From playing games to committing crimes: A multi-technique approach to predicting key actors on an online gaming forum

Jack Hughes, Ben Collier, and Alice Hutchings
Department of Computer Science and Technology
University of Cambridge
Cambridge, UK
firstname.surname@cl.cam.ac.uk

***Abstract*—We propose a systematic framework for analysing forum datasets, which contain minimal structure, and are non-trivial to analyse at scale, aiming to support future analysis of underground forum communities. We use a multi-technique approach which draws on a combination of features, including post classifications extracted using natural language processing tools, and apply clustering and predictive techniques to this dataset, to predict potential key actors—individuals who have a central role in overtly criminal activities, or activities which could lead to later offending, and hence might benefit most from interventions. We predict 49 key actors on an underground gaming-specific cheating and hacking forum, validated by observing only overlaps of techniques, combined with topic analysis, to build a classifier for key actor status. In addition, we also use these techniques to provide further insight of key actor activity. We found one cluster and two posting trajectories to contain a high proportion of key actors, logistic regression found an actor's h-index to have higher odds for prediction than other features, and partial dependence plots found reputation to have a significant change in prediction between values of 100 to 1000.**

***Index Terms*—Cybercrime, Underground Forums, Online Gaming, Pathways, Key Actors**

I. Introduction

There is increasing interest in the possible pathway from playing online games to committing cybercrime. The UK's National Crime Agency (NCA) [1] undertook a small-scale study which highlighted a potential link between online gaming and entry points into cybercrime. This identified a potential pathway beginning with the use of gaming cheats, which—depending on the tools used—can be legal. These cheats are available on online underground forums, which can lead to further interaction and potentially escalation into more serious cybercrime activity.

While there has been much research into online underground forums (e.g. [2]–[9]), their relationship with online gaming has been largely unexplored. As the identification of these potential pathways suggest that online underground forums may play an instrumental role in the shift from legal to illegal activities, there has been interest from UK law enforcement in exploring intervention activities which might deter gamers with an interest in these forums from becoming involved in illegal activities. For example, the NCA has run 'advertorials' on three UK gaming websites as part of their 'Cyber Choices' campaign [10]–[12]. These were used to inform people about the illegality of 'booter' services—a type of low-level cybercrime popular on these types of forums, which involves paying for a service to disrupt the network connection of opponents on online games [13]. The barrier to entry for carrying out these attacks is low, and previous research has found that this provides some individuals involved in online gaming with a pathway into more serious illegal activities [14].

Datasets of posts made on underground cybercrime forums collected by web scrapers are a valuable resource for researchers interested in exploring a range of research questions linked to cybercrime and security. However, these tend to constitute hundreds of thousands of messages over many years, and are therefore difficult to analyse by traditional means. In this paper, we propose a multi-technique framework for processing and carrying out analysis on forum data focused around the identification of key actors and their signature characteristics in these datasets, who are likely to benefit the most from diversion. Key actors are defined as those who are involved in activities that are likely to make them of interest to law enforcement, such as overtly criminal activities (e.g. sharing tools used for hacking), or activities which are not necessarily illegal but may lead to later offending, such as sharing tools and tutorials for cheating and cracking.

We explore this in detail through analysis of a subset of the gaming cheating and hacking community, namely active users on the MultiPlayer Game Hacking forum (MPGH), using a subset of the CrimeBB dataset [15]. This publicly accessible forum was chosen over others that are more general in nature due to its gaming-specific focus, avoiding the need to identify gaming-related discussions. Our research is intended for informing better understanding of the dynamics of these communities and the pathways taken by individuals within them, rather than for the purposes of police investigation.

This work aims to classify key actors within the forum community. Specifically, we:

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/M020320/1].

- Create a systematic data processing framework, for analysis of large, unstructured forum datasets.
- Apply natural language processing (NLP) tools to automatically classify post type.
- Cluster actors using k-means clustering, social network analysis, and group-based trajectory modelling (which models the trajectories of forum members over time). Using these methods, we identify actors in clusters with high concentrations of, or have multiple connections to, key actors. We predict these are also key actors.
- Use predictive models, namely logistic regression, random forest, and neural networks, to predict key actor status.
- Use topic analysis to cross-validate predictions.

Combining features allows for multiple indicators including interaction graphs and post contents to be modelled, to better predict key actor status on the forum.

II. Background and Related Work

Pastrana et al. [2] analyse HackForums, a popular and long-running underground forum, to predict key actors, who are defined as forum users of interest to law enforcement. To predict key actor status, they use logistic regression, social network analysis, and clustering, and cross-validate predictions using topic analysis. We adapt and expand tools created for this work on a *general* underground hacking forum, to a *gaming-specific* underground hacking forum.

Pastrana et al. [2] use latent Dirichelt allocation (LDA) for creating the per-topic word lists used for validating key actor predictions. However, Deliu et al. [16] find that LDA cannot detect 'zero days', vulnerabilities in software which the vendor is unaware of or has not been able to fix, due to sparse word usage. Therefore, LDA may not be appropriate for automatically detecting new or uncommon cybercrime-related terms used by key actors. Macdonald et al. [17] use NLP tools to determine the sentiment of key words used in posts, and find that these are limited as sentiment analysis is not perfect for modelling dialogue. They also find automated methods cannot automatically identify new terms.

Caines et al. [18] developed NLP annotation tools for HackForums data. These tools classify post according to: *post type* (e.g. if the post is a request for information, a product or service for sale, or an answer to someone's technical query); *author intent* (e.g. helpful, grateful, sarcastic, disapproving, or abusive); and *addressee* (i.e. an individual, a number of individuals, or other members of the forum as a whole).

Past research has used members' reputation scores for validating prediction results. Reputation scores are calculated using a voting system, where a member who likes or dislikes a member's post can either upvote or downvote the post respectively. High reputation scores may signal trustworthiness, and therefore such metric is often used in validation. For example, Marin et al. [7] identify key actors on a darkweb forum, and use reputation scores to validate prediction results. Also, Benjamin and Chen [19] highlight the significance of reputation scores, and create a model to predict reputation scores using a set of features including the number of replies created and the average post length, in order to predict members who have a high reputation, and could potentially be a key actor. They find reputation is influenced by the content of posts, and NLP annotation tools could help to better model post content, instead of relying on the indirect reputation scoring system. In addition, Zhang and Li [20] find reputation highly correlates with post-centred characteristics, such as the number of comments (replies) as user has made, or the average post length.

Biswas et al. [21] do not use reputation at all in their work, and instead finds member metrics (e.g. number of messages posted, posts per thread, average word count) and NLP-derived features (e.g. keyword similarity, sentiment) to be significant predictors of key actor status.

Marin et al. [7] takes a different approach and combines predictors into a hybrid prediction model, for validating results. This combines predictions with reputation scores to cross-validate key actor predictions. Pastrana et al. [2] use a similar approach, but use topic analysis for cross-validation.

Samtani and Chen [22] look at the interaction graph of the network of members, finding that the network of key actors is not connected, yet a high degree of connectedness exists among key actors. This shows that while not all key actors interact and influence each other, their strong connection to other members could be influential. Sarvari et al. [23] find similar results, with key actors having high centrality and PageRank indices. However, Johnsen and Franke [24] find centrality measures to be biased towards those who communicate more, rather than those who may be more influential. It could be the case of a key actor predominantly communicating outside of the public area of the forum, and centrality measures may not reflect this.

