Journal of Computational Social Science
https://doi.org/10.1007/s42001-024-00253-y

RESEARCH ARTICLE

An exploratory analysis of COVID bot vs human disinformation dissemination stemming from the Disinformation Dozen on Telegram

Lynnette Hui Xian Ng[1] · Ian Kloo[1] · Samantha Clark[1] · Kathleen M. Carley[1]

Received: 7 November 2023 / Accepted: 1 February 2024

Abstract

The COVID-19 pandemic of 2021 led to a worldwide health crisis that was accompanied by an infodemic. A group of 12 social media personalities, dubbed the "Disinformation Dozen", were identified as key in spreading disinformation regarding the COVID-19 virus, treatments, and vaccines. This study focuses on the spread of disinformation propagated by this group on Telegram, a mobile messaging and social media platform. After segregating users into three groups—the Disinformation Dozen, bots, and humans, we perform an investigation with a dataset of Telegram messages from January to June 2023, comparatively analyzing temporal, topical, and network features. We observe that the Disinformation Dozen are highly involved in the initial dissemination of disinformation but are not the main drivers of the propagation of disinformation. Bot users are extremely active in conversation threads, while human users are active propagators of information, disseminating posts between Telegram channels through the forwarding mechanism.

Keywords Telegram · COVID-19 · Disinformation · Network analysis · Temporal analysis · BERT · GPT

✉ Lynnette Hui Xian Ng
huixiann@andrew.cmu.edu

Ian Kloo
ipk@andrew.cmu.edu

Samantha Clark
samanthc@andrew.cmu.edu

Kathleen M. Carley
carley@andrew.cmu.edu

[1] Center for Informed Democracy and Social-Cybersecurity (IDeaS), Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

Published online: 14 March 2024

Introduction

The Coronavirus Pandemic of 2020 led to a worldwide health crisis. In a world physically interconnected via air, land, and sea travel, the virus spread across international borders exceptionally quickly. To minimize human-to-human contact and mitigate the spread of the virus, governments closed borders and enforced lockdowns where people were only allowed to leave their homes for essential tasks. In late 2020, several pharmaceutical companies developed vaccines that were effective at preventing and reducing the symptoms of COVID-19 infections, and in December 2020, the worldwide vaccination campaign began.

As the virus spread, so did many mis/disinformation narratives. These included theories that COVID-19 was a bioweapon built for the destruction of mankind, that the 5 G mobile network caused COVID-19, or that the threat posed by COVID-19 was exaggerated [1, 2]. Research has shown that the foundations of these disinformation narratives are psychological predispositions to reject information coming from authority figures and a propensity to view major events as a product of conspiracy, partisan, and ideological motivations [2]. The large amount of disinformation spread during the COVID-19 pandemic became known as an “infodemic.”

After investigating the sources of COVID-19 disinformation, the Center for Countering Digital Hate (CCDH) identified 12 people as the most responsible for propagating these narratives. They dubbed this group the “Disinformation Dozen” [3]. From over 800,000 posts extracted from Facebook and Twitter in early 2021, the CCDH reported that at least 65% of anti-vaccine content could be attributed to the Disinformation Dozen. These users were highly visible, with millions of followers, and were termed “repeat offenders” of disinformation spread.

The report produced by CCDH analyzed the influence of the Disinformation Dozen on traditional social media platforms like Facebook, Twitter, and Instagram, observing that they had high influence because of their voluminous posts and enormous following. In terms of information propagation, two main groups of users have been observed to share the content generated by the Disinformation Dozen on Twitter: low-credibility media outlets and politically linked accounts [4]. In this study, we extend the investigation of information dissemination from social media platforms to mobile media platforms, in particular, the Telegram messaging application.

While the Disinformation Dozen’s activities were initially identified on Facebook, Twitter, and Instagram, these platforms each made efforts to combat COVID-19 disinformation during the pandemic. This censorship ranged from flagging potentially misleading posts to platform bans for certain accounts [5]. While some of these interventions were effective, they also served to push spreaders of disinformation onto other platforms. Telegram is a messaging service with social media functionality that only censors content and users in the most extreme situations (e.g., removing content that is explicitly illegal). We found that 8 of the 12 Disinformation Dozen were still active on Telegram in early 2023, so we sought to investigate the disinformation networks surrounding these accounts on the platform.

Telegram has become a popular mobile messaging platform, with at least 700 million monthly active users worldwide [6]. Its large user base provides ripe ground for COVID-19 discourse. Emerging disinformation stories on Telegram have been found by identifying skepticism of claims and co-occurrence with colloquial uses of government acts enacted to combat online falsehoods [7]. Many Telegram channels have been discovered to be aggregating conspiracy theories surrounding COVID-19, evolving these theories as the channels develop, and spreading distrust in governments [8]. With the large number of coronavirus-related information circulating on Telegram, the messaging app is a huge source of disinformation transmission [9], and therefore we leverage this wealth of information in our study.

In this paper, we investigated the disinformation spread by the Disinformation Dozen on the Telegram messaging application. From the Telegram channels hosted by the Disinformation Dozen, we collected channel messages and information about forwarded and replied messages. We analyzed the temporal, topic, and network interactions among three different sets of users: the Disinformation Dozen, bot users, and regular human users, generating insights into the nature of information spread through Telegram interactions. Our key findings suggest that while the Disinformation Dozen may be the initial seed in setting up Telegram channels and initiating information spread, bots are extremely active in generating engagement through messaging replies, while humans facilitate the spread of disinformation to other channels through the forwarding mechanic.

Research questions and contributions

We ask the following Research Questions with respect to the dissemination of disinformation that stemmed from the Disinformation Dozen on Telegram:

1. RQ1: How do the temporal trends of message posting patterns differ from the Disinformation Dozen, Bots, and Human users?
2. RQ2: How do the topics posted within messages differ for the Disinformation Dozen, Bots, and Human users?
3. RQ3: How does disinformation begin spreading from the Disinformation Dozen throughout the broader Telegram network?

Drawing data from the Telegram messaging application, we analyzed the activity of disinformation dissemination that stemmed from the Disinformation Dozen.

The main contributions of our work can be summarized as follows:

1. We gathered a novel dataset that encompasses the messages published on Telegram by the Disinformation Dozen, the messages within the channels that they control, and the messages of the channels from which the original messages were forwarded.
2. We investigated the patterns of information dissemination within the Telegram messenger application through temporal, topical, and network analysis. We

uncovered distinct traits in terms of messaging patterns and dissemination across the Disinformation Dozen, bot users, and human users on Telegram.

3. We adapted a current bot detection algorithm for Telegram, designing and validating a Telegram bot detection algorithm that works with an accuracy of 72% through manual verification. We also summarized the characteristics of bot-like users through our observations from the manual annotation of Telegram users.

Structure of the paper

The structure of the paper is as follows. Having established an introduction and our research questions in "Introduction" and "Research questions and contributions", "Related work" provides a literature review of the related work pertaining to Telegram, pandemics and disinformation dissemination. "Data collection and processing" describes our technique for data collection from Telegram and establishes the terminology used in this paper. We segment the users into three groups—the Disinformation Dozen, bot users, and human users—and we describe the segmentation technique in "User group identification". We then describe our analytic methodology "Methodology and results", discuss the results in "Discussion", and finally state our conclusions "Conclusion".

Related work

Telegram as a mobile messaging platform

Telegram is a cloud-based messaging application. There are applications for desktop computers as well as multiple mobile platforms, such as iPhone and Android. It is known for its end-to-end chat encryption. The messaging app was founded in 2013 by Russian brothers, and today is registered as a company in the British Virgin Islands and as an LLC in Dubai. As of 2023, Telegram has more than 700 million monthly active users worldwide [6]. Telegram is not only used as a messaging application but also as a platform for news media dissemination and consumption. For example, Telegram was found to be used by state-run media outlets to disseminate Persian language news to the Iranian public from 2015 to 2020 [10].

Telegram has been given close scrutiny by researchers and academic groups because it has been identified to be one of the most influential recruitment and planning tools used by terrorists and extremism groups [11]. On top of that, Telegram is also actively used in politics. A study from 2020 reveals that Ukrainian politicians and their entourages maintained a series of Telegram channels that actively disseminated information relating to changing the political climate and discussing views on the Ukrainian government [12]. During the 2020 coronavirus pandemic, Telegram groups have been used to organize protests against control measures implemented by the German government [13] and mobilize white supremacists and hate [14].

