NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AN ANALYSIS OF COVID-19 MISINFORMATION ON THE TELEGRAM SOCIAL NETWORK

by

Marcus A. Garcia

September 2022

Co-Advisors: Ruriko Yoshida
Louis Chen
Tauhid Zaman,
Yale University
Second Reader: Robert A. Koyak

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2022	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE AN ANALYSIS OF COVID-19 MISINFORMATION ON THE TELEGRAM SOCIAL NETWORK	**5. FUNDING NUMBERS**
6. AUTHOR(S) Marcus A. Garcia	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

The proliferation of misinformation groups and users on social networks has illustrated the need for targeted misinformation detection, analysis, and countering techniques. For example, in 2018, Twitter disclosed research that identified more than 50,000 malicious accounts linked to foreign-backed agencies that used the social network to spread propaganda and influence voters during the 2016 U.S. presidential election. Twitter also began removing and labeling content as misinformation during the 2020 U.S. election, which led to an influx of users to social networks, such as Telegram. Telegram's dedication to free speech and privacy is an attractive platform for misinformation groups and thus provides a unique opportunity to observe and measure how unabated ideas and sentiments evolve and spread. In this thesis, we create a dataset by crawling channels and groups in Telegram that are centered around COVID-19 and vaccine conversations. For analysis, we first analyze the topics and sentiments of the data using machine learning models. Next, we analyze the time series relationship between sentiment and topic trends. Then, we look for topic relationships by clustering performed on topic-based graph networks. Lastly, we cluster channels using document vectors to identify super-groups of related conversations. We conclude that Telegram communities risk producing echo chamber effects and are potential targets for external actors to embed and grow misinformation without hindrance.

14. SUBJECT TERMS network, social, graph, misinformation, bots, telegram	**15. NUMBER OF PAGES** 79
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

AN ANALYSIS OF COVID-19 MISINFORMATION ON THE TELEGRAM SOCIAL NETWORK

Marcus A. Garcia
Lieutenant, United States Navy
BS, University of Maryland, Baltimore County, 2014

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 2022

Approved by: Ruriko Yoshida
Co-Advisor

Louis Chen
Co-Advisor

Tauhid Zaman
Co-Advisor

Robert A. Koyak
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The proliferation of misinformation groups and users on social networks has illustrated the need for targeted misinformation detection, analysis, and countering techniques. For example, in 2018, Twitter disclosed research that identified more than 50,000 malicious accounts linked to foreign-backed agencies that used the social network to spread propaganda and influence voters during the 2016 U.S. presidential election. Twitter also began removing and labeling content as misinformation during the 2020 U.S. election, which led to an influx of users to social networks, such as Telegram. Telegram's dedication to free speech and privacy is an attractive platform for misinformation groups and thus provides a unique opportunity to observe and measure how unabated ideas and sentiments evolve and spread. In this thesis, we create a dataset by crawling channels and groups in Telegram that are centered around COVID-19 and vaccine conversations. For analysis, we first analyze the topics and sentiments of the data using machine learning models. Next, we analyze the time series relationship between sentiment and topic trends. Then, we look for topic relationships by clustering performed on topic-based graph networks. Lastly, we cluster channels using document vectors to identify super-groups of related conversations. We conclude that Telegram communities risk producing echo chamber effects and are potential targets for external actors to embed and grow misinformation without hindrance.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

API	application program interface
BERT	Bidirectional Encoder Representations from Transformers
COVID-19	Coronavirus Disease 2019
DHS	Department of Homeland Security
DGB	Disinformation Governance Board
IDF	inverse document frequency
JSON	JavaScript object notation
LDA	latent Dirichlet allocation
TF	term frequency
UMAP	uniform manifold approximation and projection
URL	uniform resource locator

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

In recent years, the proliferation of misinformation on social media has created significant concern. We are still learning about the impacts of the COVID-19 pandemic and the consequences of COVID-19 and vaccination skepticism. The efforts from social media companies to mitigate the spread of misinformation on their platforms have led to mixed results. They have created a population of de-platformed users and demand for social networking platforms with less oversight. Telegram and other platforms have filled this void.

In this thesis, we build a dataset by crawling COVID-19 and vaccine-focused channels on Telegram. We use machine learning models to examine the topics and sentiments to begin the analysis. Next, we find super-groups of connected discussions by clustering channels. The time series link between sentiment and topic trends is then examined. Then, we identify topic relationships using clustering on topic-based graph networks. Finally, we look at the sentiments surrounding different topics.

We identify a large number of groups in the dataset as possible echo chambers dedicated to conversations surrounding skepticism of COVID-19 and vaccines, politics, and related topics. Second, we identify specific causes for trending topics and link them to conspiracies and anti-vaccine narratives. Third, we illustrate how specific keywords follow similar patterns in their usage, suggesting they belong to larger narratives that ebb and flow in prominence over time. Lastly, we show significant differences in sentiments connected to keyword usage, which further suggests we have created a dataset driven by anti-vaccination sentiments.

Our analysis highlights Telegram communities as vulnerable targets for misinformation campaigns due to the unregulated environment that could foster rapid growth of disinformation. Therefore, we propose follow on research to identify misinformation clusters, determine prevalent narratives, and develop nudge techniques to disrupt the cycle of misinformation and gently guide users out of the echo chambers.

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

To my wife, Caroline, thank you for your constant support. I love you and can't imagine making my way through this life without you.

To my fellow Operations Research students, thank you. Your camaraderie got me through this program.

To all my instructors, you made this the best educational experience possible.

To the Trident Room and staff, I will be back.

And last but not least, thank you to Dr. Yoshida, Dr. Zaman, Dr. Chen, and Dr. Koyak for your guidance throughout this process.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

1.1 Motivation

In recent years, the proliferation of misinformation on social media has created significant concern. Misinformation is inaccurate or deceptive information presented as fact. When the author or spreader knows this information is incorrect but spreads it with the intent to deceive others, we call this disinformation. One critical danger of misinformation is the people citing the information may not be aware that it is inaccurate and share it believing it is true.

A recent example is the intentional spread of misinformation about Coronavirus Disease 2019 (COVID-19) and vaccination which has become prevalent throughout the pandemic. According to the Center for Countering Digital Hate, approximately 65% of all anti-vaccine content on social media is generated by or connected to only twelve individuals, dubbed the "disinformation dozen." Further, these efforts are an apparent organized effort to instill doubt into efficacy and safety of COVID-19 vaccines (Srikanth 2021). This illustrates how a small number of individuals can leverage the spans of social networks to push their ideology at the expense of public health.

During the 2016 U.S. Presidential election, the Internet Research Agency, a Russian organization, generated and spread propaganda on various social media networks to influence the outcome of the election and sow division in the public. In 2018, Twitter disclosed the results of an internal review which identified 50,000 malicious automated accounts linked to the Internet Research Agency (Twitter Public Policy 2018). These accounts reached over 600,000 users in the United States during the election cycle.

Twitter's expanded policy of labeling potential misinformation with warnings, removing content that violates its terms of service, and banning users that incite violence has led to mixed results. One effect was the creation of a group of de-platformed users which led to an influx of users to other free speech-focused social media platforms such as Telegram (Bond

2021).

Telegram has seen an influx of users since its development in 2016. It is self-described as being a haven for free speech and privacy. For example, after President Donald J. Trump was banned from Twitter in 2021 for inciting violence, Brazil's right-wing President, Jair Bolsonaro, requested his citizen follow him on Telegram, significantly increasing Brazil's usage rate of Telegram (Londono et al. 2021).

In 2022, the Department of Homeland Security (DHS) tried to establish the Disinformation Governance Board (DGB), an advisory board focused on monitoring and developing policy surrounding misinformation and disinformation. However, DGB was quickly characterized as the "Ministry of Truth" and placed on hold (Hart 2022). Despite this, the establishment of the DGB illustrates that the DHS and the U.S. government recognize the potential harm to our national interests that malicious actors can levy through the weaponization of false information.

Twitter has developed policies and in-house methods for policing content that violates platform integrity and authenticity (Twitter 2022). In a sense, enforcement of these policies disrupts free and open dialog; thus, the conversations on the platform are not a good representation of natural human dialogue.

Telegram fundamentally differs from Twitter and other social media services in structure, privacy, and the platform's stance on free speech. Structurally, Twitter is one continuous feed of information that is a culmination of the tweets of accounts that a user follows and content that Twitter suggests to the user based on an algorithm that analyzes user data. Users follow other users to create a network.

Telegram is comprised of groups and channels that users elect to join. These can be analogous to natural communities that we would see in social groups and regional communities. Since Telegram is privacy-focused, it does not utilize information about a user's patterns to suggest channels and groups to users. Instead, users must seek out channels that interest them or rely on their existing social connections' suggestions and invitations to channels. Perhaps most importantly, users are free to express themselves without restriction, often anonymously. Telegram is arguably similar to human dialog as it would naturally occur in communities. The uniqueness of Telegram and the existence of fringe groups on the

platform provides a unique opportunity for analysis.

1.2 Research Questions

Some of the questions we want to answer using our dataset are as follows:

- What are the primary topics discussed in our Telegram dataset?
- Can we identify communities based on the topics they are discussing?
- When certain topics are trending, what is the driving factor for its prevalence?
- Do we see trends between topics that indicate they're developing together?
- When topics are being discussed, is it due to a polarized reaction? If so, what is the opinion expressed around these topics?

1.3 Thesis Organization

The remainder of this thesis is divided into five chapters. Chapter 2 provides more background information on Telegram and covers relevant research on misinformation and Telegram. Chapter 3 presents the analytical methods used in this thesis. Chapter 4 shows the results of our various analyses. Finally, chapter 5 concludes the thesis with an overview of our findings and suggestions for future research and work.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Literature and Background

In this chapter, we discuss the Telegram platform, its structure and how it differs from other social media platforms. We discuss previous work on misinformation on social networks, the spread of conspiracies and misinformation, techniques for countering misinformation, and other research that has been conducted on Telegram.

2.1 Telegram Platform

Telegram was launched in 2013 by Pavel and Nikolai Durov. It has approximately 700 million active monthly users (Telegram 2022a). It is an instant messaging service where users can send private messages to other users, but it also has social networking capabilities such as channels and groups that users can join.

Channels on Telegram function as a means of broadcasting information via posts that are visible to all users who subscribe to the channel. Channels have one or more administrators who control the settings and are able to make posts with content. Depending on the settings determined by the administrators, subscribers may be able to comment on these posts and reply to other users' comments.

Groups on Telegram act as a group chat which consists of a collection of users who elect to join the group. Groups can have one or more administrators or moderators with special privileges. Users can find channels to subscribe to by channel name or with a uniform resource locator (URL). Users can find groups to join by receiving invitations or URLs from other users. The messages in channels and groups are primarily text but can contain multimedia files such as images and videos. They can also be messages from other channels forwarded into a conversation.

Telegram is unique from other social platforms since it is primarily a messaging platform. The main difference from Twitter and Facebook is that Telegram users do not have public friends or followers, but instead have contacts. They also do not have a "wall" or "page" where they post messages and share content with followers or the public. However, users

do have the ability to create a channel which serves this purpose.

2.1.1 Privacy & Content Moderation

Telegram users have very robust privacy options. Their username, groups, and channels they belong to can all be hidden and kept private. Users have a "User Info" page to display their username and share an optional biography; however, users do not have to share any information here.

The main draws of Telegram for users is the privacy policy and the platform's stance on free speech. This following information is derived from the statements and documentation provided in Telegram's Privacy Policy and Frequently Asked Questions (Telegram 2022a, Telegram 2022b).

Telegram does not use user data to show users ads. They also allow for end-to-end encrypted secret chats so that Telegram cannot access the content of those secret messages. Telegram usually does not comply with content removal requests from third parties. In the case of publicly available content that originates from bots and channels they do process requests pertaining to intellectual property violations, pornography, and terrorist activity. However, they do not comply with content removal requests stemming from local restrictions on freedom of speech, regardless of local laws.

According to Telegram, they divide decryption keys into parts, and store them and the data it pertains to in different locations. The idea being that all the necessary data for any content is distributed across different jurisdictions to increase the requirements that have to be met before they are forced to turn over data. This policy has resulted in no data disclosures to third parties or governments.

