Characterizing Information Propagation in Fringe Communities on Telegram

Mohamad Hoseini[1], **Philipe de Freitas Melo**[2, 4], **Fabrício Benevenuto**[2],
Anja Feldmann[1], **Savvas Zannettou**[3]
[1]Max Planck Institute for Informatics, Germany
[2]Universidade Federal de Minas Gerais, Brazil
[3]Delft University of Technology, Netherlands
[4]Universidade Federal de Viçosa, Brazil
mhoseini@mpi-inf.mpg.de, philipe.freitas@ufv.br, fabricio@dcc.ufmg.br, anja@mpi-inf.mpg.de, s.zannettou@tudelft.nl

Abstract

Online messaging platforms are key communication tools, but are vulnerable to fake news and conspiracy theories. Mainstream platforms such as Facebook are increasing content moderation of harmful and conspiratorial content. In response, users from fringe communities are migrating to alternative platforms like Telegram. These platforms offer more freedom and less intervention. Currently, Telegram is one of the leading messaging platforms hosting fringe communities. Despite the popularity, as a research community, we lack knowledge of how content spreads over this network. Motivated by the importance and impact of messaging platforms on society, we aim to measure the information propagation within fringe communities on the Telegram network, focusing on how public groups and channels exchange messages. We collect and explore about 140 million messages from 9,000 channels and groups on Telegram. We examine message forwarding and the lifetime of the messages from different aspects. Among other things, we find inequality in content creation; 6% of the users are responsible for 90% of forwarded messages. We also discover that while the forwarding considerably amplifies the reach of messages, the spread of content within our dataset remains largely localized. Additionally, we find that 5% of the channels are responsible for 40% of the forwarded messages in the entire dataset. Finally, our lifetime analysis shows that messages disseminated in groups with numerous active users exhibit significantly longer lifespans compared to those circulated in channels.

Introduction

The advancement of online messaging platforms has facilitated a transition in our mode of communication toward a predominantly digital environment. These platforms have a huge user base and are often the main way for people to connect with friends and family, remain updated on daily news, and organize social movements. At the same time, they became a place for misinformation campaigns, extremism, and conspiracy theories. To combat the undesirable content circulating within their networks, most of the mainstream platforms implemented moderation tools to prevent this.

Although mainstream platforms such as Facebook, Twitter, and Instagram continue to host a significant portion of online content, numerous alternative online platforms have emerged. These platforms offer a "safe space for discussion", free from the intervention of major technology companies. Some platforms emerge primarily to accommodate users who have been suspended from other social networks for violating their terms of service, such as Gab (Zannettou et al. 2018; Lima et al. 2018), Parler (Aliapoulios et al. 2021), and BitChute (Trujillo et al. 2020). Along with these alternatives, especially with the use of smartphones, instant messaging platforms have gained more and more space in this environment (Hoseini et al. 2020). These messaging platforms, such as Telegram, WhatsApp, WeChat, Discord, etc., allow users to quickly chat with their contacts in a private and secure channel while also enabling group communication. Some of those apps have gained special attention recently, given their influence in events around the globe, such as the spreading of misinformation about the COVID-19 pandemic (Londoño 2021), fake news campaigns in political elections (Resende et al. 2019a), and even the influence of the Russian-Ukrainian war (Stokel-Walker 2022).

One of the primary messaging services is Telegram. Launched in 2013, Telegram is nowadays a messaging platform with approximately 700 million users (Leong 2020) and is vastly used worldwide. It provides users with various communication tools, including audio and video calls and multimedia messages with text, images, audio, videos, stickers, and URLs. Chats on this platform are structured in a one-to-one format with direct personal chats between two users but also allow one-to-many communication with channels and many-to-many with group chats. These groups and channels in Telegram can have millions of members, and users can share an invite link for other users to join and participate in the chats. This creates a rich network within the platform, connecting millions of users and favoring information propagation across groups and users.

Traditional media, businesses, and public figures use Telegram to officially publish news, share ideas, and promote products and services. At the same time, due to its popularity and lack of moderation, this platform is frequently exploited by malicious actors. They use it to perpetrate scams, disseminate conspiracy theories and misinformation campaigns, and serve as a stronghold for spreading hate speech and other harmful content (Urman and Katz 2022; Guhl and Davey 2020; Hou, Wang, and Wang 2022).

Despite the popularity of messaging platforms, as a research community, we lack knowledge of how content spreads over this network. Since the architecture of instant messaging services differs from the traditional social networks, consisting of chats, groups, and channels, it has distinctive patterns of sharing messages and propagating content (Melo et al. 2019). As instant messaging services have been increasingly used in our daily lives, serving both as communication tools and information sources, it has become more necessary to understand the processes and mechanisms behind the information that reaches millions of users' phones through these platforms with such an impact on society. Motivated by the importance and impact of these messaging platforms on society, in this study, we aim to measure the information propagation within the Telegram network, focusing on how public groups and channels exchange messages, focusing on forwarded content shared between them. Notably, we want to understand the reach of information posted on Telegram and the communication structure in such a closed environment. We aim to answer the following research questions:

- **RQ1:** How is forwarding used on Telegram? Is a small number of users responsible for sharing/forwarding a large number of messages?
- **RQ2:** What is the lifetime of content shared on Telegram? Does content persist for a long time on Telegram?

To answer these questions, we first create an extensive data collection of about 140 million messages sent in groups and channels on Telegram. The groups and channels are discovered by collecting public posts, including Telegram links from Facebook and Twitter. Using this large corpus of data to understand the dynamics and propagation of information, we analyze the dataset across different aspects. First, we analyze the difference in information propagation in channels and groups, which are two types of communication in Telegram. Then, we investigate the behavior of different types of messages, such as URLs, regular messages, forwarded messages, and direct messages. Also, we analyze the user's specific activity and the different categories of URLs of the messages scattered in the channels and groups. Finally, we analyze the content of the messages by performing toxicity and sentiment analysis, aiming to identify differences in the lifespan/forwarding patterns of toxic vs. non-toxic messages and positive vs. negative messages.

Main Findings. Among other things, we find:

- We find that 6% of the users are responsible for 90% of forwarded messages. We observed a disparity in content creation on Telegram, with a small fraction of users significantly influencing discourse. This observation underscores the necessity for user-specific moderation interventions to prevent a limited number of users from disseminating a large volume of potentially harmful information within the Telegram network.
- We observe a significant variation in the dynamics of information dissemination between groups and channels on Telegram. Our findings indicate that groups receive a larger proportion of forwarded messages compared to channels. Concurrently, messages originating from channels are more likely to be forwarded than those from groups. This suggests that channels predominantly function as the source of forwarded messages.
- Approximately 35% of the forwarded messages contain URLs, and over half of these URLs originate from news sources and two prominent social media platforms: "YouTube" and "Twitter".
- Our findings indicate that regular messages without URLs exhibit a longer lifespan than messages containing URLs, with the former lasting on average twice as long. Among messages with URLs, those referring to text messaging platforms demonstrate the greatest longevity.
- Our analysis reveals that toxic messages and messages characterized by emotional extremity, whether positive or negative, have a longer lifespan compared to non-toxic and neutral messages, respectively.
- We find that despite messages being distributed locally, the forwarding feature significantly extends their reach.

Background & Related Work

The propagation of information has been an increasingly debated subject, mainly due to the popularity of online platforms, which allow people to be reached very quickly. Recent studies focus on the dissemination of information in social media platforms such as Tiktok, Twitter, and Facebook (Ostrovsky and Chen 2020; Choi et al. 2020). They demonstrate how online media can quickly disseminate information, making this environment vulnerable to misinformation, rumors, and fake news (Vosoughi, Roy, and Aral 2018).