Disinformation has been found to be present on Telegram, ranging from COVID-19 conspiracy theories to far-right disinformation topics [11], including QAnon conspiracy theories on the origins and of the coronavirus [8]. Telegram channels impersonating celebrities or well-known services have been used to spread disinformation theories, with some messages reaching up to 1 million users [15]. Unfortunately, these fake channels and the information they spread are difficult to identify even by the most media literate users, and therefore users are susceptible to believing these falsehoods as truth [15].

Telegram provides a rich source of information with its varied channels and discussions, and we harness this information to study human behavior with respect to the dissemination of coronavirus-related disinformation during the 2020 pandemic.

Social media and pandemics

During health pandemics, authorities use social media for diagnostic efforts and public communication [16] or to access the public emotional state to aid in regional-level government decisions [17]. During the 2020 coronavirus pandemic, several governments set up their own social media channels to disseminate information and debunk fake news. Some governments leveraged Telegram channels for health-related communication. For example, the government of India used the channel “MyGovCoronaNewsdesk” and the government of Singapore used the channel “govsg”.

While social media was used for crisis communication by authorities, mis- and disinformation on the coronavirus pandemic also spread quickly and wildly on these platforms. From over 10,000 falsehoods shared on several social media and mobile messaging platforms, it is clear that disinformation was a common global problem, but these narratives originated from a small number of individuals and organizations [18]. There were also state-sponsored online disinformation campaigns that undermined socio-political systems, delegitimized public health and scientific bodies, and diverted public health responses, resulting in a combined cyber and biological pandemic [19].

Part of the disinformation spread was generated by uncertainty surrounding COVID-19 during the early stages of the pandemic, which led people to come up with cures from everyday foods [1]. Conspiracy theories about the origins of the COVID-19 virus were also common. Due to the large volume of conspiracy theories, machine learning models have been constructed to systematically identify these theories at scale [20].

Social media use during pandemics cuts both ways. The simplicity and the reach of the platforms enable governments and authorities to disseminate information relating to sanitary habits, governmental measures, and vaccination availability quickly and with little effort, using media that the public can easily access and in formats that are easily consumable [21]. On the other hand, the rich- and real-time information dissemination capability of social media allows malicious actors to thrive and disseminate disinformation at scale [19]. The study of social media during pandemics is important in this digital age as a growing amount of information dissemination and consumption takes place online.

Disinformation dissemination

During and after the COVID-19 pandemic, a string of studies identified patterns of dissemination and propagation of disinformation. Past work developed a Framework of Disinformation spread, characterizing how fake news could be disseminated through the official APIs of social media platforms and methodologically classifying the spread into four phases: network creation, profiling, content generation, and information dissemination [22]. Disinformation has been found to transmit faster than real news, and some stories achieve sustainable propagation to achieve a substantially wider reach. In contrast, the volume of real news was found to drop drastically after it was initially posted [23].

To combat the propagation of disinformation, classification models have been constructed using recurrent and convolutional networks for early detection of disinformation in hopes of identifying and addressing these narratives early [24]. These fake news detection approaches, including supervised machine learning techniques, multivariate time series models, and Bayesian learning models, aid in the detection of disinformation at scale [25].

Bot accounts, or inauthentic and sometimes automated accounts, have also been observed to actively disseminate disinformation. Bot accounts exploit the online information ecosystem to sow misinformation by posting content and interacting with each other and human users through legitimate social connections [26]. A study of bot activity within the COVID-19 vaccine discourse identified that bots were "hyper-social" users that were extremely active in contributing to low-credibility content distribution [27]. Unfortunately, humans are largely unable to distinguish well designed social bots from genuine human accounts [28], because quantitatively, the features of bots and humans appear very similar [29]. Therefore, humans do unwittingly contribute to the spread of disinformation. For example, humans have been observed to retweet (i.e., share a post on X) the low-credibility content posted by bots and humans at the same rate [30].

Additionally, humans do not always react to falsehoods by debunking or investigating them. Past work shows that while humans express skepticism about disinformation posts [7], they do not always take the time to investigate or debunk posts because they are either uninterested or believe it would take too long [31]. Humans also share disinformation stories because they are sensational and exciting, thus collectively contributing to the widespread circulation of disinformation on a greater scale [32].

Collectively, bots and human accounts contribute to widespread disinformation propagation. A disinformation campaign regarding the origin of COVID-19 as a synthesized virus from a biolab was discovered to be initiated by Russian state-funded groups and was spread through coordinated inauthentic amplification alongside the support of Russia's invasion of Ukraine among Russian-speaking users. At the same time, the narratives in this campaign were also naturally

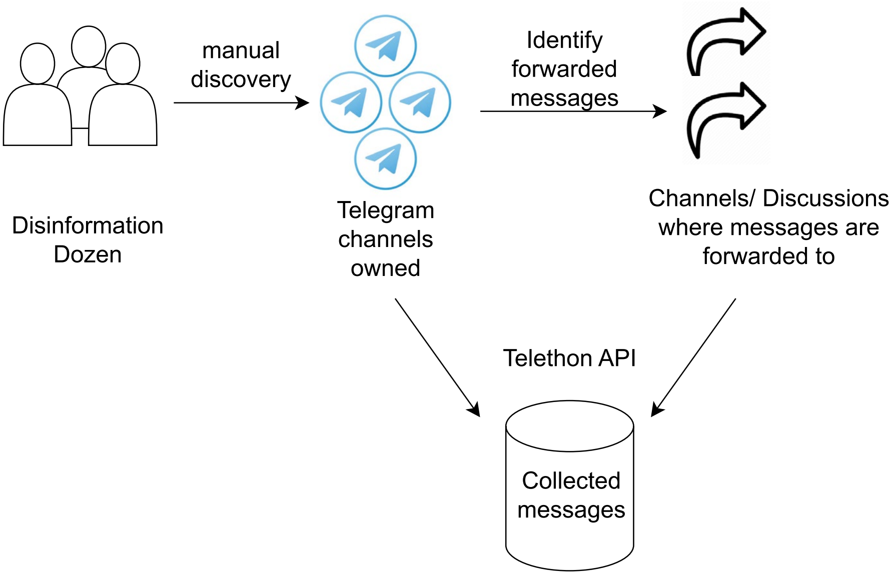


Fig. 1 Data collection pipeline

propagated by anti-vaccine and conservative English-speaking communities on X [33].

Much of the past work in this space revolves around the spread of disinformation on the platform X, previously named Twitter, and our work builds on the concepts of past work to explore the spread of disinformation on Telegram.

Data collection and processing

In this section, we describe the data collection and processing pipelines. Figure 1 illustrates our data collection pipeline.

We collected messages posted to public Telegram channels via a snowball sampling method. We first referenced the names of the Disinformation Dozen released by the Center for Countering Digital Hate [3] and manually searched their names on Telegram search engines. Only eight of the Dozen had active accounts on Telegram at the time of this study. Channels are tools used to broadcast public messages to large audiences, but it is not possible to directly access lists of users who subscribe to these channels. However, some channels also allow for comments or discussion which allows users to have public conversations about the channel content. Of the 8 Disinformation Dozen accounts, 6 of them enabled discussions on their channels. We collected both the channel content originating from the Disinformation Dozen and the associated discussion by other users.

We then computationally identified forwarded content in channel posts and discussions and identified which channels/users messages were forwarded from, forming our 1-hop channel list. We then collected channel messages from this 1-hop

Table 1 Terminology used in this paper

Terminology	Description
Disinformation dozen	A set of 12 users identified to be the source of most coronavirus disinformation
Bot users	Inauthentic users that are sometimes automated by computer software
Human users	You and me
Channel	A public broadcast list to disseminate messages to large audiences
Discussion	A forum attached to a channel where users can post content
Messages	A short text as a discrete unit of communication
Forwarding	Sending a message to another channel using the "forward" function
Replying	Responding to a message within the same channel using the "reply" function

channel list. Using a 1-hop snowball sampling method provided a way of characterizing the dissemination of disinformation that stemmed from the original seed users—the Disinformation Dozen—as it identified users that disseminated messages through forwarding and collected information from the channels that received information second-hand. This technique allowed us to discover a large number of channels and users that were previously hidden from view.

