A Theory-Driven Framework for Modeling Temporal Online Social Networks of GitHub

Completed Research

[1] Wingyan Chung, [2] Elizabeth Mustaine, [3] Jinwei Liu, [4] Mehul Vora
[1, 3, 4] School of Modeling, Simulation, and Training, University of Central Florida
[2] Department of Sociology, University of Central Florida
[1] Department of Computer Science, The University of Hong Kong
[1] Department of Decision Sciences and Managerial Econ., The Chinese University of Hong Kong
[1,2] {wchung, libby}@ucf.edu, [3,4] {jliu, mvora}@ist.ucf.edu

Abstract

Online social networks (OSNs) are becoming increasingly prevalent in human life. However, the factors that influence human behavior in OSNs are not well understood. Prior research assumes nodes in OSNs to be static and unchanged over time. Unfortunately, existing methods do not sufficiently consider important phenomena in collective social behavior. This research developed a new theory-driven framework for modeling temporal OSNs. Three models were built based on the framework to represent the effects that may possibly shape temporal online social networks. The random network model (RNM) represents randomness in the interaction among networks nodes and serves as a baseline model. Two window-based models, namely Average Aggregation Model (AAM) and Exponential Aggregation Model (EAM), represent respectively the herd effect, and the recency and primacy effects in human social behavior. To evaluate the models, we examined 42 OSNs of the GitHub software development community that committed a total of 5,499,611 events during Jan. 25-30, 2017. The results show that both EAM and AAM achieved superior performance compared with RNM. The research makes three contributions: (1) developing a new theory-driven framework for characterizing online social behavior, (2) developing and validating three models for simulating temporal OSNs using social and cognitive theories, and (3) providing empirical findings of simulating GitHub OSNs and user behavior.

Keywords

Online social network, modeling, simulation, GitHub, cognitive theory, social theory, random network model, average aggregation model, exponential aggregation model.

Introduction

Online social networks (OSNs) are increasingly prevalent in human life. Many social, political, and economic activities are facilitated by OSNs, such as electronic commerce, political mobilization, and collaborative software development. In addition to their popularity, these networks influence human behavior and decisions. Research shows that product purchase, election voting, and dietary preferences are affected by social media and online social networks (Bond et al. 2012 ; Liang et al. 2015; Zeng 2015). However, the factors that influence human behavior in OSNs are not well understood.

Prior studies of dynamic social networks assume nodes to be static and their behavior unchanged over time. However, real-world OSNs seldom follow this assumption (Marble et al. 2015). In contrast with static OSNs, temporal OSNs have nodes and links that change over time. Unfortunately, existing models lack capability to accurately predict these temporal OSN changes. Extant theories, though useful in traditional social settings, are not widely used in modeling online social networks (Li et al. 2017b).

This research developed a new theory-driven framework for modeling temporal OSNs to address the challenges. The framework incorporates social and behavioral theories to represent collective behavior, nodal interaction, cognitive biases, temporal preferences, and randomness, and allows for flexible specification of reference history and for learning from past data. Based on the framework, we developed three models (RNM, AAM, and EAM) of OSN simulation using randomness, herd effect, and recency and primacy effects on temporal OSNs. To evaluate the models, we examined 42 OSNs of the GitHub software development community that committed a total of 5,499,611 events during Jan. 25-30, 2017. The evaluation compares these models' simulated networks with actual networks extracted from the GitHub data. Results are very encouraging, offering new insights to understanding online social behavior over time. The research makes three contributions: (1) developing a new theory-driven framework for characterizing online social behavior, (2) developing and validating three models for simulating temporal OSNs using social and cognitive theories, and (3) providing empirical findings of using the models to simulate a large online social community of importance to business and software industries.

