AMoC: A Multifaceted Machine Learning-based Toolkit for Analysing Cybercriminal Communities on the Darknet

1st Chao Chen
Department of Computer Science
University of Bristol
Bristol, United Kingdom
chao.chen@bristol.ac.uk

2nd Claudia Peersman
Department of Computer Science
University of Bristol
Bristol, United Kingdom
claudia.peersman@bristol.ac.uk

3rd Matthew Edwards
Department of Computer Science
University of Bristol
Bristol, United Kingdom
matthew.john.edwards@bristol.ac.uk

4th Ziauddin Ursani
Department of Computer Science
University of Bristol
Bristol, United Kingdom
zia.ursani@bristol.ac.uk

5th Awais Rashid
Department of Computer Science
University of Bristol
Bristol, United Kingdom
awais.rashid@bristol.ac.uk

2021 IEEE International Conference on Big Data (Big Data) | 978-1-6654-3902-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigData52589.2021.9671906

***Abstract*—There is an increasing demand for expert analysis of cybercriminal communities. Cybercrime is continually becoming more complex due to the rapid development of digital technologies, on the one hand, in new types of criminal activity, such as hacking, distributing malware and DDoS attacks, and on the other hand, in digitised forms of more traditional crimes, such as email scams, phishing, identity theft, and cryptographically secured black markets. Tackling this broad array of behaviour requires tool support for multi-disciplinary investigations, and a connecting framework that can adjust flexibly to changes in the populations being studied. In this work, we present AMoC, a multi-faceted machine learning toolkit that combines structured queries, anomaly detection, social network analysis, topic modelling and accounts recognition to enable comprehensive analysis of cybercriminal communities and users. The toolkit enables the extraction of findings regarding the motivations, behaviour and characteristics of offenders, and how cybercriminal communities react to interventions such as arrests and take-downs. In our demonstration, the toolkit is deployed to analyse over 150,000 accounts from 35 underground marketplaces.**

***Index Terms*—Darknet Markets, Unsupervised Machine Learning, Anomaly Detection, Topic Modelling, Social Network Analysis, Siamese Network**

I. Introduction

Much recent research [13], [19] has highlighted that cybercrime is not solitary and anti-social activity, but one wherein online social interactions play a critical role in the recruitment, training and professional advancement of criminals. Accordingly, there is a need for tool support for the wide range of disciplines—and the variety of law enforcement practitioners—investigating these social interactions, in order to deepen our understanding of the dynamics of cybercrime.

A number of key research challenges can be identified to guide the design of such an analytical toolkit.

First, robust evidence regarding the impact and effectiveness of interventions on cyber offenders is still lacking. Prior work within the domain of cyber crime has questioned the efficiency of crackdowns and other large-scale police interventions with disappointing results [5], [6]. As a result, there is a gap in understanding which (combinations of) disruptive events have the most economic or psychological impact on cyber offender communities for the purpose of preventing cybercrime.

Secondly, prior work has shown that online social relationships are an important aspect of cyber crime [9], [13]. Offenders tend to thrive on their forum communications and try to establish a reputation within cybercriminal communities of interest. However, there is little information available about the different roles within such communities or how individuals move between them. Understanding these online peer relationships in relation to their cyber offending behaviour and identifying the key role models within these communities is essential when designing new preventative measures and (targeted) interventions [19].

A third challenge is the accounts recognition across different offender communities. A common issue for law enforcement investigators is that, arriving in a new forum or marketplace, they need to identify known suspects from other venues. Linking together evidence from different venues can be critical for an investigation. Hence, new technologies that can recognise accounts, particularly people with a prominent role, between accounts in different darknet communities could contribute to better informed and, hence, more efficient interventions.

These research challenges shape the fundamental requirements for a truly multi-disciplinary cybercrime analysis toolkit. (1) quantitative measurements of community activity and cohesion need to be exposed to enable goals such as the evaluation of interventions. (2) rich qualitative data need

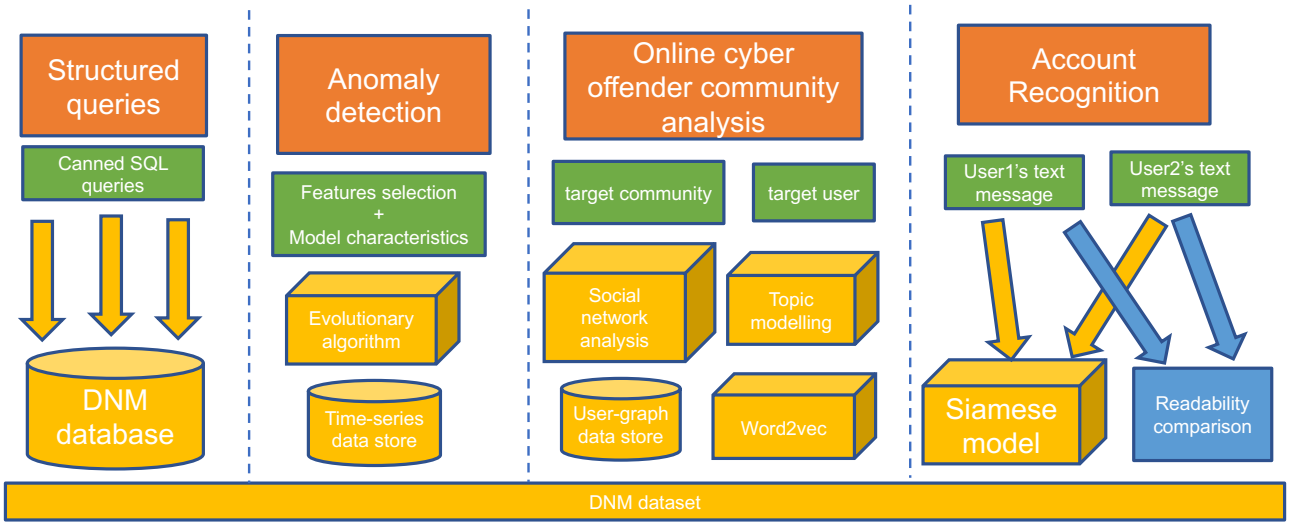


Fig. 1: AMoC toolkit overview

to be exposed and organised to enable the development of an understanding of the internal structures and patterns of cybercriminal activity. (3) features and labelled data need to be exposed to enable machine-learning components to act as support systems for analysts.

In this paper, we present a toolkit that speaks to these requirements, and demonstrate its application for automatically analysing cybercriminal communities on the darknet. Our demonstrations focus on financially-motivated cyber-dependent crime, and combine computational methods and techniques—computational stylometry, machine learning and neural networks, social network analysis, and text clustering—with qualitative analysis of discussions and interactions in darknet market (DNM) fora. More specifically, the key contributions of our work are as follows:

- An unsupervised learning based anomaly detection approach, which can indicate disruptive (or adverse) events, such as the closure of other markets, concerns about breaching of the market by law enforcement or concerns about large-scale fraud within the market. Our approach enables an automated analysis of the cascading impact of such disruptive events on the darknet ecosystem;
- A novel unsupervised learning methodology, which combines techniques from the area of Natural Language Processing (NLP) and Social Network Analysis (SNA) to automatically subdivide cybercriminal forums and marketplaces into usefully-delineated sub-communities, and identifies key users playing a prominent role in these communities;
- A Siamese neural network that automatically constructs a stylometric profile based on textual features extracted from an offender's messages and identifies stylometric similarities in other users' messages;
- Operationalisation of the above analyses and synthesis into the AMoC toolkit, a proof-of-concept software package for use in cybercrime investigations on darknet communities.

The research presented in this paper was undertaken in close collaboration with law enforcement cybercrime experts. The tools developed within our study will allow investigators to (*i*) detect cyber offenders, analyse their criminal activities and behaviour, (*ii*) assign degrees of importance and urgency to items of evidence in order to assess cyber offenders' potential danger to society and (*iii*) find useful evidence in a timely manner.

The rest of this paper is structured as follows. Section II provides an overview of the AMoC toolkit architecture. In Section III, we discuss the performance of the different components in a demonstration on a substantial dataset of darknet market (DNM) fora. In Section IV, we provide related work in the field. Finally, Section V concludes the paper, discusses the limitations of our approach and identifies directions for future research.

II. System Overview

This section discusses the AMoC toolkit and each functionality, respectively. In general, the front-end of toolkit is written in HTML, Javascript and node.js, which means it can be easily deployed as a desktop application or an online web service. The front-end communicates to four main components by json objects and the system architecture is displayed in Figure 1. We start describing the dataset we used for training our main components and how it is processed by the toolkit.

A. Data Usage

For the analysis presented in this study, we made use of over 2.5 million posts drawn from over 150,000 accounts from 35 cybercriminal communities, drawn from the DNM Corpus: a large dataset collected between 2013 and 2015 [2]. All the

DNMs have English language as their main medium of communication. In particular, we targeted discussion fora within this collection, which acted as support areas for underground marketplaces dealing in a number of different illicit goods. Communities ranged from successfully established markets with thousands of accounts (though not all were always active posters) to small sites that never moved beyond a handful of initial accounts. Table I gives a breakdown of the data available for each community.

TABLE I: Breakdown of the DNM dataset used for this study.

Community	Posts	Users
Silk Road 2	882,418	26,163
Silk Road	846,077	52,383
Evolution	509,225	33,743
Abraxas	276,300	1,607
Agora	84,914	6,153
Black Market Reloaded	80,467	7,006
Nucleus	65,175	9,478
The Hub	58,642	7,337
Pandora	49,023	8,729
Black Bank	32,817	2,381
The Majestic Garden	26,121	1,858
Utopia	14,458	4,392
Diabolus	11,456	2,151
Kingdom	10,285	856
Project Black Flag	6,131	330
Cannabis Road2	5,842	2,139
Cannabis Road3	4,905	1,903
Bungee54	3,325	1,510
Panacea	2,241	520
Tor Bazaar	2,205	902
The Real Deal	1,049	115
Hydra	937	276
Kiss	933	145
Andromeda	894	1,601
Outlaw Market	689	2,007
Revolver	660	85
Tor Escrow	490	294
Dark Bay	332	484
Doge Road	300	118
Darknet Heroes	190	793
Havana	181	77
Tom	144	4,120
Grey Road	43	24
Tortuga	37	7
Mr Nice Guy	25	6

The collected DNM dataset is split into two tables. One is the user registration table for recording user general information who registered in a forum. It contains the community, user id, title, first seen date and public key. A second table is for forum content, which includes community, user id, thread id, date, subject, category, body and quotes. The latter contains what users have posted in a forum and related user's information. The exemplar data schema is also included the front page in the AMoC toolkit that can be used to import and analyse new DNM data from other sources based on this specified data schema.

B. Structured Queries

The AMoC toolkit provides an user interface that displays the selectable functionalities can execute the canned queries on any target community or user. There is a relational database atop the DNM dataset to response our provided canned SQL statement queries, enabling the investigator to easily analyse the data by performing the following operations:

- **Basic queries**: total number of fora, registered users, and activated users & posts.
- **Aggregate queries**: top 10 users by number of posts; top 5 popular time slots (either in day of week, month of year or year) for user registration and for user posts;
- **Zoomed in views**: top 10 activated users by most posts in a target year; top 10 activated users by most posts in a target community; and zooming in on registration information and latest 10 posted messages for a target user-ID.

As we displayed in Figure 2-a, after we re-direct into the structured queries from the main entrance page, we will see three major queries (basic, aggregate and zoomed in views) have listed separated by lines. For example in Figure 2-b aggregate queries such as Top-5 popular time slots for user registration and posts, there is another option for day of weeks, month of year or year. Figure 2-c and Figure 2-d indicate the zoomed in views that enable the investigator to focus on a particular year, forum or user-id.

C. Adverse Events Detection Module

Modelling of disruptive or *adverse* events[1] can help understand the potential (cascading) impacts of law enforcement actions against DNMs. Adverse events may not deter members already well engaged in cybercriminal activity (or, at least, may not deter significant proportions of those), however, visibility of such disruptive events (and their consequences) may discourage new members from engaging in cybercriminal careers. Furthermore, DNM administrators tend to close a forum voluntarily if they suspect intrusion into the forum by law enforcement, often indicated by an unexpected influx of new members. Emulating such intrusions may trigger voluntary shutdowns of fora leading to an influx into others, thus creating a chaining effect of voluntary shutdowns.

The AMoC toolkit's adverse events detection module is based on an unsupervised learning based anomaly detection approach. More specifically, our model is based on a weighted sum of a feature set trained through an evolutionary algorithm (see [26] for a detailed overview of our approach). We provide a flowchart of the overall approach in Fig. 3.

To start the anomaly detection analysis on a target forum, one can select a number of options regarding the feature set, model characteristics with respect to time series, and the threshold limits as the model configuration. In Figure 4-a, we choose Canabis Road-3 as the target forum and selected all features, set the sample period to 1 month and the anomaly threshold to 1.5. The detection results are displayed in Figure 4-b. We can see that there is a linear model capturing all weighted sum of features set and there is an anomaly detected

[1]Some examples of adverse events are: the actions taken by law enforcement agencies against DNMs, such as the Silk Road Shutdown [17], including actions taken against its management, such as the arrest of Silk Road founder Ross Ulbricht [18] and its users (e.g., Silk Road Drug dealer Cornelis Jan "Maikel" Slomp) [14], internal fights among DNMs through rumouring, DDOS attacks or hacking (e.g., the compromise of Silk Road 2 Escrow Accounts [10]) or any negative news about DNMs in the media, such as the Gawker Blog publication about Silk Road [3].

(a) User interface for structured queries

(b) Aggregate queries

(c) Zoomed in the community

(d) Zoomed in the user

Fig. 2: Screenshots of the AMoC toolkit's structured queries

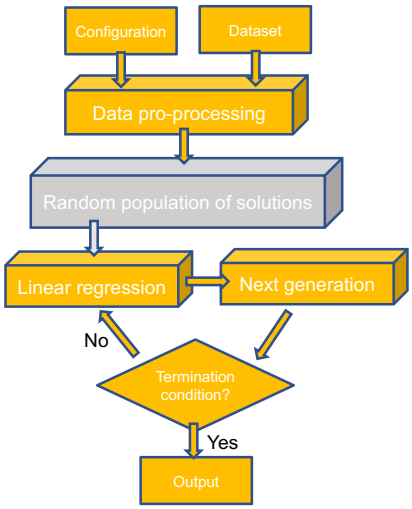


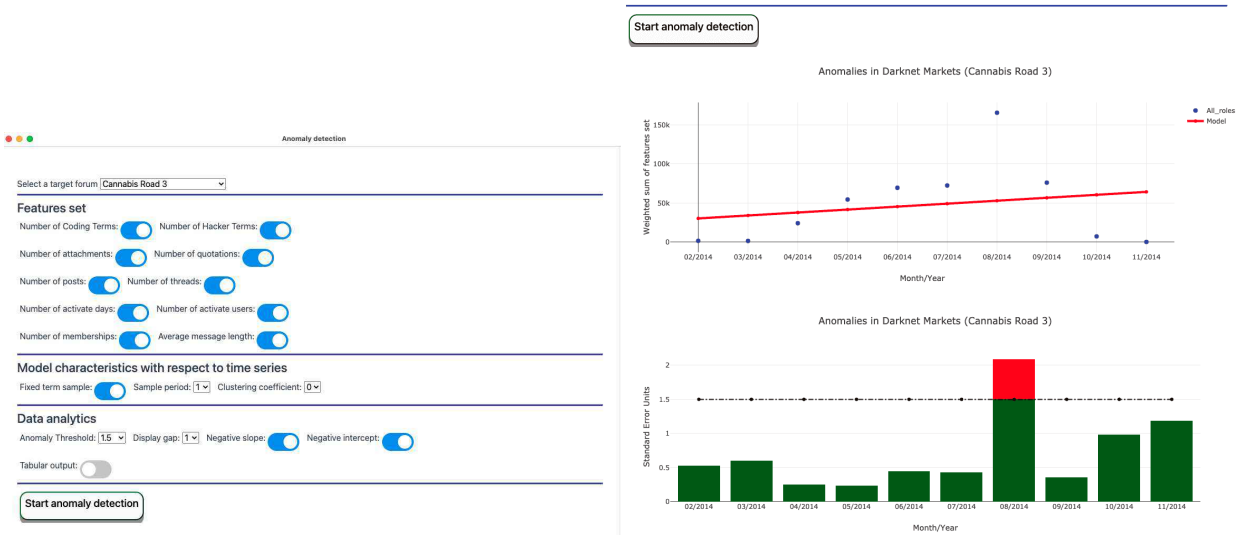
Fig. 3: Flowchart of anomaly detection algorithm

in August 2014 (highlighted in red). The model showed a standard error of over 1.5 compared to other time windows (in month) at Canabis Road-3.

D. Online Cyber Offender Community Analysis

There is comparatively little information about the roles and the separation of these roles within financially-motivated cybercrime online. As darknet markets are online fora, roles can often be conflated with membership or user types within such fora, e.g., administrator, new user, etc. Designing more efficient preventative strategies requires a more nuanced understanding of these roles, and the power relationships between them, as they emerge through and are defined by linguistic interactions. The AMoC toolkit combines novel NLP techniques for automatic topic modelling and Social Network Analysis (SNA), enabling a bottom-up approach that incorporates all users and their communications in darknet fora, and leading to a multifaceted understanding of such DNM community roles and how individuals move between them. More specifically, we apply a weighted SNA approach to identify the most important nodes in each forum (i.e., "community influencers") based on their contributions to different forum conversations (or threads). Additionally, the module automatically detects sub-communities and their members by applying Clauset-Newman-Moore greedy modularity maximization [4]. Key users are then identified by calculating the average sum of three centrality measures: degree centrality, eigen-vector centrality and the local clustering coefficient. The SNA algorithm is displayed in Algorithm 1.

As can be seen in Figure 5, an investigator can choose from a list of fora menu, and there is a text area in which a target user-id can be entered. After launching the social network analysis on a selected target forum, the toolkit displays a number of sub-communities and its users. Additionally, the most influential users in each community are displayed with



(a) (b)

Fig. 4: Screenshots of the AMoC toolkit's anomaly detection

Algorithm 1 Social network analysis on a target forum

Require: Build the user graph and find a list of communities (C).

Ensure: The weight of edge is based on the common threads.

```
for c ← 1 to C do
    for i ← 1 to N do        ▷ All nodes in this community
        I ← calculate degree centrality
        E_i ← calculate eigen-vector centrality
        LC ← calculate local clustering co-efficient
        Sorted graph's users by (I+E_i+LC)/3
    end for
end for
```

Algorithm 2 Topic modelling on a target user

Require: Target user's text messages

Ensure: Preprocessing and tokenize text messages

```
ℤ ← pre-trained word2vec
K ← 1 and −1 ← SC
while K ≤ 100 do
    cluster labels ← Kmean(K, ℤ)
    calculate θ_s ∈ [−1, 1]
    if θ_s ≥ SC then
        𝕂 ← the optimal number of cluster labels
        SC ← θ
    end if
    K ← K + 1
end while
```

their average sum of the centrality metrics mentioned above. Additionally, to enable a better understanding of the role these key users play within a community, the AMoC toolkit is able to automatically detect prominent topics discussed by these users on a forum. The topic modelling component combines

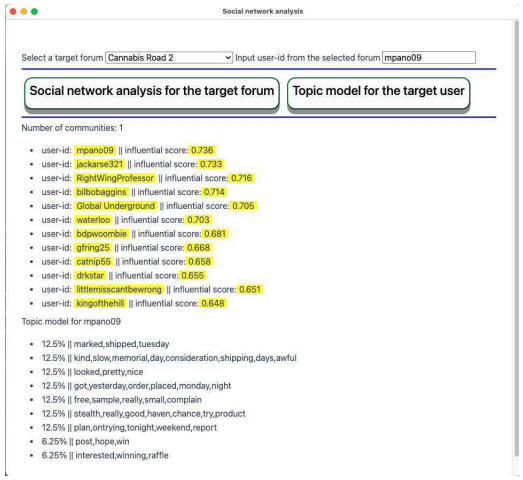


Fig. 5: Screenshot of the AMoC toolkit's community analysis at Canabis Road-2

the K-means algorithm with pre-trained word embeddings to extract the most informative words for each key topic in a target user's communications as illustrated in Algorithm 2. As is shown in Figure 5, the results of the topic modelling analysis indicate the percentage of the number of messages from this user that were attributed to each topic cluster and provides the top words of each cluster.

E. Account Recognition Analysis

Automated account recognition approaches could assist law enforcement investigators with tracking individuals across different fora or linking multiple accounts within a single forum. In this study, we used two different approaches to compare and estimate whether two seemingly different accounts may be the same person. One approach is to directly compare two input

text messages by the traditional linguistics readability indices, such as Gunning Fog index, Flesch-Kincaid index, average syllables per word, and average sentence length. The second approach draws on face recognition techniques, in which we designed a Siamese neural network model with triplet loss (see e.g., [23]).

1) Data construction: To first construct a dataset of verified linked accounts across multiple distinct DNM fora, we use self-declared PGP public keys as the identity link. There is a good reason to believe that these accounts are genuinely the same individual if they present the same public key across different fora. PGP keys are published in these fora to enable private communications, and posting a public key to which one does not own the corresponding private key would mean all incoming private messages would be unreadable, rendering the account all but useless. Therefore, we regard all accounts sharing the same public key as one positive pair, whilst others are negative pairs. It is worth noting that this leads to an extremely unbalanced dataset, as few accounts disclose their public key information. More importantly, to obtain enough text messages to train the model, we define a valid user as one who has posted at least 5 text messages under at least two different user-ids. This leads to 185 valid accounts in total that can be used for the positive samples, and 14080 accounts as the negative samples, who have public key information but only one known user-id posting at least 5 messages.

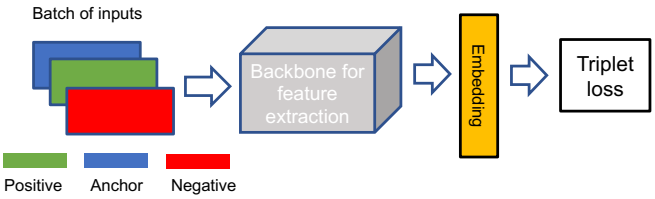


Fig. 6: The structure of Siamese network with the triplet loss

2) Model details: The structure of Siamese network with the triplet loss is displayed in Figure 6, it consists of four main components [23]. The input layer at the beginning takes a batch of three data samples, which are from the anchor (the original account's text message), positive (a matched account's text message) and negative (an unmatched account's text message), respectively, as the input. Then it was fed to a backbone neural network for the feature extraction. Similar to [23], this backbone network can be a black box and it adopts the state-of-the-art NLP neural network models, such as Recurrent Neural Networks (RNNs) [25], Bidirectional Long-Short Memory networks (Bi-LSTMs) [12], or BERT model [7]. The backbone feature extraction network results in a text message embedding, and it is a mapping from a text message to a numerical feature representations space $\mathbf{R}^d$, where d is the number of dimension for this numerical feature representations space. In the training, instead of comparing two text messages directly, a triplet loss is applied in the end to assistant with the model learning that between each pair of messages from one account to all others. In the inference, we can only use the backbone for feature extraction to generate embedding to compare.

The final layer of triplet loss has ensured that a text message embedding $f(A)$ (anchor) of a specific account is closer to all other text message embedding $f(P)$ (positive) of the same account than it is any text message embedding $f(N)$ (negative) of any other accounts. This is defined in Eq.1 [23], where α is a margin hyper-parameter that is enforced between each pair of text message embedding from one account to all other accounts. In general, it gives the diversity of text messages from one account and preserves the embedding distance to other accounts.

$$\mathbf{L}(A,P,N) = \max\left(\|f(A)-f(P)\|^2 - \|f(A)-f(N)\|^2 + \alpha,\ 0\right) \quad (1)$$

Since the account recognition analysis provides two approaches to evaluate whether two text messages are from the same or different person, in Figure 7-a, we entered a vendor's post from two different accounts, but based on the message signature ("HC"), chances are high the same account was involved. The Siamese network shows a very similar result of 0.352 (the smaller, the more similar), while the readability comparison based on different indexes generates a large score. Additionally, in Figure 7-b, we entered messages by two different vendors (two different message signatures and two different public keys). The readability comparison using Flesch, Gunning fog readability, and average syllables per word score, these two accounts showed a very small difference, but the Siamese network yielded a high result (1.731), which clearly indicates the messages were produced by two different accounts.

III. Experimental Results

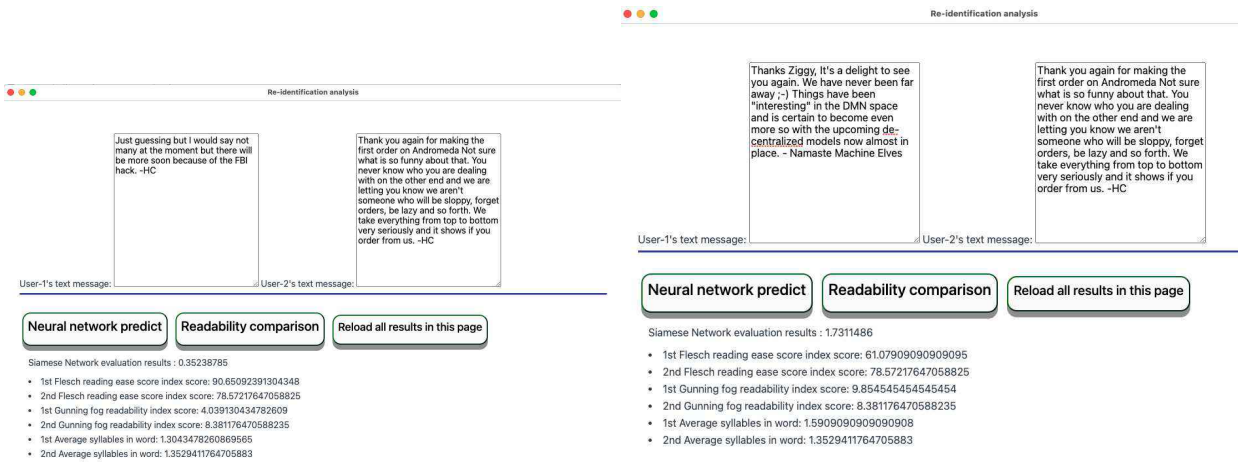
This section provides a summary of our experimental results for each of the AMoC toolkit's main machine learning components using the DNM dataset described in Section II-A.

A. Anomaly Detection

Our approach was evaluated on 35 DNM communities. To evaluate the performance of our anomaly detection module, we created a list of 10 known adverse events with specific dates, such as law enforcement interventions, and added a number of suspected adverse events that were identified through a manual Internet search against the timeline of anomalies found, which were not aligned with any known adverse event or in line with its cascading effect. The anomaly detection model has detected most anomalies that are aligned with adverse events. These adverse events connect to lots of startups and shutdowns on DNM communities, and the full details of results can be found in [26].

B. Online Cyber Offender Community Analysis

To evaluate our Social Network Analysis and topic modelling approach described in Section III-B, we targeted the Evolution forum from the DNM dataset. The Evolution dataset contained 509,225 messages written by 21,946 different users



(a) (b)

Fig. 7: Screenshots of the AMoC toolkit's account recognition analysis

in total, with on average 23.2 messages per user and 53.1 tokens per message. Each individual in the dataset contributed to on average 11.3 different threads.

The results of our SNA showed that over 95% of all users in the Evolution dataset were passive. As can be seen in Figure 8, the Evolution dataset is comprised of 4 larger sub-communities, which contained 11,725; 6,645; 1,885; and 926 users. The other sub-communities did not include more than 51 nodes. When applying a further cut-off of including users who produced at least 1,000 posts, our approach finally yielded 135 potential users of interest. Next, for each of these users, we extracted topic clusters to gain a better understanding of the thematic scope of their communications.

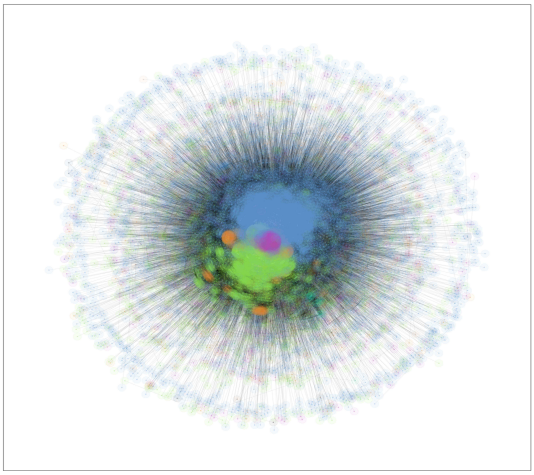


Fig. 8: Sub-communities detected in the Evolution dataset.

Because the texts provided to the topic detection model are unlabelled, no actual categorisation is performed and, hence, there is no evaluation of the accuracy of the output of the similarity algorithm. Therefore, we calculated the mean Silhouette Coefficient (SC), which is a measure to validate the consistency within the resulting clusters of data. The best SC was achieved when applying K-means using word embeddings, which resulted in an average score of 0.35.

C. Account Recognition

Evaluating the account recognition module proved to be highly challenging. The DNM dataset was extremely unbalanced in the number of matching and unmatching pairs. Not only because the majority of accounts would not disclose their public key information in the forum, but also because we require a valid account who has the public key information to have posted at least 5 messages in a forum. This led to 185 valid accounts in total that could be used for the positive samples, while we had 14,080 accounts who had public key information and also posted at least 5 messages but only in one forum as the negative samples. Therefore, we split 80% of accounts in the positive and negative samples at the same unbalanced ratio of positive and negatives so as to train the model, and 20% in order to test the model. There then remained 2,816 negative account pairs and 37 positive account pairs for our evaluation. However, in our training with the triplet loss, each anchor will combine one positive matched sample and only one unmatched negative sample, the siamese network thus would not be impeded by the extremely unbalanced dataset. In the model inference, since there is no pair for negative accounts, each negative account will combine another random account who has no public key information at this same forum to form as a pair of negative sample. This extremely unbalanced dataset could significantly affect the inference result.

As can be seen in the confusion matrix in Table II, the model successfully detected 31 matched account pairs and 1,816 non-matched account pairs at the inference stage, respectively.

TABLE II: Confusion matrix

		Actual		Support
		Matched	Non-matched	
Predict	Matched	31	6	37
	Non-matched	1000	1816	2816

TABLE III: Performance evaluation

	Precision	Recall	F1-score	Support
Matched	0.03	0.84	0.06	37
Non-matched	1	0.64	0.78	2816

Thus, the recall for the matched and non-matched achieved 0.84 and 0.64, respectively, which is remarkably important in the account recognition. However, due to the extremely unbalanced dataset for the matched and non-matched samples, the F1-score for the matched one, as expected, proved to be much more challenging than the non-matched one in Table III.

IV. Related work

Unlike other works that focuses on a single purpose to investigate the cyber crmine, AMoC combines multiple machine learning techniques together to explore the target community or user comprehensively. Here we briefly review relevant work and discuss their connections to AMoC toolkit in these areas.

A. Anomaly Detection

There are a few studies [8], [21], [22] that have applied anomaly detection on darknet data. However, their focus lay in identifying the threats posted from the darknet to legal communities. For example, exploring hacker assets in the darknets [21] and identification of hacker threats [22]. On the contrary, this work explores the anomaly detection regarding to anomalous behaviour of DNM users in response to events considered a threat to DNMs themselves, which is different to previous works in social media and network intrusion detection.

B. Online Cyber Offender Community Analysis

To our knowledge, there is only one study that focused on identifying key users in terms of roles, influence levels, and their social relationals: Huang et al. [11] proposed a topic-based social network analysis approach with unsupervised clustering methods for identifying the key members and their associated roles in the Chinese cyber fraud underground economy. Based on the results of their Latent Dirichlet Allocation (LDA) analysis, they attributed user roles based on the keywords of topics detected in user communications. However, their role typology was based on prior work in the area of the underground economy of credit card fraud [24], and, hence, limited to "attack originators", "buyers", "droppers", "shoppers", "runners", and "other sellers". Such a typology, which is mainly based on forum structure rather than the role users play in a community, or even titles assigned to users during their registration, may not reflect the actual roles users adopt in the market.

In this paper, we combine novel SNA and NLP techniques to perform a bottom-up, dynamic analysis of detecting different user roles in cybercriminal communities on the darknet and categorise offenders within this dynamic role typology based on the thematic scope of their conversations.

C. Account Recognition

Contemporary computational stylometry research typically focuses on two aspects: recognising and extracting linguistic features that are potentially discriminative for an author's *writing print* (or *stylome* [27]) and developing an efficient computational model that includes these features to automatically determine an author's identity or demographics. Although a range of feature types and computational methods have been suggested for the task, the field is dominated by studies that evaluate their computational stylometry approaches on non-deceptive datasets (see e.g., [1], [16]). However, a key issue when designing a computational stylometry approach to be used in cybercrime investigations is whether it will remain useful when it is confronted with adversarial behaviour [15]. Cyber criminals may try to hide behind multiple digital personas or a group of offenders can share a single online identity. Additionally, they might attempt to hide their true identity or imitate other (non-criminal) users and use specialised vocabulary or coded language to conceal the nature of their activities [2].

This study investigates the feasibility to recognise cyber offenders across darknet market fora and linking different user-ids within a single forum based on stylometric features in their online communications.

V. Conclusion

This article presents a multifaceted, unsupervised machine learning approach to analysing cybercriminal communities based on cyber offender communications on darknet fora. The computational approaches developed to study cybercriminal communities have been implemented into a proof-of-concept AMoC toolkit for use by law enforcement practitioners. The toolkit supports:

- Structured queries on DNMs under analysis, including basic queries (e.g., total number of posts, accounts, and communities), aggregate queries (e.g., top accounts by number of posts), and zoomed in views (e.g., top active accounts in a target community by number of posts);
- Detection of anomalies in a target community that may indicate disruptive events, e.g., closure of other markets, concerns about breaching of the market by law enforcement or concerns about large-scale fraud within the market, which then can be used to build the darknet ecosystem by connecting the adverse events through anomalies found;
- Automated analysis of cyber offender target communities, enabling the detection of sub-communities within forums or marketplaces, the identification of the key accounts within these sub-communities, and the analysis of the thematic scope of their communications;

[2]For example, illegal drug traffickers have been reported to use a widely varied terminology for selling their products [20].

- Recognition of a user based on comparing text (paragraphs or sentences) from posts in different communities. Two alternative approaches are available to the analyst to estimate whether two seemingly different accounts may be the same person: a neural network or traditional linguistics readability indices, such as Gunning Fog index, Flesch-Kincaid index, average syllables per word, and average sentence length, etc.
- Desktop app with a friendly user interface that enables a user to access the above functionalities with different parameters;
- Possibility to integrate the individual components, implementing the above analyses into existing investigative workflows within law enforcement settings.

A key limitation of evaluating the different components was the lack of ground truth labels regarding accounts recognition, their roles in a cybercriminal community (SNA), and the topics discussed in their communications (topic modelling). The team is currently liaising with law enforcement partners to enable testing of the toolkit on new and validated DNM data. Additionally, word embeddings that were pre-trained on non-darknet data might include semantic similarity assumptions that do not uphold when applied to darknet communications between cyber offenders. For example, offenders can use guarded language or specialised vocabulary in order to hide their illegal activities from law enforcement investigators. Therefore, we intend to train new word embedding models on the entire DNM dataset and include them in our experiments.

REFERENCES

[1] Georgios Barlas and Efstathios Stamatatos. Cross-domain authorship attribution using pre-trained language models. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 255–266. Springer, 2020.

[2] Gwern Branwen, Nicolas Christin, David Décary-Hétu, Rasmus Munksgaard Andersen, StExo, El Presidente, Anonymous, Daryl Lau, Delyan Kratunov Sohhlz, Vince Cakic, Van Buskirk, Whom, Michael McKenna, and Sigi Goode. Dark net market archives, 2011-2015. https://www.gwern.net/DNM-archives, July 2015. Accessed: 2021-10.

[3] Adrian Chen. The underground website where you can buy any drug imaginable. Wired, 6. 2011. [Online] www.wired.com/2011/06/silkroad-2 Accessed: 2021-10.

[4] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111, 2004.

[5] David Décary-Hétu. Police operations 3.0: On the impact and policy implications of police operations on the warez scene. *Policy & Internet*, 6(3):315–340, 2014.

[6] David Décary-Hétu and Luca Giommoni. Do police crackdowns disrupt drug cryptomarkets? a longitudinal analysis of the effects of operation onymous. *Crime, Law and Social Change*, 67(1):55–75, 2017.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[8] Yaokai Feng, Yoshiaki Hori, Kouichi Sakurai, and Jun'ichi Takeuchi. A behavior-based method for detecting distributed scan attacks in darknets. *Journal of Information Processing*, 21(3):527–538, 2013.

[9] Thomas J Holt. Lone hacks or group cracks: Examining the social organization of computer hackers. *Crimes of the Internet*, pages 336–355, 2009.

[10] Martin Horton-Eddison and Matteo Di Cristofaro. Hard interventions and innovation in crypto-drug markets: The escrow example. *Policy Brief*, 11, 2017.

[11] Shin-Ying Huang and Hsinchun Chen. Exploring the online underground marketplaces through topic-based social network and clustering. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pages 145–150. IEEE, 2016.

[12] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.

[13] Alice Hutchings and Richard Clayton. Exploring the provision of online booter services. *Deviant Behavior*, 37(10):1163–1178, 2016.

[14] Seidel Jon. World's most prolific online drug dealer 'supertrips' gets 10 years. Chicago Suntimes, 5. 2015, 2015. [Online].

[15] Patrick Juola. Authorship attribution. *Foundations and Trends® in Information Retrieval*, 1(3):233–334, 2008.

[16] Moshe Koppel, Jonathan Schler, and Shlomo Argamon. Authorship attribution in the wild. *Language Resources and Evaluation*, 45(1):83–94, 2011.

[17] Wesley Lacson and Beata Jones. The 21st century darknet market: Lessons from the fall of Silk Road. *International Journal of Cyber Criminology*, 10(1), 2016.

[18] Linda Marric. The underground website where you can buy any drug imaginable. New Scientist, 3 2021. [Online] http://www.newscientist.com/article/mg24933260-400-silk-road-review-the-true-story-of-the-dark-webs-illegal-drug-market Accessed: 2021-10.

[19] National Cyber Crime Unit, Prevent Team. Intelligence assessment: Pathways into cyber crime, 2017. Available at http://www.nationalcrimeagency.gov.uk/publications/791-pathways-into-cyber-crime/file.

[20] Samuel Nunn. 'Wanna still nine hard?': Exploring mechanisms of police bias in the translation and interpretation of wiretap conversations. *Surveillance & Society*, 8(1):28–42, 2010.

[21] Sagar Samtani, Ryan Chinn, and Hsinchun Chen. Exploring hacker assets in underground forums. In *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 31–36. IEEE, 2015.

[22] Sagar Samtani, Hongyi Zhu, and Hsinchun Chen. Proactively identifying emerging hacker threats from the dark web: A diachronic graph embedding framework (D-GEF). *ACM Transactions on Privacy and Security (TOPS)*, 23(4):1–33, 2020.

[23] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.

[24] Abhinav Singh. The underground ecosystem of credit card fraud. *Black Hat Asia*, 2015.

[25] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112, 2014.

[26] Ziauddin Ursani, Claudia Peersman, Matthew Edwards, Chao Chen, and Awais Rashid. The impact of adverse events in darknet markets: an anomaly detection approach. In *Workshop on Attackers and Cyber-Crime Operations*, 2021.

[27] Hans Van Halteren, Harald Baayen, Fiona Tweedie, Marco Haverkort, and Anneke Neijt. New machine learning methods demonstrate the existence of a human stylome. *Journal of Quantitative Linguistics*, 12(1):65–77, 2005.