Using Sampling and Cross Platform Entropy for Predicting Social Dynamics

Neda Hajiakhoond Bidoki, Gita Sukthankar
University of Central Florida

Abstract

Social media users exhibit repetitive behavior patterns that can be leveraged to predict trends in network evolution. Additionally, different social platforms have many users in common and are influenced by common events occurring in the real world. As such, it is likely that they exhibit similar trends. In this work, we use sampling, time series analysis and deep learning techniques to capture the internal and external correlations and predict future network features.

Introduction

Computational Simulation of Online Social Behavior (SocialSim) is a DARPA sponsored challenge on accurately simulating multitple social media platforms. Our team at University of Central Florida has created a high-fidelity agent-based simulation of user activity [1]. We are building a comprehensive, realistic and at-scale computational simulation of information spread and evolution in online social networks using a Deep Agent Framework (DAF) [1] to address the SocialSim problem. DAF combines multiple models into a unified architecture.

Contribution: Our research is on integrating insights from data mining into DAF to simulate the dynamics of cross-platform information propagation and network evolution at multiple levels of granularity: user, community, and population [2].

Data

Our datasets consist of GitHub, Reddit and Twitter events from three different domains:

- **Software Vulnerability:** the Common Vulnerabilities and Exposures (CVE) system provides a reference method for publicly known information-security vulnerabilities and exposures. Data in this domain was collected by pattern matching against activities related to 2600 CVEs.
- **Cryptocurrency:** data in this domain consists of all activities related to cryptocurrency coins.
- **Cybersecurity:** includes activities related to cyber security starting with seeds such as cybersecurity and netsec subreddits; all the other related subreddits were identified using Jaccard similarity based on user comments over that subreddit.

Community based Sampling and Replay

In this work we simulate network dynamics based on forecasts of community features. The methodology consists of the following steps:

1. Generate the network sequence from training data
2. Measure the communities' features over time
3. Generate the time series data for each community
4. Characterize the future network using time series forecasting model in terms of:
   (a) Communities
   (b) Network growth and cold start nodes
5. Sample each community from the historical network with the closest characteristics to the future network with respect to that community
6. Replay the sampled communities
7. Generate cold start nodes
8. Replay recently active users activities with regard to the gap between predicted network and the generated network in the previous step
9. Simulate cold start nodes' activities and connect them with respect to:
   (a) Cold start nodes degree, event type, and occurrence time distribution from historical data
   (b) Apply preferential attachment to generate links

Simulating Network Dynamics based on Forecasts of Community Features

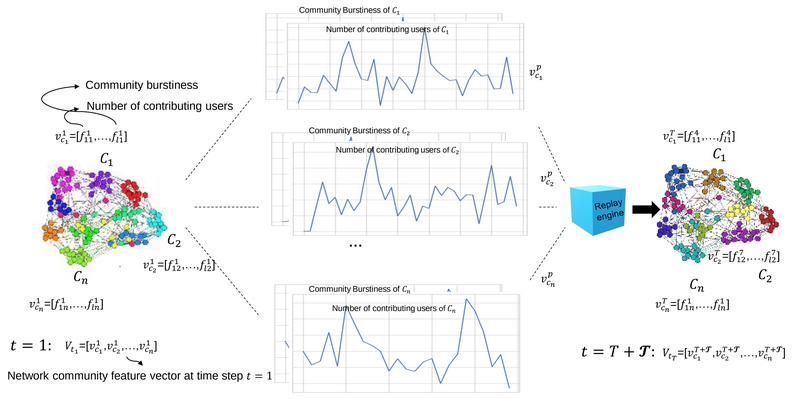

Figure 1: Community feature based prediction model framework

Cross Social Media Platform Prediction

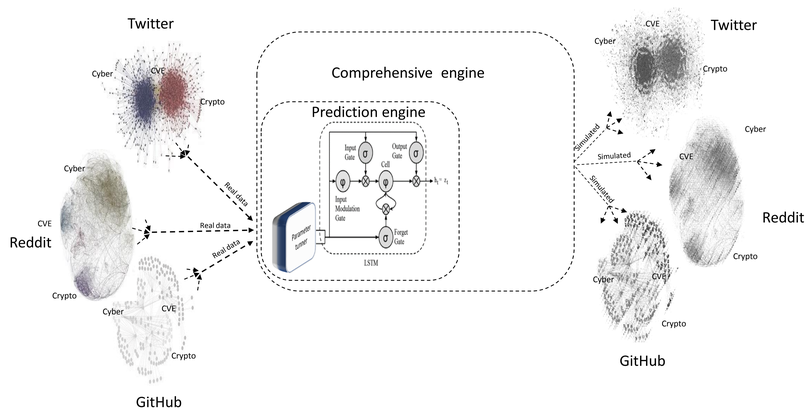

Figure 2: Cross social media platform measurement prediction used for social network platforms simulation

The cross social media platform prediction model works by augmenting each platform stream with information from other platforms:

- It is likely that social platforms show correlations at different resolutions because:
  - Different social platforms have many common users.
  - Information spreads across multiple platforms.
- Our model allows us to capture the correlations and dependencies across multiple channels.
- It can be applied at different resolutions, including individual, community, or population

Results

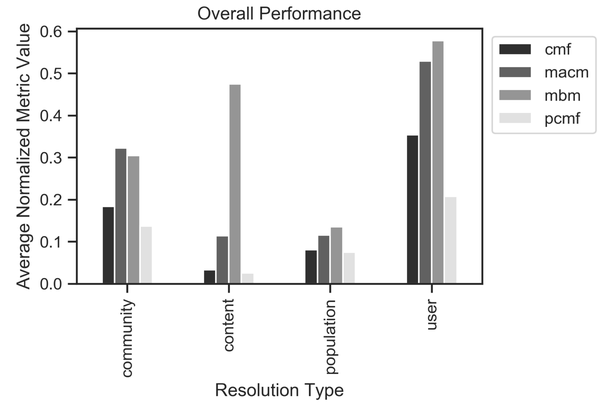

Figure 3: We compare our community feature based method's performance with two state of the art agent-based models at predicting population, user and content activity over three different data sets.

Conclusion

- This work introduces two approaches to predict and simulate the dynamics of social platforms: 1) simulating by sampling data based on forecasts of community features 2) a sequence learning model to predict the network measurements.
- We present results on incorporating community features into a replay model which outperform two agent base model at simulating the GitHub network.

References

[1] Ivan Garibay et al. "Deep Agent: Computational Social Science Centered Simulation of Online Information Environments". In: *Proceedings of International Conference on Computational Social Science*. 2019.

[2] Neda Hajiakhoond Bidoki et al. "Predicting Social Network Evolution from Community Data Partitions". In: *International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation*. Springer. 2019, To appear.

Acknowledgment

This work was partially supported by grant FA8650-18-C-7823 from the Defense Advanced Research Projects Agency (DARPA). The views and opinions contained in this work are the authors and should not be construed as official or as reflecting the views of the University of Central Florida, DARPA, or the U.S. Department of Defense.

Contact Information

- Email: Nedahaji@cs.ucf.edu