Literature Review

Temporal networks have been studied in various disciplines, including physics, mathematics, computer science, information systems, statistics, sociology, and economics, among others. The prevalence of online social networks further increases interest in this area. Various theories and methods have been developed to support understanding and prediction of collective behavior, network dynamics, nodal behavior, and link structure. The following provides a brief review of relevant background in these theories and methods.

Theoretical Background

Social Contagion Theory and Emergent Norm Theory are well known historically and still relevant currently (Reicher 2000). Both explain the ways in which the social-psychology of a crowd differs from that of individuals. Social contagion theory asserts that people normally base their behavior on the information available to them (e.g., rational thought, experience). But in a mass crowd, information is often contradictory due to a lack of agreement (Le Bon 1895). Emergent norm theory further posits that new norms happen when group leaders and members agree on a new normative status or purpose for the group (Turner and Killian 1957). One central difference between the two theories is time. Generally, mass contagion is more spontaneous and rapid than the emergence of new social norms (Quan-Haase 2016). Individuals may take time to observe the emerging norms and dynamics of a mass group, and gradually identify with the group mentality (Gino et al. 2009) and purpose (Reicher 2000).

The question of whether or not the above theories can be applied to communication and information spread in the online world has not yet been validated. However, several important differences between online communication and face-to-face communication suggest that the theories need to be empirically examined and may need modification. These differences include the quality of interaction, the speed and geographic spread of messages (Li et al. 2017a; Quan-Haase 2016). Consequently, online communication fundamentally changes the human perception of time. Elements of time that are relevant to the spread of information throughout mass groups include recency and primacy effects (Hovland 1957; Miller and Campbell 1959). Research has shown that people are more likely to join mass groups and share information when they are exposed to that information recently (recency) or when it reaches them early (primacy) (Gino et al. 2009; Ngai et al. 2015). However, research into these effects on temporal OSNs is scarce.

Methods on Temporal Social Network Modeling

Prior studies in temporal networks have examined specific properties of online social networks (Holme and Saramäki 2012; Li et al. 2017b), such as large number of participants who commit actions over short time periods (Mislove et al. 2007), global structure of online friendship network (Ugander et al. 2011), and expression of user sentiment and emotion (Lim et al. 2018). Researchers also have developed methods to automatically learn models of user behavior in temporal OSNs. Examples include modeling temporal dynamics of social ratings (Jamali et al. 2011), using a supervised-learning-based approach to link prediction (Scellato et al. 2011), inferring new links among users given a snapshot of a social network (Liben-Nowell and Kleinberg 2007), detecting OSN community based on "proximity" of nodes in a network (Lazarsfeld and Merton 1954), and joint inference for user links and attributes by leveraging the data redundancy and mutual reinforcement (Yang et al. 2017). These methods use machine learning to acquire knowledge of user behavior and then make prediction based on the learned models. On the other hand, graph-based methods characterize link structure of entities within networks. Examples include quantifying user interaction in social links (Wilson et al. 2009), building a stochastic topic model for link prediction (Barbieri et al. 2014), examining time-evolving properties of OSN subgraphs (Ardon et al. 2013), stochastic

Markov process models to represent influences on network change (Snijders et al. 2010), and extending the Exponential Random Graph Models for statistical modeling (Hanneke et al. 2010).

The above methods assume network nodes to be static and unchanged over time. However, real-world online social networks seldom follow this assumption (Wang et al. 2015). In contrast with static networks, temporal OSNs have nodes and links that change over time. Furthermore, the aforementioned studies do not sufficiently consider important social phenomena, such as herd effect, and recency and primacy effects in temporal OSNs (Marlow et al. 2013; Pelled et al. 2017; Yu et al. 2014), which have important implication of users' social and cognitive processes of decision making in OSNs.

A Theory-Driven Framework for Modeling Temporal OSNs

Temporal OSNs have different fundamental properties than static networks (Holme and Saramäki 2012), and have a much higher complexity than traditional social networks (Liao et al. 2017). Unlike traditional social networks, temporal OSNs have larger numbers of nodes and links (e.g., friends on Facebook, connections on LinkedIn, etc.) and exhibit stronger dynamics, creating more complex network structure and more frequent user interaction. Unfortunately, existing theories and methods may not adequately represent these characteristics.

