Hacker Engagement Prediction: Analysis

1st Casey Cannon
dept. Computer Science
California State Polytechnic University, Pomona
Pomona, California
clcannon@cpp.edu

2nd Michael Melkonian
dept. Computer Science
California State Polytechnic University, Pomona
Pomona, California
mamelkonian@cpp.edu

***Abstract*—With the given evolution of hacker communities among the deep/dark-web the necessity for a proactive cyber-threat intelligence system is at an all-time high to combat zero-day attacks as well as other cyber threats alike. Hence, with the collection of influence measures sourced from currently available hacker forums, machine learning models in the likes of sequential rule mining, logistical regression, and random forests can be applied toward predicting individual hacker-related activity. The goal would ultimately be to track down individuals who have a high priority in influencing deep/dark-web communities. This research seeks in assessing the effectiveness of present influence-ability measures, features, and behavior-adoption prediction techniques against a deep/dark-web hacker forum data-set.**

I. Introduction

The spread of adoption behavior between individuals within a social network was most effective with the current method of the observation and prediction of trends. The initial experimentation that has been performed against data-sets containing social-networks proves some algorithms to be better suited in successfully predicting the user engagement with the given social-network community. The implementation of these current algorithms proves to successfully predict greater than 50% of the engaged users within a social-network. However, with current research, it is important to note that the integration of temporal elements in the form of time constraints has adequately boosted results to the user-engagement prediction by 18.89%.

The goal of this paper is to compare existing user-engagement analysis with improved methods like temporal elements, to ultimately lead towards a more accurate system that promotes proactive cybersecurity.

II. Background

The fundamental prerequisites to this research are a strong understanding of social network analysis through graph-based algorithms and a strong understanding of supervised machine learning. Even though the networkx python library will be primarily responsible for social network feature extraction, knowledge of available network features is useful for planning the features of the machine learning models. Social Media Mining by Zafarani will serve as the base reference for social network analysis concepts and algorithms.

Additionally, chapters of Networks, Crowds, and Markets: Reasoning About a Highly Connected World by Easley and Kleinberg may also be used for reference. The book is framed from an interdisciplinary perspective, drawing on ideas from economics, sociology, computing, information science, and applied mathematics. Though the whole work is useful as a reference, the most relevant chapters are most likely 16, 19, and 21 (Information Cascades, Cascading Behaviors in Networks, and Epidemics, respectively).

CalSys labs performs research on several different approaches to the concept of predictive machine learning models in a dark-web environment. Some of these approaches include community finding, information cascade prediction, key-hacker identification, and user-engagement prediction. The projects may all have different end-goals, but ultimately, success in one project potentially bolsters the success of others. For example, finding key-hackers in a community could benefit prediction of user-engagement, as a model would be able to flag key-hackers and factor in their unique, potentially stronger influence on the network.

Cascade prediction is useful in flagging topics that will potentially see interaction of a certain order-of-magnitude increase. In the 2015 article by Ruocheng Guo and colleagues, their model is able to successfully predict microblogs that will grow from 50 to 500 reposts with a precision of 0.69 and recall of 0.52. This may sound middling, but these ”viral” microblogs actually only account for 2% of their dataset. This greatly outperforms what was thought of at the time as the current ”state of the art”. If combined with user-engagement prediction, which determines if specific users will interact with certain posts, one could potentially create a model that could predict which posts will garner the most attention from which users.

This research will focus on user-engagement prediction. The high-level goal of behavior adoption prediction is to take a topic and its posts, a user, and the respective features of the user's active neighborhood to classify whether the user is expected to engage with the topic in question; 1 for yes, and 0 for no.

Behavior-adoption prediction research is often seen with regard to surface-web explicit social networks. An informative work by State and Adamic seeks to track the behavior spread of adopting profile pictures supporting same-sex marriage.

