UNIVERSITÄT
DES
SAARLANDES

Universität des Saarlandes

Analyzing Information Propagation in Online Messaging Platforms

Dissertation zur Erlangerung des Grades des
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Fakultät für Mathematik und Informatik der
Universität des Saarlandes

vorgelegt von

Mohamad Hoseini

Saarbrücken, 2024

Day of Colloquium:	17 / 10 / 2024
Dean of the Faculty:	Prof. Dr. Roland Speicher
Chair of the Committee:	Prof. Dr. Sebastian Hack
Reporters:	Prof. Anja Feldmann, Ph. D.
	Prof. Dr. Ingmar Weber
	Prof. Dr. Savvas Zannettou
Academic Assistant:	Dr. Taha Albakour

Abstract

Online messaging platforms such as WhatsApp, Telegram, and Discord, each with hundreds of millions of users, hold a dominant position in the realm of communication. They play a pivotal role in shaping societal interactions. Concurrently, there is an alarming surge in the spread of misinformation, including conspiracy theories, which has tangible real-world consequences, including instances of violence. While mainstream platforms like Facebook and Twitter are intensifying content moderation to counter misinformation and fake news, users are migrating to platforms that offer more freedom, such as Telegram. However, there's a lack of systematic characterization across online messaging platforms and knowledge about users' privacy exposure. The substantial real-world impact of misinformation necessitates a study on its online dissemination to comprehend how conspiracy theories evolve and spread. Despite Telegram's widespread popularity and its role in attracting a substantial user base engaged in discussions about fringe ideologies, research on content dissemination within the platform remains limited. This raises questions about how conspiratorial content spreads among fringe communities on Telegram.

This thesis commences by addressing the lack of a comprehensive characterization across multiple platforms in the online messaging ecosystem, utilizing Twitter as the source to identify public groups. We discover public groups from WhatsApp, Telegram, and Discord shared on Twitter and monitor the characteristics of these groups to gain deeper insights into their discovery via Twitter and how they evolve. We also examine the exposure of personally identifiable information within these messaging platforms. Next, we conduct a multilingual and longitudinal analysis of Telegram groups associated with QAnon, a popular conspiracy theory. Furthermore, we delve into the QAnon discourse spanning multiple languages and uncover additional conspiracy theories that QAnon supporters engage in. Finally, we explore the information dissemination within fringe communities on Telegram. We analyze forwarding patterns and the lifespan of different types of messages. Furthermore, we examine samples of representative messages in our case study to offer additional insights into the internal dynamics of the network.

This thesis addresses the lack of comprehensive characterization of the online messaging ecosystem, particularly focusing on the propagation of misinformation and conspiracy theories within public groups. As researchers delve deeper into the intricate dynamics of online communication, this work serves as a stepping stone, encouraging further exploration and analysis. The methodologies employed and the insights derived contribute valuable perspectives on identifying and examining public groups that disseminate misinformation across online messaging platforms.

Zusammenfassung

Online-Nachrichtenplattformen wie WhatsApp, Telegram und Discord, die jeweils Hunderte Millionen von Nutzern haben, nehmen eine dominante Position im Bereich der Kommunikation ein. Sie spielen eine entscheidende Rolle bei der Gestaltung gesellschaftlicher Interaktionen. Gleichzeitig gibt es einen alarmierenden Anstieg der Verbreitung von Fehlinformationen, einschließlich Verschwörungstheorien, die spürbare Auswirkungen im realen Leben haben, einschließlich Gewalttaten. Während Mainstream-Plattformen wie Facebook und Twitter die Inhaltsmoderation verstärken, um Fehlinformationen und Fake-News entgegenzuwirken, wechseln Nutzer zu Plattformen, die mehr Freiheit bieten, wie Telegram. Es mangelt jedoch an einer systematischen Charakterisierung über verschiedene Online-Nachrichtenplattformen hinweg und an Wissen über die Privatsphäre der Nutzer. Der erhebliche Einfluss von Fehlinformationen in der realen Welt erfordert eine Untersuchung ihrer Online-Verbreitung, um zu verstehen, wie Verschwörungstheorien entstehen und sich verbreiten. Trotz der weit verbreiteten Beliebtheit von Telegram und seiner Anziehungskraft auf Nutzer, die sich mit Randideologien beschäftigen, ist die Forschung zur Inhaltsverbreitung auf der Plattform begrenzt. Dies wirft wichtige Fragen darüber auf, wie verschwörerischer Inhalt unter Randgemeinschaften auf Telegram verbreitet wird.

Diese Dissertation beginnt mit einer umfassenden Charakterisierung über mehrere Plattformen im Ökosystem der Online-Nachrichten, wobei Twitter als Quelle verwendet wird, um öffentliche Gruppen zu identifizieren. Wir entdecken öffentliche Gruppen von WhatsApp, Telegram und Discord, die auf Twitter geteilt werden, und überwachen die Merkmale dieser Gruppen, um tiefere Einblicke in ihre Entdeckung über Twitter und ihre Entwicklung im Laufe der Zeit zu erhalten. Wir untersuchen auch die Offenlegung personenbezogener Informationen innerhalb dieser Nachrichtenplattformen. Anschließend führen wir eine mehrsprachige und längsschnittliche Analyse von Telegram-Gruppen durch, die mit QAnon, einer beliebten Verschwörungstheorie, in Verbindung stehen. Darüber hinaus gehen wir auf den QAnon-Diskurs in mehreren Sprachen ein und decken zusätzliche Verschwörungstheorien auf, mit denen sich QAnon-Anhänger befassen. Schließlich untersuchen wir die Informationsverbreitung innerhalb von Randgemeinschaften auf Telegram. Wir analysieren Weiterleitungs-Muster und die Lebensdauer verschiedener Arten von Nachrichten. In einer Fallstudie untersuchen wir stichprobenartig repräsentative Nachrichten, um zusätzliche Einblicke in die internen Dynamiken des Netzwerks zu geben.

Diese Arbeit adressiert den Mangel an umfassender Charakterisierung des Online-Nachrichtenökosystems, wobei der Schwerpunkt auf der Verbreitung von Fehlinformationen und Verschwörungstheorien innerhalb öffentlicher Gruppen liegt. Während Forscher tiefer in die komplexen Dynamiken der Online-Kommunikation eindringen, dient diese Arbeit als Ausgangspunkt und ermutigt zu weiterer Exploration und Analyse. Die angewandten Methoden und die gewonnenen Erkenntnisse tragen wertvolle Perspektiven zur Identifizierung und Untersuchung öffentlicher Gruppen bei, die Fehlinformationen über verschiedene Online-Nachrichtenplattformen verbreiten.

Acknowledgements

Earning a PhD has been an experience filled with both challenges and achievements. Throughout this journey, I've been fortunate to have the support of many incredible individuals who turned potential obstacles into stepping stones towards growth.

First of all, I express my profound gratitude to my advisor, Anja Feldmann, for her constant guidance, support, and mentorship. Her expertise and insight have been invaluable, influencing not just this work but also my growth as a researcher.

I am deeply thankful to my primary collaborator, Savvas. His remarkable contributions and insights greatly enriched this thesis. Collaborating with him has been both a learning experience and a privilege.

To my wife, Aniss: your constant encouragement, patience, and understanding have been my anchor throughout this PhD journey. Your belief in me has been a cornerstone of my strength, and I am grateful for your perpetual support.

I'd like to express my gratitude to the collaborators from UFMG, Fabricio and Philipe. Their knowledge and diverse viewpoints added significant value to this research.

I extend warm appreciation to my friends and colleagues at INET. Their combined insights, valuable feedback, and fellowship have made this journey enjoyable and rewarding. Thank you Seif, Emilia, Danesh, Fariba, Florian Steurer, Florian Streibelt, Franziska, Malte, Daniel, Pascal, Ali, and, all other members of the INET family. I want to acknowledge the valuable technical discussions and support I've received from Artin Saberpour and Mahmoudreza Babaei at MPI.

Last but not least, I owe a heartfelt thank you to my parents. Their endless love, support, and sacrifices laid the foundation for all my achievements. Their teachings and values have always guided me, and this accomplishment is as much theirs as it is mine.

Publications

The core three phases of this thesis find their basis in the following peer-reviewed papers that are published in international conferences. All my collaborators are listed among the co-authors.

- Mohamad Hoseini, Philipe Melo, Manoel Júnior, Fabrício Benevenuto, Balakrishnan Chandrasekaran, Anja Feldmann, and Savvas Zannettou. "Demystifying the Messaging Platforms' Ecosystem Through the Lens of Twitter". In: *The 2020 Internet Measurement Conference (IMC)*. 2020, pp. 345–359

- Mohamad Hoseini, Philipe Melo, Fabricio Benevenuto, Anja Feldmann, and Savvas Zannettou. "On the globalization of the QAnon conspiracy theory through Telegram". In: *Proceedings of the 15th ACM Web Science Conference*. 2023, pp. 75–85

- Mohamad Hoseini, Philipe Melo, Fabricio Benevenuto, Anja Feldmann, and Savvas Zannettou. "Characterizing Information Propagation in Fringe Communities on Telegram". In: The International AAAI Conference on Web and Social Media (ICWSM). 2024.

Contents

1 Introduction 1
1.1 Contributions 8
1.2 Published Papers 11
1.3 Structure of the Thesis 11

2 Background 13
2.1 Terms and Concepts 13
2.2 Related Work 17

3 Demystifying the Messaging Platforms' Ecosystem 19
3.1 Background and Related Work 21
3.2 Methodology and Dataset 22
3.2.1 Discovering WhatsApp, Telegram, and Discord Groups 23
3.2.2 Collecting Group-Specific Metadata 24
3.2.3 Analyzing group composition and activity 25
3.3 Results 25
3.3.1 Discovering Public Groups 26
3.3.2 Activity and Evolution of Public Groups 32
3.3.3 Privacy Implications 37
3.4 Discussion 40

4 On the Globalization of QAnon Through Telegram 43
4.1 Background and Related Work 45
4.2 Methodology and Dataset 47
4.3 Results 50
4.3.1 Activity (Hypothesis 1) 50
4.3.2 Toxicity (Hypothesis 2) 53
4.3.3 Topics (Hypothesis 3) 56
4.4 Discussion 65

5 Characterizing Information Propagation in Fringe Communities on Telegram 69
5.1 Background and Related Work 71
5.2 Methodology and Dataset 73
5.3 Results 74
5.3.1 Forwarding 74
5.3.2 Life Span 80
5.3.3 Case Studies 85
5.4 Discussion 87

6 Conclusion 89
6.1 Discussions and Implications 89
6.2 Limitations 91
6.3 Ethical Considerations 92
6.4 Future Work 92

6.5 Summary 94

Bibliography 95

List of Figures 111

List of Tables 113

1
Introduction

Internet technologies are significantly influencing content creation, communication, and social interactions among people as they become more prevalent and widely used. Social media, meeting people's socialization needs, has gained popularity, fostering complex relationships and the rapid dissemination of information in society. The emergence of social media has fundamentally transformed the way people communicate and interact. Platforms like Facebook, Twitter[1], Instagram, and Telegram are now integral parts of our daily lives, providing new ways for connection, information sharing, and relationship building. There are 5 billion active social media users worldwide [3], highlighting the significant impact of online platforms on society. Characterized by its expansive scale and continually increasing reach, the online social media ecosystem has emerged as a primary medium exerting a unique and profound influence on society. This influence has infiltrated diverse aspects of people's lives, including communication, entertainment, education, politics, and business. The importance of social media in society is multi-faceted, ranging from its effect on communication and information sharing to its impact on business practices and social dynamics. Given its broad influence and continued growth, understanding and engaging with social media is crucial in the modern world.

The focus of researchers has primarily been limited to exploring online social networks such as Facebook and Twitter. In social networking platforms, users establish connections with each other by linking their respective personal profiles. In contrast with social networking platforms, an online messaging platform is a type of digital communication channel that allows real-time transmission of messages among users. Within this environment, connections are in three forms: one-to-one, one-to-many, and many-to-many interactions [4]. Although there is plenty of research studying online social networks, online messaging platforms remain significantly under-explored. Online messaging platforms have attracted billions of users worldwide, demonstrating a swift growth in their user base. These platforms act as significant channels for the dissemination of news and information on a wide variety of subjects, including politics, business, and healthcare. The average user dedicates substantial time each day to these applications, becoming increasingly exposed to the information that circulates within these digital spaces, and inevitably influenced by it. Consequently, these platforms play a substantial role in influencing real-world events.

[1] The current name of the platform is X. We use its former name throughout this thesis, as the platform was referred to as Twitter during our work.

Despite the widespread adoption of online messaging platforms, many questions remain about their impact on individuals and society. The intertwining of personal and public communication on these platforms raises significant concerns regarding data privacy and the security of user information. Questions arise about the potential privacy leakage associated with disclosing Personally Identifiable Information on these platforms. With a massive global user base, online messaging platforms serve as effective mediums for the rapid spread of misinformation, impacting public perception and potentially influencing real-world events. It prompts inquiries into whether harmful content is disseminated worldwide through these platforms and if they are employed for spreading propaganda. The impact of online messaging platforms on political discourse, elections, and civic engagement underscores the need to comprehend their specific role in shaping political narratives. A critical examination of the effectiveness of content moderation policies becomes imperative, along with addressing the challenges these platforms encounter in consistently enforcing such policies. This involves assessing whether content moderation is implemented on these platforms and, if so, determining its utility and effectiveness. Furthermore, delving into the globalization of influence reveals inquiries into how these platforms contribute to shaping narratives across diverse cultures and societies. The nature of these platforms' contribution to the broader issues we face in digital communication is also somewhat ambiguous. There is anecdotal evidence showing that online messaging platforms positively and negatively affect different aspects of people's lives. Furthermore, these platforms have been used to spread false information, propaganda [5], hate speech [6], and conspiracy theories [7], highlighting the potential risks associated with their use. Empirical observations suggest the utilization of online messaging platforms, such as WhatsApp and Telegram, in propagating misinformation related to the COVID-19 pandemic [8], as well as fostering disinformation campaigns during political elections [9].

Given the limited exploration of online messaging platforms, the unique challenges presented by their usage remain largely undelineated. Understanding the challenges associated with these platforms and their roles in societal issues can provide insights into the dynamics of online interaction and information exchange, leading to potential strategies for improvement and intervention. Our objective is to delve into the challenges associated with online messaging platforms. We aim to conduct an exploration and characterization to gain deeper insights into their role in disseminating information among online users.

In this thesis, we set out to explore several facets of the online messaging platforms' ecosystem. Our investigation is segmented into three primary areas:

1. The discovery and characterization of groups within the platforms.
2. The evolution of conspiracy theories within the platforms.
3. The dynamics of misinformation dissemination within the platforms.

How can we systematically conduct a comprehensive characterization of the platforms' ecosystem?

To comprehensively study the online messaging ecosystem, obtaining access to a substantial number of public groups across different platforms is imperative for effective characterization and meaningful group comparisons. The primary challenge we encounter in our research is the absence of a vantage point to find public groups from online messaging platforms, which significantly complicates the study of these platforms. Our initial inquiry revolves around determining where and how we can systematically identify and access public groups from these platforms on a large scale. Therefore, our first research question is formulated as follows:

Research Question 1: How can we access public groups of online messaging platforms and their messages?

The emergence of mobile-specific online messaging platforms has stimulated the dissemination of information within a novel digital ecosystem that provides users with a more profound and immersive experience. The online messaging ecosystem is composed of a diverse range of online messaging platforms, each with its unique features, resulting in a complex network of communication channels. Our understanding of this ecosystem is currently deficient, leaving critical dimensions unexplored. Knowledge gaps exist regarding key aspects, including the number, size, and activity levels of groups within online messaging platforms. Moreover, there is limited insight into topics, linguistic diversity, and geographical distribution in group dynamics across different platforms. The temporal dynamics, evolution, and lifespan of these groups, including the duration of their accessibility, are also poorly understood. The absence of in-depth understanding also applies to the subtle features that differentiate one group from another. Factors such as size, levels of activity, and topics can differ across various groups and are susceptible to changes over time. Importantly, these changes may appear differently across distinct online messaging platforms. Addressing these gaps is crucial for obtaining a profound understanding of the online messaging platforms' ecosystem. Consequently, our second research question is posed as:

Research Question 2: What are the characteristics of the groups of online messaging platforms?

In online messaging platforms where users participate in both private and public conversations, privacy concerns take precedence. Users share personal information within these platforms, prompting concerns about the potential vulnerability of this information to exposure. The protection of personal information is likely to be the primary consideration for users when choosing online messaging platforms. Then, it is crucial to determine the presence of privacy leakage in these platforms, the types of information that may be disclosed publicly, and the distinctions between various platforms concerning the protection of users' personal information. Our third research question focuses on this critical aspect:

Research Question 3: Is there any leakage of Personally Identifiable Information (PII) in online messaging platforms?

How do conspiracy theories evolve within the platforms?

The spread of conspiracy theories in social media is of particular concern, as it can have severe consequences for individuals and society. Conspiracy theories can erode trust in institutions, fuel polarization, and even incite violence. For instance, the Pizzagate conspiracy theory was the driving factor behind a shooting at a pizzeria in Washington DC in 2016 [10]. The proliferation of conspiracy theories on the Internet makes it increasingly important to understand how they spread online and how they influence individuals in the real world. Our comprehension of conspiratorial content shared on online messaging platforms is limited due to difficulties in measuring these platforms and accessing their content on a large scale. There is limited research focused on studying misinformation on online messaging platforms. Previous work evaluates the functionality of fact-checking within WhatsApp public groups [11], proposes an approach to identify misinformation shared in public groups [12], and suggests strategies to combat misinformation [13]. Despite these studies, the content and popularity of conspiracy theories within public groups on online messaging platforms remain undiscovered, making it unclear how many individuals are exposed to such content and the extent of its influence. Moreover, the speed and patterns of growth of conspiracy theories, as well as their prevalence across different languages and communities, are not well-documented. Therefore, the fourth research question we are going to answer is as follows:

Research Question 4: How do conspiracy theories evolve over time and across languages?

One significant negative aspect of online communication is the presence of toxic content that possesses considerable potential for harm. Toxic content, prevalent in online platforms, goes beyond virtual spaces, affecting mental health, community dynamics, and societal discourse. Despite the presence of numerous guidelines and regulations on online social media platforms like Twitter and Reddit aimed at preventing the posting of toxic content, such platforms still contain instances of toxic content [14]. Understanding the toxicity level of discussions in public groups related to conspiracy theories helps shed light on the relationship between conspiratorial content and toxic discourse. The degree of toxicity within conspiracy-related discussions raises questions: To what extent are the discussions characterized by toxicity? Is there an observable change in the toxicity levels of the content over time? Additionally, an exploration into the variation of toxicity levels across different languages is crucial. Are there specific languages in which discussions tend to be more toxic? Exploring the dynamics of toxicity within these conspiracy-related groups requires delving into the temporal evolution and linguistic differentiations of the shared content. Accordingly, our fifth research question is outlined as:

Research Question 5: How toxic are the conspiracy theories discussions over time and across languages?

Conspiracy theories proliferate across various topics in online communication, and the dissemination of such content carries significant consequences for individuals, groups, and societies. There is evidence of real-world harmful movements caused by conspiracy theories disseminated among people in society, especially politically related conspiracy theories that lead to radicalized and extremist actions[15] or violent intentions[16]. Within public groups of online messaging platforms, users engage in discussions on a diverse array of topics. Our focus is on shedding light on the discussions within groups associated with conspiracy theories. Do participants exclusively discuss the particular conspiracy theory at hand, or do they delve into other conspiracy theories as well? According to Goertzel's study, individuals who believe in one conspiracy theory are more likely to believe others [17]. Additionally, the themes of conversations may vary between conspiracy groups and other types of groups. The sixth research question we are going to answer is:

Research Question 6: What topics and conspiratorial content are popular among fringe communities?

How does misinformation propagate on messaging platforms?

A vast number of messages circulate within public groups on online messaging platforms every day, and among them, certain messages get viral quickly on a large scale. Viral messages have the potential to rapidly reach a broad audience, gaining widespread attention. The information conveyed to such a large audience plays a crucial role in shaping shared narratives, influencing perceptions of various issues, events, and individuals. This dissemination process, especially in the context of conspiracy theories, can contribute to spreading misinformation among a substantial number of people. In recent years, there has been growing interest in identifying and understanding the sources of the messages on online messaging platforms, particularly concerning the coordination for disseminating misinformation. We observe that many instances of false news spread rapidly on the Internet. Our understanding of how a message or a piece of information gets viral in a short time among a vast range of communities is limited. The forwarding feature is a highly effective mechanism for accelerating and expanding the viral spread of messages, particularly misinformation, within online messaging platforms. However, our understanding of how this forwarding feature specifically contributes to the spread of messages in fringe communities is limited. Furthermore, the dynamics of forwarding across different message types remain unclear. Messages with various characteristics, including toxicity, the inclusion of URLs, or distinct emotional tones, may experience diverse forwarding behaviors, necessitating in-depth investigation. Consequently, our seventh research question is formulated as:

Research Question 7: How does the forwarding feature contribute to information dissemination within fringe communities?

Within an online messaging platform, messages have the potential for reiteration, particularly through the utilization of the forwarding feature. Specific messages may experience increased frequency or an extended duration of recurrence, thereby increasing their influence. Messages with distinct characteristics might show varying lifespans. Saha et al. [18] find that messages containing fear speech tend to exhibit a longer lifespan in comparison to those without fear speech. According to [19], messages on WhatsApp that contain misinformation on WhatsApp demonstrate a significantly prolonged lifespan compared to those without misinformation. We aim to understand the longevity of various message types within our dataset, distinguishing those with extended lifespans from those with shorter durations. This investigation helps identify the characteristics that contribute to the prolonged persistence of messages. Then, our eighth research question is framed as:

Research Question 8: What is the lifespan of various types of content shared in fringe communities?

To answer the research questions mentioned above, a comprehensive step-by-step approach is adopted to systematically investigate and provide insights into each inquiry.

Demystifying the Online Messaging Platforms' Ecosystem.

The main challenge in analyzing online messaging platforms is to find groups and communities on these platforms at a large scale. We confront this challenge by leveraging Twitter as a means to pinpoint public groups from online messaging platforms. We focus on three widely used online messaging platforms: WhatsApp, Telegram, and Discord. These three online messaging platforms have established an intricate and diverse ecosystem that is widely used for spreading misinformation as well as typical news. We search for public groups associated with the three online messaging platforms on Twitter within a specific time frame. Given that Twitter serves as a rich source for these groups, we can identify a substantial number of distinct groups daily. We additionally conduct a comparison of the number of public groups identified on Twitter for each of the platforms against one another (RQ1).

We delve into the characterization of the three online messaging platforms to gain a deeper understanding of the online communication ecosystem. To examine the tweets containing group URLs, we assess language distribution in tweet texts and extract topics using topic modeling techniques. Additionally, we analyze data related to group size, creators, creation dates, and countries. Employing a monitoring system, we calculate group sizes and the duration of accessibility during our monitoring time window. Furthermore, we explore the activity within a sample set of groups, extracting information on the number and types of messages shared over time. Note that throughout these investigations, we compare the results across the three platforms (RQ2).

To examine the exposure of users' personal information, we retrieve user-related data from each platform utilizing its API. Our objective is to identify any potential privacy issues with online messaging platforms. We look for different kinds of publicly available information on these platforms such as phone numbers, accounts on other platforms, and email addresses. We seek various types of information publicly available from other users on these platforms, including phone numbers, email addresses, and accounts on other platforms (RQ3).

After characterizing three different online messaging platforms and comparing them to each other, we aim to obtain a better understanding of the evolution of theories spread inside an online messaging platform. In the second phase of our study, we explore the evolution of the QAnon conspiracy theory on Telegram.

On the Globalization of QAnon Through Telegram. To study the evolution of conspiracy theories in online messaging platforms, we focus on QAnon, one of the most popular conspiracy theories that has gained significant attention and support online. Over the recent years, the QAnon conspiracy theory has attracted an increasing number of followers globally, evolving into a movement with cult-like characteristics. As Telegram has become a destination for banned communities discussing and promoting conspiracy theories, including the QAnon movement, our study specifically centers on examining QAnon within the Telegram platform. First, we search for Telegram groups on Twitter and Facebook. Then, we collect messages shared inside these groups using Telegram API and select QAnon-related groups. We analyze the evolving activity within the groups over time, comparing it to a baseline dataset. Additionally, we explore the distribution of language among messages over time. This approach enables us to comprehend the growth patterns of the groups, discern periods of increased activity among QAnon supporters, and identify the associated countries and languages. Moreover, we seek correlations with real-world events to provide context to the observed trends (RQ4).

To perform toxicity analysis, we utilize the Google Perspective API [20] to extract toxicity scores for text messages. Employing a systematic methodology, we establish thresholds above which a piece of text is labeled as toxic. For each of the top languages, we determine the corresponding toxicity threshold. We then examine the percentage of toxic messages shared in each language over time, aiming to identify trends in the dissemination of toxic content. Our analysis includes a comparison of toxicity levels between messages in different languages and between the QAnon dataset and the baseline dataset. This approach allows us to identify languages with higher or lower toxicity, observe the time periods during which toxicity levels change among messages and across languages, and evaluate the relative toxicity of QAnon content compared to baseline content (RQ5).

To analyze the topics discussed within the groups, we first preprocess text messages and then employ a multilingual topic modeling technique to extract their underlying themes. We identify the top topics and their frequency over time in our dataset to find out the connection of QAnon content to other discussions. Additionally, we assess the popularity of these top topics across the most frequently used languages, gaining insights into the prevalence of specific themes in each language. In conjunction with topic modeling, we conduct qualitative analysis on a sample set of messages to gain a more profound understanding of the discussions (RQ6).

After analyzing the evolution of a famous conspiracy theory on Telegram, we aim to investigate how misinformation propagates among fringe communities on Telegram. In the third phase of our study, we analyze the dynamics of misinformation dissemination on Telegram, specifically concentrating on the impact of the forwarding feature and the lifespan of various message types.

Characterizing Information Propagation in Fringe Communities on Telegram. First of all, we choose to focus on Telegram primarily due to its relaxed moderation policies on sensitive and conspiratorial content. Moreover, there has been evidence indicating a migration of users from platforms such as Facebook and Twitter to Telegram [21]. This migration often occurs in response to more stringent moderation policies and account bans implemented on these other platforms.

To shed light on how frequently shared messages become popular among users of online messaging platforms, we collect and study messages from a huge set of sources of QAnon content on Telegram. Our examination encompasses diverse perspectives, enabling a comprehensive comparison between forwarded and direct messages from various angles. We extract user contribution patterns, addressing questions such as whether a small group of users is responsible for a disproportionate share of forwarded or direct messages. We extract all URLs in our dataset, resolve shortened URLs to retrieve their original long forms, extract their domains, and categorize each domain. Leveraging this information, we investigate the role of forwarding in URL dissemination and compare the distribution of URL categories between both direct and forwarded messages. Utilizing a machine learning approach, we extract sentiment from each text message, exploring the distribution of messages with different emotional tones among forwarded and direct messages. Simultaneously, we assess the toxicity of messages, comprehending how toxic messages differ in being forwarded compared to non-toxic ones. By calculating the number of groups reached by each message and comparing the distribution of reach between forwarded and direct messages, we demonstrate how forwarding contributes to expanding the reach of messages (RQ7).

We define the message's lifespan as the time interval between its initial and final appearances in our dataset. This indicates the duration for which the message persists in being reposted. Note that we conduct this analysis specifically for text messages that appear more than once in our dataset. We calculate the lifetime of each unique message that is repeated in the dataset, then delve into comparing the lifetimes of different message types to identify characteristics contributing to their longevity. The distribution of the lifetime of toxic and non-toxic messages, as well as messages with positive, negative, and neutral sentiment, is presented to examine the impact of toxicity and sentiment on their longevity. To explore the role of forwarding in extending the lifetime of messages, we assess the distribution of lifetimes among forwarded and direct messages. Additionally, we examine the lifetime of messages containing URLs in comparison to regular messages with no URLs to determine if sharing URLs has any impact on the persistence of messages (RQ8).

1.1 Contributions

Built upon a systematic methodology, this thesis conducts the collection and analysis of data from online messaging platforms, serving as the cornerstone for our comprehensive exploration. Our work significantly contributes to various dimensions of information propagation within online messaging platforms. The following discussion highlights the specific contributions made in this thesis, addressing distinct facets of the intricate landscape of information dissemination in online messaging platforms.

Demystifying the Online Messaging Platforms' Ecosystem. We comprehensively characterize three online messaging platforms—WhatsApp, Telegram, and Discord—analyzing and comparing various aspects of their groups, including composition, activity, evolution, ephemerality, topics, and privacy. Our study uncovers key insights into how these groups are discovered via Twitter and how they evolve over time. Our investigation highlights the richness of Twitter as a source for discovering group URLs on these platforms. During our data collection period, we detect a substantial number of new groups daily — 1,000 WhatsApp groups, 2,000 Telegram groups, and 6,000 Discord groups, on average (RQ1).

Examining the ephemerality of group URLs, we find that 27% of WhatsApp, 20% of Telegram, and 68% of Discord group URLs become inaccessible within 38 days. Analyzing the content of tweets containing group URLs, we characterize the differences in group themes across platforms. Notably, WhatsApp and Telegram host numerous groups dedicated to cryptocurrency discussions, while Telegram stands out for its extensive content on sex and pornography, and Discord groups focus predominantly on gaming and hentai (RQ2).

We identify privacy leaks on all three platforms, with WhatsApp exhibiting a higher prevalence than Telegram and Discord. Specifically, our findings reveal the exposure of sensitive personally identifiable information (PII) through WhatsApp, Telegram, and Discord groups. Over 54K WhatsApp users' phone numbers are identified, covering the entire user base discovered in our research. Telegram exposes a substantially lower number, with only less than 1% of the discovered users. Discord, while not revealing user phone numbers, discloses other social media accounts linked to users' Discord profiles, with 30% of Discord users having at least one linked social media account (RQ3).

On the Globalization of QAnon Through Telegram. We present the first large-scale multilingual analysis of QAnon discussions through Telegram by collecting 4.4 million messages posted in 161 QAnon groups/channels. Our analysis reveals several key insights about the QAnon activity in our dataset. In 2021, there is a significant spike in QAnon activity, showing nearly a 5x increase in both messages and senders. This growth contrasts sharply with our baseline dataset, which only saw a 2x increase. Additionally, after June 2020, German-language QAnon content surpasses its English counterpart in popularity, with averages of 55% and 28%, respectively (RQ4).

When we delve into the toxicity levels of the messages, we observe that content in Portuguese and German exhibits higher toxicity levels compared to English (8.6% of the Portuguese messages and 2.8% of the German messages are toxic, while only 1% of English messages are toxic). Moreover, the toxicity levels in English and Portuguese QAnon content are notably higher than in our baseline dataset, at 3.6 times and 1.2 times, respectively (RQ5).

Our multilingual topic modeling analysis highlights that QAnon has evolved into discussing various topics of interest within far-right movements across the globe. We find several topics of discussion like world politics, conspiracy theories, COVID-19, and the anti-vaccination movement. Beyond these quantitative findings, we undertake a qualitative analysis of a randomly selected sample of messages representing the top eight topics. This analysis is conducted to gain insight into the content of discussions within these groups. Our observations indicate that QAnon followers on

Telegram engage in sharing and discussing a broad spectrum of topics, disseminating conspiratorial and false information related to Politics and the COVID-19 pandemic. Furthermore, our analysis highlights a growing diversity within the QAnon discourse. Additionally, our findings suggest that most topics transcend linguistic boundaries, manifesting across multiple languages (RQ6).

