

THE UNIVERSITY OF CHICAGO

No Vaccine for Me:  
Investigating How Misinformation Targeting  
COVID-19 Vaccines Went Viral on Twitter

By  
Zhihan Xiong  
June 2022

A paper submitted in partial fulfillment of the requirements for the  
Master of Arts degree in the Master of Arts in Computational Social  
Science

Faculty Advisor: Sabrina Nardin

Preceptor: Sanja Miklin

## ABSTRACT

A major action to end the global pandemic of COVID-19 is the vaccine-rollout programs. However, with pervasive online misinformation targeting the vaccine, it has been difficult to reach the vaccination level required to form herd immunity. The wide spread of such misinformation can find its roots in the homophily and polarization of social media platforms, the long-lasting anti-vaccine sentiments, and the uniquely politicized narrative on pandemic in the U.S. Using an established dataset with tweets annotated with stances towards misinformation, the current research compared the features of post narrative and of their authors using natural language processing and network analysis techniques. The main findings included (1) people adopted vaccine misinformation used a less coherent but more provoking narrative, and attached more negative sentiment and concepts to the vaccines; (2) people adopted vaccine misinformation showed generally lower influence on social media; they relied more on fake news websites and followed more people holding hostile attitudes toward the vaccine, which formed newsfeeds with significantly more unreliable, anti-vaccine information that impacted their attitudes and behaviors. This research fostered the understanding of online discussion concerning vaccine misinformation, and may help platforms and public health experts to perform interventions at the right time and on the right targets when addressing vaccine hesitancy.

*Keywords:* COVID-19 vaccine, misinformation, social media, natural language processing, network analysis

# 1 Introduction

The past two years have witnessed how a global pandemic has changed the world dramatically. COVID-19 was declared by World Health Organization (WHO) as a pandemic on 11 March 2020 (MacDonald, 2015), and has caused an unprecedented public health and economic crisis globally. A major undertaking to end the pandemic is the development and distribution of vaccines. By the end of 2020, several candidate vaccines were proved safe and effective at generating immune responses, some concluded phase III study with efficacies reached 95% (Pfizer, 2020), and authorized for emergency use by World Health Organization.

To reach herd immunity through vaccination, an approximate 55% - 85% vaccine rate is required. However, reaching this vaccination level may be hindered by the long existing vaccine hesitancy (Schmidt et al., 2018), which has been fueled by a robust and widespread “infodemic” that comes almost synchronously with the pandemic (Loomba et al., 2021). For instance, a video that had been viewed more than 225,000 times on Facebook claimed that the COVID-19 vaccines contained microchips that can track people’s location. Another message widely shared on several social media platforms claimed that Pfizer admitted containing “genetic manipulation” in their vaccines which may lead to birth defects (Siwakoti et al., 2021). These groundless and misleading misinformation targeting COVID-19 vaccines focused on the necessity, safeness and importance of vaccines, and has evoked mistrust towards the vaccines among the general public.

Research on misinformation have proved it to be impactful, as misinformation on social media generally spreads faster than factual information (Vosoughi, Roy, & Aral, 2018), would causes information avoidance and less systematic processing of COVID-19 information (Islam et al., 2020), could result in long-term change in attitudes and behavior even after a brief exposure (Zhu et al., 2012), and the misinformation embedded into long-term memory has been proved resistant to future correction – fact check messages from authorities may even cause a back-fire effect and reduce people’s belief in other factual information (Ayers & Reder, 1998; Carey et al., 2020). Taken together, a successful COVID-19 vaccination rollout campaign requires understanding and addressing the spread of relevant misinformation.

The notion of misleading information overlaps with those of misinformation (information that is unintentionally false or misleading) and disinformation (information that is purposely deceiving). The present research would focus on misleading content regardless of their original intention, and would use the terms interchangeably with misinformation here mainly refers to false or misleading information which is “considered incorrect based on the best available evidence from relevant experts at the time” (Lazer et al., 2018).

The following section consists of two parts. The first three sub-sections summarize key factors for the transmission of online misinformation targeting the vaccine: the nature of social media platforms, the long-lasting anti-vaccine sentiments, and factors specific to COVID-19 and its spread in the United States. In the following sub-section, I briefly review

other relevant work on the subject and articulate the goal of the present study.

## **1.1 Social Media and Misinformation**

A significant volume of misleading information circulates on social media platforms. Such phenomenon has become so pervasive that the digital misinformation has been listed by the World Economic Forum as a main threat to our society (Howell, 2013). Social media has profoundly changed the way information is received and transmitted by the public. Research have shown that about 65% of Americans reporting that they now rely on social media as their source of news (Gottfried & Shearer, 2019). However, social media have also become important enablers and key conduits for fake news and misleading information, as it has provided an accessible medium for publishing and monetizing fake news (Guess, Nyhan, & Reigler, 2018). The ease, speed, and scalability of information dissemination on social media also reduces many of the journalism norms that oversight and constraint news spread (Lazer et al., 2018), thus allows for the rapid dissemination of groundless rumors and conspiracy theories (Quattrocio et al., 2014). In fact, Vosoughi et al. (2018) tracked 126,000 rumors spread on Twitter and found that diffusion of false information diffused had significantly higher range, speed, and depth than that of true information.

Researchers have posited that the content of misinformation is often disruptive and elicits high-arousal negative emotions (e.g., fear, anger, disgust), and such emotions are relevant to sharing behavior (Berger & Milkman, 2012). The structural features of online social media further fuel the pervasive spread of misinformation. Online information exchange networks are characterized by social homophily (i.e. the extent to which social

ties between similar people occur more often than chance alone would predict; McPherson et al., 2001) and polarization (i.e., extremely limited communication between ideologically-opposed clusters; Conover et al., 2011). Social media users' information exposure consists of news feeds shared by their self-picked friends, self-picked followees, and algorithmically ranked content which often reinforce personal preference. This biased exposure of information could lead to the formation of homogeneous, segregated and polarized *echo chamber*, i.e., social bubbles or clusters with low content diversity and strong social reinforcement. Majority of contents circulate only inside these echo chambers (Del Vicario et al., 2016). Users in such echo chambers gradually learn a shared meaning system for information process, and form a “collective framing of narratives that are often biased toward self-confirmation” (Bessi et al., 2015). Previous work has offered evidence of echo chamber around general anti-vaccine information on real-world social media platforms. That is, the anti-vaccine or vaccine-hesitant community showed a clear separation from the mainstream media community (Schmidt et al., 2018; Getman et al., 2018). It's reasonable to expect similar echo chamber around COVID-19 vaccines, which may lead to selective exposure to misinformation for certain groups of people, who would in turn be less willing to get vaccinated.

To address the pervasive misinformation on social media, the platforms have been actively deplatformed misleading content and suspended spreaders of such information (Pasquetto et al., 2020). The common practices include algorithm-driven detection and crowd-sourcing. For instance, Facebook allows user to flag unsubstantiated content so as

to improve their newsfeed algorithms. Third-party fact-checking outlets, including networks such as PolitiFact, Factcheck.org, Snopes, and Lead Stories, have also made great contribution on counteracting the trend of misinformation (Siwakoti et al., 2021). However, both methods suffer from some drawbacks. Algorithm-driven solutions executed by social media platforms raises fear that the removal of content and deplatform of users may threat free speech, especially when the proposed algorithms may not be 100% accurate (Bessi et al., 2015). Fact-checking articles, on the other hand, have an average of 13-hour lag compared to the misinformation, and the distribution of factual content generally involves fewer active users (Shao et al., 2016). Thus, better understanding of the discourse (e.g., help detect content spreading or supporting misinformation) and dynamics (e.g., help detect echo chambers around vaccine hesitancy) of misinformation spreading on social media platforms may help platforms and public health experts to perform interventions at the right time and on the right targets.

