

# Tracking Fringe and Coordinated Activity on Twitter Leading Up to the US Capitol Attack

Padinjaredath Suresh Vishnuprasad<sup>1,2,5</sup>, Gianluca Nogara<sup>1</sup>, Felipe Cardoso<sup>1</sup>,  
Stefano Cresci<sup>3</sup>, Silvia Giordano<sup>1</sup>, Luca Luceri<sup>1,4</sup>

<sup>1</sup>University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Switzerland;

<sup>2</sup>Indian Institute of Science Education and Research (IISER) Tirupati, India;

<sup>3</sup>Istituto di Informatica e Telematica CNR, Italy;

<sup>4</sup>USC Information Sciences Institute, USA;

<sup>5</sup>QUT Digital Media Research Center, Australia

vishnu.prasad@hdr.qut.edu.au, {gianluca.nogara, silvia.giordano}@supsi.ch,  
cardosodesantana.89@gmail.com, stefano.cresci@iit.cnr.it, lluceri@isi.edu

## Abstract

The aftermath of the 2020 US Presidential Election witnessed an unprecedented attack on the democratic values of the country through the violent insurrection at Capitol Hill on January 6th, 2021. The attack was fueled by the proliferation of conspiracy theories and misleading claims about the integrity of the election pushed by political elites and fringe communities on social media. In this study, we explore the evolution of fringe content and conspiracy theories on Twitter in the seven months leading up to the Capitol attack. We examine the suspicious coordinated activity carried out by users sharing fringe content, finding evidence of common manipulation techniques ranging from targeted amplification to manufactured consensus, which went likely undetected by Twitter. Further, we map out the temporal evolution of, and the relationship between, fringe and conspiracy theories, which eventually coalesced into the rhetoric of a stolen election, with the hashtag #stopthesteal, alongside QAnon-specific narratives. Our findings highlight how social media platforms offer fertile ground for the widespread proliferation of conspiracies during major societal events, which can potentially lead to offline actions and organized violence.

## Introduction

The 2020 US Presidential Election took place against the backdrop of incumbent president Donald Trump's impeachments, the COVID-19 pandemic, and racial unrest of the Black Lives Matter movement. In such a tumultuous scenario, Donald Trump challenged the integrity of the elections by diffusing claims of alleged, widespread voter fraud and claiming his victory despite the vote count (Election Integrity Partnership 2020). Activism from fringe groups and Trump supporters gained traction online in the months following the election, calling for violence against those who opposed Trump's claims (Pennycook and Rand 2021). Following Trump's instruction ("Be there, will be wild!") in one of his prominent tweets,<sup>1</sup> on January 6th, 2021, as Congress

convened to ratify the electoral college results, his supporters attacked the US Capitol, breaching the Capitol Building and causing five deaths and multiple injuries (Election Integrity Partnership 2020).

The US Capitol attack was seen as an attempted *coup d'état* (Jacobson 2021), organized and promoted online, with violent actions and repercussions offline (Election Integrity Partnership 2020; Luceri, Cresci, and Giordano 2021). In the aftermath of the attack, the US House of Representatives established a Select Committee to investigate the Capitol attack, and the role of social media in relation to the "spread of misinformation, efforts to overturn the 2020 election, domestic violent extremism, and foreign influence in the 2020 election".<sup>2</sup> In their final report,<sup>3</sup> released on December 22nd, 2022, the Select Committee examined former President Donald Trump's conduct to invalidate the election results and foster the attack to the Capitol.<sup>4</sup> Besides Trump's potentially pivotal role in driving the attack, which is out of the scope of this paper, the report highlights how Twitter, among all the subpoenaed social media platforms, served as the main channel for Trump, his supporters, and fringe groups, such as QAnon, to amplify conspiracy theories and plan the assault. The attack ensued by large-scale deplatforming, with Twitter (until November 2022), Facebook, Instagram, and Snapchat, among the others, banning Donald Trump and many of his followers, who flocked together to alt-tech social media platforms (Pennycook and Rand 2021).

The research community examined the effects of these interventions (Horta Ribeiro et al. 2021; Trujillo and Cresci 2022) and studied online conversations surrounding the election from different angles and various platforms (La Gatta et al. 2023). The Election Integrity Partnership (EIP) released an observational study tracking misleading claims related to the integrity of the electoral system from the run-up to the aftermath of the election (Election Integrity Partnership 2020). Among all the fringe groups, the QAnon con-

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup><https://www.justsecurity.org/84551/important-elements-of-the-january-6th-report/>

<sup>2</sup><https://techpolicy.press/january-6-committee-issues-subpoenas-to-social-media-platforms/>

<sup>3</sup><https://perma.cc/ZA6R-JRF2>

<sup>4</sup><https://www.justsecurity.org/84658/insiders-view-of-the-january-6th-committees-social-media-investigation/>

spiracy played a relevant role in steering conversations along the line of the election fraud rhetoric (Wang et al. 2023; Sharma, Ferrara, and Liu 2022). While a large body of research focused on the voting event, a longitudinal study that examines the unfolding of conspiracy and fringe narratives leading up to the Capitol attack is still lacking.<sup>5</sup>

In this paper, we investigate the proliferation and coordinated amplification of conspiratorial content on Twitter in the months preceding the 2020 US Election and the Capitol attack. Leveraging a dataset of over 500M election-related tweets gathered from July 2020 to January 2021, we deconstruct the *Twittersphere* to investigate the temporal evolution of, and the relationship between, fringe narratives. Further, we examine the potential use of coordinated and manipulative actions to amplify content and create an illusion of consensus around conspiracies. In particular, we address the following research questions (RQs):

**RQ1:** *Were fringe narratives amplified employing manipulation techniques? Who were the actors responsible for the manipulation, and which stories were pushed?*

**RQ2:** *What was the relationship between the various fringe narratives and conspiracy theories? How did they evolve over time and converge at the time of the Capitol attack?*

To address these RQs, we rely on and build upon computational models to chart the landscape of conspiratorial and coordinated activity surrounding the 2020 US Election, leading up to the January 6th attack, providing the following contributions:

- Employing existing approaches for the detection of coordinated and inauthentic activity (Pacheco et al. 2021; Nizzoli et al. 2021; Luceri et al. 2024), we found evidence of manipulation attempts to amplify and elicit an illusion of consensus around conspiratorial content.
- Leveraging manually curated annotations (Kennedy et al. 2022; Abilov et al. 2021), we observed how the “Stop the Steal” rhetoric was by far the narrative mostly amplified by coordinated actors. Further, we uncovered and verified that the vast majority of users involved in such manipulation actions were supporting and promoting claims about voter fraud (Abilov et al. 2021).
- We built on prior work (Sato et al. 2021) to map out the temporal evolution of fringe narratives and to draw a broad picture of their relationships. We discovered that conspiratorial and fringe content coalesced into two distinct, yet interlaced, meta-narratives embodying the “stolen election” rhetoric and the QAnon conspiracy. Our results experimentally confirm the observational findings released by the Election Integrity Partnership (2020).

---

<sup>5</sup>For the sake of simplicity, we use the terms “fringe narratives” and “conspiracy theories” interchangeably throughout the paper. However, in our results, we clarify and distinguish these two distinct instances of problematic content.

## Related Work

### Misinformation and the 2020 US Election

The 2020 US Presidential Election saw an unprecedented number of false claims and misinformation about voter fraud, including the self-proclamation of Donald Trump as the real winner (Chen et al. 2021; KhudaBukhsh et al. 2022). In this context, many studies tracked the spread of electoral misinformation in the run-up to the election and its aftermath. Diverse collaborative and participatory schemes were used to spread questionable narratives (Election Integrity Partnership 2020). Some of these were *bottom-up*, where influential media, political elites, and social media influencers broadcasted content produced by average users. Other times instead, the patterns of diffusion were *top-down*, involving large numbers of users, automated (Ferrara et al. 2020) or otherwise (Jachim, Sharevski, and Pieroni 2021), that coordinated to amplify narratives proposed by influential users.

Among the plethora of intertwined false narratives, “Stop the Steal” gave rise to an advocacy organization responsible for multiple riots across the country (Luceri, Cresci, and Giordano 2021). The “Stop the Steal” campaign spread across multiple platforms (La Gatta et al. 2023) pushing allegations that the election was stolen and inciting Trump supporters to overturn the result (Benkler et al. 2020; Election Integrity Partnership 2020): One year after the election, up to one-third of the US population still believed that the electoral outcome was not to be trusted (Matatov, Naaman, and Amir 2022; Pennycook and Rand 2021). The large body of work on the 2020 US Election also resulted in the collection and publication of several resources, such as social media datasets of annotated posts and users involved in the spread of election misinformation (Chen, Deb, and Ferrara 2022; Kennedy et al. 2022; Abilov et al. 2021). In the present study, we leverage a combination of existing resources to provide unprecedented insights into the coordinated strategies used to amplify false narratives and how these contributed to sparking the Capitol Hill attack.