Characterizing Information Propagation in Fringe Communities on Telegram.

We conduct an extensive analysis of forwarding patterns within a substantial dataset composed of public groups and channels on Telegram. Our dataset comprises approximately 140 million messages collected from around 9,000 channels and groups on Telegram. We analyze different aspects of forwarded messages along with their life span. Our analysis of the creators of the content reveals compelling patterns of user behavior and content dissemination. A mere 6% of users are responsible for generating 90% of all forwarded messages, suggesting a significant disparity in content creation. We further discover that the majority of popular messages predominantly originate from a single source. These observations underscore 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. Examining the dynamics between Telegram's groups and channels, we noted distinct differences. While groups are the recipients of a larger share of forwarded messages, channels seem to be the primary originators of content that gets forwarded. In fact, over half of the forwarded messages in groups consist of shared content, with more than 50% of these groups having around 40% of such messages. A closer look at the content being forwarded reveals that about 35% of these messages contain URLs. Surprisingly, over half of these URLs can be traced back to news outlets and two major social media platforms: "YouTube" and "Twitter". We observe that while the messages are disseminated locally, the forwarding feature has increased their reach significantly (RQ7).

While forwarded messages experience a higher repetition rate than direct messages, they also disappear more quickly compared to direct messages. Assessing the lifetime of messages shared in different types of chat, we observe a notably prolonged duration for messages exclusively disseminated within groups compared to those confined to channels. Further examination reveals that regular messages without URLs exhibit a longer lifespan than messages containing URLs, with the former lasting on average twice as long. Among all URLs, those pointing to online messaging platforms demonstrate a more extended duration than other types. A deeper dive into the nature of the messages indicates that those characterized by toxicity or extreme emotional content, whether positive or negative, have a longer lifetime compared to their neutral counterpart (RQ8).

1.2 Published Papers

Different parts of the work we present in this thesis are published at academic conferences. Our study is a collaborative process, engaging the insights of multiple researchers. In this section, we outline the primary contributions of the thesis author to the published papers and the thesis.

Demystifying the Online Messaging Platforms' Ecosystem.

Chapter 3 presents this work published at IMC 2020 in collaboration with other co-authors [1]. The main contributions of the author include (a) searching on Twitter and collecting invite URLs of public groups from three online messaging platforms, WhatsApp, Telegram, and Discord, (b) developing a monitoring system to collect metadata from WhatsApp, (c) extracting topics from teet texts using topic modeling techniques, and (d) analyzing data and visualizing the results.

On the Globalization of QAnon Through Telegram.

This work, explained in Chapter 4, is published at WebSci 2023 in collaboration with other co-authors [2]. The main contributions of the author include (a) searching on Twitter and collecting invite URLs of Telegram public groups, (b) collecting messages from Telegram groups, (c) validating Google Perspective API for multiple languages, (d) analyzing data and visualizing the results, and (e) performing qualitative analysis on a sample set of messages.

Characterizing Information Propagation in Fringe Communities on Telegram.

Detailed in Chapter 5, this work has been accepted for publication at ICWSM 2024 in collaboration with other co-authors. The main contributions of the author include (a) identifying the source chats from Telegram and collecting their messages, (b) resolving URLs, extracting their domains, and extracting the categories of the domains, (c) conducting toxicity analysis, (d) performing sentiment analysis, (e) analyzing data and visualizing the results, (f) implementing multilingual topic modeling, and (g) executing a case study.

1.3 Structure of the Thesis

The remainder of this thesis is organized into five main chapters. Chapter 2 provides background and relevant information on the online messaging platforms' environments and information propagation within these environments. This chapter also includes a literature review to contextualize the research within the existing body of knowledge. Chapter 3 explains our work on characterizing and comparing three online messaging platforms, namely WhatsApp, Telegram, and Discord. Chapter 4 presents a large-scale multilingual study on the globalization of the QAnon conspiracy theory through Telegram. Chapter 5 describes the work on characterizing information propagation in fringe communities on Telegram. The focus of this chapter is to analyze how the forwarding feature contributes to the spread of popular messages. Chapter 6 concludes the thesis by summarizing the main findings and contributions of the study. The chapter also offers implications of the research for theory and practice and highlights the significance of the study for the field.

2
Background

In this chapter, we introduce important terms and concepts crucial for understanding the ecosystem of online messaging platforms. We explain various social media platforms involved in our work and discuss the role of online messaging platforms in the dissemination of information, particularly misinformation. We also provide an overview of related work.

2.1 Terms and Concepts

In this thesis, we conduct extensive research on data collected from three widely used online messaging platforms: WhatsApp, Telegram, and Discord.

Online Messaging Platforms.

"Online messaging platforms" refer to digital services, particularly mobile applications, that enable users to exchange messages and communicate over the Internet. These platforms facilitate real-time text-based conversations and sharing of various types of messages, including images, videos, and support voice or video calls for both individuals and groups. Prominent examples of online messaging platforms include WhatsApp, Telegram, WeChat, and Discord. In the digital age, these platforms have evolved into indispensable tools for both personal and professional communication, providing a convenient and instant means of staying connected.

WhatsApp.

Launched in January 2009, WhatsApp is the largest online messaging platform with over 2 billion users [22]. It is also the second most used social media platform, following Facebook [23]. To use the platform, users must register with their phone number. Users can also use the platform via WhatsApp's Web or desktop client, but these clients require the user's mobile phone also to be connected to the Internet. The platform supports both one-on-one chats and group chats–simultaneously with up to 257 users–through chat rooms or *groups*. Administrators of a group can add others to the group either by directly making them members of the group or by sharing a group URL (or an invite link) with them. Members of a group can share or forward information in a range of different formats including text, images, videos, documents, contacts, locations, and stickers. In addition to chats, the platform supports audio and video calls, and all communications on WhatsApp are secured using end-to-end encryption.

We include WhatsApp in our study for various reasons. First, as the largest online messaging platform, WhatsApp is the mainstream communication medium for billions of people. Second, prior work indicates (ab)use of WhatsApp for disseminating false information [24] and dissemination of hateful rhetoric can incite violence in the real world [25].

Telegram.

Launched in August 2013, Telegram is an online messaging platform with approximately 700 million monthly users [26]. Similar to WhatsApp, it requires users to register with their phone numbers, and after registration allows them to communicate using its Web or desktop clients. Unlike WhatsApp, users are not required to have their phone connected to the Internet while using the Web or desktop clients. Users can create two types of chat rooms: *channels* and *groups*. Channels support a few-to-many communication pattern, where the creator and the administrators of the channel can share information with the rest of the members, and do not impose a limit on the number of members per channel. Groups, in contrast to channels, facilitate a many-to-many communication pattern, where all members of the group can share information with one another, and impose a limit of $200,000$ members per group. Both groups and channels allow users to share and forward information in a wide range of formats. In addition to facilitating the sharing of various content types such as text, images, videos, audio files, and stickers, Telegram also enables users to engage in audio and video calls. Unlike WhatsApp, not all message exchanges are end-to-end encrypted. End-to-end encryption in Telegram is only available for "secret chats," which are device-specific communication channels. Users can access the secret-chat messages only from the devices on which the chat was created, and they cannot forward messages from secret chats. Telegram possesses a critical functionality known as forwarding, which empowers users to redistribute received messages to various chats without the necessity of downloading or re-uploading the content. This feature is instrumental in facilitating the rapid and extensive propagation of messages throughout the Telegram platform.

We include Telegram in our study owing to both its growing popularity and reports indicating exploitation of the platform by bad actors, e.g., white supremacists [27] and terrorists [28]. Telegram has also received relatively less attention in academic research.

Discord.

Though it started with a focus on providing an online gaming community, Discord is nowadays used by the general public for various purposes, even including education [29]. The platform was launched in May 2015, roughly two years after Telegram and six years after WhatsApp. In contrast to WhatsApp and Telegram, users can register with an email; the platform does not require users to provide a phone number. Users can create a *server* (or *guild*) and within it, several *channels*. After joining

Table 2.1: Characteristics of the three online messaging platforms.

Characteristic	WhatsApp	Telegram	Discord
Initial release date	January 2009	August 2013	May 2015
User base	2.7 Billion	700 Million	250 Million
Clients	Mobile, Desktop, Web	Mobile, Desktop, Web	Mobile, Desktop, Web
Registration method	Phone	Phone	Email
Options for public chats	Groups	Groups and Channels	Server
Max. #members in public chats	256	200,000 for groups (unlimited for channels)	250,000 (500,000 for verified servers)
Types of content supported	Text, Sticker, Image Video, Audio, Location Document, Contact	Text, Sticker, Image Video, Audio, Location Document, Contact	Text, Sticker, Image Video, Audio, Location Document, Contact
API for data collection?	No (only Business API)	Yes	Yes
Message forwarding?	Yes (up to 5 groups)	Yes	Only available via link and only for members
End-to-end encryption	Yes	Only for “secret” chats	No

a server, users can exchange messages on the server’s channels, which support many-to-many communication patterns similar to WhatsApp groups. Users can also make audio or video calls to other members. After joining a server, a user can exchange messages with other users on the server’s channels (i.e., channels support many-to-many communication patterns similar to the groups of WhatsApp) and make audio or video calls to other users. Administrators may also restrict access to specific channels to some users. Discord’s servers can have a large number of users–up to $250,000$ by default–and some (e.g., “verified” servers of organizations, artists, or games) can host up to $500,000$ users. Verified servers are intended for organizations, artists, or games (manually verified by Discord employees). Lastly, channels in Discord do not offer end-to-end encryption.

We include Discord in this work as it is a fast-growing online messaging platform and especially attracts the young population; analyzing the platform could shed light on the use or abuse of the messaging platform by the young demographic. Discord has been used for organizing extremist rallies, e.g., the “Unite the Right” rally in Charlottesville in 2017 [30], and for disseminating potentially harmful and sensitive material, e.g., revenge porn [31].

Table 2.1 presents a comparison of the characteristics of the three online messaging platforms.

To access public groups on these platforms, we utilize data available from two online social networks, Twitter and Facebook. Subsequently, we further explain the details of these social networking platforms.

Twitter.

Twitter is a microblogging and social networking service where users can share and engage with brief messages called “tweets”. Launched in 2006, Twitter has rapidly grown to become one of the world’s leading social media platforms, serving as a real-time public communications system that spans various sectors, including politics, entertainment, academia, and more. Initially, tweets were restricted to 140 characters, though this limit was doubled to 280 characters in 2017. The platform also supports the sharing of various forms of media, including photos, videos, GIFs, and URLs. It facilitates real-time discussions, debates, and news updates, distinguishing it as a dynamic and immediate platform. There are a bunch of features introduced by Twitter such as Hashtag, Retweet, and Mention. Hashtags, represented by the '#' symbol, are a notable feature of Twitter. They allow users to categorize their tweets or search for tweets within a specific topic. Users on Twitter can "mention" other users by incorporating their usernames, denoted by the '@' symbol, within their tweets. Retweeting is another key feature on Twitter, which allows users to share someone else's tweet with other users. Twitter has gained substantial significance in today's digitally connected world. It serves as a real-time information exchange platform, making it a go-to source for news updates. It plays a critical role during global events, emergencies, and crises, where instant communication is crucial. Moreover, Twitter has significantly influenced the way we communicate. Its limitation on the length of posts has cultivated a culture of concise and efficient communication. Hashtags have also transformed online conversations, making it easier to follow and contribute to specific topics of interest. Twitter's ability to connect individuals across the globe, foster open discussions, and serve as a real-time information source underscores its significance in today's digital society. Twitter is often used by individuals and organizations to share links that direct users to a variety of other platforms. This is done to distribute information, promote content or products, or simply facilitate the connection between various online activities.

Twitter stands out as a powerful tool for discovering invite links to public groups from a variety of online messaging platforms due to its real-time nature, rapid information dissemination, hashtag functionality, and trending topics. The retweet mechanism significantly extends the reach of these links, and Twitter's concise format facilitates communication. The platform's open and public nature encourages the spontaneous formation of communities, facilitating the seamless sharing of invite links. Overall, Twitter's dynamic features position it as a valuable focal point for both exploring and disseminating invite links across diverse online messaging platforms.

Facebook.

Launched in 2004, Facebook is a leading social networking service and platform. Facebook has a massive global user base, making it one of the most popular social media platforms worldwide. The platform's users span different continents, age groups, and demographic profiles, creating a diverse and dynamic online community. Facebook's primary features allow users to create a personal profile where they can share text posts, photos, multimedia, and URLs. Users can add other users as "friends," exchange messages, join interest-based groups, and participate in various online activities. There are a few key features for communication including the 'like' button to state agreement and interest in other posts, the ability to 'share' other users' posts, and to insert 'comment' in reply to other posts. One of Facebook's most significant characteristics is its News Feed, where users see updates from their friends and the pages they follow. Facebook is not just a social media platform; it is a significant player in today's digital and social landscape, affecting various facets of personal, commercial, and societal activities. In the realm of online messaging platforms, Facebook serves as a valuable source for discovering invite links to public groups from various platforms, primarily due to its extensive user base and widespread network connections.

Facebook's extensive user base, diverse communities, and global reach make it a valuable source for finding invite links to public groups from various online platforms. Facebook's groups, pages, and search features empower users to share invite links within specific interest areas, effectively broadening their networks beyond the limitations of traditional social media channels.

2.2 Related Work

Recent studies center on the dissemination of information in social media platforms such as TikTok, Twitter, and Facebook [32, 33]. They show the nature of online media that are able to disseminate information quickly, which makes this environment susceptible to the effects of misinformation, rumors, and fake news [34]. A rich body of previous work emphasize measuring and analyzing various aspects of social networks, as well as understanding emerging social networks and Web communities. Specifically, previous work focus on mainstream social networks like Twitter [35–37], YouTube [38, 39], Reddit [40–42], Flickr [43, 44], and Facebook [45, 46]. More recently, previous work focuses on analyzing and measuring emerging fringe social networks like 4chan [47, 48], an anonymous imageboard, Gab [49, 50], an alt-right Twitter clone, and Mastodon [51], a decentralized microblog. Furthermore, motivated by the overwhelmingly large number of social networks available, previous work analyzes multiple social networks and measuring the interplay between them [52–55]

The structure of groups and chats on online messaging platforms is similar to other mainstream social networks. They are characterized by a well-connected network with pathways facilitating message transmission among groups and users [56]. Melo et al. [57] 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 in this environment.

Furthermore, the proliferation of fake news and harmful content within online messaging platforms have drawn attention. Rumors and false stories shared on WhatsApp lead to lynchings and violent acts in India [58, 59]. Regarding the pandemic of COVID-19, an infodemic is 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 floods the chats of users around the globe [60]. Studies conducted in various locations, including Pakistan [61], Spain [62], Zimbabwe [63], Brazil [64], UK [65], and India [66], raise concerns about COVID-19 misinformation on online messaging platforms, highlighting its global impact. Misinformation campaigns on these messaging services are also recognized for their significant interference in the democratic election process, particularly in countries with large user bases on these platforms. This includes instances in Nigeria [67, 68], India [12, 69], Brazil [9, 56], and their role in events such as the Capitol riot in January 2021 [70]. Moreover, on Telegram, users create groups for conspiracy theories such as QAnon that mobilize users across international borders and reach a global audience [2, 71]. Prior research suggests that people are more likely to perceive and share fake news when presented in video format, rather than as audio or text, particularly if they have a limited understanding of the subject. The surge in 'deep fakes', which are expertly manipulated videos, amplifies this issue, capitalizing on the principle of 'seeing is believing'. This becomes particularly problematic on private online messaging platforms such as WhatsApp, where small community interactions escape the regulatory scrutiny of newsfeed algorithms, making the correction of misinformation challenging [72].

3

Demystifying the Messaging Platforms' Ecosystem

The online messaging platforms' ecosystem, as an information dissemination medium, has crucial implications for society and humanity at large. This ecosystem is also an invaluable data source for analyzing and understanding emerging socio-technical issues.

Overall, as a research community, we lack a holistic view of the online messaging platforms' ecosystem. Specifically, we do not clearly understand how online messaging platforms differ from one another, or how different are the characteristics of and activities within the groups found across different platforms. How these groups grow or evolve over time, whether they are ephemeral, and if they leak personally identifiable information (PII) remain largely unknown. As these public groups in online messaging platforms are increasingly being used by non-tech-savvy people or an uninformed population, the answers to these questions, especially those concerning privacy, are key to limiting their harm to society.

Prior work on the exploitation of this ecosystem focused on specific issues, e.g., the dissemination of false information [56, 58, 73], typically within a limited sample, e.g., in a small number of political groups [74, 75], and in a specific platform, e.g., WhatsApp [9, 76, 77]. User-created groups in online messaging platforms, however, are not limited to only politics; groups for virtually every conceivable topic plausibly exist. Furthermore, virtually all of these prior work focus on a specific platform, and ignore the opportunity to compare observations across different platforms to provide a holistic picture. Restricting the focus only on a specific, large platform limits the perspective and skews the insights: Studies indicate that small, potentially fringe platforms can exert a disproportionate influence on other mainstream platforms [53, 54].

First, we characterize the online messaging platforms' ecosystem through the lens of Twitter, a prominent social media platform. To this end, we discover public groups in these platforms via Twitter and analyze their characteristics. We focus specifically on answering the following research questions.

RQ1: How can we access public groups of online messaging platforms and their messages?

RQ2: What are the characteristics of the groups of online messaging platforms?

RQ3: Is there any leakage of Personally Identifiable Information (PII) in online messaging platforms?

To answer these questions, we first discover public groups in WhatsApp, Telegram, and Discord over a period of 38 days using Twitter APIs. We gather a set of about 350,000 group URLs and collect several meta attributes for each group (e.g., number of members in the group) once per day to understand how the groups change over time. We also selectively join a random sample of 616 public groups, and we gather all the messages posted in them: Overall, we collect a set of about 8 million messages posted by 800,000 users across the 616 groups. Using this large corpus of data, we shed light on the discovery of public groups on Twitter and also analyze their commonalities and differences. We shed light on the topics of conversation in these groups using topic modeling and compare the topics across the groups discovered in WhatsApp, Telegram, and Discord. We conduct temporal analyses to investigate the changes in the composition of and activity within the discovered groups over time. Finally, we look for potential PII exposures through these groups and discuss the privacy implications of such leaks for users.

Findings. Below, we summarize the key findings of this study.

- Twitter is a rich source for discovering WhatsApp, Telegram, and Discord group URLs. During our data collection period, we discover a substantial number of new groups: Per day we find, in the median, 1,000 WhatsApp groups, 2,000 Telegram groups, and 6,000 Discord groups.
- We analyze the content of the tweets, with the group URL(s), to characterize the differences in groups across the online messaging platforms: We find, for instance, a substantial number of groups in WhatsApp and Telegram that are used extensively for discussing crypto-currencies, in Telegram on the topics of sex and pornography, and in Discord on topics related to gaming and hentai (Japanese anime pornography).
- Group URLs across all of the platforms are ephemeral. We find that 27% of WhatsApp, 20% of Telegram, and 68% of Discord group URLs become inaccessible within 38 days.
- We discover PII leaks via WhatsApp, Telegram, and Discord groups. Specifically, we find the phone numbers of over 54,000 WhatsApp users (or *all* of the discovered WhatsApp users). On Telegram we find the phone numbers of a substantially fewer number of users–509 phone numbers corresponding to 0.68% of the discovered Telegram users. Discord, in contrast to the other two, does not expose the phone numbers of users but exposes the social media accounts linked to each user's Discord account. We observe that 30% of Discord users have at least one social media account linked to their Discord profile.

Chapter Organization. The rest of this chapter is organized as follows. Section 3.1 provides the background on WhatsApp, Telegram, and Discord and reviews prior work. Section 3.2 discusses our data-collection methodology and dataset. Section 3.3 presents how WhatsApp, Telegram, and Discord groups are shared on Twitter and analyzes the activity and evolution of the discovered groups. This section also contains our assessment of privacy implications for users stemming from the utilization of these online messaging platforms. Finally, we conclude in Section 3.4.

3.1 Background and Related Work

At this point, we review previous work related to analyzing and measuring online messaging platforms like WhatsApp, Telegram, and Discord.

WhatsApp. Previous measurement studies on WhatsApp mainly focus on acquiring data from public WhatsApp groups and analyzing their content to study emerging phenomena like the spread of false information. Rosenfeld et al. [78] perform surveys to characterize the behavior of 4 million messages from 100 WhatsApp users with the goal of inferring demographic information. Then, Garimella and Tyson [79] develop a set of tools that enable the large-scale collection of WhatsApp data from public groups, finding 2,500 groups and joining 200 in order to characterize WhatsApp users in India. Bursztyn and Birnbaum [76] also find 232 partisan WhatsApp groups through searches on other platforms for both right-wingers and left-wingers in the 2018 Brazilian presidential elections. Several WhatsApp studies focus on the spread of false information, in particular during electoral periods in Brazil [9, 56, 75, 77], India [80, 81], and Ghana [82]. Resende et al. [56] analyze how information spreads on WhatsApp with more than 350 public groups related to politics in Brazil, focusing on image-based misinformation, while Maros et al. [83] characterize the content of audio messages shared on WhatsApp. Melo et al. [74] develop a system, to assist fact checkers, that gather data from 1,100 groups from Brazil and India, and daily display the most popular content (i.e., messages, images, URLs, audio, and video). Melo et al. [73] also investigate the impact of message forwarding limits on the spread of messages in WhatsApp public groups, suggesting that the limit of 5 for forwarding is not sufficient to contain the spread of viral content on the platform. Finally, a recent study [80] releases a dataset of fact-checked images shared on WhatsApp during the Brazilian and Indian Elections.

Telegram. Previous work focuses on collecting data from Telegram and studying emerging research problems. Specifically, Baumgartner et al. [84] collect and make publicly available a large-scale dataset of 27,000 Telegram groups and 317 million messages. Anglano et al. [85] and Satrya et al. [86] investigate the artifacts generated by the Telegram application, while Abu-Salma et al. [87] perform a user study to understand user perceptions related to Telegram's security. A large body of work examines the use of Telegram in Iran. Specifically, Nikkah et al. [88] study the use of Telegram by Iranian immigrants with a focus on understanding how Telegram groups are moderated. Hashemi et al. [89] perform a large-scale analysis on 900,000 Iranian channels and 300,000 Iranian groups aiming to distinguish groups into the ones that are high-quality (e.g., business-related) and low-quality (e.g., dating groups). Asnafi et al. [90] analyze the use of the Telegram platform in Iranian libraries. Akbari et al. [91] investigate the ban of the Telegram platform by Russia and Iran after Telegram refused to provide access to encrypted data posted among users of the platform. Darghani et al. [92] collect data from 2,600 Telegram groups and channels and perform a structural analysis of the content posted within those groups/channels. Naseri et al. [93] focus on the spread of news on Telegram by collecting data from five official Telegram channels (i.e., Telegram channels that are used by news outlets). Finally, previous work focuses on studying how Telegram is exploited by terrorist organizations like ISIS [94–96]. Such organizations exploit the Telegram platform for their communication purposes, to spread propaganda, and possibly to recruit new members.

Discord. Finally, we review previous research on Discord. Hamrick et al. [97] study pump and dump schemes on the cryptocurrency market by analyzing data obtained from Discord. Lacher and Biehl [98] examine the use of Discord for teaching purposes. Jiang et al. [99] study the moderation challenges that exist on Discord, in particular on voice-based channels. Similarly, Kiene and Hill [100] focus on moderation on Discord and more specifically on the use of bots for moderating content posted on Discord servers.

Remarks. Overall, previous studies are dedicated to measuring the dynamics and discourse of specific topics in each of the online messaging platforms considered. Importantly, these previous studies show that all popular online messaging platforms have been exploited for some sort of underground activities and different forms of abuse in communication systems, from misinformation campaigns to revenge porn. Despite the undeniable importance of existing efforts, they do not attempt to provide a clear big-picture understanding of the dynamics of public groups on multiple platforms and do not attempt to characterize key differences between them. We fill this gap, by performing, to the best of our knowledge, the largest multi-platform analysis of online messaging platforms by collecting and providing an in-depth study of 351,000 groups from WhatsApp, Telegram, and Discord, shared on Twitter.

3.2 Methodology and Dataset

We measure the use of different online messaging platforms and identify key differences. To this end, we use Twitter—a widely used social media platform—to discover groups from WhatsApp, Telegram, and Discord, and characterize the composition of and activity within these groups. In the rest of the section, we use the terms "groups" and "channels" interchangeably, since the distinction does not affect our analyses or findings.

Our data collection methodology consists of three steps: (1) discovering public groups from WhatsApp, Telegram, and Discord via Twitter; (2) collecting group-specific metadata; and (3) joining the discovered WhatsApp, Telegram, and Discord groups and collecting data (e.g., group metadata and messages). Below, we elaborate on each step.

Table 3.1: Overview of the online messaging platforms dataset.

	Twitter			Online Messaging Platform		
	#Tweets	#Users	#Group URLs	#Groups	#Messages	#Users
WhatsApp	239,807	88,119	45,718	416	476,059	20,906
Telegram	1,224,540	398,816	78,105	100	3,148,826	688,343
Discord	779,685	340,702	227,712	100	4,630,184	52,463
Total	2,234,128	806,372	351,535	616	8,255,069	761,712

3.2.1 Discovering WhatsApp, Telegram, and Discord Groups

All of the three online messaging platforms support public groups, and the most common way to invite other users to a public group is to share the group URL (also referred to as the "invite" URL) with them. The group URLs of each platform follow one or more distinct patterns. On WhatsApp, for instance, group URLs have the pattern "chat.whatsapp.com/`<gID>`" with `gID` representing a unique identifier of the group, which is automatically generated by the WhatsApp messenger application when the group is created. We begin our data collection by first identifying the set of URL patterns for each platform. We review the documentation of each platform and manually inspect the URLs to compile a list of six patterns utilized across these platforms. These six patterns have the following prefixes or `host` values: `chat.whatsapp.com/`, `t.me/`, `telegram.me/`, `telegram.org/`, `discord.gg/`, and `discord.com/`

We search for the occurrences of the above URL patterns between April 8 and May 15, 2020, on Twitter, using two different approaches: (a) using Twitter's Search API [101] every hour, and (b) using Twitter's Streaming API [102]. The former retrieves all matching tweets (i.e., tweets containing the URL patterns) that were shared during the past seven days (i.e., from the time at which the query was issued), while the latter retrieves matching tweets in real-time, as they are posted on Twitter. We merge the tweets obtained via both APIs since a preliminary investigation revealed discrepancies between the tweets retrieved using the two APIs.

Using the above approach, we discover 351,535 group URLs (belonging to the three online messaging platforms) from 2,234,128 tweets posted by 806,372 Twitter users (refer to the left side of Table 3.1). Per this table, we discover a larger number of group URLs from Discord (227,000) than either Telegram (78,000) or WhatsApp (45,000). A large number of Discord and Telegram groups discovered despite these platforms being smaller (in terms of the number of users) than WhatsApp, suggests that these two platforms perhaps have greater channel diversity and public accessibility compared to WhatsApp; they both also have less strict limits on group sizes compared to WhatsApp. We discover the largest number of groups from Discord presumably owing to Discord group URLs automatically expiring after a day [103]; users, hence, are likely sharing a large number of unique group URLs compared to the other platforms.

Control Dataset. We compare the tweets dataset, where applicable, against a control dataset. The control dataset comprises a random sample of 1% of all 1,797,914 tweets posted between April 8 and May 15, 2020, and obtained via Twitter's 1% Streaming API. In this case, we use the Streaming API without limiting the results to a list of matching patterns or keywords and obtaining a 1% random sample of all tweets.

3.2.2 Collecting Group-Specific Metadata

Although we can join an online messaging platform's group given its group URL, we refrain from joining hundreds of thousands of groups for three practical reasons. First, there is a limit on the number of groups a user can join before getting banned from an online messaging platform. We empirically find that the limit for WhatsApp is between 250 and 300 groups per user, while on Discord it is up to 100 servers. Second, in the case of WhatsApp the above limit translates to a need for hundreds of phones and SIM cards to join all discovered groups, limiting the scale as well as the scope of the study. Third, we intend to minimize disruptions caused by joining hundreds of thousands of groups on any platform. We, hence, take a more pragmatic approach to obtain metadata from each group without joining every one of them. Below, we explain our approach.

WhatsApp. We use WhatsApp's Web client to obtain basic information about a WhatsApp group without joining it. Specifically, we automate the process of clicking on a WhatsApp group URL and opening the landing page for the group on a browser. We refrain from clicking the "Join" button on the landing page but scrape the page to gather several details: (1) title of the group; (2) size of the group (at the time of visiting the landing page); (3) country code of the phone number of the group's creator; and (4) phone number of the group's creator.

Telegram. Similar to the method for WhatsApp, we use Telegram's Web client to obtain basic information about Telegram groups without joining them. We implement a custom scraper that obtains and parses the web page for each group to gather several details: (1) title of the group; (2) size of the group and number of members online (at the time of visiting the group's web page); and (3) whether the "chat room" is a channel or a group.

Discord. For obtaining metadata about Discord groups we use the platform's REST API [104]. For each group, we collect the (1) title of the group, (2) number of members–both in total and online–in the group, and (3) group creator and group creation date. We follow the aforementioned techniques to gather metadata on each group on all of the three platforms every day from April 8 through May 15, 2020. We commence the metadata collection for each group from the date when we discovered it and repeat it every day unless the URL is revoked; landing pages of revoked URLs clearly indicate the revocation. We also track the status of each group (i.e., check if the group URL is alive or revoked) and the number of members in the group, every day starting from the discovery date.

3.2.3 Analyzing group composition and activity

For a subset of the discovered groups, we supplement the basic group metadata with details on the structure of and activity within the groups. To this end, we select a set of group URLs uniformly at random and join them using an account for each platform. Below, we describe how we obtain data from within the groups on every platform.