Regarding messaging platforms, Resende et al. (2019b) show that WhatsApp's structure of groups and chats is similar to other mainstream social networks, with a well-connected network with paths that enable messages to travel between groups and users. Melo et al. (2019) study the virtualization of messages on WhatsApp. They investigate how messages spread through WhatsApp and highlight the epidemic process that makes messages go viral within this platform, showing that limits imposed by the system are not enough to prevent misinformation dissemination.

Furthermore, instant messaging platforms have drawn attention due to the prevalence of abuses within them, including the spread of fake news and harmful content. On WhatsApp, rumors and false stories shared on WhatsApp lead to lynchings and violent acts in India (Arun 2019; Vasudeva and Barkdull 2020). In parallel to the pandemic of COVID-19, an infodemic was also running within the app, in which a huge volume of health misinformation about the disease, against the vaccine, and with ineffective treatments against COVID-19 flooded the chats of users around the globe (Malhotra 2020). Studies from various regions such as Pakistan (Javed et al. 2022), Spain (Elias and Catalan-Matamoros 2020), Zimbabwe (Bowles, Larreguy, and Liu 2020), Brazil (Forte Martins et al. 2021), UK (Vijaykumar et al. 2021), and India (Varanasi, Pal, and Vashistha 2022) raise concerns regarding misinformation about COVID-19 on messaging platforms, indicating its status as a global issue. Misinformation campaigns circulating on these messaging services are also pointed to have an important role of interference in the democratic process of elections in countries in which those platforms have a large user base, including in

Nigeria (Cheeseman et al. 2020; Hitchen et al. 2019), in India (Reis et al. 2020; Kazemi et al. 2022), in Brazil (Resende et al. 2019b,a), and related to the American elections with and the Capitol riot in January 2021 (Solopova, Scheffler, and Popa-Wyatt 2021). Moreover, on Telegram, users create groups for conspiracy theories such as Qanon that mobilize users within networks beyond countries' limits and reach a global scale (Hoseini et al. 2023; Peeters and Willaert 2022).

Still related to extremism in this cyberspace, Guhl and Davey (2020) discuss that Telegram's lenient content moderation policies could serve as a safe space for white supremacists to disseminate and deliberate on extremist and hateful content. They analyze one million posts across 208 channels that disseminate white supremacist material, revealing endorsements for terrorism in 125 of them. Solopova, Scheffler, and Popa-Wyatt (2021) investigate online harms on Telegram, building an annotated dataset for hate speech and offensive language from a channel of Donald Trump supporters, Walther and McCoy (2021) suggest that these platforms are progressively serving as channels for disseminating hate speech and extremist violence.

Telegram has also gained considerable attention for being used by jihadist groups such as ISIS. A 2019 work on online extremism (Clifford and Powell 2019) investigate 636 Telegram pro-Islamic State channels containing English propaganda, finding these groups exploit the Telegram encrypted environment to attract sympathizers and promoters of terrorist content. There are a bunch of studies that focus on exploring the way terrorist groups such as ISIS leverage Telegram's encrypted environment (Prucha 2016; Yayla and Speckhard 2017; Shehabat, Mitew, and Alzoubi 2017). These groups harness the capabilities of the Telegram platform for communication, the dissemination of propaganda, and potentially for the recruitment of new affiliates.

Another issue regarding these platforms, in particular Telegram and Discord, is that they are also commonly used for the practice of diverse forms of digital scams (La Morgia et al. 2021). Especially, a vast range of cryptocurrency schemes to steal digital activities from users to pump-and-dump manipulation (Hamrick et al. 2021; Andryukhin 2019; Morgia et al. 2022; Gao et al. 2021; Nizzoli et al. 2020; Mirtaheri et al. 2021). The architecture of these messaging platforms facilitates such abuses by providing a private, secure, and anonymous environment for cybercriminals to interact with their customers in underground markets (Hou, Wang, and Wang 2022). These platforms often offer illicit products and services, drawing their targets to this unmoderated and closed environment (Kansaon, Melo, and Benevenuto 2022). Some authors suggest that within these platforms, a hidden underground space thrives, characterized by fakes, extremism, scams, and conspiracies, coexisting alongside regular user activity (Urman and Katz 2022; La Morgia et al. 2021).

There are also other issues with the usage of Telegram in other countries. Nikkah et al. (Nikkhah, Miller, and Young 2018) examine Telegram usage among Iranian immigrants, specifically inspecting the moderation mechanisms within these Telegram communities. Hashemi et al. (Hashemi and Chahooki 2019) undertake an extensive evaluation of 900k Iranian channels and 300k Iranian groups, aiming to categorize them based on quality, distinguishing between high-quality channels, such as those related to business, and low-quality channels, for instance, those dedicated to dating.

Towards a more panoramic view and characterization of Telegram, some previous work focuses on collecting large-scale data from Telegram and studying emerging research problems. Hoseini et al. (2023) perform a large-scale collection on Twitter of invite links for public groups from Telegram, WhatsApp, and Discord, analyzing more than 350K groups from those platforms and the peculiar characteristics of each one. They also discuss privacy flaws of the public groups advertised online. Dargahi Nobari, Reshadatmand, and Neshati (2017) collect data from 2,600 Telegram groups/channels, conducting a structural review of the posted content within these communities. Abu-Salma et al. (2017) execute a user-based study to gauge perceptions surrounding Telegram's security measures. Naseri and Zamani (2019) investigate news dissemination via Telegram, aggregating data from five official channels utilized by media outlets.

Research Gap. Social media and private messaging apps, such as Telegram are widely used by people to share information and become a key source of information propagation about different events. Considering high engagement and millions of daily users on these platforms, there is a need to understand the dynamics of these platforms and how information is created and propagated in this ecosystem. Recent studies focus on the dissemination and propagation dynamics of information in mainstream social media. However, despite its growing popularity, alternative cyberspaces such as Telegram have received relatively less attention in academic research, and we know little about the propagation of information in this ecosystem. In this work, we analyze the propagation of information among Telegram groups and channels to provide a better understanding of how information spreads in a large-scale ecosystem.

Data Collection

Instant messaging platforms present peculiar characteristics that pose unique challenges for studying them. The enclosed structure of these platforms in public and personal chats requires some strategies to find and collect data. First of all, there are two types of chat or communication environments on Telegram: channels and groups. In groups, there exists many-to-many communication among users, and each user can send messages. In channels, there is one-to-many communication or broadcast that the admin of the channel can send messages, and the other members are only able to read the messages. For both scenarios, administrators can make their chats public by sharing an invite link with other users or posting it online for others to join and participate in the discussion. Additionally, users of a chat can share messages with other chats by redirecting content through the forwarding feature. This flow of messages creates an interconnected network within Telegram, enabling large-scale information propagation through those apparently detached chats, spreading content throughout the whole network.

The first step for researchers who study phenomena in instant messaging or community-based social network platforms is to discover related and public groups to collect data from them. However, it is not a trivial task as there is no vantage point to find related groups. For this work, we gather data from QAnon communities within Telegram, since this has been shown to be a growing topic within this platform (Pasquetto et al. 2022) that exchanges messages on a global scale through hundreds of groups and channels dedicated to discussion of these conspiracy theories (Hoseini et al. 2023). Therefore, to build a large-scale data set of shared messages on Telegram, we use data of groups and channels made available by the study from Hoseini et al. (2023) as a starter point and expand it. The dataset comprises 161 Telegram public chats related to the QAnon movement from different countries. These chats were collected and filtered from an extended web search for public invite URLs posted by users on their social networks (i.e., Facebook and Twitter) between April and October 2020. We expand this dataset by discovering new chats based on messages forwarded from our initial set of chats. Specifically, we:

Collect groups metadata. Using Telegram's Web client, we obtain metadata of the chats from corresponding invite URLs including: a) Chats' title; b) Description; c) Number of members; and d) Messages sent within each chat.

