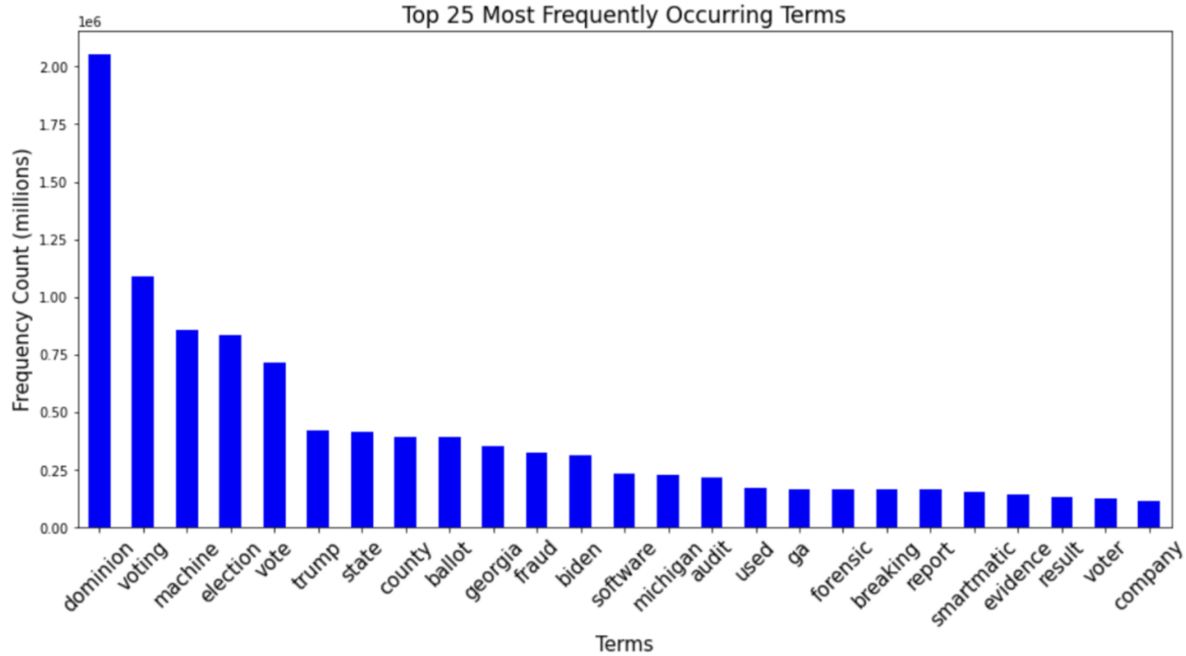
3. Data

For our analysis, tweets related to the "Dominion Conspiracy Theory" were used as the primary data source. The dataset consisted of 2.5 million tweets collected by the Center for an Informed Public (CIP) in real-time. Since the tweets were collected while the Dominion Conspiracy Theory was at large, it wasn't limited by traditional constraints. This means that tweets which were later removed or ones that were posted by users whose accounts were removed from Twitter are available in the data. Since the project aims to study the roles of users in the spread of disinformation, we removed the data pertaining to users who posted less than 5 tweets. After filtering out these users, we were left with 1.88 millions rows for further analysis.

Descriptive Analysis of the data



The figure above shows the number of tweets over the time period of Nov 11, 2020 to Feb 1, 2021. The highest number of tweets related to the Dominion conspiracy theory were posted in mid-November, almost 1.5 weeks after the election results were announced. Moreover, there were multiple spikes in the tweet data with time, which suggests that discussions related to new (fake) information had occurred on those days.



The figure above shows the top 25 most frequently occurring unigrams in the data. The terms “dominion”, “voting”, “machine”, “election”, and “vote” are understandably the top 5 occurring terms with frequencies ranging from 0.75 to 2 million.