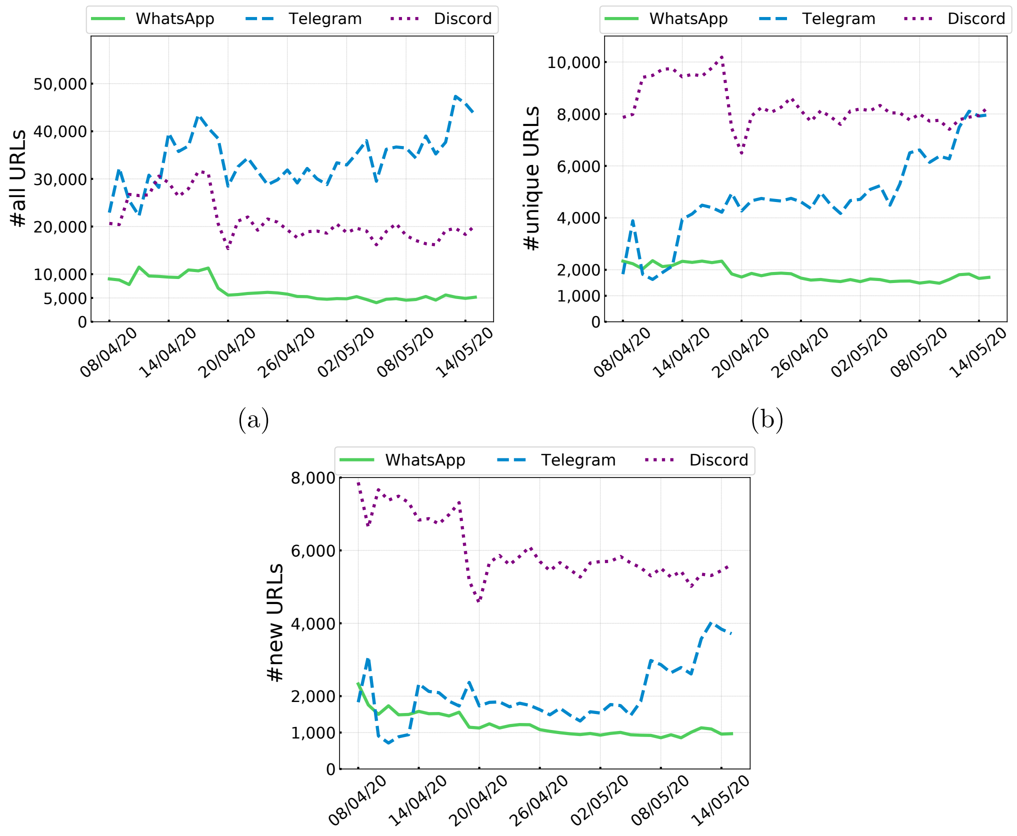
WhatsApp. WhatsApp does not provide an API to join groups or retrieve messages from within a group. As a consequence, we rely on WhatsApp's Web client to join the groups and collect data within these groups [105]. In total, we select and join 416 random public groups. Joining a group provides us with several pieces of information that are otherwise inaccessible (i.e., inaccessible without joining the groups): (1) messages shared on the groups (WhatsApp gives access to messages shared on the group, after our joining date); (2) phone numbers of the members of the group (For privacy reasons, we store only a hash of the phone numbers); and (3) creation date of the group.

Telegram. Telegram, unlike WhatsApp, provides a public API for gathering data on groups [106]. We select 100 URLs uniformly at random and join them with a new account. For each group, we collect (1) messages shared on the groups (since the group was created), (2) the creation date of the group, and (3) user profiles for the members of the group. A group administrator may opt to hide the member list from the group, and we obtain, hence, the member list only in 24 groups (out of the 100) where administrators did not exercise this option.

Discord. Although Discord provides an API for developing bots to help manage groups (e.g., run commands or send automatic messages), such a bot application has limited access to public groups. A bot is disallowed, for instance, from joining a group, albeit the group's administrator can add the bot to the group. To address the issue, we automate the process of opening the landing page of a group and joining it using a dedicated user account. We join 100 random servers (the maximum number of servers that a single user can join) and, using an application created with the user account, obtain the following data through the Discord API [107]: (1) messages on all groups on the joined servers (since the data from each group was created) and (2) user profiles for the group members.

3.3 Results

In this section, we explore the discovery of public groups on Twitter. We analyze the text of tweets containing group URLs to gain an understanding of the group's content. Subsequently, we investigate the activity within the groups and their evolution over time. Finally, we assess the privacy implications for users resulting from the use of the online messaging platforms.



(a)

(b)

(c)

Figure 3.1: Discovered group URLs on Twitter.

3.3.1 Discovering Public Groups

We aim to analyze the tweets that contain group URLs from WhatsApp, Telegram, and Discord to understand the interplay between Twitter and these online messaging platforms. The tweets also provide some context on the shared groups. We analyze how public groups are shared over time on Twitter, the prevalence in use of various Twitter features (i.e., hashtags, mentions, and retweets) when sharing groups, and the main themes of these groups by performing topic modeling on the content of the tweets.

Group Sharing Dynamics.

We begin our analyses with the number of group URLs discovered on Twitter for the three platforms (see Fig. 3.1). We report three different metrics: (1) all group URLs discovered on Twitter; (2) the number of unique group URLs per day; and (3) the number of *new* group URLs per day (i.e., excluding group URLs already observed on previous days).

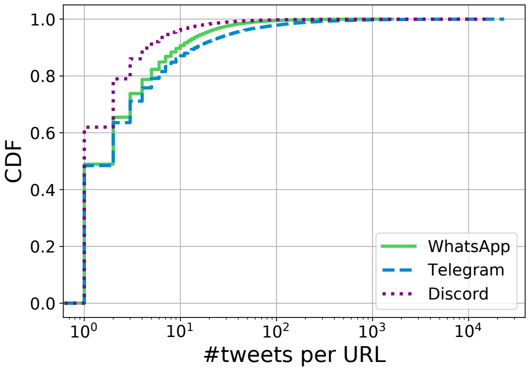


Figure 3.2: The number of tweets per group URL on Twitter.

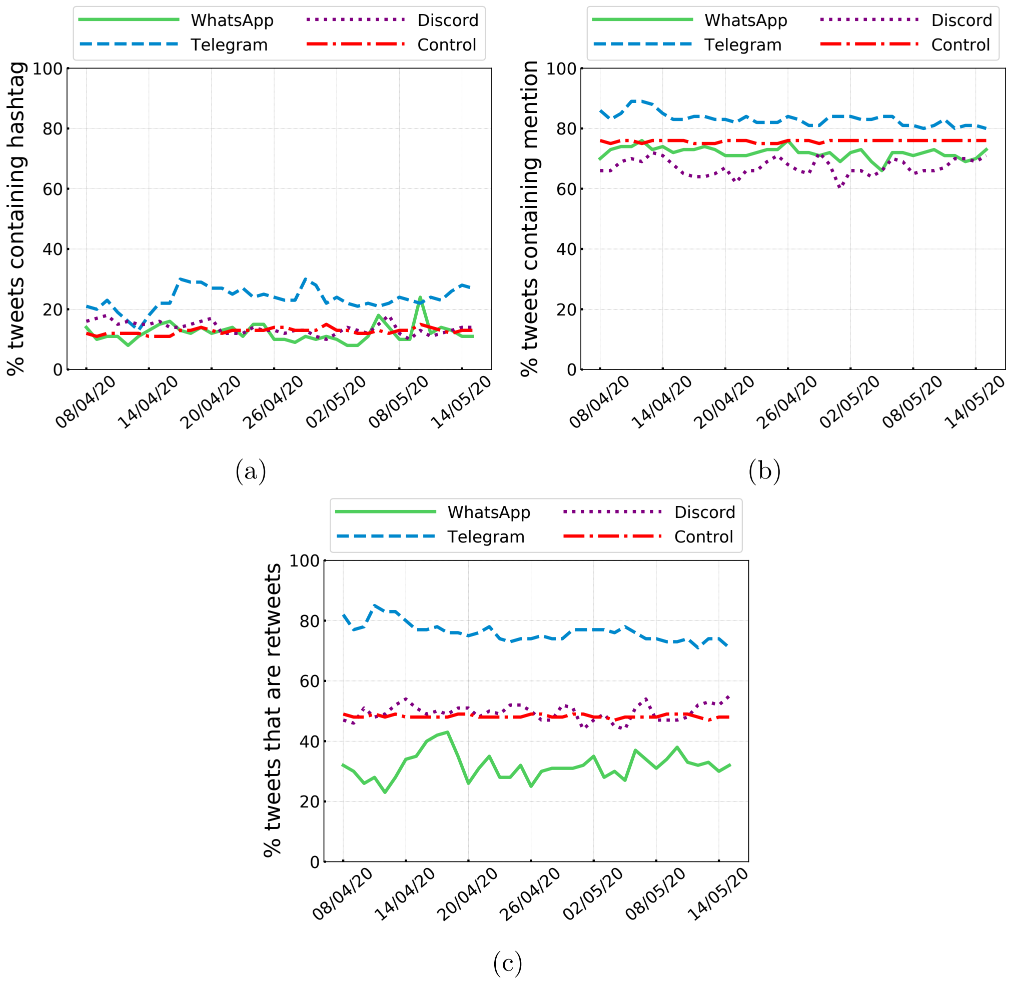
WhatsApp appears, per Fig. 3.1, to be the most "private" online messaging platform: We discover fewer group URLs belonging to WhatsApp than that of Telegram and Discord, despite WhatsApp being a much larger and widely used online messaging platform. This observation perhaps suggests that WhatsApp users are less willing to share public group URLs on Twitter compared to Telegram and Discord users. Second, we discover the largest number of group URLs for Telegram (Fig. 3.1a), with 33,864 group URLs, in the median, per day, followed by Discord with 19,970 URLs.

When it comes to discovering unique group URLs on a daily basis (Fig. 3.1b), Discord surpasses Telegram (8,090 URLs vs 4,661 URLS, in the median). These findings indicate that Telegram groups are shared more times than that Discord and WhatsApp, within the same day (see Fig. 3.1a and Fig. 3.1b). The number of newly discovered group URLs per day (Fig. 3.1c) indicates that Telegram group URLs are likely to also be shared across several days. Overall, we find that Twitter is a rich source for discovering public groups of online messaging platforms.

Fig. 3.2 sheds more light on the number of times that each group URL is shared on Twitter. Approximately half of the group URLs from WhatsApp and Telegram are shared only once, compared to 62% of the URLs in Discord. Overall, on average, each WhatsApp and Telegram group URL is shared in more tweets compared to Discord. We observe a few Telegram groups (14 in total) that were shared on a large number of (i.e., more than 10,000) tweets. We find, via manual examination, that 11 groups focus on pornography 2 on cryptocurrencies, and one is a general discussion group.

Content Analysis.

For characterizing the tweets, we use three widely used Twitter mechanisms for content broadcasting and discovery: hashtag, mention, and retweet. A hashtag is a keyword associated with a tweet that conveys a topic theme or event of interest. Users can discover tweets on a given topic by searching for a relevant hashtag, and it allows Twitter to group tweets by hashtags and broadcast them to interested users. Mentions support a "controlled" broadcast. A mention allows a user to refer to one or more users in the tweet who will be notified when the tweet is shared, increasing the likelihood of those users to read and respond. In the same vein, a retweet is a



(a)

(b)

(c)

Figure 3.3: Hashtags/Mentions/Retweets in tweets containg group URLs.

broadcast of a specific tweet to all the followers of the "retweeting" user. Next, we analyze the prevalence of the use of these mechanisms in tweets that include group URLs from the three online messaging platforms.

Per Fig. 3.3a, only a small percentage of tweets across all three platforms include hashtags. Specifically, tweets containing Telegram group URLs are more likely to include hashtags (24% of these tweets include hashtags), while for the other two platforms as well as the control dataset we observe a lower percentage of tweets with hashtags (13% for WhatsApp, 14% for Discord, and 13% for control). The lack of hashtags could perhaps be due to users intentionally restricting the tweets' visibility to their followers. Given the relatively low limit on the size of WhatsApp groups, for instance, users might intend to share a WhatsApp group only with few other people; tweets with WhatsApp groups, per Fig. 3.3a, contain fewer hashtags than those with Telegram and Discord groups. We also find that only a small percentage of tweets include more than one hashtag: 4% for WhatsApp, 10% for Telegram, 7% for Discord, and 5% for the control dataset.

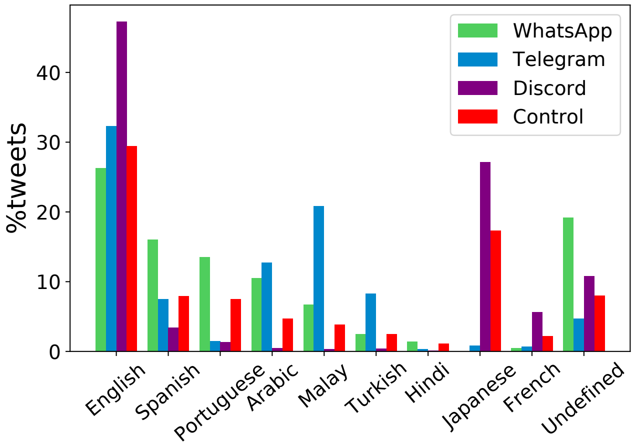


Figure 3.4: Language distribution among tweets containg group URLs.

When analyzing tweets with mentions (see Fig. 3.3b), we note a higher percentage of tweets with mentions among those containing Telegram group URLs (84%) compared to the control dataset and other platforms (73%, 68%, 76% for WhatsApp, Discord, and control, respectively), likely because Twitter users with Telegram groups are selective about the people they invite to their groups, despite the fact that they are sharing tweets in a public space. We also investigate the number of mentions per tweet finding that in general, only a small percentage of tweets include more than one mention; 20% for WhatsApp, 14% for Telegram, 15% for Discord, and 12% for the control dataset.

Lastly, our analysis of retweets (see Fig. 3.3c), shows a smaller percentage of retweets for WhatsApp (33%) than that for Telegram (76%) and Discord (50%). Twitter users are more likely to retweet posts containing group URLs from Telegram and Discord as these platforms are probably considered more public than WhatsApp.

Topic Modeling.

Next, we focus on understanding the context around the sharing of group URLs by analyzing the text of the tweets. First, we analyze the various languages that exist in our dataset. To this end, we use the language field as returned by Twitter's APIs, and observe that English is the most popular language with. Fig. 3.4 shows the percentage of tweets in each language across the three online messaging platforms: 26%, 35%, 47% for WhatsApp, Telegram, and Discord, respectively. For WhatsApp, the second and third most popular languages are Spanish (16%) and Portuguese (14%), while for Telegram it's Arabic (15%) and Turkish (8%). Interestingly, we find Discord users have a substantial number of Japanese users, as 27% of all tweets with Discord group URLs are in Japanese. These results shed light on the demographics of the users sharing the public groups and using the groups on the online messaging platforms.

Table 3.2: Topics in tweets containing WhatsApp URLs.

#	Label	Topic terms
1	**Forex Training (6%)**	learn, free, forex, training, join, trading, text, mini, class, animation
2	**Earn money from home (8%)**	home, earn, don, just, money, using, can, start, stay, google
3	**Instagram Followers Boosting (9%)**	join, followers, instagram, gain, want, money, online, group, learn, make
4	**Cryptocurrencies (7%)**	bitcoin, ethereum, crypto, currency, ads, year, like, line, people, new
5	**Earn money from home (13%)**	make, can, money, know, daily, home, earn, forex, cash, market
6	**Cryptocurrencies (5%)**	learn, cryptocurrency, make, join, days, period, another, want, day, accumulate
7	**WhatsApp group advertisement (30%)**	join, group, whatsapp, link, follow, click, please, chat, open, twitter
8	**Making money (9%)**	get, never, time, actually, income, chat, best, taking, account, full
9	**Nigeria-Related (6%)**	will, new, retweet, capital, people, now, interested, writing, nigerian, online
10	**Cryptocurrency courses (6%)**	business, ethereum, free, smart, skills, eth, million, join, training, webinar

Table 3.3: Topics in tweets containing Telegram URLs.

#	Label	Topic terms
1	**Cryptocurrencies (9%)**	bitcoin, join, sats, get, winners, sex, hours, chat, nice, come
2	**Cryptocurrencies (9%)**	usdt, giveaways, oin, winners, ollow, enter, btc, trc, trx, hours
3	**Social Network Activity (11%)**	follow, like, retweet, giveaway, tag, join, win, twitter, friends, friend
4	**Ask Me Anything/Quiz (8%)**	ama, may, will, utc, quiz, someone, wallet, don, ust, today
5	**Advertising Telegram groups (14%)**	free, join, just, telegram, money, day, channel, don, can baby
6	**Sex (13%)**	new, worth, user, brand, xpro, performer, smartphones, girls, boobs, price
7	**Giveaways (7%)**	giving, away, will, tmn, link, honor, full, butt, video, get
8	**Sex (10%)**	fuck, want, girl, click, show, trading, pussy, powerful, can, cum
9	**Advertising Telegram groups (11%)**	telegram, join, group, channel, now, below, link, get, available, opened
10	**Referral Marketing (8%)**	airdrop, open, https, tokens, wink, referral, token, earn, new, good

Table 3.4: Topics in tweets containing Discord URLs.

#	Label	Topic terms
1	**Gaming (7%)**	patreon, free, get, today, mystery, public, gaming, gamedev, indiegames, alongside
2	**Organizing online events (7%)**	will, may, hosting, week, one, time, tonight, don, night, last
3	**Gaming (5%)**	like, oin, alpha, deal, daily, art, lots, battle, raffle, nintendo
4	**Advertising Discord groups (33%)**	discord, join, server, link, can, visit, want, just, new, hey
5	**Pokemon (7%)**	united states, venonat, bite, quick, bug, full, fortnite, pikacku, confusion
6	**Advertising Discord groups (10%)**	giveaway, follow, retweet, friends, tag, join, discord, enter, fast, winners
7	**Tournaments (9%)**	good, live, launching, now, tournament, open, next, will, free, prize
8	**Giveaways (8%)**	giving, est, away, awp, will, saturday, friday, coins, many, competition
9	**Advertising Discord groups (4%)**	discord, join, make, sure, ends, chat, token, https, music, server
10	**Hentai (9%)**	join, discord, server, come, hentai, now, new, paradise, tenshi, official

To better grasp the context of the shared groups, we first extract all tweets posted in English and perform topic modeling using Latent Dirichlet Allocation (LDA) [108]. First, we focus on English, since it is the most popular language for tweets including group URLs for all three online messaging platforms. For each platform, we extract all the English tweets, remove stop words, and extract ten topics using the LDA method.

Tables 3.2- 3.4 report the topics extracted from the tweets sharing WhatsApp, Telegram, and Discord groups. For each topic, we manually assess the extracted topic terms and provide a high-level label and we also report the percentage of tweets that match each topic.

The extracted topics can be categorized into three types:

(1) *micro topics* that refer to topics that are specific to a single platform;

(2) *meso topics* that refer to topics that exist to more than a single platform; and

(3) *macro topics* that refer to topics that exist across all platforms.

For micro topics, we observe Forex Training (6% tweets), earning money from home (21%), and Instagram followers boosting (9%) topics on WhatsApp (see topics 1, 2, and 3, respectively, in Table 3.2), sex-related topics on Telegram (23%, see topics 6 and 8 in Table 3.3), and gaming (12%) and hentai-related (Japanese anime and manga pornography, 9% of all tweets) topics on Discord (see topics 1, 3, and 10 in Table 3.4). We find several meso topics related to cryptocurrencies on both WhatsApp (18%, see topics 3, 6, 10 in Table 3.2) and Telegram groups (18%, see topics 1 and 2 in Table 3.3), but not for Discord. Finally, for macro topics, we observe that across all platforms, there are topics where Twitter users try to persuade people to join their groups. For instance, see topic 7 for WhatsApp topics (30%), topics 5 and 9 for Telegram (25%), and topics 4, 6, and 9 for Discord (47%).

Interestingly, during our LDA analysis in English, we do not find any politics-related topics. We repeated our analysis with a larger number of topics (up to 50 topics per platform) and no politics-related topic emerged. This highlights that Twitter users are not sharing many politics-related groups from online messaging platforms in English, or if they do, they do not make it clear from the tweet's accompanying text.

Finally, we repeat the same analysis for other popular languages like Spanish and Portuguese but omit the results due to space constraints. We find some topics that do not emerge in our English analysis mainly due to the COVID-19 pandemic (in Spanish for WhatsApp and Telegram) and politics-related groups (in Spanish for Telegram and in Portuguese for WhatsApp).

Overall, our LDA analysis allows us to obtain insights into the content of the discovered online messaging platforms' groups by analyzing the text in the tweets sharing the group URLs. The extracted topics indicate that there are some similarities across the use of online messaging platforms, while at the same time, there are some topics where users prefer specific online messaging platforms to discuss them.

Takeaways. Twitter is a rich data source for discovering groups from WhatsApp, Telegram, and Discord. Our analyses reveal that users prefer to avoid using hashtags and only mention a small number of users in their tweets when sharing content about WhatsApp, Telegram, and Discord groups. Also, by performing topic modeling in the tweets, we find differences in the groups that are shared on Twitter from WhatsApp, Telegram, and Discord. Specifically, WhatsApp and Telegram are used for cryptocurrency discussions, Telegram for disseminating pornographic content, and Discord mainly for gaming, giveaways, tournaments, and hentai.

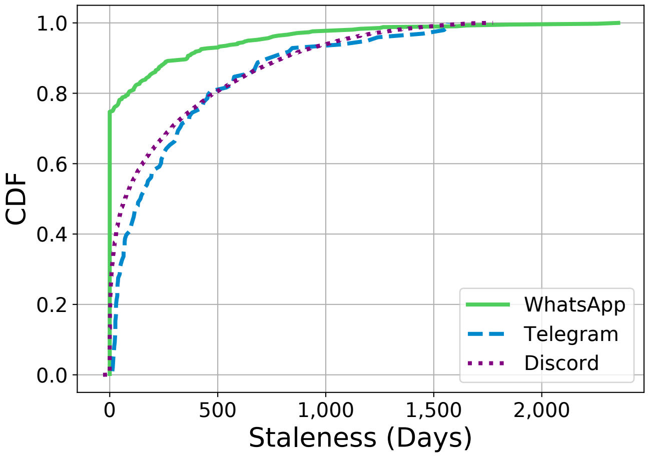


Figure 3.5: Staleness of the groups discovered on Twitter.

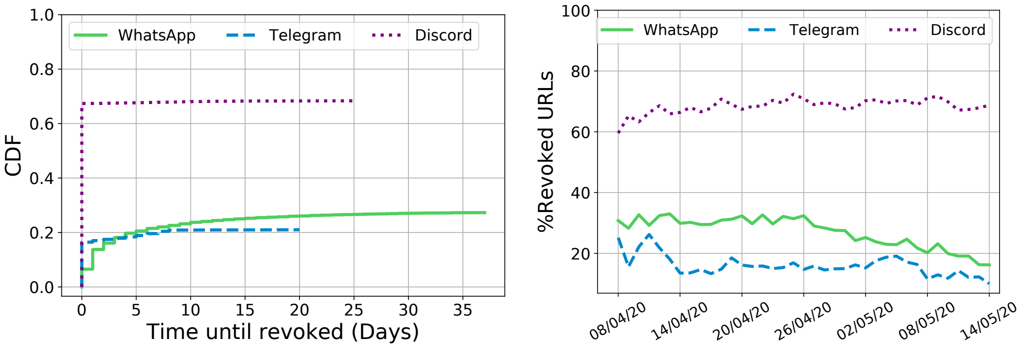
3.3.2 Activity and Evolution of Public Groups

In this section, we analyze the data obtained from the WhatsApp, Telegram, and Discord groups discovered from Twitter, with a focus on understanding the characteristics of those groups, how they change over time, and the volume of information disseminated within them.

Group Creators. For all groups from WhatsApp and Discord, the information about the creator of the group is available even without joining those groups. On the other hand, for Telegram, we are only able to obtain information about the creator for the 100 groups we join. We find that 34,078 different users created groups on WhatsApp, 49,753 users created groups on Discord, and 100 users created groups on Telegram. Also, we find that most of the users create a single group (100% for Telegram, 96% for Discord, and 93% for WhatsApp), with only a small percentage of users creating 2 groups or more (5% for WhatsApp and 4% for Discord).

Despite that, we find users that create a large number of groups (e.g., a single user created 61 groups on Discord and another 28 groups on WhatsApp). The number of users creating multiple groups on WhatsApp is larger compared to the other platforms and this is likely due to the imposed group limit (257 members). To overcome this limit, WhatsApp users are creating multiple groups with similar topics with the goal of reaching a larger audience.

Group Creation Dates. Next, we analyze the creation dates for the groups. For Discord, the creation date is available without the need to join the groups, yet for WhatsApp and Telegram we only obtain this data after joining the groups (416 for WhatsApp and 100 for Telegram). Based on the creation date, we can calculate how old the groups are at the time they are shared on Twitter. We define *staleness* as the time interval, in terms of days, between the creation date of a group and the date at which the group is shared on Twitter. In Fig. 3.5, we observe that most of the WhatsApp groups are created and shared on Twitter on the same day (76%),



(a) Accessible period of URLs

(b) Revoked URLs per day

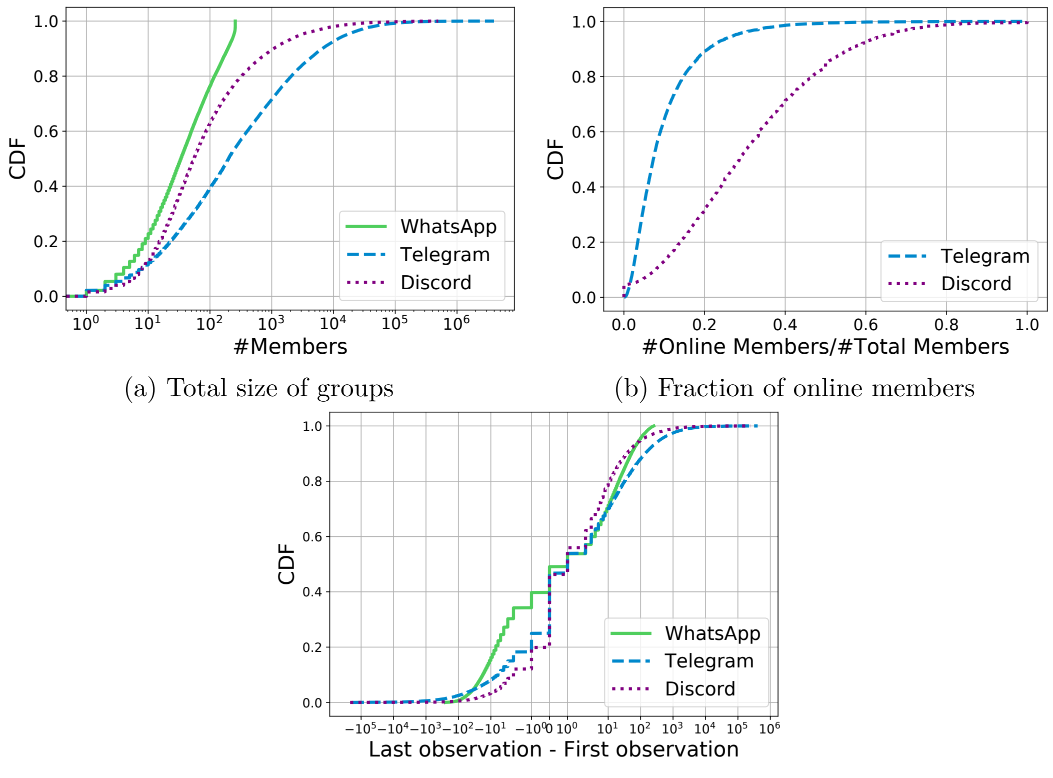
Figure 3.6: Accessibility of the groups discovered on Twitter.

while for Telegram and Discord less than 30% of the groups are shared during the groups' creation day. Also, only 10% of WhatsApp groups are older than one year compared to 29% and 26% of the groups for Telegram and Discord, respectively. The oldest group from our dataset, though, is from WhatsApp - a six-year-old group from Kuwait about the Real Madrid football team. Overall, these findings indicate that Twitter users tend to advertise older Telegram and Discord groups, compared to WhatsApp groups, and this is likely due to WhatsApp's imposed member limit (i.e., WhatsApp groups become full, hence not shared on Twitter to attract more members).

Group Countries. Since we store the country code of the creators' phone numbers for WhatsApp groups, we can investigate the group's country of origin. Note that for Discord, we do not have any information regarding phone numbers, while for Telegram we have phone numbers for only a small percentage of users, hence we limit this analysis to WhatsApp. A large number of WhatsApp groups are created by users from Brazil (BR) with 7,718 groups, followed by Nigeria (4,719), Indonesia (3,430), India (2,731), Saudi Arabia (2,574), Mexico (2,081), and Argentina (1,366). Although India is the country with the largest number of WhatsApp users (487 million, followed by Brazil with 118 million [109], it is only the 4th most popular country in our dataset. This is perhaps because our WhatsApp groups are only the ones shared on Twitter (Twitter has 22 million users in Brazil and 26 million in India [110]).

Group Revocation. On all platforms, a group URL can be revoked either manually, by an administrator, or automatically when all members leave the group or if the group URL expires (e.g., on Discord). Once revoked, no new users can use the group URL to join the concerned group and the landing page is devoid of any details except for the revocation notice. We monitor those URLs for their status and the number of their members, every day to analyze the behavior of the groups over time.

Although we cannot precisely determine whether revocation was manual or automatic, the lifetime of a group–defined as the time from discovery on Twitter until it is revoked–impacts our approach of characterizing groups based on the metadata from the landing page of its group URL. Fig. 3.6a shows the accessibility time (in days) for the revoked URLs, while Fig. 3.6b shows the percentage of revoked group



(a) Total size of groups

(b) Fraction of online members

(c) Size difference over time

Figure 3.7: The size of the groups discovered on Twitter.

URLs per day. We find that 27% of the URLs for WhatsApp groups, 20% of the Telegram group URLs, and 68% of the Discord group URLs are revoked at some time. This shows that Discord has many more revoked URLs, probably because, by default, group URLs auto-expire after a day, while a group URL from Telegram and WhatsApp lasts until the user manually revokes it or deletes the group. Therefore, Discord groups are less accessible through group URLs while the URLs we find for Telegram and WhatsApp are more likely to be accessible.

Looking at the lifetime, the time period a URL is accessible, we can observe that for many of the revoked URLs, the revocation is done before our first observation (6% of all groups for WhatsApp, 16% for Telegram, and 67% for Discord). This indicates that some groups have a very limited accessibility period, indicating the ephemeral nature of the online messaging platforms' groups. In future research focusing on collecting and analyzing datasets from online messaging platforms, it is important to consider the ephemeral nature of the groups within these platforms.

Group Members. Since users share group URLs on Twitter to entice others to join, the size of a group over time can hint at their activity and the reasons behind its revocation. To this end, we gather the number of members in each group, for each day that is accessible. We compare the distribution of the total amount of members for each platform in Fig. 3.7a. Overall, WhatsApp has much fewer members compared to the other two, because of the group size limit of 257 members. It is also worth noting that only a small percentage of WhatsApp groups (5%) reach the limit of the size. Also, we observe that Discord has fewer members than Telegram, as around 60% of Discord groups have less than 100 members while only 40% of Telegram groups have the same amount. For Telegram and Discord, we also have information about how many users within the group are actively online (provided by the platform itself via the Web client). We use this information, from our first observation, for each group to analyze the proportion of online members.

Fig. 3.7b shows that even though Telegram has more members in total, they are online in less proportion compared to Discord. We observe that around 15% of the groups on Discord have more than half of their members online, while on Telegram only a few groups have such activity. These results are likely due to the fact that Discord is a more computer/desktop-oriented platform, while Telegram is frequently used from mobile devices, hence Discord users are more likely to be online compared to Telegram users.[1]

Finally, we investigate the growth of the groups over time; Fig. 3.7c shows the distribution of the growth of the groups, which is the difference of group sizes observed on the first and the last day (i.e., prior to revocation) of observation. We can clearly observe the impact of the limit sizes for each platform in the distribution of the growth of the groups. Discord and Telegram have groups that change in more than 100,000 members during our analysis period: e.g., a Discord group for fans of the new Nintendo game “Animal Crossing” launched in March 2020, and a Telegram channel that shares movies. We can also note that there are more groups increasing in size than decreasing (51% for WhatsApp, 53% for Telegram, and 54% for Discord). This likely indicates that sharing the group URLs on Twitter helps the groups to aggregate more users. Still, some groups decreased in size (38% for WhatsApp, 24% Telegram, and 19% Discord), perhaps an indication of declining interest among the members of some groups over time.

