Topology Driven Virality Detection in Malicious Cascades

Dhanush Karthikeyan
Faculty Advisor: Dr. Marin
NSF REU Site: Big Data Security and Privacy, California Polytechnic University - Pomona

Background

When malicious information – such as links, videos, photographs, microblogs, metadata, etc – are posted online, one fundamental question arises: can it propagate to viral proportions? In this project, we demonstrate whether features extracted from the social influence of users in a cascade 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 my security specialists to prevent mass-adoptions of malware.

Motivation

- ○ We seek to construct classification models that anticipate hacktivism campaigns and mass-adoptions of cyber threats
- ○ Accomplish this task through the identification of a viral cascade in its early stages through binary classification

Problem Statement

We seek to identify and forecast the potential for a malicious forum thread to go viral. For the scope of our research, a "viral" cascade is any thread which displays a multiplier increase in user adoptions.

Data Collection

Raw data is extracted from CYR3CON, which employs human analysts to inspect anonymous forums and extract forums logs through crawlers (Fig 1). Proprietary classifiers 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 data.

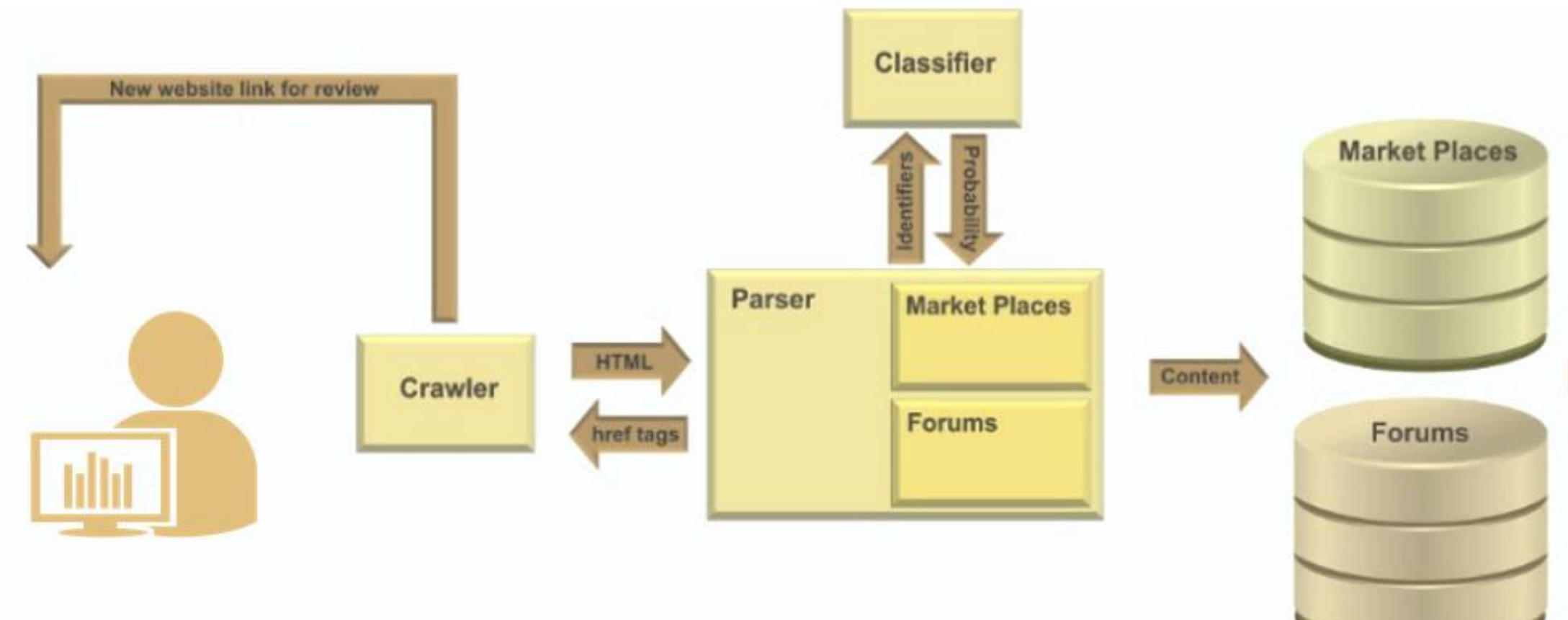

Fig 1. CYR3CON Pipeline

Data Analysis

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 (Figure 2) of the forums.

