Khaouja, Imane, Toribio-Flórez, Daniel, Green, Ricky, Rowden, Cassidy, Ang, Chee Siang and Douglas, Karen M. (2025) ***Political Communication and Conspiracy Theory Sharing on Twitter.*** **Online Social Networks and Media, 47 . ISSN 2468-6964.**

Political Communication and Conspiracy Theory Sharing on Twitter

1. Abstract

Social media has become an influential channel for political communication, offering broad reach while enabling the proliferation of misinformation and conspiracy theories. These unchecked conspiracy narratives may allow manipulation by malign actors, posing dangers to democratic processes. Despite their intuitive appeal, little research has examined the strategic usage and timing of conspiracy theories in politicians' social media communication compared to the spread of misinformation and fake news.

This study provides an empirical analysis of how members of the U.S. Congress spread conspiracy theories on Twitter. Leveraging the Twitter Historical API, we collected a corpus of tweets from members of the US Congress between January 2012 and December 2022. We developed a classifier to identify conspiracy theory content within this political discourse. We also analyzed the linguistic characteristics, topics and distribution of conspiracy tweets. To assess classifier performance, we created ground truth data through human annotation in which experts labeled a sample of 2500 politicians' tweets.

Our findings shed light on several aspects, including the influence of prevailing political power dynamics on the propagation of conspiracy theories and higher user engagement. Moreover, we identified specific psycho-linguistic attributes within the tweets, characterized by the use of words related to power and causation, and outgroup language. Our results provide valuable insights into the motivations compelling influential figures to engage in the dissemination of conspiracy narratives in political discourse.

Keywords: Conspiracy Theory, Political Communication, Twitter, NLP, Psycho-linguistic characteristics

2. Introduction

Conspiracy theories (CTs) explain significant social or political events as secret plots by malicious and typically powerful actors [Douglas and Sutton (2023); Phillips et al. (2022); Douglas et al. (2019); Enders et al. (2023)]. They are especially appealing to individuals in times of uncertainty and crisis [van Prooijen and Douglas

(2017)]. Furthermore, they provide pathways through which malevolent individuals can exert influence over or manipulate susceptible communities, potentially leading to significant repercussions for social and political dynamics [Douglas et al. (2019)]. CTs have been of interest to academic researchers since the 1960s when the assassination of US President John F. Kennedy triggered many such theories. Unlike other disinformation, CTs offer alternative explanations in opposition to conventional and well-supported explanations of events [Douglas et al. (2019)]. For example, well-known CTs posit that the NASA moon landing was faked, the Earth is flat, climate change is a hoax, and vaccines can lead to autism. Such theories ignore the overwhelming scientific evidence in these fields and instead explain events as the result of secret acts by powerful forces. The rise of social media has made it easier for people to spread CTs online [Stano (2020)]. However, our understanding of the consequences and impact of such theories, particularly for those who disseminate them, remains limited.

In this paper, we empirically investigate how CTs are shared on Twitter (currently rebranded as X) by politicians. Previous research has primarily focused on the analysis of misinformation and fake news in political discourse on one hand and the detection and analysis of conspiracist content in single-topic tweets and tweets collected through specific hashtags on the other. Notably, the main difference between fake news and CTs is that fake news is a specific type of misinformation, while conspiracy theories are a broader belief system that can be associated with various misinformation types, including fake news [CarrascoFarré (2022)]. In particular, fake news refers to news articles that are intentionally and verifiably false, and could mislead readers. However CTs aim to explain powerful people's secret plans or schemes. While fake news can be fact checked due to its falsifiable nature, CTs are often non-factual and cannot be easily verified [Aïmeur et al. (2023)].

This study empirically analyses U.S. representatives' political discourse, focusing on their use of CTs. We aimed to discern a) how politicians share such conspiracist narratives on social media (i.e., which linguistic features they use), and b) how users engage with this type of communication, compared to non-conspiracy communication. Our research was further directed toward exploring whether the use of CTs on Twitter was influenced by electoral dynamics and polarization. Unlike prior work, we provide an in-depth examination of CT propagation by influential political actors across topics and over an extensive time period.

In particular, our aim was to answer the following research questions and to test the following hypotheses:

- RQ1: Does the use of CTs in politicians' social media communication influence engagement?
- RQ2: Which linguistic features characterize CTs in politicians' social media communication?
- RQ3: Do political power dynamics influence the use of CTs in politicians' social media communication?
- H1: CT tweets elicit higher engagement (e.g., likes, shares) than non-CT tweets.
- H2: Linguistic features (e.g., causal language, negative emotions) are more prevalent in CT tweets than in non-CT tweets.
- H3: Politicians with lower political power use CTs more frequently in their social media communication than those with higher political power.

All research questions and hypotheses are addressed in Section 5. Specifically, RQ1 and H1 are addressed in Section 5.3, and RQ2, RQ3, H2, and H3 are addressed in Section 5.4.2.

We analyzed tweets posted by members of the 117th and 118th U.S. Congresses in online Twitter accounts from January 2012 to December 2022. Subsequently, we employed a ground truth sample of 2500 tweets combined with existing datasets in the literature to fine-tune a conspiracy classifier, which was then used to quantify the prevalence of conspiracy tweets. We analyzed the interactions, distribution of tweets, linguistic features, and topics of such tweets.

3. Related Work

Social media has become a crucial communication tool for U.S. politicians, with all members of Congress maintaining a professional presence on Twitter. The integration of social media into U.S. politics offers several advantages [Hong et al. (2019)]. Primarily, it facilitates direct, unfiltered communication with the public, bypassing traditional media moderation [Allcott and Gentzkow (2017)]. Additionally, it promotes a two-way dialogue between politicians and constituents regarding policy issues [Enli and Skogerbø (2013)], with engaged followers amplifying messages through shares and likes. Furthermore, social media supports political mobilization around events and causes [Theocharis et al. (2015); Yasseri et al. (2016)].

Extensive research has investigated various dimensions of political communication on Twitter. For example, the study by [Solovev and Prollochs (2022)] analyzed hate speech in political discourse and the disparities between political parties. In [Lasser et al. (2023)], the honesty component in political speech, categorized as belief speaking and fact speaking, was measured by examining keywords related to honesty in U.S. politicians' tweets.

Recent studies have also shed light on the spread of misinformation and fake news on Twitter, with a particular focus on the role of political elites. A study by [Mosleh and Rand (2022)] introduced a tool for assessing users' exposure to misinformation from political elites on Twitter. Utilizing a database of professional fact-checks, they calculated falsity scores for 816 elites based on the veracity of their statements, finding a negative correlation between users' misinformation exposure scores and the quality of news they shared. Lasser et al. (2022) emphasized the social media dissemination of low-quality news by political elites. This study, focusing on the U.S. Congress, revealed an increasing trend among Republicans in sharing links to unreliable sources from 2016 onwards, diverging from the behavior of Democrats. This divergence intensified after President Biden's election and was distinct to the United States, differing from trends observed in other Western democracies.

Another study by Tai et al. (2023) explored the role of elected officials in spreading misinformation on Twitter. By analyzing tweets from over 3,000 U.S. state lawmakers in 2020 and 2021, the study uncovered significant partisan differences, with Republicans sharing substantially more misinformation than Democrats. The findings also highlighted temporal trends in discussions on various topics and underscored the asymmetric polarization in partisan attacks.

While these existing studies enhance our understanding of misinformation dynamics and the role of political elites in political communication and public discourse on Twitter, other studies indicate that individuals who disseminate CTs may become more radicalized, engage in emotionally charged discussions, and contribute to the broader spread of misinformation, which can have significant social and psychological implications. Similarly, Langguth et al. (2023) presented the COVID-19 Conspiracy (COCO) dataset, which consists of 3,495 annotated tweets related to 12 different COVID-19 conspiracy topics. Each tweet is manually labeled to indicate its stance concerning specific CTs, offering a valuable resource for studying the spread and characteristics of conspiracy narratives on Twitter.

Furthermore, recent studies have further explored the dissemination of CTs and misinformation on social media platforms. For instance, Korenčić et al. (2024) conducted a computational analysis to differentiate between conspiracy and critical narratives in online discourse. Their research introduced a topic-agnostic annotation scheme and developed the XAI-DisInfodemics corpus, which includes annotated Telegram messages related to COVID-19 in both English and Spanish. The study's findings highlight the role of intergroup conflict, as well as the presence of violence and anger, as key indicators distinguishing conspiracist content from critical discussions.

Additionally, the spread of misinformation and CTs has been linked to the psychological and linguistic characteristics of online discourse. Korenčić et al. (2024) emphasized the importance of identifying intergroup conflict and emotional cues, such as anger and violence, in distinguishing between conspiracy and critical narratives (e.g., questioning vaccine safety without suggesting a sinister, hidden agenda). Their computational analysis of oppositional discourse on platforms like Telegram provides insights into the linguistic markers that differentiate conspiracist content from legitimate criticism.

Recent studies have further explored the dissemination of CTs and misinformation on social media platforms. For instance, Phadke et al. Phadke et al. (2022) analyzed Reddit users participating in conspiracy discussions, identifying distinct engagement pathways: steady high, increasing, decreasing, and steady low. Users with increasing or consistently high engagement exhibited signs of radicalization, such as adopting insider language and participating in insular group discussions. Conversely, those with decreasing engagement interacted in more diverse groups and showed less conformity to conspiracy communities.