In this research, we are interested in studying the factors that change online social networks over time. Understanding these factors could help to accurately simulate future networks and to predict behavior in these networks. The factors may be useful for studying human social behavior in different online settings, such as e-commerce, online political participation, and collaborative software development.

Based on design science research paradigm and methods in social-media-based informatics (Chung and Zeng 2018; Hevner et al. 2004), this research develops a new framework for modeling temporal OSNs, provides instantiations of models to simulate temporal OSN activities, and evaluate their performance in accurately and efficiently simulating online social behavior. Our research questions are (1) What are the factors that change temporal online social networks? (2) How should these factors be incorporated in the development of models to simulate the networks? (3) How do these models perform in accurately simulating real-world online social community?

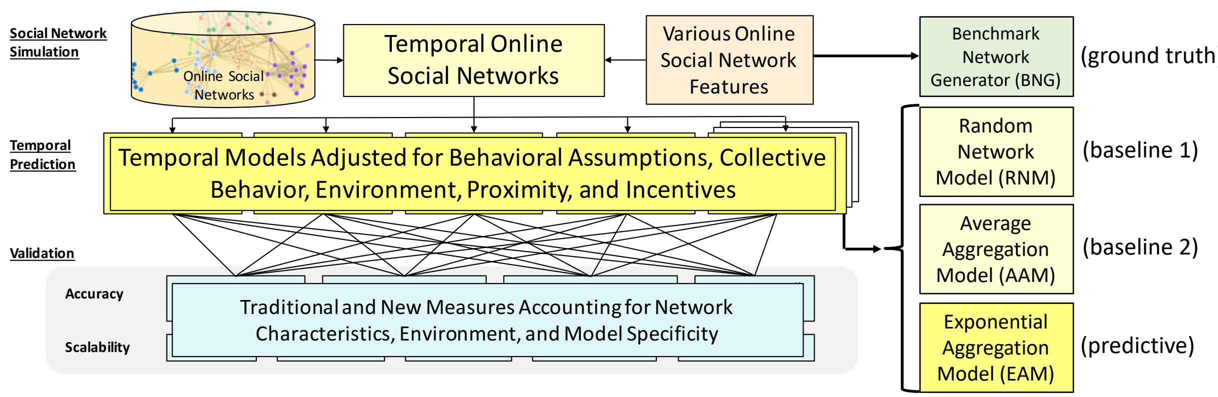

Figure 1. A Framework for Modeling Temporal Online Social Networks

The framework (see Figure 1) transforms raw data to temporal social networks, extracts various features from these networks, and then provides predictive models to project future networks based on behavioral assumptions, environment, proximity, and user preferences. Notations in graph theory are used to describe the framework's components (Erdős and Rényi 1959; Holme and Saramäki 2012). Let $G_t^{L(w)}(k)$ be an undirected bipartite network of link type *k* at time *t*. The bipartite network consists of two types of nodes such that a link must be between two nodes of different types. Also, let *L(w)* be the function that specifies the network to span the most recent *w* time periods at and before time *t*. A link belongs to a type *k* to indicate the type of relationship between the two nodes being connected. In a collaborative social network, there are usually two types of nodes that form different types of links. For example, a recommender system may have two types of nodes: user and product, which form three types of links: "viewed," "commented," and "purchased." As another example, a collaborative writing network may have two types of nodes: user and document, which form three types of links: "edit," "comment," and "create." From these examples, a general formulation of a temporal collaborative social network at time *t* for link type *k* over a time span of w can be defined as:

$$G_t^{L(w)}(k) = \{V_{u,t}^{L(w)}, V_{r,t}^{L(w)}, E_{k,t}^{L(w)}\}$$

Here, the three components of the network are the sets of type-1 nodes, type-2 nodes, and the set of links of type *k* respectively. The number of bipartite networks at time *t* is equal to the total number of link types.