### QAnon and Fringe Conspiracy Theories

Interest in tracing the diffusion and in identifying supporters of conspiracy theories is motivated by their detrimental effect on people’s political engagement and trust in authorities (Green et al. 2022), so much so that the interplay between the spread of conspiracy theories and major political events is long studied. QAnon was the main conspiracy involved in the US 2020 electoral debate (Sharma, Ferrara, and Liu 2022; Wang et al. 2023). QAnon conspiracy theories are associated with a fringe political movement that centers around the belief in an anonymous user claiming to possess a high-level security clearance within the US Government (Papasavva et al. 2021). Followers of QAnon assert that a secretive group of Satanic, cannibalistic individuals engaged in global child sex trafficking conspired against Donald Trump. The interested reader can refer to the *Appendix* for a detailed overview of QAnon’s theories.

Sharma, Ferrara, and Liu (2022) found that around 66% of all users that were active in the online debate interacted with QAnon content and that far-right QAnon supporters ac-

tively tried to persuade left-leaning users. Among the many narratives that originated from QAnon, the “Hammer and Scorecard” and “Dominion” theories were particularly popular during the 2020 Election. These narratives were pushed by a combination of prominent users (e.g., Donald Trump and other politicians), far-right communities (e.g., the Proud Boys), and other average users (Election Integrity Partnership 2020), including automated accounts (Ferrara et al. 2020; Luceri, Cardoso, and Giordano 2021).

Moreover, the proponents of these extremist conspiracies also endorsed the election-fraud claims, up to the point that QAnon conspiracies coalesced in and reinforced the “Stop the Steal” campaign, favoring the radicalization of Trump supporters (Pennycook and Rand 2021; Luceri, Cresci, and Giordano 2021). This intricate interplay poses inevitable challenges towards tracing the spread of conspiracies, understanding how they unfold (Papasavva et al. 2022), and what leads users to adopt the theories (Wang et al. 2023).

### Coordinated Online Behavior

A frequent characteristic of mis- and disinformation campaigns is the use of *coordinated activity* to boost the spread of the shared messages (Nizzoli et al. 2021). A recent body of work thus focused on detecting and characterizing coordinated online behavior as a way to surface influence campaigns (Luceri et al. 2024). The majority of existing approaches detect online coordination by identifying unexpected or exceptional similarities in the action sequences of two or more users. For example, many works model common user activities (e.g., co-retweets, temporal and linguistic similarities) to build user similarity networks where coordinated groups (i.e., communities) of users are identifiable (Weber and Neumann 2021; Pacheco et al. 2021; Nizzoli et al. 2021; Magelinski, Ng, and Carley 2022; Luceri et al. 2024). Some methods do not only distinguish between coordinated and non-coordinated users, as in a binary classification task, but also quantify the extent of coordination thus providing more nuanced results (Nizzoli et al. 2021).

The characterization of coordinated users and communities is another thriving area of research (Ezzeddine et al. 2023). This task is typically aimed at distinguishing between malicious and benign forms of online coordination (Vargas, Emami, and Traynor 2020; Hristakieva et al. 2022). In fact, coordination is a necessary ingredient not only of manipulation campaigns, but also of many other social movements, including fandoms, activists, and protesters (Nizzoli et al. 2021). Longitudinal analyses of the coordinated actions of social movements revealed that both offline and online events can spark online coordination (Varol et al. 2014; Casas and Williams 2019). Importantly, the opposite is also true, since online coordination often precedes real-world action (Steinert-Threlkeld et al. 2015). These results demonstrate the importance of characterizing the multiple forms of online coordination, given the real-world consequences that they can have. To distinguish between neutral/benign and malicious online coordination, some considered the structural properties of the coordination networks (Vargas, Emami, and Traynor 2020; Nizzoli et al. 2021; Cinelli et al. 2022; Luceri et al. 2024), the reliability of the news they

US Election	QAnon	COVID-19
#hammerandscorecard	#pizzagate	#plandemic
#sharpiegate	#qanon	
#qsntch	#qarmy	
#stopthesteal	#taketheoath	
#dobbs	#wwglwga	
#dominionsoftware	#projectveritas	
#dominion	#thegreatawakening	
#hammer	#civilwar	
#scorecard	#obamagate	

Table 1: Fringe hashtags analyzed in this paper.

shared (Cao et al. 2015), or their use of propagandist language (Hristakieva et al. 2022), to gain insights into the objectives and intent of the coordinated users. Here, we employ state-of-the-art coordination detection frameworks (Weber and Neumann 2021; Pacheco et al. 2021; Nizzoli et al. 2021) to identify users that participated in coordinated activities with the goal of spreading conspiratorial content, and we leverage curated annotations (Kennedy et al. 2022; Abilov et al. 2021) to characterize the messages and users involved in such manipulation attempts.

### Dataset & Fringe Hashtags

We leverage the first public Twitter dataset about the 2020 US Presidential Election (Chen, Deb, and Ferrara 2022), collected using election-specific keywords and gathering messages from a set of political figures, including the running candidates. This dataset was gathered in real-time by leveraging the 1% stream of tweets provided by Twitter’s streaming API. From the list of released tweet IDs (Chen, Deb, and Ferrara 2022), we applied a rehydration process to collect the full tweet objects. For our analyses, we use a subset of the whole dataset considering only tweets from July 2020 to January 2021, to cover an observation period from the final stage of the political campaign to the Capitol attack. This results in a longitudinal dataset that contains about 568 million tweets, consisting of 52% of retweets, 19% of replies, 16% of original tweets, and 13% of quotes. The rehydration allowed us to retrieve about 51.72% of the tweets collected by (Chen, Deb, and Ferrara 2022) in our window of observation. The missing tweets were likely deleted or removed by Twitter’s intervention against QAnon (Wang et al. 2023).

Leveraging previous analyses on the same dataset (Ferrara et al. 2020; Sharma, Ferrara, and Liu 2022; Wang et al. 2023; Chen, Deb, and Ferrara 2022) and diverse reports on the Capitol attack (Pennycook and Rand 2021; Election Integrity Partnership 2020; Luceri, Cresci, and Giordano 2021), we identified nineteen hashtags pushed by conspiracies and fringe communities, which we refer to as *fringe hashtags*. In this body of literature, the fringe hashtags are commonly grouped into three broad categories related to (i) QAnon – hashtags originated from and propagated by the QAnon conspiracy; (ii) US Election – hashtags encompassing a diverse set of conspiracies supporting the election fraud rhetoric; and (iii) COVID-19 – a single hashtag related to a conspiracy theory about the pandemic. These categories and corresponding hashtags are summarized in

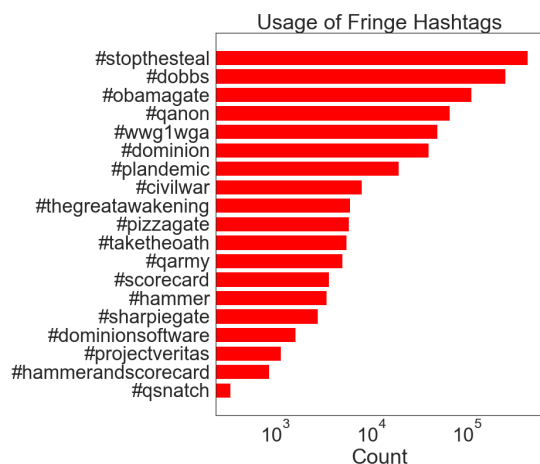


Figure 1: Distribution of fringe hashtags.

the *Appendix*. The three groups of fringe hashtags are displayed in Table 1 while Figure 1 depicts their distribution in our dataset. Notably, #stopthesteal – promoting the stolen election narrative, and #dobbs – supporting the former Fox News political commentator Lou Dobbs, were by far the most shared hashtags, followed by #obamagate, #qanon, #wwg1wga, #dominion, #plandemic, and #civilwar.

## Methodology

To address the proposed RQs, we leverage three methodologies. We explore suspicious coordinated activity relying on existing techniques that (i) identify networks of *rapid retweet* interactions – generally employed to amplify content generated by targeted users (Mazza et al. 2019; Pacheco et al. 2021), and (ii) discover clusters of users sharing similar tweets – usually shared with small variations and nuances to create an illusion of consensus (Weber and Neumann 2021). We refer to these two network-based models as *rapid retweet network* and *coppypasta network*. In regard to the suspicious coordinated activities surfaced by our methodologies, we remark that Nizzoli et al. (2021) highlighted the difference between coordination and inauthenticity, admitting the existence of a spectrum of behaviors that ranges from benign to neutral and malicious coordinated activities. However, since the hashtags analyzed in our work endorse empirically disproved conspiracy theories, we posit that the coordination found with our methodologies relates to malicious and inauthentic activities carried out with the intent to misinform.