Group Messages. Next, we analyze the collected messages from all of the joined groups. Overall, we gather 476,000 messages from WhatsApp, 3 million messages from Telegram, and 4,6 million messages from Discord. First, we compare the types of messages in each platform, as all platforms allow users to send text, images, videos, audio, stickers, and documents. Fig. 3.8 reports the percentage of the messages in each type. Unsurprisingly, text is the most shared type with 78%, 85%, and 96% of all messages on WhatsApp, Telegram, and Discord, respectively. Also, it is worth noting that WhatsApp is the platform with the largest variety of multimedia with more than 20% of multimedia messages (images, videos, audios, and stickers).[2] In particular, stickers, which are a specific format of images, represent 10% of all the collected WhatsApp messages. They are very common on WhatsApp and there are even groups dedicated to sharing exclusively stickers between users.

[1] Note that a Discord user is shown online even if the Discord Web/desktop client is running in the background.

[2] Note that our analysis only includes audio/video that is shared as messages (i.e., audio/video clips) and it does not consider audio/video calls within groups.

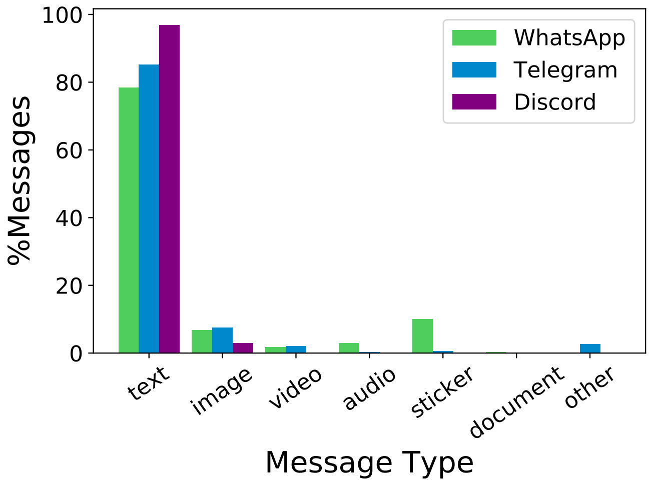


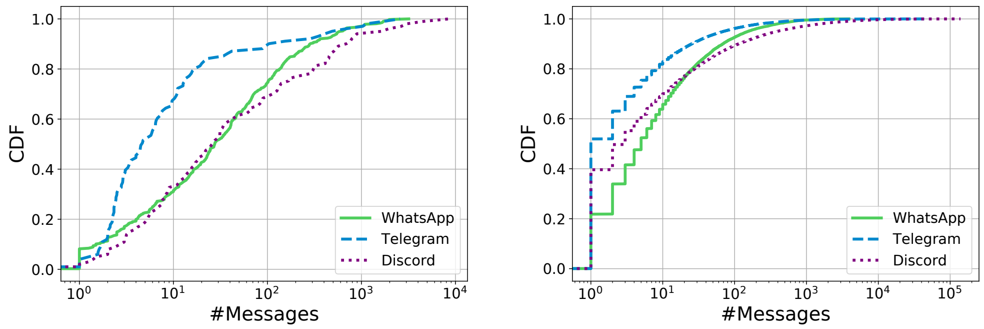
Figure 3.8: Messages types among the groups discovered on Twitter.

Note that Telegram also has a small portion of "other" types of messages including service messages (e.g., users joining/leaving groups, and editing group information). We also look into the volume of messages shared in each group and the number of messages per user. Fig. 3.9a shows the number of messages shared per day in each group for all the platforms. We report the number of messages per day since for WhatsApp we can only obtain messages shared after we joined the group, while for Telegram and Discord, we obtain messages since the group's creation date. We observe that Telegram groups are less active compared to WhatsApp and Discord. Specifically, approximately 60% of the groups have more than 10 messages a day, while just 25% of the Telegram groups have such activity. For all platforms, we can observe some groups with more than 2,000 messages per day.

The collected messages shared by 12,000 distinct users on WhatsApp, 100,000 users on Telegram, and 35,000 users on Discord. This represents, respectively, 59%, 15%, and 66% of the total number of members in the joined groups (see Table 3.1). Although we can not affirm that this represents the percentage of members sharing messages, as total size changes over time, these numbers give us a hint of the portion of active members in each platform. Discord has a higher number of active members. On the other hand, on Telegram, just a small portion of the total members share messages, probably because of channels, which allow only a small number of users to share messages (i.e., one creator and a few administrators).

Finally, we analyze the volume of messages shared per active member in Fig. 3.9b. We observe that most members share only a few messages, while some share a large volume of messages. In particular, 66% of them share up to 10 messages on WhatsApp, 70% on Discord, and 83% on Telegram.

When looking at the volume of the messages shared by the top 1% of the members (in terms of the number of messages they shared), they are responsible for 31% of all messages collected from WhatsApp, 60% of all Telegram messages, and 63% of all



(a) Mean number of messages of groups per day

(b) Number of messages per user

Figure 3.9: The number of messages shared in the groups discovered on Twitter.

Discord messages. This indicates that Telegram and Discord have a larger percentage of very active users that share a very large number of messages across groups.

Takeaways.

We show that the groups shared on Twitter are mostly “fresh”: they are shared on Twitter soon after they are created, yet a few groups are still being shared even though they were created more than a year ago. We discover that most Discord group URLs expire during the first days after being shared on Twitter, while WhatsApp group URLs last longer. Also, Telegram group URLs are less likely to get revoked.

We observe that the difference in the group size limit between the three platforms indeed impacts the size of the groups since Telegram and Discord have larger groups of up to 4 orders of magnitude compared to WhatsApp. Regarding those members, we can also note that Discord members are more active than Telegram in terms of the number of active members. The selection bias and ephemeral nature of group URLs, discovered on Twitter, have implications for studies that use such URLs.

3.3.3 Privacy Implications

In this section, we analyze the users’ privacy implications from using WhatsApp, Telegram, and Discord. When dealing with online messaging platforms, a common concern is about privacy and exposure of sensitive personally identifiable information (PII). In particular, for platforms where users are engaged in direct and closed conversations in a private and secure manner, it is important to analyze the potential PII that can be exposed by the platform.

Usually, users join these groups while being fully agnostic that various aspects of their private information are exposed by either the platforms’ interfaces or their APIs. This raises some legitimate concerns with regard to what kind of PII is exposed by each platform, and how critical and prevalent is the exposure of PII on WhatsApp, Telegram, and Discord. To this end, we collect all user-related information from each

Table 3.5: Statistics on exposed users' sensitive PII.

	WhatsApp	Telegram	Discord
Users observed	54,984	74,479 members	25,701 members
Users' Phone Numbers	54,984 (100%)	509 (0.68%)	-
Users' Social Networks	-	-	7,708 (30%)

platform and analyze them to understand the underlying privacy implications of the use of these platforms.

Each of the online messaging platforms has its own peculiarities and it requires a different approach to collect user information. On Telegram and Discord, we are able to collect user information for users who participate in groups that we also are members of. This also applies to WhatsApp, however, there is an important difference as WhatsApp exposes the phone number of group creators even before joining WhatsApp groups. To collect data related to users, for Telegram and Discord, we used the available APIs to get user information for groups we joined, while for WhatsApp we scraped the information from all discovered groups.

Table 3.5 reports the number of users whose PII is exposed for each platform. Looking at the total number of users from which we collected data, we find 21,000 WhatsApp users within the groups we joined, and 34,000 unique users that are the creators of the rest of the groups that are accessible, totaling 55,000 users. For Telegram, we collect information from 74,000 users, while for Discord we find 26,000 users. Note that for Telegram and Discord, the number of users is smaller than the total of users for groups we joined, representing 11% and 49% for Telegram and Discord, respectively. This is because, on Telegram, administrators are able to restrict access to the member list, thus users can not see who are the members of the group. For Discord, the API blocks the bot to join groups by themselves (they need to be added by an administrator) and obtain the list of members. Due to these constraints, we collect user information for users who posted at least one message within the groups we joined. Both Telegram and Discord offer official APIs that enable group management and task automation. These APIs also provide the ability to distinguish between bots and human users within the groups. In our study, we detected several bots among the users; specifically, 53 on Telegram and 102 on Discord. Although these bots constituted less than 1% of the total user population, their presence is potentially significant given the scope of our observation. We analyzed precisely one hundred groups on each platform, suggesting a potentially high density of bots within these groups.

Our data collection and analysis highlight the exposure of PII information in each platform. Alarmingly, on WhatsApp, we are able to obtain the phone numbers of *all* users that we discovered during our data collection, a total of 54,000 phone numbers. On the other hand, on Telegram we are able to only obtain the phone numbers of 509 users, which corresponds to 0.68% of all the Telegram users that participated in the groups we joined. The relatively low percentage is because Telegram hides the phone number of the users by default. A phone number is only shown within the platform if the user explicitly opts in. Finally, for Discord, since phone numbers are not required

Table 3.6: User account exposure on Discord

Platform	#Users (%)
Twitch	5,256 (20.4%)
Steam	3,158 (12.2%)
Twitter	2,287 (8.9%)
Spotify	2,080 (8.0%)
YouTube	1,712 (6.6%)
Battlenet	1,338 (5.2%)
Xbox	956 (3.7%)
Reddit	785 (3.0%)
League of Legends	617 (2.4%)
Skype	169 (0.6%)
Facebook	139 (0.5%)

for registration, we find no evidence of phone number exposure. However, we find that Discord exposes accounts that users have on other platforms: we find 8,000 users (30%) for whom we are able to obtain at least one other account that they have on other platforms, namely, Twitch, Steam, Twitter, Spotify, YouTube, Battlenet, Xbox, Reddit, League of Legends, Skype, and Facebook. Table 3.6 reports the number of users whose users' accounts are exposed for each of the other linked platforms. We find that 20% of the Discord users have linked their Twitch account, a platform used for streaming, 12% linked their Steam account, a gaming platform, while almost 9% of the users linked their Twitter account. Finally, we find that only 0.5% of the Discord users linked their Facebook accounts.

Overall, these findings have important implications for users' privacy. The exposure of PII from all these online messaging platforms can be potentially exploited by malevolent actors that aim to target users. For instance, state-sponsored actors [111, 112] that have considerable resources and can perform a much larger data collection than our study, can create profiles for all those users and target them on the same or on other social media platforms. A potential attack vector is the creation of user profiles based on the topics of the groups they participate in and then their targeting on other social media platforms via posts or advertisements with the goal of manipulating them or changing their ideology. Additionally, our findings highlight the need to raise user awareness about the privacy implications of the use of online messaging platforms like WhatsApp, Telegram, and Discord.

Takeaways. The main takeaway points from our analysis in this section are: (1) WhatsApp not only displays the phone number of all members of the groups we joined but also reveals the phone number of the groups' creators to non-members of the group; (2) Telegram exposes the phone numbers of all users that opt-in to share their phone numbers (note that by default this is turned off). Our results show that this happens only to 0.68% of the collected users; and (3) Discord exposes accounts of the same user to other platforms like Steam, Spotify, Twitter, Facebook, YouTube, etc. Our analysis shows that Discord exposes at least one social media account for 30% of the Discord users we monitored.

3.4 Discussion

We performed a large-scale characterization of public groups from WhatsApp, Telegram, and Discord shared on Twitter, a popular micro-blogging platform. Over a period exceeding one month, we systematically searched for group URLs (or invite links) on Twitter, collecting a dataset containing 350,000 URLs to groups. Our findings underscore several critical considerations for researchers examining similar platforms. By adopting a multi-platform approach to the Web ecosystem, we expose valuable insights that would otherwise be difficult to find, particularly in the context of studying individual platforms like WhatsApp in isolation. Notably, existing research efforts [53, 54] focused on phenomena such as news and memes do not consider online messaging platforms. Furthermore, we meticulously monitored the discovered groups across the three platforms, gathering measurements once per day. These coarse-grained measurements enabled us to investigate changes in group characteristics over time, including metrics such as group size.

We tried to provide answer to our first research question, RQ1: How can we access public groups of online messaging platforms and their messages? We found that Twitter serves as a rich data source for discovering public chat groups across the three online messaging platforms. Throughout our data collection period, we consistently identified a substantial number of new groups. On a daily basis, we found, on average, 1,000 WhatsApp groups, 2,000 Telegram groups, and 6,000 Discord groups.

To address the second research question (RQ2) concerning the characteristics of the groups of online messaging platforms, we conducted various analyses. These included extracting topics from tweets containing group URLs to gain insights into group content. Additionally, we investigated group accessibility over time, sharing dynamics, and group activity. Our findings successfully answered the research question. There is a substantial number of groups in WhatsApp and Telegram dedicated to discussions on crypto-currencies, while Telegram exhibited significant activity in topics related to sex and pornography, and Discord focused on gaming and hentai. The groups are shared on Twitter soon after they are created, indicating their freshness. However, a substantial portion of groups became inaccessible during the study period, with 27% for WhatsApp, 20% for Telegram, and 68% for Discord. Discord groups had a higher likelihood of expiring shortly after being shared, whereas WhatsApp group URLs tended to last longer. Furthermore, Telegram and Discord boasted larger group sizes compared to WhatsApp, with differences spanning up to four orders of magnitude. Notably, Discord exhibited a higher number of active members, whereas only a small fraction of total members on Telegram engaged in message sharing.

Finally, our investigation aimed to uncover any leakage of Personally Identifiable Information in online messaging platforms (RQ3). We examined the exposure of sensitive PII, such as phone numbers for WhatsApp and Telegram users, and linked social media accounts for Discord users. Our analysis effectively addressed the research question. We identified a substantial number of over 54,000 phone numbers for WhatsApp users. Additionally, we found exposed phone numbers for a small percentage of Telegram users (less than 1%). While we did not locate phone numbers on Discord, we managed to collect at least one linked social network account for 30% of the users analyzed. These privacy implications are alarming since online messaging platforms are often used because of their perceived security in communication and privacy. Our results highlight the need to raise public awareness regarding these privacy implications and design guidelines on how online messaging platforms can adjust to better safeguard users' privacy.

4 On the Globalization of QAnon Through Telegram

In the previous chapter, we discussed three distinct online messaging platforms, conducting a comparative analysis to address our initial three research questions. By answering these questions, we enhanced our comprehension of the online messaging ecosystem. Given the consequential impact of this ecosystem on society and the prevalent dissemination of misinformation and conspiracy theories within these platforms, our objective is to gain deeper insights into the propagation of conspiracy theories within such online messaging platforms. In this chapter, we delve into the evolution of the QAnon conspiracy theory specifically within Telegram.

One conspiracy theory that attracts high engagement from people is QAnon, which is a conspiracy theory alleging that a secret group of people (i.e., a cabal consisting of Democratic politicians, government officials, and Hollywood actors) were running a global child sex trafficking ring and were plotting against former US President Donald Trump[113]. Between 2017 and 2021, the QAnon conspiracy theory attracted many new followers across the globe and essentially evolved into a cult. Worryingly, the followers of the QAnon conspiracy theory have begun making threats or participating in violent real-world incidents (e.g., Capitol attack in 2021 [114]), hence highlighting the impact that the conspiracy theory has on the real world [115].

Motivated by the negative impact that QAnon has in the real world, mainstream platforms like Facebook, Twitter, and YouTube, started moderating and removing QAnon-related content [116–119]. Then, QAnon supporters sought new online "homes" in less-moderated platforms and migrated to other platforms like Parler and Telegram [120]. Also, QAnon became a global phenomenon; the QAnon conspiracy theory has accumulated new followers worldwide, particularly in European countries like Germany and Spain [121]. Overall, it is crucial to understand how QAnon evolved and became a global phenomenon that has not yet been investigated on a large scale by any other work. To do this, we select Telegram as the source of our study for two primary reasons. First, anecdotal evidence suggests that QAnon followers migrated to Telegram after bans on other platforms [120]. This observation aligns with our findings of the prevalence of harmful content in Telegram groups during the initial phase of our study in Chapter 3. Second, Telegram is a rapidly growing platform with worldwide coverage [122], making it an ideal platform for studying QAnon across the globe. The study detailed in Chapter 3 further substantiates the platform's rapid expansion, encompassing multiple languages within the messages.

We aim to answer these research questions:

- **RQ4:** How do conspiracy theories evolve over time and across languages?
- **RQ5:** How toxic are the conspiracy theories discussions over time and across languages?
- **RQ6:** What topics and conspiratorial content are popular among fringe communities?

Hypotheses. To address these research questions, we formulate the following hypotheses and concentrate on testing them:

- **H1 - Activity:** We hypothesize that QAnon activity in Telegram increases in volume over time (due to moderation actions on other platforms) to a larger extent compared to groups focusing on other topics. Also, we hypothesize that there are substantial changes in the popularity of the used languages over time due to anecdotal evidence suggesting QAnon is popular in Europe [121].
- **H2 - Toxicity:** QAnon Content in Telegram is more toxic compared to the content on groups/channels focusing on other topics.
- **H3 - Topics:** We hypothesize that QAnon followers discuss various topics related to Politics, they are sharing false information, and that the popularity of these topics changes over time.

We argue that these three hypotheses are equally important and need to be studied together. **H1** allows us to understand how active the QAnon movement is on Telegram and especially how this activity has evolved. In **H2** and **H3** we focus on what content is shared in QAnon groups/channels, how these discussions differ over time, how the discourse differs from previous work or other platforms, and how toxic the content is; this is equally important as it allows us to understand what the topics of discussions are and whether the QAnon discourse is becoming more toxic, which is of paramount importance given previous participation of QAnon followers in real-world violent acts. For instance, if QAnon followers subscribe to an anti-vax ideology, they are likely to participate in real-world protests associated with the anti-vax movement. Such findings can help us better prepare for dealing with such protests and potentially mitigate real-world violence.

To test the above-mentioned hypotheses, we perform a large-scale data collection and analysis of QAnon-related groups/channels on Telegram. Overall, we collect 4.4M messages shared in 161 Telegram groups/channels between September 2019 and March 2021. Using Google's Perspective API [20], we investigate the toxicity of QAnon content on Telegram and assess whether the movement is becoming more toxic over time and whether there are substantial differences across languages. Additionally, we use a multilingual BERT-based topic modeling approach [123] to study the QAnon discourse across multiple countries/languages. Finally, to strengthen our topic modeling results and gain deeper insights into the specific narratives shared by QAnon followers on Telegram, we perform a small-scale qualitative analysis based on thematic analysis [124].

Main findings. Our study provides some key findings:

- We find that QAnon activity in our dataset increased substantially during 2021 with an increase of almost 5x in terms of the number of messages and senders, while our baseline dataset has an increase of only 2x. Furthermore, by comparing content across languages, we find that German QAnon content surpassed English (on average 55% for German and 28% for English) in popularity after June 2020. Our findings support our first hypothesis **(H1)**.
- By analyzing the toxicity of QAnon-related messages in our dataset, we find that content shared in Portuguese and German is more toxic compared to English (9% of the Portuguese messages and 3% of the German messages are toxic, while for English we only have 1% of QAnon messages being toxic). At the same time, we find that QAnon content posted in English and Portuguese is more toxic compared to our baseline dataset (3.6x and 1.2x, respectively). Our results partly support our second hypothesis (**H2**), since for German we find that the baseline had 1.15x more toxic messages compared to our QAnon dataset.
- Our discourse analysis highlights that QAnon has evolved into discussing various topics of interest within far-right movements across the globe. We find several topics of discussion like world politics, conspiracy theories, COVID-19, and the anti-vaccination movement (**H3**).

Chapter Organization. This chapter is organized as follows: First, we start by presenting background and related work (Section 4.1). Next, we describe our methodology to collect our dataset (Section 4.2). Then, we present our analysis for investigating our three hypotheses related to the activity, toxicity, and topics posted by QAnon supporters (Section 4.3). We conclude in Section 4.4.

4.1 Background and Related Work

QAnon is a conspiracy theory alleging that a secret group of people (i.e., a cabal consisting of Democratic politicians, government officials, and Hollywood actors) were running a global child sex trafficking ring and were plotting against former US President Donald Trump[113]. The conspiracy theory emerged in October 2017 with a post on 4chan by a user named "Q," who claimed that he was an American government official with classified information about plots against then-President Donald Trump. "Q" continued disseminating cryptic messages about the QAnon conspiracy theory (called "Q drops") mainly on 8chan. The QAnon conspiracy theory has amassed a following in fringe Web communities like 4chan/8chan and mainstream ones like Facebook [125] and Twitter, especially after then-president Donald Trump retweeted QAnon-related content [126]. QAnon followers use their motto "Where We Go One, We Go All" (or simply wwg1wga) to tag content related to QAnon.

Over the past years, followers of the QAnon conspiracy theory have made violent threats or been linked to several incidents of real-world violence [115], with the Federal Bureau of Investigation (FBI) labeling it as a potential domestic terrorist threat [127]. In particular, on January 6th, 2021, supporters of the QAnon conspiracy theory attacked the US capitol in an attempt to overturn Donald Trump's defeat in the

2020 US elections by disrupting the Congress that was in the process of formalizing Joe Biden's victory [114]. Due to these threats and violent incidents, mainstream platforms like Facebook [118], Twitter [116], Reddit [119], and YouTube [117] started monitoring and removing QAnon-related groups, subreddits, and users. Naturally, following these content moderation interventions, supporters of the QAnon conspiracy theory flocked to other fringe Web communities with lax moderation, like Parler [128] and Gab [49, 50], or online messaging platforms like Telegram [120].

Even though the idea of the QAnon conspiracy theory is US-centric, QAnon became a global phenomenon, in particular among people with far-right ideology. In 2020, the QAnon theory spread to Europe [121]. The conspiracy theory is nowadays shared among people from Spain, Italy, the United Kingdom, and Germany, one of the most popular "representatives" in Europe [129].

Previous work investigates several aspects of the QAnon conspiracy theory. Papasavva et al. [130] analyze content toxicity and narratives in a QAnon community on Voat, finding that discussions in popular communities on Voat are more toxic than in QAnon communities. Aliapoulios et al. [128] provide a dataset of 183 million Parler posts, and they highlight that QAnon is one of the dominant topics on Parler. Miller [131] investigates a sample of QAnon-related comments on YouTube, highlighting the international nature of the movement. Garry et al. [132] explore QAnon supporters' behavior in spreading disinformation on Gab and Telegram, finding that the dissemination of disinformation is one of the main reasons for the growth of QAnon conspiracy. Hannah et al. [133] also investigate the reasons for the growth of QAnon, finding that sharing and discussing Q drops is one of the main reasons. Chandler [134] investigates how QAnon followers are influenced by Q drops, finding that Q drops focus on the perceived allies or enemies of QAnon. Planck [135] compares the QAnon community's rhetoric with a mainstream conservative community on Twitter, finding that tweets posted by QAnon supporters are more violent. Aliapoulios et al. [136] investigate a dataset of 4,900 canonical Q drops from six aggregation sites, finding inconsistencies among the drops and demonstrating that the drops have multiple authors. Ferrara et al. [137] investigate 240 million election-related tweets finding that 13% of users spreading political conspiracies (including QAnon) are bots. Sipka et al. [138] compare the language and narratives of QAnon-related content on Parler, Gab, and Twitter on a dataset of about 100k posts with the #QAnon hashtag and they find a prevalence of anti-social language on Parler, while Gab has the most conspiratorial and toxic content. Phadke et al. [139] characterize 2,000 posts from 4chan and 8chan and 1.2 million comments from 12 subreddits to understand the social imaginary within QAnon online communities and identify how their members express their belief and dissonance towards the conspiracy. Engel et al. [140] collect over 12 million posts from early QAnon users on Reddit and characterized how users engage in the QAnon conspiracy, showing they were dedicated and committed to the movement even after a massive ban of the QAnon from Reddit. Pasquetto et al. [141] examine the disinformation infrastructure of QAnon built on Italian digital media platforms by a digital ethnography over eleven months of QAnon activity on Facebook, Twitter, and Telegram communities. They observe a top-down design in Qanon structure online in which decisions are made and imposed on the community while the followers are expected to participate and share but they are not allowed to directly contribute to how information is organized or curated.

Due to its privacy policy and encrypted nature (i.e., "all data is stored heavily encrypted"), Telegram attracted the interest of dangerous organizations like terrorists [28] and far-right groups [27]. Given this history and use of Telegram, in this work, we study the QAnon conspiracy theory through the lens of the Telegram platform. Also, we select Telegram as it is popular across the globe, hence assisting us in studying the globalization of the QAnon conspiracy theory.

4.2 Methodology and Dataset

An inherent challenge when studying phenomena through platforms like Telegram is to discover groups/channels related to the topic of interest. To discover groups/channels related to QAnon, we follow the methodology explained in Chapter 3. Specifically, we: 1) search on Twitter and Facebook for URLs to Telegram groups/channels; 2) collect metadata for each group/channel; 3) select groups/channels based on QAnon-related keywords. 4) manually validate the selected groups/channels; 5) join and collect all messages from all discovered QAnon groups/channels; and 6) expand our QAnon groups/channels based on forwarded messages shared in already discovered QAnon groups/channels and repeat Step 5. Below, we elaborate on each step.

1. Discovering groups/channels. We use Twitter and Facebook to discover Telegram groups and channels. For Twitter, we follow the methodology explained in Section 3.2, while for Facebook, we use the Crowdtangle API to obtain posts including Telegram URLs [142]. For both data sources, we perform queries with three URL patterns mentioned in the previous chapter: *t.me*, *telegram.me*, and *telegram.org*. Note that the list of these patterns is not exhaustive; there is also the *tg://join?invite* pattern, however, we did not include it in our collection since our initial experiments showed that they are rarely shared on Twitter/Facebook (less than 0.1% more URLs discovered by including this specific pattern). We collect Twitter and Facebook posts, including Telegram URLs between April 8, 2020, and October 10, 2020, resulting in a total of 5,488,596 tweets and 14,004,394 Facebook posts that include a set of 922,289 unique Telegram URLs. Note that the Crowdtangle API tracks and provides data only from publicly available Groups and Pages (i.e., does not include user timeline posts).

2. Collecting group/channel metadata. We use the methodology from Section 3.2 and obtain basic group/channel metadata including: a) Name of the group/channel; b) Description of the group/channel; c) Number of members; and d) the URL type (i.e., channel or group).

3. Selecting QAnon groups/channels. The next step is to narrow down the set of groups/channels to the ones that mention QAnon. To do this, we search for the

Table 4.1: Overview of QAnon dataset.

Dataset	Source	#Groups	#Senders	#Messages
QAnon	Twitter/FB	78	92,322	3,503,381
	Forwarded	84	84	903,611
	Total	161	92,406	4,406,992
Baseline	Twitter/FB	869	195,499	7,983,230

appearance of QAnon-related keywords on Twitter/Facebook posts that shared Telegram URLs or on the group/channel metadata obtained from Step 2. We use two QAnon-related keywords: *qanon* and *wwg1wga*. The former refers to the conspiracy theory itself, while the latter is the QAnon movement's motto that refers to "Where We Go One We Go All." We select these specific keywords mainly because they are prevalent and used extensively by members of the QAnon movement. Overall, we find 204 Telegram groups/channels that include the above keywords in their group/channel metadata or any posts collected from Twitter/Facebook.

4. Validating QAnon groups/channels. Next, we validate that the selected groups/channels are related to QAnon and remove any groups/channels that are not directly related (e.g., mentioning QAnon only once because of mentions in the news). To do this, an author of this study, who has previous experience with the QAnon conspiracy theory, manually annotated the 204 groups/channels obtained from Step 3. The annotator viewed each group/channel via Telegram's Web client and spent 5-10 minutes reading the content shared in the group/channel and checking the group/channel metadata to decide whether the group/channel is related and supports the QAnon conspiracy theory. The annotator focused only on selecting groups/channels that were promoting QAnon or discussing theories related to QAnon and avoided selecting groups/channels that simply mentioned some news about QAnon but their primary focus was on another topic. Note that since many groups/channels are in languages other than English, the annotator used Google's translate functionality to translate content into English. Overall, we annotate all 204 groups/channels and find 77 QAnon groups/channels.

5. Joining and collecting messages in QAnon groups/channels. The next step in our data collection methodology is to join the QAnon groups/channels and collect all their messages. We join all QAnon groups/channels, and then we use the Telethon library [143], which uses Telegram's API [106] to collect all the messages shared within these groups. Note that we only join and collect data from public groups/channels. Initially, we collect 3.5 million messages shared in 77 QAnon groups/channels between September 1, 2019, and March 9, 2021 (see Table 4.1).

6. Expanding QAnon groups/channels. During our manual validation of the QAnon groups and channels, we observed many messages shared in QAnon groups/channels that are forwarded messages from other groups/channels. Aiming to expand our set of QAnon groups/channels, we extract all groups/channels that forwarded messages in the 77 already discovered QAnon groups/channels and manually validate (see Step 4) the top 200 groups/channels in terms of the number of forwarded messages. Note

that we only validate the top 200, as manually checking and validating the groups/channels is time-consuming. Using this approach, we discover an additional 84 QAnon groups/channels. Then, we repeat Step 5 for the newly discovered groups and collect all of their messages. Overall, by combining the initial dataset and the one after expanding the QAnon groups/channels, we obtain a set of 4.4 million messages shared in 161 QAnon groups/channels between September 1, 2019, and March 9, 2021 (see Table 4.1).

Baseline Dataset. We also collect a baseline dataset for comparing it with our QAnon dataset. To collect our baseline dataset, we follow Steps 1, 2, 3, and 5, with the only difference that we use a different set of keywords for selecting the groups/channels (note that we do not validate and manually check the groups/channels because they are not focusing on a specific topic). Specifically, we use a set of keywords obtained from First Draft [144], an organization that aims to fight disinformation on the Web. First Draft provided us with a list that includes 133 keywords/phrases[1] about important events that happened in 2020 (e.g., the US election and the COVID-19 pandemic). Overall, we join 869 groups/channels and collect 7.9 million messages shared between September 1, 2019, and March 9, 2021 (see Table 4.1).

Validation of the Perspective API

Given that the Perspective API is essentially a black box, it is important to assess its performance in our dataset, and more importantly, how well it performs across multiple languages. To do this, we extracted random samples of messages from our QAnon dataset in English, German, and Portuguese. Then we performed annotation on each message to determine whether it was toxic or not. We focus on these three languages as they are the most popular in our dataset. We extracted a random sample of 500 messages for each language while ensuring that our random sample covers the entire score range from Perspective API. We extracted 50 random messages that had a score between 0 and 0.1, 50 messages from 0.1 and 0.2, and so on. Then, we recruited three annotators (Ph.D. students and researchers) for each language; for English, the annotators were fluent in English, while for Portuguese and German, we recruited native speakers. The annotators were provided with the following definition of toxicity: “We define toxicity as a rude, disrespectful, or unreasonable comment that is likely to make someone leave a discussion” (obtained from Perspective API’s website), and were asked to independently annotate each message as toxic or not (the annotators were unable to see the actual Perspective score, they only had access to the comment itself). Then, to obtain our ground truth, we annotated each message as toxic or not based on the majority agreement of the three annotators. We also calculated the inter-annotator agreement using Krippendorff’s alpha coefficient [145]; we find 0.41, 0.43, and 0.44 for English, Portuguese, and German, respectively. The coefficient values ranging from 0.41 to 0.60 indicate that the annotators had a moderate agreement [146] across languages and highlight the subjectivity when people annotate content as toxic or not.