2.2 Literature Review

This section discusses other research on topics surrounding misinformation on social media and other background information that provide a broader view necessary to approach our research questions.

2.2.1 Misinformation

Misinformation can be centered around seemingly insignificant topics with no political or social aims; however, it can also be used by state actors to polarize the public and sow mistrust in state institutions. Badawy et al. (2018) examine the influence attempts of Russian troll accounts on Twitter to manipulate the results of the 2016 U.S. Presidential election campaign and find a widespread adoption and spread of content generated by these accounts. Pennycook and Rand (2021) conducted surveys after the 2020 U.S. Presidential election which highlighted the effectiveness of misinformation campaigns.

2.2.2 Echo Chambers

Echo chambers are environments where members have similar beliefs and where those members are unlikely to encounter opposing arguments which potentially result in confirmation bias and polarized communities (Cinelli et al. 2021).

2.2.3 Detection Methods

Since groups have increasingly used bots to spread misinformation, one area of research has been to develop methods to detect users that exhibit unique behavioral patterns that suggest the users are online bots (non-human automated users).

Ferrara et al. (2016) provide an overview of popular techniques for bot detection. Often, bots are not generating content but forwarding, retweeting, or simply interacting with human-generated content to increase its visibility and how far the desired information can spread across the network and reach new users. Network-based bot detection techniques can be based on the connectivity of an accounts friend connections or interaction patterns. Another method is to use machine learning on various features of the user accounts, such as network structure, account meta data, temporal patterns, and content metrics. A more low-tech method is to simply crowd source bot identification to human workers.

Since misinformation can also be spread by non-automated accounts, while other approaches are aimed at detecting these non-automated accounts that spread misinformation. Lee et al. (2021) create “honeypot” accounts to lure and identify candidate content polluting accounts. Then they use classification techniques to automatically detect accounts that exhibit similar behavior as content polluters. One study focuses on the behavior and social network extremist

groups consisting of humans and bots (Berger and Morgan 2015). Other studies focus on detecting content aimed at polarizing users or "toxic" content, which is used to driving users away from online groups (Sureka and Agarwal 2014; Dinakar et al. 2021).

2.2.4 Combating Misinformation

One strategy to keep users from being duped by misinformation is to educate them. Social media platforms' principal approach is to detect and apply warning labels to suspected misinformation. The issue with this technique is that the lack of a misleading label results in an implied truth effect for any information not labeled with a warning (Pennycook et al. 2019a). This means that methods for detecting and classifying misinformation would have to be perfectly accurate and applied to every incidence of disinformation, which is an impossible standard; thus, the risk of implied truth effect will always be present when using warning labels in this manner. Another way of preventing disinformation is to train people to be more critical of information accuracy. Pennycook et al. (2019b) were able to redirect user focus to the material's accuracy and improve the quality of information shared by having users score the integrity of headlines.

Another method is to utilize nudging tactics to influence consumers. Conflict is the least effective way to affect users. Assume we directly dispute people's ideas or push storylines that oppose their beliefs. In that situation, it has the unintended consequence of making subjects more entrenched in their belief system and hence more difficult to reach (Mosleh et al. 2021). Yang et al. (2020) develop an influence model to combat this, which first identifies the topic key to the user's rhetoric, mimics or "paces" their language, and gradually transitions to opposing ideas, thus "guiding" them to alternatives to their predetermined beliefs. Vendeville et al. (2022) tackles the task by introducing diverse content recommendations to users based on the user's apparent preferences.

2.2.5 Telegram Research

Since Telegram is a newer platform, research on Telegram is sparse but growing. Dargahi Nobari et al. (2017) build a dataset of Telegram messages by developing a crawler and then perform an exploratory analysis of Telegram's structure focusing on the topical nature of channels and groups. While focused on a forensic analysis of the Telegram application

itself, Anglano et al. (2017) provide a thorough overview of the privacy capabilities of Telegram messages and groups that could be leveraged by criminal groups.

Conspiracy groups, such as QAnon, have also used Telegram as a means to communicate and spread misinformation. Hoseini et al. (2021) describe a multilingual analysis focused on QAnon channels and groups, studying the evolution of topics. La Morgia et al. (2021) created a dataset that aimed to be a broader cross-section of Telegram channels, developed methods to detect fake and clone accounts and channels, and briefly explored a conspiracy movement called Sabmyk, a competitor to QAnon.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:
Methods

In this chapter, we describe the analytical methods we use in this study. We begin by discussing how we create the dataset, pre-processed text, and extract subsets of data from messages that contained keywords of interest. Next, we fit topic models to get precursory knowledge about the conversations in the data. We suspect there could be natural groupings of channels and groups. Using the vocabulary of the dataset, we show how to create representative vectors for each channel, cluster them and inspect each cluster. Next, we use the keyword data subsets and time-series information to look for connections and correlating topics that develop together. Lastly, we leverage sentiment measurement of messages to see if keyword usage is significant to the message's sentiment.

3.1 Dataset

Our first step is to obtain data from Telegram. Because Telegram does not suggest groups and channels to users, we must discover groups of interest on our own. To do this, we turn to the methodologies outlined for crawling Telegram by Dargahi Nobari et al. (2017) and Hoseini et al. (2021). We start with an initial seed set of groups or channels, and we look for references to other groups or channels within our seed set.

To create our initial seed set, we manually search for candidate groups and channels utilizing Telegram's global search feature and three unofficial Telegram directories (Telegram Channels 2021; TGStat 2021; Telegram Directory 2021). We perform a manual inspection to select entities that fell into two categories: centered around vaccine and COVID-19 topics (A.1) or focused on other topics such as sports, technology, general entertainment, or politics (A.2).

Using Telethon, a Python library used to access Telegram's application program interface (API), we fetch all messages and replies to those messages from our entities of interest. We use the algorithm presented by Yousefi (2019) as a base framework for fetching messages in JavaScript object notation (JSON) format. A sample message is shown in Fig 3.1. Note that the value corresponding to "replies" is not null, indicating replies are possible or present.

We check each message for this condition, and replies to the messages are fetched and stored.

Figure 3.1 is a sample of a single message in JSON format.


```
[
    {"_": "Message",
        "id": 23,
        "peer_id": {"_": "PeerChannel", "channel_id": 1234567890},
        "date": "2022-05-10T10:24:03+00:00",
        "message": "Here is the message text that we analyze.",
        "out": false,
        "mentioned": false,
        "media_unread": false,
        "silent": false,
        "post": true,
        "from_scheduled": false,
        "legacy": false,
        "edit_hide": true,
        "pinned": false,
        "from_id": null,
        "fwd_from": null,
        "via_bot_id": null,
        "reply_to": null,
        "media": {"_": "MessageMediaPhoto", "photo": {"_": "Photo",
        "reply_markup": null,
        "entities": [{"_": "MessageEntityBold", "offset": 0, "lengt
        "views": 678,
        "forwards": 4,
        "replies": {"_": "MessageReplies", "replies": 0, "replies_p
        "edit_date": "2022-05-10T10:24:06+00:00",
        "post_author": null,
        "grouped_id": null,
        "restriction_reason": [],
        "ttl_period": null
    }
]
```


Figure 3.1. Sample Data Format of a Single Message from a Telegram Channel. All messages from a group or channel are output as a list of comma-separated JSON entries. Note that the 'replies' key does not have a *null* entry, signifying there may be replies to this message.

We then search through the message texts for URL patterns that are links to other Telegram entities (*t.me*) and record the number of times each URL is referenced. These URLs in the body of messages can be considered advertisements for groups/channels. Often, forwarded messages posted in a channel will contain URL self-references to the original channel so that users can easily find the source. From this set of newly discovered entities, we then fetch

the messages and replies from the most referenced, setting a minimum threshold number of references to limit the size of our dataset. We complete this process for both the COVID-19 entities and general topic entities.

The resulting dataset comprises 9,726,778 Messages and 19,938,687 Replies from 895 Channels and 119 Groups.

3.1.1 Text Preparation

To prepare for topic analysis, we must standardize the text and remove extraneous information. To standardize the text, first, we lemmatize it, which converts the variant forms of a word to its base word, essentially converting words that are functionally the same to a standard form. For example, "lives" would be transformed into "life." The point is to limit the number of variations of to a smaller subset that has the same meaning or contain the same information. During this process, we also convert uppercase letters to lower case. Next, we identify and remove text that provides no relevant information to our analysis, such as stop words, punctuation, and hyperlinks. Stop words are articles, conjunctions, and order words that occur naturally in language structure but provide little information, such as "the," "it," "a," and "for." We refer to the resulting text as *clean text*.

3.1.2 Keyword Subsets

In order to analyze trends in conversation, we choose a variety of keywords we expect to occur across a spectrum of conversation surrounding politics, COVID-19 and vaccines. We intentionally select keywords that we reasonably expect could be used in discourse that range from having positive to negative sentiment.

We intentionally include some alternative spellings, acronyms and truncated forms of the keywords of interest to capture more references to the keywords across the dataset. We do not require the whole word to match, e.g., "deworm" will detect the occurrences of both "deworm" and "dewormer." We do not account for all possible misspellings.

After selecting our search terms, we identify all messages in the dataset that contain matches to each keyword string We then count the number of messages that contained at least one occurrence of our keyword per day. This gives us a time series of the daily mentions of the

keyword. We divide the daily count of each keyword by the total number of messages in the dataset resulting in the daily percentage of messages in the dataset that contain the keyword of interest.

Finally, we combined the daily count data for thematically related keywords into keyword groups. The keyword groups and search terms are shown in Table

3.1.

Table 3.1. Keyword Groups and Terms

Keyword Group Topic	Search Terms
COVID-19/Virus	covid, corona
Vaccines	mrna, vaccine, shot, pfizer, moderna, astrazeneca, biontech, jj, booster
Masks	mask, breath, n95
Alternative Medicine	deworm, hydroxychloroquine, dexamethasone, methanol, ethanol, ivermectin, saline, ultraviolet, chlorine, hcq, hydrox
Medical Concerns	stroke, clot, fertility, viral load, myocarditis, sterili, aneurysm, aneurism, infertil
Conspiracies	microchip, 5g, bioweapon
Political Left	fauci, biden, pelosi, gates, clinton, kamala, aoc
Political Right	malone, trump, mtg, mcconnell, lindell
Negative Terms	terrorist, dictator, tyrant, moron, stupid, idiot, murder
Liberal Terms	liberal, democrat
Conservative Terms	conservative, republican
Moderate Terms	centrist, moderate

3.2 Topics

Now that we have a workable dataset, we want to understand the space and nature of the conversations that are detected in it. To do this, we turn to topic modeling. Topic modeling

is a method of finding groups of similar words or documents through unsupervised machine learning. We accomplish this through latent Dirichlet allocation (LDA). LDA provides a probability distribution of topics which would produce a word or document. LDA differs from clustering in that an observation will be assigned probabilities that indicate how likely it is to belong to each topic (Blei et al. 2003).

Specifically, we use the gensim Python library to train our LDA model on the entire dataset and identify topics (Rehurek 2010). When fitting LDA models, the user must specify the number of topics. Because LDA models fit to a large corpus is time consuming, we choose to identity five topics, which is sufficient for reaching our objectives.

After the topic model is trained, we manually inspect the most frequent terms in each topic to assign a label to the pattern that model has identified. Then, we create a two dimensional visualization of the topic model, which provides insight into the frequency of the topics and how similar or dissimilar topics are from one another.

3.3 Channel and Group Commonalities

After gaining insight into the topics that occur in the dataset, we examine channel and group entities for clusters where each cluster has a commonality between its members. To do this, we use document vectors to develop a representation of the conversation that occurs within each channel or group.

3.3.1 Channel Vectors

To develop document vectors, we first create a bag-of-words model, which is the "bag" or multiset of all words in a document with the number of occurrences of each word (Yse 2019). Term frequency (TF) is the number of times a word occurs in a document divided by the total number of words in the document. TF-inverse document frequency (IDF) is a term frequency statistic to measure the importance of a word by offsetting TF by the number of documents that contain the word. From these, we build document term matrices where the row corresponds to a document, columns are words, and the values across a row are a document vector.