Extract sources of the messages. As many messages within the chats consist of forwarded content from various sources, including channels, groups, and individual users, they contain an identifier indicating the original source of the message. Then, we extract the identifiers of all sources of the forwarded messages from the set of 161 QAnon-related chats. With this step, we identify 40,000 new sources, significantly expanding our initial data. Furthermore, all these sources are related to the original set of chats because forwarded messages from them are shared within the chats we initially collected. This relationship is crucial for our study's investigation into information propagation within this ecosystem.

Collect messages of the groups. Next, using the Telegram API, we try to collect messages from the sources. It is not possible to collect messages from all of the sources since, among them, there are private chats, user accounts, and chats that are not accessible anymore. Finally, the messages from 9,139 public chats are collected in our dataset (see Table 1).

Limitations. Our dataset has important limitations. First, we are unable to assess the representativeness of our dataset because extracting a random sample of chats from Telegram is not feasible. Consequently, we rely on Telegram chats shared on platforms like Twitter and Facebook and expand our dataset using forwarded messages. Second, our initial seed of chats is related to a fringe movement, namely QAnon, thus, our dataset is likely to be biased towards chats involved in the dissemination of fringe ideologies. It is possible that our snowballing method for data collection could have identified groups/channels similar to our initial seed set. Due to this, we acknowledge that our findings apply to these particular fringe communities and probably cannot be generalized to the entire Telegram network. We believe that this is an inherent limitation that exists in almost all the studies focusing on messaging platforms like Telegram, mainly because there is no vantage point to obtain holistic or representative samples of Telegram chats. Additionally, since we collect messages after joining the groups/channels, we miss messages that have already been deleted by the users. Nevertheless, we can confirm that by expanding the dataset based on forwarded messages, we collect a large amount of data that includes many mainstream chats.

Chat Type	#Chats	#Senders	#Messages	#Forwarded messages
Channels	7,669	7,669	51,516,609	19,334,687
Groups	1,355	2,201,374	86,884,730	20,570,197
Total	9,024	2,209,036	138,401,339	39,904,884

Table 1: Overview of our Telegram data set.

Ethical considerations. Standard ethical guidelines (Rivers and Lewis 2014) are respected in dealing with the gathered data during this project. Also, the ethical review board of our institute has checked and approved our data collection and analysis. Note, that all data obtained during this project, tweets, Facebook posts, invite URLs, and messages of the Telegram chats, are publicly available data, and none of the users of the above-mentioned platforms is de-anonymized.

Results

Forwarding

Users on the Telegram platform are provided with the "Forwarding" feature. They can forward messages to other private chats, groups, or channels. Forwarded messages can be forwarded again by any user who has access to the messages. This way, messages can propagate and go viral throughout the entire Telegram platform. Here, we perform an analysis to get a better understanding of how messages get forwarded and spread through our dataset. Before describing our analysis, we will define some key terms:

- *Forwarded message:* Any message in a chat that is forwarded into the chat using the "Forwarding" feature.
- *Direct message:* Any message which is not a forwarded message.
- *Original message:* If direct message A is forwarded into a chat as message B and message B is forwarded into a chat as message C, A is the original message for forwarded messages B and C.
- *Source chat:* In the above example, the chat in which message A is shared is the source chat of forwarded messages B and C.
- *Internal source chat:* A source chat that is included in our dataset.

Forwarded messages vs. direct messages. According to Table 1, about 40 million messages (one-third of all messages) in the dataset are forwarded messages. Since we have the ID number of the original source chat for each forwarded message in our dataset, we check if there is any match between these ID numbers and the ID numbers of the 9,000 chats in our dataset. We find the original source chats of 25 million (63% of all 40 million) forwarded messages in our dataset. The forwarded messages originate from 454,980 unique source chats, and we find 7,894 of these source chats in our dataset. These 7,894 chats are the original source

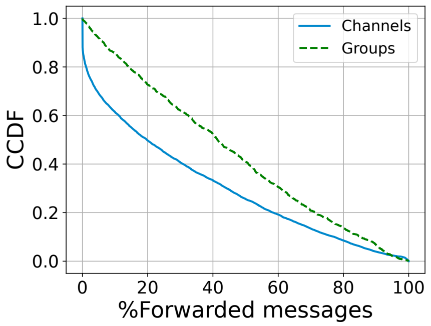

Figure 1: CCDF of the percentage of forwarded messages in the channels and groups.

chats of 63% of all forwarded messages. This implies that our dataset is mostly fed by its own sources and we have a big community of chats highly connected to each other.

Channels vs. groups. We observe different forwarding behaviors inside the groups and channels. The Complementary Cumulative Distribution Function (CCDF) of the percentage of forwarded messages within the channels and groups is shown in Fig. 1. We observe that messages get forwarded more in groups than in channels. While in half of the channels, less than 20% of the messages are forwarded messages, in half of the groups more than 40% of the messages are forwarded messages. This shows that the groups play more of the role of consumers of the messages originally created in the other sources.

The 7,894 source chats we find in our dataset consist of 7,346 channels (96% of all channels in the dataset) and 548 groups (40% of all groups in the dataset). We find the original source messages of about 25 million forwarded messages in our dataset. The original messages of 44,000 forwarded messages are found in the groups and the rest (about 25 million) are found in the channels. This shows that although the number of channels is more than five times the number of the groups, the number of forwarded messages in the dataset that are fed from the channels is more than 500 times the number of forwarded messages fed from the groups. This means that the channels have the role of producer of the forwarded messages much bigger than the groups. After finding the original messages of the forwarded messages in the dataset, we map the number of forwarded messages corresponding to these original messages for each chat. In this way, we find out each chat how many forwarded messages are responsible. We also imply to which extent the messages of each chat get spread in the entire dataset. Fig. 2 shows the CCDF of the number of forwarded messages that each chat is responsible for producing the corresponding original messages. We observe that 60% of the groups do not produce any original message of the forwarded messages. This implies that messages created in the groups don't get forwarded to the other chats at a high rate. On the other hand, the channels in the dataset have a great influence on other chats in the sense that their messages get spread widely in other communities. We extract top active channels in originating forwarded messages. The top 5% of these channels are responsible for producing the original messages of 40% of all forwarded messages in the dataset. These top active channels are leading the content shared inside the dataset.

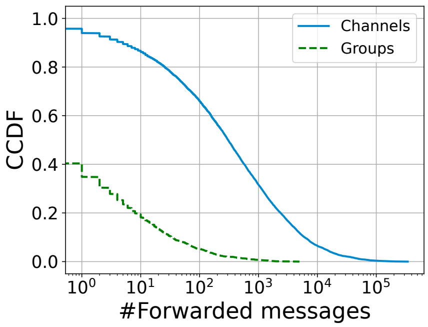

Figure 2: CCDF of the number of forwarded messages originally produced in the channels and groups.

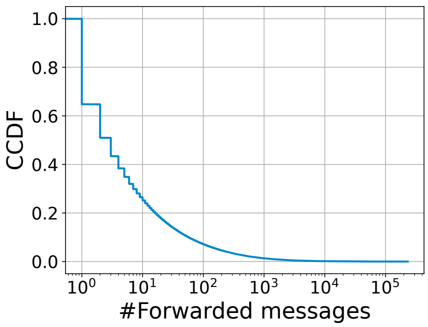

Figure 3: CCDF of the number of forwarded messages shared by each active user.

Users of the content. As channels work as broadcast communication and users are not able to share messages, we analyze the role of users only for the groups. We do not have access to the number of users in each group, but the sender of the messages. Inside all of the groups, there are 2 million unique users who have sent at least one message, and 225,000 of them have forwarded at least one message. Fig. 3 shows the number of forwarded messages by each user. About 30% of the users who contribute to forwarding messages, forward only one single message. Meanwhile, about 90% of them forward less than 100 messages. However, there are extremely active users with several thousands of forwarded messages.