Zhang et al. Zhang et al. (2021) conducted a large-scale analysis of online discussion cascades, and the spread of conspiracy and scientific information on social media. Their findings indicated that conspiracy information tends to trigger larger cascades, involves more users and generations, persists longer, and is more viral and bursty than scientific information. Additionally, conspiracy cascades contain more negative and emotional words, conveying anger, fear, disgust, surprise, and trust.

Moreover, several datasets have been utilized in research on CT detection, as illustrated in Table 1. For instance, Phillips et al. (2022) compiled a dataset comprising four types of CTs: climate change, COVID-19 origins, COVID-19 vaccine, and CTs about Jeffrey Epstein and Ghislaine Maxwell, achieving an F1 score of 0.81. Moffitt et al. (2021) examined COVID-19-related CTs by training a BERT-based classifier to identify conspiracy content in tweets, achieving an F1 score of 0.95. Additionally, Liaw et al. (2023) collected a dataset from YouTube with extensive metadata from videos on suspicious channels identified as containing CTs, achieving a precision of 78% and a recall of 86%. Similarly, Faddoul et al. (2020) gathered a list of conspiracy videos using seed channels and enriched the datasets through YouTube's recommendation algorithm.

Although these studies demonstrate the efficacy of transformer-based models in detecting conspiracy tweets, the datasets collected with hashtags known to be associated with potential CTs present challenges for generalization to other datasets.

Some studies have examined the psycho-linguistic characteristics of CTs on social media by collecting tweets from predefined lists of accounts known to actively disseminate such theories. For instance, Klein et al. (2019) and Samory and Mitra (2018) analyzed discussions on the Reddit CT forum, finding that conspiracy discussions emphasized themes of power, terrorism, deception, and government compared to control discussions. On Twitter, pro- an anti-vaccine users exhibited linguistic markers indicative of conspiracy thinking, particularly around government mistrust [Mitra et al. (2016)]. In a comparative study, Fong et al. (2021) examined prominent conspiracy theorists on Twitter alongside science advocates, identifying unique linguistic features. They found that the language used by conspiracy influencers and their followers predominantly featured negative emotions like anger, along with a higher frequency of expressions related to power, death, and religion compared to those focused on science.

Previous studies have typically collected conspiracy tweets by specifying hashtags, lists of CT influencers, or threads related to CTs. However, to the best of our knowledge, no prior work has examined the prevalence and extent of CTs in politicians' tweets.

In this paper, we aim to provide a comprehensive analysis of how U.S. politicians leverage CTs on Twitter. Unlike prior research that focused on collecting CT content through hashtags or predefined lists of influential accounts, we analyze actual tweets from elected members of Congress spanning from 2012 to 2022.

Dataset	Source	Topics	Original Label	Annotation method	Interannotator agreement	Reported F1-score
Hoaxes Dataset, Phillips et al. (2022)	Twitter	- Climate change -Epstein -Covid origins -Covid vaccine	3,100 annotated instances. with 2,336 CTs	Agreement between 3 labelers	0.478	81,3%
Younicon Dataset, Liaw et al. (2023)	Youtube	-Ethnicity, Race, and Religion -Government, Politics, and Conflict, -Science and technology, -Extraterrestrials and UFOs.	3,161 videos with 1,144 CTs	Majority vote between 3 annotators	0.4111	86,2%
Youtube Dataset, Faddoul et al. (2020)	Youtube	-	1110 annotated instances with 542 CTs	Top CTs from YouTube and Reddit	-	81,8%
Covid Dataset, Moffitt et al. (2021)	Twitter	Covid19	4,573 annotated Tweets with 924 CTs	Annotation class determined by 1 single annotator.	-	-

Table 1: Characteristics of the different human-annotated conspiracy theory datasets in the literature. '–' indicates information not reported.

This approach allows us to empirically quantify the prevalence of conspiracy narratives propagated by these influential political elites across a broad spectrum of topics over a prolonged period. Additionally, we explore how the dissemination of such content is influenced by political party affiliation, electoral cycles, and polarization within the U.S. political landscape. By integrating computational methods to detect CTs at scale with an in-depth analysis of linguistic attributes and user interaction behaviors, our work illuminates how CTs infiltrate mainstream political discourse on social media platforms. Our findings aim to inform efforts to combat the spread of misinformation and CTs propagated by influential actors.

4. Data

4.1. Data for Analysis

We analyzed tweets from members of the 117th and 118th U.S. Congresses, which convened on January 3, 2021, and January 3, 2023, respectively. Data on the members of the 118th Congress was collected from the official U.S. Congress webpage [1] and from the UC San Diego Library[2], which provides links to the senators' Twitter handles. For data collection, we gathered their tweets, along with the number of likes and retweets, from January 2012 to December 2022, using the Twitter Historical API.

For the 117th Congress, we utilized data gathered in a previous study [Lasser et al. (2023)], which comprehensively collected tweets of congressional members using the Twitter API within the same timeframe.

Out of the 605 accounts identified, 209 were inaccessible due to deletion, suspension, or being set to private. For simplicity and interpretability, our empirical analysis later focuses on tweets from Republican and Democratic members, excluding tweets from independent members. Our final sample comprised tweets from 107 Democrat members and 283 Republican members. It is important to note that this approach includes all tweets posted by a given account within the specified timeframe, not just those posted while a politician was in office. This decision was guided by the average turnover rate, approximately 12 years [Desilver (2022)], which underscores the extended duration of congressional member transitions.

We excluded retweets from our collection, resulting in a total of 1,013,152 original tweets, with a median of 2,349 tweets per account. Our dataset comprises 1,013,152 tweets from 394 users (107 Democrats, 283 Republicans). The party-level analysis shows that Republicans contributed 571,541 tweets and Democrats contributed 368,388 tweets. At the user level (Table 8), Congress members posted an average of 2,584 tweets (SD = 2,166), with activity ranging from 15 to 21,613 tweets (Table 6). We identified super-posters in Table 9 (top 5%, n=22) who posted between 3,114 and 21,613 tweets (M = 8,825, SD = 4,378).

Our dataset encompasses all tweets from these politicians, not just those suspected of containing conspiracy content. This comprehensive approach allows us to study CT dissemination within the broader context of political discourse. For comparative analysis and model training, we also incorporated existing CT datasets from the literature: Covid dataset, CMU, Youtube2, and Younicon (detailed in Table 1). These datasets were selected based on their annotation quality (requiring multiple annotators), topical coverage (spanning political and social issues), and data availability. We found that combining these diverse datasets improved our model's

[1] https://pressgallery.house.gov/member-data/members-official-twitter-handles

[2] https://ucsd.libguides.com/congress_twitter/senators

ability to identify conspiracy content across different contexts and topics, achieving a macro F1-score of 0.81 (Table 11 in appendix A).

4.2. Ground-Truth Dataset Construction

Three experts specializing in the psychology of CTs manually labeled two different subsamples of tweets by examining the textual content to identify conspiracy or non-conspiracy content. We used the majority vote to assign the CT label.

The labeling process was conducted in two stages. In the first stage, we sampled 1,000 tweets. We discovered that this initial dataset was imbalanced, with only about 4% containing CTs. This highlights the infrequent occurrence of CT-related tweets in a general context where hashtags are not specified.

To create a more balanced dataset, we initially generated pseudo-labels using machine learning models. We fine-tuned a BERT model [Devlin et al. (2019)] on the sampled tweets, dividing the data into training, validation, and test sets. This model achieved a macro F1-score of 50%. We then used the trained model to assign ”conspiracy” or ”non-conspiracy” pseudo-labels to all tweets.

Subsequently, we sampled 1,500 tweets with ’conspiracy’ pseudo-labels for manual labeling by experts. After both stages, we obtained a final dataset of 2,500 tweets, with 15% labeled as containing CTs. The annotation process resulted in an average Cohen’s Kappa of 0.55, indicating moderate inter-annotator agreement among the annotators. The dataset is available for research purposes upon request.

	CTBert		Roberta		Bert large	
Dataset	F1score	Test on politicians	F1score	Test on politicians	F1score	Test on politicians
Covid Dataset	0.83	0.23	0.86	0.59	0.84	0.53
Hoaxes Dataset	0.83	0.3	0.54	0.46	0.5	0.33
Youtube Dataset	0.77	0.27	0.59	0.56	0.67	0.57
Younicon Dataset	0.71	0.27	0.66	0.61	0.71	0.61
Combined datasets	0.81	**0.62**	0.83	0.58	0.8	0.56

Table 2: Cross Validation of Existent Datasets

5. Methodology

5.1. Conspiracy detection

5.1.1. Cross validation using existing datasets

We built a classifier to detect CTs using different combinations of existing datasets, as reported in Table 1. Specifically, we developed three classifiers: BERT [Devlin et al. (2019)], CTBERT [Mu¨ller et al. (2023)], and RoBERTa [Liu et al. (2019)], which have been shown to be effective at detecting conspiracy tweets [Phillips et al. (2022)].

Prior to classification, URLs and short tweets were removed. The words in the tweets were tokenized using the default tokenizer of each model. The classifiers were trained with a learning rate of 2e-5 and an Adam parameter of 1e-8, with a batch size of 16. The proportion of each class label was used as a parameter to assign weights for loss calculation, thereby accounting for the unequal instances of each class. Each training run was conducted for four epochs until the loss values converged. Each classifier was evaluated using five-fold cross-validation with an 80:20 train:test split, and the average performance metrics were reported. We used macro-F1 performance metrics to adjust for the imbalance in the proportion of class labels in the dataset.