For each message in a channel, we calculate its TF and TF-IDF vectors and collect them

into their respective document term matrix. Next, we average these embeddings into one representative channel vector. We repeat this for every channel and group in our dataset. This results in high-dimensional vectors that represent these channels.

3.3.2 Dimensionality Reduction

Dimensionality is the number of coordinates that it takes to represent an object. To reduce the computational complexity required to process this data and to be able to represent the data graphically, we reduce the dimensionality of these vectors to a smaller number of dimensions using uniform manifold approximation and projection (UMAP). UMAP is a non-parametric approach that preserves global structure and suited for machine learning data preparation (McInnes et al. 2018). We reduce the dimensions of our vectors to two dimensions to allow for simple visualizations.

3.3.3 Clustering

Next, we use clustering to determine our groupings. Clustering is an unsupervised machine learning technique used to split data into groups where the observations within a group are very similar, and the difference between clusters is maximized. Since an observation's attributes are often spatial, similar data are close together and in dense groups. When clustering is performed, observations will be assigned to a single cluster (Grootendorst 2021).

K-Means clustering is a centroid-based method of clustering observations, where K, provided by the user, is the number of clusters. There are K centroids, and observations are assigned to a cluster when it is closest to the centroid of that cluster than to any other cluster's centroid (Mysiak 2020).

After clustering the channels and groups into three divisions, we want to see how the clusters differ. To do this, we combine all messages from all channels and groups in each cluster, remove stop words, and produce word clouds, which allow us to visually inspect the most frequent words.

3.4 Time Series Analysis

After gaining insight to the nature of the topics that are occurring in our data, we start to consider that the topics we identify are from the entire timeline of the dataset. We want to see how the conversations evolve over time. When topics are trending, what is unique about these peaks? Do we see trends between topics that indicate they're developing together? Are topics replacing one another and narratives swapping to new topics?

3.4.1 Peak Analysis

In order to analyze the conversation when topics are trending, we look at the time series occurrence of keywords. To focus our efforts, we choose to limit the time period of our analysis to the years 2020 and 2021.

First, we identify the time a peak in the usage rate of a keyword occurs. Then we take the messages that occur within a time period covering the peak, and inspect it by creating word clouds of the most frequently used words.

3.4.2 Correlation between Keyword trends

Since we want to see if there are keywords that evolve and grow together, we can use the correlation coefficient between the daily time series counts of the keyword occurrences. Further, we can create an undirected graph network based on these scores, and then, using graph statistics, we identify clusters within the graph structure.

Since we do not know the relationship between these many pairs of time series data and there exists a strong possibility of outliers, we use the Spearman correlation coefficient for the pairwise correlations.

Graph Networks

Next, for every pair of keywords with a correlation coefficient greater than a chosen threshold of 0.7, we add both keywords as nodes and create an undirected edge between them. We do this for all keyword pairs to develop one graph.

Spectral Clustering

Next, we use spectral clustering to identify communities of keywords. Spectral clustering is a technique used to cluster data through the application of graph theory. It treats observations as nodes and performs graph partitioning. K-Means clustering assumes the observations belonging to a cluster are approximately spherical about its centroid. Spectral clustering does not require this assumption and can account for non-convex clusters of observations (von Luxburg 2007).

In order to determine the optimal number of spectral communities, k, we use modularity as our metric. Modularity is used to measure the strength of the division of the network into our defined clusters. Modularity is defined as the proportion of edges that fall within a group minus the expected proportion of edges that would do so if the edges were randomly placed (Newman 2006). We try a range of values for k, calculate the modularity of our resulting clustered network, and choose the number of clusters that results in the highest modularity. As shown in Fig 3.2, the number of communities with the highest modularity was eight.

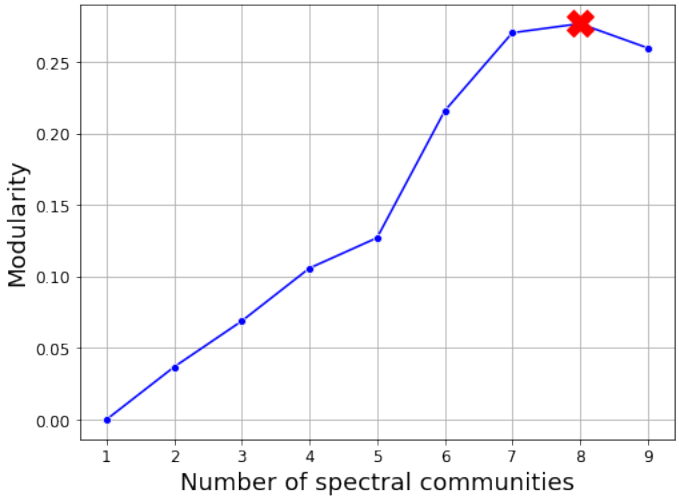


Figure 3.2. Number of Spectral Communities and Modularity

Finally, we cluster our graph into seven communities and plot the resulting graphs networks.

3.5 Sentiment

Keyword time series metrics are helpful in inspecting for trends of topics but do not necessarily tell us in what light the keywords are being referenced. When topics are being discussed heavily, is it due to a polarized reaction? If so, what is the opinion expressed around these topics when they are trending? In order to explore this, we use sentiment analysis. Sentiment is the measure of how positive or negative an author's disposition is towards an object or topic according to (Jagota 2020).

To determine the sentiment score of messages, we use a pre-trained transformer model which was trained for sentiment classification on product review text and the corresponding review's score. The possible score ranges from one to five. The specific model we use is a Bidirectional Encoder Representations from Transformers (BERT) multilingual model (NLP Town 2022).

3.5.1 Sentiment and Keyword Relationships

Now that we have sentiments of all messages, we can use this scoring to see how topics are discussed in the conversation space. We consider that the sentiment of messages that contain a keyword could be statistically different from messages that do not contain the keyword.

To do this, we take all sentiment scores of messages that contain each keyword and the sentiment scores of messages that do not. We use the t-test to test the null hypothesis that the mean sentiments are equal. Using a significance level of 0.01, we compare the resulting p-values to the significance level to determine if we can reject the null hypothesis.

We also consider that some of our keywords could be used in similar narratives and potentially have the same sentiment associated with their use. To test the null hypothesis that the mean sentiments of each keyword are not distinct from the other keywords, we us multiple hypothesis testing. We elect to use the Holm-Bonferroni method, which corrects for family-wise error rate.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:
Results

This chapter presents and discusses the results of our analysis. Section 4.1 will discuss and interpret the results of topic modelling. Section 4.2 explores the channel clusters. Section 4.3 shows the peaks in keywords over time and the conversations that occur during these peaks. Section 4.3.2 explores correlating narratives through graph networks. And section 4.4 analyzes the link between keyword usage and sentiment through hypothesis testing.

4.1 Topic Modelling

We used LDA topic modeling to identify five topics in our corpus. After fitting the topic model, we can inspect the most relevant words generated by the topic ranked according to their probability for each topic, shown in Table 4.1.

After inspecting the most relevant terms to each topic, we can identify themes between the words grouped within each topic.

- **Topic 1** has strong associations with terms that indicate it captures the conversations revolving around COVID-19 and vaccines.
- **Topic 2** includes the names of politicians, the government, and elections. Interestingly, the words channel, share, post, and telegram are present, indicating this topic might include many references to Telegram and other social media content.
- **Topic 3** is more mixed in themes, but we notice a few unique religious terms and some words linked to positivity.
- **Topic 4** has more general terms, particularly about time. This appears to be a general conversational topic.
- **Topic 5** is very different and is a mix of foreign language terms with some sports terms.

Table 4.1. Most Frequent Terms in each Topic

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
vaccine	trump	god	year	u
people	mike	thank	day	s
covid	need	like	today	arsenal
use	people	good	live	n
work	know	people	happen	de
need	election	know	time	o
know	biden	think	2	e
mask	channel	love	wait	ya
death	share	yes	player	m
year	president	look	happy	b
jab	state	time	watch	r
die	vote	right	let	arteta
system	like	thing	come	ha
professor	truth	need	new	em
dr	telegram	man	ago	d
card	think	agree	leave	que
want	post	bless	miss	q
health	want	great	3	t
school	america	come	month	messi
virus	country	pray	week	un

We turn to pyLDAvis, a Python library, to understand the relationship between topics and to produce an interactive topic model visualization (Siervert and Shirley 2015). We display the resulting visualization in Figure 4.1.

This plot visually represents the fitted LDA model. Each circle represents a topic, where the area of the circle is proportional to its topic frequency in the corpus. Topics 1, 2, and 3 are the largest and make up the bulk of the corpus. Topic 4 is smaller, and topic 5 is the

smallest.

The plot leverages multidimensional scaling, which attributes similarity to distances in the plot. Similar topics are plotted close together, and dissimilar topics are plotted further away from one another. We notice topics 1, 2, and 3 overlap, suggesting they are very similar. However, topics 4 and 5 are isolated from all other issues indicating they are very distinct.

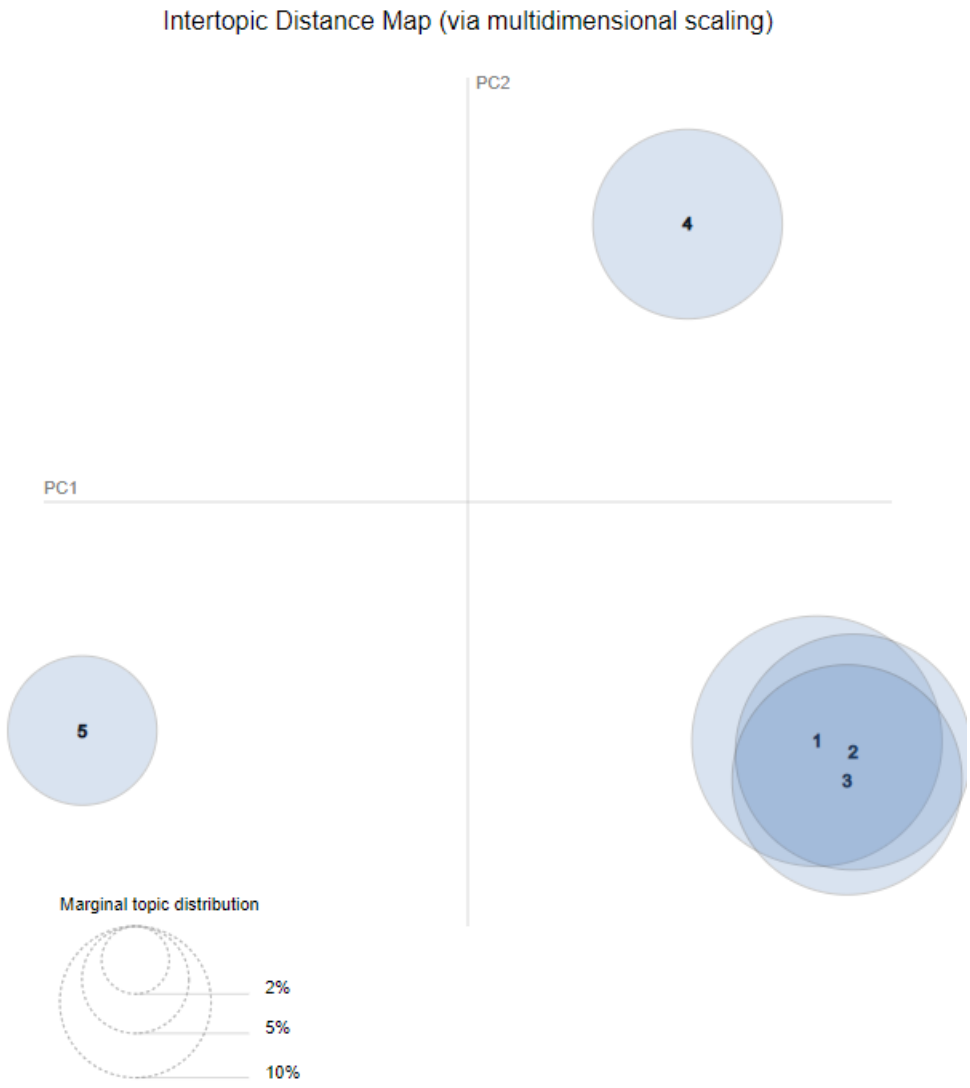


Figure 4.1. LDA Topic Model Visualization: This plot presents a visual representation of the fitted LDA model. The circles represent the topics, where the area of the circles is proportional to the topic's frequency in the corpus. Topics 1, 2, and 3 are the largest and make up the bulk of the corpus. These topics also overlap, suggesting they are similar. Topics 4 and 5 are both isolated and far away from other topics, indicating they are distinct from each other and the other topics.