Overdorf et al. [4] begins to try to detect such out-of-band private messages based on interactions on the public forum, but finds no straightforward relation between these. However, when identifying important features they used, they found NLP-derived features (e.g. bag of words used) to be more important for predictions than metadata (e.g. time spent on forum, reputation).

It is also interesting to begin to categorise key actors, for identifying subgroups who may benefit from different types of targeted intervention activities, and then to be able to work out development of each subgroup for identifying intervention points. Seebruck [25] categorises by aims, such as for profit, revenge, or recreation. Whereas Zhang et al. [26] categorises hackers by knowledge levels, such as guru, casual, and learning. Frank et al. [27] uses the guru category of key actors on one forum, to trace these across, finding a small number move information across. Park et al. [6] use different approach by looking to categorise post types by key actors using social graphs. They identify roles of consumers and talkers. They also find weighting by the amount users interact on the forum is important.

III. Methods

A. Dataset

We use a subset of the CrimeBB dataset [15], which contains data scraped from 25 different underground forums, and is available from the Cambridge Cybercrime Centre.[1] Specifically, we use data from the MultiPlayer Game Hacking (MPGH) forum, an underground discussion platform for members to discuss hacking and cheating techniques for online multiplayer games. As well as general discussions of gaming and technology, the forum provides a marketplace for trading goods and services, such as for cracked accounts, and alternative currencies. Cracked accounts include those where the credentials have been stolen, and are used for gaining free access to paid services. Alternative currencies advertised for transfer includes Bitcoin and gift vouchers such as Amazon Gift Cards.

We selected MPGH for the gaming-specific nature of the forum, to reduce complexity in analysis; other larger forums such as HackForums would require either a manual or automatic selection of gaming-related activity, which is non-trivial. As MPGH is gaming-specific, analysis can be carried out with the entire dataset.

The forum contains over 730 top-level subforums. These are high-level collections, such as "General Hacking" and "Marketplace". Contained within these are 764k threads, which each contain a linear ordered set of posts focused around a certain topic, such as the sale of an item on the marketplace. There are over 9.4M posts on the forum made by over 478k members. Of these, 18% have posted over 5 times. We refer to these as "active" members, and only these members (n=84k) are used for our subsequent analysis.

B. Ethical Considerations

Ethics approval was granted from the department's ethics committee. Furthermore, we complied with the Cambridge Cybercrime Centre's data sharing agreements. The research uses data collected from a publicly available forum. It would not be possible to gain informed consent from all members as this would be considered as spamming. As this work only analyses collective behaviour, rather than identifying individuals, under the British Society of Criminologys Statement of Ethics [28], this work falls outside the requirement of informed consent. Further precautions taken include not identifying individuals (including not publishing usernames), and presenting results objectively.

C. Key Actor Selection

Due to a lack of information about law enforcement activity on the forum, key actor selection was a manual process. We selected 84 key actors from popular areas of the forum that exhibited set criteria, such as sustained involvement in harmful or illegal behaviour, that were likely to indicate that they might benefit from diversionary approaches to prevent them from becoming involved in more serious illegal activities. Key actors included:

- 20 members who had released tools and tutorials on cracking and hacking subforums. This included key generators used for stealing software, and tutorials for credential stuffing tools used to break into accounts. Credential stuffing is prominent on the forum, including in the marketplace which sells cracked accounts, which have likely been obtained through these techniques.
- 17 members who had released tools and tutorials on gaming-specific forums, for bypassing anti-cheat mechanisms.
- 4 members who had advertised booting services.
- 21 members who were found through compilation posts on hacking tools, including tutorials and tools for UDP flooding (used in denial of service attacks), key logging, and various techniques for hacking websites.
- 25 members who had a high proportion of interactions with other key actors and were also involved in similar cybercrime-related activities. These members were manually checked to validate these before being added to the initial set of key actors.

The manual selection process began by identifying subforums likely to contain activity of interest to law enforcement. We selected threads within these with high post counts and positive replies, to avoid poor quality threads. The selection method aimed to identify a broad range of key actors, although we note they may not be representative of the whole population of the forum.

D. Data Processing Pipeline

The data processing pipeline is shown in Figure 1. The pipeline includes pre-processing and feature extraction, prediction techniques, and validation.

1) Feature Extraction: **Member features** are extracted by adapting open-source research tools developed by Pastrana et al. [2]. Features include per-member features of activity on the forum. The tools were developed for a different subset of the CrimeBB dataset, allowing the tools to be adapted to this dataset. While the tools were developed for a *general* underground hacking forum, we adapt these for analysis of the *gaming-specific* forum MPGH. In particular, MPGH does not provide reputation voting data, so such features are omitted.

NLP features are extracted using tools developed by Caines et al. [18]. These tools provide per-post annotations from the dataset. For each user, we count annotations to create per-user features, except for sentiment and token count, for which we take the mean value. The NLP tools were developed for analysing a different subset of CrimeBB, and hence we adapt these for use in MPGH. We note the tools use models trained on HackForums data, where actors may use slightly different terminology than on MPGH.

There are highly correlated features contained in the initial featureset. As later analytic approaches assume a lack of multicollinearity, we iteratively remove features with correlation of over 80%. Of the 47 features initially collected, 32 are

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

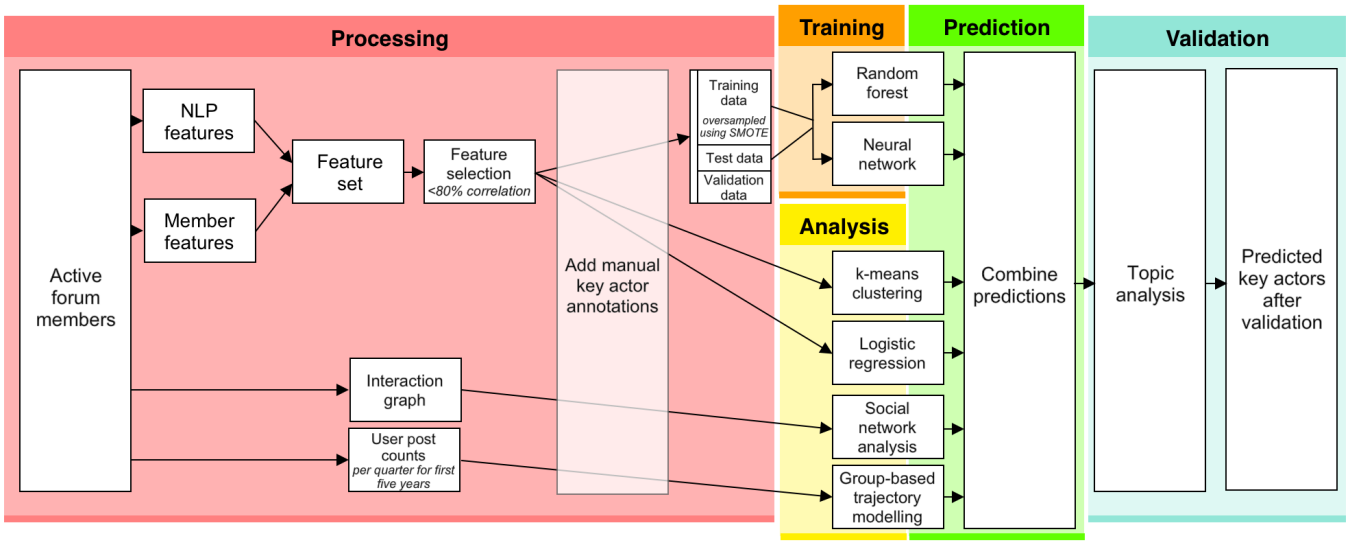

Fig. 1. Data processing pipeline

retained for analysis. The set of 32 per-member features, containing members with over five posts, with an additional feature indicating whether they are in the set of manually identified key actors ("manual key actor annotations"), shall be referred to as the dataset. The dataset contains a variety of features related to metadata and post content, for predicting and modelling actor activity, including:

