It's a Trap! Detection and Analysis of Fake Channels on Telegram

Massimo La Morgia[1], Alessandro Mei[1], Alberto Maria Mongardini[1], and Jie Wu[2]

[1]*Department of Computer Science, Sapienza University of Rome, Italy, Email: {lamorgia, mei, mongardini}@di.uniroma1.it*
[2]*Department of Computer Science and Information Sciences, Temple University, United States, jiewu@temple.edu*

***Abstract*—Telegram is a widely used instant messaging app that has gained popularity due to its high level of privacy protection and social network features like channels, which are virtual rooms where only administrators can post and broadcast messages to all subscribers. However, these same features have also led to the emergence of problematic activities and a significant number of fake accounts. To address these issues, Telegram has introduced verified and scam marks for channels, but only a small number of official channels are currently marked as verified, and only a few fakes as scams.**

In this research, we conduct a large-scale analysis of Telegram by collecting data from 120,979 different public channels and over 247 million messages. We identify and analyze fake channels on Telegram. To automatically detect fake channels, we propose a machine learning model that achieves an accuracy of 85.49%. By applying this model to our dataset, we find the main targets of fakes are political figures, well-known people such as actors or singers, and services.

***Index Terms*—Telegram, Web services, Fake channels, Instant messaging app, Automatic detection**

I. Introduction

Telegram is likely the most controversial instant messaging platform. While it gives voice to dissidents in countries without freedom of speech [47], in Indonesia, terrorists used Telegram to promote radicalism and provide instructions for carrying out attacks [37]. Neo-Nazi groups leverage Telegram to share their ideologies [36]. Crypto investors coordinate large groups to arrange market manipulations like pump and dump frauds [22]. These activities were carried out by exploiting a social network feature of Telegram: the channels. They are virtual rooms where only the administrator can write and broadcast the messages to their subscribers. However, just like what happens with fake accounts on online social networks [9, 48], fake channels are widespread in Telegram. As a fake account, a fake channel impersonates an important service or person. A fake channel, to deceive the users, usually has the exact name of the target or a slight variation of it (*e.g.,* presence of emoji in the title). It attempts to qualify itself as an official using words such as official, real, and verified or adding the verified mark on the profile image. Indeed, by leveraging the popularity and influence of a notorious company or a person, the fake channel quickly obtains a considerable number of subscribers and can begin to perform fraud or scams, spam, or spread new ideologies. Significant cases of fake channels and their dangers were those created to impersonate Coinbase [1] and Kraken [3], two popular cryptocurrency exchange sites. Here, the admins used fake channels to perpetrate scams and account takeovers. The high number of users following fake channels on Telegram raises the need to develop a specific platform detection model to warn users about possible malicious behaviors. The problem is even more notable if we consider that Telegram has become more popular daily, and, as we noticed, the verified channels are still few on the platform. To perform our study, we built two datasets: the TGDataset and the Fake Channel dataset. The first dataset, which we publicly release [21], includes over 120,000 channels gathered over a one-year period, while the second is a manually curated dataset containing only verified and fake channels. We leverage the Fake Channel dataset to understand distinctive features of verified and official channels and train a machine learning model able to detect fake channels with an F1-score higher than 85%. Then, we further assess our model on the English channel of the TGDataset. By performing a qualitative analysis of the discovered fake channels, we are able to determine the most preferred target of the fake channels and their goals. We discover that fake channels are exploited by political movements like QAnon and Sabmyk to spread their conspiracy theories. Our main contributions are the following:

- **TGDataset.** We build a new dataset made of 120,979 channels. To the best of our knowledge, TGDataset is the first collection of Telegram channels that take a snapshot of the actual Telegram ecosystem instead of focusing on a particular topic. Moreover, we release our resource [21] publicly to help researchers in further investigations.
- **Fake channels characterization**. We study the phenomenon of fake channels on Telegram, performing quantitative and qualitative analyses. Through our study, we are able to understand that fake channels mainly target political figures to spread new ideologies, sell goods and promote other channels. Moreover, we notice that although fake channels usually have fewer subscribers than their official counterparts, they still reach a large audience.
- **Fake channels detection.** We analyze the problem of fake channels detection on Telegram, comparing it with the fake accounts in other Online Social Networks. We propose three machine learning models able to detect fake channels with a weighted F1-score of 85.45%.