## **1.2 Vaccine Hesitancy and Anti-Vaccine Sentiments**

Vaccine hesitancy is a long-lasting, global phenomenon with well-documented evidence (De Figueiredo et al., 2020). Ever since the first vaccine was introduced in the 18<sup>th</sup> Century, fear and mistrust has accompanied the introduction of almost every new vaccine, leading to crippling results such as pandemic outbreaks and needless deaths (Poland & Jacobson, 2011).

The WHO SAGE Working Group on Vaccine Hesitancy concluded that vaccine hesitancy in general refers to “delay in acceptance or refusal of vaccination despite

availability of vaccination services” (MacDonald, 2015). Such hesitancy is often a combination of beliefs, attitudes and behavior. The SAGE Working Group also proposed a 3 Cs model for common vaccine hesitancy determinants: complacency, convenience, and confidence. *Confidence* refers to one’s trust in the effectiveness and safety of vaccines, in the motivation of governments that roll out the vaccine, and in the manufactures. *Complacency* refers to the circumstances when the perceived risks of a certain disease are low and the perceived necessity of the corresponding vaccines would also decrease. *Convenience* refers to the geographical, financial and temporal accessibility of the vaccine, and the quality of the vaccination service (real and/or perceived). The model was widely adopted and supported by research evidence. For instance, a semantic examination of vaccine-related news articles circulated on social media revealed that the prevalence of negative vaccine sentiment was framed around *confidence* issues: “skepticism and distrust of government organizations that communicate scientific evidence supporting positive vaccine benefits” (Kang et al., 2017).

Misinformation targeting vaccines generally evokes mistrust of government and manufacturers, conspiratorial thinking, and denial of actual risks, which are common reasoning flaws in the vaccine resistance (Poland & Jacobson, 2011). Similar mistrust fueled by misinformation and conspiracy theories targeting COVID-19 vaccines has also emerged. Some fake stories target *confidence* of general public by accusing WHO and governments for spreading a “fear pandemic” that benefits large pharmaceuticals, or by claiming that COVID-19 vaccines developed by Bill Gates contains GPS chips (Shapiro et

al., 2020). Other target *complacency* by asserting infection-acquired immunity offers better protection than vaccination (Weinzierl & Harabgiu, 2022).

### **1.3 Factors Specific to COVID-19 in U.S.**

The difficulty of combatting misinformation targeting the COVID-19 vaccines also lies in the complexity of this medical and public policy issue. The rapidly evolution of knowledge on this global pandemic and on the vaccines further exacerbate the situation, as it's hard for government officials, epidemiologists and public health experts to be completely accurate, consistent, and factual (Hossain et al., 2020).

Siwakoti et al. (2021) conducted a wide range of regional analysis on misinformation spread around during the outbreaks of COVID-19, and reported a great deal of heterogeneity. That is, the discourse and common themes of misinformation varied significantly across regions and countries. Researchers concluded that misinformation around COVID-19 in the U.S. was dominated by anti-mask and anti-vaccine narratives, and can be characterized by the extreme political polarization as the pandemic had “rapidly devolved into a partisan issue” in the U.S. The uniquely political feature is evident from several aspects: firstly, conspiracy theories around COVID-19 normally target democratic figures or suggest that policies about the pandemic (e.g., mask mandate, vaccine rollout campaign) were election ploys (Peretti-Watel et al., 2020); secondly, a significant share of misinformation that evoked anti-vaccination sentiments or promoted other false information were perpetuated by political figures. For instance, misinformation frequently stemmed from former President Trump, who once claimed hydroxychloroquine as an

effective cure for COVID-19, and had flouted public health protocols including masks.

Since a large proportion of Americans now seek health information from sources outside the formal medical system, and are increasingly relying on online sources to make health-related decisions (Jacobs et al., 2017), the impact of misinformation on social media platforms is not confined to mere online discussion. Research has established the causal relation between exposure to online misinformation targeting COVID-19 vaccines and decreased intent to vaccinate in order to protect oneself and others (Loomba et al., 2021). Besides confusion and even deadly consequences in health problems, vaccine rejection on large scale could also hinder the accomplishment of herd immunity. It's estimated that to provide herd immunity, the COVID-19 vaccines would need a vaccination level by 55% - 85%, while public survey showed only 34.2% of the public in the U.S. were willing to accept a vaccine in June 2020, and the rate fell significantly in October among all demographic groups (Kwok et al., 2020). This may be attributable to misinformation reached a spike in early October, 2020 (Siwakoti et al., 2021).

In sum, the pervasive online misinformation targeting COVID-19 vaccines stemmed from both the long-lasting anti-vaccine sentiment and the systematic features of social media platforms, and was exacerbated by the uniquely politicized environment in the U.S. The wide spread of such misinformation made it difficult to reach the vaccination levels required for herd immunity which may cause serious public health outcomes. To advocate for the uptake of vaccines brings up pressing needs to understand combat misinformation, and to communicate health information to the general public in a more acceptable manner.

#### **1.4 Contribution of the Present Work**

As discussed above, a vein of research has proved the impact of vaccine misinformation on individuals' intent to vaccinate and general vaccination level through surveys (Islam et al., 2020) and experiments (Loomba et al., 2021). The goal of this research is to offer further understanding of the discourse around misinformation target COVID-19 vaccines, and of people who accepted such misinformation. This would enable public health experts to detect target audience and perform interventions in a more timely and precise manner, which may help address vaccine hesitancy.

In real-world social media, exposure to misinformation does not necessarily lead to acceptance of the opinion, and such misinformation may have varied impact on individuals from different sociodemographic groups. Consequently, in addition to the general effect of exposure to such misinformation, the present research aimed to further investigate the narratives of people holding different stances towards the vaccine misinformation, and what features could best characterize people accepting such misinformation.

Previous researchers have applied both qualitative and quantitative methods to analyze the misinformation in the pandemic era. The first line of research focus on the cataloging misleading information through qualitative encoding. For instance, the Empirical Studies of Conflict Project (ESOC) at Princeton University collected 5613 distinct misinformation stories around the world, manually coded the narratives, motivations and keywords for each story, and conducted cross-regional comparisons on the evolution and debunking of misinformation (Siwakoti et al., 2021). Researchers from computational fields, on the other

hand, have focused on detection and annotation of misinformation on social media platforms on large scale. Researchers retrieved news articles and social media posts related to vaccine misinformation, and deployed state-of-the art algorithms such as knowledge embedding models, Language-informed Attitude Consistency-preserving Relation scoring (LACRscore) to classify whether an article/post is trustworthy (Cui & Lee, 2020), or to annotate the stances of the text towards certain misinformation (Hossain et al., 2020; Weinzierl & Harabgiu, 2022).

The current study combines qualitative analysis with computational linguistic techniques and compare the narratives of social media posts with different stances towards the vaccine misinformation. Such methodology can be applied to large-scale dataset and at the same time offer intuitive understanding of the mindset or reasoning beneath the acceptance of misinformation.