To study further the user-specific forwarding behavior, in Fig. 4, we plot what percentage of users are responsible for what percentage of forwarded messages in our dataset. We observe that 6% of the senders are responsible for forwarding 90% of the forwarded messages. When considering direct messages, we find that 18% of the senders are responsible for sharing 90% of the direct messages inside the groups, which indicates that the user-specific behavior is more concentrated to a small number of users for forwarding messages compared to direct messages.

URLs vs. regular messages. On messaging platforms, it is

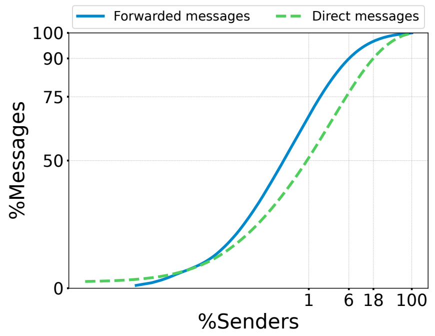

Figure 4: The percentage of forwarded and direct messages for which a percentage of senders is responsible.

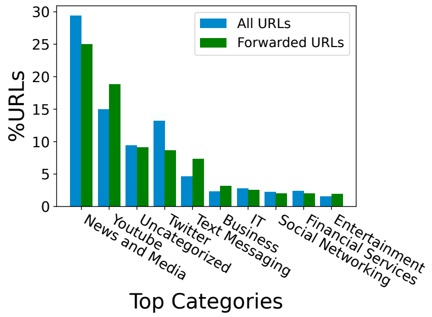

Figure 5: Percentage of top 10 URL categories in our dataset.

common among users to share URLs pointing to pieces of information instead of sharing that information inside the chats. Motivated by this, here we study the use and forwarding of messages that include URLs. We find a considerable number of URLs among text messages in our dataset. While 19% of direct messages include at least one URL in their text, 35% of forwarded messages contain URLs. This indicates that users tend to forward messages with URLs more than regular messages without a URL. We can also imply that messages with URLs have a higher chance of getting forwarded. The users try to spread the content linking to these URLs among several communities on Telegram. The topics of these URLs show the types of content users try to spread among communities. To know about the type of URLs, we try to extract the categories of the URLs.

URL categories. There are about 4 million URLs within the forwarded messages. We resolve the URLs to obtain their long version and then extract the domain for each one of them. Then the category of each domain is extracted using the Virus Total URL categorization API. Fig. 5 shows the percentage of top URL categories in forwarded and all URLs. The most common category is "News and Media" with 25%, followed by "Youtube" and "Twitter" with 19% and 9% of forwarded URLs respectively. These results confirm the findings in (Grindrod and Bovet 2022), as they found that URLs linking to "Youtube" and "Twitter" are the most shared URLs in upstream group chats. This shows that users tend to spread the news among the chats. Also, the connection between the chats in our dataset and two well-known platforms, "Youtube" and "Twitter" is indicated.

Toxicity. To evaluate the toxicity level of the content within our dataset, we employ Google's Perspective API (Perspective API 2018) to annotate each text message. We adopt the SEVERE_TOXICITY model, as recommended in Ribeiro et al. (2021), to assign a toxicity score to each message. This score serves as a numerical representation of the comment's degree of rudeness or disrespectfulness. We chose to utilize Perspective API for annotation because the API provides models that are production-ready and multilingual. As of August 2023, Perspective API supports annotation in 18 languages, enabling us to analyze text messages in diverse languages. This coverage is particularly beneficial, as it covers 96% of the text messages—each containing a minimum of five words—in our dataset. It is important to note that the use of Perspective API for toxicity assessment is not without limitations. In particular, the API has the potential to yield false positives and may also be subject to biases (Davidson, Bhattacharya, and Weber 2019). In our dataset, the occurrence of toxic messages is low for both forwarded and direct messages, 1.5% and 1.9%, respectively. These proportions suggest that toxicity is not a significant factor affecting forwarding behavior.

Sentiment. We perform sentiment analysis to determine if the emotional tone of a text message is positive, negative, or neutral. For this analysis, we employ a machine learning approach for the sentiment analysis of text messages presented in (Loureiro et al. 2022). Specifically, we utilize a pre-trained RoBERTa model fine-tuned for Twitter sentiment analysis. The model, identified by the handle *"cardiffnlp/twitter-roberta-base-sentiment-latest"* is accessed through the Hugging Face Transformers library. The text messages are processed using a sentiment analysis pipeline. This pipeline streamlines the application of the pre-trained model to our dataset. Sentiment labels and associated confidence scores are automatically generated for each text entry. We also subject our sentiment analysis methodology to validation. The method is applied to a randomly selected subset of 100 text messages. The outputs are subsequently compared against manual annotations for this sample. This validation yields an accuracy rate of 84%. Based on the results, among the forwarded messages, the sentiment is distributed as follows: 15% positive, 34% negative, and 51% neutral. On the other hand, for direct messages, the sentiment distribution is 16% positive, 42% negative, and 42% neutral. The frequency of positive sentiment is almost equal in both categories of messages. This suggests that the tendency to forward a message is not influenced by its positive sentiment. However, a meaningful difference exists in the negative sentiment category. The percentage of negative sentiment (42%) in direct messages is substantially higher than the percentage of negative sentiment (42%) in forwarded messages (34%). This indicates that messages with negative sentiment are less likely to be forwarded.

Message Reach. We define the "reach" of a message as the total number of chats where the message has been shared.

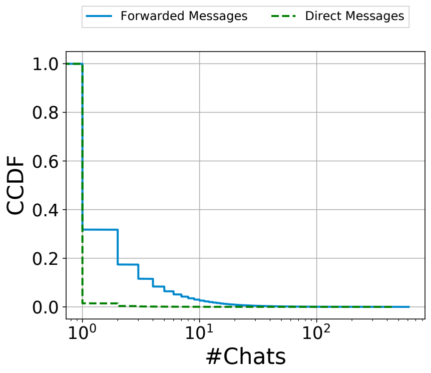

Figure 6: CCDF of the reach for forwarded/direct messages.

We aim to examine the role of forwarding behavior in influencing this reach. To do this, we conduct an analysis comparing the reach of forwarded messages to direct messages within our dataset. Fig. 6 presents the CCDF for both forwarded and direct messages. Our statistical analysis using the Kolmogorov-Smirnov test demonstrates a significant difference between the distribution of reach for forwarded and direct messages. This indicates that forwarding behavior effectively extends the reach of messages. While forwarded messages generally exhibit higher reach values, both forwarded and direct messages typically have a relatively low rate of reach in our dataset. This implies that messages in our dataset are not broadly viral; instead, they appear to be shared within a localized network of chats.

Remarks. About 29% of all messages shared inside the dataset are forwarded messages. This shows a strong connection between the content shared in different chats. The original messages of 63% of forwarded messages are produced by the chats in the dataset. This indicates that in our dataset, we have an interconnected network in which chats frequently produce and consume each other's messages. 35% of the forwarded messages contain URLs which are mostly from news sources and two well-known platforms namely “Youtube” and ”Twitter”. 28% of all of the forwarded URLs are links from “Youtube” and “Twitter”. The channels and groups have different manners of consuming and producing forwarded messages. While the groups have more of the role of a consumer of the forwarded messages, the channels have more of the role of producer of forwarded messages. The activities of the users differ massively. Although there are a lot of users with very low levels of activity, 6% of users are super active and are responsible for forwarding about 90% of forwarded messages into the groups. While the level of toxicity in messages does not significantly influence forwarding behavior, messages with a negative emotional tone exhibit a lower forwarding rate compared to direct messages. The forwarding feature significantly expands the reach of messages, even though they mainly circulate locally.