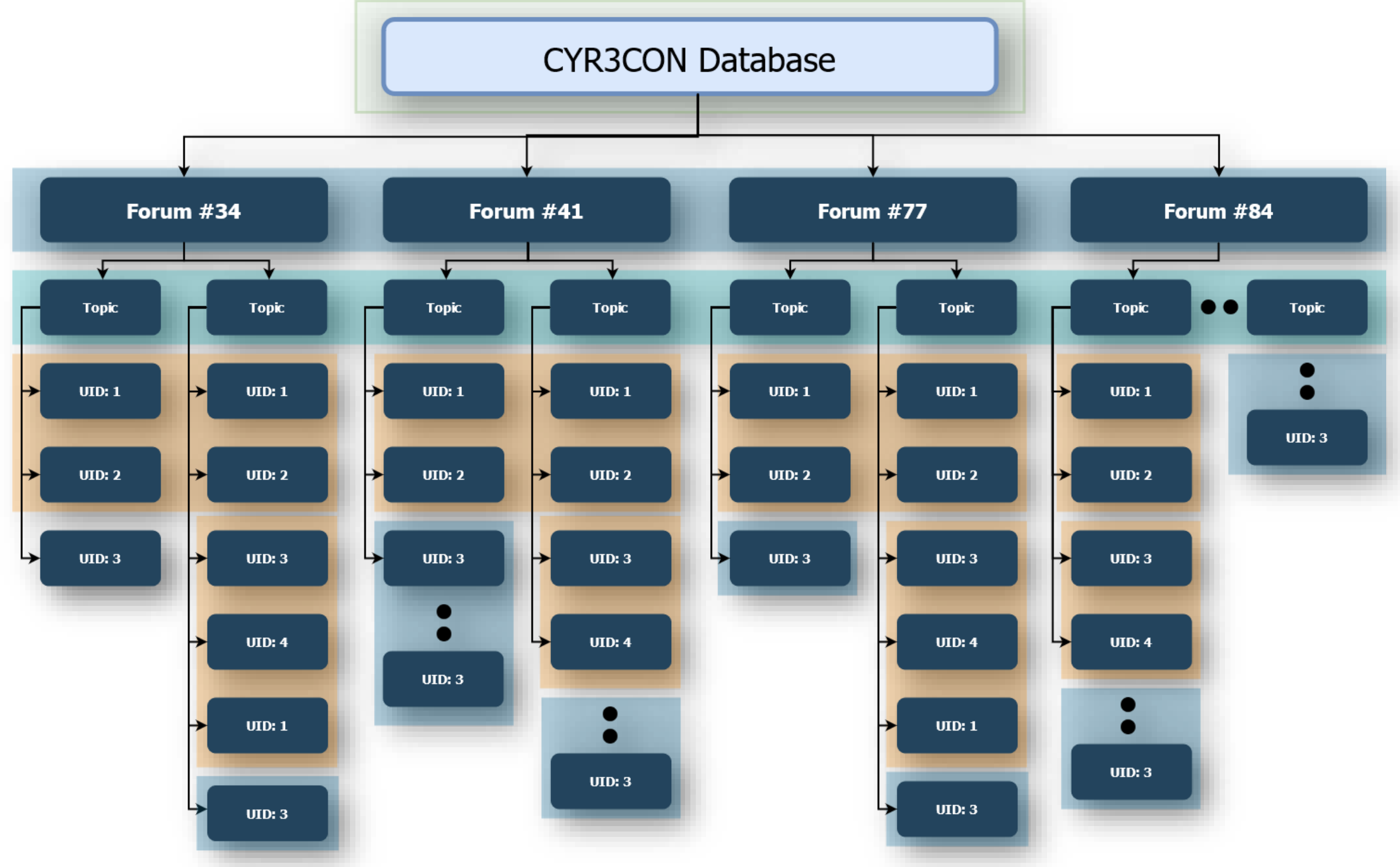

Fig 2. CYR3CON Forum Hierarchy

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 (Fig 3) 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 dataset.