- Social network analysis centrality measures. These are used to quantify a members interactions with other forum members, to model author-replier relationships between members. This includes eigenvector centrality, a measure of the influence of a node (forum member) in a network (directed graph of repliers to post authors).
- The number of posts and threads made within categories of subforum, namely hack, tech, coding, games, market, and web. Also, the number of threads within common and money categories are included, the number of posts in the graphics category, and the combined sum of threads and posts in currency exchange areas of the forum.
- Features of members, namely the number of days they have spent on the forum in total, the number of days between registering on the forum and their first post, and reputation score of the user.
- Interactions on the forum, including the total number of citations a user has made (quoted text in post replies).
- Impact metrics based on measures used in academia for quantifying research quality using citations, namely h-index and i-100 index. The h-index is defined by a member having index h if they have created h threads which have at least h replies. i-100 is defined as the number of threads a member has created with at least 100 replies.
- NLP features include post sentiment, post activity type (social post type only), post intent type, a feature indicating if a post includes a code snippet, and the average number of tokens (individual words) a user has for all of their posts. Post sentiment is a score of emotion used in each post. Post intent types include gratitude, moderate, negative, positive, private message, and vouch (posts showing support of or confidence for another member).

2) Prediction Techniques: We use a number of techniques for prediction purposes. Logistic regression, neural networks, and random forests are predictive models. We also use group-based trajectory modelling and k-means clustering to identify actors contained in clusters with a high proportion of key actors. Social network analysis uses a graph of author-replier relationships, to identify members which are connected to three or more key actors ("bridge nodes", defined by Pastrana et al. [2]). Bridge nodes and those in key actor clusters are also predicted to be key actors. The combination of prediction techniques allows predictions to be made on different types of member activity, including interactions with other members (social network analysis), and posting activity over time (group-based trajectory modelling), to build a stronger overall classifier.

Social network analysis uses tools adapted from Pastrana et al. [2] used for analysis of HackForums. This creates a directed graph using nodes for members in the forum, and edges representing replies: an edge from A to B represents a reply from A to B, either by replying to a thread, or by citing a post. This directed graph represents interactions on the forum. For the visualisation, we only plot the key actors and their closest neighbours, to make the graph easy to interpret. The closest neighbours are defined as the highest weighted five successors and five predecessors of a member, using the count of interactions between members as the weighting of edges.

We use social network analysis to model interactions between members of the forum–the interactions between key

actors and non-key actors has the potential to influence non-key actors to become interested in activities of key actors, and progress further into more serious cybercrime. The aim of this technique is to identify "bridge nodes"—nodes that connect together three or more key actors.

k-means clustering was also used by Pastrana et al. [2] for analysis of HackForums. This technique extracts clusters from the set of features, to identify groups of members with similar characteristics. The algorithm has one parameter, k, which is the number of clusters to use. To determine the optimal value of k, the Elbow method is used, which calculates a score to represent the quality of clusters for each parameter value.

Group-based trajectory modelling is a statistical technique developed by Nagin [29] that takes time-series data and groups these into different trends over time ("trajectories"). Latent trajectory groups are allowed to emerge from the data itself, rather than pre-specified, resulting in a series of trajectories, the group membership probabilities of each member, and a group assignment based on their highest probability. This technique was developed for use in criminology, and is now being applied in other fields such as clinical research [30].

We use the post counts from the dataset over time, to identify trajectories of members within four subforums, namely gaming, general, hacking, and market. The time series uses quarter-by-quarter counts of posts for the first five years' of posting for each actor. Trajectories are found for low, medium, and high activity levels in each subforum. For a member to have a low activity level for a given category, they must have posted between 20 and 99 times within the subforum. Users with a medium activity level have between 100 and 499 posts, and users with a high activity level for a category have posted over 500 times within the given subforum. Those with fewer than 20 posts are not included in the analysis.

This technique is useful as the dataset is intrinsically dynamic and the time-series approach of this method can better model the development of users over time. We adapted our STATA scripts,[2] which use a plugin based upon work by Jones and Nagin [31], for use on MPGH.

Logistic regression is used to predict potential key actors. This technique is useful for two purposes: first, we are able to obtain a model used for prediction of potential key actors, and secondly, we are able to inspect this type of model. For example, odds values can show the impact each feature has on the prediction outcome.

We use backwards stepwise logistic regression with the likelihood ratio method. This method begins with all features and iteratively removes those that do not have a significant influence on the model. Field [32] justifies the use of stepwise methods when causality is not of interest, but rather a model to fit the data, and recommends the backward method over the forward method, which has a higher risk of Type II (false negative) errors. Cases that have an undue influence on the model (where the Cooks distance is greater than 1.0) are removed [32].

Neural networks are used to train from examples to find patterns in the dataset, in order to predict potential key actors. We build a multi-layer perceptron classifier for our neural network using the Keras [33] library, and wrap this inside a scikit-learn [34] model for easier parameter tuning and model inspection. Neural networks use hidden layers to train a non-linear model, providing an advantage over logistic regression.

For this technique, the dataset is split into three sets: 64% of members are used in the training set, 20% in the test set, and 16% in the validation set. The training set is used to train each model, and the test set is used to find suitable parameters. The validation set is used to compare model performance, and is not used during training or testing stages. Neural networks require feature scaling to be performed on the dataset, and therefore features are scaled down to the range between 0 and 1.

Comparing the low number of key actors with the size of the training dataset, there is a large class imbalance. This could result in models overfitting to the small set of positive samples, or models predicting all cases as negative. To solve this issue, we oversample the training dataset using SMOTE [35], to have an equal number of positive and negative samples. This technique creates synthetic cases based on the k nearest neighbours of each selected case.

Neural networks typically have a number of hyperparameters available for tuning the model, and therefore finding the most suitable model can be computationally expensive. We perform grid search with cross validation to select the best hyperparameters and model, using F1 scoring with the test dataset to select the best performing model. F1 score is the harmonic mean of precision and recall, and is used due to the imbalanced dataset.

It is not straightforward to inspect the reasoning of neural network models, with tools limited due to the black-box nature of the technique. Therefore, we use partial dependence plots from the pdpbox library[3] to inspect the neural network. These plots are used to show the change in prediction (relative changes of prediction probabilities) as the value for a given feature increases.

Random forests consist of a number of decision trees. Decision trees are used for supervised classification, and recursively split a dataset into a tree structure. At each node in the tree, a condition (e.g. `h-index < 5`) determines which branch to take for predictions, and the weighting of positive-negative examples at a leaf node determines the prediction result. For a random forest, a weighting is used across all trees to aggregate predictions from each tree, to create an overall prediction value. To carry out prediction for a given tree, a walk is used, starting at the root node and taking branches depending the feature values, until ending up at a leaf node containing the prediction result (majority label).

[2]Available at https://github.com/JohnnyHistone/Group_trajectory_model_HF

[3]https://github.com/SauceCat/PDPbox

However, decision trees are prone to overfitting if they contain leaf nodes with a small number of examples, or underfit if leaf nodes contain a large number of examples with a half-half split of prediction classes. Typically pruning is used to remove leaf nodes which may cause issues, but this is not supported by the scikit-learn library [34]. To overcome this issue, random forests are used, which apply a random weighting across n decision trees. Note that scikit-learn applies a random weighting across the trees, compared to other implementations such as Breiman [36] where trees each vote for a particular class.

