Topology-Driven Detection of Virality in Malicious Cascades

Dhanush Karthikeyan
College of Engineering
California State Polytechnic University, Pomona
dkarthikeyan@cpp.edu

Dr. Ericsson Marin
College of Science
California State Polytechnic University, Pomona
santanamarin@cpp.edu

***Abstract*—When any malicious information—such as links, videos, photographs, microblogs, and metadata—are posted online, one fundamental question arises: can it propagate to viral proportions? In this project, we demonstrate whether social influence features extracted from a sequence of posts in a forum thread can distinguish between a viral and non-viral cascade. Through anonymous communities on secure sites and forums, there is now an unprecedented flow of ideas, malware, and exploits. The development of machine learning models to identify viral cascades in their infancy can be leveraged by security specialists to prevent mass-adoptions of malware. Through training on balanced datasets from on our forum data, we find that we get an average F1 score of 0.65.**

***Keywords*—Social Network Analysis, Cascade, Darknet**

I. Introduction

WHEN malicious information—such as links, videos, photographs, microblogs, and metadata—is posted online, one fundamental question arises: can it propagate to viral proportions?

In this project, we demonstrate whether social influence features extracted from a sequence of posts in a forum thread can distinguish between a viral and non-viral cascade. Through anonymous communities on secure sites and forums, there is now an unprecedented flow of ideas, malware, and exploits. Without visibility into this new offensive industrial base, the production pipeline is abstracted from defenders. This has prompted the development of machine learning models to identify viral cascades in their infancy, which can then be leveraged by security specialists to prevent mass-adoptions of malware.

The motivation behind this research is to construct classification models that anticipate hacktivism campaigns and mass-adoptions of cyber threats using structural features extracted from the cascade [1,2]. This task can be accomplished through the identification of a viral cascade in its early stages through binary classification. Previous research has focused on identifying cascades in social networks through a combination of temporal and structural features. However, we seek to extend the application of these features to malicious forums from the Dark Web, using features extracted from the structural characteristics of the cascade utilizing social network analysis. In addition, this analysis of our forums will be strictly topology-driven using the social network metrics (no content information from the forum posts was analyzed). Given a sequence of user posts in a thread, our problem consists in predicting whether this cascade will achieve viral proportions. For the scope of our research, a "viral" cascade is defined as any thread which displays a multiplier increase in user adoptions.

To test our approach, we train a variety of classifiers: Random Forest, AdaBoost, Naive Bayes, etc. In the data collected through the CYR3CON API, it is important to note an imbalance in the classes. The ratio between the positive (viral) class and the negative (nonviral) class heavily leans towards the negative class across datasets with various levels of cascade growth. In our training of the model, we address any bias through the one-to-one sampling of our training dataset.

This work provides the following main contributions:

1) First application of machine learning for virality detection on malicious cascades.

2) Research into the extraction of social features from malicious cascades, both viral and nonviral.

3) Comparison of different machine learning models with different levels of cascade growth.

II. Technical Preliminaries

In this section, we introduce the necessary notation and description of our malicious networks. Social networks are represented as graphs, G = (V,E) where V denotes our nodes and E represents our edges [2]. Using directed graphs, we can trace the influence of one node above another. Likewise, the use of weighted edges is utilized to determine the growth of influence of node over another.

A. Cascades in the Dark Web

In this work, a cascade can be any thread in which we have at least α number of posts in sequence, where α is a user-designated value relative to the overall number of distinct users who partake in conversation in these forums.

B. Viral vs Nonviral Cascades

Cascades can further be categorized as a viral cascade or nonviral cascade. A **viral cascade** is any thread which displays a multiplier increase in user adoptions, such as doubling or tripling from initial cascade size.

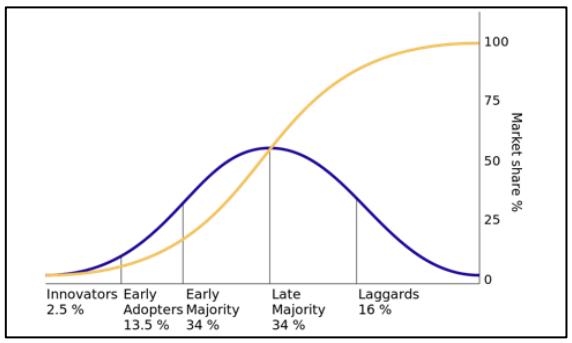

Fig. 1. Growth of viral (yellow) vs non-viral (blue) cascade according to Rogers *Diffusion of Innovations*

Within cascades, we have the following distinct groups: innovators, α early adopters, and β early majority. The innovator is essentially the root user, or the originator of the cascade. Their post initiates the cascade; thus, it is important to measure the social influence of this root user. To be identified as a potential cascade, we first look at the following:

- initial size threshold to be considered a cascade: α
- upper size threshold to be considered a cascade: β

We establish two threshold values to be considered a cascade. The first value is α which denotes the lower threshold for a group of initial adopters to be considered for a potential viral cascade. The upper threshold is denoted by β which is a measure of the early majority. This value is a multiple of the lower threshold, such as $\beta = 2\alpha$, $\beta = 3\alpha, or\ \beta = 4\alpha$. As shown in Roger's work [4], adoption of any material can either reach saturation levels, as in the case of a viral cascades, or reach a peak before weaning in interaction.