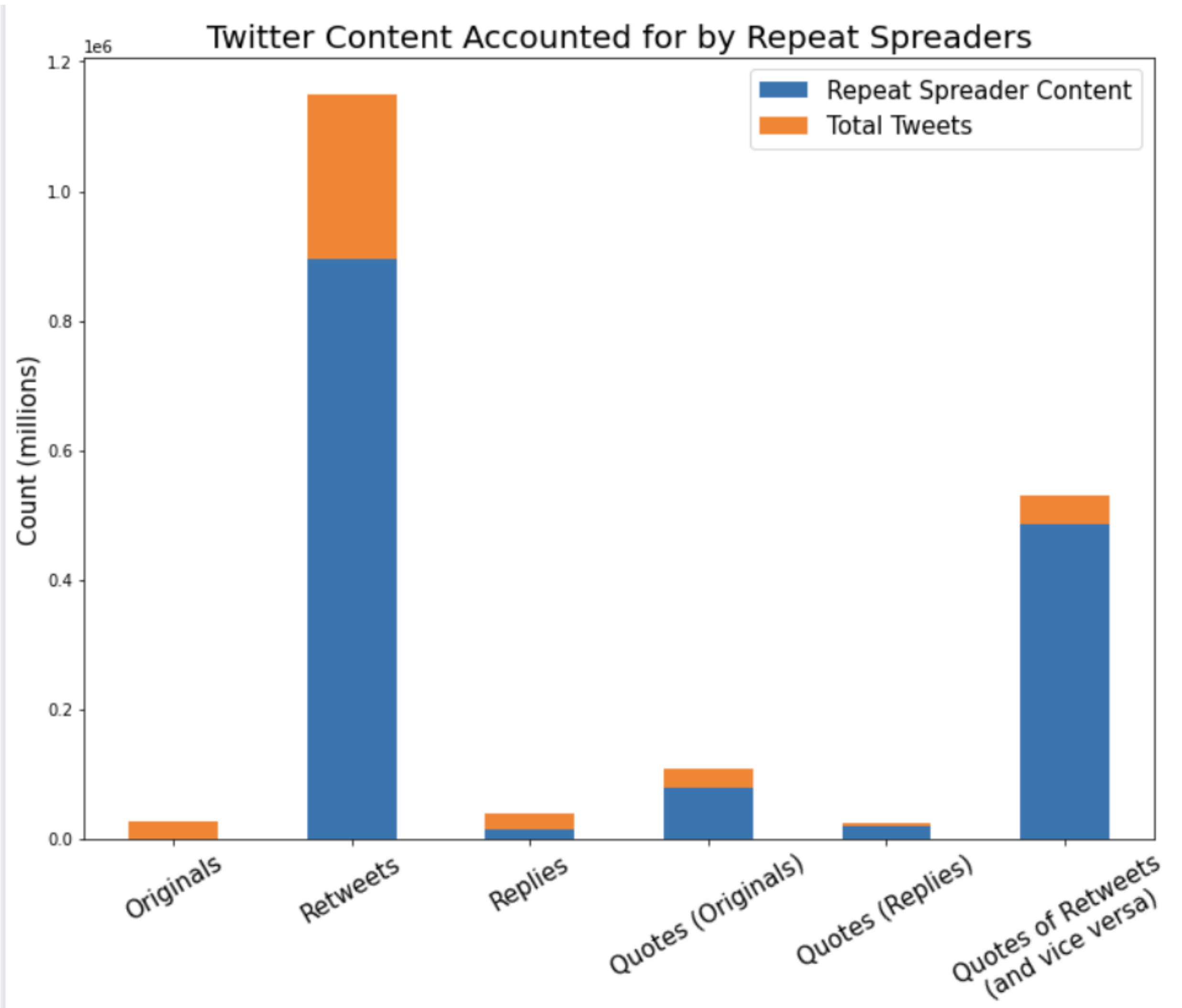
Twitter Content Accounted for by Repeat Spreaders

	Full Dataset Count	Repeat Spreader Count	Repeat Spreader %
Total tweets	1881491	1499288	79.686164
Originals	28010	843	3.009639
Retweets	1148647	895915	77.997418
Replies	40004	15654	39.131087
Quotes (Originals)	108858	80316	73.780521
Quotes (Replies)	25955	20758	79.976883
Quotes of Retweets (and vice versa)	530017	485802	91.657815

The figure above provides a summary of the Dominion dataset (“Full Dataset Count”) and a comparison showing the prevalence of “repeat spreaders” (“Repeat Spreader Count”) of mis/disinformation in the dataset. Repeat Spreaders are the accounts that were present in multiple incidents of mis/disinformation tracked by the Center for an Informed Public

(CIP). These users were tagged as repeat spreaders in the dataset and the tweets were classified into categories - original tweets, retweets, replies, original quote tweets, quote tweets as replies, quote tweets of retweets (and vice- versa). Finally, the proportion of the full sample that referenced repeat spreader content was calculated ("Repeat Spreader %"") for each of the categories.

From the figure, it can be observed that more than 70% of the tweets belonging to each category (except original and replies) were posted by repeat spreaders. Interestingly, only 3% of the total original tweets (tweets which are written by the user posting them) were posted by repeat spreaders. Hence, it shows that repeat spreaders were likely not posting any original tweets and were rather building up on the existing ones.