Further, we characterize (i) the users involved in such manipulation attempts by leveraging the annotations produced by Abilov et al. (2021), who classified users as either promoters or detractors of voter fraud narratives; and (ii) the tweets shared by these coordinated accounts relying on the curated dataset from Kennedy et al. (2022) to label messages sowing doubts in the legitimacy of the election. Finally, we build on the approach proposed by Sato et al. (2021) to chart the landscape of fringe narratives by mapping out their relationship and evolution over the months leading up to the Capitol attack. We refer to this augmented model as *Hashtag Temporal Evolution Mapping (HTEMMap)*.

## Rapid Retweet Network

Retweets represent the quickest and easiest form of social endorsement on Twitter (Metaxas et al. 2015), and they signal a high level of trust and agreement with the shared message. Users can re-share others’ content in a simple, *one-click* operation, which can even be automated (Luceri et al. 2019). However, the simplicity behind this action led to its nefarious use in orchestrated manipulation campaigns (Bessi and Ferrara 2016), where retweets were collaboratively employed to spread and artificially amplify information shared by influential users (Mazza et al. 2019). We remark that quoted tweets, which are possibly used to express criticism rather than endorsement, are not considered in this analysis.

Given the widespread misuse of retweets, in this work, we aim to identify networks of accounts that suspiciously employ retweets in a fast, repeated, and coordinated fashion. Following the methodology presented by Pacheco, Flammini, and Menczer (2020), we define a rapid retweet as a re-share of a tweet that occurs in a narrow time window. Here, we consider different time window sizes to assess the strength of this coordinated activity. According to this definition, we build a weighted *rapid retweet network*, where the user that re-shared a tweet (*source*) is connected to the user that generated the original content (*target*) if the retweet occurred within the designated time window size, and the weight represents the number of rapid retweets between the source and target users. Similarly to Pacheco, Flammini, and Menczer (2020), we also filter out occasional instances of rapid retweets that might happen by chance by considering only links with a weight larger than one.

## CopyPasta Network

Similarly to retweets, content with high similarity can be used to resonate messages and elicit the idea of widespread agreement around an idea, a piece of information, or a misleading claim. For the sake of simplicity, and leveraging the Internet jargon, we refer to tweets with high textual similarity as *coppypasta tweets*. To identify such tweets, we carry out a pair-wise comparison of the text of original content shared in our Twitter dataset, thus including replies, quotes, and original tweets. We do not consider retweets in this subset as, by definition, they have exactly the same text and cannot be considered as an instance of *coppypasta tweet*.

Calculating pairwise similarity between a huge volume of tweets is, however, a significant resource- and time-consuming task. Similarly to (Pacheco, Flammini, and Menczer 2020), we sort tweets in chronological order and compare only messages shared in close time proximity, considering a sliding window of ten tweets. For each time window, we construct a  $10 \times 10$  matrix of tweet similarity scores, which are in turn computed using the Ratclif/Obershelp algorithm (Ratclif 1988). This computation forms the basis to build the *coppypasta network*, defined as an undirected network of tweets linked to each other if their similarity score is above a certain threshold. As we thoroughly show in the following sections, this threshold is not chosen arbitrarily, but rather it is adequately selected by observing the distribution of similarity scores of both fringe and generic tweets.

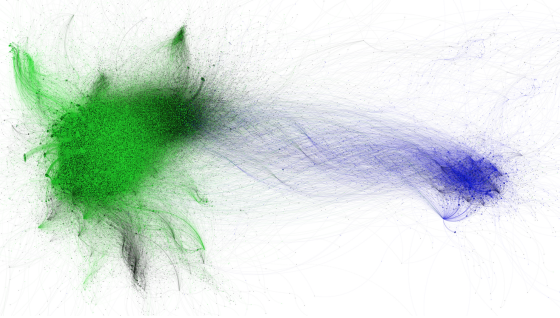


Figure 2: Retweet network of detractors (blue) and promoters (green) of the voter fraud conspiracy, along with unlabeled users (black). The node size represents the in-degree.

### Characterizing Users and Tweets

To characterize the users involved in the spread of fringe narratives, we leverage a publicly-available annotation released by Abilov et al. (2021). There, the authors classify users as either *detractors* or *promoters* of the voter fraud rhetoric and specific conspiracy theories. Leveraging a rich set of information (shared URLs, images, and YouTube videos), they discovered five distinct communities, where *detractors* sit in a unique community and *promoter* populate the other four communities. Here we use the annotations of Abilov et al. (2021) to label users as *detractors* or *promoters* and to understand their role in the coordinated inauthentic activities described before. The overlap of users with Abilov et al. (2021) resulted in a labeling of 63.25% of the users in our dataset and as we could expect, the majority of those (81.5%) represent *promoters* of the voter fraud rhetoric. It is worth noting the relevant fraction of users (18.5%) labeled as *detractors* in our dataset. This is not surprising, however, given that *detractors* of conspiracies also make use of controversial hashtags and quote misleading claims to attack or convert fringe users (Wang et al. 2023). Abilov et al. (2021) also established a connection between users’ classification and their political alignment, with AllSides and Media Bias/FactCheck serving as references. Their research revealed that *promoters* generally align with a conservative orientation, whereas *detractors* align with a liberal orientation.

Figure 2 represents the interactions between detractors and promoters with a directed, weighted retweet network, where the nodes are users, edges represent retweets, and the weight is proportional to the frequency of retweets between each pair of nodes. The network in Figure 2 is visualized using a force-directed layout (ForceAtlas2), where nodes repulse each other, whereas edges attract their nodes. This results in a natural split of the network into two communities entirely populated by users with the same opinion about the voter fraud rhetoric. Unlabeled users are balanced between the community of promoters (53.23%) and the community of detractors (45.17%).<sup>6</sup> In the next sections, we delineate the role of these two distinct communities in the manipula-

<sup>6</sup>Note that the percentages do not sum up to 100% as a small fraction of unlabeled nodes does not belong to any community.

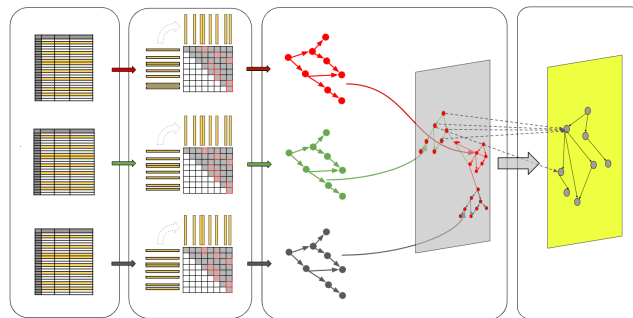


Figure 3: The HTEMap Model: The steps from, left to right, are tagset identification, similarity scoring, network building, and network collation & hashtag network inference

tion attempts of amplifying fringe narratives.

Additionally, we characterize a subset of tweets in our dataset by leveraging the annotations curated by Kennedy et al. (2022), which manually associated tweets to stories on the delegitimization of the election. Since the analysis by Kennedy et al. (2022) involves a much shorter time window than that analyzed by us (i.e., it overlaps with ours for 90 days, which corresponds to 28% of the whole time covered by our data), their annotations only cover 21.48% of all tweets in our dataset. Nonetheless, when combined with our analyses of the manipulation tactics adopted in the run-up to the election, such annotations allow us to understand which stories were artificially amplified. It is important to note that we use the dataset provided by (Kennedy et al. 2022) to complement and expand our hashtag-based study with a meta-level analysis of manually-annotated tweets. Notably, the annotations by Kennedy et al. (2022) are characterized by a significant agreement between the human coders (Fleiss’s  $\kappa$  score = 0.64). The top five misleading stories in our dataset are related to the “Stop the Steal” narrative, named *Stop the Steal Pushed* (46.71%) and *Stop the Steal rallies* (4.03%) by Kennedy et al. (2022), followed by the *Dominion* (38.87%), *Hammer and Scorecard* (4.52%), and *Sharpiegate* (1.54%) conspiracy theories.

### HTEMap Model

To delineate the relationships between the various fringe narratives and map out their temporal evolution, we augment the model proposed by Sato et al. (2021) to introduce the *Hashtag Temporal Evolution Mapping (HTEMap)* model. To construct a temporal network of hashtags, HTEMap infers an evolutionary trajectory of hashtags by first charting the relationship between tweets and hashtags across time as in Sato et al. (2021), and then building a co-occurrence network of hashtags. The architecture of the model is displayed in Figure 3 and consists of four steps.