We employed F1-macro score as our evaluation metric following established practices in CT detection research Phillips et al. (2022); Liaw et al. (2023). As the harmonic mean between precision and recall, F1-score provides a balanced assessment that captures both accurate identification and comprehensive detection of conspiracy tweets. This metric also enables direct comparison with previous approaches, with our model achieving a competitive score of 0.81.

We present the results of our three experiments in Table 2. The highest macro F1score on our ground truth politician dataset was achieved by the CTBERT classifier, which was fine-tuned on the combined datasets from the literature. The macro F1scores of the classifiers may differ from the reported performance in the original papers, as we were unable to rehydrate all tweets when reconstructing the datasets.

Classifiers trained on the existing datasets separately showed lower performance, except for the YouTube and YouTube2 datasets, which achieved a macro F1-score of 0.6 at times. One explanation for this relatively satisfactory performance is that these datasets contain a variety of CTs. However, these classifiers do not generalize well on our politician dataset, necessitating further fine-tuning to improve their performance.

5.1.2. Gradual finetuning & Layer Freezing

The performance of NLP models can degrade significantly when applied to a new domain with different data distributions [Xu et al. (2021)]. Gradual fine-tuning is proposed as a solution to this problem, as it allows the model to adapt to the target domain while retaining the knowledge learned from the source domain. Xu et al. (2021) demonstrated that gradual fine-tuning can improve NLP models for domain adaptation, consistently enhancing performance as the data distribution aligns more closely with that of the target domain. Therefore, we leveraged existing datasets in the literature for conspiracy detection and trained an initial classifier using the combined data as depicted in Table 1. The classifier was trained with a learning rate of 2e-5, an Adam parameter of 1e-8, and a batch size of 16. The proportion of each class label was used as a parameter to assign weights for loss calculation, accounting for the unequal instances of each class. Training was conducted for one epoch to avoid overfitting, and we selected the CTBERT classifier as it performed best in Table 2. Subsequently, we fine-tuned the resulting classifier using the politician dataset while freezing some layers.

Lee et al. (2019) highlight the importance of freezing some layers when fine tuning transformer-based language models for downstream tasks such as BERT and RoBERTa models. Accordingly, we experimented with different numbers of frozen layers. The second classifier was trained using the same parameters as the first, modifying only the number of epochs to five, at which point the loss values converged.

In Table 3, we observe that freezing eight layers yields the best performance, with a macro F1-score of 0.8, a precision of 0.81, and a recall of 0.78. More specifically, when inspecting performance by class, Table **4** presents strong performance metrics across both the CT and non-CTclasses. The conspiracy class achieved a precision of 0.68, recall of 0.74, and an F1-score of 0.70. The high recall indicates that the classifier correctly identifies 74% of the actual conspiracy tweets. The precision shows that 68% of tweets labeled by the classifier as conspiracies truly belong to that class, with an overall recall of 70%.

Model Name	**Training Data**	**Fine-tuning Strategy**	Frozen Layers (2nd FT)	**Precision**	**Recall**	**F1 Score**	**Remarks/Note**
BERT-Large	Politicians	One-step	-	0.76	0.75	0.75	-
Roberta	Politicians	One-step	-	0.77	0.77	0.77	-
CTBERT	Politicians	One-step	-	0.76	0.76	0.75	-
Gradual-CTBERT1	Combined +Politicians	Gradual	ALL Layers	0.58	0. 6	0.73	Refinetuned on Politician Dataset, freezing all Layers in ths second finetuning
Gradual-CTBERT2	Combined+Politicians	Gradual	9 Layers	0.78	0.78	0.78	Refinetuned on Politic ian Dataset, freezi ng 9 Layers
Gradual-CTBERT3	Combined+Politicians	Gradual	8 Layers	**0.81**	**0.78**	**0.8**	Refinetuned on Politic ian Dataset, freezi ng 8 Layers
Gradual-CTBERT4	Combined+Politicians	Gradual	7 Layers	0.79	0.79	0.78	Refinetuned on Politic ian Dataset, freezi ng 7 Layers
Gradual-CTBERT5	Combined+Politicians	Gradual	6 Layers	0.76	0.8	0.78	Refinetuned on Politic ian Dataset, freezi ng 6 Layers

Gradual-CTBERT6	Combined+Politicians	Gradual	0 Layer	0.79	0.78	0.79	Refinetuned on Politician Dataset, without freezing any Layer

Table 3: Conspiracy-detection results

Similarly, for the non-conspiracy class, the classifier achieved precision, recall, and F1-scores all above 0.89, reflecting the classifier's strong performance in correctly identifying tweets that do not contain CTs.

These positive results across both classes demonstrate that fine-tuning the CTBERT model on combined datasets from prior work, followed by politicians' tweets, allows us to correctly identify 74% of true CTs and 91% of true negative cases per class-specific metrics. The model generalizes well despite differences in conspiracy detection experiment

Class	Precision	Recall	F1-score
Conspiracy	0.68	0.74	0.70
Not Conspiracy	0.89	0.91	0.89

Table 4: Class-wise Performance Metrics (Precision, Recall, and F1-Score) for the Best-Performing Model (Gradual-CTBERT3)

topics, domains, and time. Finally, we applied the classifier to the dataset collected in Section 3 to analyze the prevalence of CTs in political discourse.

5.2. Language Characteristics & Topics

To extract the distinctive characteristics of the language used in conspiracy tweets and differentiate them from non-conspiracy tweets, we utilized the Linguistic Inquiry and Word Count (LIWC) software.

Specifically, we employed the LIWC 22 software, which includes several extensively validated dictionaries that enable researchers to make inferences about individuals' psychological states [Boyd et al. (2022)]. According to the authors, the words people use in everyday life provide insights into their psychological states, beliefs, emotions, thinking habits, lived experiences, social relationships, and personalities. There are strong theoretical reasons to suspect that certain linguistic features characterize the expression of online conspiracy conversations, as people adopt CTs in an effort to fulfill basic psychological needs [Douglas et al. (2017)]. We focused on psychological themes outlined in the LIWC dictionary that are particularly relevant to conspiracy ideation [Douglas et al. (2017); Fong et al. (2021)], including ingroup versus outgroup language (e.g., we, us, vs. they, them); cognition (e.g., cause, know); negative emotions such as anger (e.g., hate) and anxiety (e.g., nervous, afraid); and themes related to popular CTs such as narratives revolving around power (e.g., superior), death (e.g., bury, kill), and religion (e.g., church, mosque). Additionally, we examined time orientation by analyzing words related to past versus present orientation using the LIWC software to understand the temporality of events driving CTs.

We also examined the topics present in conspiracy tweets by applying nonnegative matrix factorization (NMF), as it has been shown to perform well on tweets and short texts [Chen et al. (2019); Egger and Yu (2022); Athukorala and Mohotti (2022)]. NMF works on TF-IDF transformed data by decomposing a matrix into two lower-ranking matrices and does not require prior knowledge for topic extraction, making it a suitable algorithm for social media data analysis. Compared to Latent Dirichlet Allocation (LDA), NMF has been found to produce better results in terms of topic identification and human interpretation [Egger and Yu (2022); Athukorala and Mohotti (2022)].

To determine the optimal number of topics, we applied the elbow method to the reconstruction error plot [Guven (2023)], after removing stopwords and links, and lowercasing the tweets. We found that the optimal number of topics, ranging from one to ten, is eight.

6. EMPIRICAL ANALYSIS

6.1. Conspiracy distribution in political discourse

We first analyzed the prevalence of CTs as a percentage of total tweets for each party. Conspiracy tweets represented less than 5% for both parties, with a slightly higher proportion among Republicans who tweeted more overall (see Table 5).

Regarding tweet characteristics(Table 6), CT tweets were typically longer (Democrats: M = 237.10 characters, SD = 57.14; Republicans: M = 233.49 chars, SD = 67.70) compared to non-CT tweets (Democrats: M = 184.08 chars, SD = 73.43; Republicans: M = 165.27 chars, SD = 73.80). Link presence was common across both parties (Democrats: 69.88%, Republicans: 73.78% for non-CT tweets).

6.2. Language Characteristics & Topics

Next, we examined the temporal distribution of conspiracy tweets by party separately. Figure 3 shows an increase in Democrats' conspiracy tweets from Q3 2016 to Q3 2020, period corresponding with the Trump Administration and the role of the Democrats as the opposition party. In contrast, Republicans' conspiracy tweets spiked starting Q3 2019 the beginning of the COVID-19 pandemic, continuing as Democrats took power. This is consistent with the hypothesis that "conspiracy theories are for losers", in that the party out of power tends to focus its effort on discrediting the party in power, and therefore, they might use CTs to achieve this goal Edelson et al. (2017). For example, Uscinski and Parent (2014) analyzed New York Times letters between 1890-2010, and found that Republicans blamed the Democrats and Socialists when the Democrats held the presidency, and vice versa when the Republicans were in office. They also suggested that CTs focus their narratives around foreign enemies during war times. CTs also focused more on foreign enemies during wars and the Cold War Uscinski and Parent (2014). Besides its political utility, another reason for the increase of CT content is that situational factors like losing political control could increase conspiracy thinking Douglas et al. (2019). In addition, CTs can be a coping mechanism in contexts of intergroup conflict (e.g., a political election) to protect the image of the ingroup by portraying the outgroup as an evil conspiracy agent [Biddlestone et al. (2021)].

