Characterizing Eve: Analysing Cybercrime Actors in a Large Underground Forum

Sergio Pastrana[1]✉, Alice Hutchings[1], Andrew Caines[2], and Paula Buttery[3]

[1] Cambridge Cybercrime Centre, Dept. of Computer Science & Technology,
University of Cambridge, United Kingdom
{Sergio.Pastrana,Alice.Hutchings}@cl.cam.ac.uk

[2] Theoretical & Applied Linguistics, Faculty of Modern & Medieval Languages,
University of Cambridge, United Kingdom
apc38@cam.ac.uk

[3] Natural Language & Information Processing, Dept. of Computer Science &
Technology, University of Cambridge, United Kingdom
pjb48@cam.ac.uk

Abstract. Underground forums contain many thousands of active users, but the vast majority will be involved, at most, in minor levels of deviance. The number who engage in serious criminal activity is small. That being said, underground forums have played a significant role in several recent high-profile cybercrime activities. In this work we apply data science approaches to understand criminal pathways and characterize key actors related to illegal activity in one of the largest and longest-running underground forums. We combine the results of a logistic regression model with k-means clustering and social network analysis, verifying the findings using topic analysis. We identify variables relating to forum activity that predict the likelihood a user will become an actor of interest to law enforcement, and would therefore benefit the most from intervention. This work provides the first step towards identifying ways to deter the involvement of young people away from a career in cybercrime.

Keywords: Cybercrime, Underground forums, Social behaviour, Criminal pathways

1 Introduction

Cybercrimes carried out by organized groups using custom tools with political or military motivations capture the public imagination. However, the vast majority of attacks are committed by actors with a low level of technical sophistication [24,34]. While these may receive less media attention, they can cause large financial losses and be costly to defend against [3]. This criminality is to a great extent promoted by an active underground economy where attack tools and services are traded, and cyber attacks are monetised [2].

Online underground forums bring together individuals interested in cybercrime and illicit online monetizing techniques [22,12]. In contrast with other forms of crypto-markets [31], some of the contents of these forums are legal, such as discussions relating to current events, gaming, and technology-related issues. However, these forums are also used to exchange information about deviant behaviour, and trade in goods and services with an illicit origin or application. Previous research has found these forums can provide a stepping stone towards more serious online criminal activities [13,14].

The underground economy attracts actors that are unlikely to be involved in traditional crime, but who may become involved in cybercrime [23]. For example, the use of booter services for 'DDoSing' others has become a widespread phenomenon among school-aged children, and even victims can become attackers [24]. This is due to the ease of access to hacking tools, the sense of anonymity provided by the Internet, and the perceived lack of law enforcement online.

Cybercrime has proliferated in recent years, and online forums have become a key source of data for researchers (see Section 2 for related work). While insightful, this research has mainly relied on cross-sectional data, analysing forum content from short periods of time or focussing on particular areas of cybercrime. Typically, researchers have considered only the tools and technologies adversaries use, not their motivations or personal context [10]. Understanding not only 'what' is traded in underground economies, but also 'why' and by 'whom' can provide insights into ways to tackle cybercrime from multiple perspectives. The evolution of offenders, understanding how they learn to commit crime over time, is a key aspect of this. Multidisciplinary research on the behavioural aspects of cybercrime is necessary to develop defences aimed at understanding and preventing incidents, rather than stopping or recovering from them.

In this paper, we analyse the characteristics and pathways of 'key actors'; forum users who have been linked to criminal activities, such as providing services and tools to disrupt systems and networks or using these tools to perform attacks. We use a variety of sources to identify these actors (see Section 3). While we do not publish this list for ethical reasons (see Section 6), activities linked to these key actors include providing DDoS as a service, distributing malware, operating bot shops and pay-per-install services, as well as providing services for web exploitation and account cracking. Characterizing key actors and analysing their evolution within forums is beneficial for various reasons. From a social perspective, it is the first step towards identifying ways to deter people away from criminal activities. From the cybersecurity perspective, these actors provide state-of-the art tools and techniques that can be used for attacking systems. This information can be used by response teams and security firms to focus their attention, increasing their capacity to react rapidly to new forms of attack.

We focus our study on *Hackforums*, one of the largest underground forums. *Hackforums* is well established, operating since 2007. While this forum is known to be overrun by novice teenage hackers (contemptuously dubbed 'script kiddies'), in the last few years there have been a number of high profile attacks directly related to products distributed through this forum. For example, in September 2016, the *Mirai* source code was released on the forum, which led to several related botnets being used for illegal activities such as DDoS attacks [4] or mining cryptocurrencies [21]. In the first three months of 2018, there have been at least two cases relating to *Hackforums* users before the UK courts.

We use the CrimeBB dataset, which includes *Hackforums* data spanning from 2007 onwards and contains information about 572k user accounts [27]. We start by identifying key actors on the forum. We first apply data science approaches to present a longitudinal study of these key actors. Concretely we apply social network analysis to analyse their social interests, natural language processing to classify the type of information posted, and clustering to group the actors based on forum activity. Our research uncovers common activity patterns and the pathways taken over time in terms of interests and knowledge. Second we develop tools to identify factors that predict involvement in cybercrime. These tools use social network analysis, logistic regression, and clustering to preselect a list of potential actors, and topic analysis to analyse the type of information they post. Our findings suggest that combining the different techniques helps in the prediction of potential actors. These tools can be applied to any particular cybercrime domain, so we make them publicly available. The CrimeBB dataset also contains data from other forums and is available to academic researchers through data sharing agreements from the Cambridge Cybercrime Centre.[1]

2 Background and related work

The rise of cybersecurity incidents parallels the development of underground economies, where attacking tools and services are easily accessible at low cost or even for free [2]. For example, pay-per-install services outsource the task of infecting a machine and allow miscreants to buy 'installs' for spreading their malware [6]. Other common assets that can be found in underground forums are bot shops and botnets [8], crypters and packers [30], or exploits [2].

Various authors have addressed the offenders perspective. Karami and McCoy analysed leaked databases of booter services: websites providing DDoS for hire, publicly marketed as network 'stressers', but offered in underground forums as services to perform DDoS [16]. While mostly used to take down gaming servers, booters are also used to attack medium-sized websites. Hutchings and Clayton researched the provision of denial of service attacks, interviewing and surveying the providers to ask how they began providing the services, and why [14]. They found most operators were young men from North America. They had escalated from using booter sites, to setting them up and running them themselves. They were initially exposed to booter services through gaming and hacker communities. Financial gain was the main reason for providing services, but they also reported they enjoyed the challenge of their activities.