4.2 Channel Clustering

In this section, we present the results of clustering performed on the channel vectors. Then, we inspect the most common words in the channels that are assigned to each cluster using word clouds. Finally we analyze the results to extract meaning in the cluster assignments.

To inform the selection of K for K-Means clustering, we utilize inertia curves. The selection of K using inertia curves is subjective, and there is no mathematical consensus on how to do so. Still, this is a sufficient method since we are only using the results for a subjective analysis of the frequently occurring words in each cluster.

By inspecting the inertia curves for both TF and TF-IDF embeddings in Fig 4.2 and Fig 4.3, we determine that three or five clusters provide sufficiently low inertia while keeping the number of clusters sufficiently low to be interpretable. Since the LDA topic model in Section 4.1 illustrated three topics make up a bulk of the corpus, we decide to set the number of clusters to three for both embedding methods.

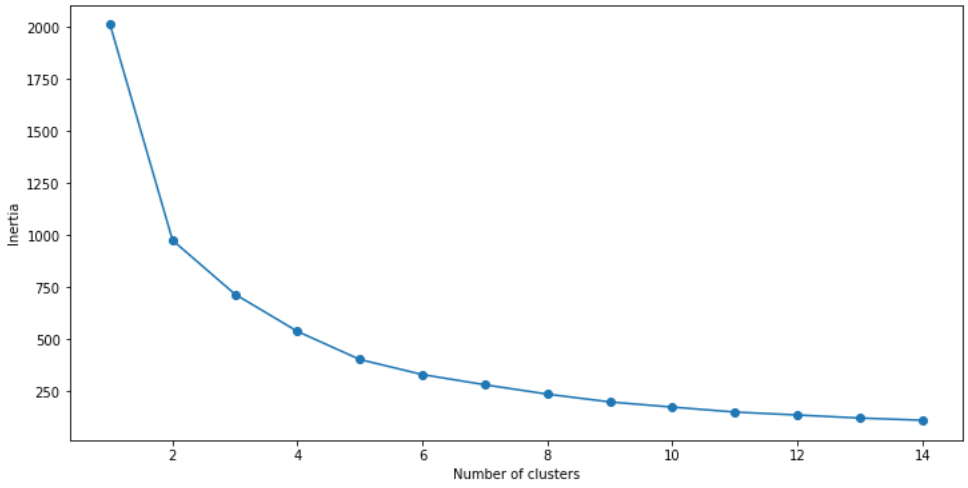


Figure 4.2. TF Inertia Curve

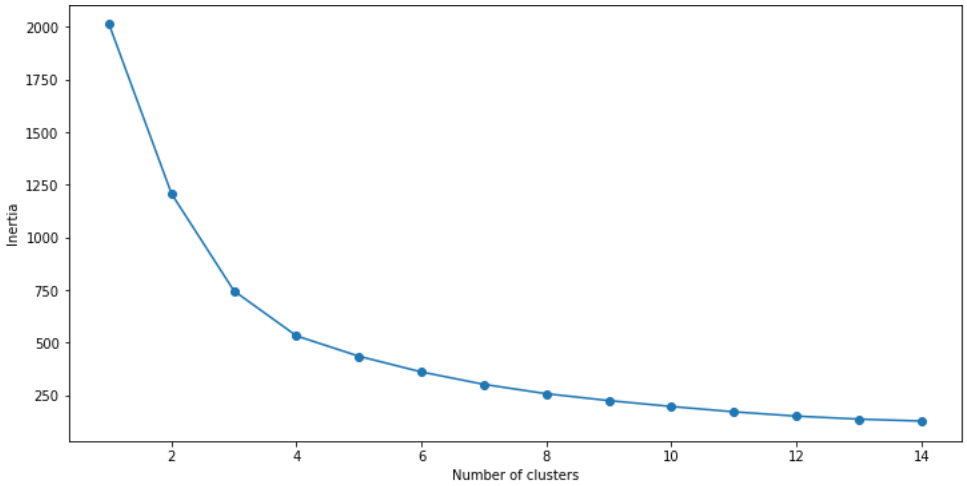


Figure 4.3. TF-IDF Inertia Curve

4.2.1 Channel Cluster Results

In this section, we present the results of K-means clustering on the TF and TF-IDF channel vectors.

TF Cluster Results

After reducing the TF channel vectors to two dimensional UMAP embeddings, they are divided into three clusters using K-means. The resulting clusters are plotted in Figure 4.4.

Visually inspecting the K-means clustering plot, there are several small groupings that are outliers in the UMAP projection. Cluster 1 does appear to find a delineation between its members and the other observations. The line that divides clusters 0 and 2 does not appear to follow a clearly natural division between the two sets of observations.

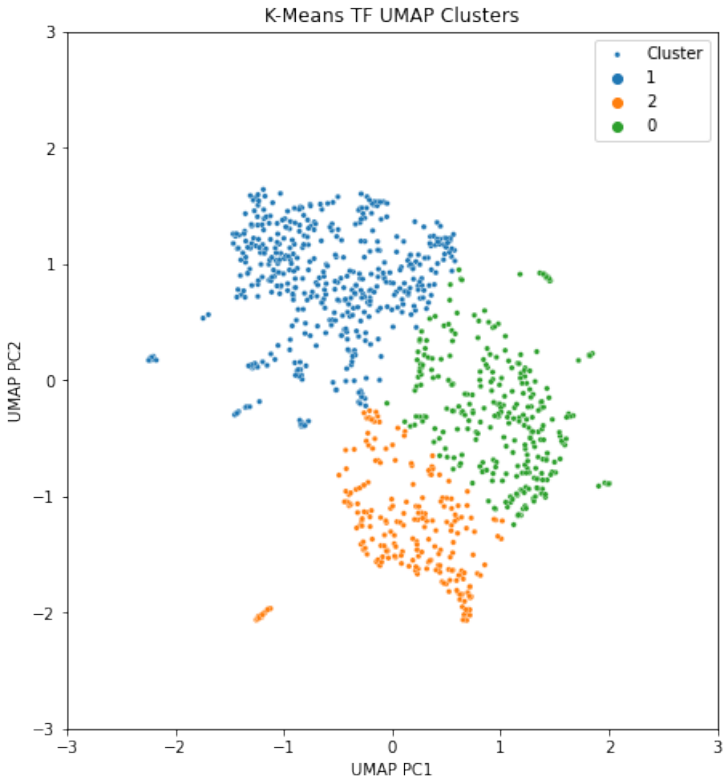


Figure 4.4. K-Means Clustering Plot of Channels Using TF Vectors: After using UMAP to reduce the channel vector dimension to two, K-Means clustering is used to identify three clusters of channels.

TF-IDF Cluster Results

We repeat the process using the TF-IDF channel vectors. The clustering results are presented in Figure 4.5.

Visually inspecting this K-means clustering plot, we see there are still several small grouping outliers in the UMAP dimensions, but their distance from the bulk of the observations has decreased significantly. The cluster assignments all seem to better follow visually apparent divisions than with the TF channel clusters.

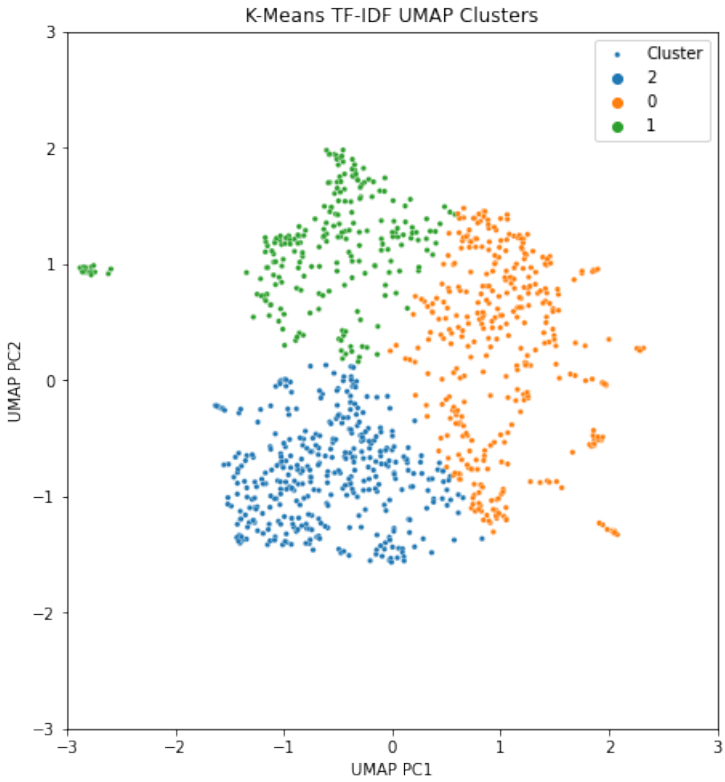


Figure 4.5. K-Means Clustering Plot of Channels Using TF-IDF Vectors: After using UMAP to reduce the channel vector dimension to two, K-Means clustering is used to identify three clusters of channels.

Comparing the cluster assignments of the TF channel vectors to the TF-IDF channel vectors, we find that 85.8% of the channel vectors are assigned to equivalent clusters.

4.2.2 Channel Cluster Word Clouds

After we perform clustering, we combine all messages in the channels that belong to each cluster and produce word clouds from the term frequencies.

These word clouds display the most frequent terms in each cluster and size the term according to frequency. The largest words are the most frequent, and the small words occur relatively less frequently. It is important to note that even the smallest words on a word cloud occurred relatively more frequently than those words that were not selected for inclusion, meaning

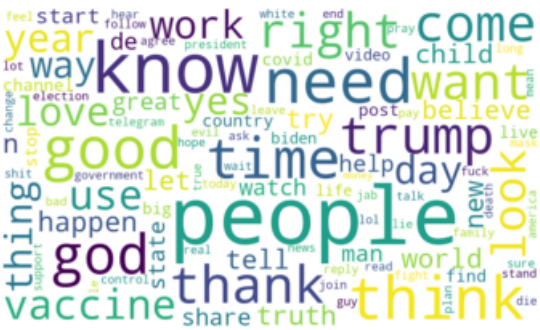
they are still a part of the conversational trends. Using visual inspection of the word clouds, we can attempt to see if there is some comparison to the identified LDA topics.

TF Cluster Word Clouds

Inspecting the word clouds for the TF clusters shown in Figure 4.6, we immediately notice that many terms exist on all three clusters but in different proportions. However, some terms appear in a similar proportion between all clusters. *People* and *know* are two that stand out as an example of this, suggesting we can disregard these terms while trying to identify unique trends in each cluster.

- **Cluster 0**: This word cloud has a big mix of terms, but it may be more conversational relative to the other two. We note it appears to have also accounted for the topic of religion with the terms of *god* and *love*.
- **Cluster 1**: The terms *vaccine*, *covid*, and *covid19* appear prominently. For smaller words, we notice *pfizer*, *health*, and *death*, suggesting this cluster is perhaps more Vaccine and COVID-19.
- **Cluster 2**: It is readily apparent that this word cloud captures the conversations about politics, due to the high frequency of *trump*, *biden*, and *election*.

Word cloud of Cluster 0: Conversational



Word cloud of Cluster 1: COVID-19 and Vaccines

Word cloud of Cluster 2: Politics

Figure 4.6. K-Means Clustering TF Word Clouds: Using the average TF channel clusters, we perform count the occurrence of terms in each cluster and produce a word cloud representing the most frequent words. The size of terms in the word clouds indicates relative frequency. We note the apparent theme of each cluster: Conversation, COVID-19 and Vaccines, and Politics.