[1] Available in `https://telegra.ph/Keywords-08-03`.

Table 4.2: Performance evaluation of Perspective API.

	English			German			Portuguese		
Thresh.	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
0.50	0.475	0.906	0.623	0.317	0.917	0.471	0.506	0.799	0.620
0.55	0.522	0.858	0.649	0.326	0.881	0.476	0.577	0.753	0.654
0.60	0.562	0.858	0.679	0.363	0.821	0.504	0.580	0.753	0.655
0.65	0.617	0.811	0.701	0.400	0.762	0.525	**0.606**	**0.740**	**0.667**
0.70	**0.686**	**0.740**	**0.712**	0.474	0.643	0.545	0.716	0.506	0.593
0.75	0.717	0.677	0.696	**0.491**	**0.643**	**0.557**	0.735	0.487	0.586
0.80	0.753	0.551	0.636	0.510	0.583	0.544	0.740	0.481	0.583
0.85	0.833	0.394	0.535	0.597	0.440	0.507	0.845	0.318	0.462
0.90	0.920	0.181	0.303	0.684	0.310	0.426	0.919	0.221	0.356

Then, to assess the performance of the Perspective API and select an appropriate threshold for each language (i.e., any message that has a Perspective score above the threshold is considered toxic), we varied the threshold and calculated standard performance metrics like precision, recall, and F1 score (see Table 4.2). Based on our validation results and performance metrics, we treat a message as toxic if it has a score over 0.7 for English, over 0.75 for German, and over 0.65 for Portuguese (thresholds with the largest F1 score, see Table 4.2). Also, our validation results show that the Perspective API does not perform the same across languages; English is the best-performing language (0.712 F1 score), followed by Portuguese (0.667 F1 score), and German (0.557 F1 score). Future work should further validate the performance of the Perspective API on a larger scale and across multiple languages/datasets.

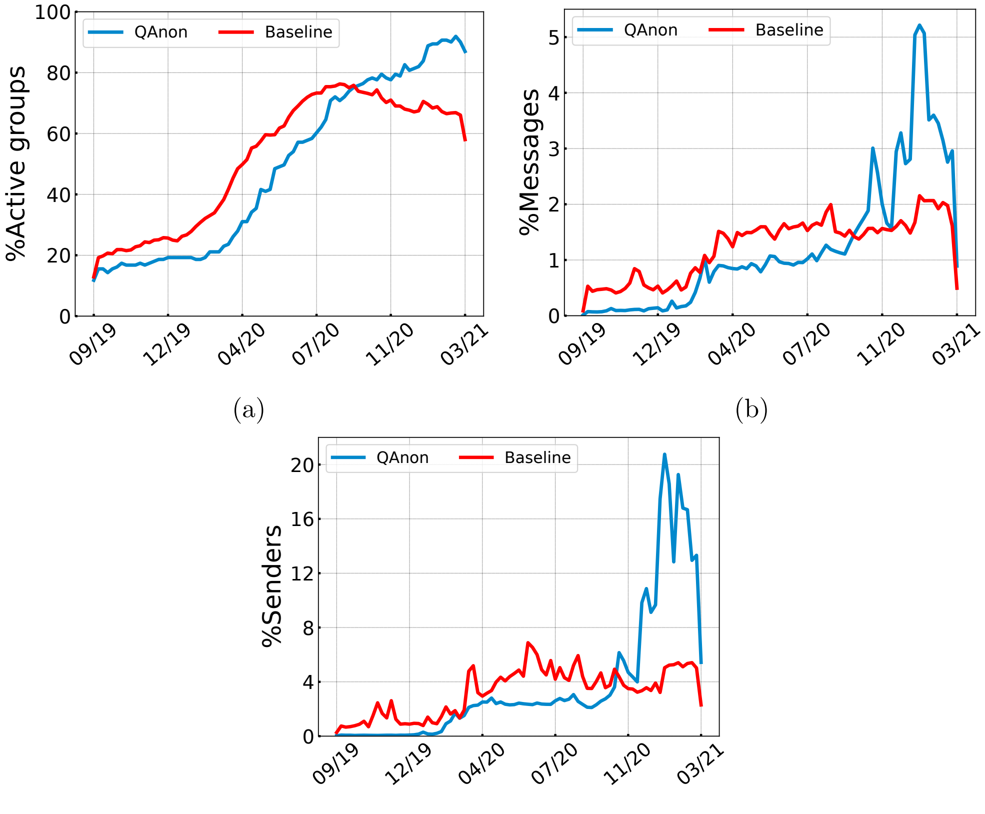
4.3 Results

In this section, we present our analysis for investigating our three hypotheses related to the activity, toxicity, and topics posted in our QAnon dataset.

4.3.1 Activity (Hypothesis 1)

We start our analysis by looking into the general activity across the QAnon groups/channels and how it differs from our baseline dataset.

Fig. 4.1 shows the percentage of active groups, messages, and senders per week in our dataset. When looking at the activity of groups over time (see Fig. 4.1a), we observe that for both QAnon and baseline datasets, we have an increasing number of active groups over time; for QAnon, we have 12% active groups by September 2019, and by March 2021 the active groups/channels increase to 86%. For the baseline dataset, we find 12% and 58% active groups for September 2019 and March 2021, respectively. These increases in the overall activity for both datasets are likely due to Telegram



(a)

(b)

(c)

Figure 4.1: Activity within QAnon and Baseline chats.

becoming more popular over time [147] and the platform is onboarding more users that create more groups/channels on various topics of interest.

When looking at the activity of messages and senders (Fig. 4.1b and Fig. 4.1c), we again observe an increase in activity for both datasets over time. Specifically, by April 2020, we have 1% of all messages for QAnon and 1.5% for the baseline dataset. These percentages increase later on and by 2021, we observe an activity of 2% for the baseline, while for QAnon, we have an activity of over 3% with specific weeks increasing even over 5%. Importantly, we observe that the QAnon activity surpasses the baseline activity by October 2020, which likely indicates that the QAnon movement on Telegram substantially increased by that time, even surpassing other topics of interest.

The larger increase in QAnon compared to the baseline is likely because Facebook [118] removed accounts and groups related to QAnon from their platforms during October 2020, hence users likely migrated to alternative platforms like Telegram. Also, for the QAnon dataset, we observe a peak in activity during early 2021 (over 5% of all messages and over 20% of the users were actively sharing messages), which coincides with the attack in the US capitol by QAnon supporters. This initial analysis indicates that the QAnon conspiracy theory is growing rapidly on Telegram in terms of the

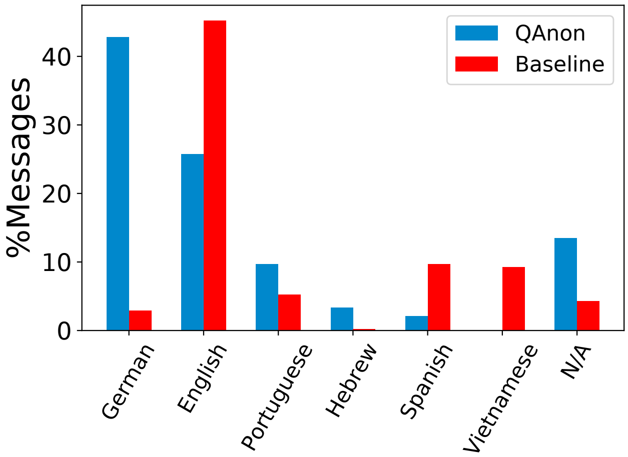


Figure 4.2: Language distribution among messages in QAnon and Baseline datasets.

number of groups/channels (almost 7x increase while baseline has 4.8x increase), the number of messages (over 5x increase while baseline has 2x), and the number of users sharing messages (over 5x increase while baseline has 2x).

Next, we analyze the languages that appear in our QAnon and baseline datasets. Fig. 4.2 shows the percentage of messages for the top five languages in our QAnon and baseline datasets (the figure includes the union of the top five languages on both datasets). We observe substantial differences in the popularity of languages across the two datasets; German is the most popular language in our QAnon dataset, with 43% of all messages (only 3% in the baseline). The most popular language is English for the baseline dataset, with 45% of all messages (26% for the QAnon dataset). Other popular languages in our QAnon dataset are Portuguese (10%), Hebrew (3%), and Spanish (2%).

Next, we look into how the popularity of the five most popular languages changed over time to understand how QAnon became a global phenomenon on Telegram. Fig. 4.3 shows the popularity of the languages over time in our QAnon dataset (we omit the figure for the baseline since there are no substantial differences in the popularity of languages in the baseline dataset). We observe that English was the most popular language between September 2019 and December 2019, with over half of the QAnon-related messages posted in English (55%), with German having a substantial percentage (39% of the QAnon-related messages). Furthermore, between February and April 2020, we observe a substantial increase in the popularity of the Portuguese language, which became the most popular language with 48% of the messages of this period, overshadowing both English and German.

This period coincides with the beginning of the COVID-19 pandemic in Brazil when the virus was first confirmed to have spread to Brazil in February 2020 [148]. Finally, after June 2020, we find that German is consistently the most popular language in our dataset, reaching 55% of the messages, followed by English (28%) and Portuguese (6%) having stable popularity.

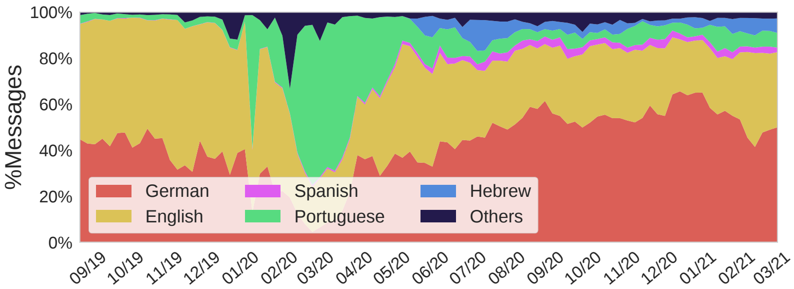


Figure 4.3: Language distribution in QAnon dataset over time.

Remarks. Our results confirm our first hypothesis. QAnon's popularity in our dataset is rapidly increasing and surpassing the baseline dataset (almost a 5x increase in messages and senders in 2021, whereas for the baseline dataset, we only find a 2x increase). Also, we observe substantial shifts in language popularity in our QAnon dataset, with English being the most popular between September 2019 and February 2020 (55%), Portuguese being the most popular between February 2020 and April 2020 (48%), while German is the most popular language after June 2020 (55%). These findings prompt the need to further investigate the multilingual aspect of conspiracy theories that become a global phenomenon like QAnon.

4.3.2 Toxicity (Hypothesis 2)

Here, we investigate the toxicity of content shared in our QAnon and baseline datasets. The QAnon movement has links with events of real-world violence, hence it is important to analyze the toxicity of QAnon discussions on Telegram. We aim to uncover whether QAnon discussions in our dataset are more toxic than other discussions and how toxicity changes over time (i.e., are QAnon discussions in our dataset becoming more toxic over time?).

Toxicity Assessment. To quantify how toxic the content in our datasets is and whether there are changes over time, we use Google's Perspective API [20] to annotate each message in our dataset with a score that reflects how rude or disrespectful a comment is. Following [21], we use the SEVERE_TOXICITY model provided by the Perspective API, mainly because it is robust to positive uses of curse words. We use Perspective API for annotating content mainly because it offers production-ready models that support multiple languages; as of May 2021, the Perspective API supports English, Spanish, French, German, Portuguese, Italian, and Russian. The Perspective API allows us to assess the toxicity of messages posted in any of the seven languages above, which corresponds to 65% of the messages in our dataset. The rest of the messages do not include any text (20% are sharing only audio, video, or images) or are in other languages (15%) that the Perspective API does not support. Note that the use of the Perspective API to assess the toxicity of content is likely to introduce some false positives or biases [149]. Previous work [150], has validated

Table 4.3: Percentage of toxic messages.

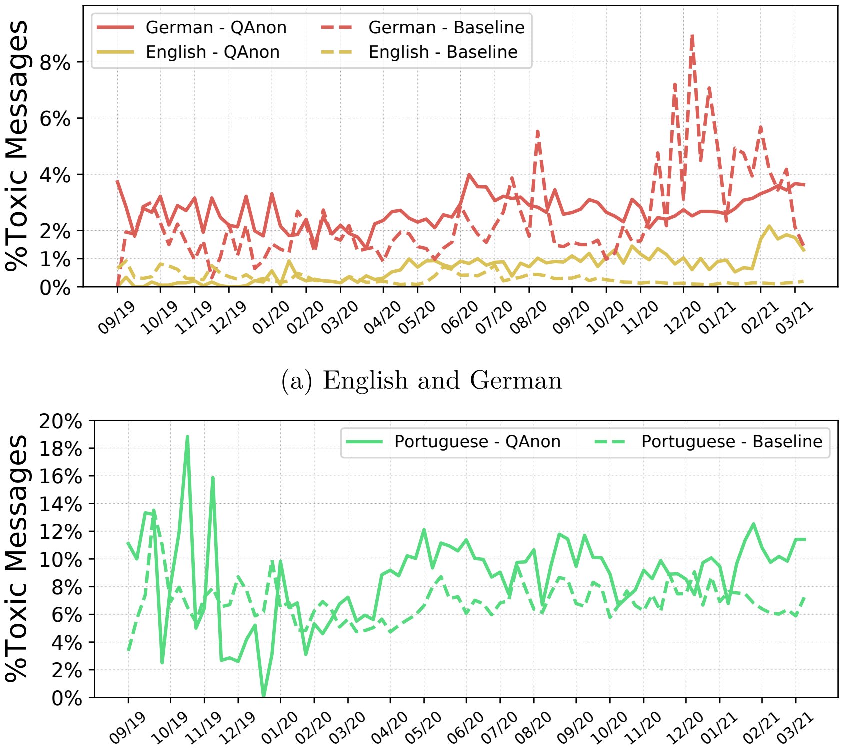
English		German		Portuguese	
QAnon	1%	QAnon	2.8%	QAnon	8.6%
Baseline	0.3%	Baseline	3.3%	Baseline	6.9%
Voat	6.5%	Voat	N/A	Voat	N/A

the performance of the Perspective API, however, it focuses mainly on the English annotations.

Given that, likely, the Perspective API performs differently across languages, we make a manual validation of the performance of the Perspective API in the three most popular languages in our dataset: English, German, and Portuguese (see Appendix for details). Based on our annotation, we treat a message as toxic if it scores over 0.7 for English, 0.75 for German, and 0.65 for Portuguese. We use these specific thresholds because our validation procedure demonstrates that we achieve the highest performance in terms of F1 score when using them. Also, we limit our analyses to the three aforementioned languages mainly because we did not validate the performance in other languages as this task is outside the scope of our study and requires the recruitment of native speakers for each language.

Results. First, we look into the prevalence of toxic messages in our QAnon dataset by comparing it with our baseline dataset, and the Voat dataset obtained from [130]. Voat was a social network that hosted many QAnon followers who migrated from other platforms (Voat was shut down in December 2020). We use Voat as a baseline because it is another platform where QAnon followers migrated to after bans from mainstream platforms, the time period of the Voat dataset is a subset of the time period in our dataset, and because QAnon was very popular on Voat before the platform's shut down [130]. Table 4.3 reports the percentage of messages that are toxic in our QAnon/baseline datasets and the above-mentioned Voat dataset. First, we observe that QAnon discussions in our dataset shared in German and Portuguese tend to be more toxic than discussions in English (2.8% for German and 8.6% for Portuguese compared to 1% for English).

These findings are particularly alarming when combined with the popularity results of these languages in our QAnon dataset. This is because Portuguese and German are overshadowing English during 2020 (see Fig. 4.3), hence less toxic discussions in English give way to more toxic discussions in Portuguese and German. Second, for English and Portuguese, we observe that QAnon discussions in our dataset are more toxic than our baseline dataset; 3.6x greater percentage for English and 1.2x greater percentage for Portuguese. On the other hand, we observe that the baseline dataset has a greater percentage of toxic messages (1.15x more) for German. Third, the QAnon Voat dataset (only available in English) has a substantially larger percentage of toxic messages than the Telegram one (Voat has a 6.4x larger percentage than Telegram).



(a) English and German

(b) Portuguese

Figure 4.4: Toxicity level in QAnon dataset over time.

This difference in the toxicity levels between Voat and Telegram is likely due to the fundamental differences between the two platforms and the audience they attract. While Voat is a fringe Web community mainly discussing conspiracy theories, Telegram is a more general-purpose and mainstream platform. Nevertheless, Voat's toxicity levels are comparable with our QAnon dataset in other languages (i.e., Portuguese), which highlights the need to monitor and further study the QAnon movement across the globe, particularly on platforms like Telegram. These results indicate that platforms like Telegram, which allow users to create their own sub-communities, can be exploited to create fringe communities that can disseminate harmful and toxic content in such prevalence comparable with other notorious communities known for the dissemination of hateful content like Voat.

We also look into how the toxicity in our QAnon and baseline datasets changes over time. Fig. 4.4 shows the weekly percentage of toxic messages. We observe that for English, we have a steady increase of toxic messages over time in our QAnon dataset; before April 2020, the percentage of toxic messages is below 1%, between April 2020 and December 2020 is stable at 1%, while during 2021, we find 2x more toxic messages (2% of all messages are toxic). For German, we observe that our QAnon dataset has a larger percentage of toxic messages between September 2019 and July 2020 (on average of 2.7% for QAnon and 1.7% for baseline), while the baseline has a substantially larger percentage after November 2020 (on average of

3% for QAnon and 5% for baseline). For Portuguese, we observe some big peaks in toxicity before January 2020, however, these peaks are likely because we only have a small number of messages during that period (see Fig. 4.3). Looking at the rest of the figure, we can observe a big increase from 6% to 12% between early 2020 and May 2020. We manually examined some of these toxic messages, finding that they are related to the COVID-19 pandemic in Brazil, including anti-vaccine conspiracies. Also, we find politics-related messages that attack two Brazilian ex-ministers who left the government during this period. Overall, similarly to English, we observe an increasing trend of toxic messages posted in Portuguese over time in our QAnon dataset.

Remarks. Our analysis partly confirms our second hypothesis; we find that QAnon discussions in our dataset are more toxic than our baseline for English and Portuguese (1.2x and 3.6x more toxic messages for English and Portuguese, respectively). Our German QAnon dataset does not support our hypothesis since we find a higher percentage of toxic messages in our baseline (1.6x more toxic messages in the baseline). Alarmingly, our results show an increase in QAnon content toxicity over time in our Telegram dataset. These findings emphasize the importance of monitoring such groups within the Telegram platform and taking moderation actions in cases where communities orchestrate campaigns that might have a negative impact in the real world (e.g., real-world violence).

4.3.3 Topics (Hypothesis 3)

Thus far, we analyzed our datasets' activity and toxicity aspects without analyzing the topics of discussion. Here, we analyze the content of the messages shared within QAnon groups/channels using two different approaches namely BERT topic modeling and qualitative analysis.

BERT Topic Modeling. To analyze QAnon discourse across multiple languages, we use a Bidirectional Encoder Representations from Transformers (BERT)-based topic modeling methodology by [123]. We use a pre-trained multilingual BERT model (distiluse-base-multilingual-cased) from [151] to embed documents from multiple languages to the same high-dimensional vector space. We select this specific model mainly because it supports 50 languages and performs well in semantic similarity tasks. Then, we use Uniform Manifold Approximation and Projection (UMAP) proposed by [152] to reduce the dimensionality of the extracted embeddings. This is an important step, as it allows us to increase the performance and scalability of the next step (i.e., clustering). Then, we group the reduced embeddings using the HDBSCAN algorithm [153]. We treat each cluster as a separate topic and then we use hierarchical reduction (i.e., iteratively combining the most similar clusters) to obtain a small number of high-level topics/clusters. Finally, to generate topic representations, we calculate the centroid of each cluster based on the embeddings of all documents in the cluster and then select the most similar words (based on the BERT embeddings of the words that appear in the documents of each cluster) that are closer to the centroid.

Table 4.4: Top topics of messages.

Topic	Terms	#Messages
Politics	trumppresidente, trumppresident, presidenciales, presidencial, senatswahlen, presidential, presidenciais, diabolsonaro, kongresswahlen, presidency, obama, presidenttrump, republicano, impeach, impeachment	353,696
Reactions	hahahahahaha, hahahaha, hahahahaha, hahahah, hahaha, ohhhh, ahhhh, mhhh, ahhh, hahah, ohhh, uhh, haha, ahh, ohh, dahingerafft, yhwh, mhh, hmmmm, oooh	257,039
Enviroment/Masks	wwf, stromaggregate, noah, kohlekraftwerke, atomkraftwerken, boooooooooom, maskenkontrolle, atomkraftwerke,mikroelektronik, kontrollgruppe	206,848
Nazis	nazideutschland, nazistas, neonazis, nazista, fascists, fascist, fascistas, polizeigesetz, fascism, nazis, fascismo, bundespolizei, faschistischen, kriminalpolizei, massenproteste	175,201
Apocalypse/Holocaust	wikileaks, killuminati, reichstagssturm, apocalisse, apocalipse, rechtsradikaler, apokalypse, johnfkennedyjr, apocalypse, weltkriegen, holocausto, doomsday, rechtsradikale, holocaust	169,319
COVID-19/Vaccines	impfenden, vacinacao, vaccinations, vaccines, impfen, impfens, vaccination, vacunarse, grippevirus, grippeviren, vacunado, vacinar, ungeimpft, virusnachweis, geimpften, impfgruppe	159,787
Video Sharing	videokanal, videobeitrag, youtubekanal, videonachricht, originalvideo, schockvideo, kurzvideos, video, videolink, beweisvideos, videointerview, youtubelink, videoschalte, videobotschaft, youtube	119,238
Information Warfare	staatsterror, infokrieg, cyberkrieg, patriotsfight, atomkrieg, terrorists, terroristas, militari, weltkrieges, weltkrieg, staatsfeind, militares, vietnamkrieg, weltkriegs, military	116,618
Satanists	satanists, satanismo, antichristen, satanisten, satanism, satanistas, antichrist, satanismus, hausdemokraten, satanist, satanistischen, anticristo, cristianismo, satanischer, satanic	105,151
Q News	wahrheitssuche, qnews, wahrheitskanal, halbwahrheiten, wahrheitssucher, justthenews, hoax, breakingnews, faktenchecker, wahrheiten, extremnews, telenews, conspiracies, freetruthmedia, q_for_you_news,	97,929

We apply this topic methodology after preprocessing all messages by removing emojis and URLs from the text and filtering out messages with an empty body (i.e., messages sharing only URLs, emojis, videos, or images). We focus only on messages posted in the top six languages in our QAnon dataset and we remove very short messages (less than 5 words). After our preprocessing steps, we end up with a set of 2.2 million messages, which is the input to our topic modeling approach. Since our topic modeling approach relies on UMAP, a stochastic technique, our approach can yield varying results on different runs. To alleviate this, we train five separate topic models and select the one that provides the highest average coherence score. For each model, we hierarchically reduce the number of topics to N (we experimented with numbers between 10 and 20) by iteratively combining the most similar clusters until we end up with N clusters (each cluster represents a high-level topic). For each model, we calculate the coherence scores for $N \in \{10, 15, 20\}$ and then select the model with the largest average coherence score. To select the number of topics to present, we again select N based on the coherence scores; we obtain the largest coherence score when $N = 10$ (0.58 vs. 0.53 and 0.52 for 15 and 20, respectively). Below, we report our analysis using the best-performing model in terms of the coherence scores.

Table 4.4 reports the ten extracted high-level topics along with the number of messages that are mapped to each topic (note that 21% of the messages are not mapped into any topic and they are considered noise), while Fig. 4.5 shows the distribution of messages into these topics per week in our dataset.

The most popular topic in our QAnon dataset is Politics (353,000 messages); by examining the terms and some messages mapped to this topic, we find political messages in various countries like the USA, Germany, Brazil, and Italy. These results compound previous findings from [130] and [131] that found political discussions and discussions of international topics in Voat's and YouTube's QAnon community. Other popular topics in our QAnon dataset are related to reacting to other messages during a discussion (257,000 messages), discussions about environmental issues and masks (206,000 messages), discussing German news and Nazis/Neonazis (175,000), as well

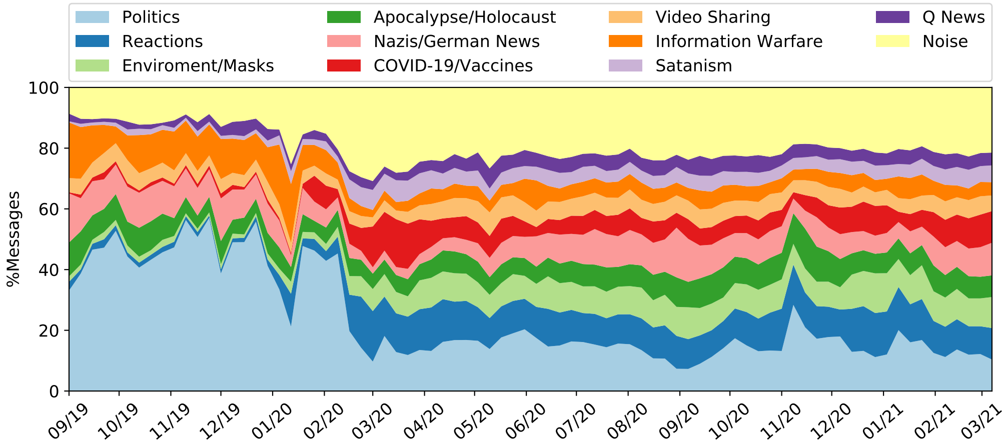
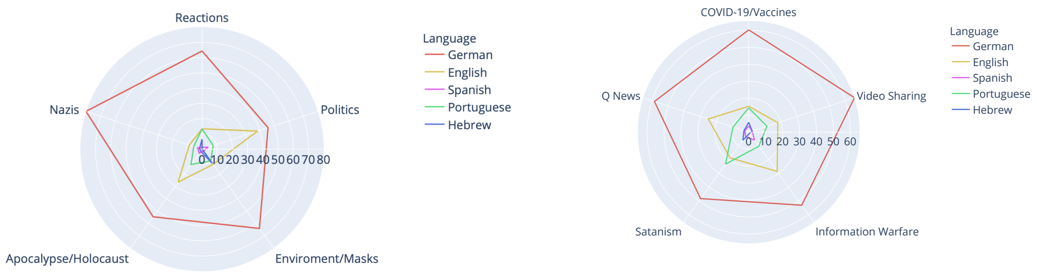


Figure 4.5: Topic distribution in QAnon dataset over time.

as historical events (holocaust) or possible future events (apocalypse) (169,000). By manually inspecting messages referring to the holocaust, we find that QAnon followers call the holocaust a hoax and have a holocaust denial approach to this specific topic. Another popular topic in our QAnon dataset is the COVID-19 pandemic and the debate around vaccines (159,000 messages).

Again, we inspect some messages on this topic. We find that QAnon followers have a strong anti-vax ideology and share a lot of false information about this subject. Some examples include messages claiming that COVID-19 vaccines make people sterile, that vaccines are a lie and a fraud, and that people with medical professions are refusing to get vaccinated because they know vaccines do not work. Also, we find several messages pointing out that the COVID-19 pandemic is a plan of Bill Gates to reduce the earth's population. Also, we find a topic related to sharing videos to disseminate QAnon ideology (119,000 messages), highlighting that videos play an integral role in QAnon. The rest of the topics are related to cryptic messages about Information Warfare (116,000), a topic that alleges that politicians are actually Satanists (105,000), and a topic for disseminating news about QAnon or Q drops (97,000). Our results confirm and reinforce anecdotal evidence presented by [121] highlighting that QAnon follows an anti-vax ideology and that they treat world politicians as arch enemies (e.g., by claiming they are Satanists).

Looking at the popularity of these topics over time (see Fig. 4.5), we find that before February 2020, QAnon discussions are mainly related to Politics, with almost 50% of the messages being on that topic. After February 2020, we observe that the popularity of the Politics topic decreases (below 20% of all messages shared per week). We observe the insurgence and the popularity of other topics like "Reactions", "COVID-19/vaccines", and "Environment/Masks". The increase in popularity of the topic reaction likely indicates an increase in users' engagement with QAnon-related messages. Additionally, we observe that topics that emerged after February 2020 are long-lasting as they have a considerable percentage of all weekly messages during the whole time period until the end of our dataset. Overall, these results highlight



(a) (b)

Figure 4.6: Topic distribution in QAnon dataset across languages.

that QAnon's discussions are evolving over time and that nowadays, QAnon is not only related to Politics, rather QAnon followers discuss a wide variety of topics that can be weaponized for spreading potentially false or harmful information (e.g., false information on vaccines and the COVID-19 pandemic).

Finally, we look into the languages of the messages in each topic to understand if topics are specific to one language and quantify how popular these topics are in each language. Fig. 4.6 shows the percentage of messages that are assigned to each topic and each language (e.g., 60% of the messages in the COVID-19/Vaccines topics are shared in German, see Fig. 4.6b). Unsurprisingly, the most popular languages in almost all topics are German and English, mainly because of their popularity in our QAnon dataset. In the Politics topic, we observe similar popularity between German and English, with 45% and 38% of all Politics messages. 7% of Politics messages are shared in Portuguese, while for the messages in Hebrew, we find that they rarely talk about Politics (only 0.1% of all Politics messages are in Hebrew). For the COVID-19/Vaccines topics, we find that Portuguese and English have similar popularity, highlighting that there is likely a lot of false information disseminated in Portuguese related to the pandemic in QAnon groups (based on our manual examinations, we find a lot of false information in that specific topic). In summary, our language-specific distributions in Fig. 4.6 indicate that most of the topics are not specific to one language. Rather they are discussed across many QAnon groups/channels and, more importantly, across many languages.

Qualitative Analysis. After analyzing the text of the messages automatically using a multilingual topic modeling approach (as described above), we aim to investigate in more detail and depth the contents of the messages using a small-scale qualitative analysis. This kind of analysis is important since it allows us to extract unique insights into specific narratives related to QAnon that are shared on Telegram, which are hard to extract with aggregate topic modeling techniques. In particular, we apply thematic coding analysis on a random sample of 400 messages extracted from our Telegram dataset. We randomly extract 50 English messages from each topic presented in Table 4.4 with the exception of the "Reactions" and "Video Sharing" topics; we exclude these topics from analysis as they are general reactions to messages or they just share videos (analyzing videos shared by QAnon followers is outside the scope of our work). We follow this sampling approach as our main goal is to qualitatively assess the contents of a diverse set of messages that span across the various topics that exist in our dataset.

Although the groups are related to QAnon, users talk about a wide range of topics. We aim to find out the actual content of the discussions inside the groups. In this step, we develop a codebook of the content frequently discussed in the groups using thematic coding [124]. This codebook provides us statements as pieces of information appearing in a substantial number of messages indicating the overall flow of information among the users. Constructing the codebook is an iterative procedure. First, we read all the messages to get familiarized with the messages. Then, we read the messages to extract meaningful pieces of information, namely codes, for each message. In the next iteration, we look at the codes to refine them to better fit the content. Next, we categorize the codes into different themes iteratively. Themes can be removed, merged, or split in different iterations, and codes can be reallocated among the themes. We continue this iterative procedure until no further change is required and the codebook is stabilized. In the end, two properties are derived for the messages: (1) topics of the content, and (2) stances of the statements. In the following, the descriptions of the properties and their corresponding themes are explained.

