Demystifying the Messaging Platforms' Ecosystem Through the Lens of Twitter

Mohamad Hoseini
Max-Planck-Institut für Informatik
mhoseini@mpi-inf.mpg.de

Philipe Melo
Universidade Federal de Minas Gerais
philmelo@dcc.ufmg.br

Manoel Júnior
Universidade Federal de Minas Gerais
manoelrmj@dcc.ufmg.br

Fabrício Benevenuto
Universidade Federal de Minas Gerais
fabricio@dcc.ufmg.br

Balakrishnan Chandrasekaran
Max-Planck-Institut für Informatik
balac@mpi-inf.mpg.de

Anja Feldmann
Max-Planck-Institut für Informatik
anja@mpi-inf.mpg.de

Savvas Zannettou
Max-Planck-Institut für Informatik
szannett@mpi-inf.mpg.de

ABSTRACT

Online messaging platforms such as WhatsApp, Telegram, and Discord, each with hundreds of millions of users, are one of the dominant modes of communicating or interacting with one another. Despite the widespread use of public group chats, there exists no systematic or detailed characterization of these group chats. There is, more importantly, lack of a general understanding of how these (public) groups differ in characteristics and use across the different platforms. We also do not know whether the messaging platforms expose personally identifiable information, and we lack a comprehensive view of the privacy implications of leaks for the users.

In this work, we address these gaps by analyzing the messaging platforms' ecosystem through the lens of a popular social media platform—Twitter. We search for WhatsApp, Telegram, and Discord group URLs posted on Twitter over a period of 38 days and amass a set of 351K unique group URLs. We analyze the content accompanied by group URLs on Twitter, finding interesting differences related to the topics of the groups across the multiple messaging platforms. By monitoring the characteristics of these groups, every day for more than a month, and, furthermore, by joining a subset of 616 groups across the different messaging platforms, we share key insights into the discovery of these groups via Twitter and reveal how these groups change over time. Finally, we analyze whether messaging platforms expose personally identifiable information. In this paper, we show that (a) Twitter is a rich source for discovering public groups in the different messaging platforms, (b) group URLs from messaging platforms are ephemeral, and (c) the considered messaging platforms expose personally identifiable information, with such leaks being more prevalent on WhatsApp than on Telegram and Discord.

IMC '20, October 27–29, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-8138-3/20/10.
https://doi.org/10.1145/3419394.3423651

CCS CONCEPTS

• **Information systems** → **Social networks**; **Data mining**; **Chat**; • **General and reference** → **Measurement**; • **Security and privacy** → **Social network security and privacy**.

KEYWORDS

Messaging Platforms, Measurement, Privacy, WhatsApp, Telegram, Discord, Twitter

ACM Reference Format:
Mohamad Hoseini, Philipe Melo, Manoel Júnior, Fabrício Benevenuto, Balakrishnan Chandrasekaran, Anja Feldmann, and Savvas Zannettou. 2020. Demystifying the Messaging Platforms' Ecosystem Through the Lens of Twitter. In *ACM Internet Measurement Conference (IMC '20), October 27–29, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3419394.3423651

1 INTRODUCTION

Over the past few years, online messaging platforms such as WhatsApp, Telegram, and Discord have become extremely popular [61], mainly because they provide a seamless, real-time communication platform that connects billions of users from different geographies and socioeconomic statuses. These messaging platforms constitute a rich and complex ecosystem, comprising a conglomerate of various messaging platforms each with its own unique characteristics. This ecosystem has, unfortunately, become an effective medium for disseminating false or malevolent information. Prior work showed, for instance, that WhatsApp played an important role in propagating false information, in particular during major real-world events such as elections in India [6, 40] and Brazil [13, 14, 38, 41, 53, 54]. Telegram has reportedly been exploited by terrorist organizations [63] and white supremacists [3], and Discord for organizing real-world violent protests [56] and disseminating harmful or sensitive material such as revenge porn [19]. These reports unambiguously suggest that the messaging platforms' ecosystem, as an information dissemination medium, has crucial implications to society and humanity at large. This ecosystem is also an invaluable data source for analyzing and understanding emerging socio-technical issues.

Prior work on the exploitation of this ecosystem focused on specific issues, e.g., the dissemination of false information [6, 41, 54], typically within a limited sample, e.g., in a small number of political

groups [38, 40], and in a specific platform, e.g., WhatsApp [13, 14, 53]. User-created groups in 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 [72, 73].

Overall, as a research community, we lack a holistic view of the messaging platforms' ecosystem. Specifically, we do not clearly understand how the 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 the 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 limit their harm on the society.

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

⋆ What is the interplay between Twitter and the different messaging platforms such as WhatsApp, Telegram, and Discord?

⋆ How the groups of these messaging platforms differ from one another and change in composition over time? How long do they remain publicly accessible?

⋆ Do the groups leak any PII and how prevalent are such leaks? What are the privacy implications for users?

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 351,535 group URLs, and, for each group, collect several meta attributes (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 8,255,069 messages posted by 753,329 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 into the topics of conversation in these groups using topic modeling, and compare and contrast the topics across the discovered groups in WhatsApp, Telegram, and Discord groups. We conduct temporal analyses to investigate the changes in 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.

(1) 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, 1111 WhatsApp groups, 1817 Telegram groups, and 5664 Discord groups.

(2) We analyze the content of the tweets, with the group URL(s), to characterize the differences in groups across the 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).

(3) Group URLs across all messaging platforms are ephemeral. We find that 27% of WhatsApp, 20.4% of Telegram, and 68.4% of Discord group URLs become inaccessible within 38 days.

(4) We discover PII leaks via WhatsApp, Telegram, and Discord groups. Specifically, we find the phone numbers of over 54K 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 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.

Paper Organization. The rest of this paper is organized as follows. Section 2 provides the background on WhatsApp, Telegram, and Discord, while Section 3 discusses our data-collection methodology and dataset. Section 4 presents how WhatsApp, Telegram, and Discord groups are shared on Twitter, and Section 5 analyzes the activity and evolution of the discovered groups. In Section 6, we present our analysis on the privacy implications for users from the use of these messaging platforms, while Section 7 reviews prior work. Finally, we conclude in Section 8.

2 BACKGROUND

In this section, we provide the necessary background information on WhatsApp, Telegram, and Discord, and present the characteristics of these messaging platforms, highlight how they differ from one another, in Table 1.

WhatsApp.