Permutation importance is applied to the model, using the eli5 library,[4] to shuffle one feature at a time in order to determine the weighting of each feature on the final result. We also use the technique SHAP [37] to inspect the model. SHAP values are useful for showing the importance of each feature for a single prediction outcome. When plotting using the library, the figure shows increasing (pink) and decreasing (blue) bars for a given prediction probability, with wider bars indicating higher importance.

We again use the training-test split dataset, where the training dataset has been oversampled using SMOTE, to predict potential key actors. The model consists of a weighted set of decision trees used to predict key actor status, using trees containing simple conditions of features at each node.

Predictions for each technique are made using the following criteria:

- Social network analysis: members connected to at least 3 key actors ("bridge nodes").
- k-means clustering: members of the cluster with the highest ratio of key actors.
- Group-based trajectory modelling: members of trajectory archetypes with a high proportion of key actors.
- Logistic regression: members with a predicted probability of 10% or more.
- Random forest and neural networks: members which are predicted by the model to be key actors.

We combine these predictions, and use topic analysis to validate results.

3) Validation Techniques: Predictions are used to identify non-key actor members who have similar characteristics of key actors. Predictions are then validated using topic analysis to confirm if predicted key actors use similar cybercrime-related terminology to those of key actors. We use topic analysis for validation and not for prediction as this would be computationally expensive to process for all forum content.

For topic analysis, we adapt code by Pastrana et al. [2], which uses latent Dirichlet allocation (LDA) to identify words used by key actors for each category of subforums. From this set, words relating to cybercrime activity are manually selected.

To validate predictions, the same method is run for all posts made by each predicted member, and to be confirmed, the member must use at least 20% of the cybercrime-related terms which key actors use. This is used to check if predicted key actors use similar cybercrime-related terms to those of key actors.

[4]http://github.com/TeamHG-Memex/eli5

IV. Results

A. Social network analysis

Figure 2 shows the initial set of key actors are well connected, except for three key actors. Note that while the manual process of key actor selection identifies key actors connected to other key actors (the yellow nodes), key actors not selected using this process are connected by at most one other node. This analysis identified 13 bridge nodes (members connected to at least 3 key actors) between key actors, which are used for predicting potential key actors. Yellow nodes were selected first based on interactions with other key actors, and then checked against the key actor criteria. We do not show the closest neighbours of yellow nodes, as these have been selected based off connections and would bias the graph in appearing to have further in-connections.

B. k-means clustering

k-means found a cluster with a greater proportion of key actors (8.3%, which includes 58.3% of all key actors) compared to other clusters. Table I shows mean values of different variable types for each cluster, including the interest categories of each group, activity in currency exchange (#CurExc), and social relation features (H = h-index, i100 = i-100 index, EV = eigenvector centrality). This cluster has spent the second-longest average time on the forum, have the greatest eigenvector value, and the highest average activity in currency exchange. They are mostly interested in the gaming, common, and marketplace subforums. The NLP features shows this cluster has the lowest average sentiment score, while post-related features are greater than other clusters. When clustering only the set of key actors, all groups are primarily involved with both common and gaming areas, with two groups also interested in the marketplace.

C. Group-based trajectory modelling

There were fairly similar trajectories found in each of the four subforums (gaming, general, hacking, and market). Overall, users tend to exhibit one of five key trajectory archetypes:

- Fickle, where members initially have high activity levels that quickly decline.
- Decliner, where members' activity levels decline steadily to zero over time.
- Sustainer, where members' activity levels are steady.
- Engager, where members' activity levels steadily increase, then decrease to zero after reaching a peak.
- Super-engager, where members' activity levels increase and then reduce to a lower, but sustained, level.

Group-based trajectory modelling shows that most key actors are not very active in any of the trajectories within the hacking area of the forum (see Table II). This is likely due to the category not being the main focus of the forum, and

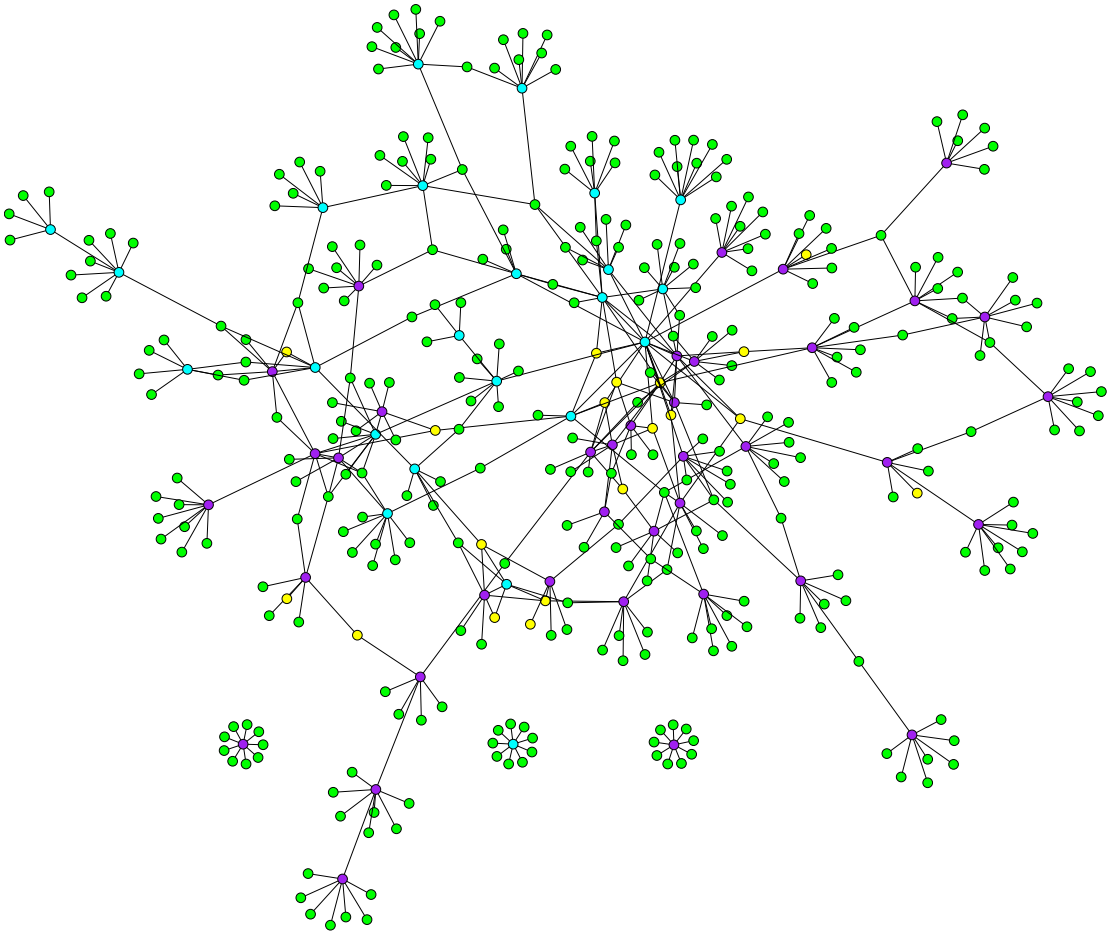

Fig. 2. Social network analysis graph of key actors and their closest neighbours (purple = general key actor, blue = key actor distributing tools and information, yellow = key actor identified through interactions with other key actors, green = non key actor)