The figure above visually describes the summary data presented in the previous figure. It can be observed here that a significant proportion of tweets are generated by "repeat spreaders" of mis/disinformation.

4. Methodology

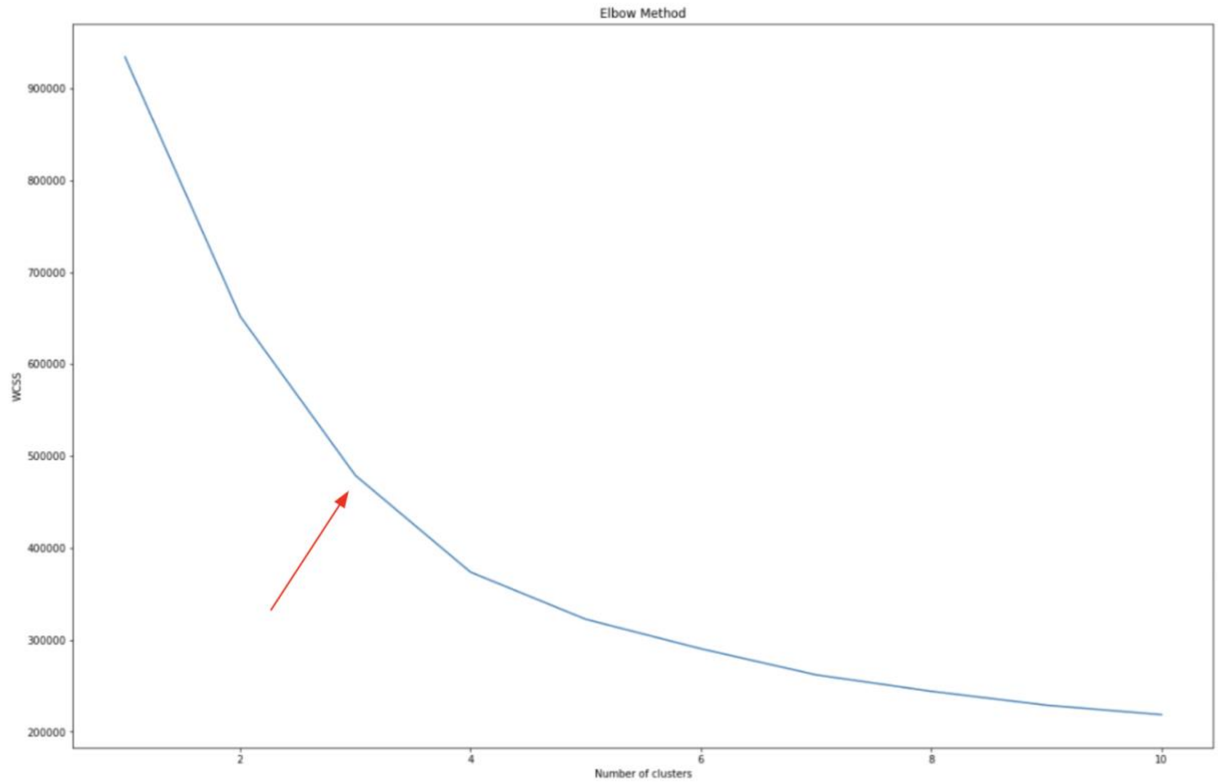
First we sampled 1/10 of the original dataset which led to a total 188,149 records because the original almost 2 million records was too large for our computer to handle efficiently. For the pre-processing process, we removed emojis, removed punctuation, removed retweets keywords, removed url, removed hashtags, removed emojis, removed mentions, removed quote tweets keywords, removed stopwords using NLTK package and finally did lemmatization.

Based on Phadke & Mitra (2021)'s method, we identified specific word clusters from LIWC (Tausczik and Pennebaker, 2010) and Moral Foundations Dictionary (Frimer et.al, 2019) dictionary. Specifically, we identified risk, reward, injustice, achievement, group identity, anger, opinion, and solicitation word clusters from two dictionaries. We used pre-trained word-to-vec (Mikolov et. al 2013) model to calculate the average vector of each word clusters. Then we used this centroid vector to calculate cosine similarity between this vector with each tweet record from our 188,149 records dataset. We generated 8 features based on the similarity between each tweet and each word cluster centroid vector. Each of these features was generated to capture different elements of speech that are associated with drives for participation, engagement in the movement, and strategies for mobilization present in social movements.

