How did an election fraud narrative spread online? Testing theories using machine learning and natural language processing

Loni Hagen
University of South Florida, FL, USA
lonihagen@usf.edu

Diego Ford
University of South Florida, FL, USA
diegoford@usf.edu

Jahnae Edwards
University of South Florida, FL, USA
jahnae@usf.edu

Ly Dinh
University of South Florida, FL, USA
lydinh@usf.edu

Nic DePaula
SUNY Polytechnic Institute, Utica, NY, USA
nfvd@proton.me

Joshua Scacco
University of South Florida, FL, USA
jscacco@usf.edu

Abstract

In this study, we investigated 1) how election fraud narratives propagated on social media, and 2) the role of influential actors in the process of building and spreading the election fraud narratives. We applied machine learning (ML) and natural language processing (NLP) methods to examine Twitter data related to an election fraud narrative following the 2020 Presidential election. We identified influential actors and found evidence for the former President's use of social media to cue group identity.

Introduction[1]

On November 7, 2020, President Biden was declared the President-elect. At the same time, election fraud narratives were spread online rapidly. Among many narratives, digital messages falsely claimed that Dominion Voting Systems were compromised to change the election results. We selected this case as a representative election fraud narrative, based on the court decision (U.S. Dominion, Inc. v. Fox News Network, LLC, 2023). This research aims to investigate how election fraud narratives regarding Dominion Voting Systems spread online following the 2020 election. Integrating social scientific and AI-based computational methods, we examine the role of influential actors in the process of building and spreading such false narratives.

Related Work

Information propagation on social media is influenced by various factors that can be analyzed through investigating technical properties such as temporal, structural, and linguistic characteristics (Kwon et al., 2013). We explore the narrative properties associated with how political elites create protagonistic and antagonistic entities in messaging. Social identity theory (SIT) concerns how individuals relate to perceived in-groups and out-groups (Tajfel, 1982). Political leaders, in attempts to cue group identity amid audience segmentation by information sources, have messaged increasingly to gender, race, ethnic, sexual, and political identities over time (Scacco & Coe, 2021). As president, Donald Trump cued group identity by constructing messages around content sources and individuals deemed as part of in-groups and out-groups (Scacco & Wiemer, 2019). Such message construction highlights perceived protagonists and antagonists and may provide coherence and fidelity for the audience (Fisher, 1985).

Researchers have adopted NLP and machine learning techniques to build predictive models to detect misinformation (Roy et al., 2023) or to predict credibility of Twitter accounts (Saxena et al., 2023) to combat false information propagation. Similarly, social science scholars have focused on understanding the mechanisms of information propagation, including how shared messaging improvisation online between elite figures and conversation participants can propagate narratives (Starbird et al., 2023). We anticipate that Trump and an affiliated set of actors co-constructed friendly content sources to highlight "in-group" entities for supporters to attend to following the election. Simultaneously, we expect an opposing co-construction of antagonistic out-groups from media outlets and individuals who challenged false Dominion messaging.

Methodology

Data Collection

Twitter data was collected through a Twitter firehose, namely Brandwatch. Brandwatch allows data collection from major social media sources. The daily limit on the total number of Twitter data is set at 50,000 by Brandwatch. We decided 'dominion voting' as the query for data collection mainly to keep the volume of data within the limit. The date range was set as Nov 1st and Nov 15th of 2020 to capture from the initiation of the campaign, and the peak of the information propagation. The dataset included a total of 29,251 unique accounts who created 5,337 original tweets and 37,114 retweets.

Methods

Network Analysis.

We constructed networks based on unique Twitter accounts (nodes), and retweets (edges). When A retweet B's post, we consider that A endorsed B's idea. Therefore, there is a directed edge from A to B in this case. For community detection, we used modularity algorithm by Newman (2006). For detecting the influential actors, we used PageRank algorithm developed by Google (Brin & Page, 1998). In the network analysis literature, PageRank identifies influential nodes by considering not just how often they are endorsed by other nodes, but also the influence level of those endorsing nodes (Easley & Kleinberg, 2010). To test the effect of @realDonaldTrump, we compared statistics of the network structure with and without the presence of @realDonaldTrump from the network. We used Gephi for network analysis, which is an open-source tool for network analysis and visualization (Bastian et al., 2009).

Named Entity Detection.

We used FlairNER with Ontonotes18-class training data, a transformer-based NLP model for named entity recognition tasks, to automatically detect locations and person names from the posts (Akbik et al., 2019; Hu et al., 2023).

Duplicate Detection.

Duplicate detection task was employed to identify near-duplicate contents to deal with tweets that are communicating identical content but with slight modification of the text. Near duplicate text detection can inform possible coordinated efforts to spread pre-formatted or identical narratives. We used FuzzyWuzzy library in Python for fuzzy matching, setting a threshold fuzz ratio of 80% to determine near-duplicate text strings between pairs of tweets.