#KeyActors	Activity		Interests				Social Relations			NLP Features								
/Count	Days	Threads/Posts	cat1	cat2	cat3	#CurExc	H	i100	EV	Sentiment	SOC	NEU	POS	NEG	MOD	PMS	GRA	VOU
49/589	1872.8	187.5/3418.8	G	C	M	4.5	21.8	2.3	0.03	0.051	754.6	5039.3	507.4	26.6	65.9	34.1	255.5	44.9
11/21275	826.6	8.8/68.6	G	M	C	0.4	2.8	0.0	0.00	0.067	14.9	110.4	10.6	0.5	1.1	2.9	9.7	2.7
20/55180	142.5	4.6/26.2	G	M	C	0.1	2.0	0.0	0.00	0.056	5.8	43.0	4.2	0.2	0.5	1.5	3.6	0.8
4/6898	1920.4	13.0/115.1	G	M	C	0.3	3.5	0.1	0.00	0.064	24.6	178.0	17.3	0.9	1.9	4.3	15.2	4.1

TABLE I

CHARACTERISTICS OF ACTIVE FORUM MEMBER GROUPS FROM K-MEANS CLUSTERING (K = 4). G=GAMING, C=COMMON, M=MARKETPLACE. SOC=SOCIAL POST TYPE, NEU=NEUTRAL INTENT, POS=POSITIVE INTENT, NEG=NEGATIVE INTENT, MOD=MODERATE INTENT, PMS=PRIVATE MESSAGE INTENT, GRA=GRATITUDE INTENT, VOU=VOUCH INTENT

there are other underground forums which are more general for these types of discussion.

Two trajectory archetypes were identified that contained a high proportion of key actors. Of the 89 actors in the high-frequency super-engager post activity in the gaming category (the red line in Figure 3), 19 are key actors. Furthermore, of the 55 actors in the high-frequency super-engager post activity in the general category (the red line in Figure 4), there are 17 key actors.

D. Logistic regression

Without any independent features in the model, 100% of cases are predicted to not be key actors. Compared to the baseline, the final model is significantly improved and is statistically better at predicting key actors ($\chi^2(15, n=83{,}942)=579.1$, $p<.001$). The final model accounts for between 0.7 and 43.8% of the variance, accurately predicting 31.0% of key actors with a low false error rate (0.01%).

Table III presents the results of the final step of the logistic regression analysis. The odds ratios indicate that for each increase in an actor's h-index, the odds of them being a key ator increases by 1.231. The frequency with which actors posted on various sections also predicts being a key actor, with each additional post in coding, gaming, market, tech and web subforums increasing the odds by 1.002, 1.001, 1.002, 1.005, and 1.113 respectively. Each additional thread generated in the

Category	Activity Level	Fickle	Decliner	Sustainer	Engager	Super-engager
Gaming	Low	4 (10734)	4 (12081)	1 (3155)	1 (5219)	-
	Mid	-	3 (2862)	8 (2791)	-	-
	High	2 (416)	-	23 (409)	7 (527)	19 (89)
General	Low	4 (1904)	-	2 (471)	4 (1217)	-
	Mid	0 (356)	3 (461)	-	0 (101)	8 (283)
	High	3 (108)	6 (197)	17 (181)	-	17 (55)
Hacking	Low	3 (314)	9 (364)	7 (100)	19 (261)	-
	Mid	-	0 (36)	4 (27)	3 (53)	-
	High	-	-	-	-	-
Market	Low	2 (6160)	14 (4380)	24 (1826)	-	-
	Mid	1 (925)	-	14 (519)	5 (956)	-
	High	-	4 (59)	5 (114)	2 (120)	-

TABLE II
NUMBER OF KEY ACTORS (ALL USERS IN PARENTHESES) FOR EACH TRAJECTORY ARCHETYPE PER CATEGORY AND ACTIVITY LEVEL

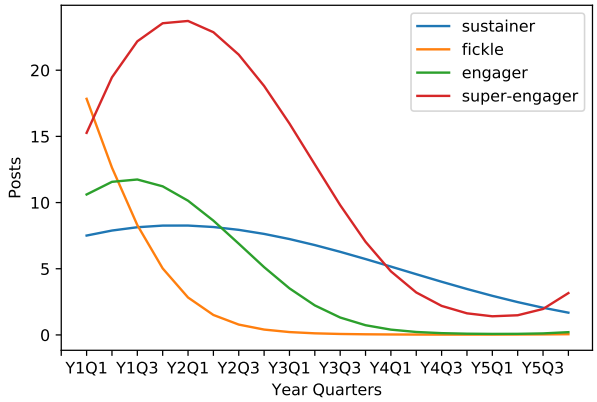

Fig. 3. Group-based trajectories for high-frequency activity in the gaming category

common and hacking subforums increased the odds of being a key actor by 1.009 and 1.077 respectively.

Generating threads in the gaming section and posting in currency exchange decreases the odds that users are key actors by .991 and .954 respectively. Other features that decrease the odds of being a key actor include showing gratitude (.997),

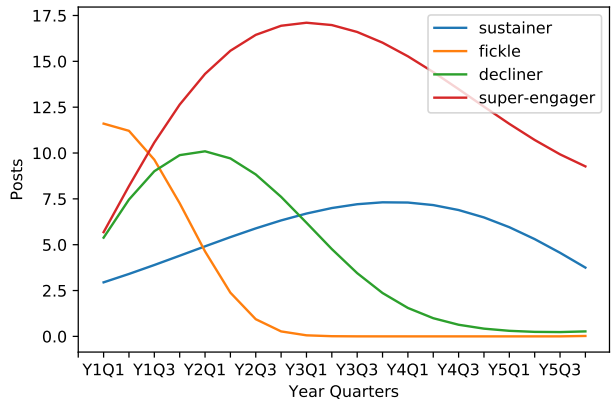

Fig. 4. Group-based trajectories for high-frequency activity in the general category

moderating posts (.980), social posts (.998), and having a high i-100 index (.867).

TABLE III
LOGISTIC REGRESSION MODEL PREDICTING KEY ACTORS

		B	S.E.	Wald	Sig.	Exp(B)	95% C.I. for Exp(B) Lower	Upper
Step 17	intent_gratitude	-.003	.001	3.858	.050	.997	.994	1.000
	intent_moderate	-.021	.006	11.626	.001	.980	.968	.991
	postType_social	-.002	.001	9.451	.002	.998	.997	.999
	CurrencyExchange	-.047	.020	5.725	.017	.954	.918	.992
	Reputation	.000	.000	9.517	.002	1.000	1.000	1.000
	h-index	.208	.024	73.487	.000	1.231	1.174	1.291
	i-100	-.143	.044	10.762	.001	.867	.796	.944
	post_coding	.002	.000	29.671	.000	1.002	1.001	1.003
	post_gaming	.001	.000	33.964	.000	1.001	1.001	1.001
	post_market	.002	.000	17.680	.000	1.002	1.001	1.003
	post_tech	.005	.002	3.907	.048	1.005	1.000	1.010
	post_web	.107	.054	3.847	.050	1.113	1.000	1.238
	thread_common	.009	.002	16.421	.000	1.009	1.005	1.013
	thread_gaming	-.009	.005	4.042	.044	.991	.982	1.000
	thread_hack	.074	.016	20.541	.000	1.077	1.043	1.112
	Constant	-8.435	.205	1685.465	.000	.000		

E. Random forest and neural network models

The random forest and neural network models predict 128 and 375 actors to be key actors, respectively. Random forest and neural network models are black-boxes, and therefore hard to inspect. We use techniques to assist our understanding of the models, including using SHAP diagrams [37] and partial dependence plots. SHAP diagrams help to identify feature importances for a given prediction result using a complex model. This technique is similar to feature importances in linear regression, but also supports non-linear models. Partial dependence plots are used to visualise how changing one feature value changes the prediction output, and can show whether the relationship of features is linear or non-linear.