TF-IDF Word Clouds

The word clouds produced from these clusters are displayed in Figure 4.7.

Upon inspecting the word clouds produced by the TF-IDF data, we notice very similar results as we discovered with the TF clustering.

- **Cluster 0**: This appears to capture the topics of general conversation and religion.
- **Cluster 1**: Again, the terms *vaccine*, *covid*, and *covid19* appear prominently and smaller terms (*pfizer*, *health*, and *death*) also reappear, indicating it is also capturing the topics of COVID-19 and Vaccines.
- **Cluster 2**: This cluster revolves around the topic of politics, with terms such as *trump*, *biden*, and *election*.

Overall, the TF-IDF cluster word clouds in Fig 4.7 to the TF cluster word clouds in Fig 4.6, we very similar results and draw the same conclusions. This clustering seems to have picked up on what we discovered from LDA topics: A tight group of topics that are highly related and have a lot of overlap in the conversations.

(a) Word cloud of Cluster 0: Conversation

(b) Word cloud of Cluster 1: COVID-19 and Vaccines

(c) Word cloud of Cluster 2: Politics

Figure 4.7. K-Means Clustering TF-IDF Word Clouds: Using the average TF-IDF channel clusters, and the term frequencies the clustered channels, we produce word clouds representing the most frequent words. The size of terms in the word clouds indicates relative frequency. We note the apparent theme of each cluster: Conversation, COVID-19 and Vaccines, and Politics.

4.2.3 Channel Cluster Discussion

In summary, it seems to indicate further that there are three clusters of channels that are highly similar. Each group shares a lot of terms with the other clusters suggesting the dataset is somewhat homogeneous across the clusters. We observe overlapping language used between clusters. For instance, we do see some conversational terms that frequently occur in all clusters, such as *people* and *know*, which is to be expected. Still, more importantly, we notice COVID-19 and Vaccine terms (*vaccine*, *covid*, etc.) occurring in high frequency in all clusters.

One note is that the TF-IDF data clusters appear to have been able to find a cluster of Politics channels that feature less prominent *COVID-19 and Vaccine* topics, but we do notice they are still present in these channels.

4.3 Time Series Analysis Results

In this section, we explore the results of our time series analysis of the mention rate of select keywords from Table 3.1.

4.3.1 Peak Analysis

Seeking to answer the question of what is unique about the conversations occurring during significant peaks in the mention rate of specific terms, we narrow our discussion down to *Vaccine* and *Alternative Medicine* terms.

We begin by inspecting the mention rate over time of *Vaccine* terms as shown in Figure 4.8.

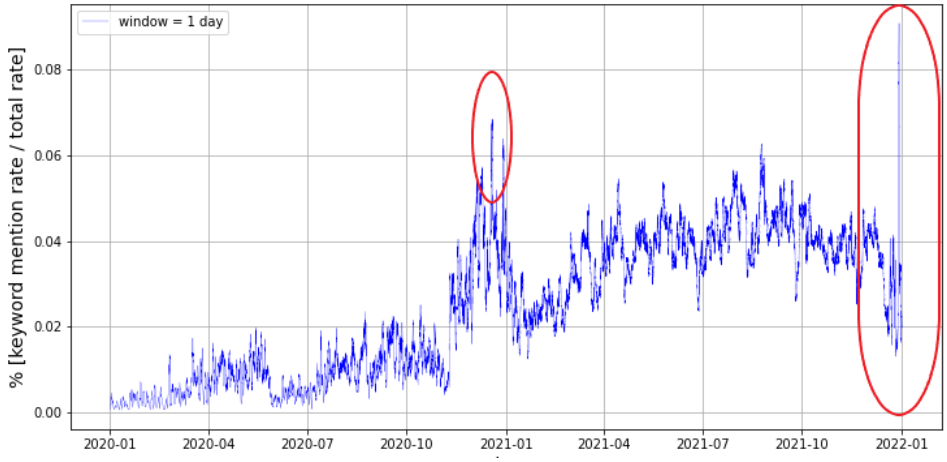


Figure 4.8. Vaccine Term Mention Rate Over Time: We notice two prominent peaks in the mention rate. The red overlayed ellipses highly the peaks which we selected to analyze. The first is on December 18, 2020 and has a gradual increase that builds for several weeks. The second occurs on December 29, 2021 and has a sharp peak that quickly appears.

We notice two prominent peaks in the mention rate. The first occurs around December 18, 2020, and has a gradual increase that builds for several weeks. The second occurs on December 29, 2021, and has a sharp peak that appears with no build-up. For both peaks, we select a 24-hour window centered on the peak, select the messages containing our *Vaccine* terms that were sent during the respective periods, and generate the word cloud in Figure 4.9 and Figure 4.10.

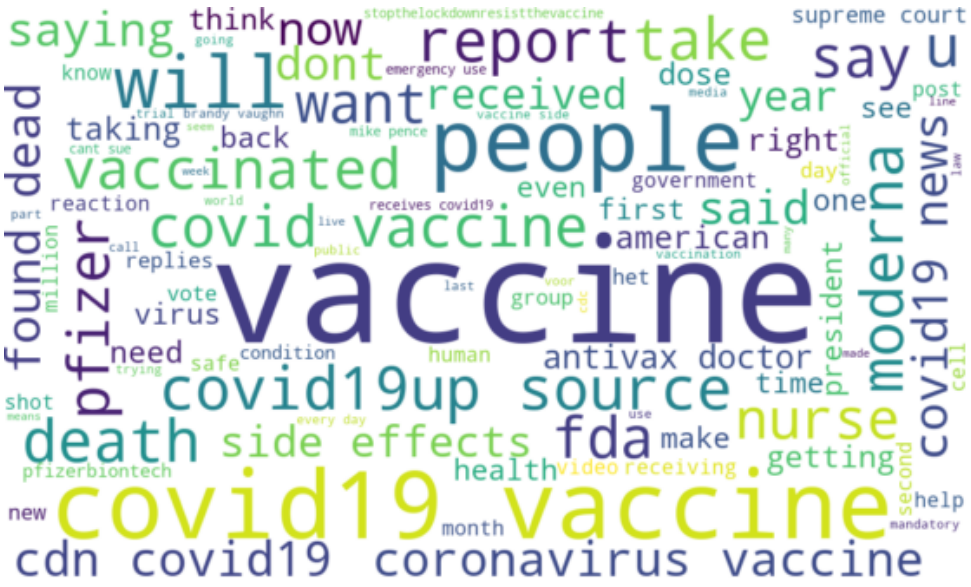


Figure 4.9. Word Cloud for Vaccine Term Messages: Created between 2020-12-18 13:13:15 to 2020-12-19 13:13:15, server time.

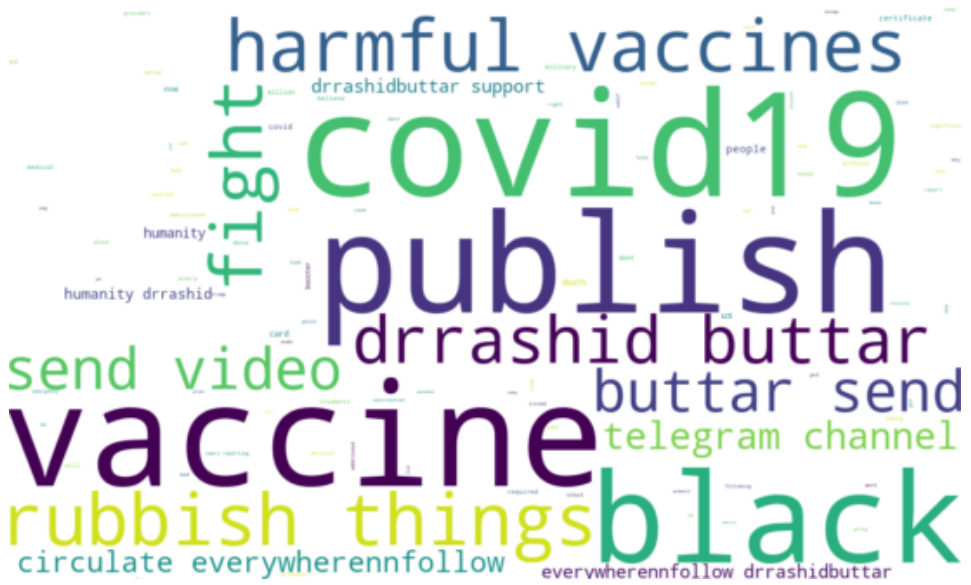


Figure 4.10. Word Cloud for Peak in Vaccine Term Usage rate from 2021-12-28 22:29:30 to 2021-12-29 22:29:30.

Analyzing the first word cloud, Fig 4.9: The words contained in the word cloud appear to be general vaccine and COVID-19 words. On the left, we notice *'found dead.'* This brings our

attention to a less frequent but specific name *'brandy vaughn'* in the word cloud. Brandy Vaughan was an anti-vaccine activist who died in December 2020, after which there was much speculation about the cause of her passing (Evon 2020). We also notice a trending hashtag, *'stopthelockdownresistthevaccine.'* These two items are possible causes of this peak. Finally, the build-up of vaccine term usage might be attributed to Pfizer submitting a Vaccine Clearance Request to the FDA in November 2020 (Park 2020).

The second peak's word cloud, shown in Fig 4.10, is much different. This one is sparse but contained multiple mentions of Dr. Rashid Buttar (*'drrashid buttar'*, *'buttar'*), a medical doctor who is a prolific personality in the anti-vaccine and conspiracy communities (Srikanth 2021). This peak could result from a viral story from a released statement or article about Dr. Rashid Buttar.

Next, we inspect the mention rate over time of *Alternative Medicine* terms as shown in Fig 4.11.

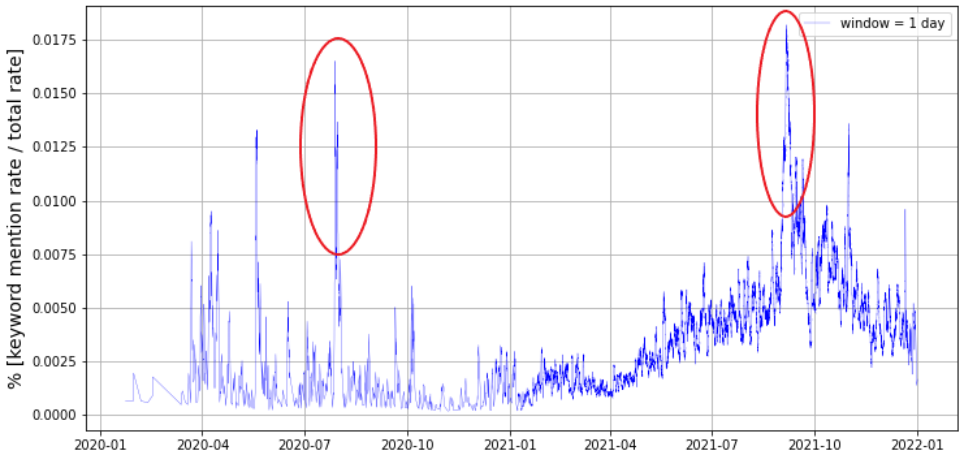


Figure 4.11. Alternative Medicine Term Mention Rate Over Time: We notice two prominent peaks in the mention rate. The overlayed red ellipses highly the peaks which we selected to analyze. The first is on December 19, 2020, and has a sharp peak that quickly appears. The second occurs on December 29, 2021 and occurs after a gradual increase in mention rate.

We notice a few prominent peaks in 2020, and decide to select one from this period, which occurs on July 28, 2020. In 2021, we identify a peak on September 5, 2021. We select a 24 hour window centered on each peak, select the messages containing our *Alternative*

Medicine terms and generate the word clouds in Figure 4.12 and Figure 4.13.