Launched in January 2009, WhatsApp is the largest messaging platform with over 2 B users [61] and the most used social media platform, second only to Facebook [47]. To use the messaging 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, image, 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 messaging platform, WhatsApp is the mainstream

Table 1: ***An overview of the characteristics of the different messaging platforms, highlighting some of the differences.***

Characteristic	***WhatsApp***	***Telegram***	***Discord***
Initial release date	January 2009	August 2013	May 2015
User base	2 Billion	400 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

communication medium for billions of people. Second, prior work indicates (ab)use of WhatsApp for disseminating false information [64] and dissemination of hateful rhetoric can incite violence in the real world [8].

Telegram.

Launched in August 2013, Telegram is a messaging platform with approximately 400 M monthly users [65]. Similar to WhatsApp, it requires users to register with their phone numbers, and after registration allows them to communicate also 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 K members per group. Both groups and channels allow uses to share and forward information in a wide range of formats. Telegram also support audio and video calls between users. 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.

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 [3] and terrorists [63]. Telegram has also received relatively less attention in academic research.

Discord.

Though it started with a focus on providing a messaging platform for the online gaming community, Discord is nowadays used by the general public for various purposes, even including education [24]. 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 a server, a user can exchange messages with others 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 250K by default—and some (e.g., "verified" servers of organizations, artists, or games) can host up to 500K users. Lastly, channels in Discord do not offer end-to-end encryption.

We include Discord in this work as it is a fast growing messaging platform and especially attracts 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 [56], and for disseminating potentially harmful and sensitive material, e.g., revenge porn [19].

3 METHODOLOGY & DATASET

We measure the use of different messaging platforms and identify key differences in their use. 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 paper, 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.

3.1 Discovering WhatsApp, Telegram, and Discord groups

All three messaging platforms support public groups (refer §2), 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 message 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 messaging platform. We review each platform's documentation and also manually examine the URLs of each platform to compile a list of six patterns employed across these messaging 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 [67] every hour, and (b) using Twitter's Streaming API [68]. 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 messaging platforms) from 2,234,128 tweets posted by 806,372) Twitter users (refer left side of Table 2). Per this table, we discover a larger number of group URLs from Discord (227K) than either Telegram (78K) or WhatsApp (45K). The large number of Discord and Telegram groups discovered despite these platforms being smaller (in terms of 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 [21]; users, hence, are likely sharing a large number of unique group URLs compared to the other messaging 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.

Limitations. The use of Twitter as the only data source for discovering public groups of the different 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.

3.2 Collecting group-specific metadata

Although we can join a 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 the messaging platform.

Table 2: *Overview of our datasets.*

	Twitter			Messaging Platforms		
	#Tweets	#Users	#Group URLs	#Joined 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

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 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 scope of the study. Third, we intend to minimize disruptions caused by joining hundreds of thousands of groups on any messaging 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 the 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. We only store the hash of the phone number, although it is available to anyone with access to the group URL.

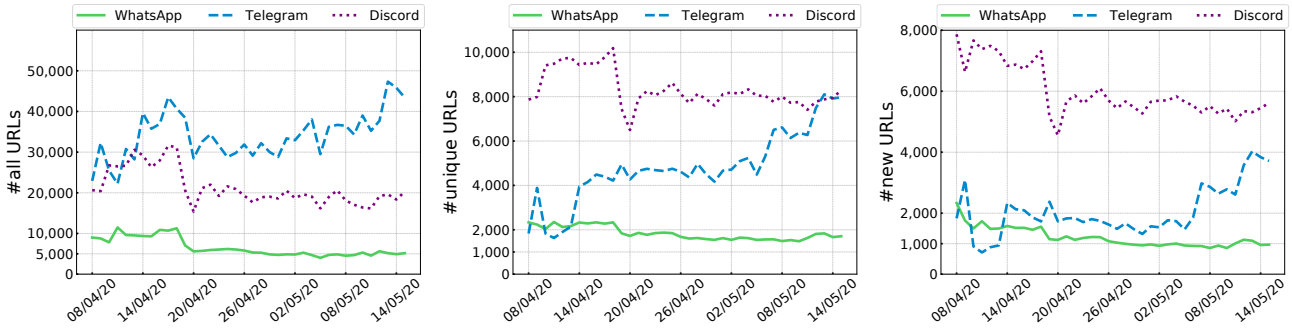
Telegram. Similar to the method for WhatsApp, we use Telegram's Web client to obtain basic information about Telegram groups without joining them. We implement a custom scraper that obtains and parses the web page for each group to gather several details: (1) title of the group; (2) size of the group 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 [22]. 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 three messaging 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.

3.3 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,



(a) *All* (b) *Unique* (c) *New (not seen in previous days)*

Figure 1: ***Number of group URLs discovered on each day during our Twitter data collection, showing that Twitter is a rich data source for discovering public groups on the different messaging platforms.***

we describe how we obtain data from within the groups on every messaging 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 [1]. 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 [66]. 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) 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 the 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 [23]: (1) messages on all groups on the joined servers (since the data each group was created) and (2) user profiles for the group members.

3.4 Ethics

We submitted our methodology to our institution's ethical review board and obtained approval prior to collecting any data. We emphasize that we (a) work only with publicly available data; (b) do not store users' phone numbers as such, but use one-way hashes of such data; (c) do not attempt to de-anonymize users from any personally identifiable information; and (d) do not attempt to link users across platforms. We follow standard ethical guidelines [55].

4 DISCOVERING PUBLIC GROUPS

In this section, we analyze the tweets that contain group URLs from WhatsApp, Telegram, and Discord for understanding the interplay between Twitter and these 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 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 messaging platforms (see Fig. 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). WhatsApp appears, per Fig. 1, to be the most "private" 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 messaging