Model Design and Implementation

Based on the framework, we developed three models to compare against the benchmark networks that were generated from the data (to be described below). These models try to capture randomness, herd behavior, and recency and primacy effects in temporal online social networks. Below we describe the characterization of these models in the context of the GitHub collaborative software network. GitHub (https://github.com/) is an online collaborative software development platform that allows users to share and edit software repositories. A repository (*repo* for short) is an independent project where users store code, documentation, resource files, and references. An owner of a repository can invite other GitHub users to be collaborators to contribute to the repo development. A user can download a repository to their local computer to edit, upload their changes, post comments, and report problems about the repository. Similar to other OSNs (e.g., Facebook, Twitter), GitHub allows users to interact directly with each other by contributing to repos, and to interact indirectly by following other users or by watching specific repos. Compared with Facebook and Twitter, GitHub has a significantly tighter social relationship among users, a longer average path length, and a higher proportion of followers who would eventually become contributors (Jiang et al. 2013).

The GitHub Online Social Network

The GitHub social network consists of two types of nodes: users and software repositories, and seven types of links: events between a user and a repository. These event types are "watch," "fork," "pull," "push," "issue," "create," and "delete." A "watch" event occurs when a user clicks the "Watch" button on GitHub to watch a repository. A "fork" event occurs when a user makes a copy (or a branch) of the repository. A "pull" event occurs when a user downloads from GitHub.com the latest updates of the repository to their local computer. A "push" event occurs when a user uploads his/her latest changes to the repository's online storage on GitHub.com. An "issue" event occurs when a user posts a comment on the repository about a repository, usually to signal some bugs or problems that need to be addressed. A "create" event occurs when a repository is created by a user. A "delete" event occurs when a repository is deleted by a user.

Based on the structure of GitHub data, we formulated the social network as a hypergraph, which consists of seven types of bipartite social networks (one for each event type). A bipartite network consists of two types of nodes: user and repository. We formulated (in each time period) seven benchmark networks that describe the seven types of events occurring between users and repos. A time period *s* is the length of time during which a network is formed from the data (*s* is set to be 1 calendar day in the study described below). We develop three models to capture the characteristics of GitHub OSNs, namely, a tight social relationship, long path length, and high contribution from users. While applied only to GitHub in this study, the models provide generic capabilities to represent various effects and features of any OSNs, as described below.

Temporal OSN Models

Random network model (RNM) characterizes the randomness of online social behavior by specifying a link probability based on prior knowledge. A simple way to set this probability is to use a uniform low probability (e.g., 0.01) to indicate a low likelihood of linkage between nodes. The probability can be learned from prior knowledge or provided by experts. A key characteristic of this model is that this probability is set in advance of modeling, and stays constant throughout the time being considered for prediction.

Average aggregation model (AAM) represents the herd effect in collective behavior where humans observe the crowd and conform to norm by following the average trend. AAM uses a window w, defined as a multiple of *s*, to learn from the recent history of data for predicting the network at *t+1* (w = 24 calendar days in our study). The model uses an averaging strategy to learn from the most recent history of w networks to predict events in the next network at $t+1$.

<u>*Exponential aggregation model (EAM)*</u> represents recency and primacy effects of human cognition to relate their behavior to time. A transmission parameter, α_k, which ranges from 0 to 1, is set for the network of link type k in a single exponential smoothing function to model the effects. If α_k is close to 1, then recent information is emphasized most in the computation of emergence probability (recency effect). If α_k is close to 0, then early information is emphasized most in computation of emergence probability (primacy effect). Similar to AAM, EAM uses a window w to learn from recent history of data. Table 1 provides pseudo code to explain the steps in AAM and EAM.