II. Background and Related Work

A. Telegram

Telegram is a popular instant messaging platform that started in 2013, with more than half a billion active users by 2021 [35]. On Telegram, users can share text messages, images, videos, audio, stickers, and files weighing up to 2 GB. Aside from the standard one-to-one messaging, Telegram provides group chats and channels. Both have a unique username on the platform, a title, and a description, and they can be private or public. While groups allow many-to-many messaging (any member can write) and have a limit of 200,000 members, channels provide one-to-many communication (only admins can post content) and unlimited subscribers. Moreover, channels do not show info about the subscribers, except the total number. Although they serve different purposes, private chats, groups, and channels are not isolated but linked through message forwarding. This functionality allows users and administrator's channels to forward content posted in a chat to a different user, group, or channel showing the author of the original message. In particular, Telegram channels are an effective solution for spreading information to a large pool of people. Indeed, several institutional public figures and companies opened an official Telegram channel to broadcast announcements and news [38]. Likewise, start to pop up on the platform channels aiming to impersonate official channels or leverage Telegram channels and groups to sell fake products or services. Telegram introduced the *verified* and the *scam* marks to face this phenomenon. Channels, groups, and bots can achieve the verified mark proving to Telegram that the profile has the verified status on at least two social media platforms (*e.g.,* TikTok, Facebook, Twitter, Instagram) [26]. Instead, Telegram flags a channel or a group as a scam if several users report it for fraud [30].

B. Telegram channels analysis

Several works focused on the Telegram ecosystem or emerging research issues related to it. Hashemi et al. [18] collect Iranian channels and groups on Telegram to identify high-quality groups, such as business groups, among low-quality groups (e.g., dating groups). They show that high-quality groups distinguish themselves from low-quality ones through longer messages and more user engagement. Nobari et al. [13] present a structural and topical analysis of messages posted on Telegram on a dataset of more than 2,000 groups or channels. This study indicates that there is no correlation between the Page Rank of channels or groups and their number of subscribers. Baumgartner et al. [6] publish a dataset of over 27,800 thousand channels and 317 million messages from 2.2 million unique users. Their dataset includes a wide range of right-wing extremist groups and protest movements. In their work, Weerasinghe et al. [44] reveal that Telegram hosts several organized groups, called pods, where each member interacts with each other's content to increase the popularity of their Instagram accounts. Other works [49, 22, 24] reveal a vast presence on Telegram of channels and groups focused on pump and dump, a cryptocurrency market manipulation. Finally, several studies focus on the activity of terrorist organizations, like ISIS, that utilize Telegram for disseminating content and recruiting followers [10, 50].

C. Fake accounts on other OSNs

Fake accounts are widespread in Online Social Networks [9, 48]. The meaning of fake account is broad as it indicates deception contained in its content and personal information [12]. Thus, fake accounts represent several types of accounts aiming to deceive a user for different purposes. These goals can be spamming, malware distribution, impersonating people, and creating artificial interaction on the platform, for instance, using bot accounts to increase the followers of the target account [11, 42]. Several works address the problem of fake accounts, especially on Twitter. Ershain et al. [14] study the fake Twitter accounts that do not belong to a real human. They propose a classifier using features based on user behavior, such as the number of tweets, the number of accounts followed, and the number of followers. The underlying idea of their classifier is that humans behave differently. A very similar problem is the one related to Bot detection on Twitter. This task is also addressed in PAN, a series of scientific events and shared tasks on digital text forensics and stylometry [20]. In the PAN context, a classifier can rely only on stylometric features to detect bot accounts, achieving an F1-score higher than 90% on multilingual settings [4]. Instead, Caruccio et al. [11] focus on the problem of fake followers, fake accounts created specifically to increase the number of followers of a target account. The author's technique relies on the Relaxed Functional Dependencies to discriminate fake accounts from real ones. Also do Cresci et al. [12] face the problem of fake followers in Twitter. After evaluating the most relevant features and rules exploited in the Twitter fake accounts detection, they discovered that it is possible to detect with high accuracy fake followers using lightweight features such as profile information and the ratio between followers and following accounts. Gupta et al. [16] addresses the problem of detecting fake accounts on Facebook. The authors propose a classifier based on features related to user activity, such as likes and comments posted, which can detect fake accounts with an accuracy of 79%. Bilge et al. [7] shows the threats of fake accounts on Facebook. In this study, the authors forge fake accounts of the target victims using public information. Then, they send a friend request to the victim's contacts from the fake account, observing that the contacted victim trusts the request of the fake account.

III. Data collection

A. The TGDataset

Existing Telegram datasets are designed for specific studies. Thus, they contain only channels related to a particular topic [19, 6] or country [18]. Conversely, our work aims to study the phenomenon of fake channels on the Telegram ecosystem. Thus, we need a dataset representing an actual snapshot of Telegram covering many popular and connected channels. For these reasons, we build the TGDataset.