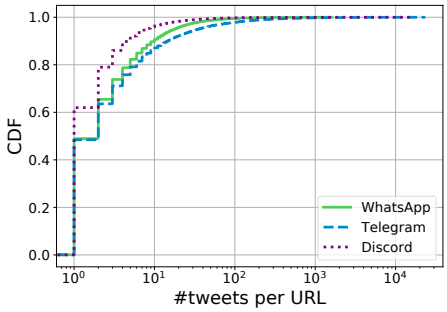
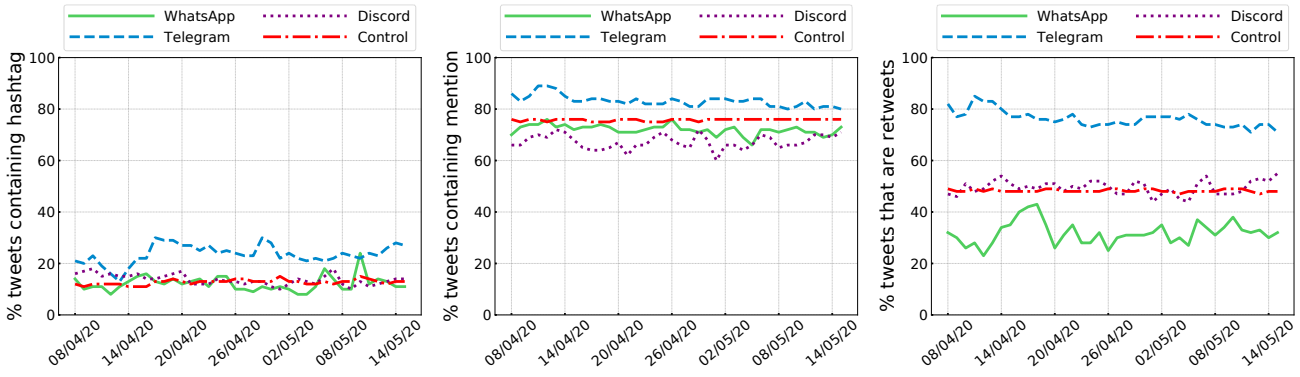


Figure 2: ***CDF of number of tweets for each group URL over the entire dataset. A large percentage of the group URLs are shared only once.***



(a) *Hashtags* (b) *Mentions* (c) *Retweets*

Figure 3: ***Percentage of tweets that contain hashtags/mentions and percentage of tweets that are retweets. Twitter users tend to not use hashtags and use mentions when sharing WhatsApp, Telegram, and Discord groups. Tweets with Telegram groups are more likely to get retweeted compared to the other platforms.***

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. 1a), with 33,864 group URLs, in the median, per day, followed by Discord with 19,970 URLs. In terms of unique group URLs discovered each day (Fig. 1b), Discord, however, surpasses Telegram (8,090 URLs vs 4,661 URLS, in the median). These findings indicate that Telegram groups are shared more number of times than that of Discord and WhatsApp, within the same day (see Fig. 1a and Fig. 1b). The number of newly discovered group URLs per day (Fig. 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 messaging platforms.

Fig. 2 sheds more light into 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 10K) tweets. We find, via manual examination, that 11 groups focus on pornography and 2 on cryptocurrencies, and one to be 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 or 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 also respond. In the same vein, a *retweet* is a broadcast of a specific tweet to all the followers of the "retweeting" user. Next, we analyze the prevalence in the use of these mechanisms in the tweets that include group URLs from the three messaging platforms.

Per Fig. 3a, only a small percentage of tweets include hashtags for all three messaging platforms. Specifically, tweets containing Telegram group URLs are more likely to include hashtags (24% of these tweets include hashtags), while for the other two messaging 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. 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.

When analyzing tweets with mentions (see Fig. 3b), we observe a larger percentage of tweets with mentions compared to that in the control dataset and the other messaging platforms (73%, 84%, 68%, 76% for WhatsApp, Telegram, Discord, and control, respectively), likely because Twitter users 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. 3c), shows that 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

Table 3: *Topics extracted from the English tweets that include WhatsApp, Telegram, and Discord group URLs.*

#	Whatsapp		Telegram		Discord	
	Label	Topic Terms	Label	Topic Terms	Label	Topic Terms
1	**Forex training (6%)**	learn, free, forex, training, join, trading, text, mini, class, animation	**Cryptocurrencies (9%)**	bitcoin, join, sats, get, winners, sex, hours, chat, nice, come	**Gaming (7%)**	patreon, free, get, today, mystery, public, gaming, gamedev, indiegames, alongside
2	**Earn money from home (8%)**	home, earn, don, just, money, using, can, start, stay, google	**Cryptocurrencies(9%)**	usdt, giveaways, oin, winners, ollow, enter, btc, trc, trx, hours	**Organizing online events (7%)**	will, may, hosting, week, one, time, tonight, don, night, last
3	**Instagram Followers Boosting (9%)**	join, followers, instagram, gain, want, money, online, group, learn, make	**Social Network Activity (11%)**	follow, like, retweet, giveaway, tag, join, win, twitter, friends, friend	**Gaming (5%)**	like, oin, alpha, deal, daily, art, lots, battle, raffle, nintendo
4	**Cryptocurrencies (7%)**	bitcoin, ethereum, crypto, currency, ads, year, like, line, people, new	**Ask Me Anything/Quiz (8%)**	ama, may, will, utc, quiz, someone, wallet, don, ust, today	**Advertising Discord groups (33%)**	discord, join, server, link, can, visit, want, just, new, hey
5	**Earn money from home (13%)**	make, can, money, know, daily, home, earn, forex, cash, market	**Advertising Telegram groups (14%)**	free, join, just, telegram, money, day, channel, don, can baby	**Pokemon (7%)**	united states, venonat, bite, quick, bug, full, fortnite, pikacku, confusion
6	**Cryptocurrencies (5%)**	learn, cryptocurrency, make, join, days, period, another, want, day, accumulate	**Sex (13%)**	new, worth, user, brand, xpro, performer, smartphones, girls, boobs, price	**Advertising Discord groups (10%)**	giveaway, follow, retweet, friends, tag, join, discord, enter, fast, winners
7	**WhatsApp group advertisement (30%)**	join, group, whatsapp, link, follow, click, please, chat, open, twitter	**Giveaways (7%)**	giving, away, will, tmn, link, honor, full, butt, video, get	**Tournaments (9%)**	good, live, launching, now, tournament, open, next, will, free, prize
8	**Making money (9%)**	get, never, time, actually, income, chat, best, taking, account, full	**Sex (10%)**	fuck, want, girl, click, show, trading, pussy, powerful, can, cum	**Giveaways (8%)**	giving, est, away, awp, will, saturday, friday, coins, many, competition
9	**Nigeria-Related (6%)**	will, new, retweet, capital, people, now, interested, writing, nigerian, online	**Advertising Telegram groups (11%)**	telegram, join, group, channel, now, below, link, get, available, opened	**Advertising Discord groups (4%)**	discord, join, make, sure, ends, chat, token, https, music, server
10	**Cryptocurrencies (6%)**	business, ethereum, free, smart, skills, eth, million, join, training, webinar	**Referral Marketing (8%)**	airdrop, open, https, tokens, wink, referral, token, earn, new, good	**Hentai (9%)**	join, discord, server, come, hentai, now, new, paradise, tenshi, official