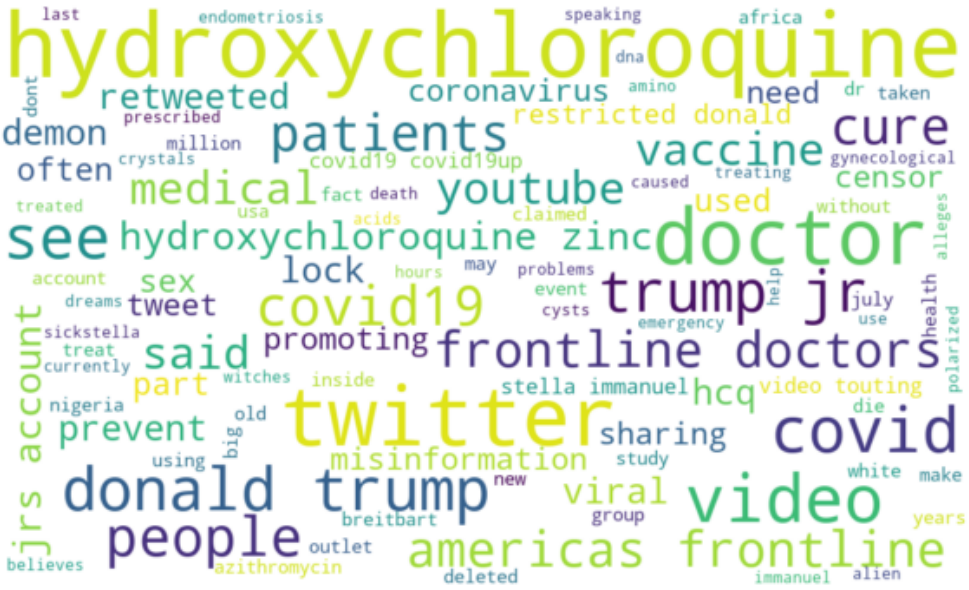


Figure 4.12. Word Cloud for Peak in Alternative Medicine Term Usage Rate from 2020-07-28 08:22:33 2020-07-29 08:22:33

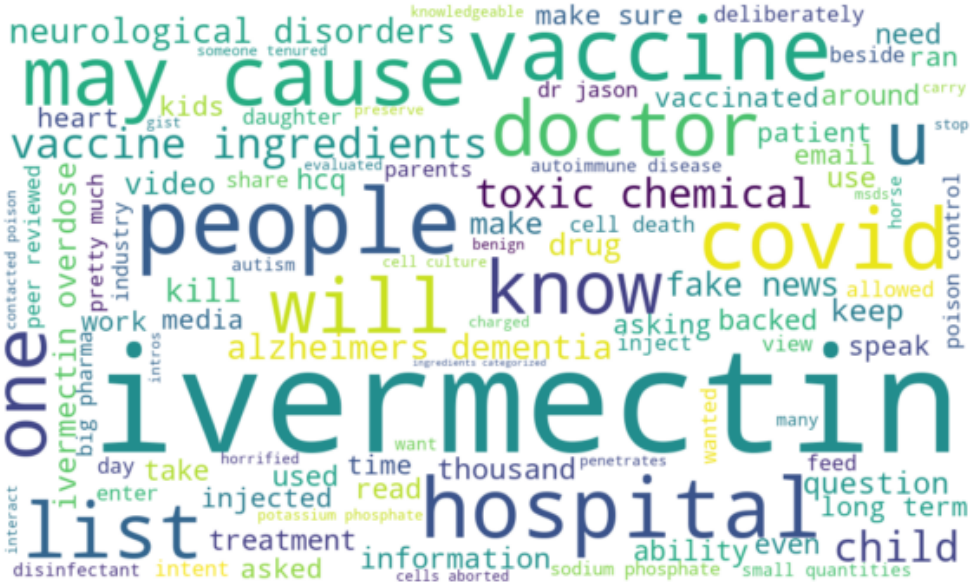


Figure 4.13. Word cloud for Peak in Alternative Medicine Term Usage Rate from 2021-09-05 04:48:01 2021-09-06 04:48:01

The word cloud in Fig 4.12 has a strong emphasis on *'hydroxychloroquine.'* We also see some interesting terms: *'cure'*, *'demon'*, *'dreams'*, *'sickstella'*, *'misinformation'*, and *'stella immanuel.'* Together, these terms lead us to a story on July 29, 2020, where President Trump retweeted a video of Dr. Stella Immanuel, a physician promoting hydroxychloroquine and a cure to COVID-19 (Stracqualursi 2020).

In Fig 4.13, the conversation has shifted to ivermectin, another alleged alternative treatment for COVID-19. Interestingly, we see mentions of *'overdose'* and *'toxic chemical.'* Earlier in the week when this peak occurs, news stories started to be published describes a surge in reports of ivermectin overdoses resulting from self-medication with the anti-parasite drug (Romo 2021).

Through analysis of trending terms and the word clouds that highlight stories that are growing within these topics, we can get contextual insight to the types of stories are being shared in the communities of Telegram.

4.3.2 Correlating Keywords

After calculating all pairwise Spearman correlation coefficients between keywords, we generate a graph network with keyword pairs with a correlation coefficient greater than or equal to 0.7. Then we use spectral clustering to partition the graph into eight communities. Figure 4.14 is the resulting graph network where each cluster is a subgraph with a unique color assignment. We then select a few of the subgraphs and plot them separately in Figures 4.15.

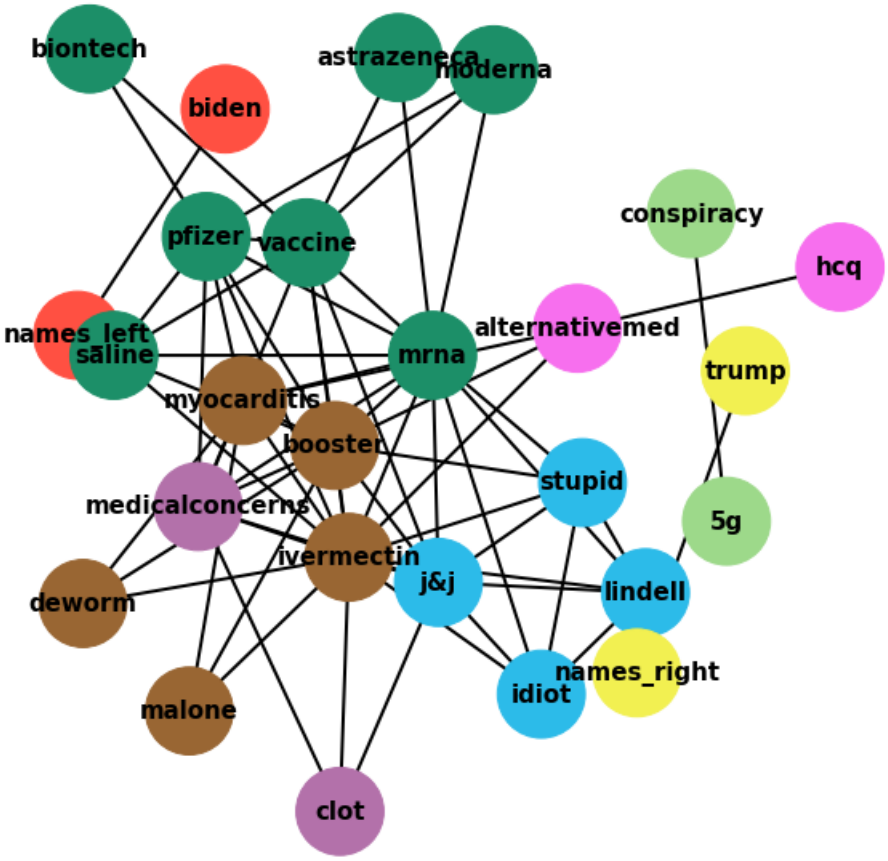
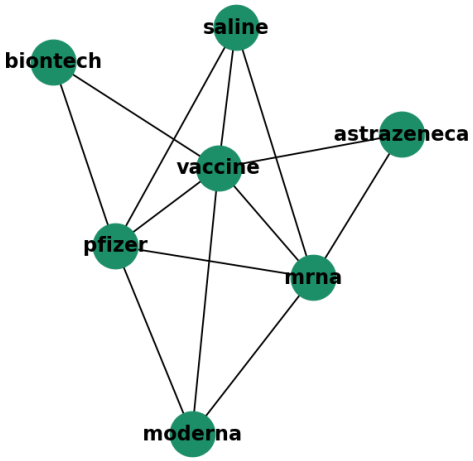
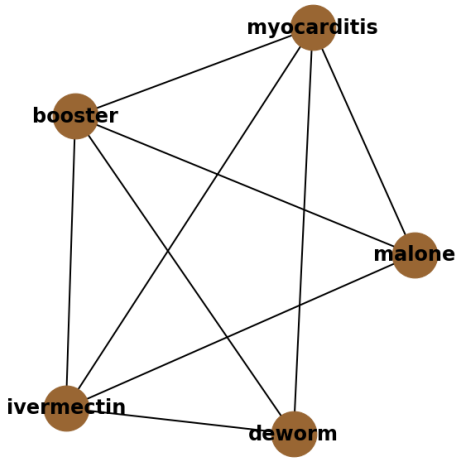


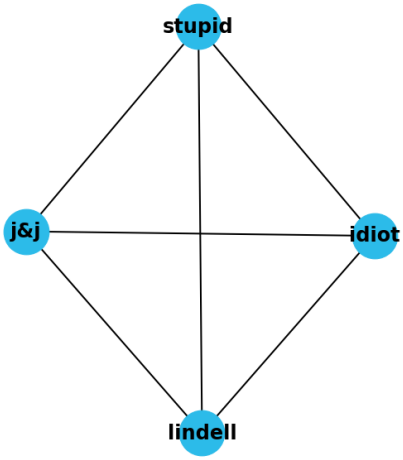
Figure 4.14. Keyword Time-Series Correlation Graph with Spectral Partitioning: We use keywords as nodes. When the correlation coefficient between two keyword usage rates is greater than or equal to 0.7, we create undirected edges between their corresponding nodes. The color coding signifies the subgraphs that result from spectral partitioning.



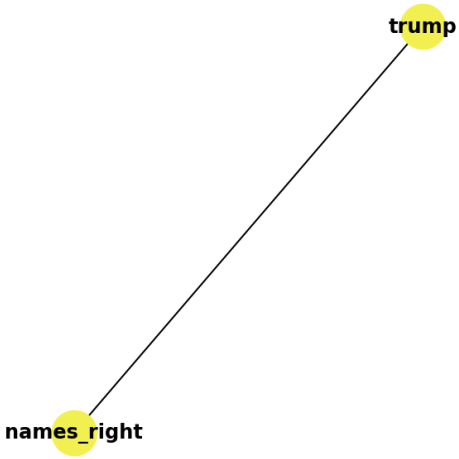
(a) Vaccine Subgraph



(b) Anti-Vaccine Subgraph



(c) Mixed Subgraph



(d) Political Right Names Subgraph

Figure 4.15. Spectral Subgraphs: We plot separate subgraphs from Fig 4.14. (a) displays a group of vaccine terms that clustered together. (b) appears to be anti-vaccination narratives. (c) is an example of mixed terms that is hard to interpret. (d) shows a set of terms, *'names_right'*, connected with one of its member terms, *'trump.'*

In Figure 4.15a, we see a collection of terms clearly related to vaccines, with vaccine brands, *'mrna'* and *'saline.'*

In Figure 4.15b, we see terms about COVID-19, but perhaps more associated with anti-vaccination and alternatives to vaccination. Myocarditis seems to be a health concern with both COVID-19 infections and vaccination side effects (Boehmer et al. (2021), Centers for Disease Control and Prevention (2022)). *'Malone'* is a reference to Robert Malone, a vaccine scientist who has helped spread vaccine hesitation (Bartlett 2021). This cluster seems to be highlighting trends of anti-vaccine narratives growing together.

Figure 4.15c is an example of a problematic subgraph to interpret and use to gain insight. It is possible that Mike Lindell used these terms together and users were quoting him, or users could be using these terms in reaction to news stories involving Mike Lindell.

In Figure 4.15d we only have two terms. *'Names_right'* is a collection of right-leaning political figures that include *'trump.'* Interestingly no other names are connected to it, suggesting Trump's name takes up most of the keyword group mentions, and thus the two have high correlations.

4.4 Sentiment

In this section, we analyze the results of hypothesis testing to find out if keyword usage has a significant impact on the sentiment of messages. We also compare the sentiments of messages that contain keywords.