While previous studies showed users experience diminishing returns of influence after repeated exposures to a behaviors, State and Adamic find that the influenceability of their studied behavior would peak at about 3 or 4 exposures, then see dimishing returns thereafter. [?] This reveals that behavior spread is variable based not only on the environment and network structure, but also on the subject matter of the behavior as well.

There are several papers that serve as a foundation to this research. One such paper is Measuring Time-Constrained Influence to Predict Adoption in Online Social Networks by Ericsson Marin, Ruocheng Guo, and Paulo Shakarian. The article both reviews existing measures for estimating influence while showcasing the improvement in these measures when applying the Susceptible Span and Forgettable Span time constraints. Furthermore, these new measures with tuned temporal constraints are used as features in contemporary adoption-prediction models, and boast improvements of up to 18.89%.

The Susceptible Span (T_{sus}) refers to the interval when people are able to receive social signals from their neighbors; intuitively, how long a user can go without an interaction between their neighbor before the influence of that connection weakens. Forgettable Span (T_{fos}) refers to the interval when people are able to remember the actions performed by their neighbors, making these actions eventually forgotten as time passes. Holistically, T_{sus} is the amount of time a node visualizes a neighbor's actions, while T_{fos} illustrates how long the node "remembers" these subsequent actions.

Without the implementation of T_{sus} and T_{fos}, the models assume "infinite" T_{sus} and T_{fos} values. This means that nodes remember every connection they've ever made and have the ability to consider every action made by that connection since the beginning of the relationship. Over short periods of time this might be okay, but intuitively, the human memory is not infinite. Additionally, adding the temporal element helps to better identify times of high and low concentrations of adoption. A survey of correlation coefficients of 10 features are then observed with varying values of T_{sus} and T_{fos}. The study finds that implementing the time constraints can increase correlation coefficients by up to 518.75% in certain cases. The study also finds that the correlation of certain features differs between Twitter and Sina Weibo. Additionally, the correlations seem dependent on the initial filtering of the user data.

This study ultimately yields an 18.89techniques also trying to predict adoption.

Another area for improvement of features lies in the active user count. In the online forums by which our research is based, users respond chronologically to topics. It is assumed that a user is replying to all those who posted previously when responding to a post. Realistically, a user may actually be replying to one or a subset of previous posts instead of all that came before. So, the accuracy of models may be improved upon if one is able to distinguish the lines of conversation.

In Interaction Coherence Analysis for Dark Web Forums by Tianjun Fu, Ahmed Abbasi, and Hsinchun Chen uses various system and linguistic features to identify interaction networks specifically in dark web forums.

III. Goals

Upon performing the research, we can note three major goals:

1) Re-evaluate correlation coefficients of key influencability measures against the deep/dark-web forum database.
2) Determine the effectiveness of existing adoption prediction models against deep/dark-web forum environments.
3) Improve on the current accuracy predictions.

The primary improvement behind the current accuracy of the predictions would be dependent upon the results from the first two goals. Reason being, once the measures in likes of temporal aspects and features are tuned for each given dataset that is being worked with, the subsequent step/goal would be to determine the optimal approach.

IV. Features

The below features are sourced from *Measuring Time-Constrained Influence to Predict Adoption in Online Social Networks* where the following aspects of the dataset are quantified for association.

1) Connectivity:
   a) Number of Influential Active Neighbors
   b) Personal Network Exposure
2) Temporal:
   a) Continuous Decay of Influence
3) Recurrence:
   a) Previous Reposts
4) Transitivity:
   a) Closed Triads
   b) Clustering Coefficients
5) Centrality:
   a) Hubs
6) Reciprocity:
   a) Mutual Reposts
7) Structural Diversity:
   a) Active Strongly Connected Components Count
   b) Active Strongly Connected Components Ratio

V. Methodology

The dataset that is currently being leveraged for this research are web-scraped hacker-related forums consisting of approximately 100 individual forums. The forums vary in aspects of language, size, and user-ship while having the largest population of users found among forums: 34, 41, 77, and 84.