Also, we use permutation importance from the eli5 library[5] with the random forest model. This technique measures the increase in prediction errors when a given feature is shuffled, to determine the influence a feature has over the prediction outcome. Permutation importance is applied to the random forest model (Table IV). These results show posts either containing code or in the coding category, and threads in the common category, have a positive weighting on the prediction in the model. Also, this found h-index to have a low weighting.

Partial dependency plots are used to inspect the neural network model (Figure 5 and 6). Plots show percentiles for

[5]http://github.com/TeamHG-Memex/eli5

	Feature	Weight	Standard Deviation
1	Posts in the coding category	0.175807	0.026433
2	Threads in the common category	0.171898	0.090516
3	Posts containing code	0.091195	0.018782
4	Posts in the tech category	0.055403	0.013412
5	Posts in the market category	0.044822	0.038840
⋮	⋮	⋮	⋮
28	Total cites	-0.031643	0.018201
29	Posts with gratitude intent	-0.034834	0.022421
30	H-index	-0.049396	0.008518
31	Threads in the coding category	-0.052596	0.002372
32	Threads in the hack category	-0.099399	0.011469

TABLE IV
PERMUTATION IMPORTANCE FOR RANDOM FOREST

each feature, to show the variation in prediction status as the feature value increases. The partial dependence plot for reputation shows a steady increase in change of prediction value, with a larger increase between scores of 100 to 1000, whereas h-index does not show any overall change, although the standard deviation begins to rapidly increase when the h-index is greater than 5.

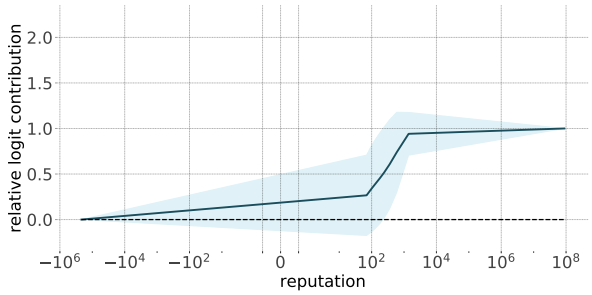

Fig. 5. Partial dependency plot of reputation

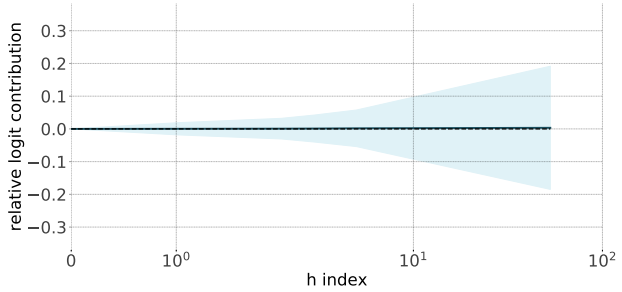

Fig. 6. Partial dependency plot of h-index

SHAP is also useful to identify important features in a single prediction instance, which can be compared to domain knowledge to validate the model is working as expected. Figure 7 shows *increasing* (pink) and *decreasing* (blue) bars for a given prediction probability, with wider bars having higher importance. This shows the features relating to a member's reputation score and total number of citations have the greatest positive influence on the prediction value, and the number of threads the member has made in the gaming category has the greatest negative influence.

Figures 8 and 9 show the distribution of two features (namely, h-index and reputation) for each set of predictions. These were selected to illustrate the differing distributions of each prediction technique. Overall, we note the distributions of predicted key actor features differ greatly for different techniques – for example, the neural network predicts a set of members with a lower distribution of h-indices than other prediction methods used. For reputation, the tail of each curve differs, highlighting the different range of members selected.

F. Validation

Predictions were made using the criteria discussed previously for social network analysis (n=13), k-means clustering (n=468), group-based trajectory modelling (n=99), logistic regression (n=53), random forest (n=128) and neural networks (n=377). Predictions are combined and validated using topic analysis.

Topic analysis is first carried out on the initial set of key actors. Figure 10 displays the various terms used by key actors across each category. Terms are selected relating to the creation or distribution of tools and tutorials, and these are compared against the set of predicted key actors. Predicted key actor results are considered validated if they use at least 20% of the selected terms used by key actors.

Of the 63 predictions of key actors where 3 or more techniques overlap, 49 are validated using topic analysis (Table V). Predictions from clustering are contained within all of the overlaps, likely caused by the greater number of predictions made by this technique. Predictions from logistic regression, random forest, neural network, and group-based trajectory modelling techniques occur in most overlaps. Social network analysis only predicts a small number of potential key actors. The threshold of overlaps could be increased to four techniques per overlap, which would be necessary if a greater number of validated key actors had been predicted at this stage. Increasing the overlap to four would predict 29 key actors, with 24 validated by topic analysis.

V. Discussion

The combination of prediction techniques helps to identify groups of predicted key actors. Results identify 7 overlaps with predictions from social network analysis, showing that bridge nodes are useful in identifying potential key actors. This suggests, as indicated by other research [2], [22], [23], that key actors in these communities may play a "mediating role", brokering connections between larger social groupings in these communities. Diversionary interventions targeted at these actors may, therefore, have a wider effect on the community as a whole.

Clustering identifies groups of similar characteristics, and the cluster used for prediction shows that predicted key actors are interested in gaming, common, and market type areas of the forum. Also, these results highlight that predicted key actors have on average a greater h-index compared to other cluster groups. This suggests that they are important parts of these communities' social structures, with connections to a wide range of other individuals on the forum.

Odds ratios found from logistic regression find an increase of h-index to increase the predicted key actors status by 1.231. Also, these show that new threads in common and hacking

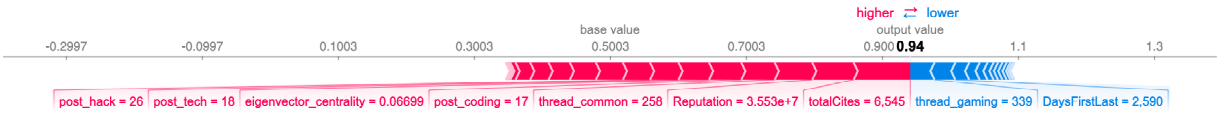

Fig. 7. SHAP values used to show feature importance for a potential key actor

SNA	Clustering	Logistic regression	Random forest	Neural network	GBTM	Predicted/Total	Avg. Distance	Farthest	Closest
	✓		✓	✓		5/6	0.5	0.34	0.61
	✓			✓	✓	4/4	0.64	0.54	0.71
✓	✓	✓				1/1	0.64	0.64	0.64
✓	✓				✓	4/4	0.66	0.5	0.82
	✓	✓	✓			3/6	0.56	0.32	0.75
	✓	✓			✓	3/3	0.42	0.29	0.57
	✓		✓		✓	5/10	0.56	0.39	0.68
✓	✓	✓		✓		1/1	0.43	0.43	0.43
✓	✓			✓	✓	1/1	0.54	0.54	0.54
	✓	✓	✓	✓		1/1	0.37	0.37	0.37
	✓	✓		✓	✓	3/3	0.57	0.39	0.75
	✓		✓	✓	✓	3/5	0.52	0.39	0.63
✓	✓		✓		✓	1/1	0.64	0.64	0.64
	✓	✓	✓		✓	4/6	0.51	0.29	0.68
✓	✓		✓	✓	✓	2/2	0.57	0.54	0.61
	✓	✓	✓	✓	✓	7/8	0.64	0.45	0.75
✓	✓	✓	✓	✓	✓	1/1	0.68	0.68	0.68
11	49	24	32	28	38	**49/63**			

TABLE V
RESULTS OF TOPIC ANALYSIS ON PREDICTIONS, WITH INTERSECTIONS OF AT LEAST THREE METHODS. SUM OF PREDICTIONS, VALIDATED BY TOPIC ANALYSIS, ARE BELOW EACH TECHNIQUE.