Party	Users	Total Tweets	CT Tweets (%)	Non-CT Tweets (%)
Democrat	107	368,388	2.45	97.55
Republican	283	571,541	3.84	96.16

Table 5: Overall Dataset Distribution by Party

Party	Tweet Type	Mean Length	SD	Median	Links (%)
Democrat	CT	237.10	57.14	257.0	60.95
	Non-CT	184.08	73.43	164.0	69.88
Republican	CT	233.49	67.70	259.0	68.41
	Non-CT	165.27	73.80	140.0	73.78

Table 6: Tweet Characteristics by Party and CT Classification

Metric	Mean	SD	Min	Median	Max
Tweets per user	2,584.6	2,166.1	15	2,448.5	21,613
CT tweets per user	86.0	108.7	0	50.0	765
CT percentage	3.65	3.99	0.00	2.24	30.47

Table 7: User-Level Activity Statistics

Metric	Mean	SD	Min	Max
Total tweets	8,825.5	4,378.5	3,114	21,613
CT tweets	196.4	194.0	8	635
CT percentage	2.42	2.71	0.14	11.54

Table 8: Super-Poster Characteristics (Top 5%, n=22)

Following our BERT-based classification of CT tweets, we conducted additional analyses to examine temporal distribution while accounting for the zero-inflation in our data. Standard mixed-effects models proved unsuitable due to convergence problems arising from both the lack of variance and highly skewed distribution in the counts of CT tweets. Therefore, we employed a Mixed-Effect Hurdle model, which is specifically designed for count data with excess of zeros (Mullahy, 1986)—a characteristic of our dataset, where many politicians did not post tweets identified as CT tweets at all. The Hurdle model consisted of two components: i) a zero-inflation component (logistic regression) modelling the probability of zero values (i.e., of not posting a CT tweet), and ii) a conditional model (regression assuming a truncated negative binomial distribution), which accounted for the non-zero values (i.e., count of CT tweets among users who posted at least 1). Both components included the fixed effects of Party and Year, as well as the Party × Year interaction term. The model also accounted for the repeated measures from the same users. Note that the proportion of zero-inflation pre-2016 was over 90%, likely due to the inactivity of the Twitter accounts of some of the politicians in our sample. Thus, we focused our analysis on the 2016-2023 time period (3,120 observations from 390 users), which, in addition to showing more consistent Twitter activity, it coincided with a politically relevant time period, specifically the beginning of the first administration of President Donald Trump.

The Hurdle model revealed three key findings. First, the zero-inflation component showed an increasing probability of posting CT tweets over time, relative to 2016 (2021: = -3.41 , $p < 0.001$). However, this increase was not conditional on the users' party affiliation, which suggests a growing generalized engagement in conspiracy sharing. Second, while we did not observe differences across parties in the probability to post CT tweets, the count component showed these differences in the amount of CT tweets posted by users from each party. Specifically, we observed significant Party × Year interactions terms, which indicated that while the count of CT tweets increased over time relative to 2016, it did so differently for Democrat and Republican politicians (see Figure 2). Noteworthy, the sign of the interaction terms reversed around 2020, indicating a shift from the period 2016-2020 when Democrats posted more CT tweets than Republicans, to the period 2020-2023 when Republicans posted more CT tweets than Democrats. This reversal aligns with changes in government, demonstrating how the post of CT tweets varied systematically with political power dynamics, becoming more frequent among politicians from the opposition party. Finally, individual-level variation was substantial (random effects variance = 1.045), indicating persistent differences in conspiracy sharing propensity across users.

6.3. Interaction with Conspiracy Theories

To address RQ1, we examined whether CT content in politicians tweets influences engagement. Confirming H1, when inspecting the average mean of likes and shares for CT vs. non-CT tweets as depicted in Figure 4, we can clearly see that CT tweets trigger more interaction than non-CT tweets. This supports research suggesting that conspiracy tweets trigger more interactions than scientific tweets [Zhang et al. (2021)].

6.4. Language Characteristics

6.4.1. Group Dynamics and Temporal Focus

In CT tweets, the ingroup is often depicted as the ”victim” of a conspiracy by the outgroup [Mashuri and Zaduqisti (2015)]. Conspiracy thinkers typically feel oppressed by a powerful outgroup, framing them as a deceitful enemy [van Prooijen and Van Lange (2014)]. Our analysis, shown in Figure 6, reveals a higher usage of outgroup language (e.g., they, them, she/he) in conspiracy tweets. The word clouds (Figure 5) further illustrate this tendency, with Democrats mentioning "Trump" and "Republicans," while Republicans mention "Democrats," "Biden," and "China."

Additionally, we measured time orientation by comparing past versus present language usage in conspiracy tweets. Contrary to previous research [Fong et al. (2021)], our findings indicate that politicians' conspiracy tweets are more focused on current events using the present tense rather than past events. This suggests that political conspiracies aim to challenge contemporary narratives and legitimacy.

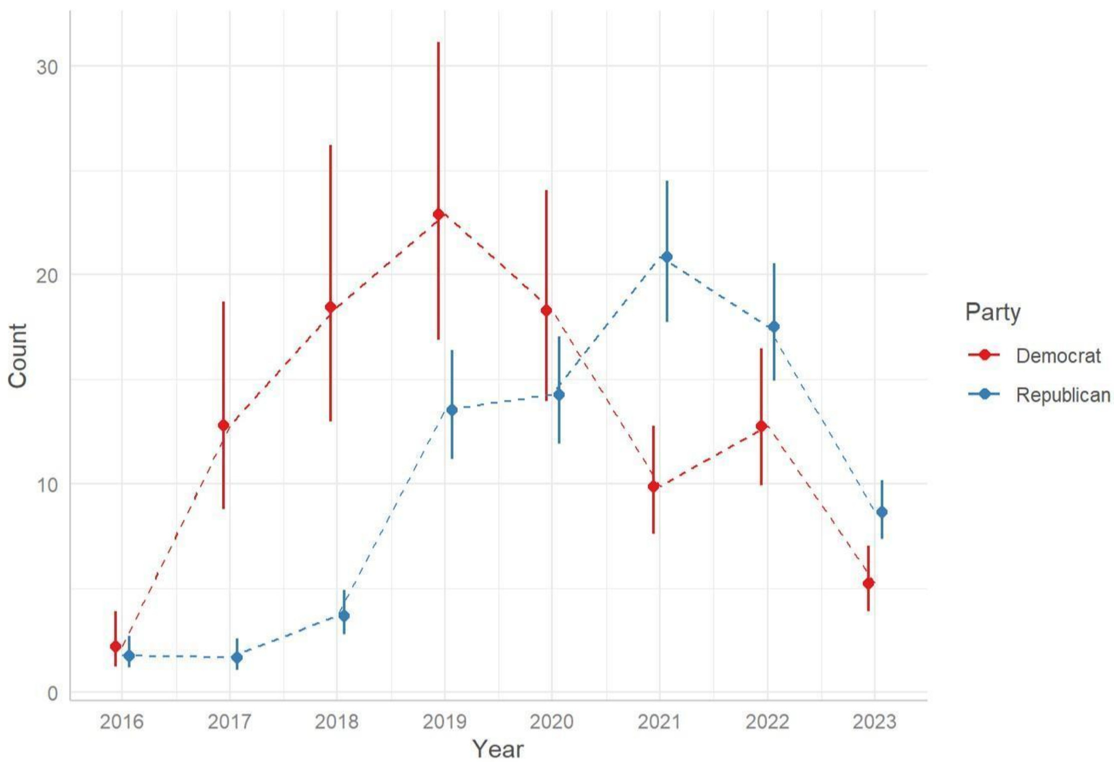


Figure 1: Predicted counts of CT tweets using Hurdle model. Error bars represent 95% CIs.

Year	Democrats	Republicans	Interpretation
2016	0.789*	0.572	Base year: Similar levels
2017	1.762*	-0.061	Democrat increase, Republican decrease

2018	2.126*	0.729	Strong Democrat dominance
2019	2.343*	2.035	High levels for both parties
2020	2.119*	2.086	High levels for both parties
2021	1.499*	2.466	Republican increase during Biden admin
2022	1.759*	2.291	Sustained Republican elevation
2023	0.864	1.584	Moderation for both parties

Note: Effects shown are from the count component of the hurdle model. Values are log counts. Significance levels: * p¡0.001, p¡0.01, * p¡0.05. Republican effects calculated by combining base effect and interaction term.

Table 9: Temporal Patterns in Conspiracy Theory Sharing by Party (2016-2023)

6.4.2. Emotional and Cognitive Analysis

To address RQ2 and RQ3–concerning the linguistic features of CTs and the potential role of powerlessness–we used the LIWC dictionary to assess two primary components in politicians' tweets: negative emotions, and cognitive processes. Confirming H3, our analysis revealed higher usage of power-related words in conspiracy tweets (Figure 7), suggesting that these tweets may help individuals express feelings of powerlessness [Douglas et al. (2019); Uscinski and Parent (2014)]. Additionally, confirming H2, conspiracy tweets showed higher usage of causal language (e.g., "how," "why," "because"), indicating that CTs offer explanations for complex events [Fong et al. (2021)].