With the implementation of the features seen above, along with temporal-constraints the behavior among the two datasets found in *Measuring Time-Constrained Influence to Predict Adoption in Online Social Networks* most notably Twitter and Sina Weibo displayed incredible behavior correlation. Coupling this with how noise filtering was integrated:

Noise Filtering Tactics:

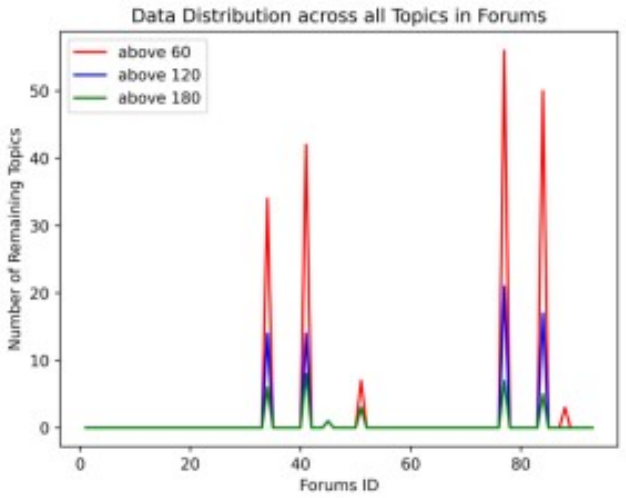

Fig. 1. High-User Count Topic Frequency by Dhanush Karthikeyan

1) Minimum Number of Retweets **(R60)**
2) Minimum Number of Hashtags **(H40)**
3) Minimum Influenceability Score **(I50)**

The current models that are seen from the *Measuring Time-Constrained Influence to Predict Adoption in Online Social Networks* document provide analysis through Logistical Regression and Random Forrests. Although these models are essentially designed to make their predictions based off of the strucutre of the social connections, the very models could be strengthened when assessing the network.

This being said, once the 10 features as well as the temporal aspects in the likes of T_{sus} and T_{fos} are tuned with respect to each dataset that is being worked on, the subsequent step would be to leverage a machine learning model in accurately predicting the user-adoption within these social-networks. As stated prior, the reference [4] had leveraged both Logistical Regression and Random Forrests however only reported results from Random Forrests. Along with this, models in the likes of Sequential-Rule-Mining could be leveraged in order to improve accuracy with regards to user to user engagement. [5]

VI. Evaluation of Results

Currently the results are still pending to be evaluated as the data from the collected deep/dark-web hacker-forums must still go through the machine-learning model.This being said, with the current data, primarily from Forum-77 it was important in conducting further analysis of the user-posts per thread. The reasoning behind this stems simply in having a need for understanding the optimal value of user-posts per thread to obtain the best possible results. As can be seen from Figure 2, the initial user-post thread threshold value is set to a low value of 5 and with this, information we can see that the majority of threads have approximately 50-70 users post at most 5 times within the specified thread. As we slightly increase this threshold value to 15, we can see that there really