Which sociodemographic groups are specifically susceptible to misleading information has received great attention in the study of combatting misinformation. For instance, Grinberg et al. (2019) found that engagement with fake news on presidential election was extremely centralized on Twitter: there were super-consumers (approximately 1% of individuals were exposed to 80% of such fake news), and super-sharers (less than 0.1% of accounts in the panel shared 80% of the fake news). They also reported the people with highest fake news sharing behavior were conservative leaning, older adults. Other research showed similar results, that the people over 65 were more likely to share political fake news on Facebook during the 2016 presidential election (Guess, Nyhan, & Reigler,

2018). However, most of these works focus on political misinformation, and whether these findings were generalizable to public health was unclear. Also, few researches focused on the features of users' social networks, which worth investigation as according to the Social Contagion Theory, individuals tend to adopt the behavior and attitudes of members within their social network (Christakis & Fowler, 2013).

Overall, this research aims to address how misinformation targeting COVID-19 vaccines has been discussed and spread on social media through answering the following research questions: (1) Based on content analysis, what are the distinguishing semantic and syntactic features of tweets that support and spread misinformation, especially when compared with those holding opposite stances (reject or even critique such misinformation)? (2) Through analyzing how users had been influenced and influence others, how misinformation targeting COVID vaccines get in and spread around online social network? What are the characteristics that could best distinguish the newsfeed of users who adopt misinformation related to COVID19 vaccines?

## 2 Dataset

To track the discussion on and the spread of misinformation targeting COVID-19 vaccines in Twitter, the present research is performed on an established dataset named CoVaxLies2, a dataset that contains not only tweets reacting to misinformation, but also annotation of users' stances (i.e., attitude users manifest) towards the misinformation. The CoVaxLies2 dataset was constructed by Weinzierl and Harabagiu (2022) from streaming

Twitter discourse during the pandemic.

The researchers first assembled a set of 47 misinformation targets (MisTs; see Table 1 for examples) questioning the necessity, safeness and importance of COVID-19 vaccines from (a) misinformation identified by the specific Wikipedia page concerning misinformation of vaccines against the SARS-CoV-2 virus, and by clinical and academic organizations, and (b) questions from the Vaccine Confidence Repository (Rosssen et al., 2019). Based on MisTs collected, they then retrieved tweets responding to the target misinformation by processing these MisTs as queries sent to Twitter API. Only tweets in English were collected. Subsequently, Weinzierl and Harabagiu (2022) asked Natural Language Processing (NLP) experts from the Human Language Technology Research Institute at the University of Texas at Dallas to judge whether the collected tweets are relevant to assembled MisTs (some tweets may be reacting to more than one MisT), and to annotate the stances of these tweets (i.e., whether the user accepted or rejected the corresponding MisTs).

Weinzierl and Harabagiu (2022) also categorized the MisTs into common themes based on discussion with public health experts. Concerns expressed in the misinformation were abstracted into nine themes (see Table 1 for a brief summary and a sample MisT for each theme). Each theme characterizes an aspect that impacts confidence or complacency in the COVID-19 vaccine, which is consistent with the 3 Cs model (MacDonald, 2015).

**Table 1.** Taxonomy of misinformation targeting COVID-19 vaccine and proportion of each theme in retrievable tweets

<i>Theme</i>	<i>Summary</i>	<i>Sample MisT</i>	<i>Proportion</i>
1	Unsafe COVID-19 Vaccine	The vaccine is a bioweapon.	17.95%
2	Adverse Events of COVID-19 Vaccine	Vaccine interacts with DNA.	17.50%
3	COVID-19 Vaccine Ingredients	The vaccines is gene therapy.	16.85%
4	Unnecessary COVID-19 Vaccine	We have natural immunity.	11.20%
5	Not effective COVID-19 Vaccine	Worse than natural immunity.	9.93%
6	Effect on Immune System	Overrides the immune system.	9.88%
7	Testing of COVID-19 Vaccine	No long-term research.	8.65%
8	Information about the Vaccine is concealed	Large pharmaceuticals conceal information about reinfections.	4.39%
9	Alternatives to COVID-19 Vaccine	Vitamins are better alternatives.	3.65%

Due to the limitation set by IRB and Twitter API policy, researchers only provided the IDs of the collected tweets and the original content need to be downloaded through the Twitter API. Therefore, I scraped the content from Twitter using the provided IDs. However, aiming to create a better and healthier environment of public conversation, Twitter actively detects and takes actions toward tweets containing misleading content (limit amplification, label or remove content). Tweets in this dataset that concern misinformation are susceptible to these actions by their nature, and consequently not all annotated tweets were retrievable

by the time of the present research. Furthermore, the original dataset contains 9296 tweets annotated with four stances (Accept, Reject, No Stance, Not Relevant). To better compare people who showed faith with those who weren't convinced, I filtered out tweets without salient stances and divide the remaining 5365 tweets into Accept and Reject group, each contains 3355 and 2010 tweets, respectively. In sum, I successfully retrieved 3827 tweets (71.33% of the original), 2179 tweets accept one or more MisTs (64.95%) and 1648 tweets were annotated as rejection (81.99%). Such disproportional retrievability is intuitive, since tweets adopting MisTs are more likely to be flagged by Twitter as harmful and get removed. The retrievable tweets were authored in the time frame from December 18th, 2020, to June 30th, 2021.

To further ensure the validity of the dataset, I randomly sampled 5% (N = 195) tweets after basic data cleaning, and examine their assigned MisTs and annotated stances. All tweets were correctly annotated to my best knowledge.

### 3 Content Analysis

With the two manually constructed corpuses consist of tweets post by people who accepted (N = 2179) and rejected (N = 1648) misinformation targeting COVID-19 vaccine, it's possible to compare the linguistic features using natural language processing techniques, e.g., to quantify and analyze the presence, meaning and connections of certain words, concepts and topics, and to ultimately investigate perception of COVID-19 vaccines and the pandemic from people with different stances.

Common psychological features among antivaccinationists including strong mistrust towards government and vaccine manufacturers, belief in conspiracy theories, low cognitive complexity in thinking patterns, and reliance on emotional anecdotes instead of scientific data (Pennebaker et al., 2003). People are more likely to convict to conspiracy theories when they have strong need to maintain a positive image of the self and the group they belong to, such need could derive from a feeling of under threat (Douglas et al., 2019). Indeed, a common story structure applied by disinformation or conspiracy theory is a straightforward “threat narrative” (Shahsavari et al., 2020; Lazic & Zezelj, 2021), which comprised of a hostile orientation (e.g., large pharmaceuticals, despotic governments) who perform purposefully threatening action (e.g., to manipulate the public or to amass profit), a proposed strategy against the threat (e.g., protests and anti-vaccination) and the outcome of applying that strategy (e.g., healthy, free from surveillance). My hypothesis is that tweets from the Accept group would show stronger negative emotion, less language complexity (due to the lower cognitive complexity), stronger identification to a community, and attach more negative concepts to vaccines and the orientation in the threat narrative (governments, scientists, vaccine manufacturers).

Specifically, I investigated the psychological features and perceptions through analyzing syntactic and semantic characteristics of the two corpuses. I applied the following computational linguistics approaches: (1) frequency count of unigrams and bigrams (i.e., a sequence of two adjacent elements from a string of tokens) and Part-of-Speech tagging, to obtain understandings of the language use through analyzing the

presence of certain linguistic elements; (2) word embedding applying Word2Vec model, a technique that maps words to vectors that can capture distance and similarities between words and concepts, and to embed core concepts in their local linguistic contexts; and (3) sentiment analysis.