Viral Cascade Prediction

In this work, a cascade can be any thread in which we have a certain number of posts in sequence, which 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 (Fig 3), such as doubling or tripling from initial cascade size.

- root user: *innovator* [1]
- initial threshold for cascade size: α
- upper threshold for cascade size: β

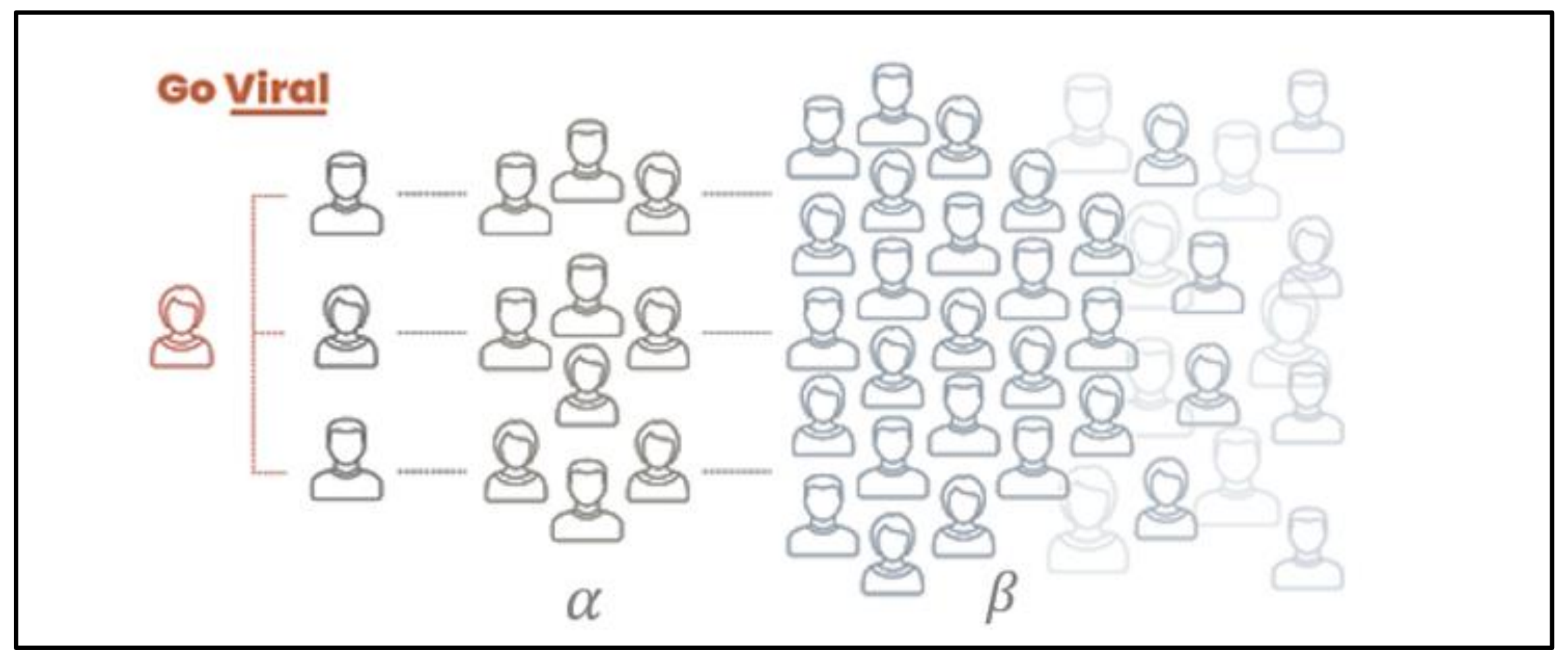

Fig 3. Viral Cascade Growth

Social Network Analysis

To extract social influence features from our cascades, social network analysis is applied to study information dissemination 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 (Figure 4). As more connections are made between any two nodes, the edge weight in 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.