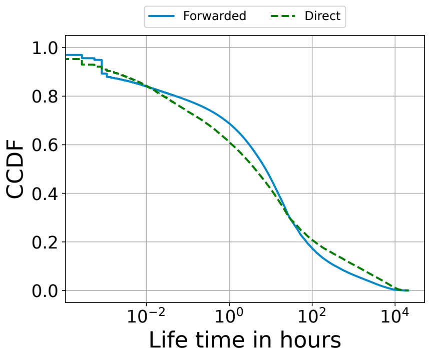

Figure 7: CCDF of the lifetime of each repeated text message for forwarded and direct ones.

Life Span

There are text messages that are repeated throughout the entire dataset. From 138 million messages in our dataset, 115 million are text messages. Out of these 115 million text messages, 10 million are unique text messages shared more than once in our dataset, appearing in 53 million messages overall. To enhance the quality of our analysis, we exclude text messages composed of fewer than 5 words. We establish the criterion of excluding messages with fewer than five words after an initial examination of our sample set revealed a high frequency of short messages. These messages, often consisting of phrases such as 'yes,' 'hi,' and 'thanks,' typically lack substantial content. To validate this approach, we investigate two subsets: 50 random and 50 frequent five-word messages. After manual annotation, 83% are considered meaningful, supporting our decision to focus on messages with at least five words for insightful analysis. Finally, we observe 8,640,142 unique text messages containing more than 4 words and shared more than once in our dataset. These text messages appeared in 39,733,986 messages in total. For repeated text messages, we define the lifetime as the time interval between their first and last appearances in our dataset. Note that repeated text messages mean the exact match between the entire string patterns of different messages. We calculate the lifetime of a message as the time difference between the first and the last appearance of the message in our dataset. We investigate how long messages from different aspects continue to be shared in our dataset.

Forwarded messages vs. direct messages. There are about 3.7 million unique forwarded messages while there are about 4.9 million unique direct messages with more than one appearance in our dataset. About 4% of unique direct text messages and 40% of unique forwarded text messages appear more than once in the dataset. This shows that forwarded text messages get repeated much more than direct text messages. Fig. 7 shows the lifetime of the forwarded and direct text messages. Running a two-sample Kolmogorov-Smirnov test discloses significant differences between the two distributions ($p < 0.01$). Overall, direct text messages that appear more than once last longer than repeated forwarded messages. We can infer that based on our findings, com-

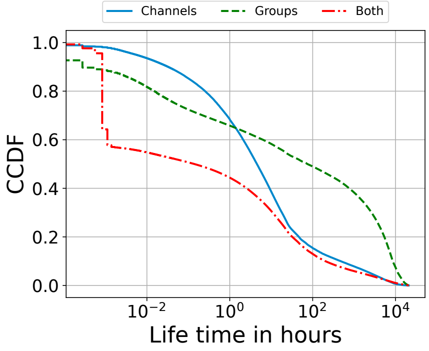

Figure 8: The CCDF of the lifetime of each repeated piece of text shared in the channels, groups, and both.

pared to direct messages, forwarded messages get repeated more frequently but in shorter time intervals. This observation suggests that the forwarding mechanism predominantly influences messages of immediate relevance or high popularity, similar to trending news. However, these messages also seem to have a shorter lifespan, potentially fading from discourse more quickly as they are replaced by newer topics.

Channels vs. groups. In this part, we evaluate and compare the lifetime of the messages shared inside the groups and channels. We aim to investigate the disparities in message lifetimes between the two distinct environments, characterized by differing numbers of users capable of sharing messages. Out of 9,895,811 unique text entries, 1,654,780 are found only in channels, 1,875,935 are shared solely in groups, and 6,365,096 are observed in both channels and groups. Fig. 8 shows the lifetime of the messages shared in the groups, channels, and both sets. We observe that about 6%, 8%, and 40% of the messages shared in both, only in the channels, and only in the groups respectively have a lifetime longer than one month. A two-sample Kolmogorov-Smirnov test also confirms that the distributions of the lifetime of the messages shared in groups and channels are significantly different ($p < 0.01$). Based on the statistics, text messages disseminated exclusively within groups exhibit longer lifetimes compared to those distributed solely in channels. On average, messages disseminated exclusively within groups last for 105 days, whereas those shared solely in channels persist for an average of 17 days. This shows that messages that are shared solely in the groups have a significantly higher chance of living longer than the ones shared in the channels. One plausible explanation for this phenomenon may be the group's structure, where every user has the capability of sharing messages. This larger set of potential senders inherently increases the likelihood of messages being shared again.

URLs vs. regular messages. Another factor that may impact the lifetime of messages is the inclusion of URLs within the message content. There are 3.7 million messages that appear more than once in our dataset and contain no URL. On the other hand, our dataset includes approximately 4.9 million messages that contain at least one URL and appear more than once. On average, regular messages with no URLs have a lifespan of 16 days, while those containing URLs persist for an average of 9 days. Fig. 9 shows the CCDF of the lifetime of the text messages with and without URLs. We perform a two-sample Kolmogorov-Smirnov test on the two distributions. The result shows statistically significant differences between the lifetime of the text messages containing at least one URL and regular messages without any URL ($p < 0.01$). Our evaluation shows that regular messages last longer than messages with URLs.

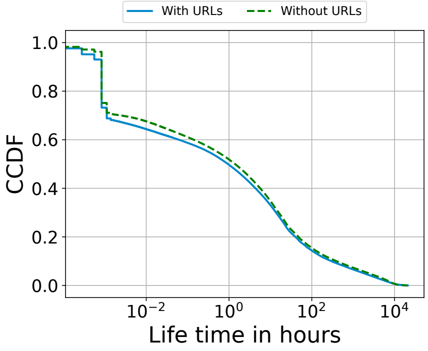

Figure 9: The CCDF of the lifetime of messages containing URLs and regular messages without URLs.

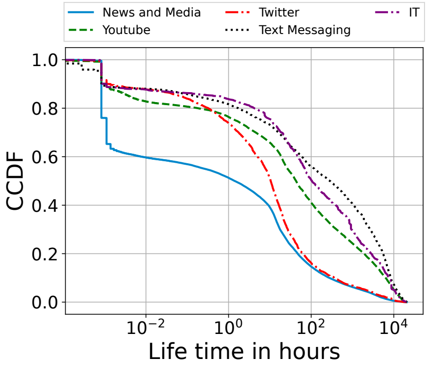

Figure 10: The CCDF of the lifetime of the URLs in the top 5 categories.

URL categories. There are 4,418,985 URLs which are appeared in our dataset more than once. Depending on the categories and topics of the URLs, users may exhibit different behaviors regarding their dissemination inside the chats. We aim to investigate how the categories of URLs impact their lifespans and to determine which categories users are more inclined to continue sharing. Based on the appearance of the URLs, we calculate the lifetime for the URLs in the top 5 categories. Fig. 10 shows the CCDF of the lifetime for the URLs with each one of the top 5 categories. On average, URLs with categories of “Text messaging”, “Information technology”, “Youtube”, “Twitter”, and “News and media” have a lifespan of 54 days, 40 days, 33 days, 8 days, and 7 days respectively. We observe that the URLs with the “News and Media” category have the shortest life-

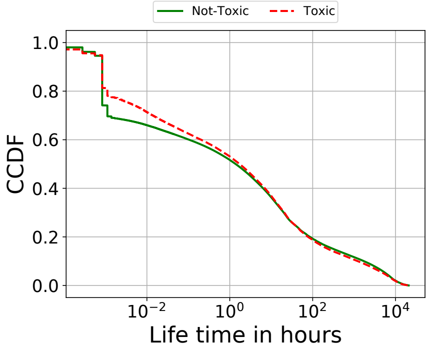

Figure 11: The CCDF of the lifetime of the toxic and non-toxic messages.