In addition to the features generated from tweet text, we also generated another 4 features focused on capturing social media behaviors based on interacting with other users/information through platform affordances. These feature vectors were generated per account and consisted of: proportion of average favorites (similar to likes on Facebook) of repeat spreader content (in place of extremist domains) of the user compared to the average favorites for all users; proportion of the average retweets of repeat spreader content compared to average retweets of all users; average replies to repeat spreader content compared to average replies to all users; and proportion of average number of all retweets and quote tweets of repeat spreader content compared to average number of all retweets/quote tweets.

After we had these 12 features in total, we normalized clustered the features using kmeans clustering. Finally we visualize our dataset by principal components analysis.

5. Results



KMeans clustering was performed on the features generated from word embeddings, LIWC comparison, and users' tweet information. The figure above shows the plot of Number of clusters against the WCSS (sum of squared distance between each point and the centroid in a cluster). Although the plot does not show a perfect elbow bend, we decided to form three clusters from the data.

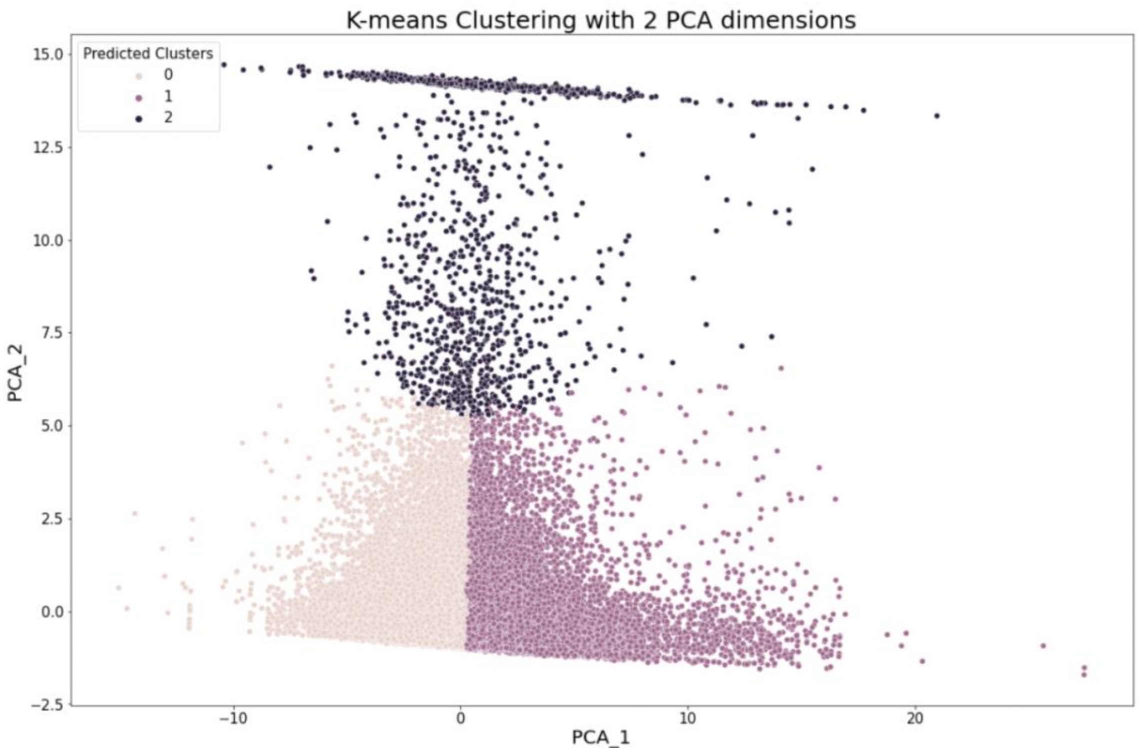


Figure: Scatter plot of identified clusters after reducing features to 2-dimensions using PCA

After KMeans clustering, 2-dimensional Principal Component Analysis was performed to visualize the clusters as seen in the figure above. To label the clusters, we selected 5 users closest to the centroid and extracted the top 10 most liked tweets of each of those users. All members of our team performed qualitative analysis of the tweets and based on the results of our analysis, came up with the following labels to classify the users:

Educator/Motivator: These were the accounts which shared intellectual content and/or portrayed a positive image of the agenda.

Amplifier/Sympathizer: These were the accounts that supported the movement but did not directly add anything to it. In the sample of tweets that we studied, Amplifiers/ Sympathizers were users who had only retweeted the twitter posts and did not post any original content.

Analyst: These were the accounts that critically analyzed the mis/disinformation content, either through supporting the conspiracy or debunking it.