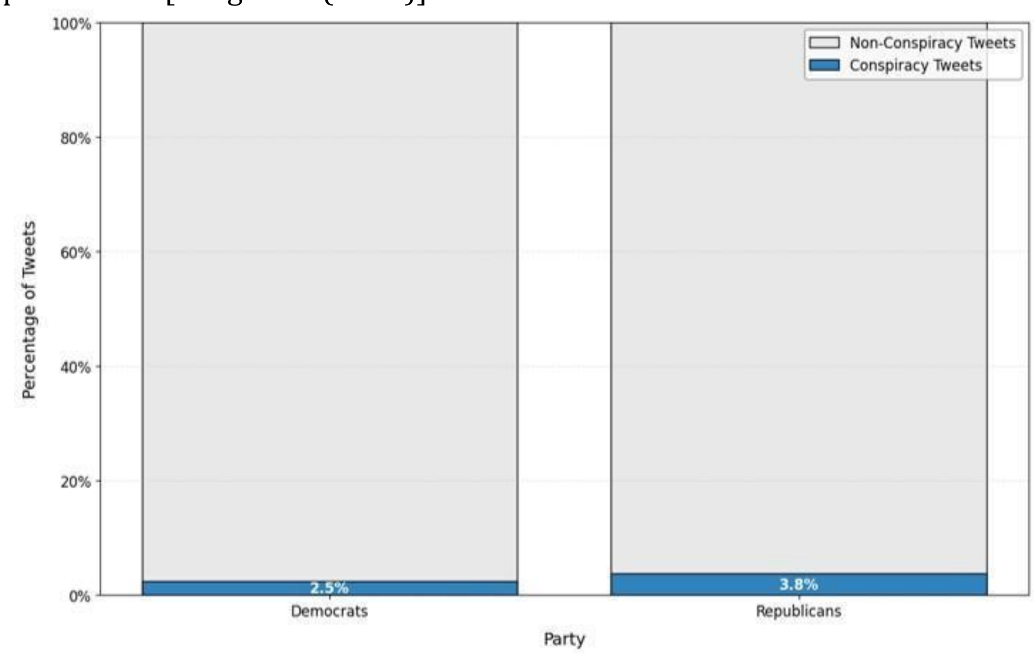
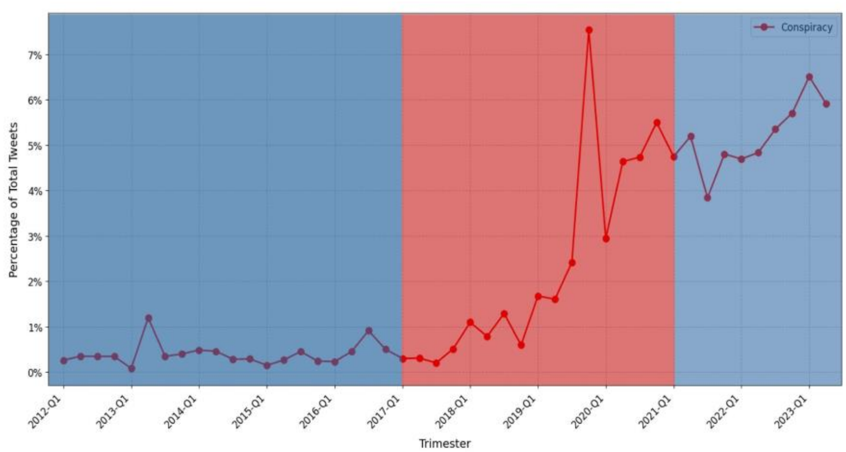


Figure 2: The ratio of conspiracy tweets from total tweets

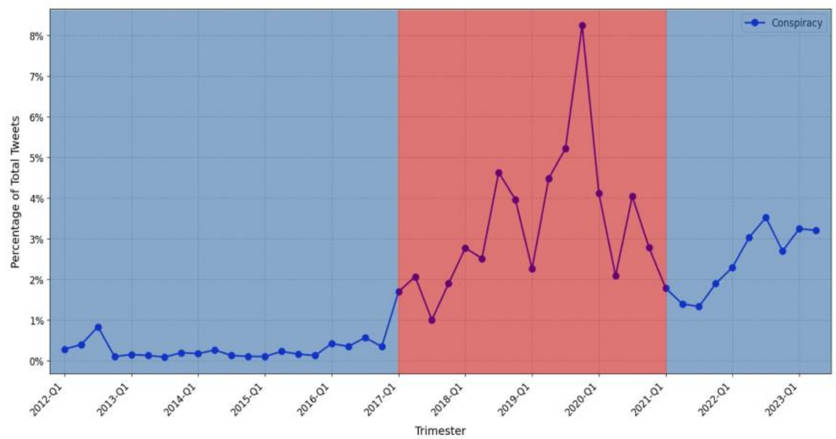
6.4.3. Topics

To understand potential confounders in the form of topical focus of tweets, we conducted a post-hoc topic analysis following our initial classification of CT tweets using a BERT-based classifier. We implemented topic modeling using Non-negative Matrix Factorization (NMF) that identified eight distinct topics across all tweets. Our analysis revealed systematic variations in topic distribution between tweets previously classified as CT and non-CT, as well as between Democratic and Republican parties, confirming topics as potential confounders. The visualization of topic distributions Figure 8 demonstrated clear differences in topical focus between CT and non-CT tweets, as well as between parties, validating the importance of considering these effects.

Figure 8 along with Table 10 illustrates the topic distribution across CT and non-CT tweets for Democrats and Republicans, revealing distinct patterns. Democratic CT tweets primarily focus on Topic 1 (Healthcare Public Service) and Topic 4 (Administration Economy), reflecting policy-oriented concerns, while Republican CT tweets emphasize Topic 8 (Border Security), highlighting security-related themes. Non-CT tweets for both parties show more balanced distributions, with Democrats emphasizing Topic 1 (Healthcare Public Service) and Topic 2 (Social Issues), and Republicans focusing on Topic 4 (Administration Economy) and Topic 5 (Updates News). Ceremonial topics such as Topic

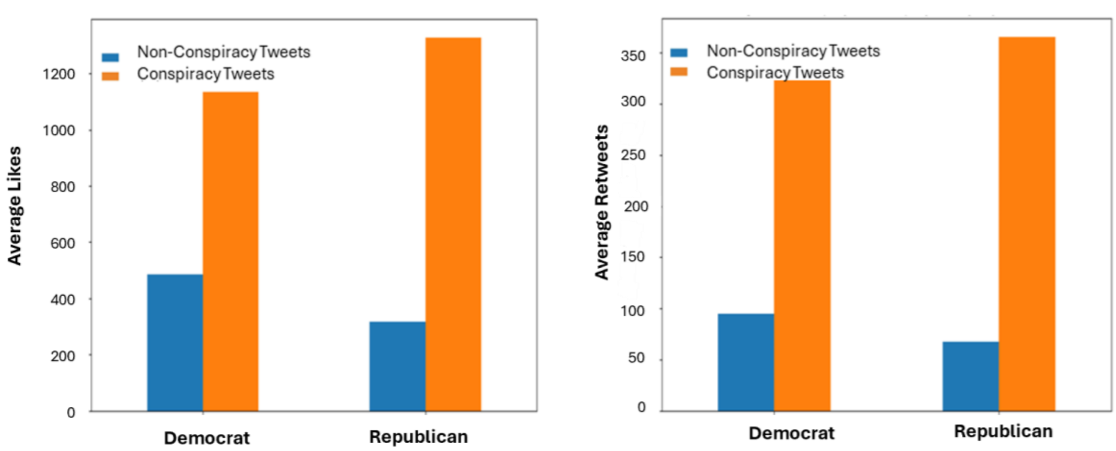


(a) The distribution of conspiracy tweets across years for Republicans



(b) The distribution of conspiracy tweets across years for Democrats

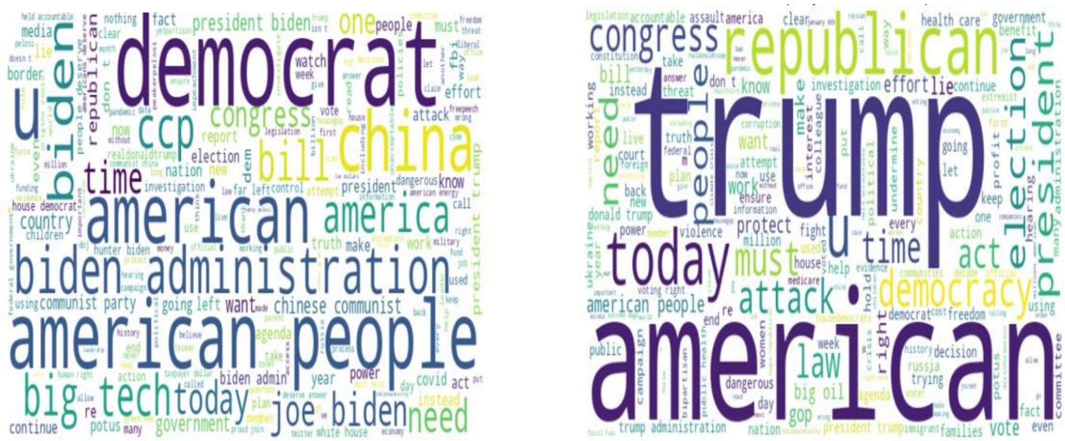
Figure 3: The distribution of conspiracy tweets across years and parties: blue intervals represent Democratic presidency, red intervals indicate Republican presidency