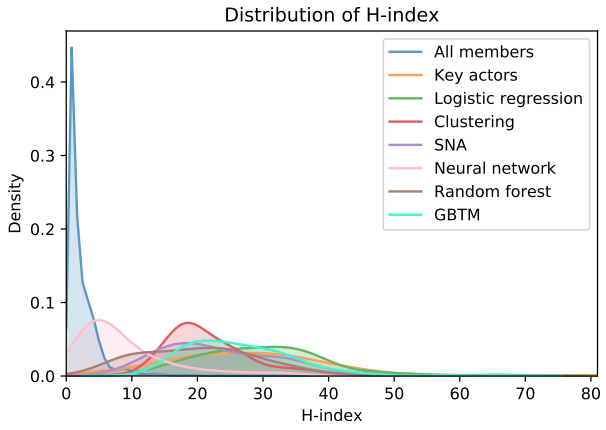

Fig. 8. Distribution of h-indices for each prediction group of actors

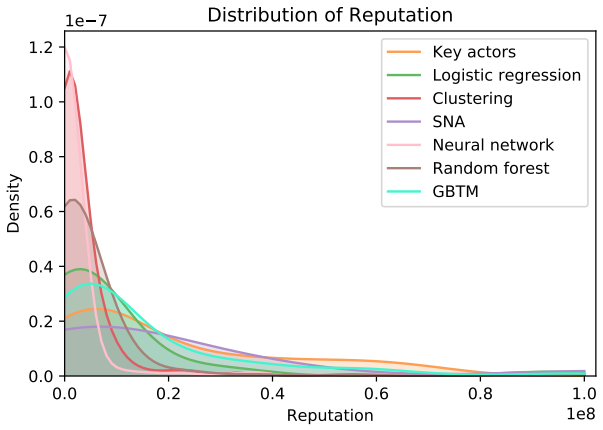

Fig. 9. Distribution of reputation scores for each prediction group of key actors

subforums increases the odds of being a key actor. However, new threads in the gaming section and posts in the currency exchange section decrease the odds of being a key actor.

Plots showing the distribution of features for random forest and neural network predictions highlight the importance of using a combination of approaches. The distribution of reputation is similar for both models, but the distribution of h-index shows the different means of the set predicted by these models.

Group-based trajectory modelling finds a high proportion of key actors with "super-engager" involvement trajectories, with high-frequency post activity for the gaming and general categories. This suggests that key actors have a characteristic trajectory of initiation on this forum, and that posting in social areas (as well as in areas more directly linked to hacking or illegal activities) is important to their involvement in these communities.

The combination of individual techniques is important, with both the combination of features to model different features available (such as post content, user activity, and interactions), and with different techniques including feature importance to signify important features in predicting. Also, analysis shows that a combination of features can provide improved prediction

html (84), thanks (83), game (83), one (83), mpgh (83), time (83), work (82), help (81), **code (80)**, name (80), post (80), thing (80), way (80), thank (79), people (79), got (79), c (79), lol (79), computer (78), version (78), thread (78), **account (78)**, **virus (78)**, man (78), **hack (77)**, day (77), copy (77), **program (76)**, ip (76), section (75), **scan (75)**, information (75), pm (74), money (73), method (73), **key (73)**, part (73), player (73), end (73), **ban (72)**, image (72), pc (72), **password (71)**, case (71), **cheat (70)**, **source (70)**, year (70), info (69), function (69), **mod (68)**, address (68), service (68), haha (68), class (68), keyboard (67), music (67), build (67), order (67), window (66), browser (66), laptop (65), news (65), card (65), **injector (65)**, weapon (65), contact (65), bump (64), **aimbot (64)**, **block (63)**, paypal (63), skype (63), mouse (63), **hacker (63)**, price (62), skill (61), range (61), flash (60), gun (60), cpu (60), troll (59), gain (59), ram (58), graphic (54), performance (54), market (53), nexon (52), board (51), string (51), trading (50), refund (48), giveaway (48), cooler (46), nx (41), budget (39), predator (38), currency (38), symmetrical (37), bitcoin (34), coin (31), xml (30), integer (27), btc (27), dim (26), eth (24), crypto (20), byval (15), congratulation (15), c++ (12), bch (7), tether (5), usdt (2)

Fig. 10. Combined list of terms used within topics (bold terms relate to the creation or distribution of tools and tutorials, and the count of key actors using these follows each term)

over a single features (such as reputation), where this can contain bias.

It is also useful to use SHAP diagrams for inspecting the model and identifying predictive features, as these show why models made certain predictions, and could be used for identifying certain intervention points.

VI. Limitations and Future Work

Despite containing over 9.4 million posts, MPGH has a smaller collection of users than some of the other forums in CrimeBB. We note, however, that our analysis is exploratory, to be later used for finding useful intervention techniques.

Key actors were initially selected using a manual process, due to a lack of existing public information relating to law enforcement activity on the forum. The manual process set out criteria for selection, and involved different areas of interest within the forum, to select a good sample of key actors. However, the selection process could benefit further from both prior knowledge of law enforcement activity, and automated annotation techniques, to build a gold-standard training dataset for analysis.

We used analysis tools by Pastrana et al. [2], created for analysis of a *general* underground forum, as we found discussion topics on the forums use a similar lexicon, and therefore the tools are able to generalise over different types of underground forums. However, different terms may be used by key actors on different forums, requiring manual human verification. Future work may build on this to automatically classify common and new terms, to detect those which may be used for predicting potential key actors, with advanced NLP tools.

Most techniques used looked at the dataset as a cross-section, except for the time-series approach with group-based trajectory modelling. It would be interesting to adapt existing methods to time-series approaches, to assist with research into cybercrime pathways. This may include adapting the topic analysis tools to include language evolution over time.

Also, further analysis into subgroups of key actors may be of interest, through further clustering techniques, including topic analysis within clusters, or time-series clustering approaches including k-means of longitudinal data.

VII. Conclusion

We took a multi-technique approach to predict key actors, members who are participating in sustained involvement of harmful or illegal behaviour, that were likely to indicate that they might benefit from diversionary approaches. This prediction approach was used to identify potential key actors, and to characterise their involvement in the community on this underground forum through a range of measures. This research both develops a more generally-applicable prediction mechanism for identifying and characterising key actors in underground forums, and elucidates a range of potential insights into the roles played by these actors in these communities and their pathways of initiation.

We proposed a systematic framework for analysing forum datasets. These require big data approaches for analysis, due to the scale and unstructured nature of the data. This could constitute the foundation of a more general approach to the study of underground forum communities, and research on hard-to-use scraped forum data.

We applied the framework for analysis of a dataset consisting of posts and threads from the forum MultiPlayer Game Hacking, selected for the forum's focus on both gaming and hacking. 49 predictions of potential key actors were found, by combining predictions from intersections of at least three different types of analysis techniques, validated using topic analysis.