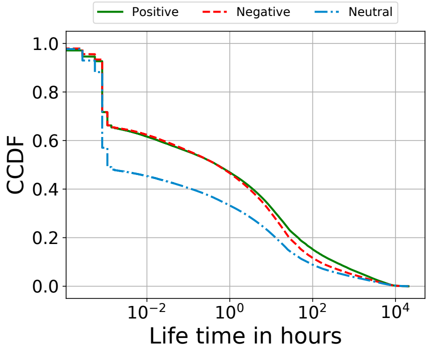

Figure 12: The CCDF of the lifetime of the messages with different sentiments.

times. About 40% of these URLs only last a few minutes. One possible reason could be the time-sensitive nature of news-related URLs. Their relevance decreases quickly and they rapidly fade away due to the emergence of newer stories or events. On the other hand, URLs with the “Text Messaging” category, which refers to text and media messaging platforms such as Telegram, have the longest lifetimes followed by the “IT” related URLs. We may imply that as URLs referring to text messaging platforms are circulating in other communities they have more chance to be shared again and live longer.

Toxicity. Fig. 11 shows the lifetime distribution of both toxic and non-toxic messages within the dataset. The result of a Kolmogorov-Smirnov test shows a significant difference between the two distributions ($p < 0.01$). Interestingly, toxic messages persist within chat environments for slightly longer than non-toxic messages. This observation is noteworthy, as one might expect that toxic messages vanish more quickly; however, our data suggest otherwise.

Sentiment. To further understand the dynamics of message longevity based on their sentiment, we examine the lifespan of messages categorized by different sentiments. Fig. 12 indicates relationships between the emotional tone of messages and their lifetime. We run the Kolmogorov-Smirnov test for each pair of the distribution of three categories of sentiment. Although the P value for all of them is lower than 0.01, the “P value” and “statistic” for the lifetime of positive and negative messages are substantially lower than the other two pairs. The results indicate that messages with emotional extremity—either positive or negative—demonstrate significantly longer lifespans within chat environments compared to those that are emotionally neutral. This suggests that emotionally charged content, irrespective of its positive or negative orientation, tends to last longer within the discussions of the chats.

Remarks. 4% of direct text messages and 40% of forwarded text messages appeared more than once showing a significantly higher repetition rate among forwarded messages. Although forwarded messages get repeated much more frequently than direct messages, they tend to vanish more quickly compared to repeated direct messages. Upon comparing the lifetime of messages shared in different chat types, we observe that messages disseminated solely within groups exhibit a remarkably longer lifespan than those distributed only within channels. More specifically, messages shared only in groups have an average lifetime of 105 days while messages shared only in channels have an average lifetime of 17 days. Regular messages without URLs last longer than messages containing URLs. Regular messages, on average, have a lifespan that is twice as long as messages containing URLs. Among all of these URLs, the ones referring to messaging platforms last longer than other types of URLs while news-related URLs fade away more quickly than the others. More specifically, URLs with the “Text messaging” category exhibit the most extensive lifespans, enduring an average of 54 days. In contrast, URLs with the “News and media” category possess the shortest lifespans, enduring an average of 7 days. Messages that are toxic or exhibit extreme emotions tend to last longer compared to those that are non-toxic and emotionally neutral.

Case Studies

After examining various aspects of the messages, in this section, we analyze five representative messages from our dataset to provide further insights into the internal dynamics of the network. Our approach for selecting these five messages comprises of three steps: 1) Text message preprocessing, 2) Extraction of the top five prevalent topics, and 3) Selection of five representative messages.

Text message preprocessing. We exclude URLs and messages containing fewer than five words to ensure the analysis focuses on meaningful content.

Extraction of the top five prevalent topics. As the discussions within the chats in our dataset are in multiple languages, we use a Bidirectional Encoder Representations from Transformers (BERT)-based topic modeling methodology by Angelov (2020) to extract the topics. This model supports 50 different languages and performs well in handling multilingual datasets. Our topic modeling technique utilizes transformer-based embeddings. Before feeding our corpus into the BERTopic model, we first need to transform our raw text data into a format the model can understand which is embeddings. SentenceTransformer and “all-

MiniLM-L6-v2" are utilized to produce embeddings. After embedding documents from multiple languages into a vector representation of the data, we reduce the dimensions of the embeddings using Uniform Manifold Approximation and Projection (UMAP) proposed by (McInnes, Healy, and Melville 2018). Then, we cluster the reduced embeddings using the HDBSCAN algorithm applying the method presented in (McInnes, Healy, and Astels 2017). Finally, we represent the topics from each cluster. The top five frequent topics among the messages are: QAnon, COVID-19, US politics, German politics, and other conspiracy theories.

Selection of five representative messages. For each topic, we select text messages that fall within the top 5% based on four criteria: frequency of occurrence, lifetime, number of senders, and number of chats in which the message appeared. Consequently, a single message is chosen from this filtered subset for our case study. Below, we elaborate on the five sample messages and compare them.

Case 1 (QAnon) *"An anon kindly translated the show. In the first chart, the facilitator tells which well-known people had a black eye and that the black eye is caused by taking Adrenochrome. In the second chart with the children, he describes how to get Adrenochrome. After severe torture, blood is drawn from the children at the time of death."*

Case 2 (COVID-19) *"The FBI arrested a Boston University professor linked to a Chinese University and Research Lab in Wuhan, who was highly paid by China. Obviously, the coronavirus is a planned bio-attack from China. A Chinese expert assures that inhaling the steam of hot water kills the Coronavirus 100 percent."*

Case 3 (US Politics) *"Breaking news Biden tortured and raped children! Trump's attorney Giuliani had previously implied it and what the New York Post understandably refused to publish now seems more confirmed. Videos and photos on Hunter Biden's laptop are said to show him sexually abusing, raping, and cruelly torturing small, underage Chinese children."*

Case 4 (German Politics) *"It is not a coincidence. Only a few days after the threat of a constitutional lawsuit, the president of the Hamburg Hotel, Franz J. Klein, is dead. Klein threatened Angela Merkel with a lawsuit before the Constitutional Court and criticized her interference with fundamental rights. We are now talking about a series of mysterious deaths of bitter Corona policy opponents."*

Case 5 (Other Conspiracy Theories) *"Arrest Bill Gates. In Texas, people demonstrated against compulsory vaccination, and for the arrest of vaccination lobbyist Bill Gates. Posters read: 'Bill Gates is a Freemason and devil worshiper.' or 'Freedom is better than Fear.'"*

All five sample messages were originally in German and have been translated into English. These samples promote conspiracy theories on different topics. These examples of misinformation reflect some of the common characteristics of fake news. They try to create strong emotions in the audience using sensitive issues such as "torturing children" (Case 1&3). They also make big claims and accuse people with no proof or reference such as portraying natural or accidental deaths as deliberate acts of murder (Case 4).

Case Study	#Messages	#Chats	#Senders	#Forwarded messages
Case 1	94	85	88	92
Case 2	557	378	475	534
Case 3	170	129	135	167
Case 4	78	67	71	77
Case 5	126	105	112	119

Table 2: Case studies overview.

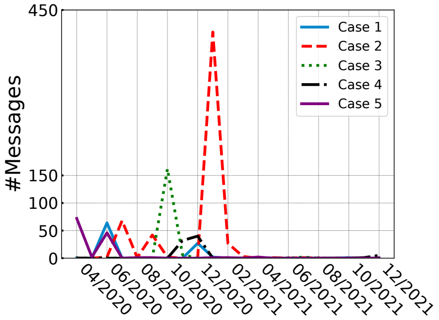

Figure 13: Number of monthly messages for each case.