(a) Mean of Likes

(b) Mean of Shares

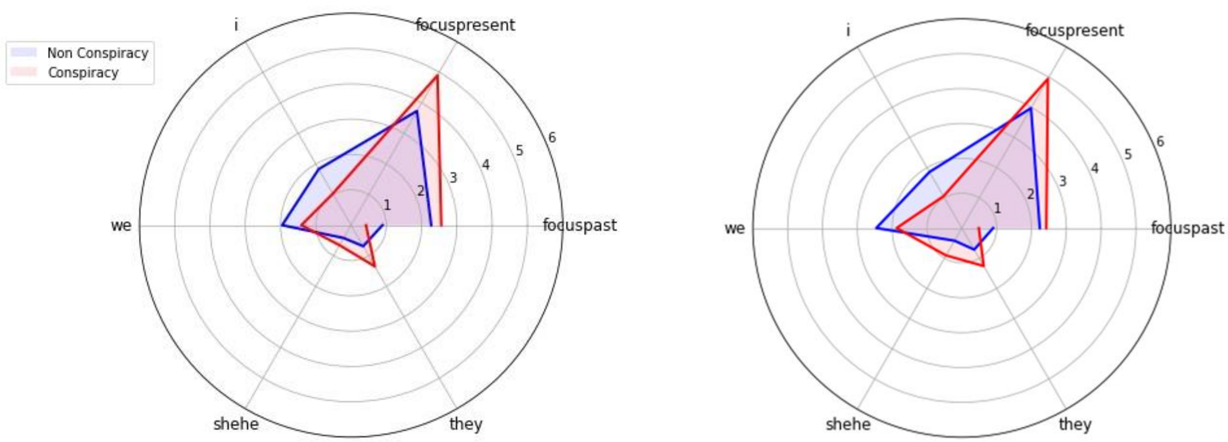
Figure 4: Mean differences in engagement between conspiracy and non-conspiracy tweets



(a) Republicans

(b) Democrats

Figure 5: Word clouds of conspiracy tweets for both parties



(a) Republicans

(b) Democrats

Figure 6: Mean differences in ingroup vs. outgroup language and time orientation between the conspiracy and non-conspiracy tweets

(a) Republicans (b) Democrats

Figure 7: Mean differences in emotion and cognitive processes between the conspiracy and non-conspiracy tweets

6 (Military Service) and Topic 7 (Celebrations) are more common in non-CT tweets for both parties, suggesting broader engagement outside polarizing themes. Overall, conspiracy-related tweets are concentrated on divisive issues, whereas non-CT tweets address a wider range of general and ceremonial topics.

7. Discussion

Our analysis revealed several key insights into the use of CTs in politicians' tweets. First, we found that politicians tended to embrace CTs more when their party was out of power (H3), confirming previous research that conspiracy theories appeal to "losers" who feel disempowered and in need of enhancing the image of their political group. Specifically, Republicans propagated more CTs when Democrats held the presidency, while Democrats did so when Republicans were in power. The more frequent reliance on conspiracy narratives among the party in the political opposition likely reflects individual efforts to enhance the party's group image by portraying political oponents as nefarious conspirators. From a psychological perspective, CTs have been argued to fulfill the psychological motivation of maintaining positive evaluations of the self and the group one belongs to (Douglas et al., 2017; Biddlestone et al., 2021). In the political domain, sharing conspiracy theories about opposing parties could help regaining the confidence of affiliated voters, consolidating their negative perceptions toward the opposing party, and fueling mobilization (Barlev & Neuberg, 2025), especially in times where political prestige has been eroded. For example, political losers of democratic elections often endorse conspiracy theories about electoral fraud,, arguably exhibiting a collective process of motivated reasoning (Uscinsky & Parent, 2014). Our findings suggests that this intentional use of conspiracy theories extends beyond electoral processes throughout the political term as a function of who is in power.

Second, we found that conspiracy tweets garnered higher engagement in terms of likes and shares compared to regular tweets (H1). This finding aligns with prior research showing that CTs draw more interactions in social media [Zhang et al. (2021); Vosoughi et al. (2018)]. Beyond their entertaining value (van Prooijen et al., 2022), the rhetorical features of CTs as narratives makes them particularly persuasive despite their epistemic flaws (Oswald, 2016), especially among those predisposed to endorse them and getting exposed to them (Uscinsky et al., 2022). This justifies the appeal that sharing conspiracy theories have for politicians - conspiracy content draw attention, an important currency within politics.

Related to the previous, a third set of findings refered to the linguistic features of CTs. We observed that CTs tweets made a higher use of causal explanations, certitude, and outgroup derogation. These features matched established attributes of

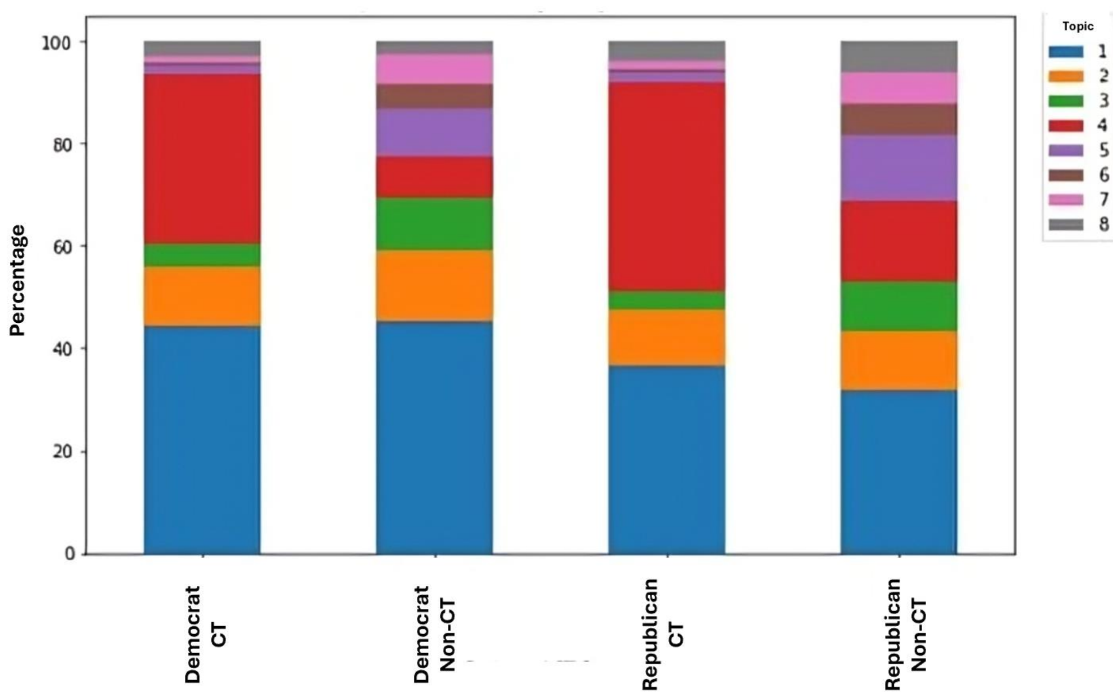


Figure 8: Topic Distribution by Party and CT Status

Category	Topics	Key Terms

Policy	Topic #1: Healthcare & Public Service	act, need, help, health, support, care, proud, work, congress, make
	Topic #8: Border & Security	border, crisis, southern, security, secure, illegal, biden, national, fentanyl, administration
Political	Topic #4: Administration & Economy	biden, president, american, energy, administration, people, americans, inflation, trump, joe
	Topic #2: Social Issues	women, family, community, jobs, security, potus, congrats, admin, joined
Ceremonial	Topic #6: Military & Service	thank, service, veterans, honor, women, nation, work, men, country, community
	Topic #7: Celebrations	happy, day, birthday, wishing, family, year, celebrate, celebrating, nation, safe
General	Topic #3: Events & Memory	today, honor, tune, day, years, live, remember, hearing, ago, join
	Topic #5: Updates & News	great, thanks, news, morning, work, students, meeting, county, state, time

Table 10: Thematic Distribution of Topics in Politicans' Tweets

conspiracy thinking (H2) and depict some of the persuasive elements of CTs. For example, the expression of causality and certainty about complex societal events, should appeal individuals with strong epistemic and existential needs (Douglas et al., 2017) If sharing CTs additionally serve to attack the rival party,, and frame them as a dangerous, deceitful enemy, this could further satisfy the need to engance one's group image. In contrast to existing research [Fong et al. (2021)], our data showed that politicians tended to focus on current events rather than bringing order to past events.

Fourth, our preliminary topic analysis suggested systematic variations in how CTs manifest across different topics and parties. By identifying eight distinct topics, we found initial evidence of differences in topical focus between CT and non-CT tweets, as well as between the Democratic and Republican parties. The topic of the CT could, therefore, be a potential confound in understanding partisan sharing of CTs.

In summary, our analysis reveals that CTs are present in political discourse and follow systematic patterns tied to power dynamics and party status. Importantly, the higher engagement rates of conspiracy tweets compared to regular tweets suggest practical incentives for their use, despite the risks that sharing CTs might poses in terms of truth, trust, and democratic governance [van Prooijen et al. (2022)].

8. Practical Implications

The present research has direct practical implications related to the understanding of political discourse strategies and the countering of the political effects of conspiracy theories and misinformation.

The findings on the trends and content of conspiracy discourse can inform media literacy curricula to emphasize how political status influences susceptibility to and promotion of CTs, helping people critically assess the motivations behind political messaging. Additionally, they might be informative for regulatory bodies of democratic elections to establish measures of monitoring of the political rhetoric and promotion of transparency and accountability regulations, especially in times of political transitions when parties in the political opposition seem more likely to rely on conspiracy narratives.