Using the t-test and a significance level of 0.01, we test the null hypothesis that the mean sentiment of messages that *contain* topic keywords is equal to the mean sentiment of messages that *do not contain* those topic keywords. We assume independence of the two groups and the sets are from normally distributed sentiment scores. We do not assume that the variance is the same in the two groups, thus we use the Welch t-test. We also perform multiple hypothesis testing to test if the mean sentiments of the topical groups are statistically different from each other using the Holm-Bonferroni correction. The results are summarized in Table 4.2, in descending order.

Table 4.2. Keyword Mean Sentiment and Hypothesis Testing Values: The mean sentiment for all messages that contained keywords from each topic and the mean sentiment for the messages that do not contain those keywords. The test statistic and p-value from the t-test of the null hypothesis that the mean sentiments are equal are displayed. Lastly, the corrected p-value from multiple hypothesis testing is included.

	(Messages with Keywords)		**(Messages without keywords)**				
Keyword Group Topic	**Mean Sentiment**	**# Messages**	**Mean Sentiment**	**# Messages**	**t-stat**	**P-Value**	**P-Value (Corrected)**
Conspiracies	2.801	115.8K	2.904	29037.7K	40.774	< .0001	< .0001
Political Right	2.664	905.9K	2.911	28248.2K	242.119	< .0001	< .0001
Alternative Medicine	2.652	136.6K	2.905	29017.0K	106.240	< .0001	< .0001
Political Left	2.455	776.4K	2.916	28377.5K	485.085	< .0001	< .0001
Masks	2.448	345.3K	2.909	28808.3K	309.442	< .0001	< .0001
COVID-19/Virus	2.433	126.9K	2.925	27884.8K	692.780	< .0001	< .0001
Vaccines	2.423	1116.9K	2.922	28037.1K	642.814	< .00010	< .0001
Medical Concerns	2.388	137.2K	2.906	29016.4K	243.106	< .0001	< .0001

The t-statistic are all very large, and the corresponding P-Values are all numerically zero, indicating we reject the null hypothesis in all cases. The mean sentiments of messages containing the topic keyword sets are not equal to the mean sentiment of messages that do not contain the keywords. We notice that all topics had lower sentiments than the average messages that did not include our topic terms. It suggests that even though these groups are discussing topics they may view positively, such as alternative treatments for COVID-19, it is still within a negative context centered around COVID-19.

The Holm-Bonferroni corrected P-Values are also zero for all keyword group topics, meaning they are all distinct in their mean sentiments. Note Table 4.2 is sorted with the mean sentiment of messages with keywords in descending order. The dataset tends to speak most positively about conspiracies. There is a drop in sentiment, and then we see that conversations involving the names of people associated with the political right are the next highest, with alternative medicine terms lagging close behind. Then there is another significant drop in sentiment, and we see people associated with the political left. The entire group of COVID-19 terms, vaccines, masks, and lastly, side effects and medical concerns are further below them.

The mean sentiments for messages containing and not containing each keyword topic from Table 4.2 are visually represented in bar plots with 95% confidence intervals in Figure 4.16.

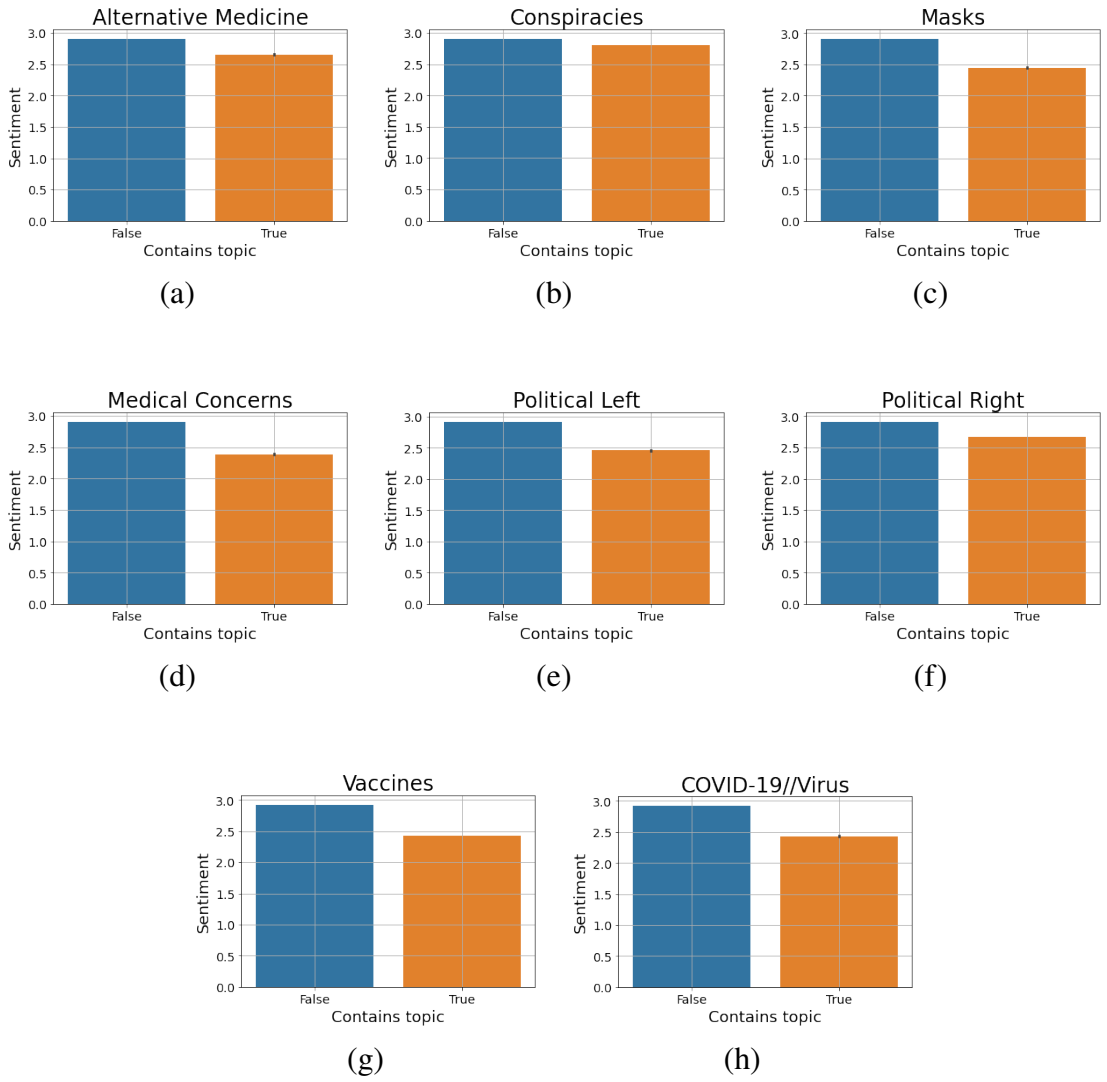


Figure 4.16. Sentiment Bar Plots by Keyword Group Topic: Mean sentiments for all messages that contain each keyword topic are displayed in bar plots with 95% confidence intervals, and compared to mean sentiments of all messages that do not contain each keyword topic. Note that the confidence intervals are small and may not be visible due to the number of observations being very large for each topic; between 115k to 29M observations.

CHAPTER 5:
Discussion

This chapter will summarize the analysis and discuss our conclusions. We also discuss the limitations of our study and dataset and suggest future research using Telegram data.

5.1 Summary

This thesis explores Telegram communities by creating a dataset of messages from highly connected channels and groups and using it to answer research questions.

What are the primary topics discussed in our Telegram dataset?
Using topic modeling, we identify and inspect the general topics being discussed, which highlighted that the channels and groups are not only connected through URL references, but they are having overlapping conversations about COVID-19 and vaccines. We discovered three primary topics with many overlapping terminologies about COVID-19 & vaccines, Politics, and Religion. We also found distinct topics consisting of general conversational terms and foreign languages.

Can we identify communities based on the topics they are discussing?
Using clustering, we separate the channels and groups into three distinguished clusters based on the terminology used in their discussions. We show the overlap of the conversations occurring between the clusters. We also show that the clusters were divided between COVID-19 & vaccines, politics, and conversation topics.

When certain topics are trending, what is the driving factor for its prevalence?
Using time series mention rates of selected keywords, we identify and analyze micro trends and determine the potential real-world reasons for these trends. We are able to identify the date and time of peaks, inspect the terms being used, and use them to reference news articles to dissect the discussed topic.

Do we see trends between topics that indicate they're developing together?
We then leverage the same time-series keyword usage rates to identify correlating topics that appear to be related narratives. We identify clusters of interrelated terms that highlight vaccine and anti-vaccine discussions that may suggest the member terms are used in discussing a common narrative.

When topics are being discussed, is it due to a polarized reaction? If so, what is the opinion expressed around these topics?
Lastly, we use sentiment analysis to highlight how specific topics are viewed and discussed by the groups. We show the sentiments of messages about particular topics are different than messages not about those topics. We also show the differences in sentiment between topics. For example, the highest sentiment topics are conspiracies and alternative treatments for COVID-19. Then, after a drop in sentiment, we see mask terms, COVID-19 virus, vaccines, and medical concerns associated with COVID-19 and vaccines. Politically, we see names associated with right-wing politics occur with higher sentiment than left-wing politics.

5.2 Conclusion

Our topic modeling and clustering analyses show tight-knit communities having three overlapping conversations that are hard to distinguish. Further, since our data collection methods only use URL references in messages to crawl Telegram, users would be exposed to the same conversations and groups we discovered. As a result, our data consists of many groups having one conversation at the intersection of politics and COVID-19. This can be problematic since it can catch users in a loop where they are only exposed to other bias-confirming echo chambers where they are unlikely to be presented with opposing viewpoints or information. This creates a good opportunity for bad actors - a vulnerable and isolated audience.

While respecting users' right to free speech, we are faced with the duty to protect our national interests and national security from external actors who may prey on these communities. In the absence of misinformation policies and standards enforced through moderation by the social media platform, we must develop methods to prevent the weaponized use of misinformation.

The first step must be to identify places where misleading information exists and can grow. This study highlights Telegram communities are a potential home for targeted misinformation campaigns.

Educating users is one way to prevent them from being tricked by misinformation. The primary method used by social media platforms is to detect and append warning labels to potential misinformation. The problem with this approach is that the absence of a misinformation label creates an implied truth effect for any information not labeled with a warning (Pennycook et al. 2019a). Since it is impossible to detect and correctly label every instance of misinformation, content labeling methods will always risk producing the implied truth effect. Another way of combating misinformation is to prime users to be more attentive to the accuracy of information. By having users rate the integrity of headlines, Pennycook et al. were able to shift user's focus to the accuracy of the information and have a lasting impact on the quality of information these users shared (2019b).

Another approach is to influence users with nudging techniques. The least effective method of affecting users is through conflict. Suppose we directly challenge a user's beliefs or push narratives in stark conflict with their beliefs. In that case, it causes a backfire effect where subjects become further entrenched in their belief system and harder to reach. An influence model presented by Yang et al. begins with identifying the topic central to the user's rhetoric, mimicking or "pacing" their rhetoric, and slowly shifting to opposing viewpoints, thus "leading" them to alternatives to their predisposed beliefs (2020).

Vendeville et al. tackles the task by determining a user's apparent preferences and introducing more diverse content than they would otherwise be exposed (2022). This again follows the theme of analyzing user preferences or beliefs and guiding them to more moderate stances.

These methods are all intended to mitigate the effects of echo chambers and misinformation. There is also the possibility that we can draw people away from the polarized narratives that are propagating in polarized communities. One method could be providing more opportunities to exit the echo chamber by introducing pathways out. In our dataset, we saw that most paths in these groups merely lead to more of the same conversations. Suppose we insert URL references and invitations that lead to more moderate conversations about similar

topics or simply more diverse groups where they might encounter more varied sources. In that case, we may be able to expose users to less biased information and thus empower them with a more diverse set of opinions and a better reference point of what is true. If we believe misinformation spreads organically, then it is possible that if we can lead users to factual information, these users could, in turn, organically influence the nature of the conversations in which they participate in the future.