Topics: The messages discuss a wide range of topics. Each message can contain more than a single topic. Overall, three high-level topics are identified: politics, COVID-19, and media. These three topics consist of different themes and codes as follows.

- **Politics**: As expected, politics is the most common topic among all the topics inside the groups. Although users mostly talk about US politics, issues about other countries are discussed too. They share news, express their opinion, and talk about consequences. They affect each other and get affected. The themes and codes related to politics are classified as follows:
  - Pro-Trump/Anti-Democrats. The main controversial discussion among politics-related messages is about the conflict between the two parties. A lot of messages support Trump's actions or accuse Joe Biden and the Democrats. Some of them see the conflict more seriously anticipating a civil war in case Democrats continue their problematic activities. Here are some code examples: *Trump rebuilt the economy and brought back jobs, Trump is the best president, God, and Trump, and Q save us, Democrats are slaughters of children, Send an army to satanic Democrats, Storm the gates of hell.*
  - The election. The 2020 United States presidential election is a trending topic among users in QAnon groups. There are different discussions around the election including encouraging to vote for Trump, expressing their belief that Trump wins the election, and also insisting on the fact that fraud has occurred in the election. Here are some code examples: *Trump wins the election, Trump, and Q said they would commit voter fraud.*
  - Foreign political issues. Users discuss various issues related to the US and foreign political issues. We observe support for relationships with other countries and also accuse politicians of collusion with other countries. Here

are some code examples: *Russia is behind a hack, thanks to the courage of American Heroes, the ISIS Caliphate has been destroyed, Chinese whistleblower about Obama & Daddy Biden.*

- **COVID-19**: Coronavirus is a hot debate in QAnon groups. People talk about different aspects of the disease, share health guidelines, discuss how COVID-19 affects the economy and lifestyle, and also politicians' reactions to Corona situation. The themes and codes related to politics are classified as follows:
  - Fake virus and vaccines. Since Coronavirus is a brand-new virus, even for virologists, there are numerous rumors about it. People have doubts about the reality of the virus and the vaccines. There are a lot of messages indicating that the vaccine and the virus are fake believing that they are intentionally made to harm people. Surprisingly, We even observe messages denying other viruses like HIV. Here are some code examples: *There is no Influenza A/B C D E F G.., no person has ever made another sick by contagion, controlling the population by the vaccine, vaccines have killed people .*
  - Restrictions rules. To control the contagion of COVID-19, governments all around the world have enforced protocols restricting people's normal way of life. They complain about these restrictions and blame politicians for setting these rules. They even suspect that these rules are set with destructive intentions. They support protests in different countries and encourage people to attend. Here are some code examples: *The government set rules to take away our freedom, forcing people to wear masks kills our freedom, Berlin: anti-Nazi lockdown protest.*
- **Media**: Policies and regulations of the mainstream media and social networks have always been a subject of debate. For example, some of the platforms removed QAnon supporter content and accounts. Users talk about their perspectives on these policies and regulations.

  We observe that people are complaining about a huge amount of fake news spread on social networks and the orientation of media and social networks towards political parties, trending issues, etc. The themes and codes related to media are classified as follows:
  - Fake news. Social media platforms have facilitated the dissemination of a high rate of fake news. Although fact-checking agencies try to debunk frequently-spread fake news, a significant number of fake news affect social media users and deceive them to forward them even further. People complain about being exposed to such a large number of fake news. Here are some examples: *We don't trust any news, Fake news all over around and in media too.*
  - Moderation and Censorship concerns. The news published by mainstream media implies support or contradiction towards a specific party or belief. Users express their disagreement with mainstream media's orientation. People also accuse social media platforms of eroding freedom of speech by applying censorship. Here are some examples: *Facebook, Google, Youtube, Twitter, MSM censor and remove accounts, the media lie about Trump, Facebook, and IG are only removing #QAnon.*

Table 4.5: Distribution of topics into the sample dataset.

	Topic	%
Politics	Pro-Trump/Anti-Democrats	28.0%
	Foreign political issues	10.0%
	The election	6.0%
COVID-19	Fake virus and vaccines	13.5%
	Restrictions rules	7.5%
Media	Fake news	9.0%
	Moderation and censorship	4.5%

Stances: Each message implies a specific stance. It could be for, against, or neutral towards a specific topic. It can be more intense and even encouraged to take specific action including attending a protest, refusing to wear a mask, or getting vaccinated. The themes and codes related to stances are classified as follows:

- Neutral Content. A high percentage of messages simply transfer information like news and instructions. Note that we do not distinguish between right and wrong claims. These kinds of messages are only neutral in their orientation towards anybody or phenomena. Here are some examples: *Zinc and Vitamin D help stop the virus, Biden administration roles out gun control plans, Suspect charged in killing of U.S. Army veteran .*
- Polarized Opinion. We observe a high percentage of the messages to be positive or negative opinions of the users combined with pieces of information. Sometimes, the language is toxic too. Here are some examples: *Satanic transgender agenda makes God mad, Masks are against our freedom.*
- Call for action. At more intense stances, users encourage people in support or opposition to a specific political party or regulation. For example, people who are running a protest invite people to attend. Here are some examples: *Don't inject the vaccine into innocent children, We are planning a huge international protest in Berlin, I invite you to attend.*

Qualitative Insights. In this section, we go through the findings of the qualitative analysis to find out the prevalence of each topic and highlight those that are popular among the users.

We observe that 28% of all messages in our sample set are related to "Pro-Trump/Anti-Democrats" which is the most popular topic. "Foreign political issues" and "The election" contribute to 10% and 6% of the messages respectively, as shown in Table 4.5. Although there are a few messages against Trump, most of the discussions related to US internal politics support Trump and accuse Democrats of hiding the truth or malfunctioning. In these discussions, users rely on Q drops and talk about the positive side of Trump's actions and plans. They introduce Democrats as evil with phrases such as "slaughters of children". They warn people to wake up and be careful about many lies spread against Trump or his supporters.

Table 4.6: The number of messages in each code.

Topic	Code	#
Pro-Trump/Anti-Democrats	"The Great Awakening"	20
	God bless Trump	16
	Democrats are evil	17
The election	Fraud is committed in the election process	10
	Misleading information for the election	5
	Trump is the winner of the election	3
Foreign political issues	Russia is menacing U.S.	4
	China is threatening U.S.	10
	Hackers and cyber-security against U.S.	7
Fake virus/vaccine	Coronavirus is fake	10
	Vaccine is harmful	18
	Coronavirus is not dangerous	6
Restrictions rules	Protests against restriction rules on streets	8
	Authorities are controlling the population by restriction rules	8
	Opposing restriction rules	10
Fake news	Media manipulates people	16
	Complaining about too many fake news within Telegram groups/channels	9
	Lies in media	5
Moderation and censorship	Pro-trump banned from social media	7
	Censorship committed by online social media platforms	4
	Government censors/bans protests	4

Messages coded with foreign political issues are mostly oriented against other countries, specifically China and Russia. The number of appearances of top codes in each second-level topic is shown in Table 4.6. For instance, a message: "Attention, American business owners! #DYK the Chinese government steals intellectual property and trade secrets from U.S. companies and researchers. Learn more about the risks the Chinese Communist Party poses to corporate America at FBI." The other common debate is about the election. There are several messages opposing the election results. We observe 10 messages directly claiming that fraud was committed in the election. Here is an example: "I honestly feel it is just theater to destroy our trust in democracy. Prime us for the one world government. Trump and Q said all along they would commit voter fraud, then they go and pull the most amateur bullsh*t you can imagine, knowing millions of us are waiting for it."[2]

Regarding COVID-19, 13.5% of the messages are related to "Fake virus and vaccines"

[2] Note that we censor offensive words in the examples.

while 7.5% of them are related to “Restrictions rules”. COVID discussions are popular among users and they talk solely about it or in relation to politics and politicians. Comparatively, a high number of messages state that Coronavirus is a fake virus and is not dangerous. In line with the previous point, anti-vaccination is a hot topic too. People bring evidence, news, and opinions to prove that vaccines are harmful and even intentionally developed to control the population. Here is an example: “The Coronavirus Hoax & The Bill Gates Vaccination Depopulation Agenda Exposed Don't buy into the fear... if you die, then it's your time. Quit stressing out about this. live your life people - Anon #bill_gates #covid”.

Users react to the protocols and restrictions due to COVID such as wearing masks and lockdown. There are a lot of messages complaining about these restrictions. They even plan protests against restrictions and invite people to join them inside the groups. People's orientation about COVID and the restrictions can be viewed through the lens of political situations. For example, observing messages stating that “Coronavirus is planned to stop Trump” or “Why is no vaccine released before the election?” brings this point to mind that people are against vaccines because of their political favorites. In more intense complaints, users call the government in many countries Nazis, Fascists, and so on only because of setting restriction rules. Here is an example: “Berlin. Anti-Nazi Lockdown protest. You can't change the world by sitting on your arse, watching TV, and clapping for your fascist government rules like a mind-controlled zombie. Arise Humans”

The other popular subject that attracts users' interest is mainstream media and social networks. “Fake news” in mainstream media social networks makes for 9% of the messages while 4.5% of them are related to “Moderation and Censorship”. Users talk about the spread of fake news on social media and show their lack of trust in the news. Some messages claim that mainstream media is under the control of politicians and political parties lying about what's really happening in the world and trying to mislead people. As a reaction to moderation in social media platforms, users accuse these platforms of violating freedom of speech. They believe that too many accounts of Trump supporters are banned because of their political orientation and that's not fair. Here is an example: “You got DC totally militarized. You got the fed reserve folding as we speak. You got Biden signing fake executive orders on a fake desk and a fake face. You got unrest abroad all of a sudden like it was organized on q, pardon the pun. You got everyone who's famous and pro-trump banned from Twitter all at the same time - Dorsey I'm sure was relieved of responsibility some time ago. This was planned to help show the masses when they do wake, how bad things were (planned for).”

Users take different stances to convey their information and opinions in the messages. In 27% of the messages, information is transferred neutrally. On the other hand, we observe messages with an orientation towards somebody, a political party, or an issue. In 61% of the messages, users clearly state their support or opposition to something such as a political party. In a more intense manner, users invite others to commit an action in support or against an opinion such as attending a protest against lockdown. We notice that 12% of the messages include a call for action.

Knowing the stance of the messages helps us to understand the extent of extremism in the media. The high percentage of polarized messages implies that users are under the pressure of radical messages. It indicates that these groups might be used to promote different ideologies among the audience.

Here is an example of a polarized message: "World Events today are far beyond Rep v Dem politics... this is BIBLICAL Prophesy battle being fulfilled before those with eyes to see and ears to hear. President Trump + Patriots worldwide + Almighty God = WINNING WINNING WINNING – and the Best is Yet to Come. Pray Patriots... our Prayer weapons are needed in this battle, they make a difference at this time in Humanity. The Truth Sets us all Free."

We observe messages designed for inviting people to take real-world actions such as protesting against lockdown and refusing to respect the rules. Here is an example of a message with a call for action: "Dear friends anywhere in Europe! Dear fighters for freedom, constitutional rights, and truth! On 29.08.2020 many of us (very many!!!) will be on the streets of Berlin to protect the peace and freedom of Europe. Please come from all countries to join us in Berlin. Berlin is important for all of us: Germany has the EU presidency, and Berlin is where the tests and numbers come from that are the tune to which the whole world has been singing for months."

Remarks. Our topic modeling and qualitative analysis confirm our third hypothesis. QAnon followers on Telegram share and discuss a wide variety of topics, and they disseminate conspiratorial or false information about Politics and the COVID-19 pandemic. Also, our analysis shows that the QAnon discourse is becoming more diverse, with the Politics topic losing popularity after February 2020 and other topics like the COVID-19 pandemic gaining a substantial share of the discussions. Also, most of the topics are not specific to one language, but rather they span across multiple languages.

4.4 Discussion

We performed the first multilingual analysis of QAnon content on Telegram. We joined 161 groups/channels on Telegram and collected a total of 4.4 million messages shared over 18 months. Using Perspective API and multilingual topic modeling, we shed light on how the QAnon conspiracy theory evolved and became a global phenomenon through Telegram.

First, our primary objective was to understand how conspiracy theories evolve over time and across languages (RQ4). Our analysis shows that QAnon content on Telegram is increasing in volume (during 2021, 5x increase in terms of messages). The number of active groups is increasing in both QAnon and Baseline datasets during the time of collecting the group URLs from Twitter and Facebook. Unlike the Baseline dataset, surprisingly we observe that the number of active groups/channels in the QAnon dataset has been increasing until March 2021. This indicates that

QAnon groups/channels are comparatively long-lasting and active. Also, a considerable increase in the percentage of messages and senders in early 2021 implies that QAnon groups/channels have a stronger reaction to events in the real world. Language analysis reveals distinct patterns: English emerges as the dominant language from September 2019 to February 2020, constituting over half of the messages; Portuguese gains prominence from February 2020 to April 2020, accounting for nearly half of the messages; subsequently, German becomes the prevailing language after June 2020, comprising over half of the messages. An immediate implication of this increased activity is the need for real-time monitoring systems that can help us track the spread of QAnon content in online messaging platforms like Telegram, similar to systems developed by [74]. This kind of system would at least allow journalists and public authorities to counter misinformation campaigns that are designed to target radical groups. These findings address the associated research question, indicating a substantial increase in volume over time compared to groups focused on alternative topics, alongside notable shifts in language preference.

Second, we tried to reveal how toxic the conspiracy theories discussions are over time and across languages (RQ5). Our toxicity analysis extends the findings from [135], which indicates that QAnon is sharing a lot of toxic and violent messages. Our analysis paints a nuanced overview of the toxicity of QAnon across multiple languages and highlights that there are substantial differences across languages. Our results and toxicity validation have several implications for researchers focusing on QAnon or hate speech. First, our results show that QAnon content in languages like German and Portuguese is significantly more toxic than English content, emphasizing the need to study this problem through the lens of languages other than English. Second, in contrast with the findings from [130], we find QAnon content being more toxic compared to the baseline for English and Portuguese, which shows the differences that exist across platforms and time. In addition, QAnon followers are likely becoming more toxic over time, particularly after multiple moderation interventions (i.e., bans) from mainstream platforms like Reddit, Facebook, and YouTube. Indeed [21] shows that moderation interventions on Reddit can lead to increased radicalization signals after users migrate to other platforms. Third, our toxicity validation highlights that production-ready models like the Perspective API perform differently across languages. This prompts the need to further study the performance of these models across languages and investigate ways to improve their multilingual aspect. We have successfully addressed our research question by demonstrating that QAnon content on Telegram exhibits a higher level of toxicity compared to content found in groups/channels focusing on other topics.

Finally, we provided answer to the question of what topics and conspiratorial content are popular among fringe communities (RQ6). Our topic modeling analysis reinforces findings from previous work [130, 131] and complements these previous efforts by investigating the same phenomenon on Telegram. We showed that QAnon on Telegram is becoming more diverse over time in terms of their discussed topics. Also, we found messages that were sharing false information across multiple languages, particularly related to the COVID-19 pandemic and international politics. This emphasizes the emerging problem of the spread of multilingual false information and the challenges in detecting and tackling it. Our work highlights the necessity of organizations that focus on fact-checking and addressing the dissemination of QAnon-related false information across languages and countries (e.g., efforts similar to the #CoronaVirusFacts Alliance focusing on the COVID-19 pandemic [154]). Our qualitative analysis provides evidence of how QAnon followers discuss various topics related to real-world events like the COVID-19 pandemic (e.g., sharing their anti-vax ideology). This highlights how QAnon evolved over time and it essentially became a “super conspiracy theory” that shares a lot of misinformative and unconfirmed information about a variety of topics like the COVID-19 pandemic and vaccines, social media platforms, and various political issues around the globe. Also, our qualitative insights on the discussions about the call for action and censorship by social media platforms are particularly interesting as it is likely to have broader implications. For instance, QAnon followers might become hostile towards these social media platforms and try to boycott them (e.g., call people to migrate to less-moderated parts of the Web). Overall, our qualitative insights highlight and paint a nuanced overview of how QAnon evolved from specific claims about a secret plan to overthrow Donald Trump to a multi-faceted and vast conspiracy theory that shares misinformative content across several topics that are of interest to society. The research question regarding the topics of the discussions is addressed by the findings presented above. Our analysis revealed that these discussions span various political topics, often involving the dissemination of false information. Furthermore, we found that the popularity of these topics shifts over time, indicating the dynamic nature of QAnon discourse and its evolving focus on political narratives.

5

Characterizing Information Propagation in Fringe Communities on Telegram

In the previous chapter, we explored the evolution of the renowned conspiracy theory, QAnon, on Telegram. Our findings revealed the rapid global dissemination of conspiratorial content on the platform. Building upon this, our focus shifts to probing the mechanisms through which misinformation proliferates within fringe communities on Telegram. Within this chapter, we examine the impact of the forwarding mechanism on the spread of misinformation, assess the lifespan of various message types, and analyze the reach, toxicity and emotional tones of messages. This investigation aims to provide a comprehensive understanding of the viral nature of misinformation on Telegram.

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 of them come especially as a response to users who have been suspended on other social networks for violating their terms of service, such as Gab [49, 50], Parler [128], BitChute [155], and Voat [21]. Along with these alternatives, especially with the use of smartphones, online messaging platforms have gained more and more space in this environment as discussed in Chapter 3. These platforms, such as Telegram, WhatsApp, WeChat, Viber, Discord, 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 [8], fake news campaigns in political elections [9], and even the influence of the Russian-Ukrainian war [156].

One of the primary messaging services is Telegram. Launched in 2013, Telegram is nowadays an online messaging platform with 800 million users [122] 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.

For this reason, many traditional media, companies, and public figures use Telegram to publish their news officially, share their ideas, and even promote some products and services. Moreover, 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 [157–159].

Despite the popularity of online 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 [57]. 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 online 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:

- **RQ7:** How does the forwarding mechanism contribute to information dissemination within fringe communities?
- **RQ8:** What is the lifespan of various types of content shared in fringe communities?

To answer these questions, we first create an extensive data collection of 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. Also, the behavior of different types of messages, such as URLs, regular messages, forwarded messages, and direct messages. Also, we analyze the users' 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.

Below, we summarize the main findings of this study.

Main Findings: Among other things, our study yields the following main findings:

- 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 observed 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.
- 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 any 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 online 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 while the messages are disseminated locally, the forwarding mechanism has increased their reach significantly.

Chapter Organization. The rest of this chapter is organized as follows. Section 5.1 provides the background on misinformation propagation in online messaging platforms and the related work, while Section 5.2 presents our data collection methodology and dataset details. Section 5.3 presents our analysis of different perspectives on information propagation. Finally, we discuss the implications and conclude in Section 5.4.

5.1 Background and Related Work

The propagation of information has been an increasingly debated subject, mainly due to the popularity of online social platforms, which allow people to be reached very quickly. Still related to extremism in this cyberspace, Guhl and Davey [158] 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 et al. [70] 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 Walther and McCoy [160] 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 study on online extremism [161] investigates 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 ways in which terrorist groups such as ISIS leverage Telegram's encrypted environment [94–96]. 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, especially Telegram and Discord, is that they are also commonly used for the practice of diverse forms of digital scams [162]. In particular, a vast range of cryptocurrency schemes to steal digital activities from users to pump and dump manipulation[97, 163–167]. The architecture of these instant messaging favors this kind of abuse as it provides a private, secure, and anonymous place for cyber criminals to communicate with their customers in underground markets [159] often offering illicit products and services in a more public network attracting their targets to this unmoderated and closed platform [168], in which some authors suggest that a hidden underground space of fakes, extremism, scams, and conspiracies coexists with the general public within the platform [157, 162].

There are also other issues with the usage of Telegram in other countries. Nikkah et al. [88] examine Telegram usage among Iranian immigrants, specifically inspecting the moderation mechanisms within these Telegram communities. Hashemi et al. [89] undertake an extensive evaluation of 900,000 Iranian channels and 300,000 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.

In the direction of a more panoramic view and characterization of Telegram, some previous work focuses on collecting large-scale data from Telegram and studying emerging research problems. Dargahi et al. [92] collect data from 2,600 Telegram groups/channels, conducting a structural review of the posted content within these communities. Abu-Salma et al. [87] execute a user-based study to gauge perceptions surrounding Telegram's security measures. Naseri and Zamani [93] 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. With high engagement and millions of daily hits 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 focused 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 how information propagates in this ecosystem. In this work, we provide an analysis of information propagation among Telegram groups and channels to provide a better understanding of how information spreads in a large-scale ecosystem.

5.2 Methodology and Dataset

Online messaging platforms present some peculiar characteristics that create unique challenges in studying such an environment. 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 is many-to-many communication among users and all of them can send messages. In channels, there is a 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, it is possible for the administrators to 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. Besides that, users of a chat can share messages with other chats by redirecting content through the forwarding mechanism. This flow of messages creates an interconnected network within Telegram, enabling large-scale information propagation through those detached chats, spreading content throughout the whole network.

In this chapter, we select to gather data from QAnon communities within Telegram, since this showed to be a growing topic within this platform [141] that exchanges messages on a global scale through hundreds of groups and channels dedicated to discussion of these conspiracy theories. Therefore, to build a large-scale data set of shared messages on Telegram, we use data of groups and channels made available from Chapter 4 as a starting 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 search on the Web for public invite URLs users post 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 to our initial set of chats. We use the methodology from Section 3.2 and obtain basic group/channel metadata including: a) Chats' title; b) Description of the chats; c) Number of members; and using the methodology from Section 4.2, d) Messages sent within each chat.

Extracting sources of the messages: Since a high number of messages inside the chats are composed of forwarded messages from other chats, including channels, groups, and even individual users, they have an identifier that indicates who is the original source of the message. Then, we extract all 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 different sources, largely expanding our initial data. Moreover, all these sources are related to the original set of chats, since there are forwarded messages from them shared within the chats we initially collected. This level of relationship is very important for this study for the task of investigating the information propagation in this ecosystem.

Collecting messages of the groups: Finally, using the Telegram API, we try to collect messages from the sources. It's not possible to collect messages from all of the sources

Table 5.1: Overview of the Telegram dataset.

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

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 as shown in Table 5.1.

5.3 Results

This chapter presents our analysis and results.

5.3.1 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 define some key terms:

- *Forwarded message:* Any message in a chat that is forwarded into the chat using the ”Forwarding” mechanism.
- *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 5.1, 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, we find 7,894 of these source chats in our dataset. These 7,894 chats are the original source chats of 63% of all

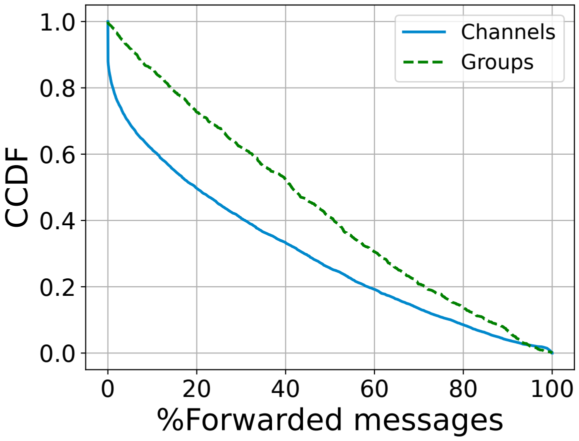


Figure 5.1: The percentage of forwarded messages in Telegram dataset.

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 into the channels and groups is shown in Fig. 5.1. 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 (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. 5.2 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

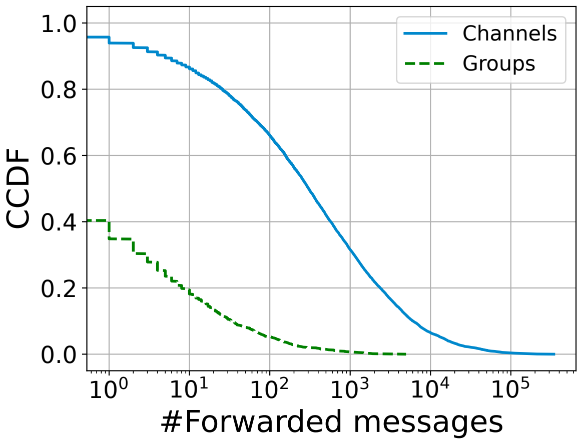


Figure 5.2: The number of forwarded messages from original messages ineach groups/channels in Telegram dataset.

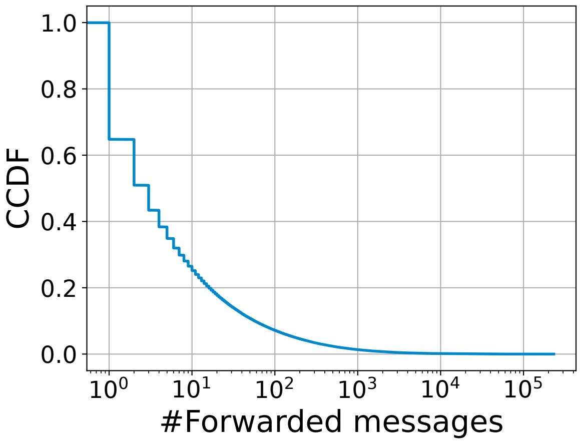


Figure 5.3: The number of forwarded messages per user in Telegram dataset.

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.

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 in the groups of our dataset. We do not have access to the number of users in each group, but the sender of the messages. Among all of the groups, there are 2 million unique users who have sent at least one message, and out of these, 225,000 users have forwarded at least one message. Fig. 5.3 shows the number of forwarded messages by each user. About 30% of the users who contribute to forwarding messages forward only a single message. Meanwhile, about 90% of them forward less than 100 messages. However, there are users who are extremely active, with several thousands of forwarded messages.

To study further the user-specific forwarding behavior, in Fig. 5.4, we plot what

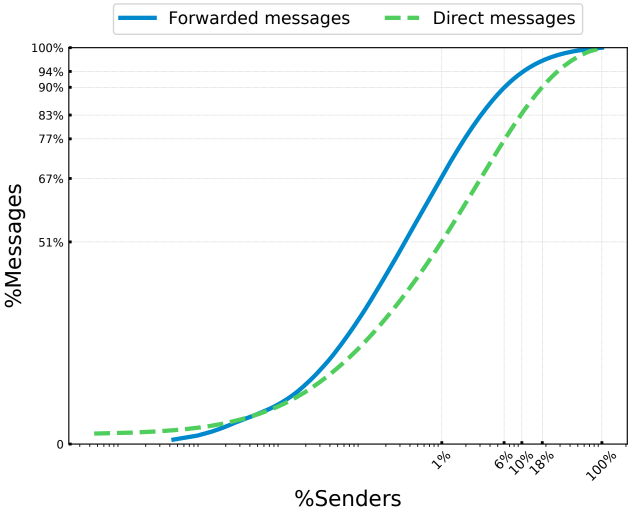


Figure 5.4: The relationship between the number of senders and shared messages in Telegram dataset.

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 online messaging platforms, it is 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: The forwarded messages in the dataset contain about 4M URLs. 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. 5.5 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 [169], as they found that URLs linking to “YouTube” and “Twitter” are the most shared URLs in upstream group chats. This shows that

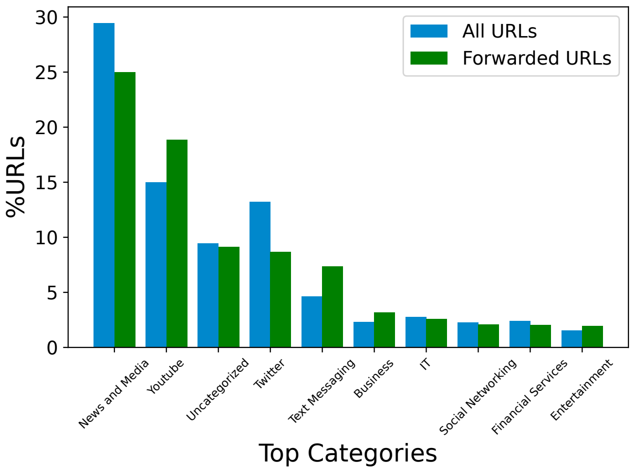


Figure 5.5: Category distribution among URLs in Telegram dataset.

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 [20] to annotate each text message. We adopt the SEVERE_TOXICITY model, as recommended in Horta Ribeiro, Jhaver, Zannettou, Blackburn, Stringhini, De Cristofaro, and West [21], 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 [149]. 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 [170]. 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

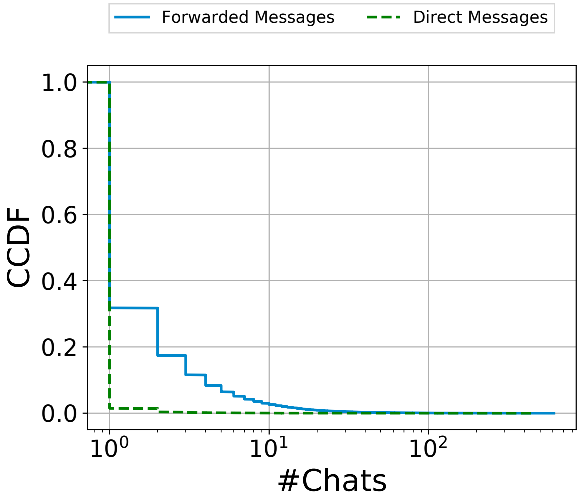


Figure 5.6: Message reach among forwarded and direct messages in Telegram dataset.

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 the same 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. 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. 5.6 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

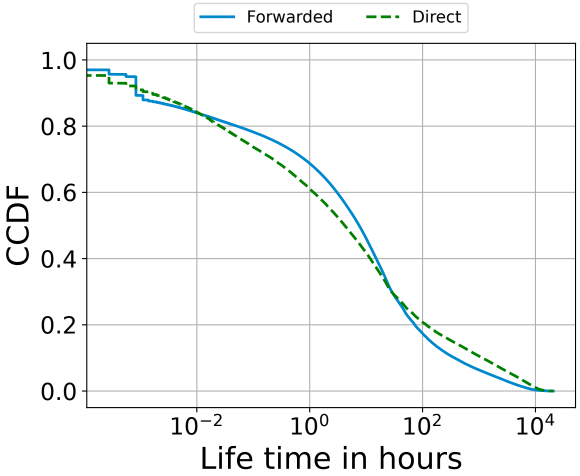


Figure 5.7: The lifetime of forward and direct messages in Telegram dataset.

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 mechanism significantly expands message reach, even though messages mainly circulate locally.

5.3.2 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. Among these 115 million text messages, there are 10 million 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 this criterion 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.

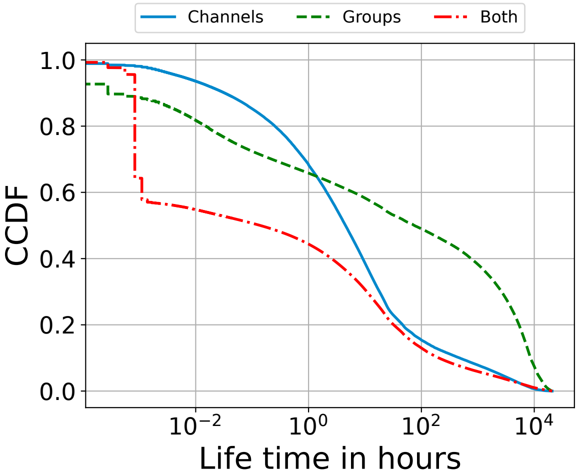


Figure 5.8: The lifetime of the messages in channels and groups in Telegram dataset.