C. *Dark Web Data*

The dark web data used for this project is retrieved with from CYR3CON, which provides an API for security researchers. This company employs human analysts to inspect anonymous forums and extract forums logs through crawlers. They then make use of proprietary classifiers to isolate the malicious data, which is then made available to security researchers. Through this API and connection to this security company, we now have access to over 2.5 million posts within the saved CYR3CON database.

D. *Social Network Analysis*

To extract social influence features from our cascades, social network analysis is applied to study information diffusion through the following process.

- A directed network is built for each selected forum.
- Extract root user and cascade features from network
- Infer viral vs nonviral cascade based on features

These social networks are generated using the NetworkX python library, which dynamically maps the CYR3CON data to generate directed graphs. Each distinct user in the malicious forum becomes a node in the network graph, from which social influence can be mapped through edges connecting nodes. Directed edges are used to map the one-way flow of influence, where users responding to a thread are perceived as being influenced by previous users. As more connections are made between any two nodes, the edge weight is incremented to reflect this growing influence. The network graphs grow dynamically as the cascade grows, allowing for analysis at any point in the growth of the viral cascade.

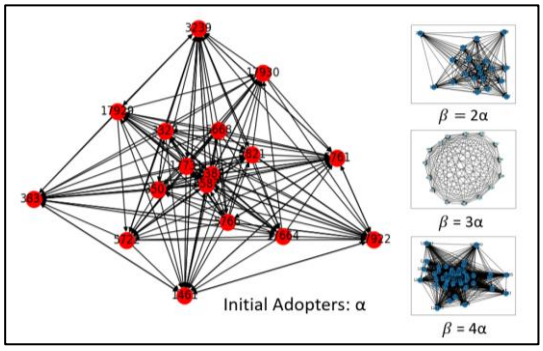

Fig. 2. Growth of a Viral Cascade visualized through NetworkX

III. DATASET

The CYR3CON data consists of anonymized chat logs from a variety of forums, populated with uniquely identifiable posts. For the scope of our analysis, we will abstract away from the specific content in these messages and instead focus on the cascade structure of the forums.

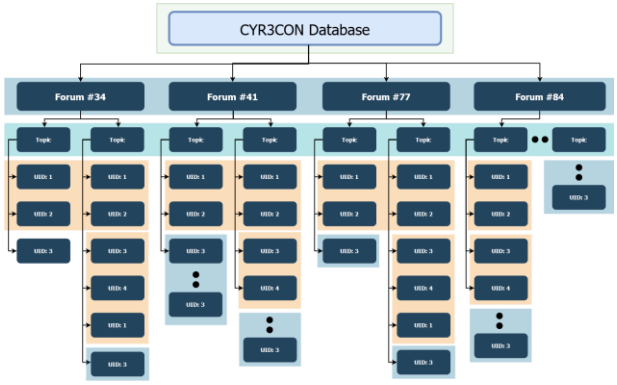

Fig. 3. Forum Hierarchy of Dark Net data aggregation

Through analysis of this data, four forums were identified that contained a high number of threads with a high number of unique users, which are ideal to extract viral cascade features.

These four forums are used to generate four different datasets, since cascade characteristics are unique to the forum. It is important to note the disproportion in our viral and non-viral classes due to the small number of viral cascades in our data. To mitigate this, our minimum cascade size threshold α is lowered to allow for more potential viral cascades. The final ratio between the positive and negative classes—viral and nonviral respectively—brings this ratio to 1:4. This is further

mitigated through down-sampling of the negative class to achieve a relatively balanced training set. In total, our final dataset consisted of 16 features as detailed in the next section.

IV. FEATURE ENGINEERING

The generated network graphs are leveraged to extract two suites of features: root user features and early adopter features. These features are visualized in Table 1, where the dichotomy of the two user groups is observed. We see how many of the features are retrieved from both user groups.