In summary, we must identify misinformation echo chambers, acknowledge users' root concerns, empower them to question the validity of false narratives, expose them to a broader array of information and try to lead out of these bubbles.

5.3 Limitations & Future Research Opportunities

The most apparent limitation of this study is the composition and size of the dataset. Since our seed channels were primarily centered on COVID-19 & vaccine themes that were the focus of our research, the resulting dataset is heavily weighted toward these conversations. Collecting more data from a greater variety of channels would provide greater opportunities for developing methods to analyze groups and identify at-risk groups and echo chamber effects. For example, analysts could use it to compare the structure and behavior of echo chamber clusters to more diverse clusters, which would aid in developing detection methods of potential echo chambers in raw Telegram data.

Another consideration is that we do not attempt to detect misinformation or validate the information in our dataset. Classification of information validity is beyond the scope of this thesis. This analysis is primarily motivated by the developing COVID-19 narratives coupled with Telegram's unique structure and policies, which could result in interesting behavior of communities on the platform. We conduct our analysis of communities centered around COVID-19 conversations due to the claimed prevalence of misinformation within related narratives.

Perhaps the most crucial opportunity would be to develop detection techniques for foreign-led groups that generate and spread targeted misinformation to disrupt our national interests. We know a primary method of these groups is to create and deploy automated bot accounts as a cost-effective means of spreading misinformation on Twitter (Badawy et al. 2018). It

stands to reason that bots could also be used for the same purposes on Telegram, so bot detection techniques should be developed and tested on Telegram data.

Use more metadata available from Telegram, which could provide an opportunity to analyze user membership across channels belonging to echo chamber clusters. Does being a member in more or fewer channels of an echo chamber cluster correlate with measurable differences in user sentiment? If we can compare the messaging of prolific influencers on Telegram to their social media posts on Twitter, are there quantifiable differences, and what does it suggest about the nature of conversations on both platforms? The answers could give insight into the effects of content moderation on sentiment and the topics discussed.

Lastly, Telegram could be an opportunity to build upon of the research developed by Pennycook et al., Yang et al., and Vendeville et al., which we discuss in Section 5.2. Developing and testing nudge techniques could disrupt the cycle of misinformation propagation or gently guide users out of the bubbles where they're stuck.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX:

A.1 Dataset Channels

Table A.1. COVID-19 and Vaccine Groups and Channels

URL
https://t.me/dailyexposenews
https://t.me/COVID1984chat
https://t.me/unvaxedlivesmatter
https://t.me/teenvaccineadversereactions
https://t.me/CovidRedPills
https://t.me/The_Library_II
https://t.me/cv19vaccine_reactions
https://t.me/covid_vaccine_injuries
https://t.me/madmixconspiraciesgroup
https://t.me/worlddoctorsalliance
https://t.me/astandintheparkbracknell
https://t.me/nocovidvaccines

Table A.2. Other Groups and Channels

URL
https://t.me/BLMProtests
https://t.me/Python
https://t.me/theprogrammingart
https://t.me/interestingAsFuck_tg
https://t.me/science
https://t.me/DeFi_ICO_Invest
https://t.me/GameFi_Launchpad
https://t.me/wildlifen
https://t.me/joinchat/PwLkvLDQacFCY7Sn
https://t.me/Manchester_city_cf
https://t.me/goal_sport_football
https://t.me/beginnersfitness
https://t.me/IMQuotes_Videos
https://t.me/Planet_Earth
https://t.me/askmenow

List of References

Anglano C, Canonico M, Guazzone M (2017) Forensic analysis of Telegram Messenger on Android smartphones. *Digital Investigation* URL http://dx.doi.org/10.1016/j.diin.2017.09.002.

Badawy A, Ferrara E, Lerman K (2018) Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE), URL http://dx.doi.org/10.1109/asonam.2018.8508646.

Bartlett T (2021) The Vaccine Scientist Spreading Vaccine Misinformation. Accessed August 12, 2022, https://www.theatlantic.com/science/archive/2021/08/robert-malone-vaccine-inventor-vaccine-skeptic/619734/.

Berger JM, Morgan J (2015) The ISIS Twitter census: defining and describing the population of ISIS supporters on Twitter.

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3:993–1022.

Boehmer T, Kompaniyets L, Lavery ea AM (2021) Association Between COVID-19 and Myocarditis Using Hospital-Based Administrative Data — United States, March 2020–January 2021. *MMWR Morb Mortal Wkly Rep 2021;70:1228–1232* DOI: http://dx.doi.org/10.15585/mmwr.mm7035e5s.

Bond S (2021) Unwelcome On Facebook And Twitter, QAnon Followers Flock To Fringe Sites (January 31), https://www.npr.org/2021/01/31/962104747/unwelcome-on-facebook-twitter-qanon-followers-flock-to-fringe-sites.

Centers for Disease Control and Prevention (2022) Myocarditis and Pericarditis After mRNA COVID-19 Vaccination. Accessed August 12, 2022, https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/myocarditis.html.

Cinelli M, Morales GDF, Galeazzi A, Quattrociocchi W, Starnini M (2021) The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118(9):e2023301118, URL http://dx.doi.org/10.1073/pnas.2023301118.

Dargahi Nobari A, Reshadatmand N, Neshati M (2017) Analysis of Telegram, An Instant Messaging Service. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2035–2038, CIKM '17 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450349185, URL http://dx.doi.org/10.1145/3132847.3133132.

Dinakar K, Reichart R, Lieberman H (2021) Modeling the Detection of Textual Cyberbullying. *Proceedings of the International AAAI Conference on Web and Social Media* 5(3):11–17, URL https://ojs.aaai.org/index.php/ICWSM/article/view/14209.

Evon D (2020) Brandy Vaughan Died of Natural Causes, Say Police. Snopes. Accessed August 12, 2022, https://www.snopes.com/news/2020/12/22/brandy-vaughan-death/.

Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The Rise of Social Bots. *Commun. ACM* 59(7):96–104, ISSN 0001-0782, URL http://dx.doi.org/10.1145/2818717.

Grootendorst M (2021) 9 Distance Measures in Data Science. Towards Data Science. Accessed May 20, 2022, https://towardsdatascience.com/9-distance-measures-in-data-science-918109d069fa.

Hart B (2022) Poorly Conceived Biden Disinformation Board Put on Pause (May 18), https://nymag.com/intelligencer/2022/05/poorly-conceived-biden-disinformation-board-put-on-pause.html.

Hoseini M, Melo P, Benevenuto F, Feldmann A, Zannettou S (2021) On the Globalization of the QAnon Conspiracy Theory Through Telegram. URL http://dx.doi.org/10.48550/ARXIV.2105.13020.

Jagota A (2020) Text Sentiment Analysis in NLP. Towards Data Science. Accessed July 2, 2022, https://towardsdatascience.com/text-sentiment-analysis-in-nlp-ce6baba6d466.

La Morgia M, Mei A, Mongardini AM, Wu J (2021) Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. URL http://dx.doi.org/10.48550/ARXIV.2111.13530.

Lee K, Eoff B, Caverlee J (2021) Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* 5(1):185–192, URL https://ojs.aaai.org/index.php/ICWSM/article/view/14106.

Londono E, Milhorance F, Nicas J (2021) Brazil's Far-Right Disinformation Pushers Find a Safe Space on Telegram (November 11), https://www.nytimes.com/2021/11/08/world/americas/brazil-telegram-disinformation.html.

McInnes L, Healy J, Melville J (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. URL http://dx.doi.org/10.48550/ARXIV.1802.03426.

Mosleh M, Martel C, Eckles D, Rand D (2021) Perverse Downstream Consequences of Debunking: Being Corrected by Another User for Posting False Political News Increases Subsequent Sharing of Low Quality, Partisan, and Toxic Content in a Twitter Field Experiment. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450380966, URL http://dx.doi.org/10.1145/3411764.3445642.

Mysiak K (2020) Explaining K-Means Clustering. Towards Data Science. Accessed May 20, 2022, https://towardsdatascience.com/explaining-k-means-clustering-5298dc47bad6.

Newman MEJ (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103(23):8577–8582, URL http://dx.doi.org/10.1073/pnas.0601602103.

NLP Town (2022) nlptown/bert-base-multilingual-uncased-sentiment. Accessed February 1, 2022 https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment.

Park A (2020) Exclusive: Pfizer CEO Discusses Submitting the First COVID-19 Vaccine Clearance Request to the FDA. Time. Accessed August 12, 2022, https://web.archive.org/web/20210129054951/https://time.com/5914139/pfizer-covid-19-vaccine-fda-authorization/.

Pennycook G, Bear A, Collins E (2019a) The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science* URL http://dx.doi.org/10.1287/mnsc.2019.3478.

Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG (2019b) Shifting attention to accuracy can reduce misinformation online. URL http://dx.doi.org/10.31234/osf.io/3n9u8.

Pennycook G, Rand D (2021) Examining false beliefs about voter fraud in the wake of the 2020 Presidential Election. URL http://dx.doi.org/10.31234/osf.io/szdgb.

Rehurek R (2010) gensim – Topic Modelling in Python, version 4.1.2. Accessed March 20, 2022. https://radimrehurek.com/gensim/intro.html.

Romo V (2021) Poison Control Centers Are Fielding A Surge Of Ivermectin Overdose Calls. NPR. Accessed August 12, 2022, https://www.npr.org/sections/coronavirus-live-updates/2021/09/04/1034217306/ivermectin-overdose-exposure-cases-poison-control-centers.

Siervert C, Shirley K (2015) pyLDAvis, version 2.1.2. Accessed March 20, 2022. https://pyldavis.readthedocs.io/en/latest/readme.html.

Srikanth A (2021) 12 prominent people opposed to vaccines are responsible for two-thirds of anti-vaccine content online: report. The Hill. Accessed August 12, 2022, https://thehill.com/changing-america/well-being/prevention-cures/544712-twelve-anti-vaxxers-are-responsible-for-two/.

Stracqualursi V (2020) Trump promotes a doctor who has claimed alien DNA was used in medical treatments. CNN. Accessed August 12, 2022, https://www.cnn.com/2020/07/29/politics/stella-immanuel-trump-doctor/index.html.

Sureka A, Agarwal S (2014) Learning to Classify Hate and Extremism Promoting Tweets. *2014 IEEE Joint Intelligence and Security Informatics Conference* 320–320.

Telegram (2022a) Telegram FAQ. Accessed February 13, 2022, https://telegram.org/faq.

Telegram (2022b) Telegram Privacy Policy. Accessed February 13, 2022, https://telegram.org/privacy.

Telegram Channels (2021) 21000+ Telegram Channels, Groups, Bots and Stickers List. Accessed December 15, 2021, https://telegramchannels.me/.

Telegram Directory (2021) Find Telegram Channels, Groups and Bots. Accessed December 15, 2021, https://tdirectory.me/.

TGStat (2021) Telegram channels and groups catalog — World — TGStat. Accessed December 15, 2021 https://tgstat.com/.

Twitter (2022) Rules and Policies. Accessed August 7, 2022, https://help.twitter.com/en/rules-and-policies.

Twitter Public Policy (2018) Update on Twitter's review of the 2016 U.S. election. Accessed January 4, 2022, https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html.

Vendeville A, Giovanidis A, Papanastasiou E, Guedj B (2022) Opening up echo chambers via optimal content recommendation. URL http://dx.doi.org/10.48550/ARXIV.2206.03859.

von Luxburg U (2007) A Tutorial on Spectral Clustering URL http://dx.doi.org/10.48550/ARXIV.0711.0189.

Yang Q, Qureshi K, Zaman T (2020) Mitigating the Backfire Effect Using Pacing and Leading. URL http://dx.doi.org/10.48550/ARXIV.2008.00049.

Yousefi A (2019) How to Get Data From Telegram Using Python. Better Programming. Accessed December 10, 2021, https://betterprogramming.pub/how-to-get-data-from-telegram-82af55268a4b.

Yse DL (2019) Your Guide to Natural Language Processing (NLP). Towards Data Science. Accessed July 2, 2022, https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1.

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

2. Defense Technical Information Center
   Ft. Belvoir, Virginia