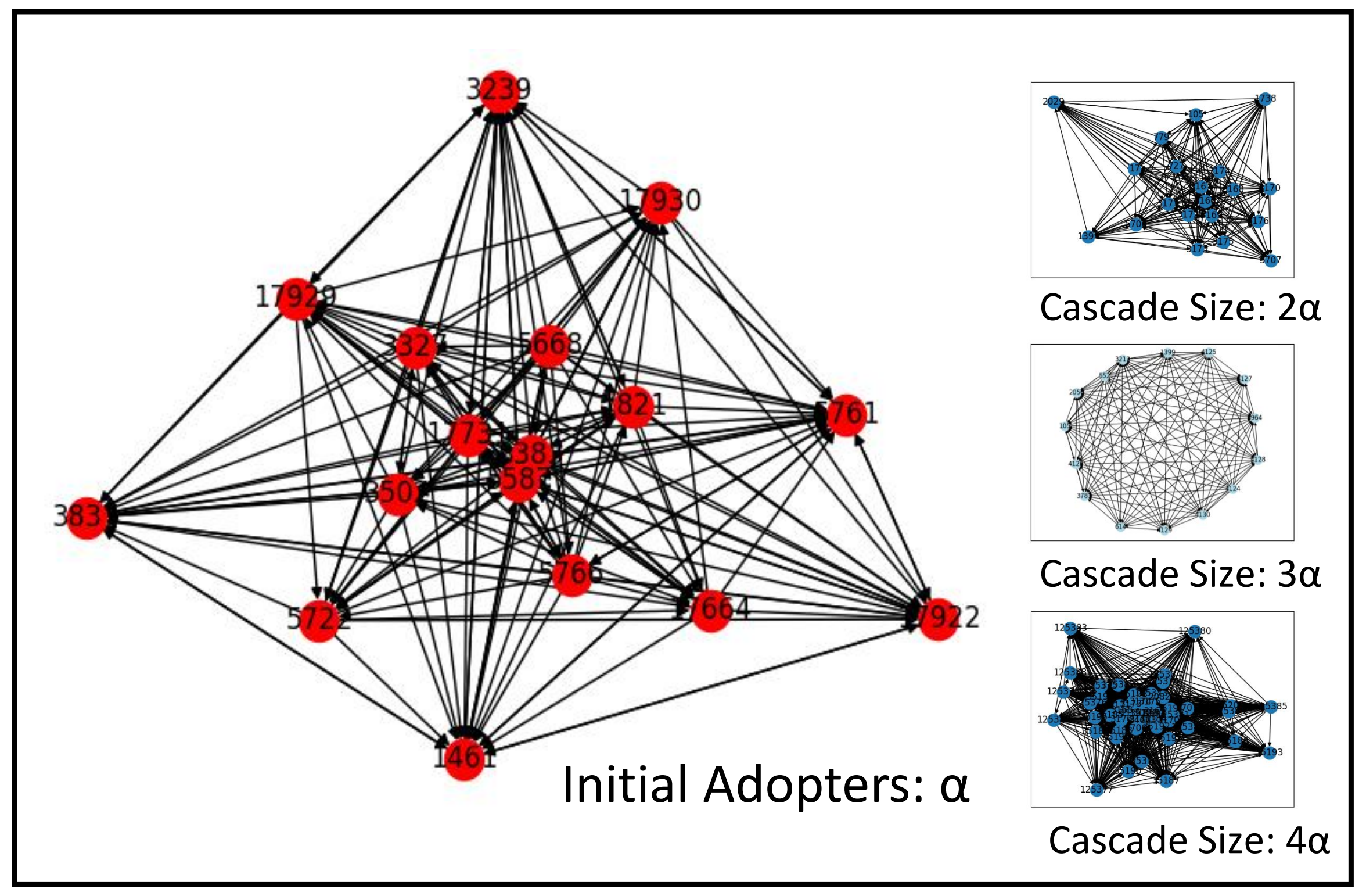

Fig 4. Phases of Viral Cascade Growth in NetworkX graph

Feature Extraction

The generated network graphs can be leveraged to extract two suites of features: root user features and early cascade features, as shown in Table 1.