Dataset construction. To explore Telegram and, in particular, the most popular and connected channels, we use a snowball approach, as previously done in [6]. We start from a list of seed channels covering different topics and expand the dataset by adding, for every forwarded message in the seed channels, the original channel of the message. To select the seed channels, we leverage Tgstat [34], a popular service that indexes more than 150,000 Telegram channels and collects statistics about them. Although Tgstat does not offer free APIs to collect the indexed channels, it freely reports the rank of the top 100 channels by the number of users. From this rank, we retrieve all the categories to which these channels belong, finding the 18 categories shown in Tab. I.

Then, we select as seeds the 10 most popular channels by the number of subscribers from each category. Overall, we obtain a total of 180 seed channels. From each seed channel, we download the last 10,000 messages through the Telethon APIs [41], an open-source Python wrapper of the official Telegram APIs. Although a channel can contain more than 10,000 messages, we decide not to download more than that. Indeed, even though Telegram's API does not have a hard limit on the number of messages that can be retrieved, the platform actively discourages the retrieval of large amounts of messages, delaying requests when retrieving more than 3,000 historical messages [39]. Since 10,000 messages cover the entire history of more than 97.84% channels, we prefer to limit the number of requests to avoid flooding the Telegram services with further requests that go beyond our primary goals. After downloading the data, we parse the messages to discover new channels analyzing the forwarded messages. Finally, to further expand the TGDataset, we use the newly discovered channels as new seeds and iterate the above-described procedure.

Data retrieved. Data collection started on 4 January 2021 and ended on 31 July 2022. Overall, the TGDataset is 235 GB in size and contains 247,662,141 messages and 120,979 different channels. Among the channels, 656 (0.53%) are verified channels, and 184 (0.15%) are scam channels. From each channel, we store the following information: The title, the description, the channelID, the creation date, the number of subscribers, and if it is marked as a scam or verified. Concerning messages, we store the channelID, the timestamp, and, in case of forwarded messages, the original channelID where the message has been posted, and the original posting date. Finally, we store the content of the text messages, while just the title and the file format of the media messages.

B. The Fake Channels dataset

To understand the main differences between fake and official channels and later train a machine learning model able to detect fake channels, we build a dataset of channels whose status (official or fake) is known with certainty. To this respect, we create the Fake Channels dataset. To build it, we use the following approach: We first leverage the *Telemetr.io* [40] services to retrieve a list of verified channels. Then, for each verified channel, we look for fake channels claiming to be the official ones, taking care to not select fan channels. At the end of this process, the Fake Channels dataset consists of 342 different channels, 184 of which are officials and 158 fakes. While selecting the channels, we ensure they are not already present in the TGDataset. In this way, we can use the Fake channel dataset as training data while developing our detector.

IV. Fake channels detection

A. Analysis of the Fake Channels dataset

As a first step toward constructing our detector model, we separately analyze the fake and verified channels contained in the Fake Channels dataset, and we use the channels of TGDataset as a reference of the average behavior of the Telegram channels. Although the TGDataset contains verified and fake channels, given its vast number of channels, we believe it can represent very well the behavior of standard Telegram channels.

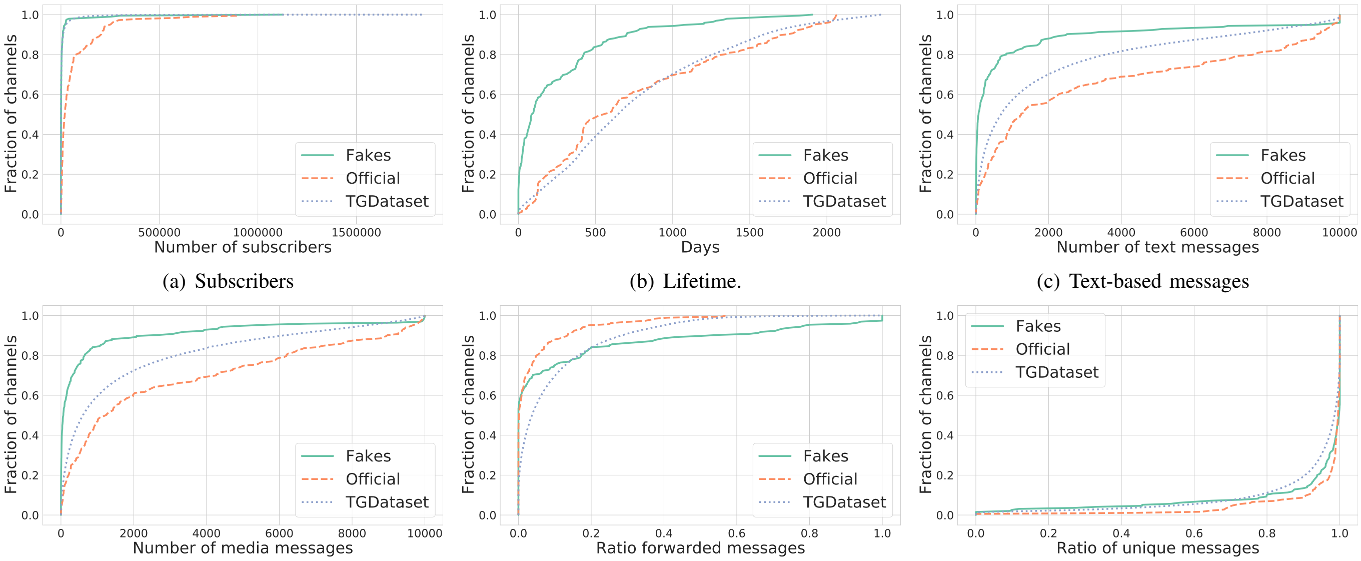
We start by studying the number of subscribers of the three sets of channels taken into account, showing them in Fig. 1(a). As we can expect, verified channels (dashed orange line), in general, have more subscribers than fake (green line) and standard channels (dotted blue line). In contrast, fake and standard channels have very similar distributions. Comparing the number of subscribers between the verified channels and their fake version, we notice that the fake channels have, on average, 10% of the number of subscribers of the corresponding verified channel. However, in our dataset, we have two cases in which the fake channels have more subscribers than the verified one. Both cases are related to *@AnuragxCricket*, a channel of the Indian fantasy cricket influencer Anurag Dwivedi. Here, the verified channel has 280,212 subscribers, while its fakes *@AnuragCricket* and *AnuragxCricket_team* have 301,742 and 1,126,330 subscribers respectively. A possible reason behind the success of the first fake could be that it was created on 2019-10-07, more than one year before the verified channel (2021-03-10). Instead, the second and bigger fake channel was created one month after (2021-04-25) the verified one. Thus this abnormal number of subscribers is less explainable. We conjecture that the fake channel achieved this success by leveraging some promotional services or the help of other fake channels, as we notice in Sec. V-A. However, we can not confirm this suspect as we do not find evidence in our dataset.

Then, we proceed with the lifetime of the channels. We define the lifetime of a channel as the time elapsed between its creation and its last message. As shown in Fig. 1(b), fake channels have a shorter lifetime (average 251.85 days) than verified (average 750.39 days) and standard channels (average 764.02 days), whereas these last two kinds of channels have similar duration. This result suggests that fakes cease to post content at a certain point as they may have been discovered or because they have reached their goals.

Finally, we analyze the type of messages shared by the channels. Fig. 1(c) and Fig. 1(d) reveal that verified channels tend to share more messages, both text-based or media-based, than the standard Telegram channels and fake channels. Verified

Table I
THE 18 CATEGORIES TO WHICH BELONG THE MOST POPULAR 100 CHANNELS ACCORDING TO TGSTAT.

Categories retrieved
Sales, Humor & entertainment, News & Mass media, Video & Movies, Business & Startups, Cryptocurrencies, Politics, Technologies, Sport, Marketing, Economics, Games, Religion, Software & Applications, Lifehacks, Fashion & Beauty, Medicine, Adults



(a) Subscribers

(b) Lifetime.

(c) Text-based messages

(d) Media-based messages.

(e) Forwarded messages

(f) Unique messages.

Figure 1. CDFs of: the number of subscribers for fake, verified and TGDataset channels (1(a)), the lifetime of the channels (1(b)), the number of text-based and media-based messages (1(c) and 1(d), respectively), ratio of forwarded messages (1(e)), and the ratio of unique messages (1(f)).

channels post on average 3,176.46 text messages and 2,892.27 media content, while fake and standard channels post 1,036.59 and 2,030.05 text messages and 862 and 1,817.82 media, respectively. The fewer messages shared by fake channels are aligned with their short life. Instead, verified channels have a lifetime similar to standard channels. Thus, the abundant number of content they produce could be a suitable feature for our classifier. A distinctive feature of fake channels is the number of forwarded messages. Fig. 1(e) shows the ratio between the forwarded messages by the channels and the total number of messages shared. As we can see, while the verified channels tend to forward few messages, fake channels are more prone to forward messages from other channels, with a fraction of fake channels (approx 18%) extensively using this Telegram functionality. Lastly, we investigate the ratio of distinct messages published over the total number of messages published by the channels (Fig. 1(f)). Here, we notice that all three kinds of channels mostly produce fresh content, with both the fake and verified channels more active in producing new content than the standard channels.

B. Features

Despite having some common traits with other OSNs' accounts, Telegram channels present limited social interaction functionalities. A key difference is that in OSNs an account can interact with others, such as commenting content of other accounts, following other accounts, appreciating content generated by other users (*e.g.,* likes), and republishing content (*e.g.,* retweeting). Instead, a Telegram channel can only post content in its channel and can not interact with anyone outside of it (*e.g.,* subscribing to other channels or texting private messages to users). Moreover, it is virtually impossible to interact with the content generated by the channels. Indeed, even if Telegram recently added the functionality to comment or react with emoticons to the content of a channel, we observe that this feature is enabled only by a tiny fraction of channels. Unlike other OSNs, Telegram discloses only the number of channel subscribers, not the list of subscribing accounts. These differences make unavailable the use of the most discriminating features to detect fake accounts on other OSNs, such as the ratio between the number of users following the account (usually low) and the number of users followed by the fake account (usually high) [11, 12] or the number of likes (given or received). In addition, some features are unique to a particular platform (*e.g.,* Twitter list or usage of Facebook application) and, therefore, cannot be used in our scenario. Nevertheless, we can adapt some features used in the previous works (e.g., biography could be considered the description of a channel) on Telegram channels and evaluate them in our scenario.

Regarding the other classification work on Telegram [18], the authors focus on detecting high-quality groups. Even in this case, we cannot utilize all their features due to differences between the channels and groups. In groups, every user can post a message like in a chat room, the list of group members is accessible, and the personal accounts of group administrators are disclosed. Conversely, in channels, only the administrator can post, and the accounts of both subscribers and channel administrators are not visible.

To build our classifier to detect fake channels, we evaluate all the features used in previous works and reproduce them in the context of Telegram channels. Moreover, we also consider what we learned in the previous subsection (*e.g.,* number of text messages published, ratio of forwarded messages) and new features specifically for this task. We tried several sets of features to build our model. In the following, we describe the features that achieved the best performance.

- **Writing style features:** average message length, average number of emojis per message, average number of non-alphanumeric characters per message, number of characters in the title and description, and average number of non-alphanumeric characters in the channel's title.
- **Temporal features**: number of text messages published in the last 3, 6, 9 months and average posting time between two consecutive messages.
- **External interaction features**: number of forwarded messages, standard deviation of the number of source channels for the forwarded messages, number of shared links, and number of duplicate messages containing at least one link.

C. *Classifiers and results*

We use the features described above to train three different models: a Random Forest classifier [8], an SVM with Linear kernel [31], and a Multilayer Perceptron (MLP) [15]. Moreover, to better assess our models, we implement two baselines. Since there are no studies dealing with fake Telegram channels, we select as the first baseline the Twitter fake account classifier that leverages the highest number of features that can also be implemented on Telegram. It is the classifier proposed by Cresci et al. [12], for which we can adapt 9 features. As the second baseline, we chose the classifier of Hashemi et al. [18] to detect high-quality groups on Telegram. Also in this case, we use only the available features on Telegram channels. To implement all the models except for the MLP, we use the Sklearn [28] Python library and tune the hyper-parameter through grid search. Instead, to implement the MLP classifier, we use Pytorch [27]. The MLP classifier is made of three linear layers with Rectified Linear Unit function (ReLU) [17] as the activation function, the Adam optimization algorithm [51] as the optimizer, and binary cross-entropy (BCE) [23] as the loss function.

We assess the models' performances through 5-fold cross-validation [2] using the weighted F1-score as the evaluation metric. Table II reports the results we achieve by the 5 different models. As we can see, the models based on the proposed features outperform the two baselines. The model that performs worst, slightly better than a random classifier (54.54% F1-score), is the one replicating the results of Cresci et al.. This result is quite expected, given the differences between the Twitter and Telegram platforms. Instead, the model proposed by Hashemi et al. achieves a weighted F1 score of 72.16%. Through the analysis of the results, it is possible to note that the precision (66.94%) and the recall (85.68%) of this classifier are unbalanced. This is due to the model's tendency to classify channels as fakes. Inspecting the weight of the features, we observe that the classifier assigns a high weight to the number of subscribers, leading to classify as fake channels with a low number of subscribers. Finally, we have the three different classifiers based on the features proposed in this work. The MLP model is the classifier that performs better, achieving an F1-score of 85.45%, outperforming the best baseline of 13 percentage points, and obtaining a good trade-off between precision and recall. Instead, both the Random Forest and the SVM model perform slightly worst than the MLP model, achieving an F1-score of 80.35% and 81.01%, respectively, but better than the baselines.

Table II
5-FOLD CROSS-VALIDATION CLASSIFICATION RESULTS.

Model	**Precision**	**Recall**	**F1 weighted**	**Accuracy**
Cresci et al.	52.94%	56.25%	54.54%	55.07%
Hashemi et al.	66.94%	85.68%	72.16%	72.79%
Random Forest	82.05%	81.03%	80.35%	81.03%
SVM linear	81.77%	81.06%	81.01%	81.62%
MLP	84.24%	85.86%	85.45%	85.49%

V. DISCOVERING FAKE CHANNELS IN THE WILD

Selection of suspicious channels. After validating our classifier, we leverage it to detect fake channels on the TGDataset. For this task, we consider only English channels, so that we can validate the channels and perform qualitative analysis. To select English channels, we perform language detection. To this end, we pre-process the messages by normalizing and polishing them. In particular, for each channel, we take into account only the pure text messages, remove mentions and get rid of numbers, hyperlinks, emoji, and messages shorter than 15 characters as they could compromise the accuracy of the tool [33, 5]. Then, we tokenize the messages using the *RegexpTokenizer* developed by NLTK [25] and provide them as input to the tool. At this point, to detect the languages of the channels, we leverage LangDetect [32], a language detection library implemented by Google with precision over 99% for 53 languages. At the end of the process, we get 21,078 English channels that account for 17.54% of the TGDataset. Hence, we collect the channels that have in their title, description or username the words *real*, *official* or *verified*. To further expand the dataset, we consider all the channels with a similar name (edit distance less than 2) to one of the verified channels. Also in this case, we manually inspect these channels to ensure they are not fan channels. In the end, we collected a set of 511 channels.

Channels evaluation. Since we do not have a ground truth for this set of channels, we check all of them manually to

Table III
RESULTS OF THE MLP CLASSIFIER ON THE TGDATASET.

	Label			
Prediction	Fake	Official	All. fake	All. official
Fake	88	28	142	0
Official	9	103	0	141

assess the results. In particular, we consider a channel:

- **Official**: if Telegram marked it as verified or there exists an official source (*e.g.,* Website, Facebook, Instagram, Twitter) of the person/service indicating the Telegram channel as the official one.
- **Fake**: if there is another channel that we consider official with the same name or an official source states that there is no official Telegram channel.
- **Allegedly fake/official**: if our classifier detects the channel as fake/official, but there is no evidence of their status. In particular, there are no channels with the same or a similar name that we consider official and the related official web pages or social media pages do not mention any Telegram channel.

Results. Tab. III reports the results we obtain after the manual investigation. Globally, we mark as fakes or officials 228 channels out of 511. In particular, among the 258 channels recognized as fakes by our model, there are 88 fakes, 142 allegedly fakes, and 28 official. Among the channels classified as official, 103 are actual official channels, 141 are allegedly official, and 9 are fakes. Thus, for the channels we have evidence of their status, our classifier was able to classify 191 channels out of 228 correctly, equivalent to an accuracy of 83.77%, which aligned with the results obtained in the cross-validation.

A. Studying fake channels

Fakes targets. The majority of the channels we verified to be fake target real people (76 out of 97). Among them, the most targeted categories are politicians (59), including nine claiming to be Donald Trump, and 17 celebrities (*e.g.,* influencers, actors, and athletes). Moreover, ten fake channels emulate news services, and seven are crypto-related services. Finally, we find four fakes pretending to be well-known companies.

Effectiveness of the fake strategy. A suitable metric for understanding fake channels' effectiveness is to examine the number of subscribers they have attracted. It emerged that the fake strategy is very effective since fakes have an average of 19,636.31 subscribers and more than 45% of them have more than 10,000 subscribers.

The goal of Fake channels. After understanding the target of the fakes, we manually inspect these channels. It turns out that 32 fakes seem to have the goal of spreading conspiracy theories, such as QAnon [45], but also new ones, like Sabmyk [43]. The latter is a conspiracy theory that proposes itself as a better alternative to QAnon and promotes a singular quasi-religion centered around a messianic figure known as Sabmyk [29]. In particular, we find 23 fake channels posting content about Sabmyk that likely belong to a greater network (about a hundred channels) spreading Sabmyk's messages according to the *"HOPE not hate"* organization [43]. Other 14 fake channels mainly advertise. There are eight fakes focused on promoting other channels sharing their invitation links and forwarding their messages. Lastly, one fake asks for funds to be sent to a wallet on Monero, a cryptocurrency focused on private and censorship-resistant transactions [46].

Status of fakes and officials. Among the 126 official channels found within the TGDataset, only 70 (55.55%) are marked as verified by Telegram. Nevertheless, there are several channels that we presume are official upon careful manual analysis but that neither appear to be verified by Telegram nor have a link to the channel on their social pages or website. Instead, the fakes marked as a scam by Telegram are only 9 out of 82 (8.53%).

VI. ETHICAL CONSIDERATIONS

In this work, we analyzed 120,979 channels on Telegram . During the data collection, we put particular effort into collecting only data belonging to Telegram's channels. Thus, neither user's personal data (username, phone number, subscribed channels) nor standard users' messages were collected. Consequently, according to our IRB's policy, we did not need any explicit authorization to perform our experiments.

VII. CONCLUSIONS AND FUTURE WORKS

In this paper, we faced the problem of fake channels on Telegram. We characterize this channel type and analyze how admins of fake channels take advantage of them. We propose a machine learning model with an F1-score of 85% in detecting fake accounts. Running our detector on a subset of TGDataset, we found 258 allegedly fake accounts, of which we could confirm 88. Given the extent of the phenomenon, the high number of subscribers, and the difficulty of distinguishing fake channels from official ones, the need for institutions, famous people, and organizations to obtain verified status for their channels is on the rise. Indeed, we notice only a few official channels leverage this opportunity.

With this work, we shed light on one of the several controversial activities running on the Telegram platform. However, we believe further investigations are needed to illuminate the Telegram ecosystem completely. Indeed, in our research, we noticed a heavy presence of channel networks that attempt to spread conspiracy theories by exploiting fake channels. Thus, it is interesting to understand how these networks are organized, how they evolve over time, and which is their target audience.

ACKNOWLEDGMENTS

This work was partially supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

References

[1] *Anatomy of a telegram scam*. https://blog.coinbase.com/anatomy-of-a-telegram-scam-9fd3dfb8c310.

[2] Davide Anguita et al. "The 'K'in K-fold cross validation". In: *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*. i6doc. com publ. 2012, pp. 441–446.

[3] *Another Phishing scam 'Kraken Official Telegram Channel'*. https://steemit.com/cryptocurrency/@techstack/another-phishing-scam-kraken-official-telegram-channel.

[4] Andrea Bacciu et al. "Bot and gender detection of Twitter accounts using distortion and LSA. Notebook for PAN at CLEF 2019". In: *Working Notes Papers of the CLEF 2019 Evaluation Labs volume 2380 of CEUR Workshop*. 2019.

[5] Timothy Baldwin and Marco Lui. "Language identification: The long and the short of the matter". In: *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. 2010, pp. 229–237.

[6] Jason Baumgartner et al. "The Pushshift Telegram Dataset". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 14. 2020, pp. 840–847.

[7] Leyla Bilge et al. "All your contacts are belong to us: automated identity theft attacks on social networks". In: *Proceedings of the 18th international conference on World wide web*. 2009, pp. 551–560.

[8] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[9] Qiang Cao et al. "Aiding the detection of fake accounts in large scale social online services". In: *9th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 12)*. 2012, pp. 197–210.

[10] Zhenfeng Cao et al. "Dynamical patterns in individual trajectories toward extremism". In: *Available at SSRN 2979345* (2017).

[11] Loredana Caruccio, Domenico Desiato, and Giuseppe Polese. "Fake account identification in social networks". In: *2018 IEEE international conference on big data (big data)*. IEEE. 2018, pp. 5078–5085.

[12] Stefano Cresci et al. "Fame for sale: Efficient detection of fake Twitter followers". In: *Decision Support Systems* 80 (2015), pp. 56–71.

[13] 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.

[14] Buket Erşahin et al. "Twitter fake account detection". In: *2017 International Conference on Computer Science and Engineering (UBMK)*. IEEE. 2017, pp. 388–392.

[15] Matt W Gardner and SR Dorling. "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences". In: *Atmospheric environment* 32.14-15 (1998), pp. 2627–2636.

[16] Aditi Gupta and Rishabh Kaushal. "Towards detecting fake user accounts in facebook". In: *2017 ISEA Asia Security and Privacy (ISEASP)*. IEEE. 2017, pp. 1–6.

[17] Kazuyuki Hara, Daisuke Saito, and Hayaru Shouno. "Analysis of function of rectified linear unit used in deep learning". In: *2015 international joint conference on neural networks (IJCNN)*. IEEE. 2015, pp. 1–8.

[18] Ali Hashemi and Mohammad Ali Zare Chahooki. "Telegram group quality measurement by user behavior analysis". In: *Social Network Analysis and Mining* 9.1 (2019), pp. 1–12.

[19] Mohamad Hoseini et al. "On the Globalization of the QAnon Conspiracy Theory Through Telegram". In: *arXiv preprint arXiv:2105.13020* (2021).

[20] Mike Kestemont et al. "Overview of the Cross-domain Authorship Attribution Task at PAN 2019". In: *CLEF 2019 Labs and Workshops, Notebook Papers*. Ed. by Linda Cappellato et al. CEUR-WS.org, Sept. 2019.

[21] Massimo La Morgia, Alesssandro Mei, and Alberto Maria Mongardini. *TGDataset*. 2023. URL: https://github.com/SystemsLab-Sapienza/TGDataset.

[22] Massimo La Morgia et al. "Pump and Dumps in the Bitcoin Era: Real Time Detection of Cryptocurrency Market Manipulations". In: *2020 29th International Conference on Computer Communications and Networks (ICCCN)*. IEEE. 2020, pp. 1–9.

[23] Shie Mannor, Dori Peleg, and Reuven Rubinstein. "The cross entropy method for classification". In: *Proceedings of the 22nd international conference on Machine learning*. 2005, pp. 561–568.

[24] Massimo La Morgia et al. "The Doge of Wall Street: Analysis and Detection of Pump and Dump Cryptocurrency Manipulations". In: *ACM Trans. Internet Technol.* (2022). Just Accepted. ISSN: 1533-5399. DOI: 10.1145/3561300. URL: https://doi.org/10.1145/3561300.

[25] *NLTK RegexpTokenizer*. https://www.nltk.org/_modules/nltk/tokenize/regexp.html. 2021.

[26] *Page Verification Guidelines*. https://telegram.org/verify. 2022.

[27] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035.

[28] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[29] *Sabmyk Network: Founder of bizarre new religion targeting QAnon believers 'unmasked' by Hope Not Hate*. https://www.independent.co.uk/news/world/europe/sabmyk-network-qanon-conspiracy-theories-b1820639.html.

[30] *Scammers in telegram and how to report*. https://www.telegramadviser.com/scammers-in-telegram-and-how-to-report/.

[31] Bernhard Schölkopf et al. “New support vector algorithms”. In: *Neural computation* 12.5 (2000), pp. 1207–1245.

[32] Nakatani Shuyo. *Language Detection Library for Java*. 2010. URL: http://code.google.com/p/language-detection/.

[33] P Sibun and JC Reynar. “Language determination: Examining the issues”. In: *Proceedings of the 5th Annual Symposium on Document Analysis and Information Retrieval*, pp. 125–135.

[34] *Telegram Analytics*. https://tgstat.com/. 2021.

[35] *Telegram FAQ*. https://telegram.org/faq\#q-what-is-telegram-what-do-i-do-here. 2021.

[36] *Telegram the latest safe haven for white supremacists*. https://www.adl.org/blog/telegram-the-latest-safe-haven-for-white-supremacists. 2019.

[37] *Telegram to block terror channels after Indonesian ban*. https://www.bbc.com/news/business-40627739.

[38] *Telegram, the powerful COVID-19 choice of communications by many governments*. https://www.channelnewsasia.com/commentary/coronavirus-covid-19-government-telegram-whatsapp-fake-news-info-936061.

[39] *TelegramClient*. https://docs.telethon.dev/en/latest/modules/client.html?highlight=iter_messages\#telethon.client.messages.MessageMethods.iter_messages.

[40] *TelemeterIo*. https://telemetr.io/en/channels. 2021.

[41] *Telethon’s Documentation*. https://docs.telethon.dev/en/latest/.

[42] Kurt Thomas et al. “{Trafficking} Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse”. In: *22nd USENIX Security Symposium (USENIX Security 13)*. 2013, pp. 195–210.

[43] *Unmasked: the QAnon ‘messiah’*. https://www.hopenothate.org.uk/unmasked-the-qanon-messiah//.

[44] Janith Weerasinghe et al. “The pod people: Understanding manipulation of social media popularity via reciprocity abuse”. In: *Proceedings of The Web Conference 2020*. 2020, pp. 1874–1884.

[45] Mike Wendling. *QAnon: What is it and where did it come from?* https://www.bbc.com/news/53498434. 2021.

[46] *What is Monero (XMR)?* URL: https://www.getmonero.org/get-started/what-is-monero/.

[47] *Why journalists and dissidents turn to Telegram*. https://www.indexoncensorship.org/2021/06/telegram/. 2021.

[48] Cao Xiao, David Mandell Freeman, and Theodore Hwa. “Detecting clusters of fake accounts in online social networks”. In: *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*. 2015, pp. 91–101.

[49] Jiahua Xu and Benjamin Livshits. “The anatomy of a cryptocurrency pump-and-dump scheme”. In: *28th {USENIX} Security Symposium ({USENIX} Security 19)*. 2019, pp. 1609–1625.

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

[51] Zijun Zhang. “Improved adam optimizer for deep neural networks”. In: *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. IEEE. 2018, pp. 1–2.