Before any further analysis, I cleaned the tweets data by removing unnecessary and uninformative content to highlight only key attributes suitable for NLP models. Text data normally contains punctuations, high frequency words of low relevance and information (i.e., stop words). Tweets, specifically, also contain other elements including URLs, hashtags and emojis. The cleaned text was then tokenized and lemmatized. Tokenization refers to the process of splitting sentences into list of words, while lemmatization is the process of removing inflectional endings and returning the base or dictionary form of a word (“lemma”) through the use of a complete vocabulary and morphological analysis of words.

### **3.1 Word Frequency and POS Tagging**

#### *Methods*

To investigate whether the language use differs between the Accept and Reject group, I analyzed computational linguistic features using techniques including word counting and Part of Speech (POS) tagging. What role does a word plays in the sentence, known as its POS, is a critical feature of text. The Spacy library offer a powerful tool for tagging POS for lists of words in English. The results can be aggregated and used to calculate linguistic features. Besides the most common words under each part of speech, word counting also

enables other types of analysis in language use: the syntactic features of sentences. For instance, the systematically disproportional use of nouns and verbs may indicate whether the text generators focus more on objects or on actions (Choi, 2000), the use of self-referential pronouns (*I, me, my*) which shows the attention of text, and the frequent use of conjunctions (*and, however, but*) indicates the effort of maintaining a coherent narrative (Graesser et al., 2004).

To measure the linguistic facet of each group, I used the Linguistic Inquiry and Word Count (LIWC) toolkit (Tausczik & Pennebaker, 2010). LIWC is a versatile toolkit for text analysis on linguistic and psychological facets. It contains well-defined, scientifically validated dictionaries for many categories of words including pronouns, verbs and emotions, and works well even with short content such as tweets (Pennebaker et al., 2013). For instance, the self-referential pronoun category contains words associated with “I”. Given a corpus, LIWC can calculate the frequency for words of interest and compute the relative percentage levels for specific categories.

### *Results*

Word clouds in Figure 1 offers an intuitive understanding of the commonly used words in each group of tweets and higher frequency is represented by larger font size. To better illustrate the features of each group, I filtered out the tokens that appear in both groups.



frequency (see Table 2 for the most common words in four categories). As for most commonly adjectives, nouns, verbs, Accept and Reject group tweets showed great overlap. When examine for most frequent bigrams (two-word collocations), however, 1/3 of the top collocations were different. Interestingly, the frequent bigrams unique to the Reject group includes “Alter DNA”, “Bell’s Palsy” and “Bill Gates”, which are relevant to concerns common among misinformation targeting COVID-19 vaccine. For instance, many misinformation circulating on social media spread conspiracy theory that Bill Gates will implant microchips to track people who have been vaccinated, or that the vaccine causes Bell’s palsy, a rumor that echoes the long-lasting vaccination hesitancy. However, the contexts (i.e., words commonly appear before and after the keyword) of these words are different. The most common action attached to “Bill Gates” in Accept group is “admits”, mostly followed by “COVID vaccine is unsafe / alters DNA / will kill”. As in Reject group, “Bill Gates” is mostly followed “is/does”, and then a corrective information such as “is/does not planning to microchip the world / not have control over all vaccines”.

Besides the most frequent words, I also compared the proportion of each POS in the corpus (relative to the total word count), and the use of inner- and outer-group pronouns (relative to all pronouns) between the two groups. Since the sample size is large and population variance is unknown, I used  $z$  test for mean comparison.

The two groups showed difference in percentage of verbs ( $z = 6.4413, p < .001$ ), and appeared similar on Adjectives and Nouns. This result indicates that the Accept group focus less on actions. Similarly, the Accept group showed a significantly lower use of

conjunctions ( $z = 4.1986, p < .001$ ), indicating a more impulsive, less-organized narrative.

**Table 2.** Most common bigrams / words in each POS for Accept and Reject groups (words unique to the group underlined)

<i>POS</i>	<i>Group</i>	<i>Most common words/bigrams unique to each group</i>
Adjective	Accept	Experimental, Positive
	Reject	Side, Genetic
Noun	Accept	Therapy
	Reject	Infection
Verb	Accept	Do, Know, Die
	Reject	Contain, Alter, Fight
Bigram	Accept	Natural Immunity; Gene therapy
	Reject	Alter DNA; Bell Palsy; Bill Gates

As for the use pronouns, I compared the proportion of the inner-group pronouns (*I, me, we, us*) and outer-group pronouns (*they, them*) in each group. Interestingly, although the two groups had little difference on the use of inner-group pronouns, the Accept group showed significantly higher use on the outer-group pronouns ( $z = -6.5192, p < .001$ ). The activation of social identities as a group member would evoke negative emotions towards outer-group members (Coleman & Williams, 2013). The disproportionally use of such pronouns can be attributable to belief in conspiracy theories in which there is always

a hostile exterior group plotting against the group members, the threaten can come from government, the Democratic party, mainstream news websites, and capitalists such as large pharmacies or Bill Gates. For instance, a popular fake news source acts against COVID-19 vaccine is named ‘Children’s’ Health Defense” with the mission to protect children from “health epidemics by working aggressively”.

### **3.2 Word Embedding**

#### *Method*

While word counting calculates linguistic features is based on tokenized and isolated words, other NLP techniques enable analysis based on the contexts of these words. The way to capture the trends within the corpus is *word embedding*, a technique to map each word to a vector in which semantically similar words have vectors with lower cosine distance. Although both try to analyze words in a certain context, word embedding is different from topic modeling, as the latter relies mainly on term cooccurrence in an unordered bag-of-word model, while word embedding requires to retain the sentence structure or a certain word window within the texts. The process of word embedding is to take a developed corpus, specify an underlying number of dimensions, and to train a model with a neural network auto encoder. In the present study, I would apply the Gensim implementation of *Word2Vec* algorithm that embeds or describes objects of interest in their local linguistic contexts, so as to learn about the discursive culture that produced these concepts.

The Gensim implementation of Word2Vec enables both model architectures for

computing continuous vector representations of words proposed by Mikolov et al. (2013) that is Continuous Bag-of-Words (CBOW) model and Continuous Skip-gram Model. According to Mikolov et al. (2013), CBOW is trained to predict a single word from context of a certain window size, while Skip-gram predicts the contexts of a single word and allows different contexts for homonyms. Although the task of CBOW is much simpler and leads to a faster convergence of the model, Skip-gram has better performance in capturing semantic relationships even for words that are morphologically different. Additionally, as Skip-gram relies on single words input, it is less susceptible to overfitting frequent words, while CBOW is prone to this flaw. Based on its higher efficiency on documents and better performance on capturing semantical similarities, the present study would use the Skip-gram model. Another parameter for Gensim implementation of Word2Vec is the size of the vector used to represent a single word (i.e., the size of the word window or context). I would tune this parameter based on the outcome of similarities scores between the words “COVID” and “Covid-19” as these two refer to the same disease.

Words embedding also enables the identification of semantic change. The process is to align the dimensions of multiple embeddings arrayed across time, and to identify the change of loadings within word vectors for focal words (Hamilton et al., 2016). In sum, word embedding techniques works better in this situation as there are specific objects of interests. By using an object as an anchor, inspecting its context in different corpuses, we can have better understanding on attitudes of the two groups towards both the pandemic and the vaccine defending against it.