For repeated text messages, we calculate the lifetime of the unique text messages by considering 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. In this section, we investigate how long messages from different aspects continue to be shared in the dataset during the time period of our experiment.

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. 5.7 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, compared to direct messages, forwarded messages get repeated more frequently but in a shorter period of time. 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. 5.8 shows the lifetime of

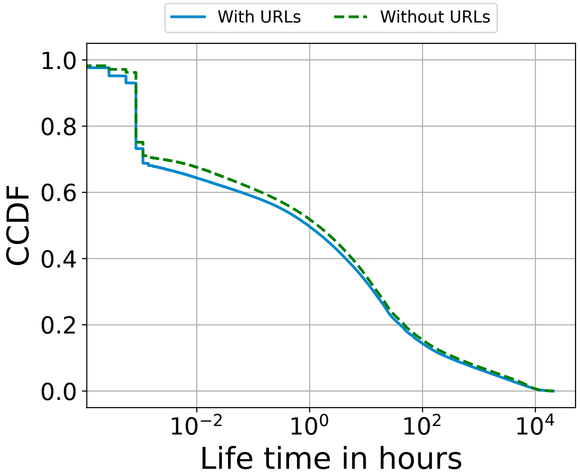


Figure 5.9: The lifetime of messages with/without URLs in Telegram dataset.

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 have a lifespan of 105 days, in contrast, 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 arise from the group's structure, in which every user has the capability to share 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. 5.9 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.

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

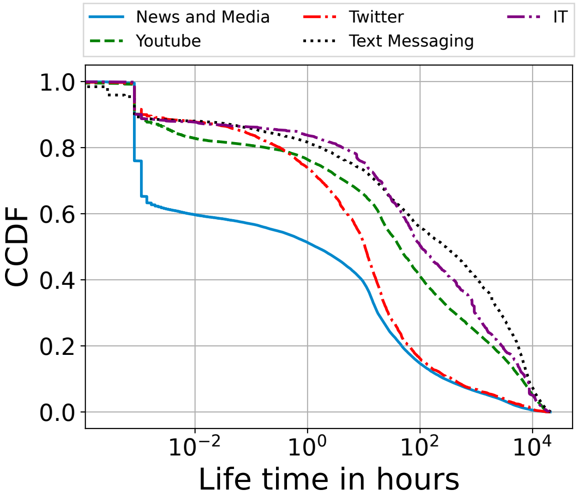


Figure 5.10: The lifetime of the URLs from top categories in Telegram dataset.

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. 5.10 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 lifetimes. 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. 5.11 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. 5.12 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

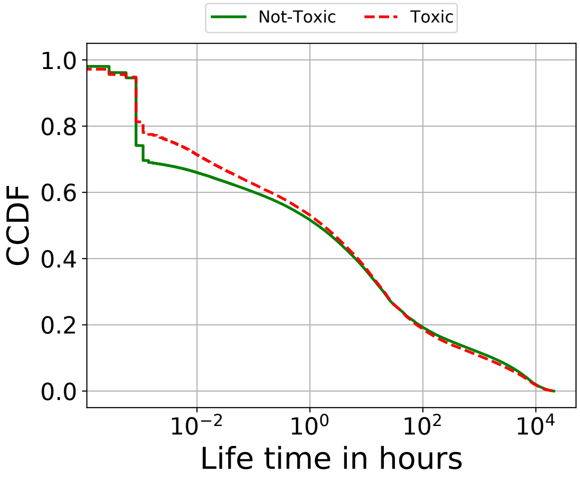


Figure 5.11: The lifetime of toxic and non-toxic messages in Telegram dataset.

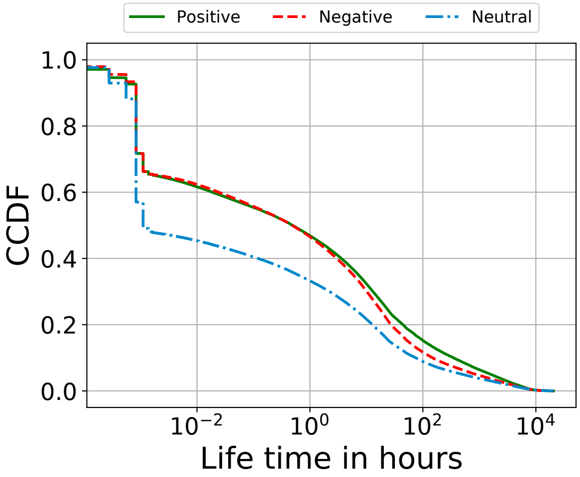


Figure 5.12: The lifetime of messages with different sentiments in Telegram dataset.

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 than direct messages, they vanish more quickly compared to repeated direct messages. Upon comparing the lifetime of messages shared in different types of chat, 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 online 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.

5.3.3 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 consists 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 [123] 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 [152]. Then, we cluster the reduced embeddings using the HDBSCAN algorithm applying the method presented in [153]. 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 for us today what the show is about. In the first chart, the facilitator tells which well-known people had a black eye and*

Table 5.2: Case studies overview.

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

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) *"Finally, the FBI arrested a Boston University professor associated with the Chinese University and Chinese Research Laboratory in Wuhan, who was highly paid by China. Now it is clear that the coronavirus is a planned bio-attack carried out by China. A Chinese expert assures everyone 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 can no longer be explained by coincidences. Only a few days after the threat of a constitutional lawsuit, the incumbent is now President of the Hamburg Hotel and Restaurant Association Franz J. Klein 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 are originally in German and we have translated them into English. These samples are promoting 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 and 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).

As we see in Table 5.2, 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

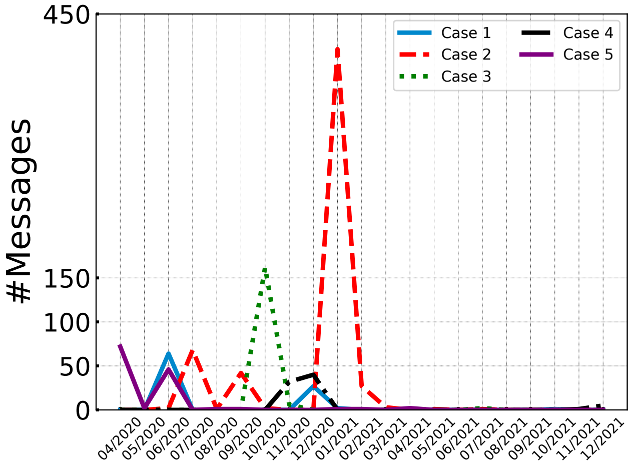


Figure 5.13: The number of appearances of each message from the case study.

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. 5.13 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.

5.4 Discussion

We performed a large-scale analysis to measure information propagation within the Telegram network, gathering a substantial dataset comprising approximately 140 million messages shared on over 9,000 public Telegram chats.

First, we tried to reveal how forwarding mechanism contribute to information dissemination within fringe communities (RQ7). Our analysis aimed to understand how Telegram users use the forwarding mechanism to propagate information across chats. 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 find that despite the localized dissemination of messages within our dataset, the forwarding mechanism plays a significant role in expanding their reach. Our results addressed the research question concerning the role of the forwarding mechanism in information dissemination within fringe communities. We find that the forwarding mechanism in Telegram has various impacts on message spread patterns, such as enabling certain users to become ”superspreaders” and increasing message reach through frequent forwarding. Additionally, our examination of a sample set of messages indicates the risk of rapid and widespread misinformation distribution due to the simplicity and quickness of the forwarding mechanism.

Next, our objective was to uncover the lifespan of various types of content shared in fringe communities (RQ8). Our analysis addressed this inquiry by revealing meaningful distinctions in the dynamics of various message types within the dataset. We find that forwarded messages, while repeated more often than direct messages, exhibit a shorter lifespan. Furthermore, messages tend to have a longer lifespan within groups compared to channels. Additionally, the lifespan of messages containing URLs varies depending on the categories of their URLs. Moreover, messages that are toxic or exhibit extreme emotions tend to last longer compared to those that are non-toxic and emotionally neutral.

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.

6 Conclusion

This dissertation illuminates the pivotal role that online messaging platforms and their respective ecosystems play in the propagation of both information and misinformation within the cyber realm. We undertook three distinct investigations, each focusing on a different aspect of these platforms: from detailing their inherent characteristics to understanding their role in globalizing and propagating conspiracy theories. In the course of this investigation, we utilized innovative methods and analyzed expansive datasets - a distinctive approach within the field that yielded fresh, insightful results. In conclusion, this dissertation has deepened our understanding of the dynamics of online messaging platforms, shedding light on their role in information propagation, and paving the way for effective strategies to mitigate the spread of harmful content. Additionally, ethical considerations and the inherent limitations of our work are discussed.

6.1 Discussions and Implications

Characterizing the Online Messaging Platforms' Ecosystem.

Online messaging platforms such as WhatsApp, Telegram, and Discord are immensely popular, connecting billions of users worldwide. Our work aimed to address the research gap in understanding the online messaging platforms ecosystem. Our extensive research on public groups across various online messaging platforms including WhatsApp, Telegram, and Discord, reveals important insights about their characteristics and significant differences. We have successfully leveraged Twitter, one of the prominent social media platforms, as a rich resource for discovering a significant number of these public groups. During a period of 38 days, we found over 350,000 unique group URLs (RQ1). Our findings underline the importance of multi-platform analysis, as focusing on a single platform could limit the depth and breadth of insights gathered. We demonstrated the ephemeral nature of the group URLs, as a significant percentage of these groups become inaccessible over a relatively short period of time, suggesting the need for robust and real-time data collection mechanisms for a more holistic view of these platforms' ecosystems. Additionally, we compared the three online messaging platforms in terms of activity, evolution, and topics to comprehensively characterize the online messaging platforms ecosystem (RQ2). Our analysis also exposed the vulnerability of personally identifiable information across these platforms. We discovered that a substantial number of phone numbers were accessible on WhatsApp. Discord, while not exposing phone numbers, did present leaks in the form of linked social media accounts. This alarming exposure of sensitive personal data on platforms that are often chosen for their perceived security underscores a pressing need for increased public awareness regarding these privacy implications. Additionally, it calls for urgent attention from the operators of these platforms to strengthen their privacy measures (RQ3).

On the Globalization of QAnon.

QAnon is a pervasive and notable conspiracy theory that has gained traction across various regions of the world, influencing individuals' perceptions and leading to tangible real-world incidents. In our subsequent analysis, we examined the development and spread of QAnon on Telegram, a thriving hub for relocated QAnon followers. Our study provided the first comprehensive, multilingual examination of QAnon's presence and activity on the Telegram platform. Drawing from a substantial dataset of 4.4 million messages posted across 161 QAnon Telegram groups and channels, our analysis unveiled an evolving global phenomenon in which QAnon's activity, influence and global reach are expanding. We observed QAnon discussions being shared among users with a variety of languages. The longevity and activity of QAnon groups surpass those of the baseline, indicating a tenacious and reactive community that can significantly impact real-world events, as evidenced by its reactions to the developments of 2021. This reveals a pressing need for real-time monitoring systems and cross-language fact-checking initiatives to track and combat the spread of such toxic content in online messaging platforms like Telegram (RQ4). We identified a notable prevalence of toxicity in QAnon discussions, especially in languages such as German and Portuguese. The observed toxicity levels surpass those found in English, suggesting a troubling diversification of QAnon's harmful influence across language barriers. Moreover, our findings regarding the potential increased hostility of QAnon supporters towards mainstream social media platforms pose implications for the entire digital ecosystem (RQ5). Furthermore, our work showed QAnon content becoming more varied in its focus, encompassing wide-ranging topics such as world politics, COVID-19, and anti-vaccination ideologies, to name a few. This highlights the evolution of the theory into a 'mega conspiracy', spreading misinformation across various pertinent social and political themes (RQ6).

Mis-information Propagation on Telegram.

We presented a comprehensive study of information propagation within the Telegram network, offering crucial insights into the significant role this platform played in contemporary communication. Through the analysis of a substantial dataset of 140 million messages shared across over 9,000 public Telegram chats, we revealed the extent of content concentration. We found that a minor proportion of users, referred to as 'superspreaders,' were responsible for the vast majority of forwarded messages. This discovery had profound implications, particularly considering the potential misuse of Telegram for distributing harmful content like hate speech or misinformation. The platform could mitigate the proliferation of such content by moderating these highly active users. We discerned distinct forwarding behavior based on the type of chat, with channels predominantly serving as the source of forwarded messages, while groups more often acted as the recipients. We also found forwarding to be a highly effective mechanism that substantially enhances the reach of messages. Furthermore, we executed a case study on five sample messages, each representing a prevalent topic among the text messages on Telegram. Our examination revealed that misinformation on this platform tends to elicit a high emotional response and frequently involves baseless claims, with forwarding serving as the principal mechanism for their propagation (RQ7). We further observed that messages shared in groups tended to have a longer lifespan compared to those shared in channels, which signified the risk of widespread and lasting misinformation due to the forwarding mechanism's efficiency. Our analysis indicated that messages displaying toxicity or emotional extremes, whether positive or negative, exhibit a prolonged lifespan compared to their non-toxic or neutral counterparts, respectively. Remarkably, our case study reveals that the sample messages displayed considerable longevity, often experiencing a resurgence in popularity in response to related real-world events (RQ8). This study underscored the significant issue of misinformation and conspiracy theory dissemination on Telegram, accentuating the pressing necessity for further research and strategic interventions to curb the proliferation of such harmful content.

6.2 Limitations

Naturally, our work has some limitations. In Chapter 3, we rely solely on Twitter to discover groups from WhatsApp, Telegram, and Discord, hence we are unaware of a large number of publicly available groups. Despite this fact, Twitter is a very large and mainstream social network that we use to make our best effort attempt to discover a large number of groups from WhatsApp, Telegram, and Discord, and mitigate potential biases. Another limitation arises from the fact that we join and collect data from only a limited number of groups from WhatsApp, Telegram, and Discord, mainly because these online messaging platforms have specific constraints that prevent us from scaling up our data collection. Namely, WhatsApp requires a large number of mobile phones and SIM cards, Discord requires the creation of multiple user accounts, and Telegram's API is rate-limited. Note that this is a limitation that exists in every study that collects data from online messaging platforms. Additionally, the use of Twitter as the only data source for discovering public groups of the different online messaging platforms potentially introduces some bias in our sample. Where applicable, we clearly state the implications of sample bias for inferences and also provide a control dataset to facilitate an accurate interpretation of our results. We make the best effort to mitigate potential biases that might affect our findings.

In Chapter 4, the QAnon data collection and dataset have some limitations. First, as with all studies that focus on online messaging platforms like Telegram and WhatsApp [80, 171], we cannot assess how representative our collected dataset is. This is because there is no single vantage point to discover all Telegram groups/channels; due to this, we focus only on groups/channels shared on Twitter and Facebook. Therefore, we likely miss QAnon groups/channels simply because they were not shared on Twitter or Facebook. Second, the dataset is biased towards more recent groups/channels that were active in 2020. Hence, we likely miss some groups/channels that were created before 2020 and eventually became inactive. Finally, our keyword filtering is based on just two keywords, which indicates that we initially miss QAnon groups/channels that do not use these keywords. We mitigated this by expanding our dataset based on forwarded messages.

In Chapter 5, the Telegram dataset has also some important limitations that are worth mentioning. First, we are unable to assess the representativeness of our dataset, mainly because there is no way to extract a random sample of chats from Telegram, hence 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, hence our dataset is likely to be biased towards chats involved in the dissemination of fringe ideologies. It's 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 can not be generalized to the entire Telegram network. We believe that this is an inherent limitation that exists in almost all the studies focusing on online messaging platforms like Telegram, mainly because there is no vantage point to obtain holistic or representative samples of Telegram chats. Also, given that we collect the messages after joining the groups/channels we miss messages that are already deleted by the users. Nevertheless, we can confirm that by expanding the dataset based on forwarded messages, we collect a large dataset that includes many mainstream chats.

6.3 Ethical Considerations

We submitted our methodology to our institution's ethical review board and obtained approval prior to collecting any data. We emphasize that (a) we work only with publicly available data; (b) we do not store users' phone numbers as such, but use one-way hashes of such data; (c) we do not attempt to de-anonymize users from any personally identifiable information; and (d) do not attempt to link users across platforms. Overall, we follow standard ethical guidelines [172] throughout our data collection and analysis. Note, that all data obtained during this project, tweets, Facebook posts, invite URLs, and messages shared inside online messaging platforms, are publicly available data.

6.4 Future Work

Interplay among online messaging platforms: In this dissertation, we characterized three online messaging platforms and investigated the evolution and propagation of misinformation within one of them. However, we have not assessed misinformation propagation across multiple platforms, and we lack an understanding of how each platform is influenced and fed by others, as well as how each platform influences and provides content for others. A comprehensive exploration of the interconnection among online messaging platforms could be the focus of future research. This entails delving into the cross-platform virality of misinformation and the potential for misinformation originating on one platform to go viral through others. Uncovering how misinformation traverses across multiple platforms and goes viral is crucial for understanding the mechanisms of misinformation dissemination and the interconnected nature of the online messaging ecosystem.

Lifespan of Messages in the Ecosystem: While we examined the lifespan of messages within Telegram, we do not know if the messages have circulated in other platforms before reaching Telegram or if they will continue to be shared in other platforms after vanishing from Telegram. A focused analysis of the lifespan of messages across platforms could be an area to explore in future studies, considering factors such as message content, user engagement, and community dynamics.

Investigation of Images and Videos: While conducting content analysis in this thesis, our focus was solely on text messages, leading to a lack of knowledge about the information contained in image and video messages. Given the substantial presence of images and videos on online messaging platforms, there is a notable opportunity for future studies to explore the analysis of these media types, providing a more comprehensive understanding of content dynamics.

Role of each Platform in Misinformation Dissemination: In our study, we identified specific groups as primary contributors to a significant volume of messages related to conspiratorial content. The exploration of source-destination relationships among platforms through cross-platform experiments can illuminate the distinct contributions of each platform in producing and disseminating misinformation. Subsequent research in this realm has the potential to unravel the diverse roles and influences of different platforms.

Coordinated Misinformation Campaigns: Our investigation brought to light the prevalence of false news shared by groups of users. Uncovering coordinated networks of users and understanding their strategies for spreading particular types of misinformation is invaluable. Future research should consider analyzing the forwarding feature's role in amplifying the impact of coordinated campaigns, providing deeper insights into the mechanisms at play.

Automated Activities: Our observations revealed accounts rapidly disseminating a substantial number of messages within a short period. However, our study did not delve into the realm of automation within communities. Future research could explore the detection of bots or automated mechanisms spreading misinformation across various platforms, unveiling potential relationships between them.

Relationship between Real-World Events and Online Activities: Our study identified coincidences between real-world events and heightened activity within groups. A crucial topic for future research involves examining the correlation between real-world events, especially harmful actions, and the online activities contributing to the spread of misinformation. Conducting a cross-platform study can provide a more comprehensive and insightful analysis of these complex dynamics.

Privacy Leakage in Advertisement: Our findings revealed a concerning amount of personal information from users exposed in public groups, raising concerns about the unauthorized usage of users' information for advertisements. It is suggested that future research delves into examining the prevalence of targeted ads on online messaging platforms and their affiliated networks. This investigation should assess potential privacy violations associated with the use of users' Personally Identifiable Information (PII), both within individual platforms and across interconnected networks.

6.5 Summary

We provided a comprehensive exploration of online messaging platforms' role in the spread of information and misinformation. It is structured around three distinct investigations: characterizing the online messaging platforms' ecosystem, the globalization of the QAnon conspiracy theory, and studying misinformation propagation on Telegram. We initiated our research with a meta-analysis of three prominent online messaging platforms, developing a unique methodology to identify a substantial number of their public groups via the lens of Twitter. This essential first step set the stage for our subsequent investigations. Next, we explored the global evolution of a particular conspiracy theory, examining its spread amongst diverse global communities. Armed with a deeper understanding of how conspiracy theories evolve and globalize, we shifted our focus to the dynamics of message propagation within these fringe communities. We analyzed the frequency and virality of content dissemination, emphasizing the critical role of forwarding in the widespread distribution of conspiratorial content. In conclusion, this dissertation has deepened our understanding of the dynamics of online messaging platforms, shedding light on their role in information propagation, and paving the way for effective strategies to mitigate the spread of harmful content.

Bibliography

[1] Mohamad Hoseini, Philipe Melo, Manoel Júnior, Fabrício Benevenuto, Balakrishnan Chandrasekaran, Anja Feldmann, and Savvas Zannettou. "Demystifying the Messaging Platforms' Ecosystem Through the Lens of Twitter". In: *The 2020 Internet Measurement Conference (IMC)*. 2020, pp. 345–359 (cit. on pp. ix, 11).

[2] Mohamad Hoseini, Philipe Melo, Fabricio Benevenuto, Anja Feldmann, and Savvas Zannettou. "On the globalization of the QAnon conspiracy theory through Telegram". In: *Proceedings of the 15th ACM Web Science Conference*. 2023, pp. 75–85 (cit. on pp. ix, 11, 18).

[3] *Number of internet and social media users worldwide*. `https://www.statista.com/statistics/617136/digital-population-worldwide/`. Accessed: 2024-04-05 (cit. on p. 1).

[4] Andrew J Flanagin. "IM online: Instant messaging use among college students". In: *Communication Research Reports* 22.3 (2005), pp. 175–187 (cit. on p. 1).

[5] Shakuntala Banaji, Ramnath Bhat, Anushi Agarwal, Nihal Passanha, and Mukti Sadhana Pravin. "WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India". In: *London School of Economics and Political Science* (2019) (cit. on p. 2).

[6] Matteo Vergani, Alfonso Martinez Arranz, Ryan Scrivens, and Liliana Orellana. "Hate speech in a telegram conspiracy channel during the first year of the COVID-19 pandemic". In: *Social Media+ Society* 8.4 (2022) (cit. on p. 2).

[7] Stijn Peeters and Tom Willaert. "Telegram and digital methods: Mapping networked conspiracy theories through platform affordances". In: *M/C Journal* 25.1 (2022) (cit. on p. 2).

[8] Ernesto Londono, Flávia Milhorance, and Jack Nicas. *Brazil's Far-Right Disinformation Pushers Find a Safe Space on Telegram*. 2021 (cit. on pp. 2, 69).

[9] Gustavo Resende, Philipe Melo, Julio C. S. Reis, Marisa Vasconcelos, Jussara M. Almeida, and Fabrício Benevenuto. "Analyzing Textual (Mis)Information Shared in WhatsApp Groups". In: *Proceedings of the 10th ACM Conference on Web Science*. 2019, pp. 225–234 (cit. on pp. 2, 17, 19, 21, 69).

[10] Faiz Siddiqui and Susan Svrluga. *NC man told police he went to DC pizzeria with gun to investigate conspiracy theory*. 2016 (cit. on p. 4).

[11] Julio CS Reis and Fabrício Benevenuto. "Supervised Learning for Misinformation Detection in Whatsapp". In: *Proceedings of the Brazilian Symposium on Multimedia and the Web*. 2021, pp. 245–252 (cit. on p. 4).

[12] Julio CS Reis, Philipe Melo, Kiran Garimella, and Fabrício Benevenuto. "Can WhatsApp Benefit from Debunked Fact-Checked Stories to Reduce Misinformation?" In: *Harvard Kennedy School (HKS) Misinformation Review* (2020) (cit. on pp. 4, 17).

[13] Julio CS Reis, Philipe Melo, Kiran Garimella, and Fabrıcio Benevenuto. "Detecting misinformation on WhatsApp without breaking encryption". In: *Association for the Advancement of Artificial Intelligence* (2020) (cit. on p. 4).

[14] Nahiyan Bin Noor, Niloofar Yousefi, Billy Spann, and Nitin Agarwal. "Comparing Toxicity Across Social Media Platforms for COVID-19 Discourse". In: *arXiv preprint arXiv:2302.14270* (2023) (cit. on p. 4).

[15] Anni Sternisko, Aleksandra Cichocka, and Jay J Van Bavel. "The dark side of social movements: Social identity, non-conformity, and the lure of conspiracy theories". In: *Current opinion in psychology* 35 (2020), pp. 1–6 (cit. on p. 4).

[16] Bettina Rottweiler and Paul Gill. "Conspiracy beliefs and violent extremist intentions: The contingent effects of self-efficacy, self-control and law-related morality". In: *Terrorism and Political Violence* 34.7 (2022), pp. 1485–1504 (cit. on p. 4).

[17] Ted Goertzel. "Belief in conspiracy theories". In: *Political psychology* (1994), pp. 731–742 (cit. on p. 5).

[18] Punyajoy Saha, Binny Mathew, Kiran Garimella, and Animesh Mukherjee. ""Short is the Road that Leads from Fear to Hate": Fear Speech in Indian WhatsApp Groups". In: *Proceedings of the Web conference.* 2021, pp. 1110–1121 (cit. on p. 5).

[19] R Tallal Javed, Mirza Elaaf Shuja, Muhammad Usama, Junaid Qadir, Waleed Iqbal, Gareth Tyson, Ignacio Castro, and Kiran Garimella. "A first look at COVID-19 messages on WhatsApp in Pakistan". In: *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.* 2020, pp. 118–125 (cit. on p. 5).

[20] Perspective API. `https://www.perspectiveapi.com/`. Accessed: 2023-06-05. 2018 (cit. on pp. 7, 44, 53, 78).

[21] Manoel Horta Ribeiro, Shagun Jhaver, Savvas Zannettou, Jeremy Blackburn, Gianluca Stringhini, Emiliano De Cristofaro, and Robert West. "Do platform migrations compromise content moderation? evidence from r/the_donald and r/incels". In: *Proceedings of the ACM on Human-Computer Interaction* 5 (2021), pp. 1–24 (cit. on pp. 7, 53, 66, 69, 78).

[22] Statista. *Most popular global mobile messenger apps as of January 2024, based on number of monthly active users.* `https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/`. Accessed: 2024-04-05. 2024 (cit. on p. 13).

[23] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, DA Levy, and Rasmus Kleis Nielsen. *Reuters institute digital news report 2018. Reuters institute for the study of journalism.* 2019 (cit. on p. 13).

[24] Tardaguila, Cristina and Benevenuto, Fabricio and Ortellado, Pablo. *Fake News Is Poisoning Brazilian Politics. WhatsApp Can Stop It.* `https://www.nytimes.com/2018/10/17/opinion/brazil-election-fake-news-whatsapp.html`. Accessed: 2023-07-04. 2018 (cit. on p. 14).

[25] Bassi, Simi and Sengupta, Joyita. *Lynchings sparked by WhatsApp child-kidnap rumours sweep across India.* `https://www.cbc.ca/news/world/india-child-kidnap-abduction-video-rumours-killings-1.4737041`. Accessed: 2023-07-04. 2018 (cit. on p. 14).

[26] Telegram. *400 Million Users, 20,000 Stickers, Quizzes 2.0 and €400K for Creators of Educational Tests.* `https://telegram.org/blog/400-million`. Accessed: 2023-07-04. 2020 (cit. on p. 14).

[27] Anti-Defamation League. “Telegram: The Latest Safe Haven for White Supremacists”. In: *ADL. December* 2 (2019) (cit. on pp. 14, 47).

[28] Rebecca Tan. *Terrorists’ love for Telegram, explained.* `https://www.vox.com/world/2017/6/30/15886506/terrorism-isis-telegram-social-media-russia-pavel-durov-twitter`. Accessed: 2023-07-04. 2017 (cit. on pp. 14, 47).

[29] Discord. *How to use Discord for your classroom.* `https://blog.discord.com/how-to-use-discord-for-your-classroom-8587bf78e6c4`. Accessed: 2023-07-04. 2020 (cit. on p. 14).

[30] Kevin Roose. *This Was the Alt-Right’s Favorite Chat App. Then Came Charlottesville.* `https://www.nytimes.com/2017/08/15/technology/discord-chat-app-alt-right.html`. Accessed: 2023-07-04. 2017 (cit. on p. 15).

[31] Joseph Cox. *The Gaming Site Discord Is the New Front of Revenge Porn.* `https://www.thedailybeast.com/the-gaming-site-discord-is-the-new-front-of-revenge-porn`. Accessed: 2023-07-05. 2018 (cit. on p. 15).

[32] Adam M Ostrovsky and Joshua R Chen. “TikTok and its role in COVID-19 information propagation”. In: *Journal of adolescent health* 67.5 (2020), p. 730 (cit. on p. 17).

[33] Daejin Choi, Selin Chun, Hyunchul Oh, Jinyoung Han, Ted Kwon, et al. “Rumor propagation is amplified by echo chambers in social media”. In: *Scientific reports* 10.1 (2020), pp. 1–10 (cit. on p. 17).

[34] Soroush Vosoughi, Deb Roy, and Sinan Aral. “The spread of true and false news online”. In: *science* 359.6380 (2018), pp. 1146–1151 (cit. on p. 17).

[35] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. “What is Twitter, a social network or a news media?” In: *Proceedings of the 19th international conference on World wide web.* 2010, pp. 591–600 (cit. on p. 17).

[36] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna Gummadi. “Measuring user influence in twitter: The million follower fallacy”. In: *Proceedings of the international AAAI conference on web and social media.* Vol. 4. 1. 2010, pp. 10–17 (cit. on p. 17).

[37] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Rosenquist. “Understanding the demographics of Twitter users”. In: *Proceedings of the international AAAI conference on web and social media*. Vol. 5. 1. 2011, pp. 554–557 (cit. on p. 17).

[38] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. “I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system”. In: *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. 2007, pp. 1–14 (cit. on p. 17).

[39] Flavio Figueiredo, Fabrício Benevenuto, and Jussara M Almeida. “The tube over time: characterizing popularity growth of youtube videos”. In: *Proceedings of the fourth ACM international conference on Web search and data mining*. 2011, pp. 745–754 (cit. on p. 17).

[40] Eric Gilbert. “Widespread underprovision on Reddit”. In: *Proceedings of the 2013 conference on Computer supported cooperative work*. 2013, pp. 803–808 (cit. on p. 17).

[41] Philipp Singer, Fabian Flöck, Clemens Meinhart, Elias Zeitfogel, and Markus Strohmaier. “Evolution of reddit: from the front page of the internet to a self-referential community?” In: *Proceedings of the 23rd international conference on world wide web*. 2014, pp. 517–522 (cit. on p. 17).

[42] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. “The pushshift reddit dataset”. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 14. 2020, pp. 830–839 (cit. on p. 17).

[43] Meeyoung Cha, Alan Mislove, and Krishna P Gummadi. “A measurement-driven analysis of information propagation in the flickr social network”. In: *Proceedings of the 18th international conference on World wide web*. 2009, pp. 721–730 (cit. on p. 17).

[44] Alan Mislove, Hema Swetha Koppula, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. “Growth of the flickr social network”. In: *Proceedings of the first workshop on Online social networks*. 2008, pp. 25–30 (cit. on p. 17).

[45] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. “On the evolution of user interaction in facebook”. In: *Proceedings of the 2nd ACM workshop on Online social networks*. 2009, pp. 37–42 (cit. on p. 17).

[46] Yabing Liu, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. “Analyzing facebook privacy settings: user expectations vs. reality”. In: *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 2011, pp. 61–70 (cit. on p. 17).