<u>Input</u>

- Node: repos: V(r) and users: V(u)
- Link: events (u-r) of type *k*: E(*k*)
- Each bipartite network of type *k* contains nodes and links
  - G_t (k) = { V(u, t), V(r, t), E(k, t) }
- Bipartite networks $G_{tc-(w+1)*s}(k)$, $G_{tc-(w+2)*s}(k)$, ..., $G_{tc}(k)$ are formed in the reference history w.
  - t_c represents the current time.

<u>Processing</u>

- For each event type *k*, compute probability of emergence.
- Define P_k to be the emergence probability that event k occurs over window w.
- Predict event k at t_{c+s} by considering emergence probability P_k during reference history w, where A_{ijkt} represents recent activities (a boolean value indicating presence (=1) or absence (=0) of event *k* between user *i* and repo *j* at time *t*).
- In AAM: $P_k = (\sum_{\{tc-(w+1)*s,\ tc\}} A_{ijkt}) / w \quad \forall\, i, j$
  - Averaging allows for modeling herd effect in collective behavior
- In EAM: $P_k = (\sum_{\{tc-(w+1)*s,\ tc\}} b_t A_{ijkt}) \quad \forall\, i, j$
  - b_t varies with time according to a simple exponential smoothing function
  - Allows for assigning recency and primacy effects of human cognition
- For one-time events (watch, fork, create, delete), the models compute the emergence probability over all users; For multi-time events (pull, push, issue), the models compute the emergence probability over the specific user and repo.
- Moving the window forward allows simulation of subsequent event networks.

<u>Output</u>: New network, $G_{tc+s}(k)$, for event type *k*.

Table 1. A Window-based Algorithm for the AAM and EAM

Empirical Study and Findings

This section describes an empirical study of the GitHub temporal social networks and reports on the results. The data consist of all GitHub events from January 1st to 30th 2017. The raw data was in the form of hourly activity data on GitHub portal. From these data, we extracted user-id, repository-id and event-type with timestamps; these data were aggregated per day to create daily networks of each event type (links are events; nodes are repos and users). The process was automated by a Benchmark Network Generator (BNG) that we developed to produce actual OSNs that were then compared with OSNs simulated by the three proposed models. It should be noted that BNG is not a predictive model; but rather a process to extract from raw data the OSNs of the actual events. Thus, BNG produces ground truth for use in comparison.

For performance evaluation, two categories of metrics were used. Network-level metrics include average node degree, total node degree, network density, edge count, user node count, and repo node count. Scalability was measured by a model's running time and by the change in predictive accuracy over time. These metrics support objective evaluation in the following three comparisons of model performance.

Comparison 1: BNG vs. RNM

The first step of the analysis was to compare BNG against the Random Network Model (RNM) to examine the effect of randomness in simulating temporal OSNs. Figure 2 shows the comparison of the performance on network-level metrics between the ground-truth (produced by BNG) derived from the data spanning one hour (9/25/2017) and the randomly generated networks for each of the selected seven event types (watch, fork, push, pull, issues, create, delete). The figure shows that RNM generated the user-repository events with a uniform probabilistic distribution. The results show that RNM captured the randomness in activities but did not recognize behavioral disparities in different event-types and did not incorporate richer features of networks.