First, for every hashtag, we extract a hashtag dataset (in short *tagset*), which includes the tweets embedding the target hashtag, sorted in reverse-chronological order.

Second, we perform a pairwise comparison of the tweets in the tagset. We represent each tweet with a  $n$ -dimensional vector (in our setting  $n = 19$ , given that we are considering

19 fringe hashtags) whose elements,  $b_k$ , represent the number of occurrences of the  $k$ -th hashtag in the tweet. We build a similarity matrix  $S$ , where every element  $S_{i,j}$  represents the cosine similarity between the  $n$ -dimensional representations of tweets ( $i$ ) and ( $j$ ).

Third, we use the similarity matrix  $S$  to construct a graph of tweets. Specifically, we consider only the upper-triangular matrix of  $S$  to select the most similar pair of tweets and connect their corresponding nodes with a directed edge that goes from the older one to the newer one. Once this operation for every tweet in the tagset is performed, we obtain an evolutionary tree for the target hashtag. By repeating the previous steps for all the hashtags of interest, we obtain a rich suite of distinct, temporally directed *tree graphs* that connects tweets containing the hashtags under analysis. It should be noted that, as tweets can have multiple hashtags, a tweet might be included in multiple tree graphs.

Finally, to infer the relationship between the hashtags, we construct a bipartite network composed of tweets and hashtags. We combine the nineteen tree graphs to get the *tweet layer* of the bipartite network, where every node represents a unique tweet and edges connect tweets as in the tree graphs. The *hashtag layer* comprises the examined nineteen hashtags. In the bipartite network, an edge exists between a tweet and a hashtag if the former embeds the latter. Thereby, the topology of the diverse tree graphs modulates the connections (and their weights) in the bipartite network. From this bipartite graph, we extract a hashtag network, named *HTEMap network*, based on the co-occurrence of the hashtags in the tweet layer, which encompasses both the temporal evolution and relationship between the fringe hashtags.

## Results

We first discuss results related to the inauthentic activities carried out by coordinated accounts to push conspiratorial content, and then we portray the broad picture of the relationship and temporal evolution of fringe hashtags generated by the HTEMap model.

### Coordinated Activity (RQ1)

To address RQ1, we analyze the techniques used to amplify fringe narratives and we characterize the tweets and users involved in coordinated inauthentic activities.

**Rapid Retweet Network.** We commence our analysis by exploring the *rapid retweet network*, which offers an *interactome* of the interactions between the users that coordinately amplified fringe content by means of rapid retweets. As a starting point, we consider a time window of 60 seconds to define a rapid retweet. Figure 4 portrays the force-directed layout of the rapid retweet network, where nodes represent users and edges indicate rapid retweets. The node size is proportional to the in-degree (i.e., the number of times a user was retweeted), whereas nodes are colored according to the annotations from Abilov et al. (2021). The thickness of the edges is proportional to the number of rapid retweets between nodes, while the color abides by the color of the source node (i.e., the author of the retweet). Figure 4 highlights that the rapid retweet network is composed of network structures exhibiting a star topology. In our scenario,

this means that a single user is retweeted by a multitude of other accounts. Lowering the size of the time window to 30 seconds, from the inset in Figure 4, we also notice how the skeletal structure of the rapid retweet network persists, with some of the most prominent accounts still at the center of their star topology and retweeted by a non-negligible number of followers, suggesting a robust, coordinated effort to target content generated by these accounts. Notably, star-like interaction structures were found in recent literature as evidence left behind by coordinated online manipulations (Nizzoli et al. 2020). By manually inspecting the users sitting in the center of such amplification bubbles, we identify some prominent and influential actors, including verified accounts with millions of followers. Among the others, we recognize various Republican politicians (e.g., Dr. Kelli Ward and Ed Martin), alt-right media outlets (e.g., RSB Network), and political commentators such as Tom Fitton – the president of Judicial Watch, a right-leaning activist group – and the former Fox News host, Lou Dobbs – also known for his controversial opinion about COVID-19 vaccines and his support to the voter fraud conspiracy theory (Battaglio 2021). In this regard, it is also worth noting that #dobbs is one of the most prevalent hashtags in our dataset, as shown in Figure 1.

Further, looking at the node color, based on the annotations from (Abilov et al. 2021), we find that 83% of the users involved in the rapid retweet network are labeled as promoters of the voter fraud conspiracy theory, which highlights the propensity of these accounts in trusting conspiratorial content and their alleged intent of resonating such narratives in a coordinated fashion. Besides characterizing the actors responsible for this targeted amplification, we are interested in uncovering the narratives pushed through their activity. The hashtag #stopthesteal (43.08%) results to be the most shared, followed by #dobbs (31.4%), #obamagate (14.12%), and #qanon (3.6%). Leveraging the annotations from Kennedy et al. (2022), we identified only a small set of misleading stories in the tweets boosted by this coordinated action, with the “Stop the Steal” rhetoric standing out with 15,285 tweets, followed by the *Dominion* (8,663 tweets), and *Hammer and Scorecard* (1,111 tweets) conspiracies.

**Copypasta Network.** To build the copypasta network, we first need to select an adequate threshold to identify similar tweets, which we also name *copypasta tweets*. To this aim, we observe the distribution of similarity scores in the analyzed tweets. From Figure 5(a), we note a bimodal distribution in the similarity scores of both fringe tweets (i.e., messages containing fringe hashtags) and a randomly sampled subset of messages (from our rehydrated dataset) that do not contain fringe hashtags. This surfaces a natural threshold = 0.7 for identifying copypasta tweets, which is also consistent with previous work (Pacheco, Flammini, and Menczer 2020; Topinka 2022). Interestingly, the distribution related to the sample of generic tweets shows less prevalence of copypasta tweets, indicating that this manipulation technique was mainly used to push fringe narratives.

In Figure 5(b), we further characterize the tweets containing fringe hashtags in contrast with generic tweets, also considering the copypasta activity. In particular, we disentangle tweets in different sharing activities (original tweets,

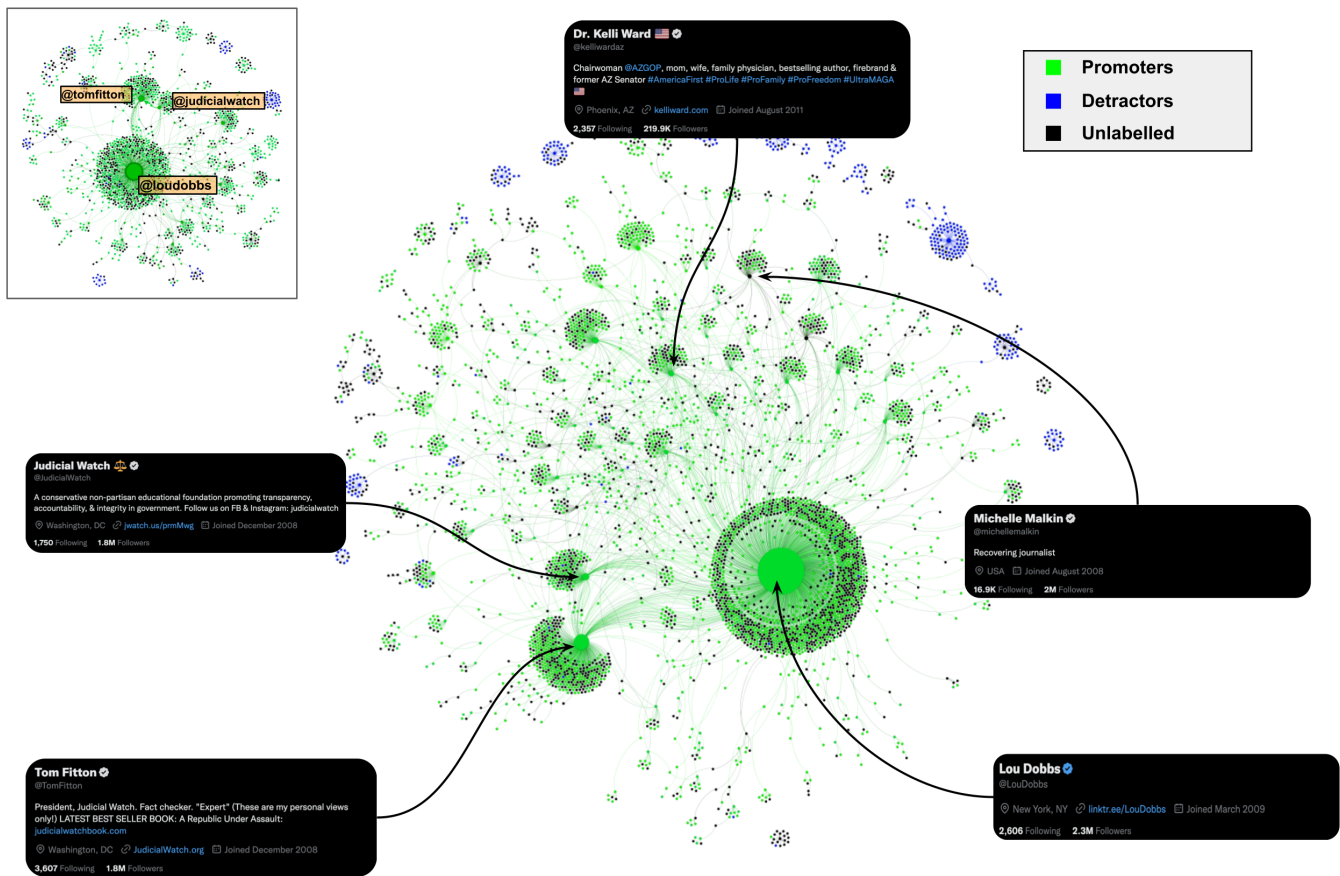


Figure 4: The rapid retweet network. Node size represents the in-degree, whereas the node color characterizes users as promoters or detractors leveraging annotations from (Abilov et al. 2021). The inset represents an instance of the rapid retweet network with a shorter time window (30 seconds).