Sood and Enbody analyse the provision of cybercrime tools and services, identifying three type of actors in underground communities: *providers* or *producers*, *advertisers*, and *buyers* [30]. Based on our analysis of underground forums, we add two new roles. First, *re-distributors* of modified versions of public

[1] https://www.cambridgecybercrime.uk/

or leaked malware. This role includes users involved in the provision of encrypted malware binaries, aimed at avoiding detection by antivirus software. The second role we dub *teachers*: actors who provide tutorials for configuration and use of various attack tools, sometimes accompanied by help-desk services.

According to the criminological theory of differential association, criminal activities are normal behaviours learnt in interaction with others [33]. Learning takes place by associating with others in personal groups. The content of what is learnt includes specific techniques to commit crime, as well as the 'definitions' (mindset) favourable to committing crime [37]. In relation to cybercrime, there is evidence that offenders associate with each other in physical space [18], but also online, particularly through the use of online underground forums [10,15,38].

Understanding offender pathways allows society to consider the most appropriate ways to divert potential offenders away from crime. For example, the UK's National Crime Agency (NCA) [23] debriefed young people involved in cybercrime activities, and found many were first exposed through their interest in gaming. The NCA have subsequently been working with the video gaming industry to deliver preventative interventions.

Underground forums serve as an entry point into cybercrime for potential offenders. These forums also allow non-technical actors to learn how to commit offences and develop their skills [29]. Normally these forums have well-defined categories like "Hacking" or "Market". Where authors are most active provides insights into their interests and expertise [25]. Forum members have a public profile with information such as the registration date, last access or time spent. Most underground forums are publicly accessible on the surface web or the 'dark web' (e.g. through Tor hidden services).

The success of underground economies relies on trust and informal social control [1]. Various authors have analysed these behaviours using social network analysis (SNA), for example to analyse the evolution of members in terms of posts and private messages [22] or to understand specialization and developments of subcommunities [11]. The use of natural language processing (NLP) to analyse underground forums is also a recurring technique, e.g. to analyse post sentiment [19] or to identify the assets being traded or the currencies used [29,28].

3 Dataset

In this work we use the CrimeBB dataset [27], which contains data collected from various underground forums. We focus our study on *Hackforums*, the largest forum contained in this dataset, with more than 30m posts[2] made by 572k user accounts over more than 10 years. *Hackforums* is divided into nine categories: Hacking, Technology, Coding, Gaming, Web, Market, Money (a miscellaneous category for all sorts of money making methods), Graphics and Common (which

[2] We refer to a whole website as a *forum*, on which pages are set aside for discussion of defined topics in *boards*, with users participating in conversation *threads* via individual *posts*.

includes boards for discussion about various topics, such as entertainment or politics, and boards intended for forum rules and suggestions).

3.1 Key Actors

We use the term ‘key actor’ to refer to forum users who have been linked to cybercrime activities, such as distributing malware, offering off-the-shelf tools to perform denial-of-service attacks or using these tools to attack others. A number of approaches were utilised to identify key actors who are or have been active on *Hackforums*. These approaches required manual effort and thus are not scalable. In Section 5 we propose tools to automatically identify likely key actors.

1. Media sources were searched to identify reports relating to *Hackforums* users being arrested or prosecuted for cybercrime activities (media included official notifications from law enforcement agencies; forum threads; social media and blog posts made by security researchers). We used Google extended search to look for sources including keywords such as ‘arrested’ or ‘prosecuted’ and ‘hackforums’. Results often included the pseudonym used by the actor in the forum. This method yielded 49 key actors.
2. A private security and intelligence company, Flashpoint, provided usernames considered to be of interest due to their activities. This yielded 9 key actors.
3. For each actor identified using the methods above, we used SNA to find their ‘closest’ neighbours (users of the forum who they interact with the most). Then, we manually analyse the activity of these neighbours looking for evidence of involvement in cybercrime activities (for instance, evidence of providing illegal material such as malware or ‘booter’ services). This method yielded a further 22 actors.
4. The final set of key actors are those providing tools aiming at disrupting systems and/or networks. To identify these actors we had two approaches:
   - We searched *Hackforums* for threads advertising the top 300 Remote Access Trojans (RATs) reported in [36]. Again, from manual inspection we identified the owners/coders of RATs and the re-distributors of modified versions (e.g. encrypted binaries aimed at avoiding antivirus detection). We discarded actors who we believed to be only purporting to be an owner (a *stealer*); and also actors distributing an infected version of a binary with the intent of compromising other forum members. This method yielded 35 key actors (there was some overlap with actors previously extracted).
   - We used ‘compilation’ threads from *Hackforums*, where popular tools and services are listed accompanied with the corresponding thread where it was first advertised. This method yielded 15 key actors.

In total, these methods yield 130 actors of interest to law enforcement: of these, we were able to identify the accounts of 113 within the dataset. The missing accounts might be due to accounts being removed or changes of the pseudonyms which we were unable to track. Also, it should be noted that various accounts might belong to the same actor.

4 Characterizing key actors

Having identified 113 key actors, we applied a number of different data science approaches to analyse the forum activity of these users, including NLP, SNA, and machine learning algorithms.

4.1 Natural language processing

Due to the massive size of the dataset (more than 30m posts), it is not possible to manually code the data. We use NLP tools to classify posts into categories. Classification of interests and expertise of members enables the identification of topics related to cybercrime offences, such as learning to attack systems or trading in stolen accounts. The data poses interesting problems for NLP techniques. The language used by members of underground forums includes technical jargon and non-standard means of expression. Contributors include non-native speakers of English, and short texts in which information is conveyed in deliberately concise ways. In this work we analyse the behavioural evolution of our identified actors, firstly building a binary classifier to identify questions in CrimeBB.

Three annotators manually labelled 2,200 posts selected from a range of boards, with substantial inter-annotator agreement for post type (see more details in [7]). We use the annotated dataset to train and test the classifier, with a training subset of 175 annotated threads from various boards, and a test subset of 186 annotated threads from another board (to prevent overfitting). For each thread, we extract features using a set of statistical techniques and a set of heuristics, having found this hybrid approach to work best [7]. The former include the number of replies, the number of links in the first post (both to external sources and to other threads in the forum), the length of the first post and a set of unigram features extracted from text. We convert every thread title and post into a document-term matrix (a matrix of counts with each word occurring as column values, and each of the documents as a row). We strip punctuation, convert to lower case characters, ignore numbers and exclude stop words. Finally, word counts are transformed using TF-IDF ('term frequency inverse document frequency'), a weighting that promotes words occurring fairly frequently in few documents above those occurring highly frequently but ubiquitously across CrimeBB [32].

The heuristics are formed through our expertise in analysing forum data. Concretely, for each thread we get the frequency of particularly interesting keywords in the heading and first post (examples of these keywords are "looking for", "I need help" or "I have a question"). Finally, we also account for the number of question marks in the heading.