Visualization.

Outcomes from the NLP and machine learning techniques we applied are numeric values. Network visualization is an effective tool to find patterns and to efficiently communicate the results (Hagen et al., 2019; Shmueli et al., 2017). To visualize influential actors in the network, we used ForceAtlas 2 and Network 3D algorithms using Gephi.

Results & Discussions

Influential actors.

Based on the PageRank value, the most influential actor in this retweet network is @realDonaldTrump (President Trump). The next level influential nodes belong to @EmeraldRobinson (Emerald Robinson, a former conservative reporter for NEWSMAX), @CodeMonkeyZ (a conservative digital strategist), and @LLinWood (Lin Wood, a former attorney).

Trump effect.

The first day @realDonaldTrump was detected in our data was on November 12th. By comparing network structure with @realDonaldTrump and without, we found Trump effects to this conversation network. First, President Trump brought 4,755 actors to the conversation community. Compared to the network with Trump, when we excluded Trump account, the total number of accounts decreased by 20 percent. Second, otherwise fringe conspiracy theory on election fraud may have become major news with Trump's endorsement. With Trump's participation, overall network diameter, network size, number of weakly connected components, and modularity increased. These structural indicators suggest that fringe ideas were becoming increasingly prominent in the news.

Fox News was treated as an out-group.

Trump, in this case study, treated OANN as an in-group to build his agenda of election fraud and treated Fox News as an out-group. Our visualization analysis, with the support of the court document (U.S. Dominion, Inc. v. Fox News Network, LLC, 2023), shows that this tactic could have been effective on building the election fraud narratives.

Locations indicating coordinated actions.

The comparison of frequencies of location mentions between near-duplicate and no-duplicate texts shows that some location names such as Arizona, China, Nevada, Pennsylvania were repeated frequently. This indicates that tweets including these location names could have been supported by possible coordinated efforts.

Limitation and Future Direction

This work has several limitations, partially due to its preliminary nature. Data was collected in 2023, two years after the case. Some of the Twitter accounts could have been deleted or suspended, which then we were not able to collect. In the future, we plan to systematically validate the findings further, and to investigate the mechanisms by which election fraud narratives spread rapidly online.

Acknowledgment

This work was partially supported with a grant from the Florida Center for Cybersecurity.

References

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP. In W. Ammar, A. Louis, & N. Mostafazadeh (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)* (pp. 54–59). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-4010

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), Article 1. https://doi.org/10.1609/icwsm.v3i1.13937

Fisher, W. R. (1985). The narrative paradigm: An elaboration. *Communications Monographs, 52*(4), 347-367.

Hagen, L., Keller, T. E., Yerden, X., & Luna-Reyes, L. F. (2019). Open data visualizations and analytics as tools for policy-making. *Government Information Quarterly*, *36*(4), 101387. https://doi.org/10.1016/j.giq.2019.06.004

Hu, X., Zhou, Z., Li, H., Hu, Y., Gu, F., Kersten, J., Fan, H., & Klan, F. (2023). Location Reference Recognition from Texts: A Survey and Comparison. *ACM Computing Surveys*, *56*(5), 112:1-112:37. https://doi.org/10.1145/3625819

Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013). Prominent Features of Rumor Propagation in Online Social Media. *2013 IEEE 13th International Conference on Data Mining*, 1103–1108. https://doi.org/10.1109/ICDM.2013.61

Newman, M. E. J. (2006). Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577–8582.

Roy, P. K., Tripathy, A. K., Weng, T.-H., & Li, K.-C. (2023). Securing social platform from misinformation using deep learning. *Computer Standards & Interfaces*, *84*, 103674. https://doi.org/10.1016/j.csi.2022.103674

Saxena, N., Sinha, A., Bansal, T., & Wadhwa, A. (2023). A statistical approach for reducing misinformation propagation on twitter social media. *Information Processing & Management*, *60*(4), 103360. https://doi.org/10.1016/j.ipm.2023.103360

Scacco, J. M., & Coe, K. (2021). The ubiquitous presidency: Presidential communication and digital democracy in tumultuous times. Oxford University Press.

Scacco, J. M., & Wiemer, E. C. (2019). The president tweets the press: President-press relations and the politics of media degradation. In R. Davis & D. Taras (Eds.), *Power Shift? Political Leadership and Social Media* (pp. 17-32). Routledge.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl, K. C., Jr. (2017). *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/usf/detail.action?docID=5014630

Starbird, K., DiResta, R., & DeButts, M. (2023). Influence and Improvisation: Participatory Disinformation during the 2020 US Election. *Social Media + Society*, *9*(2), 20563051231177943. https://doi.org/10.1177/20563051231177943

U.S. Dominion, Inc. v. Fox News Network, LLC (2023).

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual review of psychology, 33*(1), 1-39.