Table. 1. Feature Engineering Breakdown

Analyzed User(s)	Features Extracted
Root User	• degree centrality
	• out_degree centrality
	• out-edge eigenvector centrality
	• out degree weight
Early Adopters (α)	• avg number of neighbors
	• avg out degree
	• time elapsed to α
	• avg out degree weight sum
	• average page rank
	• group out degree centrality
	• group_betweenness centrality
	• group closeness centrality
	• average time to adoption

V. RESULTS

Various classification models were trained and tested on the balanced dataset, resulting in the F1 scores displayed in Figure 5. The classification models chosen are Random Forest, Adaboost, Gaussian Naive Bayes, Multinomial Naive Bayes, and Bernoulli Naive Bayes. They were tested with all four forums with a cascade growth of α to 2α. These results are depicted in Figure 4, where we see the aggregate F1 score of each model averaged out across all four forums.

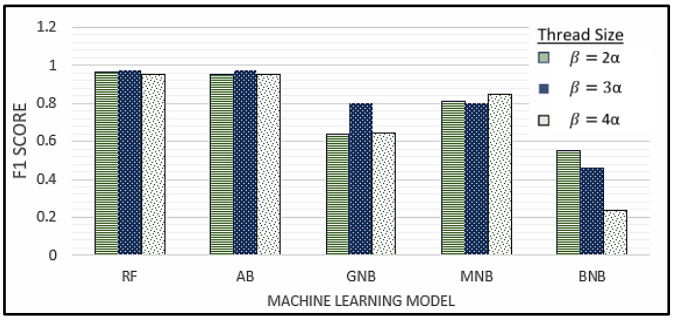

Fig. 4. F1 scores when $\boldsymbol{\alpha}$ = 30 and varying $\boldsymbol{\beta}$ cascade growth

These results are promising and indicate the possibility of cascade-based features to classify virality in malicious cascades. Future work will include the application of classification models on unbalanced datasets and the use of time-related spans to generate dynamic charts for more specific predictions. It is important to note that this testing was done with a balanced dataset, which is not indicative of the imbalanced observed in the database.

VI. CONCLUSION

In this work, we found promising results with the use of cascade-perspective features in a topology-driven scheme. With further testing and optimization using an imbalanced dataset, these can be tested in the wild for virality detection. Ideally, this project can be extended to include user adoption features, which entails cascade prediction based on individual user metrics. We can also make use a combination of temporal and structural features are key predictors of cascade growth. Making use of temporal spans, it is possible to create dynamic social networks to measure social influence at any given time, which is better indicative of real-world influence.

VII. ACKNOWLEDGEMENT

I wish to acknowledge the NSF REU site at California Polytechnic University, Pomona for their support in this research initiative. With NSF funding and the direction of the PI, Dr. Chen, it has been possible for me to learn more about social network analysis and the application of machine learning. I would like to thank my advisor, Dr. Marin, whose guidance made this project possible and who has developed the steps to take this project to the next level. I would also like to FengRu Yang, my research colleague who provided extensive support with the application of machine learning.

REFERENCES

[1] Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., & Leskovec, J. (2014, April). Can cascades be predicted?. In Proceedings of the 23rd international conference on World wide web (pp. 925-936).

[2] Ericsson Marin, Ruocheng Guo, and Paulo Shakarian. 2020. Measuring Time-Constrained Influence to Predict Adoption in Online Social Networks. ACM Trans. Soc. Comput. 3, 3, Article 13 (May 2020), 26 pages.

[3] Guo, Ruocheng, et al. "Toward early and order-of-magnitude cascade prediction in social networks." Social Network Analysis and Mining 6.1 (2016): 1-18.

[4] Rogers, Everett M. Diffusion of innovations (1st ed.). New York: Free Press of Glencoe. OCLC 254636. (1962).

Dhanush Karthikeyan is a senior at California Polytechnic University, Pomona who is double-majoring in Electrical Engineering and Computer Engineering. After industry experience in aerospace and consumer technologies, he worked on his start-up to commercialize NASA patents. After working on IOT and biomedical sensor products, his interests are now set on applying machine learning to cybersecurity and social network analysis.

Dr. Ericsson Santana Marin is an Assistant Professor in the Computer Science Department at California State Polytechnic University - Pomona (Cal Poly Pomona). He earned his B.S. in Computer Science from Pontifical Catholic University of Goias, Brazil (2001), a specialization in Software Quality Assurance and Management from Pontifical Catholic University of Goias, Brazil (2013), a M.S. in Computer Science from Federal University of Goias, Brazil (2013), and a Ph.D. in Computer Science from Arizona State University, USA (2020), where he proposed a hacker-centric perspective to empower cyber-defense. After defending his Ph.D. dissertation in April 2020, he joined Cal Poly Pomona in Fall 2020. He has also worked from 2001 until 2010 in his own software factory− Marin Solutions− where his team designed custom-built, requirements-oriented, and high-performance software solutions for different types of companies.