replies, quotes) and measure their proportion for three different sets of tweets: generic tweets, fringe tweets, and fringe tweets with cypypasta activity (from now on, simply *cypypasta tweets*). We do not consider retweets in this evaluation to carry out a fair comparison with cypypasta tweets, which cannot include retweets by definition.

From Figure 5(b), we observe remarkably high proportions of quotes in the cypypasta activity, suggesting that this sharing activity might be used to diffuse problematic content, as further indicated by the suspiciously high volume of fringe tweets. Both fringe and cypypasta tweets include a significant fraction of replies (despite being less prominent with respect to generic tweets), which resembles the activity of misbehaving users during recent geo-political events (Pierri et al. 2023). Interestingly, the volume of original content shared in cypypasta tweets is narrow, which might suggest the potential employment of automated accounts to reproduce existing content and broadcast it via replies and quotes. This investigation is, however, out of the scope of this paper.

Based on these insights, we construct a cypypasta network incorporating tweets with similarity  $> 0.7$ . Figure 6(a)

shows the cypypasta network, where nodes represent tweets, which are connected if they have a similarity score  $> 0.7$ . We observe clusters of tweets independent of each other, representing chains of cypypasta activity. Each cluster represents a set of tweets sharing the same or very similar text. Nodes are colored based on the hashtag (or combination of hashtags) embedded in the tweet. We note a huge volume of cypypasta tweets embedding the #stopthesteal hashtag (76.25%), indicating how this hashtag was heavily amplified by means of different manipulation techniques.

To have a better understanding of the narratives amplified by this coordinated activity, we leverage the annotations curated by Kennedy et al. (2022). In Figure 6(b), we highlight the tweets that pushed claims about election delegitimization by coloring network nodes based on the specific theory shared. It can be noticed how a relevant fraction ( $\sim 60\%$ ) of clusters contains at least one colored node, which suggests a relevant prevalence of misleading claims shared in a coordinated way through cypypasta activity. Similarly to the results related to the rapid retweet network, “Stop the Steal” represents the narrative most amplified by coordinated accounts (83% of the tweets including misleading claims).

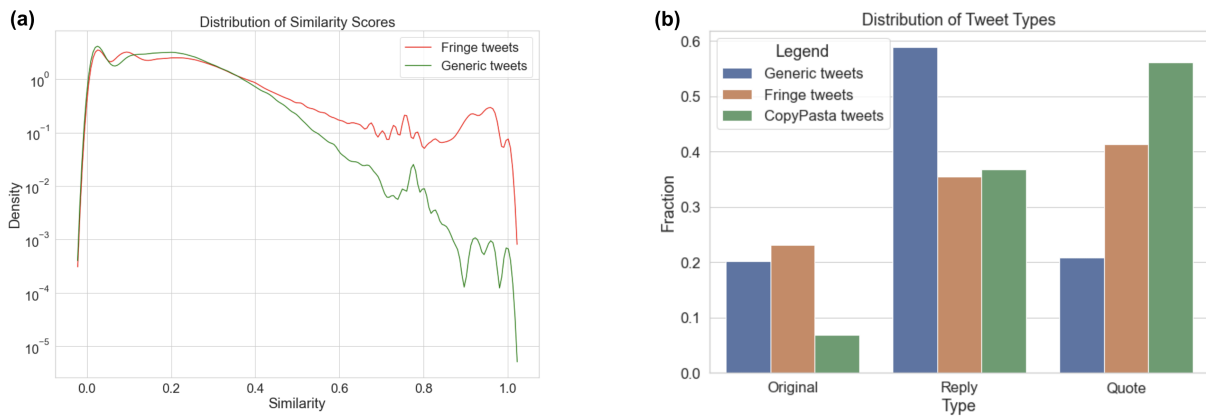


Figure 5: CopyPasta Activity: (a) Distributions of similarity scores of generic tweets and fringe tweets. (b) Distribution of sharing activities across generic, fringe, and copypasta tweets.

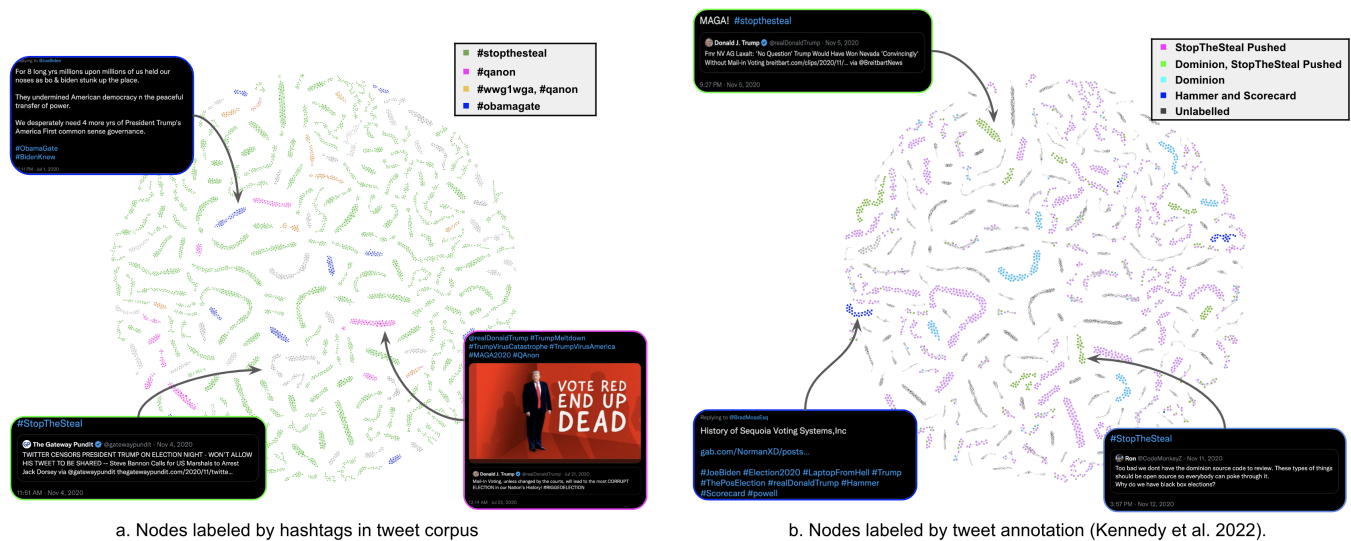


Figure 6: Copypasta network with tweets labeled by (a) hashtags, and (b) misleading claims about the election (Kennedy et al. 2022). A qualitative analysis of the shared content reveals that these tweets promote calls to action, support, and petitions.

By manually inspecting the largest clusters of tweets, we observe a variety of conspiratorial flavors in copy-pasta tweets, including calls to action, petitions, support via retweets, and trolling behavior via replies. A qualitative analysis was conducted by examining the text, embedded media objects, and URLs within the copypasta clusters. This analysis revealed that these tweets encourage various types of engagement, including calls to action such as marches, tweet-storms, and link-sharing. They also provide support through donations, in-person rallying, and organizing activities, as well as petitions involving legal challenges, fundraising, and engagement with elected Republican representatives. Figure 6 showcases several examples of these tweets, with usernames hidden to protect privacy.

Focusing on the actors responsible for this coordinated activity, we find that 95% of the users involved in the spread of copypasta tweets are labeled as promoters of the voter fraud

conspiracy theory (Abilov et al. 2021), once again highlighting their pivotal role in this deceptive effort. Interestingly, this set of users has a tiny overlap with the accounts involved in the rapid retweet network, with only 48 accounts engaged in both coordinated activities, indicating that diverse groups of users exhibit different inauthentic behaviors. This finding aligns with similar patterns discovered in recent work (Wang et al. 2023; Luceri et al. 2024) regarding the diverse roles and tactics of inauthentic actors in the artificial amplification of questionable content.