To collect data from Telegram, we used the Python Telethon API.[1] Using Python scripts, we scraped posts from the Disinformation Dozen Telegram channels, their associated groups, and any channels/groups that were linked to the initial set of channels/groups by a content forward. The resulting data are almost exclusively in the English language. We collected data from January to June 2023, obtaining a total of 7,711,975 messages from 10,633 channels that were written by 335,088 unique users.

In our analysis, we only accessed data from public channels and made no attempts to access private channels or chats. In terms of users, we only attempted to identify the accounts of the Disinformation Dozen. For the other users, we only observe public activity and we made no effort to identify these users beyond their public participation in the channels and time period stated above. In presenting our work, we redact user names so that the users are not able to be identified, further preserving user privacy.

Terminology

This study performs an examination across multiple Telegram channels. In Table 1, we define some of the terminology used in this study and the terminology unique to the social interactions on Telegram.

[1] https://docs.telethon.dev/en/stable/.

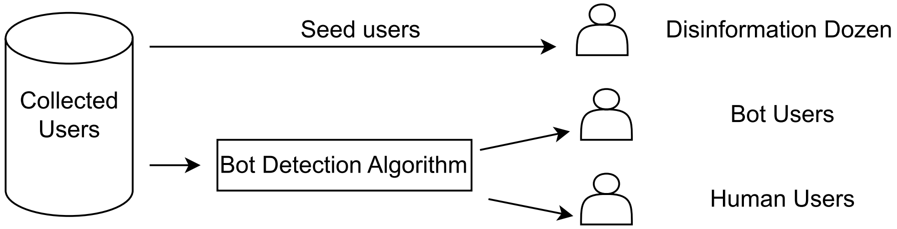


Fig. 2 User group identification pipeline

User group identification

Within this study, we segment the users into three types: (1) the Disinformation Dozen identified by the Center for Countering Digital Hate as key spreaders of COVID disinformation [3]; (2) Bot users as highly active inauthentic users; and (3) Human users as authentic users of the platform. Within this paper, we study these three user groups, comparing the similarities and differences in terms of their message posting patterns, message topics, and network interactions. In this section, we describe each of the three types of users and how we methodologically identified these users. We begin by illustrating the user group identification pipeline in Fig. 2.

Disinformation Dozen

The CCDH published a report on March 24, 2021, establishing 12 users termed as the "Disinformation Dozen", who are identified as being responsible for large-scale dissemination of COVID-19 disinformation and anti-vaccine claims circulating on social media platforms [3].

This group of 12 people is mostly comprised of medical professionals and activists who have been found to create original disinformation content on X. Their disinformation has been found to be more influential in conservative groups when compared to more liberal users [4]. On Instagram, 96.13% of their account pages redirect to the login page, suggesting that they might have been banned from the platform. Only 1.05% of mentions of the account pages are replayable with complete post images [34].

Studies surrounding the Disinformation Dozen have not been extensive, and this work contributes to current literature on the Disinformation Dozen by examining their role in the spread of information. We extend past work on the Disinformation Dozen by focusing on how disinformation propagated from their seed messages on Telegram rather than studying the content of their posts and/or the sources of information they relied upon [4].

We begin by describing the 12 influencers. Table 2 lists the names, occupations, and statuses on Telegram of the Disinformation Dozen. The names and occupations are adapted from [4]. The profiles of these 12 people involve authoritative profiles such as physicians and alternative medicine activists. We note

Table 2 Name, occupation, and Telegram user status of the Disinformation Dozen in 2023

Name	Occupation	Status
Joseph Mercola	Osteopathic physician	Active
Robert Kennedy Jr	Environment supporter	Not found
Christiane Northrup	Obstetrics and gynecology physician	Active
Rashid Buttar	Osteopathic physician	Active
Erin Elizabeth	Alternative medicine activist	Active
Kelly Brogan	Alternative medicine activist	Active
Sayer Ji	Alternative medicine activist	Active
Kevin Jenkins	Head of a no-vaccination group	Not found
Sherri Tenpenny	Osteopathic physician	Active
Ben Tapper	Chiropractor	Active
Ty and Charlene Bollinger	Alternative medicine activist	Not found
Rizza Islam	Anti-vaccination activist	Not found

which of these dozen are "active" on Telegram, indicating that we have discovered their profiles and channels. For those that we were unable to find, we list them as "Not found" rather than "Inactive" because it is possible they are operating under a different name on Telegram that we were unable to discover.

Bot users

To detect bot users, we used the concept of bot detection, which defines "bots" as user accounts that appear to be inauthentic. Bot detection algorithms typically make use of data from the user account, such as linguistic styles of user posts, user biography description, temporal information, and social network information to form feature sets used to differentiate bots from authentic accounts [35]. These feature sets are fed into supervised machine learning algorithms, including random forests, support vector machines, logistic regressions, neural networks, and deep learning methods [36].

Bot users have been observed to spread and amplify disinformation [26]. For example, in the issue of the US Elections, bots on the social media platform X were observed spreading disinformation involving themes like "Stop the Steal", manipulating public opinion with their online narratives [37]. The 2020 coronavirus pandemic also revealed a series of bot campaigns that served to distort information [38] and spread conspiracy theories [20].

Literature on Telegram bots generally focuses on techniques for developing bots for message exchange, chatting [39, 40] and monitoring server spaces [41]. An observational study analyzing the role and impact of Telegram bots in the Islamic State's online ecosystem revealed that Telegram bots identified through manual selection have two main roles: facilitating discussion and exchanging/augmenting content distribution efforts [42].

Table 3 Statistics of messages, username, and screenname between X and Telegram

	X	Telegram
Number of words in messages	10.09 ± 32.96	18.35 ± 33.39
Number of punctuation in messages	5.26 ± 8.34	3.98 ± 8.31
Number of characters in username	11.02 ± 2.7	10.67 ± 3.56
Numbers in username	1.11 ± 2.16	1.01 ± 1.49
Capitals in username	1.29 ± 1.62	1.48 ± 1.65
Numbers in screen name	0.12 ± 0.59	0.13 ± 0.66
Capitals in screen name	2.27 ± 2.72	1.89 ± 1.25
Number of words in screen name	2.07 ± 1.19	1.77 ± 0.75

All values between X and Telegram are statistically insignificant at the $p < 0.05$ value by a two-tailed *t*-test, therefore the length of messages and the style of usernames and screen names are similar between the two mediums

We build on the idea that bots on Telegram can facilitate discussion and disinformation spread and identify the extent and impact of their activities on the stories put forth by the Disinformation Dozen.

Bot user identification through bot detection algorithm

In the absence of an existing robust bot detection methodology for Telegram, we adapted a bot detection algorithm that has been validated on both X (previously named Twitter) and Reddit.

To do so, we acquired a random subset of 3000 Telegram messages and a random subset of 3000 Tweets collected from X. The Tweets collected from X were collected during the same time period using the streaming API, filtered for the hashtags #covid and #vaccine. This dataset was previously used to analyze the changes in stance toward the coronavirus vaccine [43].

With these datasets, we compared the messages, usernames and screen names, for these are the common fields between both platforms. Table 3 shows a statistics of comparison. We observed similarities in the average number of words and punctuation in messages, as well as the number of characters, numbers and capital letters in username and screen name. Further, we observed that there is no significant difference within the statistics at the $p < 0.05$ level when we perform a two-tailed t-test between the statistics derived from X and Telegram. The length of the messages and the style of the usernames and screen names are similar between the two mediums. Therefore, we can use the bot detection algorithm tuned for the platform X for our Telegram data.

Because messages and user names on the Telegram messenger were similar to those on X, we chose to adapt a bot detection algorithm that has been trained on data from X to annotate our Telegram users called Botbuster [29]. The BotBuster algorithm uses a mixture-of-experts algorithm concept where each data field has its own prediction, and the predictions of all the data fields are aggregated together for the final bot prediction. The algorithm can take in a total of six input fields: user name, screen name, post, user metadata, and user description. This helps BotBuster handle the differing inputs from several social media platforms—X, Instagram, and Reddit. Since the Telegram platform does not have all the data fields that BotBuster can leverage, being able to activate a subset of experts to derive a final bot prediction is ideal for our Telegram dataset. BotBuster also facilitates reading input from a pre-collected dataset rather than pulling data live, allowing us to work on the data that we have collected. Finally, BotBuster does not identify bots based on a temporal point-of-view, therefore it reduces the problem of false positive detection on highly active human users. There may still be highly active human users that are detected as bots, but these users are likely to have features that are similar to bots.

