Using Sampling and Cross Platform Entropy for Predicting Social Dynamics*

Neda Hajiakhoond Bidoki and Gita Sukthankar

Department of Computer Science
University of Central Florida, Orlando, FL
nedahaji|gitars@cs.ucf.edu

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 the future features of the networks.

Keywords: social network analysis · time series · deep learning

1 Introduction and Related Work

Social networks have become the mirror of what is happening in the real world, making them a valuable tool for the scientific study of human behavior. An effective prediction approach for tracking the dynamics of social dynamics is critical for a variety of social media tasks including advertising, comment moderation, and bot detection. The proposed project focuses on predicting social dynamics using historical patterns, data from related social media platforms, and their cross entropy. The previous works employ graph theory and algebraic topology to guide agent-based models [1–3] to simulate the future of social platforms.

2 Current Work

My doctoral research will use time series analysis in combination with deep learning methods to forecast the future features of a social network extracted from a social media platform.

2.1 Data Description

We consider three large online social networks that are commonly employed as communication channels: Twitter, Reddit and GitHub. Our data set covers

* Supported by grant FA8650-18-C-7823 from the Defense Advanced Research Projects Agency (DARPA).

years 2015-2017 and is partitioned into three different domains: CVE, Crypto and Cyber. CVE contains the networks of people who work in the realm of software vulnerabilities. Crypto data represents the activities associated with the content and users that communicate primarily about crypto currencies, and the Cyber data set is related to activities within the realm of cyber-threats.

2.2 Approach

My approach is to use sampled historical data to forecast the future of three social media platforms: GitHub, Reddit and Twitter. Also, inspired by the recent success of deep neural networks in a wide range of computing applications along side with the availability of social network information, we propose a sequence learning model to predict the network measurements using parallel series of information from multiple channels. Long Short Term Memory (LSTM) is one the most advanced recurrent neural networks that uses a gating mechanism to store the relevant information; it is also well suited to handle multidimensional data. When investigating internal cross entropy, channels are streamed from nodes within a single platform and when investigating cross platform entropy, streams are drawn from different platforms.

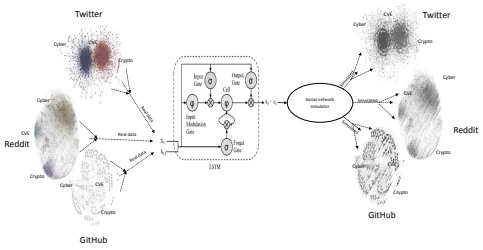

Fig. 1: Cross platform prediction model

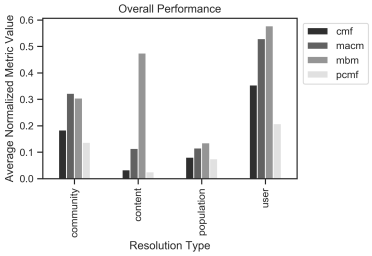

Fig. 2: Overall performance for each model at each resolution.

2.3 Novelty and Significance

Due to the dearth of data, there has been less research on cross platform propagation trends. The proposed deep learning based approach is able to capture the cross entropy between event streams within a platform as well as across platforms regardless of the cause. Our replay approach can be employed as an augmentation technique for other models while significantly reducing their time and space complexity [4].

2.4 Results

Prediction With Respect to a Single Platform An important component of our approach is investigating the advantage of directly "replaying" the past

data to simulate the future. We showed that this approach accurately captures aspects of the data distributions at different granularity levels than a more general model. We are investigating the usage of the different types of features, both temporal and community level to find the best data portions to replay. Figure 2 shows our approaches (named as PCFM and CFM respectively) as compared to other benchmarks [5, 6]. Additionally, we use our deep learning based model to capture cross nodes correlations within a platform [7].

Prediction With Respect to Multiple Platforms Researchers characterize networked structures in terms of individual users, communities, and the content within the network as well as the relationships or interactions that they have as their links. We use this cross network/domain correlation within three social networks. We evaluate our model with regard to its propagation of an information unit such as the spread of information about Common Vulnerabilities and Exposures [8].

References

1. M. Morris, *Local rules and global properties: Modeling the emergence of network structure.* na, 2003.
2. J. Wang, C. Jiang, S. Guan, L. Xu, and Y. Ren, "Big data driven similarity based u-model for online social networks," in *GLOBECOM 2017-2017 IEEE Global Communications Conference.* IEEE, 2017, pp. 1–6.
3. J. Qiu, J. Tang, H. Ma, Y. Dong, K. Wang, and J. Tang, "Deepinf: Social influence prediction with deep learning," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* ACM, 2018, pp. 2110–2119.
4. I. Garibay, M. Schiappa, G. C. Anagnostopoulos, C. Bouwens, S. Fiore, H. Keathley, A. Mantzaris, J. T. Murphy, W. Rand, A. Salter, M. Stanfill, G. Sukthankar, N. Baral, A. Berea, C. Gunaratne, N. B. Hajiakhoond, G. Fair, J. Jasser, C. Jayalath, E. Mutlu, O. Newton, A. Rajabi, S. Saadat, S. R. Scheinert, C. Senevirathna, and X. Winter, Rachel Zhang, "Deep agent: Computational social science centered-simulation of online information environments," 2019.
5. N. Hajiakhoond Bidoki, M. Schiappa, G. Sukthankar, and I. Garibay, "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, p. To appear.
6. ——, "Simulating network dynamics by sampling data based on forecasts of community features," in *International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation.* Springer, 2019, p. To appear.
7. N. Hajiakhoond Bidoki, G. Sukthankar, K. Heather, and I. Garibay, "A cross-repository model for predicting popularity in github," in *IEEE CPS in the proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI'18).* IEEE, 2018.
8. N. Hajiakhoond Bidoki and G. Sukthankar, "Deepburst: Cross network burst prediction usingrecurrent neural networks," in *International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation.* IEEE Press, 2019, p. To appear.