Analysis included three types of clustering techniques, namely k-means clustering, social network analysis, and group-based trajectory modelling. In addition to the predictions from these, three predictive models were also used, namely logistic regression, random forest, and neural networks. However, our analysis may be limited by the initial set of key actors, and future work should improve on this area.

Use of topic analysis for checking prediction results, and overlaps of prediction sets, assisted in validating predictions. NLP features used did not have a large positive or negative influence for many of the models, due to high correlation with other features, and similar values across different users. Group-based trajectory models identified two trajectories which contained a high proportion of key actors. Future work in this area could model development of cross-forum pathways over time, by developing time-series analysis tools, including the use of other time-series based machine learning techniques.

Acknowledgments

We thank the Cambridge Cybercrime Centre for access to the CrimeBB dataset. We also thank Daniel R. Thomas, Sergio Pastrana, Andrew Caines, Paula Buttery, Richard Clayton, Alastair Beresford, Ross Anderson, and our other colleagues at the Cambridge Cybercrime Centre.

REFERENCES

[1] National Crime Agency. Pathways Into Cyber Crime. http://www.nationalcrimeagency.gov.uk/publications/791-pathways-into-cyber-crime/file, 2017.

[2] Sergio Pastrana, Alice Hutchings, Andrew Caines, and Paula Buttery. Characterizing eve: Analysing cybercrime actors in a large underground forum. In *Proceedings of the 21st International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*, pages 207–227. Springer, 2018.

[3] Ahmed Abbasi, Weifeng Li, Victor Benjamin, Shiyu Hu, and Hsinchun Chen. Descriptive analytics: Examining expert hackers in web forums. *Proceedings of the IEEE Joint Intelligence and Security Informatics Conference (JISIC)*, pages 56–63, 2014.

[4] Rebekah Overdorf, Carmela Troncoso, Rachel Greenstadt, and Damon McCoy. Under the underground: Predicting private interactions in underground forums. *arXiv preprint arXiv:1805.04494*, 2018.

[5] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. An analysis of underground forums. In *Proceedings of the 2011 ACM SIGCOMM Internet Measurement Conference (IMC)*, pages 71–80. ACM, 2011.

[6] Andrew J Park, Richard Frank, Alexander Mikhaylov, and Myf Thomson. Hackers Hedging Bets: A Cross-Community Analysis of Three Online Hacking Forums. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 798–805. IEEE, 2018.

[7] Ericsson Marin, Jana Shakarian, and Paulo Shakarian. Mining key-hackers on darkweb forums. *Proceedings of 1st International Conference on Data Intelligence and Security (ICDIS)*, pages 73–80, 2018.

[8] Sadia Afroz, Vaibhav Garg, Damon McCoy, and Rachel Greenstadt. Honor among thieves: A common's analysis of cybercrime economies. In *2013 APWG eCrime Researchers Summit*, pages 1–11, Sep. 2013.

[9] Andreas Haslebacher, Jeremiah Onaolapo, and Gianluca Stringhini. All your cards are belong to us: Understanding online carding forums. In *2017 APWG Symposium on Electronic Crime Research (eCrime)*, pages 41–51, April 2017.

[10] PC Gamer. Criminal hacking is not a game worth playing. https://www.pcgamer.com/uk/criminal-hacking-is-not-a-game-worth-playing/, 2019.

[11] Kotaku. What's the deal with cybercrime? https://www.kotaku.co.uk/2019/04/08/whats-the-deal-with-cybercrime, 2019.

[12] Games Radar. Upgrade your gaming: DDoS attacks, cracking, and how to protect yourself from online cyber crime. https://www.gamesradar.com/au/upgrade-your-gaming-ddos-attacks-cracking-and-how-to-protect-yourself-from-online-cyber-crime/, 2019.

[13] Mohammad Karami and Damon McCoy. Rent to pwn: Analyzing commodity booter ddos services. *Usenix login*, 38(6):20–23, 2013.

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

[15] Sergio Pastrana, Daniel R. Thomas, Alice Hutchings, and Richard Clayton. CrimeBB: Enabling Cybercrime Research on Underground Forums at Scale. In *Proceedings of The Web Conference 2018*, Lyon, France, 2018.

[16] Isuf Deliu, Carl Leichter, and Katrin Franke. Collecting cyber threat intelligence from hacker forums via a two-stage, hybrid process using support vector machines and latent Dirichlet allocation. *Proceedings of IEEE International Conference on Big Data (Big Data)*, pages 5008–5013, 2018.

[17] Mitch Macdonald, Richard Frank, Joseph Mei, and Bryan Monk. Identifying digital threats in a hacker web forum. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 926–933. IEEE, 2015.

[18] Andrew Caines, Sergio Pastrana, Alice Hutchings, and Paula J. Buttery. Automatically identifying the function and intent of posts in underground forums. *Crime Science*, 7(1):19, Nov 2018.

[19] Victor Benjamin and Hsinchun Chen. Securing cyberspace: Identifying key actors in hacker communities. *IEEE International Conference on Intelligence and Security Informatics: Cyberspace, Border, and Immigration Securities (ISI)*, pages 24–29, 2012.

[20] Xiong Zhang and Chenwei Li. Survival Analysis on Hacker Forums. In *SIGBPS Workshop on Business Processes and Service*, pages 106–110. 2013.

[21] Baidyanath Biswas, Arunabha Mukhopadhyay, and Gaurav Gupta. "Leadership in action: How top hackers behave" A big-data approach with text-mining and sentiment analysis. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, volume 9, 2018.

[22] Sagar Samtani and Hsinchun Chen. Using social network analysis to identify key hackers for keylogging tools in hacker forums. *IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data (ISI)*, pages 319–321, 2016.

[23] Hamed Sarvari, Ehab Abozinadah, Alex Mbaziira, and Damon McCoy. Constructing and analyzing criminal networks. *Proceedings IEEE Symposium on Security and Privacy*, January:84–91, 2014.

[24] Jan William Johnsen and Katrin Franke. Identifying central individuals in organised criminal groups and underground marketplaces. In *International Conference on Computational Science*, pages 379–386. Springer, 2018.

[25] Ryan Seebruck. A typology of hackers: Classifying cyber malfeasance using a weighted arc circumplex model. *Digital Investigation*, 14:36–45, 2015.

[26] Xiong Zhang, Alex Tsang, Wei T. Yue, and Michael Chau. The classification of hackers by knowledge exchange behaviors. *Information Systems Frontiers*, 17(6):1239–1251, 2015.

[27] Richard Frank, Myfanwy Thomson, Alexander Mikhaylov, and Andrew J. Park. Putting all eggs in a single basket: A cross-community analysis of 12 hacking forums. In *IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 136–141, 2018.

[28] British Society of Criminology. Statement of ethics. http://www.britsoccrim.org/ethics/, 2015.

[29] Daniel S. Nagin. *Group-based modeling of development*. Harvard University Press, 2005. ISBN: 9780674016866.

[30] Daniel S Nagin and Candice L Odgers. Group-based trajectory modeling (nearly) two decades later. *Journal of Quantitative Criminology*, 26(4):445–453, 2010.

[31] Bobby L Jones and Daniel S Nagin. A note on a stata plugin for estimating group-based trajectory models. *Sociological Methods & Research*, 42(4):608–613, 2013.

[32] Andy Field. *Discovering Statistics Using SPSS*. London: SAGE Publications, 2nd edition, 2005.

[33] François Chollet et al. Keras. https://keras.io, 2015.

[34] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[35] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[36] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[37] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv e-prints*, page arXiv:1802.03888, Feb 2018.