## *Results*

I trained two separate word2vec models for tweets in Accept and Reject group using the *Gensim* library. As mentioned before, I chose the Skip-gram model and tuned the size parameter separately, and both models showed the best performance when it was set to 150. Inside the word2vec objects, each word has a unique vector. And similarities between words can be obtained by calculating the cosine distances between the corresponding vectors. As the current study focuses on the narrative around COVID-19 and the vaccines, I used “COVID”, “vaccine”, and “vax” (the most common contraction for vaccine) as the keywords. Table 2 shows words that are most similar with the keywords, with words unique to each group underlined. Here are some interesting findings:

Firstly, the Accept group tweets related both terms for vaccine to “infection” while the same pattern does not appear in the Reject group. When take a closer look at tweets with sentence(s) containing both words, the tweets in the “Accept group” were essentially comparing the immunity induced by vaccine and infection. A representative claim read, “Natural infection in healthy people 50 and under are better off than getting a vaccine. Natural immunity lasts 17 plus years longer. These new vaccines don’t have anywhere near that.” Some asserted that the COVID-19 vaccine is a gene therapy rather than infection prevention. The comparison between natural immune system with the efficiency of vaccine is among the nine common topics of misinformation. In short, while the people rejecting misinformation find vaccine a reliable source of immunity, people accepted misleading

information found vaccine comparable with infection induced immunity: both are safe but the latter is superior for its long-lasting effect. Also, while people from Reject group shows higher similarities between COVID and fatal outcomes, the Accept group shows no similar pattern. Such lower perception of risk is in line with the *Complacency* component in 3 Cs model (MacDonald, 2015).

**Table 3.** Most similar words for Accept and Reject groups (words unique to the group underlined)

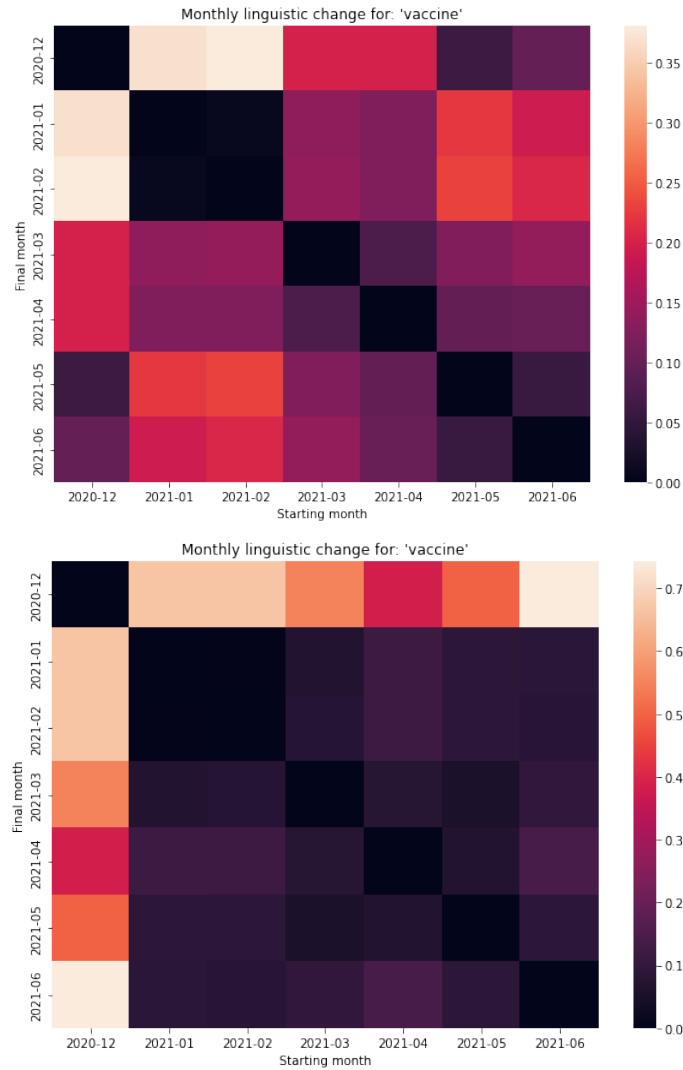
<i>Keyword</i>	<i>Group</i>	<i>Most similar words</i>
COVID	Accept	<u>Effect, Possible, Concerned, Government, Worried, Having, Unknown, People, Aware</u>
	Reject	<u>Die, Dying, Want, Vaccine, Higher, Injuries, Death, Greater, Think</u>
Vaccine	Accept	<u>Testing, Getting, Trials, AMP, People, Unsafe, Doctors, Astrazeneca, Government</u>
	Reject	<u>Effect, People, Disease, Know, Actually, Concerned, Possible, Worried, Fear</u>
Vax	Accept	<u>Cheap, Available, Evidence, Injuries, Age, Anti, Government, Time, Death</u>
	Reject	<u>Science, Developing, Far, Age, Technology, Time, Telling, Doses, Created</u>

Secondly, tweets in the Accept group related both COVID and Vaccines to the government, and relate the vaccine to Astrazeneca (one of the vaccine manufacturers) while the Reject group did not. People related their discontentment on vaccine requirements with their earlier discontentment on limitations and inconvenience brought by COVID, and reproved that both the pandemic and the vaccine are “just another political money scam” and government control.

Additionally, while tweets in the Reject group linked mostly neutral (e.g., time, people) and positive (e.g., technology, science) words to the COVID-19 vaccine, people in the Accept group attached the vaccine with negative outcome including *death*, *unsafe*, and *Injuries*. In sum, word embedding enables understanding of key concepts in their local contexts, and revealed the mental model of COVID-19 vaccines for people holding different stances towards misinformation, which is in line with common misinformation targeting the vaccine.

Beside the qualitative analysis on the meaning of key concepts, word embedding also enables quantitative investigation on the semantic *change* of focal words. The changes were calculated using  $1 - \text{cosine similarity}$ , with lower value equivalent to smaller change. Specifically, I investigated whether the semantic meaning of *vaccines*, the core concept of this research, had changed across time. Figure 2 shows heatmaps on monthly linguistic changes for the word 'vaccine' in two corpora respectively. The word vaccine remained stable in the timeframe included for both corpora. For Accept group, the change on the semantic meaning of this word was only salient between Dec 2020 and the first two months in 2021. As for Reject group, the semantic meaning of vaccine in Dec 2020 differs changes drastically, but remained almost intact afterwards. Such turning point is reasonable. As summarized by a timeline of COVID-19 constructed by the American Journal of Managed Care (AJMC), CDC approved urgent use of several major COVID-19 vaccines in the first quarter of 2021 (AJMC, 2021), and the actual roll out of vaccines may lead to semantic changes. And if 2020 was dominated by how the pandemic spread across the world, 2021

seemed to focused more on manufacturing and distributing vaccines, getting people vaccinated and ultimately ending the pandemic.



**Figure 2** Monthly linguistic change for 'Vaccine' in tweets for Accept (top) and Reject (bottom) group. The lighter color represents larger change.

### 3.3 Sentiment Analysis

#### *Method*

Another valuable feature of text data is sentiment. Sentiment analysis is a text analysis method that quantify the subjective attitudes, stances and emotions of the speaker/writer.