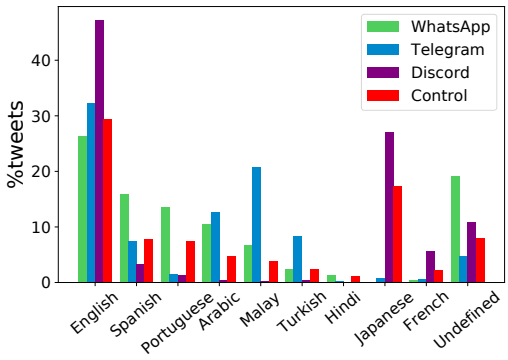


Figure 4: *Percentage of tweets for each language. English is the most popular language on tweets sharing WhatsApp, Telegram, and Discord group URLs.*

that English is the most popular language with . Fig. 4 shows the percentage of tweets in each language across the three 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 its 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 into the demographics of the users sharing the public groups and using the groups on the messaging platforms.

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) [12]. First, we focus on English, since it is the most popular language for tweets including group URLs for all three messaging platforms. For each platform, we extract all the English tweets, remove stop words, and extract ten topics using the LDA method. Table 3 reports 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 messaging platform; (2) *meso topics* that refer to topics that exist to more than a single messaging platform; and (3) *macro topics* that refer to topics that exist across all messaging 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), sex-related topics on Telegram (23%, see topics 6 and 8), and gaming (12%) and hentai-related (japanese anime and manga pornography, 9% of all tweets) topics on Discord (see topics 1, 3, and 10). We find several meso topics related to cryptocurrencies on both WhatsApp (18%, see topics 3, 6, 10) and Telegram groups (18%, see topics 1 and 2), but not for Discord. Finally, for macro topics, we observe that across all messaging 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.[1] This highlights that Twitter users are not sharing many politics-related groups from 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 WhatsaApp).

Overall, our LDA analysis allow us to obtain insights into the content of the discovered 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 messaging platforms, while at the same time there are some topics where users prefer specific 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

[1] We repeated our analysis with a larger number of topics (up to 50 topics per messaging platform) and no politics-related topic emerged.

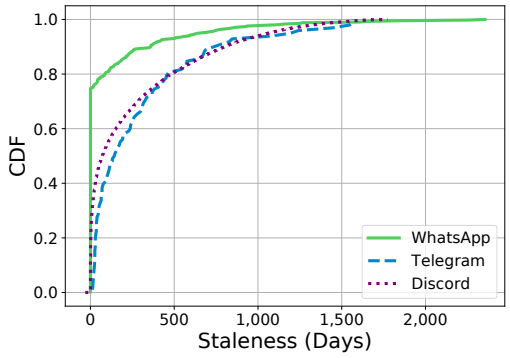


Figure 5: ***Staleness: time difference between the appearance of the group on Twitter and its creation date. Older Telegram and Discord groups are shared on Twitter, while shared WhatsApp groups are "fresh".***

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 cryptocurrencies discussions, Telegram for disseminating pornographic content, and Discord mainly for gaming, giveaways, tournaments, and hentai.

5 ACTIVITY AND EVOLUTION OF PUBLIC GROUPS

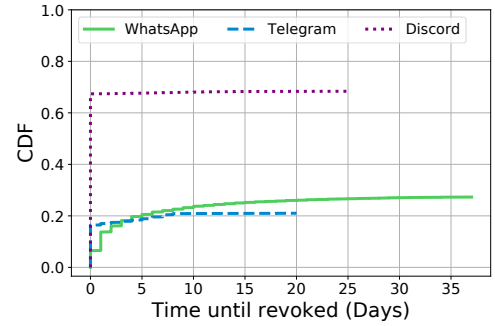
In this section, we analyze the data obtained from the WhatsApp, Telegram, and Discord groups discovered from Twitter, with a focus on understanding the characteristics of those groups, how they change over time, and the volume of information disseminated within them.

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, 95.9% for Discord, and 92.7% for WhatsApp), with only a small percentage of users creating 2 groups or more (5.3% for WhatsApp and 3.6% 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 one 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 to reach a larger audience.

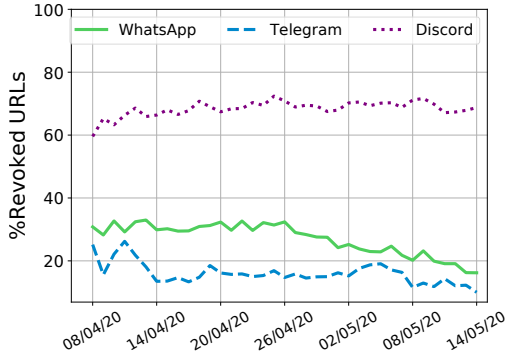
Group Creation Dates. Next, we analyze the creation dates for the groups. For Discord, the creation date is available without the need to join the groups, yet for WhatsApp and Telegram we only obtain this data after joining the groups (416 for WhatsApp and 100 for Telegram). Based on the creation date, 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. 5, we observe that most of the WhatsApp groups are created and shared on Twitter on the same day (76%), 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 25.6% 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 (see Section 6), hence we limit this analysis on 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 (340 million, followed by Brazil with 99 million [61], 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 8.15 million users in Brazil and 7.91 million in India [62]).

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 a 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. 6a shows the accessibility time (in days) for the revoked URLs, while Fig. 6b shows the percentage of revoked group URLs per day. We find that 27.3% of the URLs for WhatsApp groups, 20.4% of the Telegram group URLs, and 68.4% of the Discord group URLs are revoked at some time. This shows that Discord has much 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 (6.4% of all groups for WhatsApp, 16.3% for Telegram, and 67.4% for Discord). This indicates that some groups have a very limited accessible period, indicating the ephemeral nature that messaging platforms' groups have. The ephemeral nature of messaging platforms' groups should



(a) *Accessible period of URLs*



(b) *Revoked URLs per day*

Figure 6: ***Analysis on how long groups are accessible until they get revoked. A substantial percentage of groups are not accessible during our first observation (especially Discord groups).***

be taken into consideration in future research focusing on collecting and analyzing datasets from messaging platforms.