We use a Linear SVM to build a classifier. Again, the selection of the algorithm is based on previous experimentation with the dataset [7]. For evaluation we use the usual metrics for information retrieval, i.e. precision, recall and F1. Precision measures the fraction of actual questions retrieved among the total of questions retrieved (including false positives). Recall, or sensitivity, measures the fraction of questions retrieved among the total number of actual questions in the dataset. Finally, the F1 score combines in a single measure both precision and recall. Our classifier has Precision=0.88, Recall=0.85 and F1=0.86. While these metrics can be improved, the classifier is accurate enough to automatically identify question threads, a task which would otherwise be infeasible due to the size of the dataset.

4.2 Social network analysis

We designed and developed SNA tools to facilitate study of the forums at different levels of granularity, per board, per topic of interest, per year, etc. We build the social network by processing the public interactions of the members. This network is represented as a directed graph, where nodes are the members of the forum and edges their interactions. We define a directed edge from node V to node W if there is a *reply* from V to W. There are two possible forms of reply: a) when V explicitly cites a post made by W; and b) when V replies in a thread initiated by W. When available, we use information from reputation votes given between members to classify the interactions as positive, negative or neutral.

We use classical SNA metrics such as centrality degrees to analyse the network, i.e. in-degree (fraction of nodes its incoming edges are connected to), out-degree (fraction of nodes its outgoing edges are connected to), and eigenvector (measure of the influence of a node in a network). Additionally, we compute the following metrics to measure the popularity of the forum users: total number of replies; h-index (a member with h-index=n is author of n threads having at least n replies); and the i-10-index, i-50-index and i-100-index (i.e. the number of threads with at least 10, 50 and 100 replies respectively). These metrics are used in academia to measure the productivity and impact of a scholar. We adopt them to analyse underground forums for the same purpose.

We also developed tools to analyse the interests of forum members. This allows us to study the networks of actors interested in particular topics. Interests can be calculated for a given period, so we can analyse the evolution of different actors (e.g. a member initially interested in gaming related boards who then moved to hacking related boards). The interest of a member M in a board B is calculated as:

$$\mathrm{I}(M,B) = \mathrm{N}_{\mathrm{T}}(M,B) * 3 + \mathrm{N}_{\mathrm{P}}(M,B)$$

Where $\mathrm{N}_{\{\mathrm{T},\mathrm{P}\}}(M,B)$ denotes the number of {threads, posts} written by M in B. We assign triple weight to threads since initiating a thread represents a greater interest than posting a reply.

4.3 Machine Learning - Clustering

Machine learning techniques can be applied to extract common characteristics from a dataset. We apply k-means clustering to group the actors based on their activity [17]. K-means partitions a set of n samples into k clusters (with $k << n$).

We extract a set of 44 features for each actor, which can be classified as measures relating to forum activity, social relations, and reputation measures.

Measures relating to forum activity includes the number of days between the first and last post, the number of posts and threads in each category and the number of posts and threads in the currency exchange board. We explicitly include the currency exchange board (which is part of the marketplace) as it characterizes the financial activities of the actors.

Network centrality measures are obtained from SNA. These include out-degree, in-degree, eigenvector, h-index, and i-10 and i-100 indices.

Reputation measures are taken from the reputation systems used on the forum. These include the overall reputation bestowed and prestige scores (prestige is an forum metric based on activity). There are also counts for the number of positive, negative, and zero-value reputation votes each account received.

Then, using the feature set we perform clustering using k-means. After applying the Elbow method [35] to analyse the within-group sum of squares for various values of k, we set $k = 5$.

4.4 Results

Using the tools described above, we first analyse the social relations established between key actors and their closest neighbours. Second, we analyse their common characteristics by splitting them into groups using k-means clustering. Finally, we analyse their pathways by looking for changes in their interests and the number of questions posted as they spend more time in the forum.

Social relations Figure 1 shows the social network involving the key actors.[3] The actors identified from media sources and Flashpoint are filled in red, the ones identified from network analysis are orange and the those linked to malware distribution are blue. Colours of the edges represent the sentiment of the relationship, calculated from the reputation votes sent to each other. Most key actors are closely connected to each other, and most relationships are positive. Actors obtained from different sources are closely or even directly connected. For example, the detail in Figure 1 shows a member identified as malware distributor (in blue) which is directly connected to one identified through SNA (in orange) and very close to at least two actors identified from media sources (in red).

Some close neighbours (for example, the nodes tagged as 'Bridge' in the detail from Figure 1) are connected to more than one key actor, and act as 'bridges' for connecting different groups. These actors are of interest since they might be influential for or influenced by key actors in criminal activity. Accordingly, we use these actors for our prediction study presented in Section 5.

Characterization Table 1 shows the average values for each of the five clusters obtained by k-means. There is a small group of 5 actors who have the highest measures of forum activity, are highly reputed (though they also receive high negative votes), and have rich social relations. These 5 actors are popular

[3] For the sake of visualization, the figure only shows the key actors and their five closest repliers and replied neighbours (filled in green).

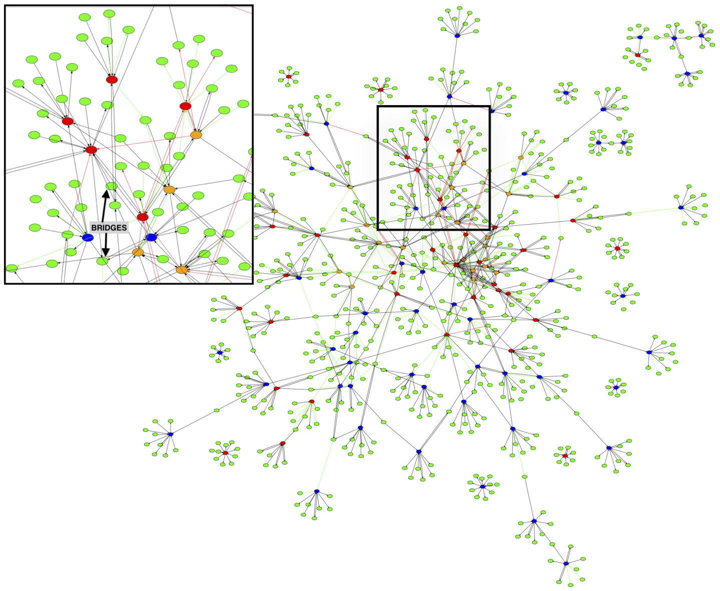

Fig. 1. Social network graph involving key actors and their closest neighbours (green nodes). Red nodes are those identified from media sources and Flashpoint, orange are those identified through SNA and blue are those linked to the distribution of malware. The colours in the edges represent the sentiment of the relationship (red=negative, green=positive and black=neutral/unknown)

(due to the high values of their H and i indexes), have influence in the network, and are well known in the community. The remaining clusters have also been active for long time (more than 2 years) but differ in quantity of posts and threads. The clusters are also differentiated by their areas of interests.