To perform sentiment analysis on sets of tweets, I used the Valence Aware Dictionary and sEntiment Reasoner (VADER), a lexicon and parsimonious rule-based model for sentiment analysis developed by Hutto and Gilbert (2014) which is specifically attuned for social media text. Being a pre-trained model, it can be applied directly to unlabeled data. Texts from social media platforms may contain non-text data including symbols, multiple punctuations (“!!!”) or emojis, VADER can handle them in its lexicon, and it’s also sensitive to unique sentiment expressions on social media platforms. As suggested by its name, VADER calculates the sentiment score of a text by mapping lexical features to emotion intensities pre-settled in a dictionary, and summing up the intensity of all words. VADER can measure the polarity (positive, negative or neutral) and intensity (level or strength) of emotions separately.

### *Results*

VADER assigned each pre-processed tweet with a compound sentiment score and a polarity classification (see Table 4 for results of the two groups). As polarity type (positive, negative, and neutral) is a discrete variable, I used Chi square test to investigate whether there is difference in the distribution of polarity types between the two groups of tweets, and the results indicate no significant difference ( $\chi^2 = 4.44, p > .05$ ). As for comparing the mean compound sentiment score, I applied Z test and the Accept group showed significantly lower mean compound score ( $z = 2.64, p < .01$ ), indicating a general leaning towards negative emotional expressions. These results are in line with findings from related work. For instance, studying social media engagement around

COVID-related news, Cui and Lee (2020) showed that tweets related to fake news tend to be more negative and have stronger sentiment polarities.

**Table 4.** Number of tweets in each class and average sentiment score for two groups

<i>Group</i>	<i>Negative</i>	<i>Positive</i>	<i>Neutral</i>	<i>Total #</i>	<i>Average Score</i>
Accept	962(44.15%)	952(43.69%)	265(12.16%)	2179	-0.0026**
Reject	759(46.06%)	668(40.53%)	221(13.41%)	1648	0.0390

## 4 Information Network and Stances

Besides the content and topic of the tweets, it's also informative to examine who were tweeting and why. Specifically, the social contagion among users' social network, i.e., how the misinformation targeting COVID-19 vaccines was discussed and spread within the social networks, also worth attention. According to McPherson et al., (2001), similarity breeds connection, and such homophilous relations in turn have powerful limitation on the information people receive and the attitudes they form, that is, individuals tend to adopt the behavior and attitudes of members within their social network. This tendency is generalizable to health-related decisions: health attitudes and behavior can be transmitted across social network, affecting the health outcomes of the network members (Smith & Christakis, 2008). A scoping review collated evidence from 24 studies and concluded that participants who have been frequently exposed to positive attitudes towards vaccinations “held more positive vaccination attitudes and greater likelihood to get vaccinated or

vaccinate their child” (Konstantinou et al., 2021). It is reasonable to expect similar effect of vaccine-related misinformation. The authors of tweets within the dataset may be influenced by attitudes and behaviors of people in their social network, and may subsequently pass down the attitudes to their followers. Grinberg et al. (2019) estimated Twitter users’ news feed using tweets posted by their followees, as individuals were potentially exposed to these tweets. The present study would follow this definition and investigate users’ newsfeed through investigating their followees and tweets posted by highly visible followees.

In this part, I analyzed the network in which the misinformation was shared and discussed among online social network. This includes analysis on (1) the influence of tweets within the dataset, and the influence of the people posted them, (2) the upstream newsfeed of the users, including the content they shared or quoted within the tweets, and the posts of their followees (referred to as their “friends” by Twitter) which consist of a great portion of users’ newsfeed on Twitter.

#### **4.1 User Influences**

To exclude the outliers of this information network that showed no further influence, only tweets that got at least one favorite or one retweet were included. The influence of the user and of a single tweet was measured separately (see Table 5 for the results).

The influence of tweets was measured by the count of favorites and the count of retweets a post got. Both counts reflect the influence of a post, but their meaning is different. Retweeting a post is a direct sharing behavior and would make the post visible on the

timeline of the user and their followers, although the attitude towards the original post is not necessarily supportive. Clicking favorite of a tweet, on the other hand, shows genuine acceptance to the original post. Although tweets liked by a user would appear on their profile, it wouldn't always appear on users' timelines, making 'like' a less direct measure of influence. See Table 5 for the mean and variance of these measures. I conducted Z tests to compare the means of two groups. In general, the Accept group tweets get significantly less favorites ( $z = 3.67, p < .001$ ), but no statistically significant difference on retweet counts ( $z = 1.87, p > .05$ ). The higher number of favorites indicate that tweets in the Reject group have mildly higher influence.

The influence of users was measured by the number of followers, proportion of verified accounts, and number of friends. Verified accounts would have a blue badge besides the account name, and according to Twitter, verification signals that the account is recognized as "real, credible, authentic and of interest to the public" (Twitter, 2021). As for the number of friends, research has proved that this index is negatively correlated with perceived influence of a user. This phenomenon may be attributable to the fact that people would perceive following fewer others as a prove that the user is "less susceptible to outside influence and thus more autonomous" (Aral & Walker, 2012). As shown in Table 5, there are significantly more verified users in the Reject group ( $\chi^2 = 58.94, p < .001$ ). Users in the Reject group have significantly more followers ( $t = 3.10, p < .01$ ) and marginally less friends ( $t = 1.89, p = .058$ ). The results indicate that tweets in the Reject group got larger exposure.

However, the results should be interpreted with great care. As discussed in the dataset description, only 71.33% of the tweets were retrievable (64.95% for the Accept group tweets) as Twitter actively suspended accounts posting misleading information and removed suspicious posts. It is possible that tweets or users with larger influence would be assigned higher weights and get noticed and censored sooner, which would consequently bias the influence indexes towards a lowering direction.

**Table 5.** Influence indexes for users in the two groups (mean and standard deviation)

<i>Group</i>	<i>Verified</i>	<i>Follower **</i>	<i>Following</i>	<i>Favorite count***</i>	<i>Retweet count</i>
Accept	7.78%	7578.74 (78531)	3594.05 (10472)	22.4 (120.01)	7.11 (23.62)
Reject	27.73%	15717.55 (90335)	2477.52 (6184)	52.44 (251.59)	13.99 (72.95)

## 4.2 Unreliable sources

### *Method*

A direct way to trace back the source to which a tweet is responding is to examine the URLs (if any) it contained. URLs contained in tweets can be inter-platform (i.e., the tweet quotes or refers to another tweet) or external to Twitter when users share content such as articles or videos from other websites. With tweet IDs collected by Weinzierl and Harabagiu (2021), Twitter API enables entity detection for tweets downloaded, and this further allows for retrieving URLs from the tweets. Besides a mediator to retrieve original contents, URLs alone contain useful information such as the original domains, which to

some extent predict the reliability of the content. I retrieved URLs contained in tweets from Accept and Reject groups respectively. Due to the Twitter length limit, many users tend to include shorten URLs. Consequently, I retrieved the expanded URLs through head requests before extracting the domains to which they belong.

Lazer et al. (2018) defined fake news as news articles that have the trappings of legitimately produced news but “lack the news media’s editorial norms and processes for fact-checking” and can be false or misleading. The attribution of “fakeness” is thus not at the level of the story but at that of the publisher. Previous researchers have iteratively constructed and maintained a set of websites taken from preexisting lists of fake news sources and trustworthy mainstream sources constructed by fact-checkers, journalists and academics (Allcott & Gentzkow, 2017; Guess, Nyhan, & Reigler, 2018; Grinberg et al., 2019). Here, unreliable sources are websites that have been identified as publishing almost exclusively fabricated stories. When cross-referenced with the constructed list, it’s possible to judge whether a URL contains misleading information.