### Temporal Evolution (RQ2)

To answer RQ2, we leverage the HTEMap model for charting the landscape of the fringe hashtags, their relationship, and temporal evolution across the seven months preceding the Capitol attack. The HTEMap model returns a temporal network of hashtags, namely the HTEMap network, which

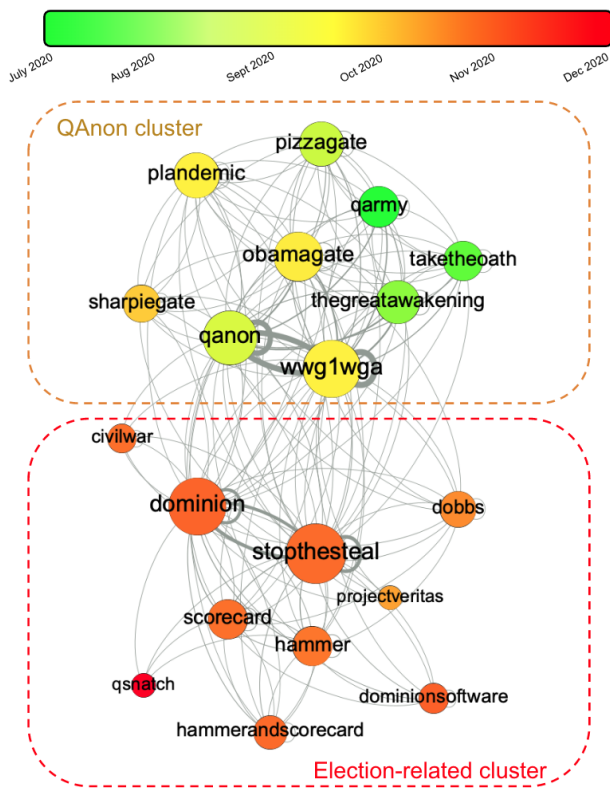


Figure 7: The HTEMap network.

is displayed in Figure 7 with a force-directed layout. The node size is proportional to the eigenvector centrality, and the edge weight represents the frequency of co-occurrence between the two hashtags in the HTEMap network. Similarly to what was done in recent longitudinal analyses of information operations (Kriel and Pavliuc 2019), the temporal dimension is represented by the node color. It should be noted that the HTEMap network encompasses by design the temporal aspect in the construction of the evolutionary trees. Here, for visualization purposes, we consider the median of the tweets' creation date for every set of chronologically-sorted tweets containing the hashtags under analysis.

From Figure 7, we observe two distinct communities, also confirmed through community extraction performed with the Louvain methodology (Blondel et al. 2008), suggesting that there are two distinct, yet interconnected narratives shared in the *Twittersphere* during our period of observation. Interestingly, each community almost entirely corresponds to one of the categories of hashtags detailed in Table 1, reflecting election-specific conspiracies and QAnon narratives. The #plandemic hashtag sits in the QAnon cluster, allegedly because of the connection with other QAnon conspiracies. We also recognize that the most central nodes (in terms of eigenvector centrality) are #qanon, #wwg1wga, #stopthesteal, and #dominion, which connect Election-specific theories with the QAnon conspiracy. It is also worth noting the role of #dobbs and #civilwar in bridg-

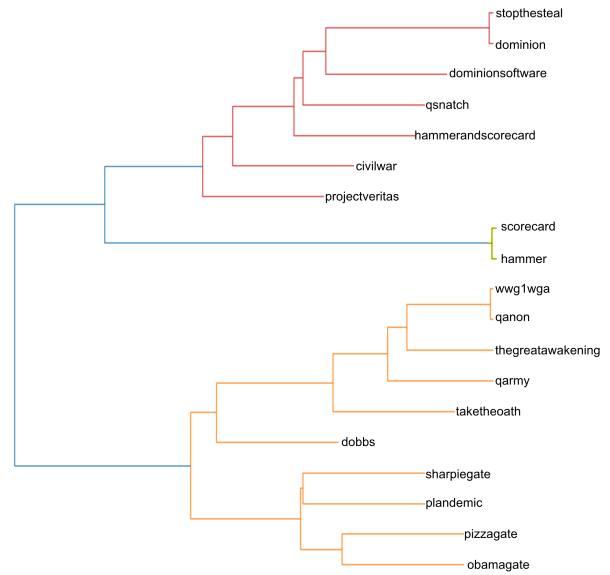


Figure 8: Evolutionary dendrogram of the fringe hashtags.

ing the two groups of conspiracies together. The former was amplified with the corresponding account (@LouDobbs) with a targeted amplification activity, as shown in the results related to the rapid retweet network. Instead, the latter was used by several influential spreaders to advertise the January attack and, as confirmed by our results, became viral on Twitter during the first days of November (Election Integrity Partnership 2020; Luceri, Cresci, and Giordano 2021). The interconnection of #civilwar with the QAnon theories further confirms the involvement of the conspiracy in the organized violence behind the Capitol attack.<sup>7</sup> Taking the temporal dimension into account, in Figure 7 we also observe that QAnon-specific hashtags were prevalent much earlier than the election-specific ones. In fact, almost all the election-specific fringe hashtags got traction around Election day (November 3rd, 2020), which is in line with previous research (Election Integrity Partnership 2020).

To deepen our understanding of the relationships between these hashtags, we perform a hierarchical clustering based on the adjacency matrix of the HTEMap network. This results in the dendrogram of fringe narratives shown in Figure 8. To interpret it, we focus on the width at which any two hashtags are branched together. We observe that the branch that joins #stopthesteal and #dominion is one of the smallest in the dendrogram, suggesting that they are among the most similar nodes in the HTEMap network. We recall that similarity in this context captures different dimensions embedded in the HTEMap network – that is, the temporal evolution and the topological co-occurrence of the hashtags.

From Figure 8, we observe the emergence of two distinct branches of conspiracy theories. This further confirms a separation between the co-occurrence of election misinformation and fringe conspiracy theories. In addition, it also sur-

<sup>7</sup><https://www.theguardian.com/us-news/2021/nov/17/qanon-shaman-jacob-chansley-sentenced-capitol-attack-role>

faces the existence of two unique, yet interconnected, narratives in the tweetspace. On the one hand, #stopthesteal and #dominion branch coalesce together all the election-specific conspiracies, including the 2017 QAnon theory called “Hammer and Scorecard”, which gained traction again in the weeks after the 2020 election (Election Integrity Partnership 2020). On the other hand, QAnon-specific hashtags pivoted around the hashtags #qanon and #wwglwga but also included the “gate” and Covid-19 conspiracies.

## Discussion

### Conclusions and Future Work

We carried out a longitudinal analysis of the interplay between conspiracy theories and fringe narratives in the run-up to and afterward of the US 2020 Presidential Election. By employing an innovative combination of methods, we surfaced and analyzed two widely used manipulation tactics: rapid retweets and cospypasta activity. Furthermore, we leveraged existing annotated resources to provide a characterization of the users involved in such manipulations, recognizing different groups of promoters of the voter fraud rhetoric acting in diverse deceptive efforts. Finally, we reconstructed the temporal trajectory of the fringe theories that eventually coalesced into the “Stop the Steal” narrative, pinpointing its interconnection with QAnon and the Capitol attack, also advertised with the hashtag #civilwar.

Among the key findings of our work is that (i) the rapid retweet networks artificially and coordinatedly amplified multiple fringe narratives (RQ1), (ii) the most artificially amplified users were right-wing commentators and politicians that endorsed the election fraud and QAnon conspiracies (RQ1), (iii) more than 76% of the cospypasta tweets and more than 96% of the users involved in cospypasta, promoted the “Stop the Steal” rhetoric (RQ2), and (iv) the QAnon conspiracies gained traction well before the election and quickly evolved into election fraud conspiracies after the polling (RQ2). Overall, our results provide insightful answers to our research questions about the adoption of manipulation techniques to amplify fringe narratives and about the interplay between the various fringe narratives and conspiracy theories. Furthermore, they shed new light on the complex socio-technical phenomena that led to one of the most destabilizing attacks on US democracy.

In the future, we aim to extend our comprehensive analyses by also considering additional platforms and other forms of online manipulation. To this end, the possibility of complementing the temporal aspects of this work with multi-platform analyses represents a promising direction of research that holds the potential for shedding further light on the tactics used to organize the Capitol attack. Other avenues of research involve uncovering more complex coordination tactics and considering a longer time window to study how the conversations fully converge after the Capitol attack.