Group Members. Since users share group URLs on Twitter to entice others to join, the size of a group over time can hint of their activity and the reasons behind its revocation. To this end, we gather the number of members in each group, for each day that are accessible. We compare the distribution of total amount of members for each platform in Fig. 7a. Overall, WhatsApp has much less 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 less members than Telegram, as around 60% of Discord groups have less than 100 members while only 40% of the Telegram 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. 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.[2]

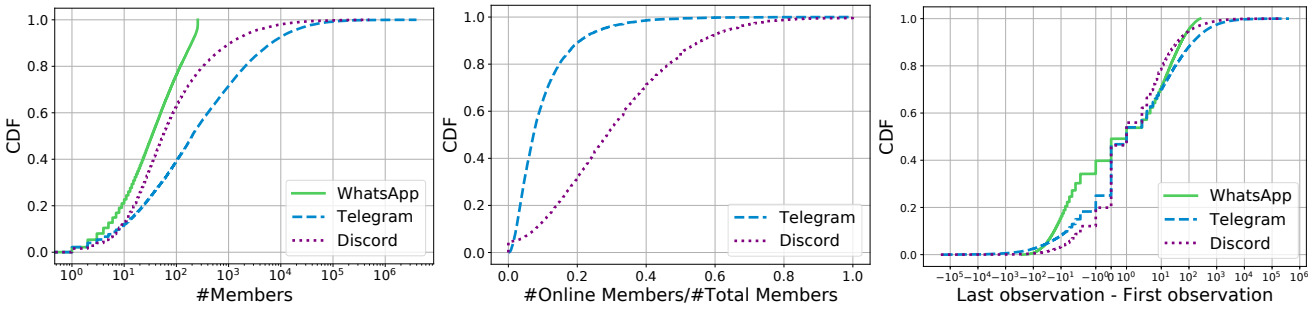
Finally, we investigate the growth of the groups over time; Fig. 7c shows the distribution of the growth of the groups, which is the difference of group sizes observed on the first and the last day (i.e., prior to revocation) of observation. We can clearly observe the impact of the limit sizes for each 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% Telegram and 54% 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 a declining interest among the members of some groups over time.

Group messages. Next, we analyze the collected messages from all the joined groups. Overall, we gather 476,059 messages from WhatsApp, 3,148,826 messages from Telegram, and 4,630,184 messages from Discord. First, we compare the types of messages in each messaging platform, as all platforms allow users to send text, images, videos, audios, stickers, and documents. Fig. 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).[3] 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. Note that Telegram also has a small portion of "other" types of messages including service messages (e.g., users joining/leaving group, editing group information).

We also look into the volume of messages shared in each group and the number of messages per user. Fig. 9a shows the number of messages shared per day in each group for all the messaging 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 are shared by 12,434 distinct users on WhatsApp, 100,504 users on Telegram, and 34,543 users on Discord. This represents, respectively, 59.4%, 14.6% and 65.8% of the total number of members in the joined groups (see Table 2). Although we

[2]Note that a Discord user is shown online even if the Discord Web/desktop client is running in the background.
[3]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.



(a) *Total size of groups* (b) *Fraction of online members* (c) *Size difference over time*

Figure 7: ***Distributions of number of members per group, percentage of online members over all members, and group size change over time (between first and last observation).***

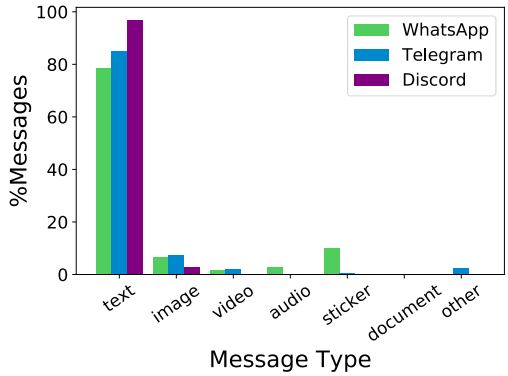
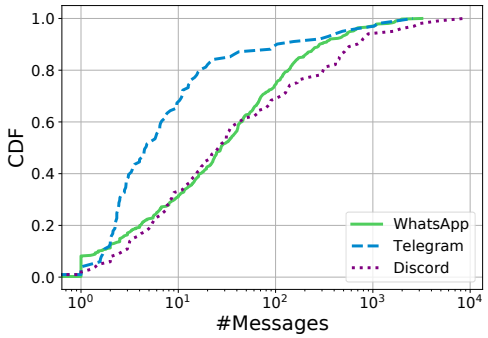
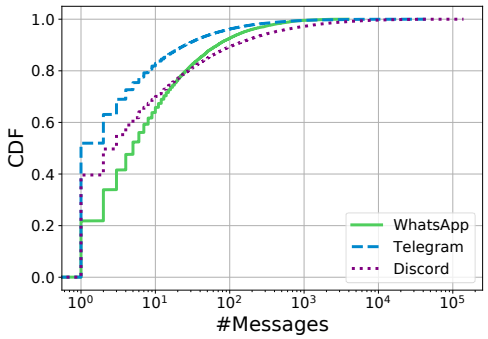


Figure 8: ***Percentage of messages in each message type. Text is the predominant type across all messaging platforms.***



(a) *Mean number of messages for each group per day*



(b) *Messages per user*

Figure 9: ***Distribution of the number of messages shared in the groups of each platform and their users.***

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., creator and few administrators).

Finally, we analyze the volume of messages shared per active member in Fig. 9b. We observe that most members share only a few messages, while some share a large volume of messages. In particular, 65.8% of them share up to 10 messages for WhatsApp, 70.1% for Discord and 82.9% for Telegram. When looking at the volume of the messages shared by the top 1% of the members (in terms of 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 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 of Discord group URLs expire during the first days after shared on Twitter, while WhatsApp group URLs last longer. Also, Telegram group URLs are less likely to get revoked.

We observe that the difference of the group size limit between the three platforms indeed impacts the size of the groups, since Telegram and Discord have larger groups 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 online members. The selection bias and ephemeral nature of group URLs, discovered on Twitter, has implications for studies that use such URLs.

Table 4: *Statistics on users' sensitive PII that are exposed from the three messaging platforms.*

	WhatsApp	Telegram	Discord
Users observed	20,906 members 34,078 creators	74,479 members	25,701 members
Users' Phone Numbers	54,984 (100%)	509 (0.68%)	-
Users' Social Networks	-	-	7,708 (30%)

Table 5: Number and percentage of Discord users whose their accounts on other platforms are exposed.