### *Results*

In sum, the 2179 tweets in the Accept group contain 1018 unique URLs, among which 704 are from exterior sources and 87 come from confirmed unreliable sources (12.36%). The 1648 tweets in the Reject group, on the contrary, contain 710 unique URLs, among which 551 come from exterior sources and 6 from unreliable domains (1.09%). The rate of sharing links from unreliable sources is significantly higher for the Accept group ( $Z = -8.558, p < .001$ ). The estimated portion of suspicious content is relatively conservative,

as unlisted domains may also contain unverified or misleading information.

**Table 6.** Most common exterior domain of URLs contained in tweets (unreliable sources underlined)

<i>Accept Group</i>	<i>Reject Group</i>
www.youtube.com (40)	www.cdc.gov (25)
<u>www.newsbreak.com (15)</u>	www.reuters.com (14)
<u>www.dailymail.co.uk (11)</u>	www.nytimes.com (13)
www.nytimes.com (10)	www.usatoday.com (10)
<u>rumble.com (8)</u>	www.nhs.uk (9)
<u>www.bitchute.com (8)</u>	www.youtube.com (7)

Table 6 shows the most frequently visited domains in each group. While all top 6 websites shared by people rejected misinformation targeting COVID-19 vaccine were main stream news websites or official sources, people who accepted the misinformation commonly shared contents from websites publishing almost exclusively fake news to confirm or support their belief. Among the four underlined unreliable sources, News Break, Rumble and Bit Chute are unreliable sources identified by researchers, while Daily Mail has long been criticized as inaccurate and “guilty of spreading disinformation with an often right-leaning bias” (Umpathy, 2022).

### **4.3 Social Media Newsfeeds**

#### *Method*

Another possible source of the Twitter users' attitudes towards vaccine / vaccine-related misinformation is the information to which they were exposed, in this case, their upstream newsfeeds on social media platforms. Social media users' information exposure consists of posts shared by their friends, their self-picked followees, and algorithmically ranked content which often reinforce personal preference. A feasible and accurate estimation of one's newsfeed is to obtain the people they follow.

Due to the rate limit of Twitter API (developers are only allowed to request for 15 users' information every 15 minutes), I randomly sampled 645 (17.86%, 406 from the Accept group, 239 from the Reject group) unique users and obtained their profile using the Tweepy library. To detect the upstream newsfeed of these users, I obtained the friends of sampled users and found the accounts followed by most sample users. To further capture and quantify their attitudes towards vaccine, I downloaded the timelines of these accounts which contain at most 3000 tweets and filtered out irrelevant ones, leaving a total 1930 tweets which contain keywords such as "vaccine" or "vax". Subsequently, I conducted general sentiment analysis with VADER.

#### *Results*

I obtained followees of sampled users from Accept and Reject group, respectively, and assembled accounts appeared most frequently, i.e., accounts mostly followed by (and thus highly visible to) tweet author in that specific group. The top 10 most popular accounts of

the Reject group are Democratic politicians (Barack Obama and Joe Biden), official accounts of CDC and mainstream news websites (e.g., New York Times, BBC, CNN). Correspondingly, the top 10 most popular accounts in the Accept group are Republicans and Trump supporters (Elon Musk, and two archives of Trump administration accounts), and some are noted anti-vaccine activist -- Robert Kennedy Jr, for instance, were referred by the CNN as “a leader of the anti-vaccination movement in the U.S.” (Bergen, 2022). Table 7 contained a brief summary of these figures and a sample tweet from their post relative to COVID-19 vaccine. All accounts other than the archive of administration show obvious resistance towards the vaccines, and some overtly promote misleading information (See Table 7 for the Top 10 most popular accounts among the Accept group and a sample tweet from each account), although sentiment analysis using VADER shows a generally neutral sentiment score (Mean = 0.003).

**Table 7.** Accounts that most users in the Accept group follow

<i>Screen name</i>	<i>Summary</i>	<i>Sample tweet related to vaccine</i>
Elon Musk	Entrepreneur, founder of Tesla	-
Tucker Carlson	Host of Fox News	"This has nothing to do with the vaccine. It's about power."
Donald Trump Jr.	Political activist; Child of former President Trump.	"Enough of the tyrannical rules!!! It's time to DEMAND Biden stop dividing Americans with radical vaccine mandates."
Candace Owens	Writer. "Black people don't have to be Democrats – still."	"It will pass– because the FDA essentially works for the vaccine manufacturers through benefiting financially from authorizes their products."
President Trump 45 Archived	Archive of a Trump Administration account.	"From the instant the Coronavirus invaded our shores, we raced into action to develop a safe and effective vaccine..."
Rand Paul	Senator for Kentucky; Republican	"Tell Fauci to keep his hands off our children–no vaccine mandates!"
Dr. Simone Gold	"Doctor,lawyer,warrior"; founder of AFLDS.org	"It's heartbreaking that children have suffered so much under medical tyranny, now they want to inject them with emergency-authorized vaccines? Parents need to say: No."
Jack Posobiec	Host of Human Events Daily; conservative Republican	"I've never taken a single Covid vaccine"
The White House 45 Archived	Archive of a Trump Administration account.	"More good news on vaccines: Novavax on Monday announced it is beginning a Phase 3 trial of its coronavirus vaccine"
Robert Kennedy Jr	Founder of Children's Health Defense	"FAA's COVID vaccine mandate for pilots violates federal regulations + places pilots and passengers at risk."

## 6 Conclusion

In general, the present research investigated the narrative and spread of misinformation target COVID-19 vaccines on Twitter by analyzing corpuses with different stances towards this misinformation. Firstly, with NLP techniques, the present research revealed the distinguishing semantic and syntactic features of narrative of misinformation supporters. Although the language elements had overlap due to the similar topics, people adopted such misinformation tended to use more verbs, less conjunctions, and more outer-group pronouns, indicating a less coherent narrative and a strategy to evoke vaccine resistance using a threat narrative; they also tended to attach 'vaccine' with the concept of infection, government and death, and expressed generally more negative expressions. Additionally, through comparing the influence and newsfeed of users with distinct stances, people in the Accept group had generally lower influence, and received more unverified, anti-vaccine information in their newsfeed judged by the URLs within tweets and people they followed.

Similar with other research investigating misinformation using authentic social media data, this research was limited and potentially biased by the deplatforming operations of Twitter (Pasquetto et al., 2020). Although the original dataset was constructed using streaming API, not all tweets were retrievable by the time of this research and this may lead to a conservative assessment on the influence of the misinformation supporters, and an incomplete corpus. Moreover, the current dataset was collected using certain misinformation targets as queries, consequently, the users and their followers do not

necessarily form a complete social network. Future research can refine the data collecting process to form complete networks, e.g., to include more layers of follower - followee relationship, or to detect and investigate existing, antivaccination-related communities. Complete networks would enable Social Network Analysis (SNA) that can further investigate the dynamics of misinformation spreading around social media platforms.