### Limitations

Our work presents a few limitations. First, the rehydration allowed us to retrieve only 51.72% of the complete dataset collected by Chen, Deb, and Ferrara (2022). This is mainly

due to the content moderation and account banning performed by Twitter during the Election period and, in particular, after the moderation interventions against QAnon-specific content (Sharma, Ferrara, and Liu 2022). For this reason, our findings represent a lower-bound estimate of coordinated inauthentic activity performed during the observation period. Nonetheless, despite these limitations, our analysis uncovered evidence of multiple manipulation techniques ranging from targeted amplification to manufactured consensus, which likely went undetected by Twitter.

Second, our analyses focus on common manipulation tactics, but novel and more sophisticated strategies might have been employed (Luceri et al. 2024; Cinelli et al. 2022; Tardelli et al. 2023). Similarly, we do not consider the activity of software-controlled accounts (i.e., bots) and state-sponsored human operators (Luceri, Giordano, and Ferrara 2020; Ezzeddine et al. 2023), who might have played a role in the amplification of fringe content (Ferrara et al. 2020). Finally, a considerable amount of fringe actors have migrated from Twitter to low-moderated spaces, such as alt-tech platforms like Gab and Parler, during the observation period, thus potentially affecting our results that are solely based on Twitter. This calls for broader research encompassing more diverse online platforms in the information ecosystem, which will be the objective of our next endeavors.

### Broader Perspective, Ethics, and Competing Interests

We believe our work will have a positive broader impact on the understanding of the complex interplay between conspiracy theories and fringe narratives. These online phenomena showed potential for developing into serious offline issues, as in the case of the Capitol attack. To this end, our work can inform future efforts to curb these and other issues. At the same time, however, among the conceivable ethical risks of our work is the possibility that some of our findings could be used for political reasons. Additional risks may derive from the unintended disclosure of sensitive user information contained in the datasets used in this work since such information could be exploited to carry out targeted attacks. To mitigate these risks, we relied on publicly available and pseudonymized data, and we carried out analyses at an aggregated level. Moreover, the observational and retrospective nature of our analyses implies that no users were harmed as part of this research. The authors declare no competing interests. In view of scientific transparency, we make our code available on GitHub<sup>8</sup> to allow reproducibility and replicability of our work.

### Acknowledgements

This project was partially funded by the Emergent Ventures India Grant (EV #1802) within the project Search 4 Computational Hate (S4CH). The authors gratefully acknowledge support by the Swiss National Science Foundation through the projects (CRSII5\_209250) *Call for Regulation support In Social Media* (CARISMA) and *Detecting Troll Activity in Online Social Networks* (CRSK-2\_195707).

<sup>8</sup><https://github.com/psvp28/s4ch>

## References

- Abilov, A.; Hua, Y.; Matatov, H.; Amir, O.; and Naaman, M. 2021. VoterFraud2020: A multi-modal dataset of election fraud claims on Twitter. In *AAAI ICWSM*.
- Battaglio, S. 2021. Fox News cancels Lou Dobbs' show; pro-trump host not expected to be back on air.
- Benkler, Y.; Tilton, C.; Etling, B.; Roberts, H.; Clark, J.; Faris, R.; Kaiser, J.; and Schmitt, C. 2020. Mail-in voter fraud: Anatomy of a disinformation campaign. Technical report, Harvard University, Berkman Klein Center.
- Bessi, A.; and Ferrara, E. 2016. Social bots distort the 2016 US Presidential election online discussion. *First monday*, 21(11-7).
- Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*.
- Cao, C.; Caverlee, J.; Lee, K.; Ge, H.; and Chung, J. 2015. Organic or organized? Exploring URL sharing behavior. In *ACM CIKM*.
- Casas, A.; and Williams, N. W. 2019. Images that matter: Online protests and the mobilizing role of pictures. *Political Research Quarterly*, 72(2): 360–375.
- Chen, E.; Chang, H.; Rao, A.; Lerman, K.; Cowan, G.; and Ferrara, E. 2021. COVID-19 misinformation and the 2020 US presidential election. *HKS Misinformation Review*.
- Chen, E.; Deb, A.; and Ferrara, E. 2022. #Election2020: The first public Twitter dataset on the 2020 US presidential election. *Journal of Computational Social Science*, 5(1).
- Cinelli, M.; Cresci, S.; Quattrociocchi, W.; Tesconi, M.; and Zola, P. 2022. Coordinated inauthentic behavior and information spreading on Twitter. *Decision Support Systems*.
- Election Integrity Partnership. 2020. The Long Fuse: Misinformation and the 2020 Election.
- Ezzeddine, F.; Ayoub, O.; Giordano, S.; Nogara, G.; Sbeity, I.; Ferrara, E.; and Luceri, L. 2023. Exposing influence campaigns in the age of LLMs: a behavioral-based AI approach to detecting state-sponsored trolls. *EPJ Data Science*, 12(1): 46.
- Ferrara, E.; Chang, H.; Chen, E.; Muric, G.; and Patel, J. 2020. Characterizing social media manipulation in the 2020 U.S. presidential election. *First Monday*.
- FORCE11. 2020. The FAIR Data principles. <https://force11.org/info/the-fair-data-principles/>.
- Geburu, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Green, J.; Hobbs, W.; McCabe, S.; and Lazer, D. 2022. Online engagement with 2020 election misinformation and turnout in the 2021 Georgia runoff election. *PNAS*, 119(34).
- Horta Ribeiro, M.; Jhaver, S.; Zannettou, S.; Blackburn, J.; Stringhini, G.; De Cristofaro, E.; and West, R. 2021. Do platform migrations compromise content moderation? Evidence from r/The\_Donald and r/Incels. In *ACM CSCW*.
- Hristakieva, K.; Cresci, S.; Da San Martino, G.; Conti, M.; and Nakov, P. 2022. The spread of propaganda by coordinated communities on social media. In *ACM WebSci*.
- Jachim, P.; Sharevski, F.; and Pieroni, E. 2021. Troll-Hunter2020: Real-time detection of trolling narratives on Twitter during the 2020 US elections. In *ACM IWSPA*.
- Jacobson, L. 2021. PolitiFact - Is this a coup? Here's some history and context to help you decide.
- Kennedy, I.; Wack, M.; Beers, A.; Schafer, J. S.; Garcia-Camargo, I.; Spiro, E. S.; and Starbird, K. 2022. Repeat spreaders and election delegitimization: A comprehensive dataset of misinformation tweets from the 2020 US election. *Journal of Quantitative Description: Digital Media*, 2.
- KhudaBukhsh, A. R.; Sarkar, R.; Kamlet, M. S.; and Mitchell, T. M. 2022. Fringe news networks: Dynamics of US news viewership following the 2020 presidential election. In *ACM WebSci*.
- Kriel, C.; and Pavliuc, A. 2019. Reverse engineering Russian Internet Research Agency tactics through network analysis. *Defence Strategic Communication*, 6.
- La Gatta, V.; Luceri, L.; Fabbri, F.; and Ferrara, E. 2023. The interconnected nature of online harm and moderation: investigating the cross-platform spread of harmful content between youtube and Twitter. In *Proceedings of the 34th ACM conference on hypertext and social media*, 1–10.
- Luceri, L.; Cardoso, F.; and Giordano, S. 2021. Down the bot hole: Actionable insights from a one-year analysis of bot activity on Twitter. *First Monday*.
- Luceri, L.; Cresci, S.; and Giordano, S. 2021. Social Media against Society. *The Internet and the 2020 Campaign*.
- Luceri, L.; Deb, A.; Giordano, S.; and Ferrara, E. 2019. Evolution of bot and human behavior during elections. *First Monday*.
- Luceri, L.; Giordano, S.; and Ferrara, E. 2020. Detecting troll behavior via inverse reinforcement learning: A case study of russian trolls in the 2016 us election. In *AAAI ICWSM*.
- Luceri, L.; Pantè, V.; Burghardt, K.; and Ferrara, E. 2024. Unmasking the web of deceit: Uncovering coordinated activity to expose information operations on twitter. In *Proceedings of the ACM Web Conference 2024*.
- Magelinski, T.; Ng, L.; and Carley, K. 2022. A synchronized action framework for detection of coordination on social media. *Journal of Online Trust and Safety*, 1(2).
- Matatov, H.; Naaman, M.; and Amir, O. 2022. Stop the [image] Steal: The role and dynamics of visual content in the 2020 US election misinformation campaign. In *ACM CSCW*.
- Mazza, M.; Cresci, S.; Avvenuti, M.; Quattrociocchi, W.; and Tesconi, M. 2019. RTbust: Exploiting temporal patterns for botnet detection on Twitter. In *ACM WebSci*.
- Metaxas, P.; Mustafaraj, E.; Wong, K.; Zeng, L.; O'Keefe, M.; and Finn, S. 2015. What do retweets indicate? Results from user survey and meta-review of research. In *AAAI ICWSM*.