As we see in Table 2, an overwhelmingly high proportion of messages in all the cases are forwarded messages (94%-99%). This demonstrates that forwarding is a primary mechanism for information propagation on this platform, especially regarding misinformation and conspiratorial content. The prominence of forwarding raises concerns about the potential for rapid and widespread dissemination of misinformation. Due to the ease and speed at which a message can be forwarded, misinformation can quickly reach many users. Also, when messages are forwarded from trusted contacts or groups, they may be perceived as more credible, leading to a higher likelihood of acceptance and further forwarding.

The sample messages have lasted about one year in our dataset. Fig. 13 shows the number of appearances of the messages in each month during 21 months of the lifetime. Typically, the trend observed in the dissemination of messages containing misinformation exhibits a single peak. These messages are widely shared, promoted, forwarded, and popular. In certain instances, previously circulated misinformation may regain popularity due to events in the real world that relate to their content. For instance, misinformation about COVID-19 experienced a third significant increase, as shown in the figure, in January 2021. This increase coincides with the global start of COVID-19 vaccination. The lifetimes of these sample cases indicate that different types of misinformation could live for a long time and be discussed within the platform.

These case studies underscore the extent of misinformation and conspiracy theory propagation within Telegram's fringe communities. They highlight the need for further research and potentially targeted interventions to curb the spread of such harmful content.

Conclusion

In this paper, we performed a large-scale analysis to measure information propagation within the Telegram network. We collected a large-scale dataset of about 140 million messages shared on over 9,000 public Telegram chats. Then, we undertake an analysis to understand how Telegram users use the forwarding feature to propagate information across chats, as well as a lifespan analysis to analyze how persistent and long-lived content is on Telegram.

Among other things, we find that a small percentage of users (6%) are responsible for 90% of all the forwarded messages in our dataset, which indicates that within the Telegram platform, there is a small percentage of users that are "superspreaders" of content. This critical finding can have significant implications given that Telegram is also exploited nowadays for disseminating potentially harmful information, such as hateful content or misinformation. For instance, platforms like Telegram can potentially moderate a few users who are actively forwarding a large amount of harmful content, which will significantly decrease the spread of harmful content within the Telegram network.

Also, our analysis shows significant differences in forwarding behavior based on the type of chat (group or channel). In particular, based on our dataset, we find it more likely that a forwarded message originates from a channel rather than a group. At the same time, we find that groups are the recipients of more forwarded messages compared to channels (in 50% of the groups, we find more than 40% of the messages being forwarded, while for channels, we find only 20%). Through reachability analysis, we found that despite the localized dissemination of messages within our dataset, the forwarding feature plays a significant role in expanding their reach. Finally, when it comes to lifespan, we find that, in general, messages shared in groups have a larger lifespan compared to messages shared in channels. Investigating a sample set of messages indicates the risk of rapid and widespread misinformation distribution due to the simplicity and quickness of the forwarding mechanism.

This study provides insight into understanding information propagation among Telegram groups and channels. The next steps can include investigating the interconnections between different social media platforms. The analysis of how information is consumed or supplied by other platforms sheds more light on the big picture of information propagation in the online world.

References

Abu-Salma, R.; Krol, K.; Parkin, S.; Koh, V.; Kwan, K.; Mahboob, J.; Traboulsi, Z.; and Sasse, M. A. 2017. The Security Blanket of the Chat World: An Analytic Evaluation and a User Study of Telegram. Internet Society.

Aliapoulios, M.; Bevensee, E.; Blackburn, J.; De Cristofaro, E.; Stringhini, G.; and Zannettou, S. 2021. An Early Look at the Parler Online Social Network. In *ICWSM*.

Andryukhin, A. 2019. Phishing Attacks and Preventions in Blockchain Based Projects. In *2019 International Conference on Engineering Technologies and Computer Science (EnT)*, 15–19.

Angelov, D. 2020. Top2vec: Distributed Representations of Topics. *arXiv:2008.09470*.

Arun, C. 2019. On WhatsApp, rumours, lynchings, and the Indian Government. *Economic & Political Weekly*, 54(6).

Bowles, J.; Larreguy, H.; and Liu, S. 2020. Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe. *PLOS ONE*, 15(10): 1–11.

Cheeseman, N.; Fisher, J.; Hassan, I.; and Hitchen, J. 2020. Social Media Disruption: Nigeria's WhatsApp Politics. *Journal of Democracy*, 31(3): 145–159.

Choi, D.; Chun, S.; Oh, H.; Han, J.; Kwon, T.; et al. 2020. Rumor propagation is amplified by echo chambers in social media. *Scientific reports*, 10(1): 1–10.

Clifford, B.; and Powell, H. 2019. Encrypted extremism: Inside the English-speaking Islamic state ecosystem on Telegram. *The George Washington University Program on Extremism*.

Dargahi Nobari, A.; Reshadatmand, N.; and Neshati, M. 2017. Analysis of Telegram, an instant messaging service. In *CIKM*.

Davidson, T.; Bhattacharya, D.; and Weber, I. 2019. Racial bias in hate speech and abusive language detection datasets. In *Third Workshop on Abusive Language Online*.

Elias, C.; and Catalan-Matamoros, D. 2020. Coronavirus in Spain: Fear of 'Official' Fake News Boosts WhatsApp and Alternative Sources. *Media and Communication*, 8(2): 462–466.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Forte Martins, A. D.; Cabral, L.; Chaves Mourão, P. J.; Monteiro, J. M.; and Machado, J. 2021. Detection of Misinformation About COVID-19 in Brazilian Portuguese WhatsApp Messages. In *Natural Language Processing and Information Systems*, 199–206.

Gao, B.; Wang, H.; Xia, P.; Wu, S.; Zhou, Y.; Luo, X.; and Tyson, G. 2021. Tracking Counterfeit Cryptocurrency End-to-End. *Measurement and Analysis of Computing Systems*, 4(3).

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Grindrod, P.; and Bovet, A. 2022. Organization and evolution of the UK far-right network on Telegram. *Applied Network Science*.

Guhl, J.; and Davey, J. 2020. A safe space to hate: White supremacist mobilisation on telegram. *Institute for Strategic Dialogue*, 1–20.

Hamrick, J.; Rouhi, F.; Mukherjee, A.; Feder, A.; Gandal, N.; Moore, T.; and Vasek, M. 2021. An examination of the cryptocurrency pump-and-dump ecosystem. *Information Processing & Management*, 58(4): 102506.

Hashemi, A.; and Chahooki, M. A. Z. 2019. Telegram group quality measurement by user behavior analysis. *Social Network Analysis and Mining*, 9(1): 33.

Hitchen, J.; Fisher, J.; Hassan, I.; and Cheeseman, N. 2019. Whatsapp and Nigeria's 2019 Elections: Mobilising the People, Protecting the Vote. *Portal Africa*.

Hoseini, M.; Melo, P.; Benevenuto, F.; Feldmann, A.; and Zannettou, S. 2023. On the Globalization of the QAnon Conspiracy Theory Through Telegram. In *WebSci*, 75–85.

Hoseini, M.; Melo, P.; Júnior, M.; Benevenuto, F.; Chandrasekaran, B.; Feldmann, A.; and Zannettou, S. 2020. Demystifying the Messaging Platforms' Ecosystem Through the Lens of Twitter. In *IMC*.

Hou, Y.; Wang, H.; and Wang, H. 2022. Identification of Chinese dark jargons in Telegram underground markets using context-oriented and linguistic features. *Information Processing & Management*, 59(5): 103033.

Javed, R. T.; Usama, M.; Iqbal, W.; Qadir, J.; Tyson, G.; Castro, I.; and Garimella, K. 2022. A deep dive into COVID-19-related messages on WhatsApp in Pakistan. *SNAM*, 12(1): 1–16.

Kansaon, D. P.; Melo, P. D. F.; and Benevenuto, F. 2022. "Click Here to Join": A Large-Scale Analysis of Topics Discussed by Brazilian Public Groups on WhatsApp. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, 55–65.