Table 1. Feature Extraction

Target	Features Extracted
Root User	• degree centrality • out_degree centrality • out-edge eigenvector centrality • cumulative 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

Preliminary Results

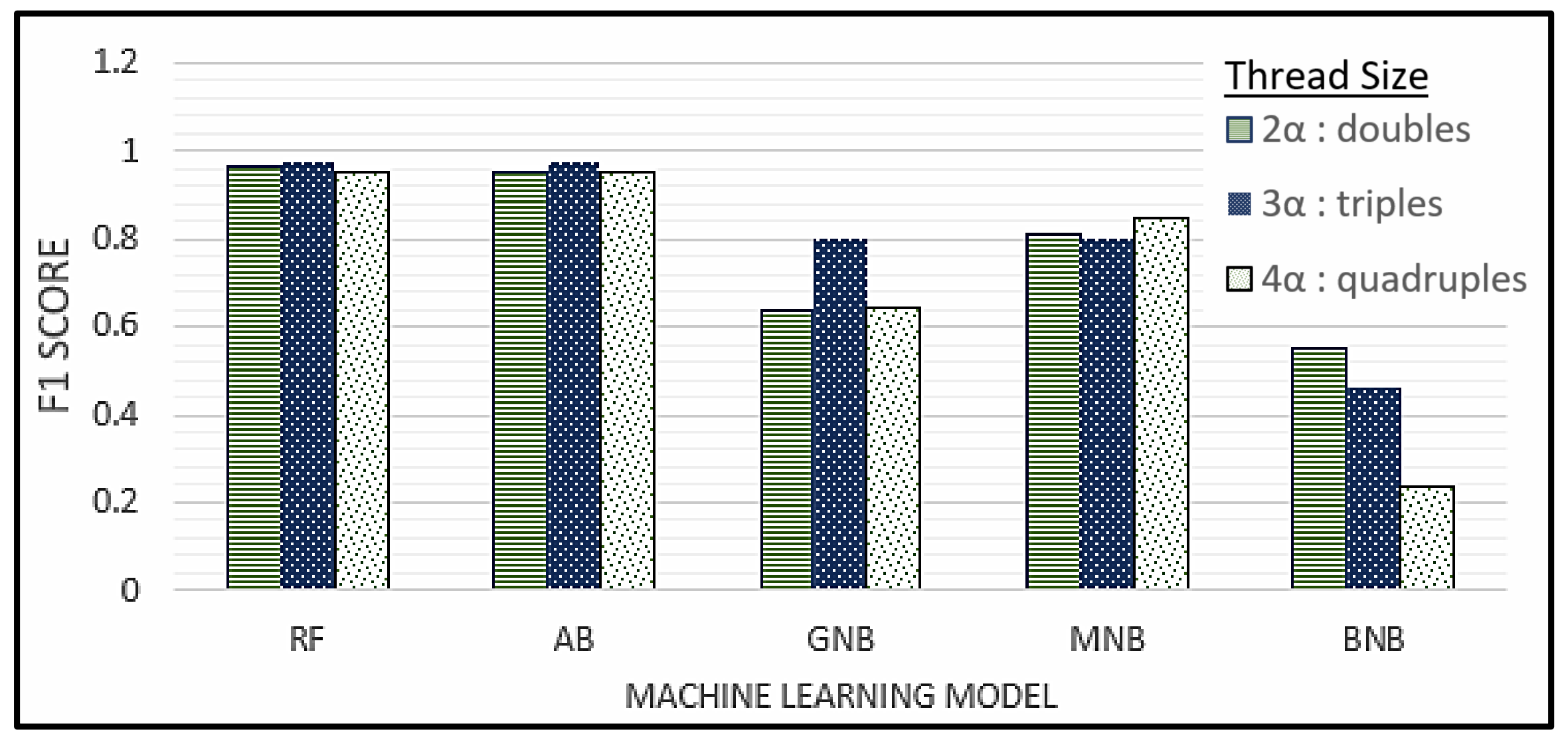

Fig 5. F1 scores when α = 30 and varying β cascade growth

Various classification models were trained and tested on the balanced dataset, resulting in the F1 scores displayed in Figure 5. 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.

References

[1] 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.

[2] 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.