The cluster with 20 actors is most interested in the market section (followed by the common section). They are the most active group in currency exchange and have high social relationship measurements (e.g. on average they have 7.2 threads with more than 100 replies). Overall, actors in this cluster are likely to be known in the community as prolific market traders.

The clusters with 27 and 37 members have similar interests (mostly in market and hacking, but also in common and coding categories), though one has higher reputation (mostly positive) and social relations (e.g. they have more than twice the number of threads with at least 10 replies). Finally, the least active cluster, which is composed of 24 actors, is interested firstly in hacking and then in the market sections, with negligible posts in currency exchange.

Overall, cluster analysis suggests key actors are mostly characterized by their interest in the market, common, and hacking areas. Also, they can be grouped by their forum activity, with some being more active and popular, and thus well known within the community, while others are less active, do not participate in the common sections of the forum and are less popular.

Table 1. Average values for key actors grouped in 5 clusters. The Interests columns show the top 3 categories and number of posts/threads in currency exchange. W=Web, G=Game, D=Code, T=Tech, C=Common, H=Hack, $=Money, X=Graphics, M=Market. +=positive reputations, 0=neutral reputations and -=negative reputations. EV=Eigenvector

#KeyActors	Activity		Interests				Reputation	Social relations			
	Days	Threads/Posts	cat1	cat2	cat3	#CurExc	Total (+/0/-)	H	i10	i100	EV
27	1298.4	74.1/1138.4	M	H	C	3.9/7.6	229.8 (61.3/2.3/4.3)	10.4	15.4	1.1	0.00
37	1595.0	163.8/3338.1	M	C	D/H	6.4/19.9	482.8 (230.9/7.4/6.9)	17.6	41.7	3.0	0.01
5	1951.0	831.0/18086.2	C	M	H	23.8/125.4	896.8 (578.2/68.8/99.0)	53.6	373.0	23.2	0.04
24	796.4	18.0/413.0	H	M	C/D	0.0/1.0	120.1 (58.0/2.4/3.2)	5.0	4.5	0.3	0.00
20	1895.7	383.6/10989.2	M	C	H	27.4/141.8	667.9 (311.6/27.0/48.3)	28.4	99.8	7.2	0.02

Evolution We track the interests of the actors since they were registered until their last visit (if enabled on their profile) or last post. We compute their interests in each board and then aggregate them per category and per year. To analyse temporal evolution, we measure the interests at the beginning, middle and end of the period each actor has been active. The beginning is defined as the year of their first post, the end is the year of their last post, and the middle is the period in between. We then calculate the evolution of interests between these periods by computing *transitions of interest*. Concretely, a transition of interest from a category C_i in time t_0 to a category C_j in time t_1 is calculated as:

$$T(C_i^{t_0} \rightarrow C_j^{t_1}) = \sum_{\forall A \in K} (|S^{t_0}| - \beta_i^{t_0}) * \lambda_i^{t_0} + (|S^{t_1}| - \beta_j^{t_1}) * \lambda_j^{t_1}$$

Where K is the set of all the key actors, S^{t_n} denotes the set of all categories of interest for actor A in time t_n, $\lambda_i^{t_n}$ denotes the normalized interest of actor A in category C_i in time t_n, and $\beta_i^{t_n}$ is the relative position of category i regarding the ordered list of categories by score in time t_n (i.e., the top category has β_i^t equal to 1, the second equal to 2 and so on). The above equation weights the categories of interest per actor according to the amount of posts and threads posted in each category with respect to the rest.

Figure 2 shows the aggregated transitions for all key actors. Overall, actors are most interested in the hacking, market, and common categories. Over their time in the forum, there is a slight increase of interest in the coding and technology sections, and a decrease in the gaming sections. From this figure we can draw several conclusions. First, in general actors are active participants in non-criminal related boards, such as those from the common category. This suggests their criminal activity runs in parallel or comes after other interests (e.g. entertainment or gaming), and they are involved in other activities within the community. Second, their high interest in the marketplace and money sections indicates they may have financial motivations. Third, as they get older and more experienced in the forum they are less likely to engage in gaming boards.

Prior research has found forums are used for sharing information and learning about cybercrime and deviant activities [10,15,38]. Thus, we analyse the

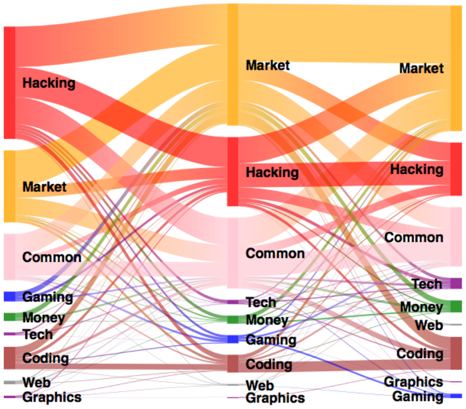

Fig. 2. Evolution of interests of key actors from initial (left), halfway (middle) and end (right) of their activity in *Hackforums*

evolution of the actors in terms of the number of questions (or requests for information) posted across time. In order to track evolution, this analysis includes the 34 key actors who have been posting for at least 4 years. We count the number of posts and number of questions posted for each year since they wrote their first post.

Figure 3 shows the proportion of questions posted per year with respect to the total number of questions posted. Each row represents a different actor (the top row shows the aggregation of the 34 actors). Most actors posted more than half of all their questions during their first or second year of activity in the forum. However, there are other actors (e.g. A1, A2, and A3) that keep posting questions at a similar rate after 5 or 6 years of activity. We can confirm these actors posted more questions in the early stages of their activity in the forums.

5 Predicting key actors

We analyse over a decade of data from *Hackforums* to identify those variables relating to forum activity that predict the likelihood a user will eventually be an actor of interest to law enforcement. Actors were selected for inclusion if they had been active since 2009, and had made more than five posts on the forum. This way, we do not consider old and low profile actors which would otherwise introduce noise in our analysis. After the forum administrator was excluded from the dataset, there were 245,636 cases extracted.[4] Our prediction framework is based on two steps: using multiple approaches to select potential key actors based on their forum activity, and predicting which of these are key actors based on the key terms used in their posts. We first combine the outputs from a logistic regression model, k-means clustering and SNA to identify actors

[4] The administrator is a well known actor in *Hackforums*

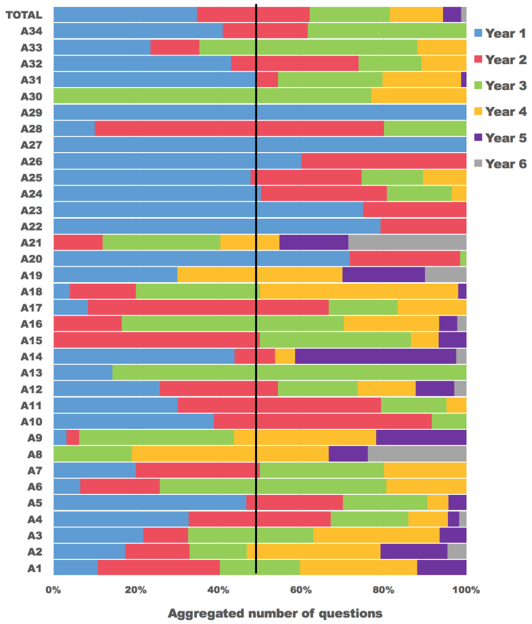

Fig. 3. Proportion of the number of questions posted per year with respect to the total questions posted

that are potentially involved in criminal activity. Second, we use topic analysis to confirm whether these users are engaged in cybercrime related activity, such as trading in illegal goods and services.

5.1 Logistic regression

We analyse the data applying backward stepwise logistic regression, using the likelihood ratio method. This method starts with a model which includes every independent variable, gradually removing every variable which does not have a significant impact on the dependent variable. Field [9] justifies the use of stepwise methods when carrying out exploratory research, in which there is no previous research on which to base hypotheses for testing, as well as situations in which causality is not of interest, but rather a model to fit the data. Both these justifications apply for this research. Field also recommends that if stepwise methods are to be used, then the backward method is the better option, as the forward method has a higher risk of Type II (false negative) errors. Logistic regression is often used in medical research, for example, to identify the risk factors associated with a disease within the wider population.

Logistic regression models predict a categorical outcome, in this case key actor status. Measures of forum activity, network centrality measures, and reputation measures (see Section 4.3) were considered for inclusion as predictor variables, however due to multicollinearity issues, a number were excluded (an assumption of logistic regression is that independent variables are not highly correlated). The independent variables included in the initial model are the number of days posting, reputation, prestige, posts and threads in the various categories, h-index, i-50-index, i-100-index, and number of positive, negative and zero-value reputation votes received.

As recommended by Field [9], 5 cases were removed as an analysis of the residuals indicated they had an undue influence on the model (Cook's Distance>1) Without any independent variables in the model, 100% of cases are predicted to not be key actors. The final model is significantly improved and is statistically better at predicting key actors (χ^2(15, n=245,631)=641.2, p<.001). The final model accounts for 34.1% of the variance, accurately predicting 11.1% of known key actors with a low false error rate (0.00%). While predicting 12 out of 108 key actors may seem low, it is from a pool of almost a quarter of a million cases. While not all the variance in the model will be explained through a user's forum activity, these significant results suggest this is an approach worthy of further exploration. The analysis also provides predicted probabilities for each user, which we explore further in Section 5.3.

Table 2 presents the results of the final step of the logistic regression analysis. The table includes regression coefficients, Wald statistics, odds ratios, and 95% confidence intervals for odds ratios for each of the 15 predictors retained in the model. The odds ratios, shown as Exp(B), show how the odds of being in one outcome category changes when the predictor variable increases by one unit.

Table 2. Logistic regression model predicting key actors

							95% C.I. for Exp(B)	
		B	S.E.	Wald	Sig.	Exp(B)	Lower	Upper
Step 15	DAYS_POSTING	.001	.000	19.407	.000	1.001	1.000	1.001
	REPUTATION	.001	.000	7.712	.005	1.001	1.000	1.001
	PRESTIGE	.006	.001	37.754	.000	1.006	1.004	1.008
	POSTS_HACK	.001	.000	25.397	.000	1.001	1.000	1.001
	POSTS_MARKET	.002	.000	65.945	.000	1.002	1.001	1.002
	POSTS_GAME	-.006	.001	15.670	.000	.994	.991	.997
	POSTS_GRAPHICS	-.009	.005	3.639	.056	.991	.982	1.000
	POSTS_CODE	.0005	.000	5.144	.023	1.0005	1.0001	1.0008
	POSTS_COMMON	-.0005	.000	4.945	.026	0.9995	.9991	0.9999
	POSTS_MONEY	-.003	.002	3.718	.054	.997	.994	1.000
	POSTS_CURRENCY_EXCHANGE	-.006	.003	6.041	.014	.994	.988	.999
	THREADS_GRAPHICS	-.044	.029	2.339	.126	.957	.905	1.012
	THREADS_COMMON	-.007	.003	5.637	.018	.993	.987	.999
	H_INDEX	.178	.017	108.025	.000	1.195	1.155	1.236
	NEGATIVE_REPUTATION	.018	.006	7.383	.007	1.018	1.005	1.031
	Constant	-9.372	.191	2397.372	.000	.000		

The odds ratios indicate that for each additional reputation and prestige point bestowed, the odds a user is a key actor increases by 1.001 and 1.006 respectively. For every additional day actors are posting, the odds they are key actors increases by 1.001. The frequency in which actors posted on various sections also predicts being a key actor, including posts in hacking (odds increased by 1.001 for each post), market (1.002) and code (1.0005) sections. Posts in some sections decrease the odds that users are key actors, including gaming (0.994), graphics (0.991, but this variable is not significant), common (0.9995), money (0.997, but not significant), and currency exchange (0.994). New threads initiated in the common and graphics sections decreases the odds a forum user is a key actor by 0.993 and 0.957 respectively, although graphics is not significant. An increase in a user's h-index increases their likelihood of being a key actor by 1.195. Key actors can also be predicted by their negative reputation (odds increase by 1.018 for each negative reputation).

5.2 Clustering

In addition to the logistic regression, we apply k-means clustering to the subset of more than 245k *Hackforums* actors. Table 3 shows the average values for each cluster applying k-means, using k=14, and which clusters the 113 key actors are grouped in. In the smallest cluster, 22 of 223 actors are key actors (9.9%). Actors from this cluster are very active, positively reputed and popular, and are most interested in the market, common, hacking, and gaming sections. Another small cluster of 2387 actors contains 31 key actors (1.3%). The profile is similar to the previous one, although the measurements are lower. Finally, the bulk of key actors (31) fall in a cluster with more than 10k actors, which is relatively smaller than other clusters. Again, the interests are within the market, common, hacking and gaming sections.

Most of the key actors are enclosed within the clusters with the fewest number of actors (relative to other clusters). This finding is interesting since it allows to reduce the amount of actors requiring thorough investigation when looking for criminal activity.

Table 3. Average values for actors grouped in 14 clusters. The Interests columns show the top 3 categories and number of posts/threads in currency exchange. W=Web, G=Game, D=Code, T=Tech, C=Common, H=Hack, $=Money, X=Graphics, M=Market. +=positive reputations, 0=zero reputations and -=negative reputations. EV=Eigenvector

#KeyActors /	Activity		Interests				Reputation	Social relations			
Total	Days	Threads/Posts	cat1	cat2	cat3	#CurExc	Total (+/0/-)	H	i10	i100	EV
1/8397	388.9	6.6/50.2	T/H	H/C	C/M	0.0/0.1	1.3 (0.4/0.0/0.1)	2.2	0.5	0.0	0.00
32/10323	1322.2	114.5/1310.2	M/C	C/M	G/H	3.5/9.8	113.9 (50.0/3.2/5.0)	11.6	17.4	0.5	0.00
0/4590	326.2	5.3/48.0	W	H	M/C	0.0/0.1	1.8 (0.6/0.1/0.1)	1.5	0.3	0.0	0.00
13/55364	338.6	7.3/46.4	H	M	C	0.0/0.1	0.7 (0.5/0.1/0.2)	2.3	0.5	0.0	0.00
9/41774	518.7	13.9/109.9	M	H/C	C/H	0.3/1.3	9.6 (3.4/0.3/0.5)	2.9	1.2	0.0	0.00
1/24202	310.9	5.7/56.2	G	H/C	M/H	0.0/0.1	2.0 (0.8/0.1/0.3)	1.9	0.7	0.0	0.00
0/36392	246.8	6.9/75.4	C	H	M	0.0/0.2	2.5 (1.1/0.2/0.4)	2.1	1.0	0.0	0.00
0/3474	296.3	3.8/90.6	T	H	C	0.0/0.1	4.1 (1.0/0.1/0.1)	1.1	0.3	0.0	0.00
0/14050	339.4	4.2/46.6	$	H	M/C	0.0/0.1	0.9 (0.4/0.1/0.1)	1.3	0.4	0.0	0.00
22/223	2111.7	611.2/11614.6	C	M	G/H	30.7/187.6	1170.7 (711.8/20.8/31.5)	32.2	162.8	8.2	0.03
3/9177	403.4	7.7/75.9	D	H	C	0.0/0.1	3.1 (1.1/0.1/0.2)	2.2	0.6	0.0	0.00
0/4845	302.2	6.9/71.0	X	H/C	M/H	0.0/0.1	5.1 (1.2/0.1/0.1)	2.1	0.8	0.0	0.00
31/2387	1723.8	295.9/4339.6	C	M	G	11.5/31.8	360.5 (170.2/9.8/13.8)	19.3	57.9	1.9	0.01
1/30437	215.8	0.2/18.2	H	M	C/$	0.0/0.0	0.2 (0.1/0.0/0.1)	0.1	0.0	0.0	0.00

5.3 Predicting actors using topic analysis

So far we have characterized and predicted actors based on features relating to forum activity, reputation and social behaviour. This provides a subset of actors who share common forum behaviour with those linked to illegal activities. To further refine the list of potential key actors, we pose the following research questions: *What* are the key actors talking about? Can we classify actors based on their topics of conversation? Next, we analyse the most frequent topics used by key actors. Then, we perform topic analysis on a selection of potential key actors obtained from the logistic regression, social network analysis and clustering.

Analysis of topics used by key actors. We use topic analysis to extract the most common terms from threads initiated by each actor. Topic analysis is an information retrieval task which produces wordlists summarised with a topic. Concretely, we apply latent Dirichelt allocation (LDA) to obtain the topics and terms that best represent the language used for each actor. Given a set of documents, LDA extracts the topics that best describe these documents [5]. A document is composed with the heading and first post of each thread initiated by an actor. We preprocess the data by tokenizing it, removing stop words, punctuation characters and numbers. Then, we extract the nouns using a Part-of-Speech (POS) tagger using the Penn Treebank tagset [20]. Using common NLP tools with low-resource language corpora presents limitations. Nevertheless, for this particular task the application of the POS tagger for extracting nouns reduces the number of noisy words.

For each actor, we extract 4 topics with 7 words per topic, resulting in 28 terms. Table 4 shows the most frequent terms used by the key actors (we show those used by more than five actors). The most common term is 'rat' (Remote Access Trojan) which could be expected given a bulk of key actors were identified due to their links with RAT coding. Various terms relate to offensive tools, such as 'bot', 'booter', 'crypter' and 'fud' ('fully undetectable'). Words related to commerce include 'paypal', 'btc' (Bitcoin), 'lr' (Liberty Reserve, a digital currency provider which was shut down in 2013), 'free', and 'cheap'. Also noteworthy is the high frequency of the words 'help', 'need' or 'question'.

Table 4. Most frequent terms used by the key actors. In parentheses are the number of key actors using each term. In bold are terms related to cybercrime.

rat (46), help (45), paypal (43), need (36), free (34), btc (34), **account** (33), thread (31), lr (28), server (26), new (25)
crypter (25), pp (25), source (23), **fud** (23), service (22), **bot** (21), question (20), hf (16), code (15), steam (15), site (14)
shell (14), cheap (14), money (14), skype (14), **booter** (13), window (12), anyone (12), tut (12), file (12), uid (11), someone (11)
system (10), vbnet (10), vpn (10), **installs** (10), please (10), member (10), php (10), problem (10), **ddos** (10), password (10)
website (10), update (10), setup (9), minecraft (9), email (9), game (9), vps (9), facebook (8), list (8), proxy (8), design (8)
darkcomet (8), **keylogger** (8), irc (8), java (8), coder (8), day (8), time (8), net (7), post (7), product (7), tool (7), beta (7)
sale (7), **exploit** (7), people (7), bitcoin (7), buying (7), **stealer** (6), version (6), **stresser** (6), live (6), feature (6)
botnet (6), domain (6), signature (6), shop (6), black (6), omc (6), web (6), year (6), support (6), official (6), youtube (6)

Selection of potential key actors. After analysing the most frequent terms used by key actors, we repeat the topic analysis with a subset of *potential* key actors identified from our previous analyses. The logistic regression provides predicted probabilities for each forum user. We obtain a subset (named *LogReg*) by selecting those with a predicted probability of 10% or more of being a key actor (n=88). From the clustering analysis we select 201 users (named *Clust*) from the cluster which contained the highest ratio of key actors (see Table 3). Finally, from our social network analysis, we select 42 actors (named *SNA*) directly connected with at least 3 key actors (see Figure 1). There are common actors between subsets: 10 actors appear in the three subsets; 26 appear in the *LogReg* and *Clust* subsets, but not the *SNA*; and 7 appear in the *SNA* and *Clust* subset, but not the *LogReg*. There are no overlaps between only the *LogReg* and *SNA* subsets. The final subset of potential key actors includes 285 forum users.

Predicting key actors. We apply topic analysis to the potential actors, extracting their 28 most common terms. We then measure the number of common terms with those obtained for the key actors to get a similarity score. This score is calculated as the number of terms matching the list of terms from key actors (Table 4) divided by the total number of terms extracted for the actor. However, similarities may be due to commerce-related terms (e.g. 'btc' or 'cheap') or forum-related terms ('need', 'thread' or 'help'). Thus, we also look for particularly interesting terms related with hacking (highlighted in Table 4).

As a prediction threshold, we establish a minimum distance of 0.2 (i.e. at least a 20% of the terms must match with those observed in the key actors) and a minimum number of 2 keywords observed.[5] Table 5 summarizes our findings. Using these thresholds, we predict 22 actors from the *LogReg* subset, 34 from the *Clust* subset and 9 from the *SNA* subset. We also predict 8 actors from the overlap of the *LogReg* and *Clust* subsets. From the 10 actors that were common in the three subsets, 7 are predicted to be key actors. The closest members to key-actors according to their topics are those identified with clustering. However, only 20% of users from this subset are predicted to be key actors. Meanwhile, 42% of the users from the logistic regression subset are predicted to be key actors. Our findings suggest combining different data science techniques assists in the prediction of potential key actors.

Overall, from the list of 285 potential key actors, 80 are predicted to be of interest. Our estimation confirms i) these are actors with a similar activity profile, interest and social behaviour as those identified manually, and ii) they talk about similar, hacking-related terms. Thus, we can conclude that these actors are either involved or close to involvement in cybercrime activities, and thus might benefit the most from intervention. Also, monitoring these actors could be of interest for security firms and intelligence agencies. A manual analysis of the forum activity of these actors confirms that they are all providing or asking for illegal assets and services such as malware, booters or stolen accounts.

[5] These thresholds were chosen after exploratory experimentation with the dataset and manually inspecting the results

Table 5. Summary of prediction using topic analysis

Subset	Predicted/Total (%)	Avg. distance	Farthest	Closest
LogReg	22/52 (42.31)	0.43	0.10	0.72
Clust	34/165 (20.61)	0.66	0.29	0.93
SNA	9/25 (36.00)	0.57	0.36	0.75
LogReg & Clust	8/26 (30.77)	0.63	0.36	0.89
SNA & Clust	0/7 (0.00)	0.66	0.50	0.79
LogReg & Clust & SNA	7/10 (70.00)	0.60	0.43	0.68

6 Ethical considerations

The research methodology was designed with ethical considerations at the forefront. The department's research ethics committee gave their approval for the research project. Furthermore, we complied with the Cambridge Cybercrime Centre's data sharing agreements. While the data are publicly available (and the forum users are aware of this), it could be used by malicious actors, for example to deanonymize users based on their posts. It was impossible for us to obtain informed consent from users as that would require us to identify them first. In accordance with the British Society of Criminology's Statement on Ethics, this approach is justified as the dataset is collected from the public Internet, and is used for research on collective behaviour, without aiming to identify particular members. Further precautions taken include not identifying individuals (including not publishing usernames), and presenting results objectively.

7 Limitations

We have presented a longitudinal study of behavioural aspects of key actors in underground forums. This research has attempted to overcome the significant difficulties in this challenging area of research. However, a number of limitations remain. First, results are based on the observation of a single forum. Thus, we do not analyse actors operating on other forums, nor do we measure actors' activities that occur off-forum. Future work will analyse cross-forum behaviour. Second, we focus on external sources to identify key actors, and thus our results could be biased by feedback from these sources. Moreover, the proportion of identified key actors in the forum is low, hindering the use of reliable classification techniques such as supervised machine learning. Third, our definition of the social network relies on public interactions. Unlike previous works [22,11], we do not use private messages to refine social relations. Recent work shows that public and private relations differ [26]. Finally, evaluating if the predicted actors are actually involved into criminal activities is not straightforward, even with manual analysis. Investigations into actors to produce evidence they are involved in cybercrime is a matter for law enforcement. Instead, our research helps to focus the spotlight, with the aim of informing crime prevention efforts.

8 Conclusion

Underground forums are one of the key pillars for the rise of underground economies. The sense of anonymity they provide, together with the ease of access to attack tools and services make these forums attractive places for young, non-skilled people to learn about hacking. Analysing the evolution of these low-level hackers makes it possible to consider early intervention approaches, with the aim of deterring their involvement away from criminal activities. Additionally, understanding who the key actors are, and what new tools they provide, is helpful for rapidly adapting to new forms of attack. For example, antivirus vendors could monitor those providing tools aimed at bypassing detection and new variants of malware.

We have conducted a large scale analysis of key actors from one of the largest English-speaking underground forums. We have evidence of online social connections between these key actors, and our research uncovers various common roles for these key actors. For example, some are well known in the community and actively participate in non-illicit sections. Others are less active and focus their activity in the market and monetization processes. Also, we note an evolution of interests towards more market and hacking related topics, as well as a decrease in threads requesting help or asking questions.

Finally, we have developed tools for detection and prediction of actors involved in cybercrime activities. These tools help to identify user accounts that might require further investigation by law enforcement and security firms monitoring underground communities and also for early deployment of new countermeasures or adaptation of existing ones. The tools used during this research are publicly available in our git repository.[6]

The purpose of our research is not to track and pursue criminal offenders, but understand who is at risk of becoming involved in crime, so as to apply intervention approaches at early stages. Identifying those that might be at risk of becoming involved in crime is critical for early intervention. Preventing young people from becoming involved in cybercrime will be of benefit for them later in life, as contact with the criminal justice system can be a stigmatising experience, affecting later job prospects and legitimate opportunities.

9 Acknowledgements

We thank the anonymous reviewers for their insightful comments. We also thank our colleagues from the Cambridge Cybercrime Centre for access to the CrimeBB dataset and their invaluable feedback, and Flashpoint, for assistance relating to actors of interest. This work was supported by The Alan Turing Institute's Defence and Security Programme [grant DS/SDS/1718/4]; and the UK Engineering and Physical Sciences Research Council (EPSRC) [grant EP/M020320/1].

[6] https://github.com/CCC-NLIP/DataSciForCybersecurity

References

1. Afroz, S., Garg, V., McCoy, D., Greenstadt, R.: Honor among thieves: A common's analysis of cybercrime economies. In: eCrime Researchers Summit. pp. 1–11. IEEE (2013)
2. Allodi, L.: Economic factors of vulnerability trade and exploitation. In: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. pp. 1483–1499. ACM (2017)
3. Anderson, R., Barton, C., Böhme, R., Clayton, R., Van Eeten, M.J., Levi, M., Moore, T., Savage, S.: Measuring the cost of cybercrime. In: The Economics of Information Security and Privacy, pp. 265–300. Springer (2013)
4. Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J.A., Invernizzi, L., Kallitsis, M., Kumar, D., Lever, C., Ma, Z., Mason, J., Menscher, D., Seaman, C., Sullivan, N., Thomas, K., Zhou, Y.: Understanding the Mirai Botnet. In: Proceedings of the 26th USENIX Security Symposium. pp. 1093–1110. Vancouver, BC (2017)
5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research **3**(Jan), 993–1022 (2003)
6. Caballero, J., Grier, C., Kreibich, C., Paxson, V.: Measuring pay-per-install: The commoditization of malware distribution. In: Proceedings of the 20th USENIX Security Symposium. pp. 13–13. Berkeley, CA, USA (2011)
7. Caines, A., Pastrana, S., Hutchings, A., Buttery, P.: Automatically identifying the function and intent of posts in underground forums. In submission
8. Chang, W., Wang, A., Mohaisen, A., Chen, S.: Characterizing botnets-as-a-service. ACM SIGCOMM Computer Communication Review **44**(4), 585–586 (2014)
9. Field, A.: Discovering Statistics Using SPSS. London: SAGE Publications, 2nd edn. (2005)
10. Franklin, J., Paxson, V., Perrig, A., Savage, S.: An inquiry into the nature and causes of the wealth of Internet miscreants. In: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (2007)
11. Garg, V., Afroz, S., Overdorf, R., Greenstadt, R.: Computer-supported cooperative crime. In: International Conference on Financial Cryptography and Data Security. pp. 32–43. Springer (2015)
12. Holt, T.J.: Subcultural evolution? examining the influence of on- and off-line experiences on deviant subcultures. Deviant Behavior **28**(2), 171–198 (2007)
13. Hutchings, A.: Cybercrime trajectories: An integrated theory of initiation, maintenance, and desistance. In: Crime Online: Correlates, Causes, and Context, pp. 117–140. Carolina Academic Press (2016)
14. Hutchings, A., Clayton, R.: Exploring the provision of online booter services. Deviant Behavior **37**(10), 1163–1178 (2016)
15. Hutchings, A., Holt, T.J.: A crime script analysis of the online stolen data market. British Journal of Criminology **55**(3), 596–614 (2015)
16. Karami, M., McCoy, D.: Rent to Pwn: Analyzing commodity booter DDoS services. Usenix login **38**, 20–23 (2013)
17. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory **28**(2), 129–137 (1982)
18. Lusthaus, J., Varese, F.: Offline and local: The hidden face of cybercrime. Policing **advanced access** (2017)
19. Macdonald, M., Frank, R., Mei, J., Monk, B.: Identifying digital threats in a hacker web forum. In: International Conference on Advances in Social Networks Analysis and Mining. pp. 926–933. IEEE/ACM (2015)

20. Marcus, M.P., Marcinkiewicz, M.A., Santorini, B.: Building a large annotated corpus of English: The Penn Treebank. Computational Linguistics **19**(2), 313–330 (1993)
21. McMillen, D., Alvarez, M.: Mirai IoT botnet: Mining for bitcoins? Security Intelligence **10** (2017)
22. Motoyama, M., McCoy, D., Levchenko, K., Savage, S., Voelker, G.M.: An analysis of underground forums. In: Proceedings of the ACM SIGCOMM conference on Internet Measurement Conference. pp. 71–80 (2011)
23. National Crime Agency: Pathways into cyber crime (2017), https://perma.cc/897P-GZ3R
24. Noroozian, A., Korczyński, M., Gañan, C.H., Makita, D., Yoshioka, K., van Eeten, M.: Who gets the boot? Analyzing victimization by DDoS-as-a-service. In: International Symposium on Research in Attacks, Intrusions, and Defenses. pp. 368–389 (2016)
25. Nunes, E., Diab, A., Gunn, A., Marin, E., Mishra, V., Paliath, V., Robertson, J., Shakarian, J., Thart, A., Shakarian, P.: Darknet and deepnet mining for proactive cybersecurity threat intelligence. In: Conference on Intelligence and Security Informatics (ISI). pp. 7–12. IEEE (2016)
26. Overdorf, R., Troncoso, C., Greenstadt, R., McCoy, D.: Under the underground: Predicting private interactions in underground forums. arXiv preprint arXiv:1805.04494 (2018)
27. Pastrana, S., Thomas, D.R., Hutchings, A., Clayton, R.: CrimeBB: Enabling cybercrime research on underground forums at scale. In: Proceedings of The Web Conference (WWW). ACM (2018)
28. Portnoff, R.S., Afroz, S., Durrett, G., Kummerfeld, J.K., Berg-Kirkpatrick, T., McCoy, D., Levchenko, K., Paxson, V.: Tools for automated analysis of cybercriminal markets. In: Proceedings of 26th International World Wide Web conference (2017)
29. Samtani, S., Chinn, R., Chen, H.: Exploring hacker assets in underground forums. In: International Conference on Intelligence and Security Informatics (ISI). pp. 31–36. IEEE (2015)
30. Sood, A.K., Enbody, R.J.: Crimeware-as-a-service: A survey of commoditized crimeware in the underground market. International Journal of Critical Infrastructure Protection **6**(1), 28–38 (2013)
31. Soska, K., Christin, N.: Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In: Proceedings of the 24th USENIX Security Symposium (2015)
32. Spärck-Jones, K.: A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation **28**, 11–21 (1972)
33. Sutherland, E.H.: White Collar Crime: The Uncut Version. New Haven: Yale University Press (1949)
34. Thomas, D.R., Clayton, R., Beresford, A.R.: 1000 days of UDP amplification DDoS attacks. In: APWG Symposium on Electronic Crime Research (eCrime). IEEE (2017). https://doi.org/10.1109/ECRIME.2017.7945057
35. Thorndike, R.L.: Who belongs in the family? Psychometrika **18**(4), 267–276 (1953)
36. Valeros, V.: A study of RATs: Third timeline iteration (2018), https://perma.cc/REB5-JFNR
37. Vold, G.B., Bernard, T.J., Snipes, J.B.: Theoretical Criminology (5th ed.). New York: Oxford University Press, Inc (2002)
38. Zhang, X., Tsang, A., Yue, W.T., Chau, M.: The classification of hackers by knowledge exchange behaviors. Information Systems Frontiers pp. 1–13 (2015)