Kazemi, A.; Garimella, K.; Shahi, G. K.; Gaffney, D.; and Hale, S. A. 2022. Research note: Tiplines to uncover misinformation on encrypted platforms: A case study of the 2019 Indian general election on WhatsApp. *HKS Misinformation Review*.

La Morgia, M.; Mei, A.; Mongardini, A. M.; and Wu, J. 2021. Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. *arXiv preprint arXiv:2111.13530*.

Leong, D. 2020. 700 Million Users and Telegram Premium. telegram.org/blog/700-million-and-premium/pt-br?setln=en. Acessed on Nov. 28, 2022.

Lima, L.; Reis, J. C.; Melo, P.; Murai, F.; Araujo, L.; Vikatos, P.; and Benevenuto, F. 2018. Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system. In *ASONAM*.

Londoño, E. 2021. Brazil's Far-Right Disinformation Pushers Find a Safe Space on Telegram. https://www.nytimes.com/2021/11/08/world/americas/brazil-telegram-disinformation.html.

Loureiro, D.; Barbieri, F.; Neves, L.; Anke, L. E.; and Camacho-Collados, J. 2022. TimeLMs: Diachronic Language Models from Twitter. *arXiv preprint arXiv:2202.03829*.

Malhotra, P. 2020. A Relationship-Centered and Culturally Informed Approach to Studying Misinformation on COVID-19. *Social Media + Society*, 6(3): 2056305120948224.

McInnes, L.; Healy, J.; and Astels, S. 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11): 205.

McInnes, L.; Healy, J.; and Melville, J. 2018. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426*.

Melo, P.; Vieira, C. C.; Garimella, K.; de Melo, P. O. V.; and Benevenuto, F. 2019. Can WhatsApp Counter Misinformation by Limiting Message Forwarding? In *International Conference on Complex Networks and Their Applications*, 372–384. Springer.

Mirtaheri, M.; Abu-El-Haija, S.; Morstatter, F.; Steeg, G. V.; and Galstyan, A. 2021. Identifying and Analyzing Cryptocurrency Manipulations in Social Media. *IEEE Transactions on Computational Social Systems*, 8(3): 607–617.

Morgia, M. L.; Mei, A.; Sassi, F.; and Stefa, J. 2022. The Doge of Wall Street: Analysis and Detection of Pump and Dump Cryptocurrency Manipulations. *ACM Trans. Internet Technol.*

Naseri, M.; and Zamani, H. 2019. Analyzing and Predicting News Popularity in an Instant Messaging Service. In *SIGIR*, 1053–1056.

Nikkhah, S.; Miller, A. D.; and Young, A. L. 2018. Telegram as An Immigration Management Tool. In *Companion of CSCW*, 345–348.

Nizzoli, L.; Tardelli, S.; Avvenuti, M.; Cresci, S.; Tesconi, M.; and Ferrara, E. 2020. Charting the Landscape of Online Cryptocurrency Manipulation. *IEEE Access*, 8: 113230–113245.

Ostrovsky, A. M.; and Chen, J. R. 2020. TikTok and Its Role in COVID-19 Information Propagation. *Journal of adolescent health*.

Pasquetto, I. V.; Olivieri, A. F.; Tacchetti, L.; Riotta, G.; and Spada, A. 2022. Disinformation as Infrastructure: Making and maintaining the QAnon conspiracy on Italian digital media. *CSCW*.

Peeters, S.; and Willaert, T. 2022. Telegram and Digital Methods: Mapping Networked Conspiracy Theories through Platform Affordances. *M/C Journal*, 25(1).

Perspective API. 2018. https://www.perspectiveapi.com/.

Prucha, N. 2016. IS and the Jihadist Information Highway–Projecting Influence and Religious Identity via Telegram. *Perspectives on Terrorism*, 10(6).

Reis, J. C.; Melo, P.; Garimella, K.; and Benevenuto, F. 2020. Can WhatsApp benefit from debunked fact-checked stories to reduce misinformation? *HKS Misinformation Review*.

Resende, G.; Melo, P.; C. S. Reis, J.; Vasconcelos, M.; Almeida, J. M.; and Benevenuto, F. 2019a. Analyzing Textual (Mis)Information Shared in WhatsApp Groups. In *WebSci*.

Resende, G.; Melo, P.; Sousa, H.; Messias, J.; Vasconcelos, M.; Almeida, J.; and Benevenuto, F. 2019b. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *The World Wide Web Conference*, 818–828.

Ribeiro, M. H.; Jhaver, S.; Zannettou, S.; Blackburn, J.; De Cristofaro, E.; Stringhini, G.; and West, R. 2021. Does Platform Migration Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. In *CSCW*.

Rivers, C. M.; and Lewis, B. L. 2014. Ethical research standards in a world of big data. *F1000Research*, 3.

Shehabat, A.; Mitew, T.; and Alzoubi, Y. 2017. Encrypted jihad: Investigating the role of Telegram App in lone wolf attacks in the West. *Journal of Strategic Security*, 10(3): 27–53.

Solopova, V.; Scheffler, T.; and Popa-Wyatt, M. 2021. A Telegram corpus for hate speech, offensive language, and online harm. *Journal of Open Humanities Data*, 7.

Stokel-Walker, C. 2022. Russia's battle to convince people to join its war is being waged on Telegram. https://www.technologyreview.com/2022/09/24/1060005/russia-war-ukraine-conscription-telegram/. Acessed on Nov. 28, 2022.

Trujillo, M.; Gruppi, M.; Buntain, C.; and Horne, B. D. 2020. What is BitChute? Characterizing the "Free Speech" Alternative to YouTube. In *HT*.

Urman, A.; and Katz, S. 2022. What they do in the shadows: examining the far-right networks on Telegram. *Information, Communication & Society*, 25(7): 904–923.

Varanasi, R. A.; Pal, J.; and Vashistha, A. 2022. Accost, accede, or amplify: attitudes towards COVID-19 misinformation on WhatsApp in India. In *CHI*, 1–17.

Vasudeva, F.; and Barkdull, N. 2020. WhatsApp in India? A case study of social media related lynchings. *Social Identities*.

Vijaykumar, S.; Jin, Y.; Rogerson, D.; Lu, X.; Sharma, S.; Maughan, A.; Fadel, B.; de Oliveira Costa, M. S.; Pagliari, C.; and Morris, D. 2021. How shades of truth and age affect responses to COVID-19 (Mis) information: randomized survey experiment among WhatsApp users in UK and Brazil. *Humanities and Social Sciences Communications*, 8(1): 1–12.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *science*, 359(6380): 1146–1151.

Walther, S.; and McCoy, A. 2021. US Extremism on Telegram: Fueling Disinformation, Conspiracy Theories, and Accelerationism. *Perspectives on Terrorism*, 15(2): 100–124.

Yayla, A. S.; and Speckhard, A. 2017. Telegram: The mighty application that ISIS loves. *International Center for the Study of Violent Extremism*.

Zannettou, S.; Bradlyn, B.; De Cristofaro, E.; Kwak, H.; Sirivianos, M.; Stringini, G.; and Blackburn, J. 2018. What is gab: A bastion of free speech or an alt-right echo chamber. In *The Web Conference*.

Ethics Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socioeconomic divide, or implying disrespect to societies or cultures? Yes.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, we explain in the parts of the paper why the methods are suitable for the intended use (e.g., the Perspective API is a production-ready tool that supports many languages and the sentiment analysis tool performs with an acceptable performance in our dataset).
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? No, because as described in Section , we do not have access to representative samples from Telegram, hence we can not make any claims about the representativeness of the data.
   (e) Did you describe the limitations of your work? Yes, see Section .
   (f) Did you discuss any potential negative societal impacts of your work? We do not anticipate any negative societal impact of this work.
   (g) Did you discuss any potential misuse of your work? We do not anticipate a potential misuse of this work.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? NA
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? NA
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA