Dark Web Dialogues: Analyzing Communication Platform Choices of Underground Forum Users

Raphael Hoheisel*, Tom Meurs*, Marianne Junger*, Erik Tews†, Abhishta Abhishta*
Email: r.e.hoheisel@utwente.nl, t.w.a.meurs@utwente.nl, m.junger@utwente.nl, e.tews@utwente.nl, s.abhishta@utwente.nl
**BMS, HBE University of Twente* Enschede, The Netherlands
†*EEMCS, SCS University of Twente* Enschede, The Netherlands

***Abstract*—We investigate the utilization of private communication platforms by underground forum users. We aim to bridge the knowledge gap regarding user preferences for communication platforms employed for private conversations within illicit contexts. We employ social network analysis, topic modeling and statistical analysis on over 7.5 million posts and 260 thousand messages from a popular underground forum. We identify prevalent communication platforms and investigate the relationship between the context in which users share contact details and their social networks in relation to platform preferences. Our contributions include an overview of prominent communication platforms used by forum users, highlighting Telegram's predominant popularity. We show that in hacking related topic users choose platforms that provide higher security and privacy levels. Lastly, findings from our statistical model indicate a significant relationship between the centrality of users in the social network and their choice of communication platform. We provide valuable insights for law enforcement agencies, helping them make strategic decisions and plan interventions combating cybercrime.**

***Index Terms*—underground forums, hacking forums, cybercrime, communication, messengers, social network analysis**

I. Introduction

Pathways into (cyber-)crime involve communication between actors, to gain knowledge, make new connections or to coordinate criminal activities.

In recent years, cyber-attacks have become increasingly coordinated, with responsible groups exhibiting a growing sophistication comparable to conventional business structures [1]. The majority of interactions within these communities occur online, primarily through forums and messaging platforms [2]. For law enforcement, this underscores the necessity of gaining comprehensive insights into the communication methods and preferred platforms of individuals involved in malicious activities to effectively investigate, disrupt, and prevent such activities. Malicious actors utilize online platforms to recruit individuals for their operations, and those interested in topics such as committing online fraud, developing malware, or executing cyber-attacks can find corresponding online communities in 'underground' or 'hacking forums' [3], [4].

Underground forums provide users with malicious intent with an ideal place to find, buy, and sell illegal goods or services [5], [6]. The public posts on these forums often discuss offered services and provide valuable insights into this illicit ecosystem. Numerous studies have focused on extracting threat intelligence from these posts, linking similar users, and identifying key actors [7]–[12]. More comprehensive information regarding criminal services or user-specific details is typically exchanged through private messages on forums [5], [13] or via other private channels, such as Telegram or Jabber [14]. Sharing contact information online is therefore essential for malicious actors aiming to attract buyers, despite the associated risk of disclosing information that could aid law enforcement in identifying them. Additionally, these actors may decide to circulate their contact details at multiple platforms to facilitate easier communication with potential clients.

Rational choice theory [15] suggests that we can best understand malicious actors' decision-making as a type of bounded rationality, where they trade off profits, effort, and risks. However, these may not necessarily correspond to the long term costs (efforts) and benefits for these actors. In the context of choosing a communication platform, rational choice theory could imply that malicious actors trade off the ease-of-use and popularity of the platform against security risks and the effort required to set up a presence on the platform. We hypothesize that the decision making process follows the principles of the rational choice theory, when a malicious actor takes decisions on which contact details to share, where on the forum to share them, or whether or not to share them privately/publicly or both.

For our research, the Dutch Police provided us access to a database of a popular underground forum. The database contains over 7.500.000 public posts from the years 2015-2022 and around 260.000 private messages of users of the forum from 2020-2022. In this study, we define communication platforms as communication means that are used by forum users to communicate with each other apart from the underground forum. Examples are WhatsApp, Telegram or Discord. We begin our research with identifying the platform contact details that are shared in public posts and private messages on the underground forum. Then, we study to what extent users share contact details of multiple communication platforms. We assume that the rational behind sharing contact details to multiple platforms is to facilitate communication with users who prefer different communication platforms (higher benefit due to a wider audience). However, this comes with the cost of managing multiple platforms and leaking more information to law enforcement (higher effort/costs and risk). Therefore, we expect that only few users will share contact details for multiple platforms. With this analysis will study the following research questions:

RQ 1: Which are the most prominent (external) communication platforms on the underground forum?

RQ 2: How common is it to share contact details of multiple communication platforms?

An underground forum is not necessarily a uniform community, but can consist of sub-forums concerning different topics, such as cryptocurrencies, drugs, hacking, or malware [16]. As a consequence, we are also interested whether these sub-communities have different preferences when it comes to communication platforms. Underground forums exhibit diversity not only in the topics discussed but also in the proficiency levels of their users. Typically, only a small group of highly skilled users is present on a forum [9]. According to [17], actors are aware of the legality of their action on underground forums. Following the rational choice perspective, we assume that forum users prefer different communication platforms in different contexts (communities). In a context related to illegal services or data, users might prefer to use a more privacy and security focused communication platform, whereas, in other contexts ease-of-use or general popularity of a communication platform might be more relevant. For law enforcement, prominent and proficient users are of more interest compared to other users, which increases the risk of prosecution for such prominent actors. As a result, we assume that proficient users try to balance between signaling other users their expertise and trustworthiness and keeping a low profile. In addition, such users are likely also using more security and privacy focused communication platforms. Consequently, we consider user characteristics that may indicate their proficiency and activity levels in order to assess how these attributes influence platform preference. Using topic modeling, social network analysis and statistical analysis, we study the subsequent two research questions.

RQ 3: To what extent does platform preference depend on the context of a post or message it is shared in?

RQ 4: How do characteristics related to a user's activity help in understanding platform preferences?

We believe this is the first publication to examine the communication platform preferences of users on an underground forum with this level of detail. In the present study, we investigate which user IDs of communication platforms, such as Jabber, Telegram and WhatsApp users share on the underground forum. Additionally, we study the context in which they are mentioned and characteristics of the users mentioning them. We combine various types of analyses and techniques including, topic modeling, descriptive analysis, social network analysis, and statistical analysis.
The contributions of our study are the following:

- We identified six popular communication platforms that are used by forums users for private communication, with Telegram being the most popular one.
- We show and explain with the rational choice model how forums user weigh risks, effort and benefits when choosing communication platforms.
- Our results show that, in contexts that are closely related to malicious activities (hacking related), forum users that share contact details, choose platforms that provide a high level of security but require more effort to use them.
- Our results indicate that forum users who choose communication platforms that provide a relatively higher level of privacy are more popular (important) while having fewer public engagements on the forum.

To the best of our knowledge, this is the first work that has access to such recent data on underground forums including both public and private communication. The insights derived from our study may be highly beneficial for law enforcement agencies. These findings can assist in strategic decision-making and the planning of more effective interventions. For instance, the establishment of a task force to monitor public malware development groups on Telegram could be justified when evidence suggests that this platform is gaining traction for such activities. Understanding the general popularity of various platforms, as well as their usage within specific communities and by malicious actors of varying levels of sophistication, is crucial for these efforts.

II. Related Works

Data analysis from the dark web and hacker forums is a crucial area of research that offers insights into the cybercrime ecosystem. Researchers employ various methods to extract, analyze, and comprehend user relationships on these online platforms. Prior to our study, researchers have used topic modeling and social network analysis to characterize and find proficient users of underground forums.

[18] focused on identifying prominent figures in hacker communities through social network analysis. Utilizing data from hacker forums, their findings showed hackers that who contributed to the cognitive advancement of their community or were considerably active had the highest reputations. Further, [7] conducted in-depth analyses of cybercrime actors and activities in underground forums. Their 2018 study provided detailed characterizations of the actors involved in these forums, and in 2019, they delved into specific cybercrime tactics [19]. [20] applied natural language processing to automatically categorize the function and intent of posts in underground forums. A more recent study by [21] applied social network analysis and topic modeling to identify influential users on underground forums using the CrimeBB dataset. Their results show that a user's reputation does not necessarily reflect a users influence on the forum. In this study, we use these social network analysis to get an understanding of how influential users are and topic modeling to show how communication platform preferences differ between the context in which contact details are shared.

Various researchers investigated underground forums and their use for extracting knowledge about the ecosystem and users. [22] highlighted a high turnover of users, with a shift in forum activities from hacking to e-whoring and market-focused discussions. This trend indicates an increasing focus on financially driven cybercrime. The research, utilizing data from the CrimeBB dataset and applying latent Dirichlet allocation for topic modeling, illustrates the application of digital drift theory to explain the evolving community composition and interests.

Studies by [23], [24], dive into more personal information that users of underground forums share and how this relates to their malicious activities, namely music and disclosures of autism. These studies highlight the variety in discussions of underground forums emphasizing the relevance of these forums also for social interactions.

Regarding communication platform use in illicit environments, [25] studied the choice of social media platforms for communication in the drug market. The researchers concluded that it is an interplay between various social and personal circumstances and that there are trade-offs between security and convenience. For example, for lower level drug deals, such as for cannabis, people did not see the need for encrypted messaging platforms and chose platforms that are easier to use. In addition, the authors mention that especially for people buying drugs via an encrypted messenger, their choice was driven by the fact the the seller was only available through this platform.

While previous research focused more on the activities and influence of users and the services that are offered on underground forums, we advance the current literature by looking at services that users prefer more specifically. In this study, we investigate the contact details for private communication users share on a popular underground forum and how this correlates with the context/topic and user position in the social network.

III. Methodology

A. The Underground Forum Database

TABLE I

Newly registered forum users, public posts, and private messages per year that mention/contain a user ID (UID) of a communication platform.

Year	Newly registered users sharing UIDs	Posts UID hits	Messages UID hits
2015	47	4	0
2016	58	3	0
2017	353	156	0
2018	628	964	0
2019	1.264	4.071	0
2020	2.410	10.331	5.420
2021	2.997	24.876	8.312
2022-03	283	5.487	1.126
Total (UID)	8.040 (1%)	45.892 (0.6%)	14.858 (6%)
Total (all)	760.000	7.500.000	260.000

Access to the data we use in our research was provided by the Dutch Police. The data contains user account information, private messages, and public posts of a large popular underground forum. The data on posts spans a timeframe of 7 years (2015-2022) and for private messages 2 years (April 2020-Mar 2022). The database contains more than 7.500.000 public posts and 260.000 private messages ranging from March 2015 to February 2022. In total, there are around 760.000 registered users, however, more than two thirds of these users are active on the forum for less than one day.

TABLE II

Regular expression matches of messenger user IDs in public posts and private messages.(*) The regular expressions for these usernames are very broad so there are likely more false positives. Multiple matches per post/message are possible.

Messenger	Matches Post	Matches PM
Telegram	28.650	9.484
Discord	19.274	3.819
Skype*	7.398	652
XMPP/Jabber	6.259	2.024
WhatsApp	3.552	115
Wickr*	470	338
ICQ	838	737
Other	39	37
TOX	119	76

The language of the forum is English, however, communication in the private messages also happens in other languages, such as Russian and other European languages.

B. Commonly used Communication Platforms

To assess the rational decision making process of forum users for choosing a specific communication platform, we first describe popular communication platforms in more detail. We also highlight the differences in the platform features according to three categories we see as relevant for the forum users choice. That is security and privacy, the features set to manage larger communities, and how convenient these platforms are to use.

Telegram is a cloud-based free instant messenger that was released in 2013. It is available for the most common mobile and desktop platforms such as Android, iOS, Windows, macOS and Linux. Telegram allows to have group chats, channels, voice and video conferences as well as bots for automating services. It also supports privacy features such as end-to-end encrypted messages and self deleting messages [26].

Discord is a free instant messaging, voice and video conference platform launched in 2015. It also runs on all popular operating systems and in web browsers. Discord is mainly organized in so-called ‘servers’ which can include several chat rooms, but also allows private chats. Chats are encrypted, but not end-to-end.

Jabber also known as XMPP (Extensible Messaging and Presence Protocol), is an open, free, and flexible communication protocol designed for real-time messaging, presence information, and instant communication. Developed by the Jabber open-source community and standardized by the IETF, XMPP supports decentralized operation, meaning any user can run their own server, fostering privacy and security. It enables interoperability among different messaging systems and supports various features like voice and video calls, file transfer, and multi-user chat. Its extensible nature allows developers to add new functionalities, making it a versatile and widely used protocol in various communication platforms.

ICQ is a free instant messaging service launched in 1996. It uses a proprietary protocol for messaging, allowing users to send text messages, share files, and engage in voice and video calls. Over the years, ICQ has evolved to include modern features such as mobile app support, social networking integration, and enhanced security measures like encryption. In 2010 it was acquired from the previous owner AOL by the Russian Main.RU group, and in 2024 it was announced that ICQ will shut down later that year.

WhatsApp is a free, cross-platform messaging app that allows users to send text messages, voice messages, images, videos, documents, and make voice and video calls. It uses the internet to facilitate communication, ensuring end-to-end encryption for privacy and security. The app supports group chats, location sharing, and integration with contacts. It is widely used globally, providing a seamless communication experience across Android, iOS, and desktop platforms.

TOX Messenger is a free, open-source messaging app designed for privacy and security, offering end-to-end encrypted text messaging, voice and video calls, and file sharing. It operates on a decentralized network, eliminating the need for central servers and reducing the risk of data breaches. Tox is available on multiple platforms, including Windows, macOS, Linux, and mobile devices. Its focus on user privacy ensures no data collection, making it a reliable choice for secure and private communications.

As pointed out by [25], ease-of-use and security play a role in the choice of communication platform of people interacting in illicit markets. All six messengers, except Discord, support end-to-end encryption for texts making it impossible for the company behind the messenger to see the content. Jabber is unique in this regard as many features such as encryption of community features depend on the implementation of the specific Jabber client. To give a high level overview of the difference between these platforms, we label their features according to three categories, security, community and messaging. We do not consider features that were introduced after the end date of our database (begin 2022). Table III shows the communication platform categories. We define the categories and features as follows:

Security: Text and and voice are end-to-end encrypted, decentralized platform, setup can be done anonymously by default.
Community: Allows automation via bots, allows large groups (over 1000 members), community features such as sub-groups, moderation, and member verification.
Messaging: High popularity in the general public, one main application that can be downloaded for different platforms.

C. Data Preprocessing

We apply the preprocessing steps mentioned in Figure 1.

1: We extract metadata related to user, posts, and messages from the forum database.

2: We extract posts and messages from the database.

3: In the initial preprocessing stage, we remove citations and HTML tags from the text of both posts and messages. We focus our analysis exclusively on private conversations between two users, excluding messages users send to themselves. Additionally, we remove system notifications which user receive as private messages (∼40k messages). We assume these messages are not intended for sharing contact details and do not represent active conversations between users. To filter out system or automated notifications, we discard messages containing phrases such as ’reputation from’ or those flagged as staff messages.

TABLE III
FEATURE OVERVIEW OF THE POPULAR COMMUNICATION PLATFORMS. ● - PROVIDES ALL OF THE FEATURES, ◐ - PROVIDES SOME OF THE FEATURES, ○ - PROVIDES NONE OF THE FEATURES.

Platform	Community	Security	Messaging
Telegram	◐	◐	●
Discord	●	◐	◐
XMPP/Jabber	◐	●	○
WhatsApp	○	◐	●
ICQ	◐	◐	◐
TOX	○	●	○

4: To find popular communication platforms on the forum, we employ a multi-faceted approach. This includes knowledge from experts (police), exploring the forum data itself and information on generally popular messaging platforms. As a next step, we use regular expressions to search in text messages and posts for the selected communication platforms. For the regular expressions, we allow a Damerau-Levenshtein distance [27], [28] of one in the messenger name if it is longer than four characters. The purpose of the distance measure is to account for writing errors. In addition, we include the keyword ‘tg’ to search for Telegram mentions as we observed that this is a frequently employed abbreviation. To determine whether a user also mentioned a user ID when mentioning a messenger we created regular expression patterns (see Table VII in the appendix) for each messenger and search the surrounding text where a messenger is mentioned. With user ID we refer to a set of characters that is unique to a user on a platform and can be used to identify and contact a user. User IDs are generally a sequence of specific characters that follow a specific pattern defined by the platform. A user ID can also be a phone number of email address. Searching the whole message or post text for a user ID can in some cases lead to false positives. For example, users may mention a phone number that is calling them frequently and also mention that they prefer Jabber over WhatsApp for communication. In this example, the text mentions both a phone number and WhatsApp, however, they are not related and are not contact information of this user. To prevent most of such mismatches, we do not search the whole post or message for a user id, but create a text window with the length of the maximum number of characters of each messenger and use that as an upper bound and lower bound (plus 10 extra characters, in case the user id is not directly mentioned before or after the messenger name).

5: To prepare messages for the topic modelling, we perform additional preprocessing steps, namely removing text in quotation blocks, HTML tags, special characters, URLs, emails, platform names, platform user IDs, single digits and strings that contain digits.
6: We create multiple dataframes for the different classes of data.
7: We compute additional features, outlined in §III-D and §III-E.

We reduced the number of originally chosen communication platforms to Telegram, Jabber, Discord, ICQ and TOX. Hereby, the results for XMPP and Jabber were merged together, as Jabber is the older but still popular name of the XMPP protocol. Messengers that had less than 100 matches in posts and messages, Keybase, Matrix, and Signal were combined under the name 'other'. Messengers such as Skype, Wickr and Vipole were excluded from the analysis. These three allow users to have usernames that consist of letters, number and certain special characters. Consequently, numerous usernames become indistinguishable from ordinary words when analyzed by our techniques. This indistinguishability results in a significant number of errors. Therefore, we decided to exclude such usernames, acknowledging that this decision adversely affects the results.

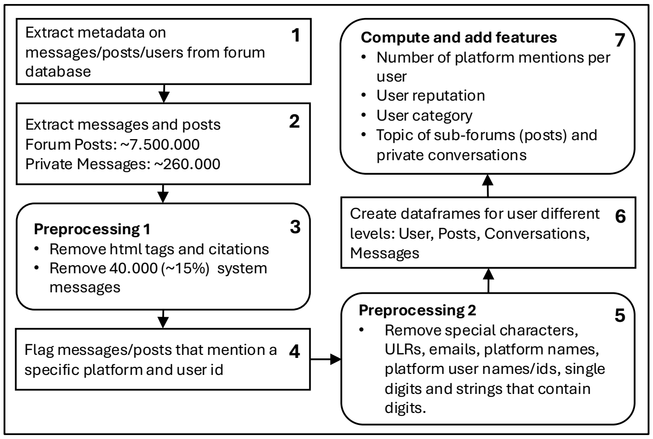
Extract metadata on messages/posts/users from forum database
1
Extract messages and posts
Forum Posts: ~7.500.000
Private Messages: ~260.000
2
Preprocessing 1
• Remove html tags and citations
• Remove 40.000 (~15%) system messages
3
Flag messages/posts that mention a specific platform and user id
4
Preprocessing 2
• Remove special characters, ULRs, emails, platform names, platform user names/ids, single digits and strings that contain digits.
5
Create dataframes for user different levels: User, Posts, Conversations, Messages
6
Compute and add features
• Number of platform mentions per user
• User reputation
• User category
• Topic of sub-forums (posts) and private conversations
7

Fig. 1. Overview of the data processing steps.

D. Analyzing users' forum position

To study our users' choices, we extract several user characteristics that we either took directly from the database or created ourselves.

Centrality Measures (Social Network Analysis): We use NetworkX Python library [29] to create directed graphs from the interactions on the forum and also to calculate the centrality measures. Hereby, we only consider interactions that mention contact details of a communication platform. We created two separate graphs, one for the interactions of posts and one for private messages. If a user replies to a post, the edge goes from the replier to the original poster. For private conversations, edges go from the sender to the receiver of a private message. After creating the graphs we calculated the in-degree-centrality, out-degree centrality and betweenness centrality.

In-Degree (private messages): A high value means that the user is contacted often.
In-Degree (public posts): A high value means the a users original posts get many of replies.
Out-Degree (private messages): A high values means that the user often reaches out.
Out-Degree (public messages): A high values means that the user very active in commenting under many different original posts.
Betweenness Centrality (private messages): A high value means the user is chatting with many people that do not chat with each other (broker).
Betweenness Centrality (public messages): A high value means that many users that comment this user's original posts do not comment posts of each other.
PageRank (private messages): A high value means the user receives messages from users who also receive many messages (influential users).
PageRank (public messages): A high value means that a users original posts receive many replies from users who also receive many comments on their posts.

To statistically validate the influence of centrality measures on platform choice, we employ a probit regression model using the Statsmodels Python library [30]. We incorporate the user's reputation value alongside the centrality measures in the model. We include the centrality measures from users' posts and messages. Furthermore, we confine the statistical analysis to users who disclose contact details in both posts and messages. As too few users mention contact details of TOX, WhatsApp, ICQ, and Other, we do not consider them for this analysis. We scale the independent variables (centrality measures and user reputation) using the MinMaxScaler from the scikit-learn Python library [31]. The dependent variable (communication platform) is binary encoded, either a user mentions a certain platform (1) or not (0).

Forum Features: Some features we use for our model are taken directly from the forum database, the reputation and group of a user. These features are not completely objective and users might have manipulated them [10]. Forum user can have different types of badges, membership types, or awards which represent different aspects. Some reflect the activity of a users, that is the number of posts a users writes. Others can be bought and do not directly reflect a users activity. Paying for a special membership allows users to, for example, send more messages, give more reputation, and change the username more often. We assume that these badges, memberships, or award indicate how much effort users invest into their presence on the forum. For our analysis, we group users into four categories:

- 'special membership': users that paid to get a special membership (2122 users)

- ‘executive role’: these are moderators or administrators (7 users)
- ‘award’: users that have an award which directly reflects their posting activity. There are different awards for different amounts of posts (440 users)
- ‘other’: every other user (4264 users)

Banned users are excluded here and when users have an award and a paid membership, we group them to the paid category only.

E. Topics

To determine the topic of posts we look at where in which thread and sub-forum the post was placed. On forum, users can make posts in different sub-forums which relate to different topics, such as gaming, anime, politics, hacking, or forum related discussions. There is a chance that the thread/post does not fit to the sub-forum, however, there is some moderation in place that makes sure users post in the right forum. Therefore, we assume that the threads/posts are in the right sub-forum. We manually label and then merge sub-forums to the following higher-level topics: ‘Hacking’, ‘Other’, ‘Social’, ‘Other Hobbies’, ‘Tech’, ‘Tutorials/Learning’, ‘Gaming’, and ‘Forum internals’. The labeling was done by two people and has high Inter-Annotator Agreement. See Table V for an overview of the number of posts per topic that contain a user ID and a short description of each topic.

To assess the topic of a message or conversation in which a platform is mentioned we use the BERTopic library [32] in combination with a multilingual pre-trained topic model ‘distiluse-base-multilingual-cased-v1’ [33]. We merge all messages of a conversation into one ‘document’ and shorten or remove a conversation according to the following criteria: (1) if the conversation is shorter than 10 words, we remove it (following [34]). (2) If the conversation is larger that the maximal token size of the topic model (128), we shorten it. We shorten the conversation around the messages that contain a user ID to a size of 128 words. This means that one conversation could be split into multiple documents. However, this only applies to around 3% of all the selected conversations. Using the automatic topic reduction, the model returned 14 topics, from which we manually merged several topic, resulting in 9 final topics. The three most representative words of each of the topics and the corresponding document count are presented in Table IV. Mapping the documents back to conversations results in 6387 conversations with a topic.

IV. Results

A. Choice of Communication Platform

The results of our analysis on the choice of communication platforms show that there is a difference in popularity of which platform user IDs forum user share in public posts and private messages. Additionally, we show that most users of this forum mention user IDs of just one communication platform.

Table II provides an overview of how often certain communication platforms are mentioned on the forum. The number of messages or posts that contain at least on of the mentioned communication platforms are 14.858 and 45.892 respectively. However, only 6.236 users mention a user ID in a post, 3.414 in a private message, and 1.618 users mention a user ID in a post and a message. Figure 2, shows that Telegram is not just leading in the total number of posts and messages that contain a user ID of Telegram, but also in the number of users that mention it. With 8.032 users that mention a users IDs in either a post or message, 67% of them mention one from Telegram. Discord is also popular, being mentioned by around 36% of users that mention user IDs. Figure 2 also highlights that most users only share user IDs of one of the platforms. Interestingly, users that mention user IDs of less popular platforms such as Jabber,WhatsApp and ICQ, often also mention Telegram. This could mean that Telegram functions as a sort of common ground platform as backup when users do not use one of the other platforms, or those platform user IDs are shared in special situations.

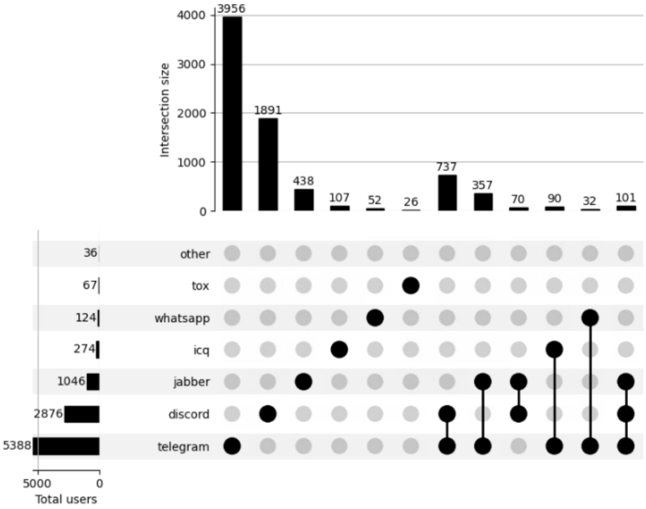

Fig. 2. Overlap in communication platform user IDs that are mentioned by users in public posts and private messages. An intersection size < 20 was ignored in this graph in order to keep it readable.

Figure 3 shows the cumulative trend of how often a user IDs of a platform is mentioned in private messages. Figure 4 shows the same for public posts. The appendix includes additional graphs showing the trend for platforms mentions when there is not user ID in the text (Figure 10 and Figure 11). To create these timelines, we selected the time-span from 2020-02-01 until 202-01-31 so that there is data for every day in each month for both posts and messages.

For platforms such as WhatsApp, TOX and ICQ, the marker size is, compared to the others, more often around 100%, possibly due to the fact that they are less popular and not mentioned frequently.

A large increase in platform occurrence combined with a small increase in marker size, can indicate that a small number of users mention this platform very often, potentially spamming other users with requests to add them to their contact list. This can be seen for WhatsApp in October 2020 (Figure 3), for TOX in December 2021 (Figure 4) and for Jabber in October

2021 (Figure 4).

The two most often named platforms, Telegram and Discord, show a relatively steady trend in all four figures, with Telegram showing an increasing popularity towards the end of the time frame in posts. WhatsApp follows a similar trend as Telegram when it comes to posts. In contrast, TOX, shows a stronger rise in popularity from beginning 2021 for messages and a few months, starting in June, for posts. However, the total number post/messages mentioning TOX low (195). Compared to the other platforms, the Jabber trend lines vary the most between messages and posts. In posts the trend line resembles a convex development while in the messages the trend line seems concave. This development is caused by one user writing over 3.000 posts containing a Jabber user ID between October 2021 and February 2022.

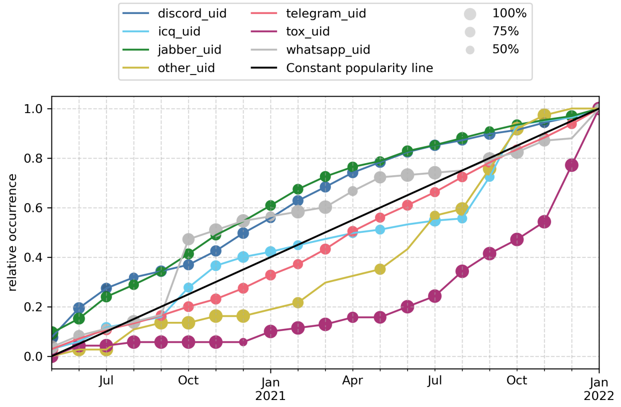

Fig. 3. Relative platform user ID (uid) mentions in private messages over time. The marker size refers to the percentage of new users that mention the platform compared to the previous months[2].

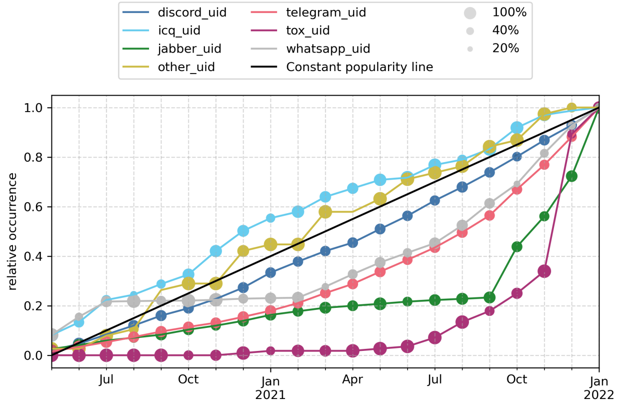

Fig. 4. Relative platform user ID (uid) mentions in public posts over time.The marker size refers to the percentage of new users that mention the platform compared to the previous months.

In conclusion, Telegram & Discord are the most frequently referenced platforms, with a consistently high level of popularity in messages and an increasing one in public posts. 80% of the forum users who mention a user ID of a communication platform only do so for a single platform. We observe that only 20% of users share user IDs to multiple platforms, this is in line with our hypothesis that the decision making process of forum users follows the principles of the rational choice theory. By only sharing user IDs of one platform users reduce the risk of identification and reduce the effort of managing multiple communication platforms which increases their benefit.

B. Context of Communication Platforms

In this section we want to investigate how the number of posts and messages containing a communication user ID changes per topic. We conduct this analysis on both the private message and public posts. With regard to posts, we look at which are the topics of the sub-forums where contact details are posted. Regarding the messages, we analyze the results of the topic model described in III-E. The analysis highlights that the popularity of platforms changes per topic.

TABLE IV
RESULTS OF THE TOPIC MODEL. IT PROVIDES TOPICS OF PRIVATE CONVERSATIONS OF FORUM USERS, IN WHICH THEY SHARE CONTACT INFORMATION.

Representative Words	Document Counts	Description
combos-combo-bases	178	User discussing combo data, gaming related.
data-csv-interested	327	Users signaling interest in unspecified data.
db-database-dbs	953	Users discussing deals for purchasing databases.
exam-writeup-hi	3779	Diverse topic, discussing study material related to hacking.
germany-countries-numbers	186	Discussing data of phone numbers of various countries.
india-vietnam-thailand	364	Offering data of people from Asian countries.
leaks-protonmail-market	145	Users offering leaked data.
mail-lists-plenty	166	Discussing data containing corporate emails.
sample-samples-send	289	Users requesting data of various types.

Table V shows that posts related to 'hacking' contain the most contact details, almost twice as many as the second place 'other'. This could be attributed to the observation, that many users that start a thread pertaining to the 'hacking' topic offer services or data. Consequently, both the initial authors and users responding to their posts provide contact information. Figure 5 shows that the messaging platforms Telegram, Jabber and TOX are preferably shared in posts related to this topic. The user IDs of others, such as ICQ and Discord are more evenly spread in multiple topics. Interestingly, WhatsApp IDs (phone numbers) are in over 90% of the cases shared in a post in a 'social' sub-forum. In addition, Figure 7 shows that most users share user IDs only in subforums belonging to the same topic. This suggests, that different user groups (based on their interest in the forum) have different communication platform preferences. Figure 12 in the appendix shows the proportion

[2] The colored graphs and figures in this paper use the color-blind-friendly colors taken from [35].

TABLE V
PUBLIC POSTS CONTAINING CONTACT DETAILS PER TOPIC (SUB-FORUM).

Sub-Forum Topic	Posts with contact details	Description
Hacking	21.444	Discussions about services or data related to hacking or online fraud.
Other	12.159	Various discussions that do not fit the other categories.
Social	3.984	Services and information related to social media.
Other Hobbies	3.523	Discussions related to anime, music, or pornographic content.
Tech	2.409	Discussions related to computer software and hardware.
Tutorials/Learning	1.517	Tutorials and tips related to software and malicious activities.
Gaming	432	Various gaming related discussions
Forum Internals	424	Forum announcements, introductions or other forum related discussions.

of communication platforms per topic.

The sharing behavior of platforms user names and ids in private messages is shown in Figure 6. The topics of the conversations in which users share contact details are mostly related to signaling interest in or asking for more details about data or information that was likely acquired through or can be used for malicious activities. The majority (59%) of the analyzed discussions belong to just on topic. Nevertheless, are contact details of platforms such as Jabber or Telegram prominent also in other topics. The reason ICQ is so prominent in the 'mail-lists-plenty' topic, is because of a spam wave offering different mail lists. Our results indicate that *users take the context into account in which they share a user ID in their decision making of which platform to choose. From the perspective of a malicious actor, topics related to hacking are more relevant to law enforcement so sharing contact details in this context bears more risk. The results show that in such a context, forum users prefer communication platforms which support a higher level of privacy and security.* Previously, we show that in public posts users often share user IDs if messengers with less security and privacy, e.g. Discord and WhatsApp. This seems counter intuitive to the rational choice theory, which assumes that actors make decisions that result in a high benefit to them, while having minimal risk and effort. However, this section shows that the context in which forums users share contact details (user IDs) of less privacy focused communication platforms, is likely perceived as less risky by the forums users. Discord contact details for example, are only in 29% of the cases shared publicly in a hacking related topic. This would emphasize again that the decision making process of forums users towards communication platforms can be explained with the principles of the rational choice theory.

C. Influence of User Characteristics on Platform Choice

In the previous section we outlined, how the communication platform popularity differs between the topic of the post conversation and sub-forum they are shared in. In this section we further investigate the connection between the in-forum reputation and activities of users and which platforms they are more likely to mention. Hereby, we focus on the reputation value of users on the forums, the centrality values from the social network analysis, and the category of users, as described in §III-D.

The results indicate that the reputation and the interactions of users relate to their choice of platform. Users who mention Jabber user IDs are likely more notorious users compared to those who mention Telegram or Discord user IDs.

Figure 8 shows that the share of users in the special membership category is higher for platform user IDs mentioned in private messages than for public posts. For both posts and messages, users who mention Jabber contact details are more often users who paid for a special membership.

Table VI shows that the lower the reputation of a user, the higher the likelihood of that user shares a Jabber or Discord user ID. In contrast, the higher a users reputation, the more likely that they mention a Telegram user ID. Besides, the more user share contact details in replies to public posts, the more likely is it that they mention a Telegram user ID.

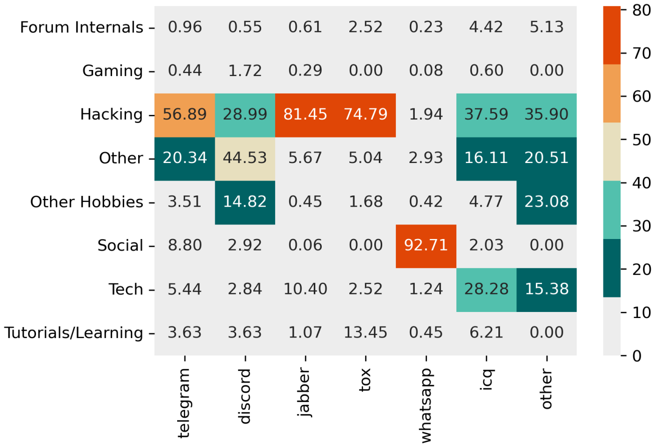

Fig. 5. Communication platform contact details shared in posts per topic (normalized per platform).

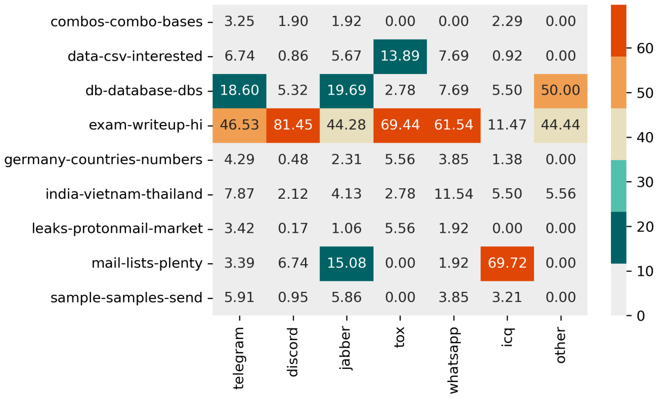

Fig. 6. Communication platform contact details shared in messages per topic (normalized per platform).

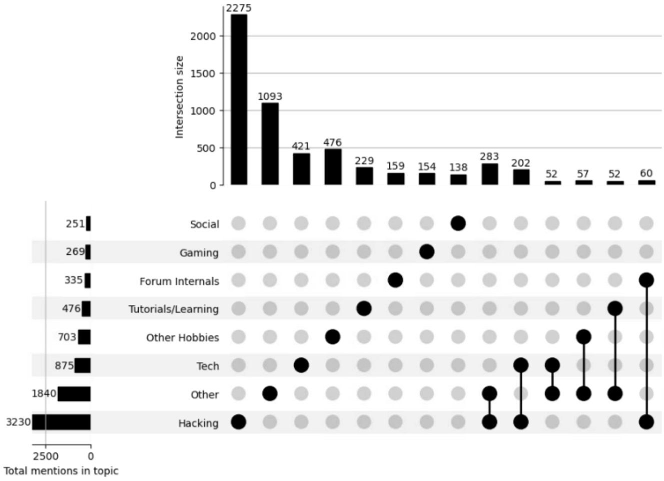

Fig. 7. Overlap of topics in which users posting user IDs. An intersection size < 52 was ignored for this plot.

TABLE VI
SIGNIFICANT RESULTS FROM MODELING THE RELATIONSHIP BETWEEN USER CHARACTERISTICS AND PLATFORM MENTIONING (PROBIT REGRESSION MODEL) OF THE THREE MOST OFTEN MENTIONED MESSENGERS. (P) REFERS TO THE POSTING AND (M) TO THE PRIVATE MESSAGING ACTIVITIES OF A USER.

Independent Variables	Telegram Coefficient	Discord Coefficient	Jabber Coefficient
reputation	7.24***	-2.37***	-6.55***
in-degree (M)	-0.98	1.94	7.15***
out-degree (P)	9.56***	-1.19	0.34
pagerank (P)	-0.19	-0.06	-3.37*
pagerank (M)	1.28	-0.30	4.52**

$^{***}p \leq 0.001$, $^{**}p \leq 0.01$, $^{*}p \leq 0.05$

The in-degree and pagerank regarding private messages of a user positively influences the likelihood that a user shares Jabber contact information. A high in-degree value generally indicates that the users is more popular and in the network. In combination with a high pagerank value, this means that a user is likely also more influential as the user is contacted by many user who are themselves popular in the network. On the underground, popular and influential users likely have information or data other users are interested in (see Table IV and [13]). The negative coefficient of the reputation value and pagerank value for public posts of users who mention Jabber, indicates that Jabber users receive fewer public engagement. This could suggest that user who prefer Jabber try to avoid public attention. However, we also showed that users who mention Jabber are also more often paying forum users which would give them some visibility as this is publicly displayed.

The results presented here indicate that users who mention Jabber user IDs are more notorious in the sense that they are more popular/influential and more determined towards the forum. Though such users do not necessarily have the highest reputation which falls in line with the findings of [21]. Conversely, Telegram seems to be popular with users who are actively sharing contact details in public and receiving positive reputation on the forum. From the rational choice perspective, it would be beneficial for notorious actors, who are more likely to be targeted by law enforcement, to reduce their public visibility as much as possible and to choose communication platforms that provide a high level of privacy and security.

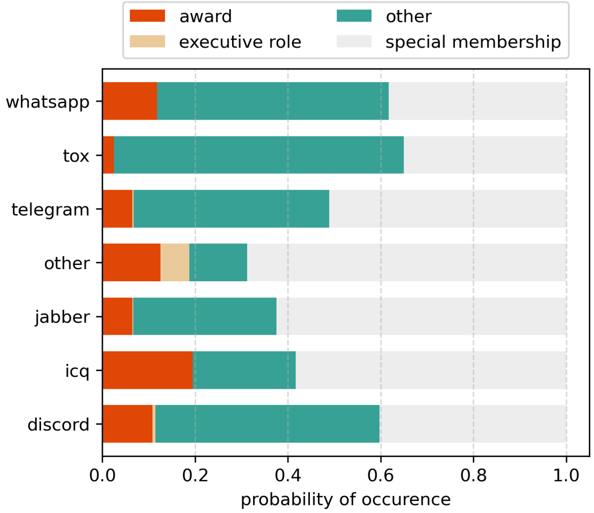

Fig. 8. Stacked bar plot of the distribution of categories of users mention a specific platform in private messages per platform.

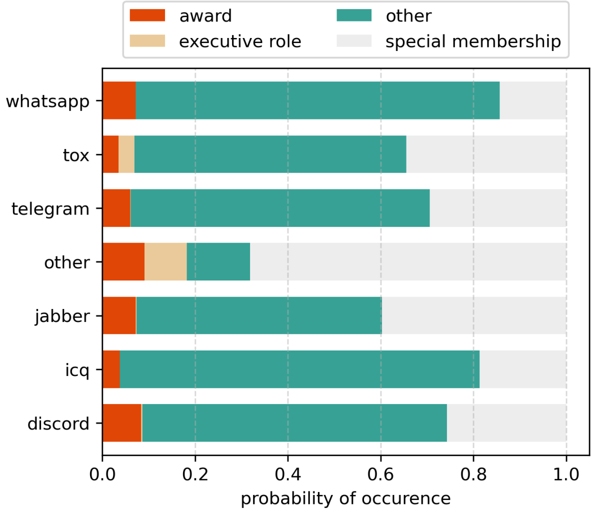

Fig. 9. Stacked bar plot of the distribution of categories of users mention a specific platform in public posts per platform.

V. DISCUSSION AND CONCLUSION

In this research, we studied a popular underground forum, to better understand which are the most popular communication platforms for private communication, whether these differ per topic and how sharing behavior correlates with user characteristics.

To study **RQ 1**, we designed regular expressions to find various communication platforms and corresponding user IDs that users mention in public posts and private messages.

Overall, Telegram contact details are shared the most often in posts and messages. Moreover, Telegram has the highest number of individual users that mention a user ID in either a post or a message. The second most often shared user IDs are from Discord, followed by Jabber, WhatsApp, ICQ, TOX, and others. Besides being popular for private (direct) communication, in recent years, Telegram and Discord have become important market places, social exchange, and community building platforms, possibly replacing the more 'traditional' underground forums/markets [36]–[38]. Both platforms have features that makes it easier to organize large communities, for example allowing group sizes of up to 200k (Telegram), allowing different forms of communication, the exchange of various types of files, and the possibility to integrate bots, to do specific tasks. The growing prevalence of these platforms within illicit markets potentially explains their widespread adoption for private communication, enabling users to interact seamlessly across various online communities without switching platforms. Another explanation as to why certain platforms are more popular than others could be that this reflects the popularity of those platforms in the general public.

Examining messenger app downloads worldwide this hypothesis does not hold. While Telegram is also popular, the download numbers of WhatsApp are marginally to much higher than those of Telegram in the years between 2019 and 2022 [39]. Moreover, recent numbers of monthly active users show that WhatsApp is generally the more (most) popular messaging platform [40]. In our data, 67% of all users that mention a platform user ID, mention one from Telegram and only around 1.5% one from WhatsApp. This disparity indicates that the communication platform preferences of users of the underground forums are likely more influenced by other factors, such as specific features of Telegram that makes it more beneficial. Actors (users) take situational risks and benefits into account in their decision making, according to the rational choice theory. We therefore expect that most user choose a platform that provides them a higher level of anonymity without sacrificing ease-of-use. Telegram is then the rational choice for users of an underground forum. As a next step, leading to **RQ 2** and **RQ 3**, we investigated whether the popularity of certain platforms is uniform throughout the forum, or whether this differs per sub-forum topic. Our results show that around 80% of the forum users mention just one platform. In instances where multiple platforms are cited, Telegram frequently emerges as a common mention. In addition, our analysis of topics, shows that users usually mention contact details in sub forums related to just one topic. The most popular platform can differ per topic, Discord for example, is the most mentioned platform in gaming related posts which corresponds to it being very popular also in the general gaming community [41]. These results are in line with our hypothesis that forum users take the context where they share contact details into account. For a malicious actor, aspects such as security and anonymity of a communication platform are less relevant in a gaming related context in comparison to a hacking related context. To disrupt malicious activities, it is helpful to target the most influential and sophisticated players. Consequently, we investigated forum internal reputation and activity features and how this corresponds to the platforms users mention (**RQ 4**). The results show that per platform, these user characteristics vary. Our results suggest that user groups, characterized by their activity levels and popularity, have preferences for different communication platforms for private conversations. For example. users who mention Jabber IDs, seem to be more popular and influential, however, they have fewer public engagement.

Understanding what influences the choice of users of underground forums in an important step towards fully understanding social interactions and constraints in an illicit environment. Knowing why and which users choose which platforms can helpful to develop better policing and intervention strategies. Our results show that the majority of users of this underground forum prefer to use Telegram, followed by Discord. Therefore, it is important for law enforcement to follow developments of communities involved in malicious activities on these platforms. Particularly because these platforms are not just used for private communication but also as markets for illegal goods ans services or as a new form of underground forums [36]–[38]. In perspective of the rational choice theory we draw the following conclusions. Most actors (users) choose communication platforms for private communication that offer a high level of security and privacy while also being convenient to use even for large groups (community and messenger category). More influential actors in the context of malicious activities choose platforms which come with higher costs in terms of ease-of-use and general adoption, yet provide them with higher security and anonymity. Thus, influential actors invest more effort for their private communications for having the benefit of reducing the likelihood of their identification.

In this study we gave a first insight into messenger platform preferences in an underground forum. We also provided a statistical analysis to show which factors influence the choice of communication platforms for private conversations.

VI. Limitations and Future Work

Our study comes with several limitations:

Uncertainties There are multiple uncertainties connected with the data and how we extract features. We are not always certain that a platform mention reflects the intention to use it for communication. We tried to account for this by also extracting user names and ids and only considering a platform when a suitable username was mentioned. However, a user could also mention the contact details of someone else.

Accuracy Regular expressions introduce inaccuracies. We also use a distance measure to match platform names even when there are minor writing errors in them. While this allows us on one hand to find more true positive matches it also introduces more false positives. This is one of the reasons we only considered posts and messages that had a positive match for a platform and user ID. The topic modeling as well as the merging of sub-forums into higher level topics do not result in perfect representations of the actual topics. Nevertheless, we think it is accurate enough to get a better understanding of the general topics in which contact details are exchanged.

Missing Platforms/Messengers Our initial selection pool was comprised of popular platform names and names we found during exploratory research on the forum. Given that vast amount of data, it is likely that we did not take all platforms into account that are mentioned on the forum. However, the scope of this study was not to have a complete insight of all occurring platforms, but to focus on the most popular ones. In future studies it would be interesting to investigate whether and how to find platforms that are not (yet) very often mentioned to get a complete picture of the communication platform landscape.

In the future we would like to address the mentioned limitations, for example, improving the regular expressions to identify communication platforms more accurately in combination with increasing the number of platforms. In addition, it would be interesting for future work to study the reasons for choosing a platform and the users who choose a certain platform in more details. For instance, we showed that user who mention Jabber are according more popular/influential. In a the next step it would be interesting to study the difference in users who mention Jabber in comparison to users who prefer Telegram or WhatsApp. Is there a difference in services they offer? How active are they on the forum? What is their area of expertise? Regarding the choice of the platform, it would be beneficial to further investigate how platform features, for example, privacy, security and ease-of-use influence the popularity of the platforms among certain groups. Or is the main reason for the popularity of certain platforms that proficient/influential users prefer specific platforms and other users adapt to their preferences?

VII. Ethical Considerations

The forum data for this study is not public and there is no consent from the users to have their data analyzed. As there are so many users involved, we did not see it as feasible to try to get consent for analyzing their data. As a result, we took additional care to not publish information that could be linked to specific persons. We asked legal and law enforcement experts to review the paper to make sure that no personal information is published. Access to the forum data was only given to two members of the project with security clearance. The other team members received aggregated results. During our research we followed the principles of ethical research as described in the Menlo report [42]. We also decided against disclosing the forum name for the privacy of the participants of the forum.

References

[1] T. Meurs, "COORDINATE: A model to analyse the benefits and costs of coordinating cybercrime," *Journal of Internet Services and Information Security*, vol. 12, no. 4, pp. 1–22, 2022-11. DOI: 10.58346/JISIS.2022.I4.001. [Online]. Available: https://jisis.org/wp-content/uploads/2022/12/I4.001.pdf (visited on 06/18/2024).

[2] B. Collier, R. Clayton, A. Hutchings, and D. Thomas, "Cybercrime is (often) boring: Infrastructure and alienation in a deviant subculture," *The British Journal of Criminology*, vol. 61, no. 5, pp. 1407–1423, 2021. DOI: 10.1093/bjc/azab026.

[3] E. R. Leukfeldt, E. R. Kleemans, and W. P. Stol, "Origin, growth and criminal capabilities of cybercriminal networks. An international empirical analysis," en, *Crime, Law and Social Change*, vol. 67, no. 1, pp. 39–53, 2017-02, ISSN: 0925-4994, 1573-0751. DOI: 10.1007/s10611-016-9663-1. [Online]. Available: http://link.springer.com/10.1007/s10611-016-9663-1 (visited on 06/18/2024).

[4] E. R. Leukfeldt and T. J. Holt, "Examining the Social Organization Practices of Cybercriminals in the Netherlands Online and Offline," en, *International Journal of Offender Therapy and Comparative Criminology*, vol. 64, no. 5, pp. 522–538, 2020-04, ISSN: 0306-624X, 1552-6933. DOI: 10.1177/0306624X19895886. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0306624X19895886 (visited on 06/18/2024).

[5] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. M. Voelker, "An analysis of underground forums," en, in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, Berlin Germany: ACM, 2011-11, pp. 71–80, ISBN: 978-1-4503-1013-0. DOI: 10.1145/2068816.2068824. [Online]. Available: https://dl.acm.org/doi/10.1145/2068816.2068824 (visited on 04/04/2023).

[6] M. Yip, N. Shadbolt, and C. Webber, "Why forums? an empirical analysis into the facilitating factors of carding forums," in *Proceedings of the 5th Annual ACM Web Science Conference*, ser. WebSci '13, New York, NY, USA: Association for Computing Machinery, 2013-05, pp. 453–462, ISBN: 978-1-4503-1889-1. DOI: 10.1145/2464464.2464524. [Online]. Available: https://dl.acm.org/doi/10.1145/2464464.2464524 (visited on 04/04/2023).

[7] S. Pastrana, A. Hutchings, A. Caines, and P. Buttery, "Characterizing eve: Analysing cybercrime actors in a large underground forum," in *Proceedings of the 21st Research in Attacks Intrusions Symposium (RAID '18)*, ser. Lecture Notes in Computer Science, vol. 11050, Crete, Greece: Springer, 2018, pp. 207–227. DOI: 10.1007/978-3-030-00470-5_10.

[8] M. Schafer, M. Fuchs, M. Strohmeier, M. Engel, M. Liechti, and V. Lenders, "BlackWidow: Monitoring the Dark Web for Cyber Security Information," in *2019 11th International Conference on Cyber Conflict (CyCon)*, Tallinn, Estonia: IEEE, 2019-05, pp. 1–21, ISBN: 978-9949-9904-5-0. DOI: 10.23919/CYCON.2019.8756845. [Online]. Available: https://ieeexplore.ieee.org/document/8756845/ (visited on 07/18/2023).

[9] C. Huang, Y. Guo, W. Guo, and Y. Li, “HackerRank: Identifying key hackers in underground forums,” *International Journal of Distributed Sensor Networks*, vol. 17, no. 5, p. 155 014 772 110 151, 2021-05, ISSN: 1550-1477. DOI: 10.1177/15501477211015145. [Online]. Available: http://journals.sagepub.com/doi/10.1177/15501477211015145.

[10] M. Campobasso, R. Rădulescu, S. Brons, and L. Allodi, *You Can Tell a Cybercriminal by the Company they Keep: A Framework to Infer the Relevance of Underground Communities to the Threat Landscape*, 2023-06. [Online]. Available: http://arxiv.org/abs/2306.05898 (visited on 06/15/2023).

[11] S. Afroz, A. C. Islam, A. Stolerman, R. Greenstadt, and D. McCoy, “Doppelgänger Finder: Taking Stylometry to the Underground,” in *2014 IEEE Symposium on Security and Privacy*, 2014-05, pp. 212–226. DOI: 10.1109/SP.2014.21.

[12] X. Wang, P. Peng, C. Wang, and G. Wang, “You Are Your Photographs,” in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ACM, 2018-05, pp. 431–442, ISBN: 978-1-4503-5576-6. DOI: 10.1145/3196494.3196529. [Online]. Available: https://dl.acm.org/doi/10.1145/3196494.3196529.

[13] Z. Sun, C. E. Rubio-Medrano, Z. Zhao, T. Bao, A. Doupé, and G.-J. Ahn, “Understanding and Predicting Private Interactions in Underground Forums,” en, in *Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy*, Richardson Texas USA: ACM, 2019-03, pp. 303–314, ISBN: 978-1-4503-6099-9. DOI: 10.1145/3292006.3300036. [Online]. Available: https://dl.acm.org/doi/10.1145/3292006.3300036 (visited on 04/03/2023).

[14] U. Akyazi, M. van Eeten, and C. H. Gañán, “Measuring cybercrime as a service (caas) offerings in a cybercrime forum,” 2021. [Online]. Available: https://weis2021.econinfosec.org/wp-content/uploads/sites/9/2021/06/weis21-akyazi.pdf.

[15] D. Cornish and R. Clarke, “Rational choice approaches to crime,” *The reasoning criminal: Rational choice perspectives on offending*, pp. 1–16, 1986.

[16] I. Pete, J. Hughes, Y. T. Chua, and M. Bada, “A Social Network Analysis and Comparison of Six Dark Web Forums,” in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, 2020-09, pp. 484–493, ISBN: 978-1-72818-597-2. DOI: 10.1109/EuroSPW51379.2020.00071. [Online]. Available: https://ieeexplore.ieee.org/document/9229679/.

[17] M. Bada and Y. T. Chua, “Understanding Risk and Risk Perceptions of Cybercrime in Underground Forums,” in *2021 APWG Symposium on Electronic Crime Research (eCrime)*, Boston, MA, USA: IEEE, 2021-12, pp. 1–11, ISBN: 978-1-66548-029-1. DOI: 10.1109/eCrime54498.2021.9738790. [Online]. Available: https://ieeexplore.ieee.org/document/9738790/ (visited on 04/17/2023).

[18] V. Benjamin and H. Chen, “Securing cyberspace: Identifying key actors in hacker communities,” in *Proceedings of the IEEE Conference on Intelligence and Security Informatics (ISI ’12)*, 2012.

[19] S. Pastrana, A. Hutchings, D. Thomas, and J. Tapiador, “Measuring ewhoring,” in *Proceedings of the Internet Measurement Conference (IMC ’19)*, Amsterdam, Netherlands, 2019, pp. 463–477. DOI: 10.1145/3355369.3355597.

[20] A. Caines, S. Pastrana, A. Hutchings, and P. J. Buttery, “Automatically identifying the function and intent of posts in underground forums,” *Crime Science*, vol. 7, no. 19, 2018.

[21] A. A. Paracha, J. Arshad, and M. M. Khan, “S.U.S. You’re SUS!—Identifying influencer hackers on dark web social networks,” *Computers and Electrical Engineering*, vol. 107, p. 108 627, 2023-04, ISSN: 00457906. DOI: 10.1016/j.compeleceng.2023.108627. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0045790623000526.

[22] J. Hughes and A. Hutchings, “Digital Drift and the Evolution of a Large Cybercrime Forum,” in *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Delft, Netherlands: IEEE, 2023-07, pp. 183–193. DOI: 10.1109/EuroSPW59978.2023.00026. [Online]. Available: https://ieeexplore.ieee.org/document/10190639/ (visited on 09/06/2023).

[23] A. Talas and A. Hutchings, “Hacker’s paradise: Analysing music in a cybercrime forum,” in *2023 APWG Symposium on Electronic Crime Research (eCrime)*, 2023, pp. 1–14. DOI: 10.1109/eCrime61234.2023.10485503.

[24] J. Man, G. A. Siu, and A. Hutchings, “Autism disclosures and cybercrime discourse on a large underground forum,” in *2023 APWG Symposium on Electronic Crime Research (eCrime)*, 2023, pp. 1–14. DOI: 10.1109/eCrime61234.2023.10485504.

[25] R. Van Der Sanden, C. Wilkins, M. Rychert, and M. J. Barratt, “‘Choice’ of social media platform or encrypted messaging app to buy and sell illegal drugs,” en, *International Journal of Drug Policy*, vol. 108, p. 103 819, 2022-10, ISSN: 09553959. DOI: 10.1016/j.drugpo.2022.103819. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0955395922002353 (visited on 12/04/2023).

[26] Telegram, *Telegram messenger*. [Online]. Available: https://telegram.org/.

[27] V. I. Levenshtein *et al.*, “Binary codes capable of correcting deletions, insertions, and reversals,” in *Soviet physics doklady*, Soviet Union, vol. 10, 1966, pp. 707–710.

[28] F. J. Damerau, “A technique for computer detection and correction of spelling errors,” *Commun. ACM*, vol. 7, no. 3, pp. 171–176, 1964-03, ISSN: 0001-0782. DOI: 10.1145/363958.363994. [Online]. Available: https://doi.org/10.1145/363958.363994.

[29] N. developers, *NetworkX - NetworkX documentation*. [Online]. Available: https://networkx.org (visited on 12/02/2023).

[30] statsmodels, *Statsmodels.discrete.discrete_model.probit*, 2024. [Online]. Available: https://www.statsmodels.org/stable/generated/statsmodels.discrete.discrete_model.Probit.html.

[31] scikit-learn, *Minmaxscaler*, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html.

[32] M. P. Grootendorst, *Home BERTopic*, en. [Online]. Available: https://maartengr.github.io/BERTopic/index.html (visited on 12/02/2023).

[33] UKPLab, *Sentence trasnformers*, 2024. [Online]. Available: https://www.sbert.net/docs/sentence_transformer/pretrained_models.html.

[34] C. Doogan, W. Buntine, H. Linger, and S. Brunt, "Public perceptions and attitudes toward covid-19 non-pharmaceutical interventions across six countries: A topic modeling analysis of twitter data," *J Med Internet Res*, vol. 22, no. 9, e21419, 2020-09, ISSN: 1438-8871. DOI: 10.2196/21419. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32784190.

[35] P. Wright, *ColourBlind: A Collection of Colour-blind-friendly Colour Tables*, 2017-08. DOI: 10.5281/zenodo.840393.

[36] D. Lummen, *Is telegram the new darknet? a comparison of traditional and emerging digital criminal marketplaces*, 2023-03. [Online]. Available: http://essay.utwente.nl/94687/.

[37] T. Garkava, A. Moneva, and E. R. Leukfeldt, "Stolen data markets on Telegram: A crime script analysis and situational crime prevention measures," *Trends in Organized Crime*, 2024-04, ISSN: 1936-4830. DOI: 10.1007/s12117-024-09532-6. [Online]. Available: https://doi.org/10.1007/s12117-024-09532-6.

[38] R. van der Sanden, C. Wilkins, M. Rychert, and M. J. Barratt, "The use of discord servers to buy and sell drugs," *Contemporary Drug Problems*, vol. 49, no. 4, pp. 453–477, 2022. DOI: 10.1177/00914509221095279. eprint: https://doi.org/10.1177/00914509221095279. [Online]. Available: https://doi.org/10.1177/00914509221095279.

[39] M. Singh, *Telegram tops 700 million users, launches premium tier*, 2022. [Online]. Available: https://techcrunch.com/2022/06/19/telegram-tops-700-million-users-launches-premium-tier/.

[40] W. A. Social, DataReportal, and Meltwater, *Most popular global mobile messenger apps as of april 2024, based on number of monthly active users (in millions)*, 2024. [Online]. Available: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/.

[41] B. Fekete, *Discord Exec on the Chat Service's Impressive Growth, Partnering with Xbox and More - Newsweek*, 2018-05. [Online]. Available: https://www.newsweek.com/discord-celebrates-three-years-bringing-gamers-together-impressive-stats-929524 (visited on 07/03/2024).

[42] E. Kenneally and D. Dittrich, "The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research," en, *SSRN Electronic Journal*, 2012, ISSN: 1556-5068. DOI: 10.2139/ssrn.2445102. [Online]. Available: http://www.ssrn.com/abstract=2445102 (visited on 12/06/2023).

Appendix

Table VII shows the regular expressions used to find the user names/ids of the selected messengers. For each messenger keyword, we randomly selected 30 post and 30 messages and manually confirmed whether the matches were correct.

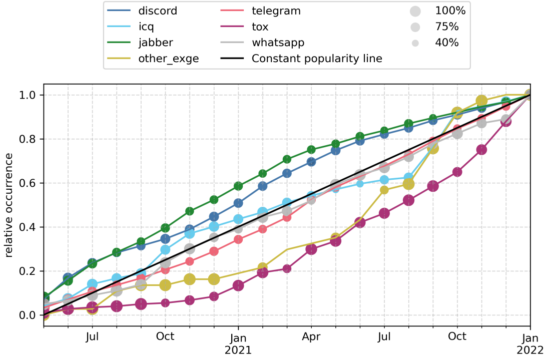

Fig. 10. Relative platform mentions in private messages over time.

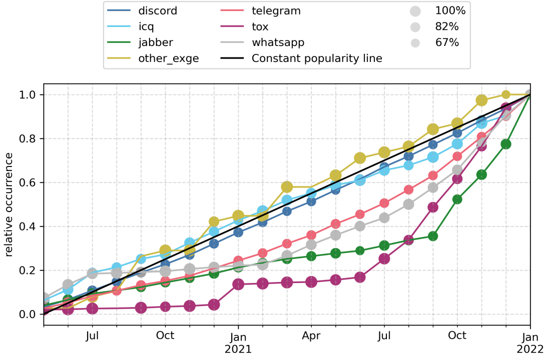

Fig. 11. Relative platform mentions in public posts over time.

TABLE VII
REGULAR EXPRESSIONS USED TO IDENTIFY MESSENGER USER NAMES/IDS AND THE F_1 SCORE FOR EACH COMBINATION.

<table>
<tr><th>Messenger Keyword</th><th>Regular Expression</th></tr>
<tr><td>Telegram</td><td>\W(?:@|(?:(?:(?:https?://)?t(?:elegram)?)\.me\/))(\w{4,32})</td></tr>
<tr><td>Tg</td><td>\W(?:@|(?:(?:(?:https?://)?t(?:elegram)?)\.me\/))(\w{4,32})</td></tr>
<tr><td>WhatsApp</td><td>(?:^|\s|:|;)(\+? ?(?:\d ?[\.\-\(\)]? ?){10,15})
(?:$|\s|(?:\.[$\W])|,|!|\?)</td></tr>
<tr><td>TOX</td><td>[a-fA-F0-9]{76}</td></tr>
<tr><td>Discord</td><td>.[^\s]{2,38}#[0-9]{4}</td></tr>
<tr><td>Jabber</td><td>[^@\s]+@[^@\s]+\.[^@\s]+</td></tr>
<tr><td>XMPP</td><td>[^@\s]+@[^@\s]+\.[^@\s]+</td></tr>
<tr><td>ICQ</td><td>[0-9]{6,9}|\W@\w{4,32}|((https?://)?icq\.im\/)\w{4,32}</td></tr>
</table>

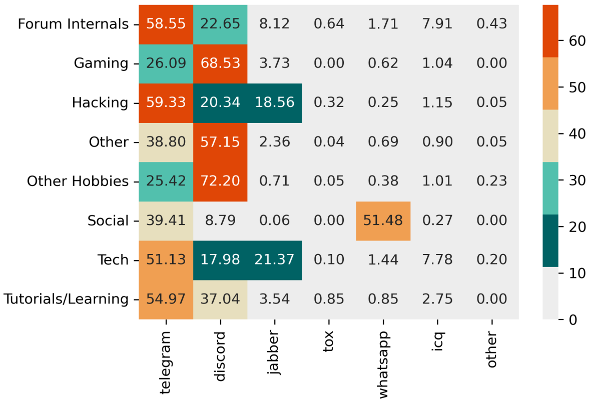

Fig. 12. Communication platform contact details shared in public posts per topic (normalized per topic).