The observed use of CT tweets further suggests that political parties and campaign strategists assume CTs to have favorable boosting effects for their candidates and political discourse. This assumption might not be totally unwarranted, considering that in the US context, sharing conspiracy theories has been observed to favourably influence the interpersonal perceptions of part of the political spectrum (i.e., to appear as a "rogue" political outsider who can effect change; Green et al., 2023. However, as previously noted, CTs could have (un-)intended broader long-term eroding effects on the public's institutional trust and political engagement (Douglas & Sutton, 2023) that responsible politicians and political advisors should not overlook.

9. Limitations and Future Work

This study has several limitations that warrant further discussion. First, the class label imbalance in the training data presents a challenge, as conspiracy-related tweets are relatively rare in datasets not filtered by hashtags or keywords. This imbalance may lead to reduced classification performance, particularly for the minority class. While we addressed this through weighted loss functions and fine-tuning, future work should explore more balanced sampling strategies or synthetic data augmentation to improve performance.

Second, although our ground truth annotation was carried out by experts in the psychology of conspiracy theories, the resulting Cohen's Kappa of 0.55 indicates only moderate inter-annotator agreement. This highlights the inherent ambiguity and subjectivity in labeling conspiracy content, which may affect model robustness. Future work should consider refining annotation guidelines and employing iterative or crowd-based labeling strategies with consensus mechanisms.

Third, reproducibility of our results is partially constrained by the nature of Twitter data. Some tweets may no longer be accessible due to deletion, account suspensions, or changes in platform policies. Additionally, our access to the Twitter Historical API—which is no longer openly available—may pose challenges for researchers aiming to replicate or extend our work. To mitigate this, we have made our annotated accessible upon request.

Fourth, while our findings provide insight into the U.S. political landscape, the generalizability to other sociopolitical contexts or platforms (e.g., Facebook, Telegram) remains an open question. Future research should assess whether similar dynamics hold across different countries, political systems, or platform affordances. Additionally, we consider important to explore topical variations in more depth, potentially using more sophisticated topic modeling approaches and larger datasets to better understand how CTs manifest differently across political topics and party lines.

10. Conclusion

Our analysis offers new insights into politicians' tactical use of conspiracy theories on social media. We find compelling evidence that electoral dynamics and loss of power motivate politicians to embrace conspiracy narratives to energize supporters, attack opponents, explain failures, and sow policy doubts. The findings demonstrate both the allure and risks of conspiracy theorizing in the political domain. Further research should build on these findings to deepen our understanding of this challenging phenomenon at the intersection of politics, communication, and misinformation. Tackling the causes and impacts of political conspiracy theories will only grow in importance for scholars, journalists, policymakers, and informed, engaged citizens.

11. Acknowledgments

This research was supported by the European Research Council Advanced Grant "Consequences of conspiracy theories - CONSPIRACY FX" Number: 101018262, awarded to Karen M.Douglas.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used CLAUDE in order to proofread the paper. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Appendix A. Tweet Statistics by Party and Year

Our final preprocessed dataset consists of 1,013,152 original tweets from 394 users (107 Democrats and 283 Republicans), with a median of 2,349 tweets per account. At the party level, Republicans contributed 635,562 tweets, while Democrats contributed 371,460 tweets. At the user level (Table 7), members of Congress posted an average of 2,584 tweets (SD = 2,166), with activity ranging from 15 to 21,613 tweets (Table 5). We identified "super-posters" in Table 8 (top 5%, *n* = 22), who posted between 3,114 and 21,613 tweets (M = 8,825, SD = 4,378) as depicted in Table 11..

Party	**Year**	**Active Users**	**Total Tweets**	**Mean Tweets/User**	**Std Dev Tweets/User**	**Min Tweets**	**Max Tweets**
Democrat	2012	16	5,096	318.50	367.42	51	1,507
	2013	21	8,541	406.71	340.11	64	1,445
	2014	23	9,166	398.52	383.31	5	1,717
	2015	25	12,176	487.04	480.97	60	2,270
	2016	26	13,590	522.69	459.01	33	2,182
	2017	32	20,527	641.47	499.82	1	2,434
	2018	35	22,217	634.77	569.40	25	2,962
	2019	53	30,274	571.21	432.07	45	2,283
	2020	80	55,987	699.84	466.19	2	2,376
	2021	100	82,304	823.04	411.18	95	2,369
	2022	104	86,813	834.74	487.66	1	2,301
	2023	88	21,697	246.56	167.30	1	666
Republican	2012	26	6,484	249.38	289.68	1	1,124
	2013	37	13,040	352.43	457.77	8	2,371
	2014	41	14,421	351.73	360.10	16	1,930

Party	Year	Active	Total	Mean	Std Dev	Min	Max
	2015	54	15,941	295.20	268.99	1	1,146
	2016	64	20,786	324.78	415.37	1	2,430
	2017	92	30,779	334.55	389.21	4	3,251
	2018	112	35,931	320.81	309.48	1	2,651
	2019	149	60,068	403.14	300.88	1	2,943
	2020	175	85,862	490.64	345.62	23	2,869
	2021	230	137,025	595.76	375.00	9	3,014
	2022	241	150,707	625.34	485.53	1	2,442
	2023	269	57,598	214.12	177.00	1	1,058

Table 11: Tweet Statistics by Party and Year (2009-2023)

12. References

Allcott, H., Gentzkow, M., 2017. Social media and fake news in the 2016 election. Journal of economic perspectives 31, 211–236. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418.

Athukorala, S., Mohotti, W., 2022. An effective short-text topic modelling with neighbourhood assistance-driven NMF in Twitter. Social Network Analysis and Mining 12, 89. URL: https://doi.org/10.1007/s13278-022-00898-5, doi:10. 1007/s13278-022-00898-5.

Aïmeur, E., Amri, S., Brassard, G., 2023. Fake news, disinformation and misinformation in social media: a review. Social Network Analysis and Mining 13, 30. URL: https://link.springer.com/10.1007/s13278-023-01028-5, doi:10.1007/s13278-023-01028-5.

Barlev, M., & Neuberg, S. L. (2024). Rational reasons for irrational beliefs. American Psychologist.

Biddlestone, M., Green, R., Cichocka, A., Sutton, R., Douglas, K., 2021. Conspiracy beliefs and the individual, relational, and collective selves. Social and Personality Psychology Compass 15. doi:10.1111/spc3.12639. place: United Kingdom Publisher: Wiley-Blackwell Publishing Ltd.

Boyd, R.L., Ashokkumar, A., Seraj, S., Pennebaker, J.W., 2022. The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin , 1–47.

Carrasco-Farr´e, C., 2022. The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. Humanities and Social Sciences Communications 9, 1–18. URL: https://www.nature.com/articles/s41599-022-01174-9. publisher: Palgrave.

Chen, Y., Zhang, H., Liu, R., Ye, Z., Lin, J., 2019. Experimental explorations on short text topic mining between LDA and NMF based Schemes. Knowledge-Based Systems 163, 1–13. URL: https://www.sciencedirect.com/science/article/pii/S0950705118304076, doi:10.1016/j.knosys.2018.08.011.

Desilver, D., 2022. Some members of Congress have been there for decades, but seats typically change hands more frequently.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota. pp. 4171–4186. URL: https://aclanthology.org/N19-1423, doi:10.18653/v1/N19-1423.

Douglas, K.M., Sutton, R.M., 2023. What Are Conspiracy Theories? A Definitional Approach to Their Correlates, Consequences, and Communication. Annual Review of Psychology 74, 271–298. URL: https://doi.org/10.1146/ annurev-psych-032420-031329, doi:10.1146/annurev-psych-032420-031329. eprint: https://doi.org/10.1146/annurev-psych-032420-031329.

Douglas, K.M., Sutton, R.M., Cichocka, A., 2017. The Psychology of Conspiracy Theories. Current Directions in Psychological Science 26, 538–542. URL: https://doi.org/10.1177/0963721417718261, doi:10.1177/0963721417718261. publisher: SAGE Publications Inc.

Douglas, K.M., Uscinski, J.E., Sutton, R.M., Cichocka, A., Nefes, T., Ang, C.S., Deravi, F., 2019. Understanding conspiracy theories. Political psychology 40, 3–35. Publisher: Wiley Online Library.

Edelson, J., Alduncin, A., Krewson, C., Sieja, J.A., Uscinski, J.E., 2017. The Effect of Conspiratorial Thinking and Motivated Reasoning on Belief in Election Fraud. Political Research Quarterly 70, 933–946. URL: https://doi.org/10.1177/1065912917721061, doi:10.1177/1065912917721061. publisher: SAGE Publications Inc.

Egger, R., Yu, J., 2022. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Frontiers in Sociology

7. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9120935/, doi:10.3389/fsoc.2022.886498. publisher: Frontiers Media SA.

Enders, A.M., Uscinski, J.E., Seelig, M.I., Klofstad, C.A., Wuchty, S., Funchion, J.R., Murthi, M.N., Premaratne, K., Stoler, J., 2023. The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation. Political Behavior 45, 781–804. URL: https://doi.org/10.1007/s11109-021-09734-6, doi:10.1007/s11109-021-09734-6.

Enli, G.S., Skogerbø, E., 2013. Personalized campaigns in party-centred politics: Twitter and Facebook as arenas for political communication. Information, communication & society 16, 757–774. Publisher: Taylor & Francis.

Faddoul, M., Chaslot, G., Farid, H., 2020. A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos. ArXiv URL: https://www.semanticscholar.org/paper/e9b8266ec5d94660dde5f89b4e01586dcddf2515.

Fong, A., Roozenbeek, J., Goldwert, D., Rathje, S., van der Linden, S., 2021. The language of conspiracy: A psychological analysis of speech used by conspiracy theorists and their followers on Twitter. Group Processes & Intergroup Relations 24, 606–623. Publisher: SAGE Publications Sage UK: London, England.

Green, R., Toribio-Flórez, D., Douglas, K. M., Brunkow, J. W., & Sutton, R. M. (2023). Making an impression: The effects of sharing conspiracy theories. Journal of Experimental Social Psychology, 104, 104398.

Guven, E., 2023. Decision of the Optimal Rank of a Nonnegative Matrix Factorization Model for Gene Expression Data Sets Utilizing the Unit Invariant Knee Method: Development and Evaluation of the Elbow Method for Rank Selection. JMIR Bioinformatics and Biotechnology 4, e43665. URL: https://bioinform.jmir.org/2023/1/e43665, doi:10.2196/43665. company: JMIR Bioinformatics and Biotechnology Distributor: JMIR Bioinformatics and Biotechnology Institution: JMIR Bioinformatics and Biotechnology Label: JMIR Bioinformatics and Biotechnology Publisher: JMIR Publications Inc., Toronto, Canada.

Hong, S., Choi, H., Kim, T.K., 2019. Why do politicians tweet? Extremists, underdogs, and opposing parties as political tweeters. Policy & internet 11, 305–323. Publisher: Wiley Online Library.

Klein, C., Clutton, P., Dunn, A.G., 2019. Pathways to conspiracy: The social and linguistic precursors of involvement in Reddit's conspiracy theory forum. PloS one 14, e0225098. Publisher: Public Library of Science San Francisco, CA USA.

Korenčić, D., Chulvi, B., Casals, X.B., Toselli, A., Taul´e, M., Rosso, P., 2024. What distinguishes conspiracy from critical narratives? A computational analysis of oppositional discourse. Expert Systems 41, e13671. URL: https://onlinelibrary.wiley.com/doi/10.1111/exsy.13671, doi:10.1111/exsy.13671.

Langguth, J., Schroeder, D.T., Filkukova, P., Brenner, S., Phillips, J., Pogorelov, K., 2023. COCO: an annotated Twitter dataset of COVID19 conspiracy theories. Journal of Computational Social Science 6, 443–484. URL: https://link.springer.com/10.1007/s42001-023-00200-3, doi:10.1007/s42001-023-00200-3.

Lasser, J., Aroyehun, S.T., Carrella, F., Simchon, A., Garcia, D., Lewandowsky, S., 2023. From alternative conceptions of honesty to alternative facts in communications by US politicians. Nature Human Behaviour , 1–12URL: https://www.nature.com/articles/s41562-023-01691-w, doi:10.1038/s41562-023-01691-w. publisher: Nature Publishing Group.

Lasser, J., Aroyehun, S.T., Simchon, A., Carrella, F., Garcia, D., Lewandowsky, S., 2022. Social media sharing of low-quality news sources by political elites. PNAS Nexus 1, pgac186. URL: https://doi.org/10.1093/pnasnexus/pgac186, doi:10.1093/pnasnexus/pgac186.

Lee, J., Tang, R., Lin, J., 2019. What Would Elsa Do? Freezing Layers During Transformer Fine-Tuning. URL: http://arxiv.org/abs/1911.03090. arXiv:1911.03090 [cs].

Liaw, S., Huang, F., Benevenuto, F., Kwak, H., An, J., 2023. YouNICon: YouTube's CommuNIty of Conspiracy Videos. URL: http://arxiv.org/abs/2304.05274, doi:10.48550/arXiv.2304.05274. arXiv:2304.05274 [cs].

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. URL: http://arxiv.org/abs/1907.11692, doi:10.48550/arXiv. 1907.11692. arXiv:1907.11692 [cs].

Mashuri, A., Zaduqisti, E., 2015. The effect of intergroup threat and social identity salience on the belief in conspiracy theories over terrorism in Indonesia: Collective angst as a mediator. International Journal of Psychological Research 8, 24–35. doi:10.21500/20112084.642. place: Colombia Publisher: University of San Buenaventura.

Mitra, T., Counts, S., Pennebaker, J., 2016. Understanding anti-vaccination attitudes in social media, in: Proceedings of the International AAAI Conference on Web and Social Media, pp. 269–278. Issue: 1.

Moffitt, J.D., King, C., Carley, K.M., 2021. Hunting conspiracy theories during the COVID-19 pandemic. Social Media+ Society 7, 20563051211043212. Publisher: SAGE Publications Sage UK: London, England.

Mosleh, M., Rand, D.G., 2022. Measuring exposure to misinformation from political elites on Twitter. Nature Communications 13, 7144. doi:10.1038/ s41467-022-34769-6.

Mullahy, J., 1986. Specification and testing of some modified count data models. Journal of econometrics 33, 341–365. URL: https://www.sciencedirect.com/science/article/pii/0304407686900023. publisher: Elsevier.

Müller, M., Salathé, M., & Kummervold, P. E. (2023). Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. Frontiers in artificial intelligence, 6, 1023281.

Phadke, S., Samory, M., Mitra, T., 2022. Pathways through conspiracy: the evolution of conspiracy radicalization through engagement in online conspiracy discussions, in: Proceedings of the international AAAI conference on web and social media, pp. 770–781. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/ 19333.

Phillips, S.C., Ng, L.H.X., Carley, K.M., 2022. Hoaxes and Hidden agendas: A Twitter Conspiracy Theory Dataset: Data Paper, in: Companion Proceedings of the Web Conference 2022, pp. 876–880.

van Prooijen, J.W., Douglas, K.M., 2017. Conspiracy theories as part of history: The role of societal crisis situations. Memory Studies 10, 323–333. doi:10.1177/ 1750698017701615.

van Prooijen, J.W., Spadaro, G., Wang, H., 2022. Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships. Current Opinion in Psychology 43, 65–69. URL: https://www.sciencedirect.com/science/article/pii/S2352250X21000828, doi:10.1016/j.copsyc.2021.06.013.

van Prooijen, J.W., Van Lange, P.A., 2014. 13 The social dimension of belief in conspiracy theories. Power, politics, and paranoia: Why people are suspicious of their leaders , 237Publisher: Cambridge University Press.

Van Prooijen, J. W., Ligthart, J., Rosema, S., & Xu, Y. (2022). The entertainment value of conspiracy theories. British Journal of Psychology, 113(1), 25-48.

Samory, M., Mitra, T., 2018. Conspiracies Online: User Discussions in a Conspiracy Community Following Dramatic Events. Proceedings of the International AAAI Conference on Web and Social Media 12. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/15039, doi:10.1609/icwsm.v12i1.15039. number: 1.

Solovev, K., & Pröllochs, N. (2022, April). Hate speech in the political discourse on social media: Disparities across parties, gender, and ethnicity. In Proceedings of the ACM web conference 2022 (pp. 3656-3661).

Stano, S., 2020. The Internet and the Spread of Conspiracy Content1, in: Routledge Handbook of Conspiracy Theories. Routledge. Num Pages: 14.

Tai, Y.C., Buma, R., Desmarais, B.A., 2023. Official yet questionable: examining misinformation in U.S. state legislators' tweets. Journal of Information Technology & Politics 0, 1–13. URL: https://doi.org/10.1080/19331681.2023. 2257682, doi:10.1080/19331681.2023.2257682. publisher: Routledge eprint: https://doi.org/10.1080/19331681.2023.2257682.

Theocharis, Y., Lowe, W., Van Deth, J.W., Garc´ıa-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, 202–220. Publisher: Taylor & Francis.

Uscinski, J.E., Parent, J.M., 2014. American conspiracy theories. Oxford University Press.

Uscinski, J., Enders, A. M., Klofstad, C., & Stoler, J. 2022. Cause and effect: On the antecedents and consequences of conspiracy theory beliefs. Current opinion in psychology, 47, 101364.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359, 1146–1151. URL: https://www.science.org/doi/full/10.1126/science. aap9559, doi:10.1126/science.aap9559. publisher: American Association for the Advancement of Science.

Xu, H., Ebner, S., Yarmohammadi, M., White, A.S., Van Durme, B., Murray, K., 2021. Gradual Fine-Tuning for Low-Resource Domain Adaptation, in: Ben-David, E., Cohen, S., McDonald, R., Plank, B., Reichart, R., Rotman, G., Ziser, Y. (Eds.), Proceedings of the Second Workshop on Domain Adaptation for NLP, Association for Computational Linguistics, Kyiv, Ukraine. pp. 214–221.URL: https://aclanthology.org/2021.adaptnlp-1.22.

Yasseri, T., Margetts, H., John, P., Hale, S., 2016. Political turbulence: How social media shape collective action. Princeton University Press.

Zhang, Y., Wang, L., Zhu, J.J.H., Wang, X., 2021. Conspiracy vs science: A large-scale analysis of online discussion cascades. World Wide Web 24, 585–606. URL: https://doi.org/10.1007/s11280-021-00862-x, doi:10.1007/ s11280-021-00862-x.