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%)

6 PRIVACY IMPLICATIONS

In this section, we analyze the users' privacy implications from using WhatsApp, Telegram, and Discord. When dealing with social media platforms, a common concern is about privacy and exposure of sensitive personally identifiable information (PII). In particular, for messaging 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 are joining 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 regards 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 messaging platform and analyze them to understand the underlying privacy implications from the use of messaging platforms.

Each of the 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 that participate in groups that we also are members of. This also applies for 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 4 reports the number of users whose PII are exposed for each messaging platform. Looking at the total number of users which we collected data, we find 20,906 WhatsApp users within the groups we joined, and 34,078 unique users that are the creators of the rest of the groups that are accessible, totaling 54,984 users. For Telegram, we collect information for 74,479 users, while for Discord we find 25,701 users. Note that for Telegram and Discord, the number of users is smaller than the total of users for groups we joined, representing 10.8% and 49% for Telegram and Discord, respectively. This is because on Telegram, administrators are able to restrict the access to the member list, thus users can not see who are the members of the group. For Discord, the API blocks bots 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.

Our data collection and analysis highlights the exposure of PII information in each platform. Alarmingly, on WhatsApp we are able to obtain the phone number of *all* users that we discovered during our data collection, a total of 54,498 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 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 7,708 users (30%) for who 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 5 reports the number of users whose users' accounts are exposed for each of the other linked platforms. We find that 20.4% 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 account.

Overall, these findings have important implications to user's privacy. The exposure of PII from all these messaging platforms can be potentially exploited by malevolent actors that aim to target users. For instance, state-sponsored actors [74, 75] 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 and then their targeting on other social media platforms via posts or advertisements with the goal to manipulate them or change their ideology. Also, our results highlight the need to raise user awareness about the privacy implications from the use of 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.

7 RELATED WORK

In this section, we review previous work related to analyzing and measuring online social networks, as well as online messaging platforms like WhatsApp, Telegram, and Discord.

Social Networks. A rich body of previous work focus on 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 [15, 34, 43], YouTube [16, 25], Reddit [9, 28, 60], Flickr [17, 42], and Facebook [37, 69]. More recently, previous work focus on analyzing and measuring emerging fringe social networks like 4chan [11, 31], an anonymous imageboard, Gab [36, 71], an alt-right Twitter clone, and Mastodon [51], a decentralized microblog. Furthermore, motivated by the overwhelmingly large number of social networks available, previous work focus on analyzing multiple social networks and measuring the interplay between them [18, 44, 72, 73]

WhatsApp. Previous measurement studies on WhatsApp mainly focus on acquiring data from public WhatsApp groups and analyze their content to study emerging phenomena like the spread of false information. Rosenfeld et al. [57] perform surveys to characterize the behavior of 4M messages from 100 WhatsApp users with the goal of inferring demographic information. Then, Garimella and Tyson [27] develop a set of tools that enable the large-scale collection of WhatsApp data from public groups, finding 2.5K groups and joining 200 in order to characterize WhatsApp users in India. Bursztyn and Birnbaum [13] 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 [14, 38, 53, 54], India [26, 52], and Ghana [45]. Resende et al. [54] 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. [39] characterize the content of audio messages shared on WhatsApp. Melo et al. [40] develop a system, to assist fact checkers, that gathers data from 1.1K groups from Brazil and India, and daily displays the most popular content (i.e., messages, images, URLs, audio, and video). Melo et al. [41] also investigate the impact of message forwarding limits on the spread of messages in WhatsApp public groups, suggesting that the limit of 5 for forward is not sufficient to contain the spread of viral content in the platform. Finally, a recent study [52] 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. [10] collect and make publicly available a large-scale dataset of 27K Telegram groups and 317M messages. Anglano et al. [5] and Satrya et al. [58] investigate the artifacts generated by the Telegram application, while Abu-Salma et al. [2] 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. [48] study the use of Telegram by Iranian immigrants with a focus on understanding how Telegram groups are moderated. Hashemi et al. [30] perform a large-scale analysis on 900K Iranian channels and 300K 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. [7] analyze the use of the Telegram platform in Iranian libraries. Akbari et al. [4] 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. [20] collect data from 2.6K Telegram groups and channels and perform a structural analysis of the content posted within those groups/channels. Naseri et al. [46] 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 [50, 59, 70]. Such organizations exploit the Telegram platform for their communication purposes, to spread propaganda, and possibly recruit new members.

Discord. Finally, we review previous research on Discord. Hamrick et al. [29] study pump and dump schemes on the cryptocurrencies market by analyzing data obtained from Discord. Lacher and Biehl [35] examine the use of Discord for teaching purposes. Jiang et al. [32] study the moderation challenges that exist on Discord, and in particular on voice-based channels. Similarly, Kiene and Hill [33] focus on moderation on Discord and more specifically in the use of bots for moderating content posted on Discord servers.

Remarks. Overall, previous studies are dedicated on measuring the dynamics and discourse of specific topics in each of the messaging platforms considered. Importantly, these previous studies show that all popular 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 about the dynamics of public groups on multiple platforms and do not attempt to characterize key differences of them. In this work, we fill this gap, by performing, to the best of our knowledge, the largest multi-platform analysis of messaging platforms by collecting and providing an in-depth study of 351K groups from WhatsApp, Telegram, and Discord, shared on Twitter.

8 CONCLUSION

In this paper, we performed a large-scale characterization of public groups from WhatsApp, Telegram, and Discord shared on Twitter, a popular micro blogging platform. We searched for group URLs (or invite links) on Twitter for all three platforms for a period of over a month and obtained a set of approximately 351K URLs to groups. By performing topic modeling on the tweets including group URLs, we were able to understand the content of these groups and the differences that exist between these messaging platforms.

Although these platforms are also designed for private conversations, we find that Twitter is a rich data source for discovering public chat groups. Our findings highlight several points that need to be considered by the research community focusing on similar platforms. First, we show that by taking a multi-platform view of the Web ecosystem, we can extract meaningful insights that otherwise will be hard to deduce if we were studying, for instance, WhatsApp in isolation. Indeed, recent research efforts aimed to

study this phenomena [72, 73] in the context of news and memes, however, they do not consider instant messaging platforms.