Nizzoli, L.; Tardelli, S.; Avvenuti, M.; Cresci, S.; and Tesconi, M. 2021. Coordinated behavior on social media in 2019 UK General Election. In *AAAI ICWSM*.

Nizzoli, L.; Tardelli, S.; Avvenuti, M.; Cresci, S.; Tesconi, M.; and Ferrara, E. 2020. Charting the landscape of online cryptocurrency manipulation. *IEEE Access*, 8.

Pacheco, D.; Flammini, A.; and Menczer, F. 2020. Unveiling Coordinated Groups Behind White Helmets Disinformation. *ACM WWW Companion*.

Pacheco, D.; Hui, P.-M.; Torres-Lugo, C.; Truong, B. T.; Flammini, A.; and Menczer, F. 2021. Uncovering coordinated networks on social media: Methods and case studies. In *AAAI ICWSM*.

Papasavva, A.; Aliapoulos, M.; Ballard, C.; De Cristofaro, E.; Stringhini, G.; Zannettou, S.; and Blackburn, J. 2022. The gospel according to Q: Understanding the QAnon conspiracy from the perspective of canonical information. In *AAAI ICWSM*.

Papasavva, A.; Blackburn, J.; Stringhini, G.; Zannettou, S.; and Cristofaro, E. D. 2021. "Is it a Qoincidence?": An exploratory study of QAnon on Voat. In *The Web Conference*. Pennycook, G.; and Rand, D. G. 2021. Examining false beliefs about voter fraud in the wake of the 2020 presidential election. *HKS Misinformation Review*, 2.

Pierrri, F.; Luceri, L.; Chen, E.; and Ferrara, E. 2023. How does Twitter account moderation work? Dynamics of account creation and suspension during major geopolitical events. *EPJ Data Science*.

Ratclif, J. W. 1988. Pattern Matching: the Gestalt Approach.

Sato, H.; Hashimoto, Y.; Oka, M.; and Ikegami, T. 2021. Exploration in Evolutionary Space by Hashtag Evolution on a Social Network Service. In *ALIFE*.

Sharma, K.; Ferrara, E.; and Liu, Y. 2022. Characterizing online engagement with disinformation and conspiracies in the 2020 US presidential election. In *AAAI ICWSM*.

Steinert-Threlkeld, Z. C.; Mocanu, D.; Vespignani, A.; and Fowler, J. 2015. Online social networks and offline protest. *EPJ Data Science*, 4(1): 1–9.

Tardelli, S.; Nizzoli, L.; Tesconi, M.; Conti, M.; Nakov, P.; Da San Martino, G.; and Cresci, S. 2023. Temporal dynamics of coordinated online behavior: Stability, Archetypes, and Influence. *arXiv:2301.06774*.

Topinka, R. 2022. The politics of anti-discourse: Copypasta, the Alt-Right, and the rhetoric of form. *Theory & Event*, 25(2): 392–418.

Trujillo, A.; and Cresci, S. 2022. Make Reddit Great Again: Assessing community effects of moderation interventions on r/The\_Donald. In *ACM CSCW*.

Vargas, L.; Emami, P.; and Traynor, P. 2020. On the detection of disinformation campaign activity with network analysis. In *ACM CCSW*.

Varol, O.; Ferrara, E.; Ogan, C. L.; Menczer, F.; and Flammini, A. 2014. Evolution of online user behavior during a social upheaval. In *ACM WebSci*.

Wang, E. L.; Luceri, L.; Pierrri, F.; and Ferrara, E. 2023. Identifying and characterizing behavioral classes of radicalization within the qanon conspiracy on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 890–901.

Weber, D.; and Neumann, F. 2021. Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining*, 11(1): 1–42.

## Paper Checklist

### 1. For most authors...

- (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes, refer to Introduction for context and Broader Perspective, Ethics, and Competing Interests for our impacts**
- (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? **Yes**
- (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes, and this is clarified in the Methodology Section**
- (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes, and this is clarified in the Methodology Section**
- (e) Did you describe the limitations of your work? **Yes, in the Limitations Section**
- (f) Did you discuss any potential negative societal impacts of your work? **Yes, refer to Broader Perspective, Ethics, and Competing Interests for our impacts**
- (g) Did you discuss any potential misuse of your work? **Yes, refer to Broader Perspective, Ethics, and Competing Interests**
- (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **Yes, refer to Broader Perspective, Ethics, and Competing Interests where we discuss ethics. We also open-source our work keeping sensitive data contained.**
- (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes, and our work conforms to the guidelines**

### 2. Additionally, if your study involves hypotheses testing...

- (a) Did you clearly state the assumptions underlying all theoretical results? **NA**
- (b) Have you provided justifications for all theoretical results? **NA**
- (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **NA**

- (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
  - (e) Did you address potential biases or limitations in your theoretical framework? NA
  - (f) Have you related your theoretical results to the existing literature in social science? NA
  - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
- (a) Did you state the full set of assumptions of all theoretical results? NA
  - (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes, our code is available on GitHub](#)
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
  - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
  - (f) Do you discuss what is “the cost“ of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
- (a) If your work uses existing assets, did you cite the creators? [Yes, and all works are cited in the paper.](#)
  - (b) Did you mention the license of the assets? [Yes, and in relevant sections, we have attributed the Creative Commons License](#)
  - (c) Did you include any new assets in the supplemental material or as a URL? [Yes, our methods are reproducible and can be accessed on GitHub.](#)
  - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? [Yes, we follow the guidelines from FAIR, Twitter Inc., and Creative Commons wherever relevant.](#)
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes, refer to Broader Perspective, Ethics, and Competing Interests. Personally identifiable information hidden for users except prominent political figures with an already existing digital footprint.](#)
- (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA
  - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects...
- (a) Did you include the full text of instructions given to participants and screenshots? NA
  - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
  - (d) Did you discuss how data is stored, shared, and de-identified? NA

## Appendix: Narratives and Hashtags

**QAnon.** QAnon conspiracy theories involve a fringe movement that believes in an anonymous government insider and alleges a conspiracy involving Satanic child sex trafficking targeting Donald Trump. This conspiracy theory gained significant attention beyond online communities and managed to connect with major political figures. The movement’s most famous slogan, “Where we go one, we go all,” gave rise to the hashtag [#wwg1wga](#). Other well-known hashtags associated with QAnon include [#pizzagate](#), which relates to the unfounded claims of child abuse by prominent Democrats, [#taketheoath](#), which encouraged QAnon followers to pledge their allegiance as “digital soldiers” endorsed by influential members, [#projectveritas](#), referring to a far-right activist group known for engaging in disinformation campaigns, [#thegreatawakening](#), drawing parallels between the supposed battle between Trump and a cabal of satanic cannibalistic pedophiles to the concept of a biblical Great Awakening, and [#civilwar](#), which calls for a second American civil war and is often associated with white supremacist movements. Finally, [#obamagate](#) emerged as a claim made by Trump, alluding to the Watergate scandal, where he alleged that former President Barack Obama wiretapped him.

**US Election.** Narratives surrounding a “stolen” election began circulating even prior to the actual vote. One notable example is the emergence of the [#sharpiegate](#) narrative, which quickly spread across Twitter and various online communities. This narrative alleged that a strategy was employed in conservative precincts, where felt-tip pens (sharpies) were used to render ballots unreadable. Despite repeated debunking of the [#sharpiegate](#) claims, the rumors gained significant traction online and even prompted Trump supporters to protest outside the Maricopa County Elections Department. The protesters chanted the slogan “Stop the steal!” during the demonstrations, which echoed through various platforms with the hashtag [#stopthesteal](#). This hashtag eventually gave rise to a movement centered around the

concept of election theft. Alongside #stopthesteal, several other conspiracy theories emerged, each with their corresponding hashtags. The hashtags **#hammer** and **#scorecard** propagated claims that a software called Hammer was capable of breaching secure networks, while Scorecard allegedly manipulated the overall vote count against Trump. Similarly, the hashtag **#hammerandscorecard** combined these two theories. The hashtag **#dominion** became associated with the belief that Dominion Voting Systems, a manufacturer of voting machines, had “deleted” millions of votes cast in favor of Trump. Lastly, the hoax hashtag **#qsnatch** circulated claims that malware was actively “stealing the election” across multiple states.

**COVID-19.** The hashtag **#plandemic** is associated with a COVID-19 conspiracy theory that claims the global pandemic is a deliberately orchestrated hoax by the pharmaceutical industry and philanthropist Bill Gates. This hashtag has served as a vehicle for various anti-vaccine conspiracy theories as well. It is crucial to note that the dissemination of conspiracy theories related to the pandemic began with changes made to the electoral process in response to the COVID-19 outbreak, aiming to mitigate transmission risks.