We first harmonized the data field names extracted from Telegram to the BotBuster algorithm convention. Then, we ran the users that are not the Disinformation Dozen through the BotBuster algorithm. The algorithm provides a bot likelihood score between [0, 1] that indicates the probability of the user being a bot. We use the threshold value of 0.5, where the user is determined to be a bot if the bot likelihood is equal to or above 0.5 and determined to be a human if the bot likelihood is below the 0.5 threshold value. With these parameters, the BotBuster algorithm determined our Telegram dataset to have 29.9% bots and 70.1% humans.

Annotation verification through manual annotation

To verify the bot/human labels generated by the BotBuster algorithm, we performed manual annotation on a subset of data. From the BotBuster output, we randomly selected a 0.1% sample by stratified sampling, thus ensuring that the proportion of bots/human users in the sample matches the proportions that are reflected by the BotBuster algorithm. In total, we extracted 2767 data points.

Two of the authors manually annotated the data points, reading through the user names and messages of each user before labeling the user as a bot or a human. In the event of a disagreement, a third annotator served to break the tie. All three annotators are native English speakers.

We also calculated the inter-annotator agreement between the first two annotators using Cohen's Kappa score. This score ranges from $[-1, +1]$ and serves as an indication of the proportion of annotations that were not in agreement due to random chance. We only compared the first two annotators, with the third annotator serving as a tie-breaker. We obtained a Cohen Kappa score of 0.92. This score is close to 1, indicating sufficient agreement between the two annotators [44, 45].

Finally, we harmonized the manual annotations through maximum pooling, taking the most common out of the three scores. These scores are thus regarded as the gold standard labels for this set of data. We compared the BotBuster-generated labels, finding a model $F1$-score of 72%. This is a reasonable accuracy given that the original model that was fine-tuned on Twitter data performed with an $F1$ of 73%.

From our observations, we determined a few characteristics of bot-like Telegram users. Note that while we present some examples in the list below, we present mild and neutral examples to avoid strong or disturbing language.

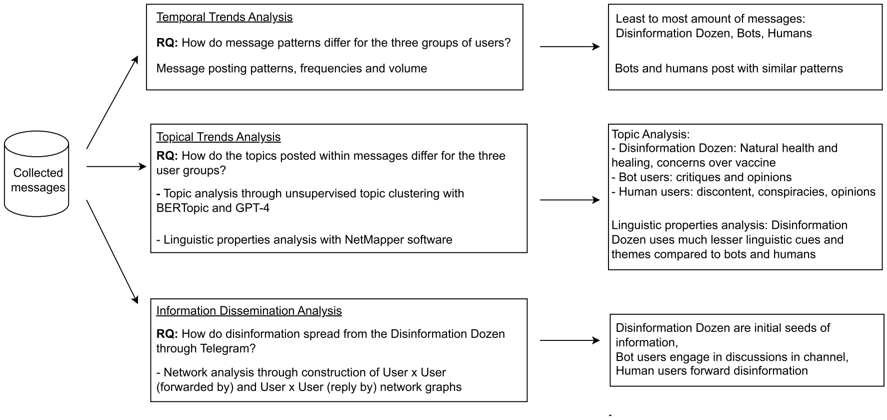

Collected messages
Temporal Trends Analysis
RQ: How do message patterns differ for the three groups of users?
Message posting patterns, frequencies and volume
Least to most amount of messages: Disinformation Dozen, Bots, Humans
Bots and humans post with similar patterns
Topical Trends Analysis
RQ: How do the topics posted within messages differ for the three user groups?
- Topic analysis through unsupervised topic clustering with BERTopic and GPT-4
- Linguistic properties analysis with NetMapper software
Topic Analysis:
- Disinformation Dozen: Natural health and healing, concerns over vaccine
- Bot users: critiques and opinions
- Human users: discontent, conspiracies, opinions
Linguistic properties analysis: Disinformation Dozen uses much lesser linguistic cues and themes compared to bots and humans
Information Dissemination Analysis
RQ: How do disinformation spread from the Disinformation Dozen through Telegram?
- Network analysis through construction of User x User (forwarded by) and User x User (reply by) network graphs
Disinformation Dozen are initial seeds of information,
Bot users engage in discussions in channel,
Human users forward disinformation

Fig. 3 Overview of methodology and brief results

1. Copying and pasting links in multiple messages, where the bulk of the links are the same except for the URL parameters and query strings; or keeps posting the exact same link
2. Repeating the same message multiple times
3. Writing in capital letters and the same number of exclamation marks for every message sent (e.g., “THATS RIGHT, THANK YOU!!!”)
4. Repetitive and short messages (e.g., “I agree with you”, “Inbox me please”)

Human users

All other users that were not labeled manually as the Disinformation Dozen, nor algorithmically as a bot user, were labeled as human users.

Methodology and results

With the Telegram data collected and annotated, we analyzed the data along three slices to answer the three research questions: temporal analysis, topical analysis, and network analysis. In this section, we detail our methodology and results for the three slices of analysis. We begin by illustrating our methodology and providing an overview of our results in Fig. 3.

Temporal trends

To answer Research Question 1 (How do message posting patterns differ for the three groups of users?), we use a temporal analysis approach. Temporal analysis provides an over-time understanding of the frequency of message sharing, showcasing the regularity and abnormalities of posting patterns across time. We plotted the

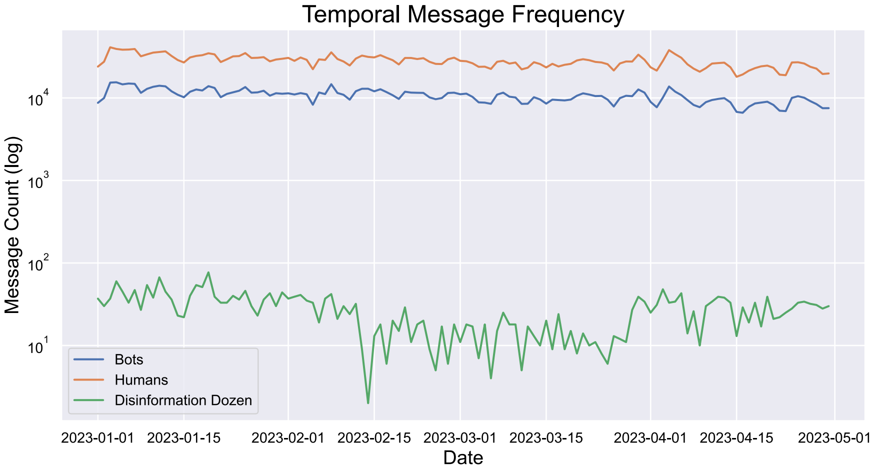


Fig. 4 Temporal post frequency per user group. The Disinformation Dozen posts least frequently, while humans post most frequently. Bots and humans post with similar frequency patterns, while the Disinformation Dozen posts the least

frequency of posts of each of the three user groups across time, grouped by day, in Fig. 4.

From manual inspection, we observe that the posting patterns of bots and humans are extremely similar. We, thus, compare their similarity using a Pearson correlation metric.[2] The Pearson correlation coefficient measures the linear relationship between the two posting frequency arrays as signals. It returns a value between $[-1, +1]$. Values closer to $+1$ indicate a positive correlation, values closer to -1 indicate a negative correlation, and values at 0 imply there is no correlation between the two signals.

From the bot and human signals, we obtain a Pearson correlation value between the bot and human signals of 0.983, with a p-value of 4.05E−90. Since the p-value is less than 0.05, we conclude that the result is statistically significant and that the posting patterns of the bot and human users are positively correlated. That is, if there is an increase in the frequency of bot messages, there is also an increase in the frequency of human messages.

Topical trends

To answer Research Question 2 (How do the topics posted within messages differ for the Disinformation Dozen, bots, and human users?), we use narrative analysis. We do so with two main techniques: topic analysis and linguistic analysis.

[2] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html.

Table 4 Generated topics for Disinformation Dozen

S/N	BERTopic label	GPT label
1	Greenmed health natural	Natural Health and Healing
2	Dies suddenly	Unexpected Deaths and Causes
3	ios medium medical	Online Medical Discussion
4	Covid masks dod	Covid masks and legal actions
5	Covid vaccine vaers	Vaccine Safety and Efficacy Concerns

Table 5 Generated topics for bot users

S/N	BERTopic label	GPT label
1	Military justice law government	Government and Justice Critiques
2	Biden trump president desantis	Political Affairs and opinions
3	Channel telegram join	Social Media Conspiracy Discussions
4	Vaccine jab covid	Covid Vaccine Opinions
5	Mask wearing	Wearing Masks and Attitudes

For topic analysis, we developed topic clusters for each of the three user groups using unsupervised topic analysis. We used two popular preexisting algorithms in our methodology: BERTopic [46] and OpenAI GPT-4 model[3] to extract topic sets from the messages. BERTopic returns a series of words that are most commonly used within a cluster of messages, while GPT provides a natural language interpretation of the messages as topics. For each user group, we input the set of messages that are authored by the users in the group into both topic clustering algorithms and extract the topics.

We derive the topic clusters for BERTopic by embedding each message into a vector using the all-MiniLM-L6-v2 transformer[4] from HuggingFace sentence-transformer. These embeddings are used as input for a BERTopic model that uses a TF-IDF transformer and a hdbscan model to segregate embeddings into clusters. The model finally returns the key phrases that represent each topic cluster. For both algorithms, we evaluated topic groups using the kMeans clustering algorithm, setting the hyperparameter K to 40, which was established by the elbow method using the KneeLocator function.

We derive the topic clusters for GPT-4 in the following fashion. We first embed each message into a vector using the same all-MiniLM-L6-v2 sentence transformer. The vector embeddings are sorted into 40 clusters through the kMeans clustering algorithm. We then extracted the text documents of each cluster and prompted a GPT-4 model with these documents. The prompt used is as follows:

[3] https://platform.openai.com/docs/models/.

[4] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2.

Table 6 Generated topics for human users

S/N	BERTopic label	GPT label
1	Trump biden president tucker	Trump Supporters, Against Biden
2	Military white police people	Racial and Political Discontent
3	Vaccine covid jabs	Covid and Vaccine Controversies
4	Video channel rumble https	Conspiracy video links
5	Mask wearing looks dog	Opinions on Masks and Appearances

> “““““
> I have a topic that contains the following documents: {DOCUMENTS HERE}
> Based on the information above, extract a short (5 words or fewer) topic label in the following format: topic: <topic label>
> ”””””

We present the top 5 topics by BERTopic and GPT labels on the three user group types in Tables 4, 5 and 6. We chose 5 topics as they represent the most salient topics determined through manual inspection. Both models generally output similar topic clusters, but the GPT labels present a more natural language interpretation. In general, the Disinformation Dozen posts messages related to vaccination, discouraging vaccination through medical discussions, vaccine safety and efficacy concerns, and unexpected deaths. Instead, they promote natural health and healing solutions. Bot users generally authored opinion messages, such as critiques toward the government, political affairs, and the coronavirus vaccine, suggesting that these bots were programmed to put forth messages consistent with specific ideologies. Human users focus on the Trump/Biden support, as well as other societal divides, such as racial and political divides.

We also performed a linguistic properties analysis using the Netmapper software.[5] The NetMapper software reads a message and counts the frequency of terms belonging to lexical categories, such as abusive, absolutist, positive, and negative terms, as well as terms relating to themes such as family, crime, and finance. These categories are built on a psycholinguistic theory that associates particular words with behavioral and cognitive states [47]. We derive quantitative cues from NetMapper that represent linguistic properties and themes. We then calculate the mean and standard deviation of each of the cues per user group (Disinformation Dozen, bots, and humans) and compare the differences between the groups.

Figure 5 shows the comparison of linguistic cues per user group. We find that the Flesch–Kincaid reading difficulty of the messages by all three groups are very low, for short online messages are not generally very complex. The largest use of linguistic cues is the positive and negative terms, which could be used in discouraging vaccination and encouraging natural health cures. Another commonly used linguistic feature is the 3rd person pronoun, e.g., “we”, which gives a sense of

[5] http://www.casos.cs.cmu.edu/tools/data.php.

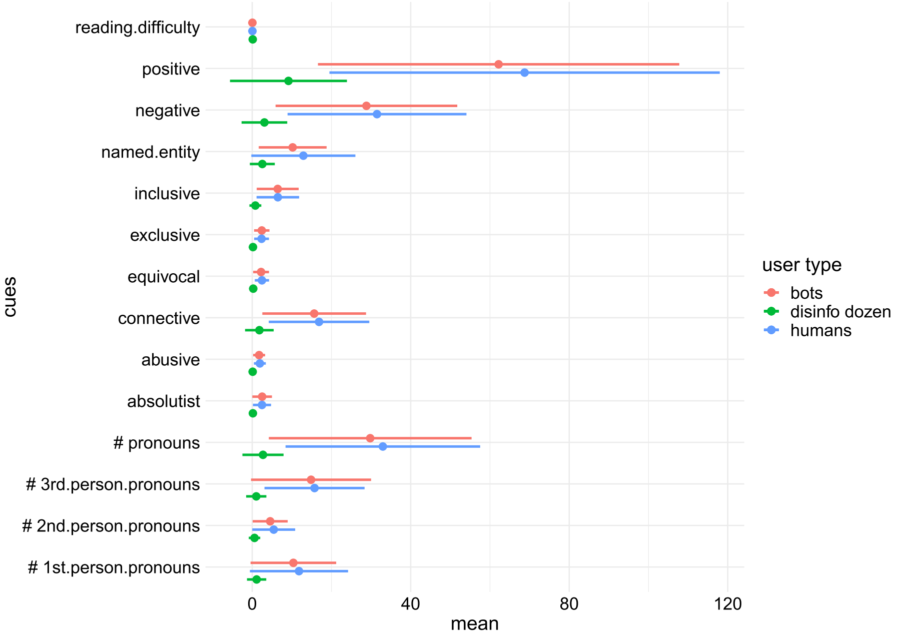


Fig. 5 Comparison of linguistic cues among the three user group

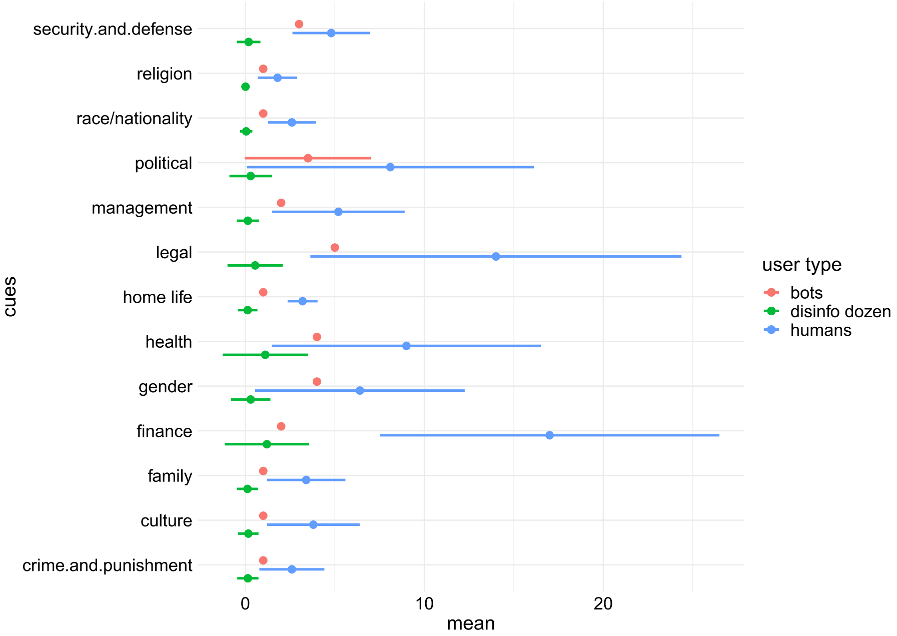


Fig. 6 Comparison of themes among the three user group

community and that we are all in it together [43]. The 1st person pronoun, e.g., “I”, is also frequently used, which provides a personal touch with personal anecdotes and opinions [48].

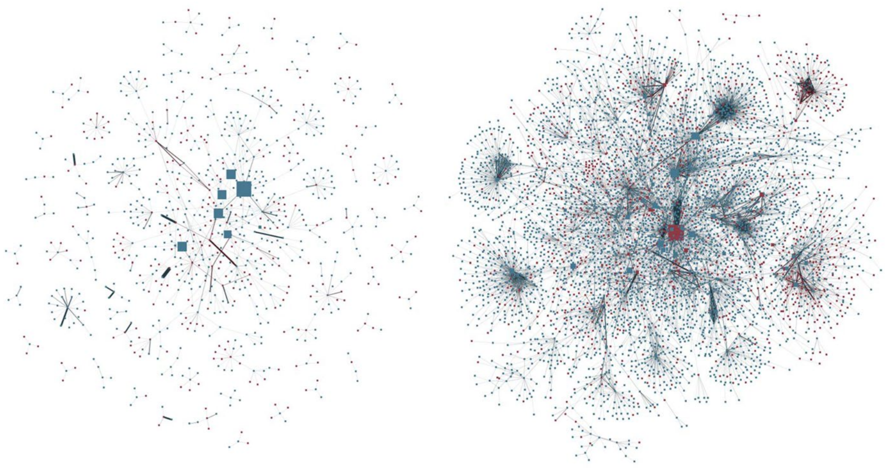
Figure 6 shows the comparison of themes per user group. Human users have the most varied uses of themes, in particular, finance, health, political, and legal themes. This is followed by bot users, which tend to have a lot of political and legal themes, followed by health and gender themes. The Disinformation Dozen has the smallest use of theme-based words within their messages, but the most talked about theme is the health theme. This is likely because this group concentrates their messages around a focused theme and does not take part in side conversations.

Information dissemination patterns

We analyzed how information is disseminated in Telegram using a network analysis approach. This answers Research Question 3: How does disinformation begin spreading from the Disinformation Dozen throughout the Telegram messenger network? Network analysis provides us a bird’s-eye view of the interactions between users and channels on Telegram [8]. Visualizing user-to-user communication and interactions through a network perspective provides information about where the hubs and spokes of information flow are and the control of aggregation and dissemination of information.

We constructed two network diagrams: (a) User x User network by message forwarding and (b) User x User network by message reply. These two diagrams represent Telegram users as nodes and construct a user-to-user diagram (i.e., User x User). Two users are joined together by a link if they have an interaction between them. For the first diagram, the interaction is message forwarding; for the second diagram, the interaction is message replies. The width of the links between two users represents the frequency of the interaction that has occurred during the timeframe that we collected data. We also segregated bots and human users by means of a red and blue color scheme, respectively.

We used the betweenness centrality measure to size the nodes, where a larger node size corresponds with a larger betweenness centrality value. The betweenness centrality value indicates the extent to which a node lies on the path of information flow. It is normally calculated as a fraction of the shortest paths between node pairs that pass through a single node [49]. The higher the betweenness centrality of the node, the more information flows through it, and thus that user is influential as a hub for disseminating disinformation. Past work that studied the difference in betweenness centrality between bots and humans observed that bots are more likely to be on the bridges of information flow, reflecting higher betweenness centrality values [50]. Bots have been observed to occupy 7–16% of the top users with the highest betweenness centrality ranking, implying that they play the role of bridging and mediating information diffusion [51].



(a) User x User network by Message Forwarding

(b) User x User network by Message Reply

Fig. 7 Network graphs of user–user interaction with the forward and reply mechanism. The network is pruned to contain interaction links between users that have at least a frequency of 5. Blue nodes represent human users, while Red nodes represent bot users. The nodes are sized by the betweenness centrality measure. The Disinformation Dozen users are not observed in these graphs as they do not frequently forward other users' messages or reply to other users

Table 7 Comparison of average betweenness centrality values between bot and human users

User x user network	Bot users	Human users
By message forwarding	1.6E−4 ± 2.1E−3	1.9E−4 ± 3.9E−3
By message reply	1.5E−8 ± 2.7E−8	2.5E−10 ± 3.6E−8

The network graphs and the betweenness centrality measures were calculated and plotted with the ORA software.[6] We observe that the network constructed by messaging forwarding interactions, depicted in Fig. 7, while sparse (network density = 5E−7), is dominated by human users, suggesting that human users are the user group that actively forward messages. The network constructed by messaging replying interactions is a lot denser (network density = 5.8E−5), which suggests heavier usage of the replying function compared to the forwarding function. That network is also dominated by bot users as influential users in the network, suggesting that bot users are highly active and generate lots of engagement for the coronavirus discourse.

For each of the two network graphs plotted, we calculated the average betweenness centrality values of the bot and human users and presented them in Table 7. The quantitative calculations of the network centrality metric match the visual inspection of the network graphs: human users have higher betweenness centrality

[6] http://www.casos.cs.cmu.edu/projects/ora/software.php.

in the message-forwarding network, while bot users have higher betweenness centrality in the message-replying network. This suggests that in terms of message forwarding, human users are more likely to facilitate the flow of information, while in terms of message replies, bot users are more likely to accelerate discussions.

Discussion

In this work, we studied the information dissemination patterns that stemmed from the Disinformation Dozen, a group of 12 users that the Center for Countering Digital Hate (CCDH) identified were key spreaders of COVID-19 disinformation online. We examined the disinformation spread through three lenses: temporal, topical, and network.

Our data collection timeline overlaps with the timeline in which the CCDH had performed their data collection and investigation. We posit that the CCDH performed their investigation during that time frame because it is likely the time frame that the Disinformation Dozen were highly active.

We began our study by annotating bot users on Telegram, adapting a current bot detection algorithm to take in Telegram data as input. Given that the message length and user names are similar across the social media platform X and the Telegram messenger, we were able to adapt the current algorithm for use within this study. The similarity of messages shows that users, in general, posted with similar writing styles on both platforms. This finding provided justification for the use of tools that were developed for other social media platforms like X, Facebook, and Reddit, adapting them to an understudied and emerging platform.

From the temporal lens, we observed that the Disinformation Dozen posted the fewest messages, followed by bots, than humans. The Disinformation Dozen were not the spreaders of disinformation, nor did they generate discourse. Rather, they were the originators of disinformation stories, which were propagated through the Telegram medium by bots and human users.

From the topical lens, we found that the Disinformation Dozen posted messages on natural health and healing and concerns over the vaccine, bot users typically critique and post opinions, while human users posted their opinions, discontent and conspiracy theories. From these topical divides, we posit that Disinformation Dozen was trying to promote natural health cures over the coronavirus vaccination. With this specific information range, the topics put forth by the Disinformation Dozen were the narrowest, and they used the least varied linguistic cues. Bot users were programmed to post critical remarks and opinions. Due to their range of topics, from political affairs to governments to social media, they had a variety of themes and linguistic styles within their messages. Human users talked generally about their opinions and discontent, therefore they touched on the widest range of topics and used the most varied linguistic styles.

From the network analysis lens, we observed that the Disinformation Dozen were not actually the main users involved in the propagation of disinformation on Telegram but rather drove the initial information dissemination. They were able to command large followings with their channels, mainly because of their background as established medical professionals or activists. This demonstrates a real-world example of authority bias, a type of behavioral bias where the opinions and instructions of authority figures are unquestioned [52]. Authority bias is common in the medical field with authority figures, professional experts, and hierarchy. It can be seen in the doctor–patient relationship, which many of the Disinformation Dozen mirrors [53]. Therefore, using their medical expertise as an authority, the Disinformation Dozen seeded their channels and allowed other users to generate discourse and forward information.

These forwarding users included bot users who were active using the platform's replying mechanic, creating a huge amount of engagement around COVID-19 disinformation within the original channels. The user set also includes human users who were active in forwarding disinformation between platforms. Whether they forwarded this information because they truly believed the disinformation or because they found it amusing [54], the act of forwarding propagates the disinformation to another set of users on other channels. We posit that the bots may have participated more in conversations within the channel rather than forwarding information because their software programming was set to analyze conversations from and reply to only one specific channel.

The bot behavior we observe on Telegram is somewhat surprising in that it is distinct from common bot behavior on other platforms, like X. In other social media, it is common to find bots functioning as bridges between users and groups (e.g., using @ mentions to link users) [55, 56]. In many cases, these bridging bots do not post content at all. This paradigm is completely reversed on Telegram, with bots serving primarily as content contributors and not regularly attempting to bridge users/groups.

Discovering content on Telegram is much more constrained than it is on other platforms such as X or Reddit because there is no algorithmic news feed or any content discovery mechanism beyond message forwarding. Interestingly, this constraint seems to impact bot behavior, and it suggests that bots may have more limited capability to influence social networks on Telegram than on platforms that have greater support for discovery and connection-making.

The information creation and sharing behavior of the Disinformation Dozen, bots, and humans were similarly observed in the platform X by past works. A previous study observed how the Disinformation Dozen acted as sources of information by creating original content while interacting very little with other users [4]. Meanwhile, other users had a fair amount of user-to-user interaction and substantially forwarded content.

We also observed separated clusters of communities in the network diagrams, as well as clusters connected together by bridging users. This is similar to past work observing patterns of information dissemination on Telegram. While communities on Telegram tend to interact and communicate within their own groups, there is substantial information sharing between communities and ideologies through the forwarding mechanism, facilitating information spread [57].

While the volume and virality of disinformation on online platforms are alarming, we take comfort in studies that show that the social media platforms are taking action to remove such posts: less than 4% of the Instagram posts by the Disinformation Dozen can be replayed, as they have been removed from the archive by the Instagram platform [34]. However, it is much more difficult to police sets of conversational channels in the Telegram messenger setup, and our work shows that many of the messages are still alive up to six months after the post dates. While Telegram has shown to have removed and banned some posts containing disinformation and conspiracy theories, many cloned channels and messages appear quickly, suggesting further work is required to moderate the messenger [15].

Limitations and future work

Our work is not without its limitations, and thus one should exercise caution before using it to inform further policy. We present some of the limitations in this section and the avenues for future work.

Finding channels of the Disinformation Dozen is relatively easy because this group of users wants to be found. They want to broadcast their (dis)information to as large an audience as possible and thus use similar user names and descriptions across social media platforms. Still, we did not manage to find all the dozen on Telegram, and the presence of which might bring results that are different from the ones observed. We were only able to locate 8 of the 12 Disinformation Dozen on Telegram. It is possible that the other 4 individuals have Telegram accounts that are not identifiable but are still relevant in the medical disinformation space. We expect that our snowball sampling approach would have found these accounts (and it very well may have), but we do not have any way to verify the identities of these users. Additionally, the snowball sampling method might introduce bias in the data collection. The snowball is constructed by the conversation within the channels manned by the Disinformation Dozen and, therefore, reflects users who are actively involved with the 12 original users. These issues might limit the breadth of the study's insights, therefore caution should be used in extrapolating our results, as there may be users and channels that separately spread disinformation and propagate anti-vaccine ideologies but are not directly linked to our initial set of seed channels.

While our bot detection algorithm has achieved an accuracy of 72% on the dataset, the algorithm used was trained on a series of datasets from the platform X. Tweets from X, while similar to Telegram messages, are not exactly the same, and therefore to further improve the accuracy, the algorithm could be modified to be trained directly on Telegram datasets, to better fine-tune the algorithm to handle Telegram-specific data.

Finally, our study focuses on data primarily in the English language. Telegram is used by users all around the world, and therefore disinformation can be propagated in multiple languages. Future work would be to investigate the propagation of coronavirus disinformation that stems from the Disinformation Dozen in languages other than English. Such work will provide insight into the global reach of the disinformation propagation.

Conclusion

The fear of the unknown and the doubts about the efficacy of the vaccine, combined with government-imposed isolation generated by the COVID-19 pandemic, created a fertile environment for disinformation to flourish.

The spread of disinformation online is a concerning phenomenon, especially when it involves public health, as it weakens the government's abilities to control and eradicate the pandemic [18]. In this study, we examined the spread of COVID-19 disinformation that stemmed from an influential group of disinformation spreaders on Telegram by identifying and studying the roles of three user group types: the Disinformation Dozen, bot users, and human users.

Contrary to the report by CCDH, we found that the Disinformation Dozen were not the most prolific in spreading coronavirus-related and anti-vaccination disinformation on Telegram. Instead, the Disinformation Dozen initiated disinformation themes, while bot users were essential in sustaining the conversations, and human users were crucial in disseminating information to other Telegram channels.

This study also demonstrated some unique and unexpected human and bot behavior on Telegram compared to other social media. Specifically, we found that community bridging behavior was primarily human-driven on Telegram. Additionally, we showed that a large community centered on medical disinformation was able to form on Telegram in spite of the platform's somewhat limited functionality for users to discover new content. These differences highlight the importance of future study into Telegram (and other similar platforms). The research community has focused on platforms like X over the past 10 years due to ease of data access, but our findings suggest these studies may not generalize as well as previously thought. Moreover, with platforms like X and Reddit no longer supporting the research community through platform-provided APIs, Telegram is becoming a more useful platform for future work.

Within this work, we developed a repeatable methodology for analyzing Telegram messages and users across temporal, topical, and network domains. We hope our work motivates and serves as a springboard for future analysis of disinformation propagation on Telegram and other less-studied social media platforms.

Author contributions Lynnette conceptualized the study and performed the analysis and writing. Ian collected the data and performed the analysis. Samantha performed data annotation and analysis. Kathleen reviewed the work. All authors have agreed upon the manuscript.

Funding Open Access funding provided by Carnegie Mellon University. This material is based upon work supported by the U.S. Army Research Office and the U.S. Army Futures Command under Contract no. W911NF-20-D-0002, Office of Naval Research (Bothunter, N000141812108), US Army Scalable Technologies for Social Cybersecurity (W911NF20D0002), Air Force Research Laboratory/CyberFit (FA86502126244), Office of Naval Research Scalable Tools for Social Media Assessment (N000142112229). The content of the information does not necessarily reflect the position or the policy of the government and no official endorsement should be inferred.

Availability of data and code Data and code can be obtained from contacting the corresponding author, in accordance to the sharing policy of Telegram.

Declarations

Conflict of interest The authors declare that they have no conflict of interests.

References

1. Ng, L. H. X., & Carley, K. M. (2021). "The coronavirus is a bioweapon'': Classifying coronavirus stories on fact-checking sites. *Computational and Mathematical Organization Theory, 27*(2), 179–194.
2. Uscinski, J. E., Enders, A. M., Klofstad, C., Seelig, M., Funchion, J., Everett, C., Wuchty, S., Premaratne, K., & Murthi, M. (2020). Why do people believe COVID-19 conspiracy theories? *Harvard Kennedy School Misinformation Review, 1*(3).
3. CCDH. (2021). *The Disinformation Dozen—Center for Countering Digital Hate | CCDH—counterhate.com*. https://counterhate.com/research/the-disinformation-dozen/. Accessed 25 Oct 2023.
4. Nogara, G., Vishnuprasad, P. S., Cardoso, F., Ayoub, O., Giordano, S., & Luceri, L. (2022). The disinformation dozen: An exploratory analysis of COVID-19 disinformation proliferation on twitter. In *Proceedings of the 14th ACM web science conference 2022* (pp. 348–358). Association for Computing Machinery.
5. Krishnan, N., Gu, J., Tromble, R., & Abroms, L. C. (2021). Research note: Examining how various social media platforms have responded to COVID-19 misinformation. *Harvard Kennedy School Misinformation Review, 2*(6), 1–25.
6. Forbes. (2023). *Pavel Durov—forbes.com*. https://www.forbes.com/profile/pavel-durov/?sh=77a6811e14c5. Accessed 26 Oct 2023.
7. Ng, L. H. X., & Loke, J. Y. (2020). Analyzing public opinion and misinformation in a COVID-19 Telegram group chat. *IEEE Internet Computing, 25*(2), 84–91.
8. Willaert, T., Peeters, S., Seijbel, J., & Van Raemdonck, N. (2022). Disinformation networks: A quali-quantitative investigation of antagonistic Dutch-speaking Telegram channels. *First Monday*, 27(5).
9. Sosa, J., & Sharoff, S. (2022). Multimodal pipeline for collection of misinformation data from telegram. In *Proceedings of the thirteenth language resources and evaluation conference* (pp. 1480–1489). Marseille: European Language Resources Association.
10. Al-Rawi, A. (2022). News loopholing: Telegram news as portable alternative media. *Journal of Computational Social Science, 5*(1), 949–968.
11. Walther, S., & McCoy, A. (2021). US extremism on Telegram. *Perspectives on Terrorism, 15*(2), 100–124.
12. Khaund, T., Hussain, M. N., Shaik, M., & Agarwal, N. (2020). Telegram: Data collection, opportunities and challenges. In *Annual international conference on information management and big data* (pp. 513–526). Springer.
13. Weigand, M., Weber, M., & Gruber, J. (2022). Conspiracy narratives in the protest movement against COVID-19 restrictions in Germany. A long-term content analysis of Telegram chat groups. In *Proceedings of the fifth workshop on natural language processing and computational social science (NLP+ CSS)* (pp. 52–58). Association for Computational Linguistics.
14. Guhl, J., & Davey, J. (2020). *A safe space to hate: White supremacist mobilisation on telegram* (Vol. 26). Institute for Strategic Dialogue.