We meticulously monitored discovered groups from three platforms, gathering measurements once per day, and used these coarse-grained measurements for investigating the changes in the characteristics of the groups (e.g., number of members) over time. Our analysis highlights the ephemeral nature of groups, as during the course of this study 27% of the groups become inaccessible for WhatsApp, 20% for Telegram, and 68% for Discord. This phenomenon prompts the need to design and develop robust, scalable, and real-time data collection solutions that will enable the research community to obtain a more holistic and complete overview of the messaging platforms' ecosystem. We also joined a sample consisting of hundreds of those groups for all three platforms, and provided a characterization of the activity within the groups.

Finally, we analyzed the exposure of sensitive PII on all three platforms, particularly phone numbers for WhatsApp and Telegram, and linked social media accounts for Discord users. For WhatsApp, even without an account, we could collect an impressive number of over 34K phone numbers. Moreover, after joining groups, we obtain another 20K phone numbers. We also found exposed phone numbers for a small portion of users on Telegram (less than 1%). Finally on Discord, we were unable to find phone numbers, however we collected at least one linked social network account for 30% of the users analyzed. These privacy implications are alarming, since messaging platforms are often used because of their perceived security in communication and privacy. Our results highlight the need to raise awareness of the public related to these privacy implications and design guidelines on how messaging platforms can adjust to better safeguard users' privacy.

Limitations. Naturally, our work has some limitations. First, we rely solely on Twitter to discover groups from WhatsApp, Telegram, and Discord, hence we are unaware for a large number of publicly available groups. Despite this fact, Twitter is a very large and mainstream social network that we use to make a best effort attempt to discover a large number of groups from WhatsApp, Telegram, and Discord, and mitigate potential biases. Second, we join and collect data from only a limited number of groups from WhatsApp, Telegram, and Discord, mainly because these 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, while Telegram's API is rate-limited. Overall, this is a limitation that exists in every study that collects data from messaging platforms.

Future Work. As part of our future work, we aim to expand our data collection so that we discover WhatsApp, Telegram, and Discord groups shared on other mainstream and popular social networks like Facebook and Instagram. Also, we aim to undertake a focused data collection within groups by selecting groups related to specific interesting topics like politics and COVID-19, with the goal to study the propagation of information across WhatsApp, Telegram, Discord, and assess the prevalence of toxic content shared within such groups (i.e., by leveraging Google's Perspective API [49]). Finally, we aim to investigate whether the exposure of PII from the messaging platforms is exploited by spammers or other malevolent actors that aim to target or manipulate users.

Acknowledgments. We thank Hamed Haddadi, our shepherd, and the anonymous reviewers for their insightful feedback. Philipe Melo, Manoel Junior, and Fabricio Benevenuto were partially supported by grants from Fapemig, CNPq, and CAPES.

REFERENCES

[1] 2018. WhatsApp Wrapper. https://github.com/mukulhase/WebWhatsapp-Wrapper.
[2] Ruba Abu-Salma, Kat Krol, Simon Parkin, Victoria Koh, Kevin Kwan, Jazib Mahboob, Zahra Traboulsi, and M Angela Sasse. 2017. The Security Blanket of the Chat World: An Analytic Evaluation and a User Study of Telegram. Internet Society.
[3] ADL. 2019. Telegram: The Latest Safe Haven for White Supremacists. https://www.adl.org/blog/telegram-the-latest-safe-haven-for-white-supremacists.
[4] Azadeh Akbari and Rashid Gabdulhakov. 2019. Platform Surveillance and Resistance in Iran and Russia: The Case of Telegram. *Surveillance & Society* 17, 1/2 (2019), 223–231.
[5] Cosimo Anglano, Massimo Canonico, and Marco Guazzone. 2017. Forensic analysis of telegram messenger on android smartphones. *Digital Investigation* 23 (2017), 31–49.
[6] Chinmayi Arun. 2019. On WhatsApp, Rumours, and Lynchings. *Economic & Political Weekly* 54, 6 (2019), 30–35.
[7] Amir Reza Asnafi, Shima Moradi, Mohadeseh Dokhtesmati, and Maryam Pakdaman Naeini. 2017. Using mobile-based social networks by Iranian libraries: The case of Telegram Messenger. *Libr. Philos. Pract* 2017, 1 (2017).
[8] Bassi, Simi and Sengupta, Joyita. 2018. 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.
[9] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 830–839.
[10] Jason Baumgartner, Savvas Zannettou, Megan Squire, and Jeremy Blackburn. 2020. The Pushshift Telegram Dataset. In *ICWSM*.
[11] Michael Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Greg Vargas. 2011. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. (2011).
[12] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
[13] Victor S Bursztyn and Larry Birnbaum. 2019. 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 (ASONAM '19)*.
[14] Josemar Alves Caetano, Gabriel Magno, Marcos Gonçalves, Jussara Almeida, Humberto T. Marques-Neto, and Virgílio Almeida. 2019. Characterizing Attention Cascades in WhatsApp Groups. In *Proceedings of the 10th ACM Conference on Web Science (WebSci '19)*. ACM, 27–36.
[15] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, P Krishna Gummadi, et al. 2010. Measuring user influence in twitter: The million follower fallacy. *Icwsm* 10, 10-17 (2010), 30.
[16] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. 2007. 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*. 1–14.
[17] Meeyoung Cha, Alan Mislove, and Krishna P Gummadi. 2009. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web*. 721–730.
[18] Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, and Eric Gilbert. 2017. 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*. 3175–3187.
[19] Joseph Cox. 2018. 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.
[20] Arash Dargahi Nobari, Negar Reshadatmand, and Mahmood Neshati. 2017. Analysis of Telegram, An Instant Messaging Service. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2035–2038.
[21] Discord. 2020. https://support.discord.com/hc/en-us/articles/208866998-Invites-101.
[22] Discord. 2020. Discord API. https://discord.com/developers/docs/resources/guild.
[23] Discord. 2020. Discord OAuth API. https://discord.com/developers/docs/topics/oauth2.

[24] Discord. 2020. How to use Discord for your classroom. https://blog.discord.com/how-to-use-discord-for-your-classroom-8587bf78e6c4.

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

[26] Kiran Garimella and Dean Eckles. 2017. Image based Misinformation on WhatsApp. In *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)*.

[27] Kiran Garimella and Gareth Tyson. 2018. WhatApp Doc? A First Look at WhatsApp Public Group Data. In *Twelfth International AAAI Conference on Web and Social Media (ICWSM '18)*.

[28] Eric Gilbert. 2013. Widespread underprovision on Reddit. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 803–808.