## REFERENCE

- AJMC. A Timeline of COVID-19 Vaccine Developments in 2021. (2021, June 3).  
<https://www.ajmc.com/view/a-timeline-of-covid-19-vaccine-developments-in-2021>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-36.
- Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, 337(6092), 337-341.
- Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, 5(1), 1-21.
- Bergen, P. (2022, January 25). The threat Robert F. Kennedy Jr. poses to Covid-19 vaccination efforts. *CNN*. <https://www.cnn.com/2022/01/24/opinions/robert-kennedy-dangerous-vaccine-misinformation-bergen/index.html>
- Berger, J., & Milkman, K. L. (2012). What makes online content viral?. *Journal of marketing research*, 49(2), 192-205.
- Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one*, 10(2), e0118093.
- Carey, J. M., Chi, V., Flynn, D. J., Nyhan, B., & Zeitzoff, T. (2020). The effects of corrective information about disease epidemics and outbreaks: Evidence from Zika

and yellow fever in Brazil. *Science advances*, 6(5), eaaw7449.

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, 27(1), 69-96.

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in medicine*, 32(4), 556-577.

Coleman, N. V., & Williams, P. (2013). Feeling like my self: Emotion profiles and social identity. *Journal of Consumer Research*, 40(2), 203-222.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 5, No. 1, pp. 89-96).

Cui, L., & Lee, D. (2020). Coaid: Covid-19 healthcare misinformation dataset. *arXiv preprint arXiv:2006.00885*.

De Figueiredo, A., Simas, C., Karafillakis, E., Paterson, P., & Larson, H. J. (2020). Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. *The Lancet*, 396(10255), 898-908.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554-559.

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.

Getman, R., Helmi, M., Roberts, H., Yansane, A., Cutler, D., & Seymour, B. (2018).

Vaccine hesitancy and online information: the influence of digital networks. *Health Education & Behavior*, 45(4), 599-606.

Gottfried, J., & Shearer, E. (2019). News use across social media platforms 2016.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix:

Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, 36(2), 193-202.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake

news on Twitter during the 2016 US presidential election. *Science*, 363(6425), 374-378.

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence

from the consumption of fake news during the 2016 US presidential campaign. *European Research Council*, 9(3), 4.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings

reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Hossain, T., Logan IV, R. L., Ugarte, A., Matsubara, Y., Young, S., & Singh, S. (2020).

COVIDLies: Detecting COVID-19 misinformation on social media.

Howell, L. (2013). Digital wildfires in a hyperconnected world. *WEF report*, 3(2013), 15-

94.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for

sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Jacobs, W., Amuta, A. O., & Jeon, K. C. (2017). Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. *Cogent Social Sciences*, 3(1), 1302785.

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine*, 35(29), 3621-3638.

Konstantinou, P., Georgiou, K., Kumar, N., Kyprianidou, M., Nicolaides, C., Karekla, M., & Kassianos, A. P. (2021). Transmission of vaccination attitudes and uptake based on social contagion theory: A scoping review. *Vaccines*, 9(6), 607.

Kwok, K. O., Lai, F., Wei, W. I., Wong, S. Y. S., & Tang, J. W. (2020). Herd immunity—estimating the level required to halt the COVID-19 epidemics in affected countries. *Journal of Infection*, 80(6), e32-e33.

Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. H. M., Hasan, S. M., Kabir, A., ... & Seale, H. (2020). COVID-19—related infodemic and its impact on public health: A global social media analysis. *The American journal of tropical medicine and hygiene*, 103(4), 1621.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.

Lazić, A., & Žeželj, I. (2021). A systematic review of narrative interventions: Lessons for

countering anti-vaccination conspiracy theories and misinformation. *Public Understanding of Science*, 30(6), 644-670.

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021).

Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature human behaviour*, 5(3), 337-348.

MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34), 4161-4164.

Marsden, P. V. (1993). The reliability of network density and composition measures. *Social Networks*, 15(4), 399-421.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 415-444.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Pallas, F., Stauffer, D., & Kuhlenkamp, J. (2020, December). Evaluating the Accuracy of Cloud NLP Services Using Ground-Truth Experiments. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 341-350). IEEE.

Pasquetto, I. V., Swire-Thompson, B., Amazeen, M. A., Benevenuto, F., Brashier, N. M., Bond, R. M., ... & Yang, K. C. (2020). Tackling misinformation: What researchers could do with social media data. *The Harvard Kennedy School Misinformation Review*.

Pennebaker, James W., Matthias R. Mehl, and Kate G. Niederhoffer. 2003. Psychological

aspects of natural language use: Our words, our selves. *Annual review of psychology* 54, 1, 547-577.

Peretti-Watel, P., Seror, V., Cortaredona, S., Launay, O., Raude, J., Verger, P., ... & Ward, J. K. (2020). A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. *The Lancet Infectious Diseases*, 20(7), 769-770.

Pfizer. Pfizer and BioNTech conclude phase 3 study of COVID-19 vaccine candidate, meeting all primary efficacy endpoints. (2020, December 18).  
<https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-conclude-phase-3-study-covid-19-vaccine>

Poland, G. A., & Jacobson, R. M. (2011). The age-old struggle against the antivaccinationists. *N Engl J Med*, 364(2), 97-9.

Quattrocioni, W., Caldarelli, G., & Scala, A. (2014). Opinion dynamics on interacting networks: media competition and social influence. *Scientific reports*, 4(1), 1-7.

Rossen, I., Hurlstone, M. J., Dunlop, P. D., & Lawrence, C. (2019). Accepters, fence sitters, or rejecters: Moral profiles of vaccination attitudes. *Social Science & Medicine*, 224, 23-27.

Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., & Quattrocioni, W. (2018). Polarization of the vaccination debate on Facebook. *Vaccine*, 36(25), 3606-3612.

Scott, J. (1988). Social network analysis. *Sociology*, 22(1), 109-127.

Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R., & Roychowdhury, V. (2020). Conspiracy in the time of corona: Automatic detection of emerging COVID-19

conspiracy theories in social media and the news. *Journal of computational social science*, 3(2), 279-317.

Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016, April). Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web* (pp. 745-750).

Shapiro, J. N., Oledan, J., & Siwakoti, S. (2020). ESOC COVID-19 Disinformation Tracking Report. *Empirical Studies of Conflict*. <https://esoc.princeton.edu/publications/esoc-covid-19disinformation-tracking-report>.

Siwakoti, S., Yadav, K., Thange, I., Bariletto, N., Zanotti, L., Ghoneim, A., & Shapiro, J. N. (2021). Localized misinformation in a global pandemic: Report on COVID-19 narratives around the world. *Empirical Study of Conflict Project, Princeton University*, 1-68.

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. *Annu. Rev. Sociol.*, 34, 405-429.

Tausczik, Y.R., and Pennebaker, J.W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29, 1, 24-54.

Twitter. (n.d.). *About Verified Accounts*. <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>

Umopathy, V. U. (2022, February 1). *Is the Daily Mail reliable*. The Factual. <https://www.thefactual.com/blog/is-the-daily-mail-reliable/>

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news

online. *Science*, 359(6380), 1146-1151.

Weinzierl, M., & Harabagiu, S. (2022). Identifying the Adoption or Rejection of

Misinformation Targeting COVID-19 Vaccines in Twitter Discourse. *arXiv preprint arXiv:2202.09445*.

Xu, L., Ma, Z., Li, Y., Pang, Z., & Xiao, S. (2021). Antibody dependent enhancement:

Unavoidable problems in vaccine development. *Advances in immunology*, 151, 99-133.

Zhu, B., Chen, C., Loftus, E. F., He, Q., Chen, C., Lei, X., ... & Dong, Q. (2012). Brief

exposure to misinformation can lead to long-term false memories. *Applied Cognitive Psychology*, 26(2), 301-307.