15. La Morgia, M., Mei, A., Mongardini, A. M., & Wu, J. (2021). Uncovering the dark side of telegram: Fakes, clones, scams, and conspiracy movements. arXiv preprint. arXiv:2111.13530
16. Liu, B. F., & Kim, S. (2011). How organizations framed the 2009 H1N1 pandemic via social and traditional media: Implications for US health communicators. *Public Relations Review, 37*(3), 233–244.
17. Ng, H. X. L., Lee, R. K.-W., & Awal, M. R. (2020). I miss you babe: Analyzing emotion dynamics during COVID-19 pandemic. In *Proceedings of the fourth workshop on natural language processing and computational social science* (pp. 41–49). Online. Association for Computational Linguistics.
18. Caliskan, C., & Kilicaslan, A. (2023). Varieties of corona news: A cross-national study on the foundations of online misinformation production during the COVID-19 pandemic. *Journal of Computational Social Science, 6*(1), 191–243.
19. Bernard, R., Bowsher, G., Sullivan, R., & Gibson-Fall, F. (2021). Disinformation and epidemics: Anticipating the next phase of biowarfare. *Health Security, 19*(1), 3–12.
20. Moffitt, J., King, C., & Carley, K. M. (2021). Hunting conspiracy theories during the COVID-19 pandemic. *Social Media+ Society, 7*(3), 20563051211043212.
21. Zheng, L. (2013). Social media in Chinese government: Drivers, challenges and capabilities. *Government Information Quarterly, 30*(4), 369–376.
22. Ng, L. H., & Taeihagh, A. (2021). How does fake news spread? Understanding pathways of disinformation spread through APIs. *Policy & Internet, 13*(4), 560–585.
23. Pal, A., & Chua, A. Y. (2019). Propagation pattern as a telltale sign of fake news on social media. In *2019 5th International conference on information management (ICIM)* (pp. 269–273). IEEE.
24. Liu, Y., & Wu, Y.-F. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32).
25. Hakak, S., Khan, W. Z., Bhattacharya, S., Reddy, G. T., & Choo, K.-K. R. (2020). Propagation of fake news on social media: Challenges and opportunities. In *Computational data and social networks: 9th International conference, CSoNet 2020, Dallas, TX, USA, December 11–13, 2020, proceedings* (Vol. 9, pp. 345–353). Springer.
26. Shao, C., Ciampaglia, G. L., Varol, O., et al. (2018). The spread of low-credibility content by social bots. *Nat Commun*, *9*, 4787. https://doi.org/10.1038/s41467-018-06930-7.
27. Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019). Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Social Media + Society, 5*(3), 2056305119865465.
28. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th international conference on world wide web companion* (pp. 963–972). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva.
29. Ng, L. H. X., & Carley, K. M. (2023). Botbuster: Multi-platform bot detection using a mixture of experts. In *Proceedings of the international AAAI conference on web and social media* (Vol. 17, pp. 686–697).
30. Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications, 9*(1), 1–9.
31. Geeng, C., Yee, S., & Roesner, F. (2020). Fake news on Facebook and Twitter: Investigating how people (don't) investigate. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–14). Association for Computing Machinery.
32. Wen, S., Jiang, J., Xiang, Y., Yu, S., Zhou, W., & Jia, W. (2014). To shut them up or to clarify: Restraining the spread of rumors in online social networks. *IEEE Transactions on Parallel and Distributed Systems, 25*(12), 3306–3316.
33. Alieva, I., Ng, L. H. X., & Carley, K. M. (2022). Investigating the spread of Russian disinformation about biolabs in Ukraine on Twitter using social network analysis. In *2022 IEEE international conference on big data (big data)* (pp. 1770–1775). IEEE.
34. Bragg, H., Jayanetti, H. R., Nelson, M. L., & Weigle, M. C. (2023). Less than 4% of archived Instagram account pages for the disinformation dozen are replayable. In *Proceedings of ACM/IEEE joint conference on digital libraries (JCDL)*. Santa Fe, NM.
35. Feng, S., Tan, Z., Wan, H., Wang, N., Chen, Z., Zhang, B., Zheng, Q., Zhang, W., Lei, Z., Yang, S., et al. (2022). Twibot-22: Towards graph-based twitter bot detection. *Advances in Neural Information Processing Systems, 35*, 35254–35269.

36. Heidari, M., James Jr., H., & Uzuner, O. (2021). An empirical study of machine learning algorithms for social media bot detection. In *2021 IEEE international IOT, electronics and mechatronics conference (IEMTRONICS)* (pp. 1–5). IEEE.
37. Chang, H.-C. H., Chen, E., Zhang, M., Muric, G., & Ferrara, E. (2021). Social bots and social media manipulation in 2020: The year in review. arXiv preprint. arXiv:2102.08436
38. Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., Epstein, D. H., Leggio, L., & Curtis, B. (2021). Bots and misinformation spread on social media: Implications for COVID-19. *Journal of Medical Internet Research, 23*(5), 26933.
39. Domashnev, P., Alexeev, V., Lavrukhina, T., & Nazarkin, O. (2019). Usage of telegram bots for message exchange in distributed computing. *International Journal of Open Information Technologies, 7*(6), 67–72.
40. Oliveira, J. C., Santos, D. H., & Neto, M. P. (2016). Chatting with Arduino platform through telegram bot. In *2016 IEEE international symposium on consumer electronics (ISCE)* (pp. 131–132). IEEE.
41. Idhom, M., Fauzi, A., Alit, R., & Wahanani, H. E. (2018). Implementation system telegram bot for monitoring Linux server. In *International conference on science and technology (ICST 2018)* (pp. 1089–1093). Atlantis Press.
42. Alrhmoun, A., Winter, C., & Kertész, J. (2023). Automating terror: The role and impact of telegram bots in the Islamic State's online ecosystem. *Terrorism and Political Violence*. https://doi.org/10.1080/09546553.2023.2169141.
43. Ng, L. H. X., & Carley, K. M. (2022). Pro or anti? A social influence model of online stance flipping. *IEEE Transactions on Network Science and Engineering, 10*(1), 3–19.
44. Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23.
45. Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics, 34*(4), 555–596.
46. Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint. arXiv:2203.05794
47. Tausczik, Y. R., & 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.
48. Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology, 33*(2), 125–143.
49. Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social Networks, 27*(1), 39–54.
50. Ng, L. H. X., & Carley, K. M. (2023). A combined synchronization index for evaluating collective action social media. *Applied Network Science, 8*(1), 1.
51. Cai, M., Luo, H., Meng, X., Cui, Y., & Wang, W. (2023). Network distribution and sentiment interaction: Information diffusion mechanisms between social bots and human users on social media. *Information Processing & Management, 60*(2), 103197.
52. Howard, J., & Howard, J. (2019). Bandwagon effect and authority bias. In *Cognitive errors and diagnostic mistakes: A case-based guide to critical thinking in medicine* (pp. 21–56). Cham: Springer.
53. Silvester, C. (2021). Authority bias. In *Decision making in emergency medicine: Biases, errors and solutions* (pp. 41–46). Cham: Springer.
54. Duffy, A., Tandoc, E., & Ling, R. (2020). Too good to be true, too good not to share: The social utility of fake news. *Information, Communication & Society, 23*(13), 1965–1979.
55. Gilani, Z., Farahbakhsh, R., Tyson, G., Wang, L., & Crowcroft, J. (2017). Of bots and humans (on Twitter). In *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017* (pp. 349–354). Association for Computing Machinery.
56. Samper-Escalante, L. D., Loyola-González, O., Monroy, R., & Medina-Pérez, M. A. (2021). Bot datasets on Twitter: Analysis and challenges. *Applied Sciences, 11*(9), 4105.
57. Kloo, I., & Carley, K. M. (2023). Social cybersecurity analysis of the telegram information environment during the 2022 invasion of Ukraine. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (pp. 23–32). Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.