[47] Gabriel Hine, Jeremiah Onaolapo, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Riginos Samaras, Gianluca Stringhini, and Jeremy Blackburn. “Kek, cucks, and god emperor trump: A measurement study of 4chan’s politically incorrect forum and its effects on the web”. In: *Proceedings of the*

International AAAI Conference on Web and Social Media. Vol. 11. 1. 2017, pp. 92–101 (cit. on p. 17).

[48] Michael Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Greg Vargas. “4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community”. In: *Proceedings of the international AAAI conference on web and social media*. Vol. 5. 1. 2011, pp. 50–57 (cit. on p. 17).

[49] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. “What is gab: A bastion of free speech or an alt-right echo chamber”. In: *Companion Proceedings of the The Web Conference*. 2018, pp. 1007–1014 (cit. on pp. 17, 46, 69).

[50] Lucas Lima, Julio CS Reis, Philipe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto. “Inside the Right-Leaning Echo Chambers: Characterizing Gab, an Unmoderated Social System”. In: *2018 ieee/acm international conference on advances in social networks analysis and mining*. 2018, pp. 515–522 (cit. on pp. 17, 46, 69).

[51] Aravindh Raman, Sagar Joglekar, Emiliano De Cristofaro, Nishanth Sastry, and Gareth Tyson. “Challenges in the decentralised web: The mastodon case”. In: *Proceedings of the Internet Measurement Conference*. 2019, pp. 217–229 (cit. on p. 17).

[52] Alan Mislove, Massimiliano Marcon, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. “Measurement and analysis of online social networks”. In: *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. 2007, pp. 29–42 (cit. on p. 17).

[53] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtelris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. “The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources”. In: *Proceedings of the 2017 Internet Measurement Conference*. 2017, pp. 405–417 (cit. on pp. 17, 19, 40).

[54] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. “On the Origins of Memes by Means of Fringe Web Communities”. In: *Proceedings of the Internet Measurement Conference*. 2018, pp. 188–202 (cit. on pp. 17, 19, 40).

[55] Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, and Eric Gilbert. “The bag of communities: Identifying abusive behavior online with preexisting internet data”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 3175–3187 (cit. on p. 17).

[56] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. "(Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures". In: *The World Wide Web Conference*. 2019, pp. 818–828 (cit. on pp. 17, 19, 21).

[57] Philipe Melo, Carolina Coimbra Vieira, Kiran Garimella, Pedro OS Vaz de Melo, and Fabrício Benevenuto. "Can WhatsApp Counter Misinformation by Limiting Message Forwarding?" In: *International Conference on Complex Networks and Their Applications*. 2019, pp. 372–384 (cit. on pp. 17, 70).

[58] Chinmayi Arun. "On WhatsApp, rumours, lynchings, and the Indian Government". In: *Economic & Political Weekly* 54.6 (2019) (cit. on pp. 17, 19).

[59] Feeza Vasudeva and Nicholas Barkdull. "WhatsApp in India? A case study of social media related lynchings". In: *Social Identities* 26.5 (2020), pp. 574–589 (cit. on p. 17).

[60] Pranav Malhotra. "A Relationship-Centered and Culturally Informed Approach to Studying Misinformation on COVID-19". In: *Social Media and Society* 6.3 (2020), pp. 1–4 (cit. on p. 17).

[61] R Tallal Javed, Muhammad Usama, Waleed Iqbal, Junaid Qadir, Gareth Tyson, Ignacio Castro, and Kiran Garimella. "A deep dive into COVID-19-related messages on WhatsApp in Pakistan". In: *Social Network Analysis and Mining* 12.1 (2022), pp. 1–16 (cit. on p. 17).

[62] Carlos Elias and Daniel Catalan-Matamoros. "Coronavirus in Spain: Fear of 'Official' Fake News Boosts WhatsApp and Alternative Sources". In: *Media and Communication* 8.2 (2020), pp. 462–466 (cit. on p. 17).

[63] Jeremy Bowles, Horacio Larreguy, and Shelley Liu. "Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe". In: *PLOS ONE* 15.10 (2020), pp. 1–11 (cit. on p. 17).

[64] Antônio Diogo Forte Martins, Lucas Cabral, Pedro Jorge Chaves Mourão, José Maria Monteiro, and Javam Machado. "Detection of misinformation about covid-19 in brazilian portuguese whatsapp messages". In: *International Conference on Applications of Natural Language to Information Systems*. 2021, pp. 199–206 (cit. on p. 17).

[65] Santosh Vijaykumar, Yan Jin, Daniel Rogerson, Xuerong Lu, Swati Sharma, Anna Maughan, Bianca Fadel, Mariella Silva de Oliveira Costa, Claudia Pagliari, and Daniel Morris. "How shades of truth and age affect responses to COVID-19 (Mis) information: randomized survey experiment among WhatsApp users in UK and Brazil". In: *Humanities and Social Sciences Communications* 8.1 (2021), pp. 1–12 (cit. on p. 17).

[66] Rama Adithya Varanasi, Joyojeet Pal, and Aditya Vashistha. "Accost, Accede, or Amplify: Attitudes towards COVID-19 Misinformation on WhatsApp in India". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–17 (cit. on p. 17).

[67] Nic Cheeseman, Jonathan Fisher, Idayat Hassan, and Jamie Hitchen. "Social Media Disruption: Nigeria's WhatsApp Politics". In: *Journal of Democracy* 31.3 (2020), pp. 145–159 (cit. on p. 17).

[68] Jamie Hitchen, Jonathan Fisher, Idayat Hassan, and Nic Cheeseman. "Whatsapp and Nigeria's 2019 Elections: Mobilising the People, Protecting the Vote". In: *Portal Africa* (2019). Accessed: 2023-07-05. URL: `https://www.africaportal.org/publications/whatsapp-and-nigerias-2019-elections-mobilising-people-protecting-vote/` (cit. on p. 17).

[69] Ashkan Kazemi, Kiran Garimella, Gautam Kishore Shahi, Devin Gaffney, and Scott A Hale. "Research note: Tiplines to uncover misinformation on encrypted platforms: A case study of the 2019 Indian general election on WhatsApp". In: *Harvard Kennedy School (HKS) Misinformation Review* (2022) (cit. on p. 17).

[70] Veronika Solopova, Tatjana Scheffler, and Mihaela Popa-Wyatt. "A Telegram corpus for hate speech, offensive language, and online harm". In: *Journal of Open Humanities Data* 7 (2021) (cit. on pp. 18, 71).

[71] Stijn Peeters and Tom Willaert. "Telegram and Digital Methods: Mapping Networked Conspiracy Theories through Platform Affordances". In: *M/C Journal* 25.1 (2022) (cit. on p. 18).

[72] S Shyam Sundar, Maria D Molina, and Eugene Cho. "Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps?" In: *Journal of Computer-Mediated Communication* 26.6 (2021), pp. 301–319 (cit. on p. 18).

[73] Philipe Melo, Carolina Coimbra Vieira, Kiran Garimella, Pedro OS de Melo, and Fabrício Benevenuto. "Can WhatsApp Counter Misinformation by Limiting Message Forwarding?" In: *Proc. of the Int'l Conference on Complex Networks and their Applications.* 2019 (cit. on pp. 19, 21).

[74] Philipe Melo, Johnnatan Messias, Gustavo Resende, Kiran Garimella, Jussara Almeida, and Fabrício Benevenuto. "Whatsapp Monitor: A Fact-Checking System for Whatsapp". In: *Proceedings of the International AAAI Conference on Web and Social Media.* Vol. 13. 2019, pp. 676–677 (cit. on pp. 19, 21, 66).

[75] Caio Machado, Beatriz Kira, Vidya Narayanan, Bence Kollanyi, and Philip Howard. "A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections." In: *Companion proceedings of the 2019 World Wide Web conference.* 2019, pp. 1013–1019 (cit. on pp. 19, 21).

[76] Victor S Bursztyn and Larry Birnbaum. "Thousands of Small, Constant Rallies: A Large-Scale Analysis of Partisan WhatsApp Groups". In: *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.* 2019 (cit. on pp. 19, 21).

[77] Josemar Alves Caetano, Gabriel Magno, Marcos Gonçalves, Jussara Almeida, Humberto T. Marques-Neto, and Virgílio Almeida. "Characterizing Attention Cascades in WhatsApp Groups". In: *Proceedings of the 10th ACM Conference on Web Science.* 2019, pp. 27–36 (cit. on pp. 19, 21).

[78] Avi Rosenfeld, Sigal Sina, David Sarne, Or Avidov, and Sarit Kraus. "WhatsApp usage patterns and prediction models". In: *ICWSM/IUSSP Workshop on Social Media and Demographic Research.* 2016 (cit. on p. 21).

[79] Kiran Garimella and Gareth Tyson. "Whatapp doc? a first look at whatsapp public group data". In: *Proceedings of the international AAAI conference on web and social media.* Vol. 12. 1. 2018 (cit. on p. 21).

[80] Julio CS Reis, Philipe Melo, Kiran Garimella, Jussara M Almeida, Dean Eckles, and Fabrício Benevenuto. "A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections". In: *Proceedings of the international AAAI conference on web and social media.* Vol. 14. 2020, pp. 903–908 (cit. on pp. 21, 91).

[81] Kiran Garimella and Dean Eckles. "Image based Misinformation on WhatsApp". In: *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media.* 2017 (cit. on p. 21).

[82] Andrés Moreno, Philip Garrison, and Karthik Bhat. "Whatsapp for Monitoring and Response During Critical Events: Aggie in the Ghana 2016 Election". In: *14th International Conference on Information Systems for Crisis Response and Management.* 2017 (cit. on p. 21).

[83] Alexandre Maros, Jussara Almeida, Fabrício Benevenuto, and Marisa Vasconcelos. "Analyzing the use of audio messages in Whatsapp groups". In: *Proceedings of the web conference.* 2020, pp. 3005–3011 (cit. on p. 21).

[84] Jason Baumgartner, Savvas Zannettou, Megan Squire, and Jeremy Blackburn. "The pushshift telegram dataset". In: *Proceedings of the international AAAI conference on web and social media.* Vol. 14. 2020, pp. 840–847 (cit. on p. 21).

[85] Cosimo Anglano, Massimo Canonico, and Marco Guazzone. "Forensic analysis of telegram messenger on android smartphones". In: *Digital Investigation* 23 (2017), pp. 31–49 (cit. on p. 21).

[86] Gandeva Bayu Satrya, Philip Tobianto Daely, and Muhammad Arif Nugroho. "Digital forensic analysis of Telegram Messenger on Android devices". In: *International Conference on Information & Communication Technology and Systems.* 2016, pp. 1–7 (cit. on p. 21).

[87] Ruba Abu-Salma, Kat Krol, Simon Parkin, Victoria Koh, Kevin Kwan, Jazib Mahboob, Zahra Traboulsi, and M Angela Sasse. "The Security Blanket of the Chat World: An Analytic Evaluation and a User Study of Telegram". In: Internet Society. 2017 (cit. on pp. 21, 72).

[88] Sarah Nikkhah, Andrew D Miller, and Alyson L Young. "Telegram as an immigration management tool". In: *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing.* 2018, pp. 345–348 (cit. on pp. 21, 72).

[89] Ali Hashemi and Mohammad Ali Zare Chahooki. "Telegram group quality measurement by user behavior analysis". In: *Social Network Analysis and Mining* 9.1 (2019), p. 33 (cit. on pp. 21, 72).

[90] Amir Reza Asnafi, Shima Moradi, Mohadeseh Dokhtesmati, and Maryam Pakdaman Naeini. “Using mobile-based social networks by Iranian libraries: The case of Telegram Messenger”. In: *Libr. Philos. Pract* 2017.1 (2017) (cit. on p. 21).

[91] Azadeh Akbari and Rashid Gabdulhakov. “Platform surveillance and resistance in Iran and Russia: The case of Telegram”. In: *Surveillance and Society* 17.1/2 (2019) (cit. on p. 21).

[92] Arash Dargahi Nobari, Negar Reshadatmand, and Mahmood Neshati. “Analysis of Telegram, An Instant Messaging Service”. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* 2017, pp. 2035–2038 (cit. on pp. 22, 72).

[93] Mohammad Naseri and Hamed Zamani. “Analyzing and Predicting News Popularity in an Instant Messaging Service”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 2019, pp. 1053–1056 (cit. on pp. 22, 72).

[94] Nico Prucha. “IS and the Jihadist Information Highway–Projecting Influence and Religious Identity via Telegram”. In: *Perspectives on Terrorism* 10.6 (2016) (cit. on pp. 22, 72).

[95] Ahmet S Yayla and Anne Speckhard. “Telegram: The mighty application that ISIS loves”. In: *International Center for the Study of Violent Extremism* (2017) (cit. on pp. 22, 72).

[96] Ahmad Shehabat, Teodor Mitew, and Yahia Alzoubi. “Encrypted jihad: Investigating the role of Telegram App in lone wolf attacks in the West”. In: *Journal of Strategic Security* 10.3 (2017), pp. 27–53 (cit. on pp. 22, 72).

[97] JT Hamrick, Farhang Rouhi, Arghya Mukherjee, Amir Feder, Neil Gandal, Tyler Moore, and Marie Vasek. “An examination of the cryptocurrency pump-and-dump ecosystem”. In: *Information Processing & Management* 58.4 (2021), p. 102506 (cit. on pp. 22, 72).

[98] Lisa Lacher and Cydnee Biehl. “Using discord to understand and moderate collaboration and teamwork”. In: *Proceedings of the 49th ACM Technical Symposium on Computer Science Education.* 2018, pp. 1107–1107 (cit. on p. 22).

[99] Jialun Aaron Jiang, Charles Kiene, Skyler Middler, Jed R Brubaker, and Casey Fiesler. “Moderation Challenges in Voice-based Online Communities on Discord”. In: *Proceedings of the ACM on Human-Computer Interaction* (2019), pp. 1–23 (cit. on p. 22).

[100] Charles Kiene and Benjamin Mako Hill. “Who Uses Bots? A Statistical Analysis of Bot Usage in Moderation Teams”. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* 2020, pp. 1–8 (cit. on p. 22).

[101] Twitter. *Twitter’s Search API.* `https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets`. Accessed: 2023-07-05. 2020 (cit. on p. 23).

[102] Twitter. *Twitter's Streaming API.* https://developer.twitter.com/en/docs/tweets/filter-realtime/overview. Accessed: 2023-07-05. 2020 (cit. on p. 23).

[103] Discord. https://support.discord.com/hc/en-us/articles/208866998-Invites-101. Accessed: 2023-07-06. 2020 (cit. on p. 23).

[104] Discord. *Discord API.* https://discord.com/developers/docs/resources/guild. Accessed: 2023-07-05. 2020 (cit. on p. 24).

[105] *WhatsApp Wrapper.* https://github.com/mukulhase/WebWhatsapp-Wrapper. Accessed: 2023-07-05. 2018 (cit. on p. 25).

[106] Telegram. *Telegram API.* https://core.telegram.org/method/channels.joinChannel. Accessed: 2023-07-04. 2020 (cit. on pp. 25, 48).

[107] Discord. *Discord OAuth API.* https://discord.com/developers/docs/topics/oauth2. Accessed: 2023-07-04. 2020 (cit. on p. 25).

[108] David M Blei, Andrew Y Ng, and Michael I Jordan. "Latent dirichlet allocation". In: *Journal of machine Learning research* (2003), pp. 993–1022 (cit. on p. 30).

[109] Brian Dean. *WhatsApp User Statistics 2024: How Many People Use WhatsApp?* https://backlinko.com/whatsapp-users. Accessed: 2024-04-05. 2023 (cit. on p. 33).

[110] Statista. *Leading countries based on number of X (formerly Twitter) users as of January 2024.* https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/. Accessed: 2024-04-05. 2024 (cit. on p. 33).

[111] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. "Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web". In: *Companion proceedings of the 2019 world wide web conference.* 2019, pp. 218–226 (cit. on p. 39).

[112] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. "Who let the trolls out? towards understanding state-sponsored trolls". In: *Proceedings of the 10th acm conference on web science.* 2019, pp. 353–362 (cit. on p. 39).

[113] Julia Carrie Wong. *What is QAnon? Explaining the bizarre rightwing conspiracy theory.* https://bit.ly/3tCCzA8. Accessed: 2023-07-04. 2018 (cit. on pp. 43, 45).

[114] David Gilbert. *QAnon Led the Storming of the US Capitol.* https://bit.ly/3k8I7iP. Accessed: 2023-07-04. 2021 (cit. on pp. 43, 46).

[115] Lois Beckett. *QAnon: a timeline of violence linked to the conspiracy theory.* https://bit.ly/3Cb9a3i. Accessed: 2023-07-04. 2020 (cit. on pp. 43, 45).

[116] BBC. *Twitter suspends 70,000 accounts linked to QAnon.* https://bbc.in/2RlDyGn. Accessed: 2023-07-04. 2021 (cit. on pp. 43, 46).

[117] Brandy Zadrozny and Ben Collins. *YouTube bans QAnon.* `https://nbcnews.to/3ybJC4H`. Accessed: 2023-07-04. 2020 (cit. on pp. 43, 46).

[118] BBC. *Facebook bans QAnon conspiracy theory accounts across all platforms.* `https://bbc.in/3hqARxH`. Accessed: 2023-07-04. 2020 (cit. on pp. 43, 46, 51).

[119] Brandy Zadrozny and Ben Collins. *Reddit bans Qanon subreddits after months of violent threats.* `https://nbcnews.to/3wgGOBR`. Accessed: 2023-07-04. 2018 (cit. on pp. 43, 46).

[120] Ej Dickson. *The QAnon Community Is in Crisis - But On Telegram, It's Also Growing.* `https://bit.ly/3k6752e`. Accessed: 2023-07-05. 2021 (cit. on pp. 43, 46).

[121] Mark Scott. *QAnon goes European.* `https://www.politico.eu/article/qanon-europe-coronavirus-protests/`. Accessed: 2023-07-05. 2020 (cit. on pp. 43, 44, 46, 58).

[122] Statista. *Number of monthly active Telegram users worldwide from March 2014 to July 2023.* `https://www.statista.com/statistics/234038/telegram-messenger-mau-users/`. Accessed: 2024-04-05. 2023 (cit. on pp. 43, 69).

[123] Dimo Angelov. "Top2vec: Distributed representations of topics". In: *arXiv:2008.09470* (2020) (cit. on pp. 44, 56, 85).

[124] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2 (2006), pp. 77–101 (cit. on pp. 44, 60).

[125] Ari Sen and Brandy Zadrozny. *QAnon groups have millions of members on Facebook, documents show.* `https://nbcnews.to/33D9GIO`. Accessed: 2023-07-05. 2020 (cit. on p. 45).

[126] Tina Nguyen. *Trump isn't secretly winking at QAnon. He's retweeting its followers.* `https://politi.co/3z8NYJo`. Accessed: 2023-07-05. 2020 (cit. on p. 45).

[127] The Philadelphia Inquirer. *FBI calls QAnon a domestic terrorist threat.* `https://bit.ly/3nx7vkf`. Accessed: 2023-07-05. 2020 (cit. on p. 45).

[128] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. "An early look at the parler online social network". In: *arXiv preprint arXiv:2101.03820* (2021) (cit. on pp. 46, 69).

[129] Katrin Bennhold. *QAnon Is Thriving in Germany. The Extreme Right Is Delighted.* `https://nyti.ms/3fh824m`. Accessed: 2023-07-05. 2020 (cit. on p. 46).

[130] Antonis Papasavva, Jeremy Blackburn, Gianluca Stringhini, Savvas Zannettou, and Emiliano De Cristofaro. ""Is it a Qoincidence?: An Exploratory Study of QAnon on Voat". In: *The Web Conference.* 2021 (cit. on pp. 46, 54, 57, 66).

[131] Daniel Taninecz Miller. "Characterizing QAnon: Analysis of YouTube comments presents new conclusions about a popular conservative conspiracy". In: *First Monday* (2021) (cit. on pp. 46, 57, 66).

[132] Amanda Garry, Samantha Walther, Rukaya Rukaya, and Ayan Mohammed. "QAnon Conspiracy Theory: Examining its Evolution and Mechanisms of Radicalization". In: *Journal for Deradicalization* (2021) (cit. on p. 46).

[133] Matthew Hannah. "QAnon and the information dark age". In: *First Monday* (2021) (cit. on p. 46).

[134] Kylar J Chandler. "Where We Go 1 We Go All: A Public Discourse Analysis of QAnon". In: *McNair Scholars Research Journal* (2020) (cit. on p. 46).

[135] Samuel Planck. "Where We Go One, We Go All: QAnon and Violent Rhetoric on Twitter". In: *Locus: The Seton Hall Journal of Undergraduate Research* 3.1 (2020), p. 11 (cit. on pp. 46, 66).

[136] Max Aliapoulios, Antonis Papasavva, Cameron Ballard, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Jeremy Blackburn. "The gospel according to Q: Understanding the QAnon conspiracy from the perspective of canonical information". In: *AAAI International Conference on Web and Social Media*. 2021 (cit. on p. 46).

[137] Emilio Ferrara, Herbert Chang, Emily Chen, Goran Muric, and Jaimin Patel. "Characterizing social media manipulation in the 2020 US presidential election". In: *First Monday* (2020) (cit. on p. 46).

[138] Andrea Sipka, Aniko Hannak, and Aleksandra Urman. "Comparing the Language of QAnon-related content on Parler, Gab, and Twitter". In: *14th ACM Web Science Conference*. 2022, pp. 411–421 (cit. on p. 46).

[139] Shruti Phadke, Mattia Samory, and Tanushree Mitra. "Characterizing social imaginaries and self-disclosures of dissonance in online conspiracy discussion communities". In: *Proceedings of the ACM on Human-Computer Interaction* (2021), pp. 1–35 (cit. on p. 46).

[140] Kristen Engel, Yiqing Hua, Taixiang Zeng, and Mor Naaman. "Characterizing Reddit Participation of Users Who Engage in the QAnon Conspiracy Theories". In: *Proceedings of the ACM on Human-Computer Interaction* (2022), pp. 1–22 (cit. on p. 46).

[141] Irene V Pasquetto, Alberto F Olivieri, Luca Tacchetti, Gianni Riotta, and Alessandra Spada. "Disinformation as Infrastructure: Making and maintaining the QAnon conspiracy on Italian digital media". In: *Proceedings of the ACM on Human-Computer Interaction* (2022), pp. 1–31 (cit. on pp. 46, 73).

[142] Crowdtangle. *Crowdtangle API*. `https://github.com/CrowdTangle/API/wiki`. Accessed: 2023-07-05. 2020 (cit. on p. 47).

[143] Telethon. *Telethon's Documentation*. `https://docs.telethon.dev/en/latest/`. Accessed: 2023-07-05. 2017 (cit. on p. 48).

[144] First Draft. `https://firstdraftnews.org/`. Accessed: 2023-07-05. 2020 (cit. on p. 49).

[145] Klaus Krippendorff. "Computing Krippendorff's alpha-reliability". In: *Departmental Papers (ASC), University of Pennsylvania*. 2011 (cit. on p. 49).

[146] J Richard Landis and Gary G Koch. “The measurement of observer agreement for categorical data”. In: *biometrics* (1977), pp. 159–174 (cit. on p. 49).

[147] Statista. *Downloads of Telegram*. `https://www.statista.com/statistics/1260684/telegram-global-downloads-by-region/`. Accessed: 2023-07-05. 2021 (cit. on p. 51).

[148] Brazil Ministry of Health. *Brasil confirma primeiro caso do novo coronavirus*. `https://bit.ly/2VFiFIh`. Accessed: 2023-07-05. 2020 (cit. on p. 52).

[149] Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber. “Racial bias in hate speech and abusive language detection datasets”. In: *Third Workshop on Abusive Language Online*. 2019 (cit. on pp. 53, 78).

[150] Sam Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A Smith. “Realtoxicityprompts: Evaluating neural toxic degeneration in language models”. In: *The 2020 Conference on Empirical Methods in Natural Language Processing*. 2020 (cit. on p. 53).

[151] Nils Reimers and Iryna Gurevych. “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”. In: *Conference on Empirical Methods in Natural Language Processing*. 2019 (cit. on p. 56).

[152] Leland McInnes, John Healy, and James Melville. “Umap: Uniform manifold approximation and projection for dimension reduction”. In: *arXiv:1802.03426* (2018) (cit. on pp. 56, 85).

[153] Leland McInnes, John Healy, Steve Astels, et al. “hdbscan: Hierarchical density based clustering.” In: *Journal of Open Source Software* 2.11 (2017), p. 205 (cit. on pp. 56, 85).

[154] Poynter. *Fighting the Infodemic: The #CoronaVirusFacts Alliance*. `https://www.poynter.org/coronavirusfactsalliance/`. Accessed: 2023-07-05. 2021 (cit. on p. 67).

[155] Milo Trujillo, Maurício Gruppi, Cody Buntain, and Benjamin D. Horne. “What is BitChute? Characterizing the “Free Speech” Alternative to YouTube”. In: *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. 2020, pp. 139–140 (cit. on p. 69).

[156] Chris Stokel-Walker. *Russia’s battle to convince people to join its war is being waged on Telegram*. 2022 (cit. on p. 69).

[157] Aleksandra Urman and Stefan Katz. “What they do in the shadows: examining the far-right networks on Telegram”. In: *Information, Communication & Society* 25.7 (2022), pp. 904–923 (cit. on pp. 70, 72).

[158] Jakob Guhl and Jacob Davey. “A safe space to hate: White supremacist mobilisation on telegram”. In: *Institute for Strategic Dialogue* (2020), pp. 1–20 (cit. on pp. 70, 71).

[159] Yiwei Hou, Hailin Wang, and Haizhou Wang. “Identification of Chinese dark jargons in Telegram underground markets using context-oriented and linguistic features”. In: *Information Processing & Management* 59.5 (2022), p. 103033 (cit. on pp. 70, 72).

[160] Samantha Walther and Andrew McCoy. “US Extremism on Telegram: Fueling Disinformation, Conspiracy Theories, and Accelerationism”. In: *Perspectives on Terrorism* 15.2 (2021), pp. 100–124 (cit. on p. 71).

[161] Bennett Clifford and Helen Powell. “Encrypted extremism: Inside the English-speaking Islamic state ecosystem on telegram”. In: *The George Washington University Program on Extremism* (2019), pp. 27–53 (cit. on p. 72).

[162] Massimo La Morgia, Alessandro Mei, Alberto Maria Mongardini, and Jie Wu. “Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements”. In: *arXiv preprint arXiv:2111.13530* (2021) (cit. on p. 72).

[163] AA Andryukhin. “Phishing attacks and preventions in blockchain based projects”. In: *International Conference on Engineering Technologies and Computer Science*. 2019, pp. 15–19 (cit. on p. 72).

[164] Massimo La Morgia, Alessandro Mei, Francesco Sassi, and Julinda Stefa. “The doge of wall street: Analysis and detection of pump and dump cryptocurrency manipulations”. In: *ACM Transactions on Internet Technology* 23.1 (2023), pp. 1–28 (cit. on p. 72).

[165] Bingyu Gao, Haoyu Wang, Pengcheng Xia, Siwei Wu, Yajin Zhou, Xiapu Luo, and Gareth Tyson. “Tracking counterfeit cryptocurrency end-to-end”. In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4.3 (2020), pp. 1–28 (cit. on p. 72).

[166] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, Maurizio Tesconi, and Emilio Ferrara. “Charting the Landscape of Online Cryptocurrency Manipulation”. In: *IEEE Access* 8 (2020), pp. 113230–113245 (cit. on p. 72).

[167] Mehrnoosh Mirtaheri, Sami Abu-El-Haija, Fred Morstatter, Greg Ver Steeg, and Aram Galstyan. “Identifying and Analyzing Cryptocurrency Manipulations in Social Media”. In: *IEEE Transactions on Computational Social Systems* (2021), pp. 607–617 (cit. on p. 72).

[168] Daniel Pimentel Kansaon, Philipe De Freitas Melo, and Fabrício Benevenuto. ““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*. 2022, pp. 55–65 (cit. on p. 72).

[169] P Grindrod and A Bovet. “Organization and evolution of the UK far-right network on Telegram”. In: *Applied Network Science* 7 (2022) (cit. on p. 77).

[170] Daniel Loureiro, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke, and Jose Camacho-Collados. “TimeLMs: Diachronic Language Models from Twitter”. In: *arXiv preprint arXiv:2202.03829* (2022) (cit. on p. 78).

[171] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. “(Mis) information dissemination in WhatsApp: Gathering, analyzing and countermeasures”. In: *The World Wide Web Conference*. 2019, pp. 818–828 (cit. on p. 91).

[172] Caitlin M Rivers and Bryan L Lewis. “Ethical research standards in a world of big data”. In: *F1000Research* 3 (2014), p. 38 (cit. on p. 92).

List of Figures

3.1 Discovered group URLs on Twitter. 26
3.2 The number of tweets per group URL on Twitter. 27
3.3 Hashtags/Mentions/Retweets in tweets containg group URLs. 28
3.4 Language distribution among tweets containg group URLs. 29
3.5 Staleness of the groups discovered on Twitter. 32
3.6 Accessibility of the groups discovered on Twitter. 33
3.7 The size of the groups discovered on Twitter. 34
3.8 Messages types among the groups discovered on Twitter. 36
3.9 The number of messages shared in the groups discovered on Twitter. . 37

4.1 Activity within QAnon and Baseline chats. 51
4.2 Language distribution among messages in QAnon and Baseline datasets. 52
4.3 Language distribution in QAnon dataset over time. 53
4.4 Toxicity level in QAnon dataset over time. 55
4.5 Topic distribution in QAnon dataset over time. 58
4.6 Topic distribution in QAnon dataset across languages. 59

5.1 The percentage of forwarded messages in Telegram dataset. 75
5.2 The number of forwarded messages from original messages ineach groups/channels in Telegram dataset. 76
5.3 The number of forwarded messages per user in Telegram dataset. . . . 76
5.4 The relationship between the number of senders and shared messages in Telegram dataset. 77
5.5 Category distribution among URLs in Telegram dataset. 78
5.6 Message reach among forwarded and direct messages in Telegram dataset. 79
5.7 The lifetime of forward and direct messages in Telegram dataset. . . . 80
5.8 The lifetime of the messages in channels and groups in Telegram dataset. 81
5.9 The lifetime of messages with/without URLs in Telegram dataset. . . 82
5.10 The lifetime of the URLs from top categories in Telegram dataset. . . 83
5.11 The lifetime of toxic and non-toxic messages in Telegram dataset. . . . 84
5.12 The lifetime of messages with different sentiments in Telegram dataset. 84
5.13 The number of appearances of each message from the case study. . . . 87

List of Tables

2.1 Characteristics of the three online messaging platforms. 15

3.1 Overview of the online messaging platforms dataset. 23
3.2 Topics in tweets containing WhatsApp URLs. 30
3.3 Topics in tweets containing Telegram URLs. 30
3.4 Topics in tweets containing Discord URLs. 30
3.5 Statistics on exposed users' sensitive PII. 38
3.6 User account exposure on Discord 39

4.1 Overview of QAnon dataset. 48
4.2 Performance evaluation of Perspective API. 50
4.3 Percentage of toxic messages. 54
4.4 Top topics of messages. 57
4.5 Distribution of topics into the sample dataset. 62
4.6 The number of messages in each code. 63

5.1 Overview of the Telegram dataset. 74
5.2 Case studies overview. 86