Discussion/Implications/Limitations

While our efforts to reliably translate an extremist user-role taxonomy from Facebook to Twitter proved nontrivial due to differences in platform-specific affordances and social norms, our adapted methodology showed some promise in identifying mis/disinformation roles and trends within the Twitter environment. First, our analysis indicated some observable overlap in behavior across roles, most notably a high level of co-occurrence between "educator" and "motivator" behaviors. In addition, we were able to identify and construct a novel "analyst" user role which appeared to be unique to our Dominion dataset. Further study is required to determine whether this role is present across other mis/disinformation-based narratives as well.

One of the more foundational questions coming out of our analysis aimed to better understand the original methodology we had worked to adapt: Was the taxonomy identified by Phadke and Mitra an expression of extremist movements, an expression of Facebook, or an expression of extremist movements on Facebook? Understanding the nuanced differences between these interpretations could prove to have significant implications for any associated research.

It is also important to recognize the potential limitations of our methodologies. While we chose to use repeat spreaders as a proxy for extremist domains, there is not a perfect overlap between these groups and this could lead to some issues when translating to other domains. The generalizability of our results is also limited by the fact that disinformation narratives do not exist in isolation and are highly interrelated. To properly establish a user role taxonomy for mis/disinformation, we need a better picture of the "movement" instead of a specific strategic element. We could address this limitation in future works by expanding our analysis to include additional disinformation datasets. By doing so, we could work to identify which user roles are present across different narratives and track how these roles may change over time. Finally, we would also look to build on our work by engaging in further experimentation, validation, and refinement of our selected features and user roles.

References

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001

Benkler, Y., Tilton, C., Etling, B., Roberts, H., Clark, J., Faris, R., Kaiser, J., & Schmitt, C. (2020). Mail-In Voter Fraud: Anatomy of a Disinformation Campaign. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3703701

Collins, B. (n.d.). *QAnon's Dominion voter fraud conspiracy theory reaches the president*. NBC News. Retrieved December 25, 2021, from https://www.nbcnews.com/tech/tech-news/q-fades-qanon-s-dominion-voter-fraud-conspiracy-theory-reaches-n1247780

Election Integrity Partnership. (2021). *The long fuse: Misinformation and the 2020 election*. https://purl.stanford.edu/tr171zs0069

Frimer, R Boghrati, J Haidt, J Graham, and M Dehgani. 2019. Moral foundations dictionary for linguistic analyses 2.0. Unpublished manuscript (2019)

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2021). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, *53*(1), 232–246. https://doi.org/10.3758/s13428-020-01433-0

Howard, P., & Hussain, M. (2013). *Democracy's Fourth Wave?: Digital Media and the Arab Spring*. Oxford University Press. https://oxford-universitypressscholarship-com.offcampus.lib.washington.edu/view/10.1093/acprof:oso/9780199936953.001.0001/acprof-9780199936953

Kelly Garrett, R. (2006). Protest in an Information Society: A review of literature on social movements and new ICTs. *Information, Communication & Society*, *9*(2), 202–224. https://doi.org/10.1080/13691180600630773

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. https://doi.org/10.48550/arXiv.1301.3781

Nemer, D. (2021). Disentangling Brazil's Disinformation Insurgency. *NACLA Report on the Americas*, *53*(4), 406–413. https://doi.org/10.1080/10714839.2021.2000769

Ong, J., & Cabañes, J. V. (2018). *Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines*. https://doi.org/10.7275/2CQ4-5396

Phadke, S., & Mitra, T. (2021). Educators, Solicitors, Flamers, Motivators, Sympathizers: Characterizing Roles in Online Extremist Movements. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 310:1-310:35. https://doi.org/10.1145/3476051

Postill, J. (2014). Democracy in an age of viral reality: A media epidemiography of Spain's indignados movement. *Ethnography*, *15*(1), 51–69. https://doi.org/10.1177/1466138113502513

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare* (First edition.). Farrar, Straus and Giroux.

Starbird, K., Arif, A., & Wilson, T. (2019). Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 127:1-127:26. https://doi.org/10.1145/3359229

Swenson, A. (n.d.). *Smartmatic does not own Dominion Voting Systems*. AP NEWS. Retrieved December 25, 2021, from https://apnews.com/article/fact-checking-9740535009

Theocharis, Y., Lowe, W., van Deth, J. W., & García-Albacete, G. (2015). Using Twitter to mobilize protest action: Online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, *18*(2), 202–220. https://doi.org/10.1080/1369118X.2014.948035

Viebeck, E. (n.d.). Trump campaign debunked Dominion conspiracy theories, internal memo shows, days before backers kept spreading them. *Washington Post*. Retrieved June 6, 2022, from https://www.washingtonpost.com/politics/2021/09/22/trump-dominion-giuliani-powell-memo/