[29] JT Hamrick, Farhang Rouhi, Arghya Mukherjee, Amir Feder, Neil Gandal, Tyler Moore, and Marie Vasek. 2018. An examination of the cryptocurrency pump and dump ecosystem. *Available at SSRN 3303365* (2018).

[30] Ali Hashemi and Mohammad Ali Zare Chahooki. 2019. Telegram group quality measurement by user behavior analysis. *Social Network Analysis and Mining* 9, 1 (2019), 33.

[31] Gabriel Emile Hine, Jeremiah Onaolapo, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Riginos Samaras, Gianluca Stringhini, and Jeremy Blackburn. 2016. Kek, cucks, and god emperor trump: A measurement study of 4chan's politically incorrect forum and its effects on the web. *arXiv preprint arXiv:1610.03452* (2016).

[32] Jialun Aaron Jiang, Charles Kiene, Skyler Middler, Jed R Brubaker, and Casey Fiesler. 2019. Moderation Challenges in Voice-based Online Communities on Discord. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[33] Charles Kiene and Benjamin Mako Hill. 2020. 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 Extended Abstracts*. 1–8.

[34] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web*. 591–600.

[35] Lisa Lacher and Cydnee Biehl. 2018. Using discord to understand and moderate collaboration and teamwork. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 1107–1107.

[36] Lucas Lima, Julio CS Reis, Philipe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto. 2018. 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 (ASONAM)*. IEEE, 515–522.

[37] Yabing Liu, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 61–70.

[38] Caio Machado, Beatriz Kira, Vidya Narayanan, Bence Kollanyi, and Philip Howard. 2019. 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 (WWW '19)*. ACM, 1013–1019.

[39] Alexandre Maros, Jussara Almeida, Fabricio Benevenuto, and Marisa Vasconcelos. 2020. Analyzing the Use of Audio Messages in WhatsApp Groups. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan*. ACM. https://doi.org/10.1145/3366423.3380070

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

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

[42] Alan Mislove, Hema Swetha Koppula, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. 2008. Growth of the flickr social network. In *Proceedings of the first workshop on Online social networks*. 25–30.

[43] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosenquist. 2011. Understanding the demographics of twitter users.. In *ICWSM*.

[44] Alan Mislove, Massimiliano Marcon, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. 2007. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. 29–42.

[45] Andrés Moreno, Philip Garrison, and Karthik Bhat. 2017. Whatsapp for Monitoring and Response During Critical Events: Aggie in the Ghana 2016 Election. In *Proc. of 14th Int'l Conf. on Information Systems for Crisis Response and Management (ISCRAM'17)*.

[46] Mohammad Naseri and Hamed Zamani. 2019. 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*. 1053–1056.

[47] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, and Rasmus Kleis Nielsen. 2019. Reuters Institute Digital News Report 2019 . Reuters Institute for the Study of Journalism. , 18–19 pages.

[48] Sarah Nikkah, Andrew D Miller, and Alyson L Young. 2018. Telegram as An Immigration Management Tool. (2018).

[49] Perspective API. 2018. https://www.perspectiveapi.com/.

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

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

[52] Julio Reis, Philipe Melo, Kiran Garimella, Jussara M. Almeida, Dean Eckles, and Fabricio Benevenuto. 2020. 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 (ICWSM '20)*.

[53] Gustavo Resende, Philipe Melo, Julio C. S. Reis, Marisa Vasconcelos, Jussara M. Almeida, and Fabrício Benevenuto. 2019. Analyzing Textual (Mis)Information Shared in WhatsApp Groups. In *Proceedings of the 10th ACM Conference on Web Science (WebSci '19)*. ACM, 225–234.

[54] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. 2019. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *The World Wide Web Conference (WWW '19)*. ACM, 818–828.

[55] Caitlin M Rivers and Bryan L Lewis. 2014. Ethical research standards in a world of big data. *F1000Research* 3 (2014).

[56] Kevin Roose. 2017. 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.

[57] Avi Rosenfeld, Sigal Sina, David Sarne, Or Avidov, and Sarit Kraus. 2016. WhatsApp usage patterns and prediction models. In *ICWSM/IUSSP Workshop on Social Media and Demographic Research*.

[58] Gandeva Bayu Satrya, Philip Tobianto Daely, and Muhammad Arif Nugroho. 2016. Digital forensic analysis of Telegram Messenger on Android devices. In *2016 International Conference on Information & Communication Technology and Systems (ICTS)*. IEEE, 1–7.

[59] Ahmad Shehabat, Teodor Mitew, and Yahia Alzoubi. 2017. Encrypted jihad: Investigating the role of Telegram App in lone wolf attacks in the West. *Journal of Strategic Security* 10, 3 (2017), 27–53.

[60] Philipp Singer, Fabian Flöck, Clemens Meinhart, Elias Zeitfogel, and Markus Strohmaier. 2014. 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*. 517–522.

[61] Statista. 2019. Most popular global mobile messenger apps as of October 2019, based on number of monthly active users. https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/.

[62] Statista. 2020. Leading countries based on number of Twitter users as of April 2020. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.

[63] Rebecca Tan. 2017. Terrorists' love for Telegram, explained. https://www.vox.com/world/2017/6/30/15886506/terrorism-isis-telegram-social-media-russia-pavel-durov-twitter.

[64] Tardaguila, Cristina and Benevenuto, Fabricio and Ortellado, Pablo. 2018. Fake News Is Poisoning Brazilian Politics. WhatsApp Can Stop It. https://www.nytimes.com/2018/10/17/opinion/brazil-election-fake-news-whatsapp.html.

[65] Telegram. 2020. 400 Million Users, 20,000 Stickers, Quizzes 2.0 and €400K for Creators of Educational Tests. https://telegram.org/blog/400-million.

[66] Telegram. 2020. Telegram API. https://core.telegram.org/method/channels.joinChannel.

[67] Twitter. 2020. Twitter's Search API. https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.

[68] Twitter. 2020. Twitter's Streaming API. https://developer.twitter.com/en/docs/tweets/filter-realtime/overview.

[69] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. 2009. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*. 37–42.

[70] Ahmet S Yayla and Anne Speckhard. 2017. Telegram: The mighty application that ISIS loves. *International Center for the Study of Violent Extremism* (2017).

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

[72] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. 2018. On the Origins of Memes by Means of Fringe Web Communities. In *Proceedings of the Internet Measurement Conference 2018*. ACM, 188–202.

[73] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtelris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn.

2017. 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*. ACM, 405–417.

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

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