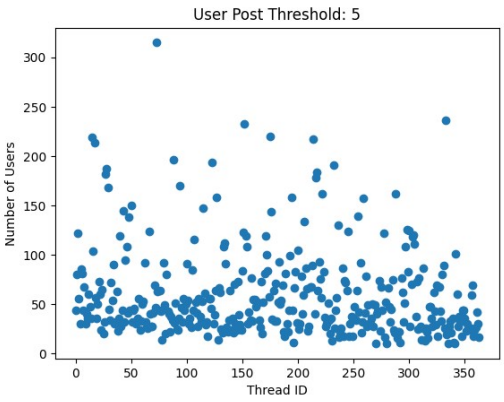

Fig. 2. User-Post Threshold Set to 5

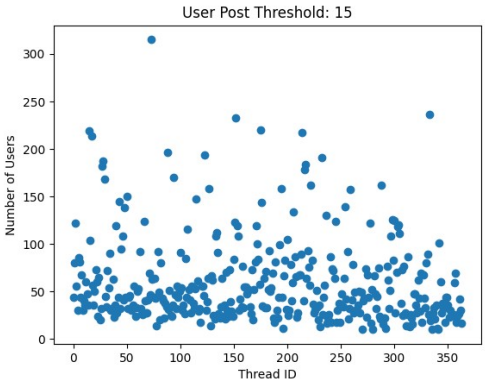

Fig. 3. User-Post Threshold Set to 15

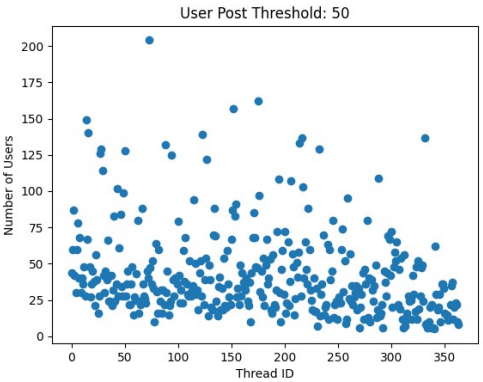

Fig. 4. User-Post Threshold Set to 50

is not much change, hence our initial threshold value can be relaxed to 15 without any cutoff of vital users-per-thread.

However, this being said, as we increase the threshold value to 50 and even 100, we can begin to see a notable reduction in the quantity of user-posts per thread. This signifies that the optimal user-posts per thread value for this particular hacker-forum, Forum-77, lies somewhere between 15 and 30 user-posts per thread. This threshold value provides confidence in retaining the majority of user-posts within any given thread from the Forum-77 dataset.

VII. Conclusion

We have seen the results of the traditional implementation of algorithms that provide findings of user-engagement within a forum holding a success-rate of greater than 50%. Also, we have noted that with the integration of temporal elements in the likes of T_{sus} and T_{fos} has proven an additional improvement to the current system by 18.89%. The availability in tuning temporal values coupled with the 10 different features covered in Section 4 (Methodology), can be leveraged to apply towards all forums: 34, 41, 77, and 84. With this, we can obtain more robust results from the machine-learning model to ultimately have a clearer understanding of the prospective behavior-adoption present in modern deep/dark-web hacker forums.

We compared the classification results of five machine learning algorithms: SVM, Neural Network, Random Forest, Logistic Regression and Ada Boost. Due to clean data, all the models obtained by each algorithm had high accuracy of above 80%. However, Ada Boost provided the most accurate model having accuracy of 88.4% We attempted to take it further and ensemble (SVM, Random Forest, Logistic Regression and Ada Boos), but this did not lead to a higher accuracy. This model was set up giving one vote each algorithm leaving us to believe the algorithms were getting the same incorrect predictions on the same instances. Possibly this means there is a unknown attribute we are missing from our data, or that we need more data to better our accuracy.

VIII. References

[1] D. Easley and J. Kleinberg. Networks, crowds and markets: Reasoning about a highly connected world. Cambridge University Press, 2019.

[2] R. Zafarani, M. A. Abbasi, and H. Liu. Social Media Mining: An introduction. Cambridge University Press, 2014.

[3] E. Nunes, A. Diab, A. Gunn, E. Marin, V. Mishra, V. Paliath, J. Robertson, J. Shakarian, A. Thart, and P. Shakarian. Darknet and deepnet mining for proactive cybersecurity threat intelligence. In 2016 IEEE Conference on Intelligence and Security Informatics (ISI), pages 7–12, 2016. doi: 10.1109/ISI.2016.7745435.

[4] E. Marin, R. Guo, and P. Shakarian. Measuring time-constrained influence to predict adoption in online social networks. ACM Transactions on Social Computing, 3(3):1–26, 2020.

[5] E. Marin, M. Almukaynizi, E. Nunes, J. Shakarian, and P. Shakarian. Predicting hacker adoption on dark-web forums using sequential rule mining, 2014. URL https://ieeexplore.ieee.org/ document/8672225/.