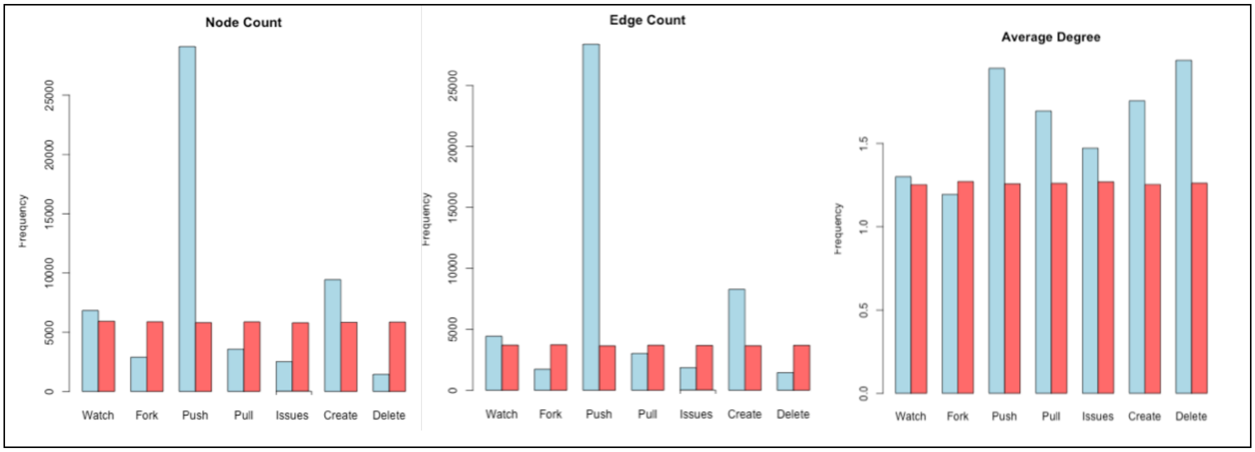

Figure 2. Comparing between Benchmark (blue) and Predictions by RNM (red)

Comparison 2: BNG vs. AAM vs. EAM

The next step was to examine herd effect and recency and primacy, used in AAM and EAM respectively, in predicting future OSNs. We evaluated the performance of AAM and EAM and compared their results against those of BNG. We chose a training period of 24 days (w=24) and testing time period of 1 day (s=1 day). For example, using data from January 1 to January 24, 2017, a model would predict the network events of January 25. This prediction was repeated for 6 days (N=6) by sliding the window w forward for each day, generating the simulated networks for January 25 to January 30, 2017. For EAM, based on event type, the transmission parameter was set between 0.5 and 0.9, assigning moderate to high recency effects to the predictions as suggested by prior research (Bampis et al. 2017; Wei and Carley 2015). Figure 3 represents the comparison between BNG, AAM and EAM using the training data of the selected 7 event-types from January 6 to 29 for the prediction for each event types on January 30; whereas Figure 4 represents the comparison of AAM and EAM with BNG results for “push” events spanning over 6 days.

Figures 3 and 4 show that EAM generally performed better than AAM across all network measures (node counts, degree and density) for multi-time events (push, pull, and issue). In addition, EAM was able to generate the relative differences in activities of different network types. For instance, the EAM-simulated counts of user nodes, repo nodes, and edges and average degrees preserve the ordering of the actual counts and actual average degrees. In case of single-time events (Create, Delete, Fork and Watch), the EAM-simulated metrics values deviate further from those of the benchmark, whereas the relatively flat AAM-simulated metrics values are closer to those of the benchmark. This could be due to values of transmission parameters for those events and the relative lower predictability of these events. The results suggest that recency and primacy effects are generally more accurate in predicting multi-time events, whereas herding effect is generally more accurate in predicting to single-time events.

Comparison 3: Scalability of Implementation – BNG vs. AAM vs. EAM

Below, we present the system performance comparison of AAM, EAM and BNG. Following two tables indicate the computational complexity of each model. Table 2a lists the number of events generated by each model and Table 2b lists the cumulative execution time for each model to compute 42 networks (1 network per event-type per day: 7 event-types and 6 days). In Table 2a, we see that the total number of events generated follows the pattern: BNG > EAM > AAM. In Table 2b, we see that the execution time of AAM, EAM and BNG follows the pattern: EAM > AAM > BNG. This is because BNG does not actually simulate the network, and AAM considers historical data by simply averaging the historical data values; however, EAM requires to iteratively calculate the exponential function, which is more time consuming.

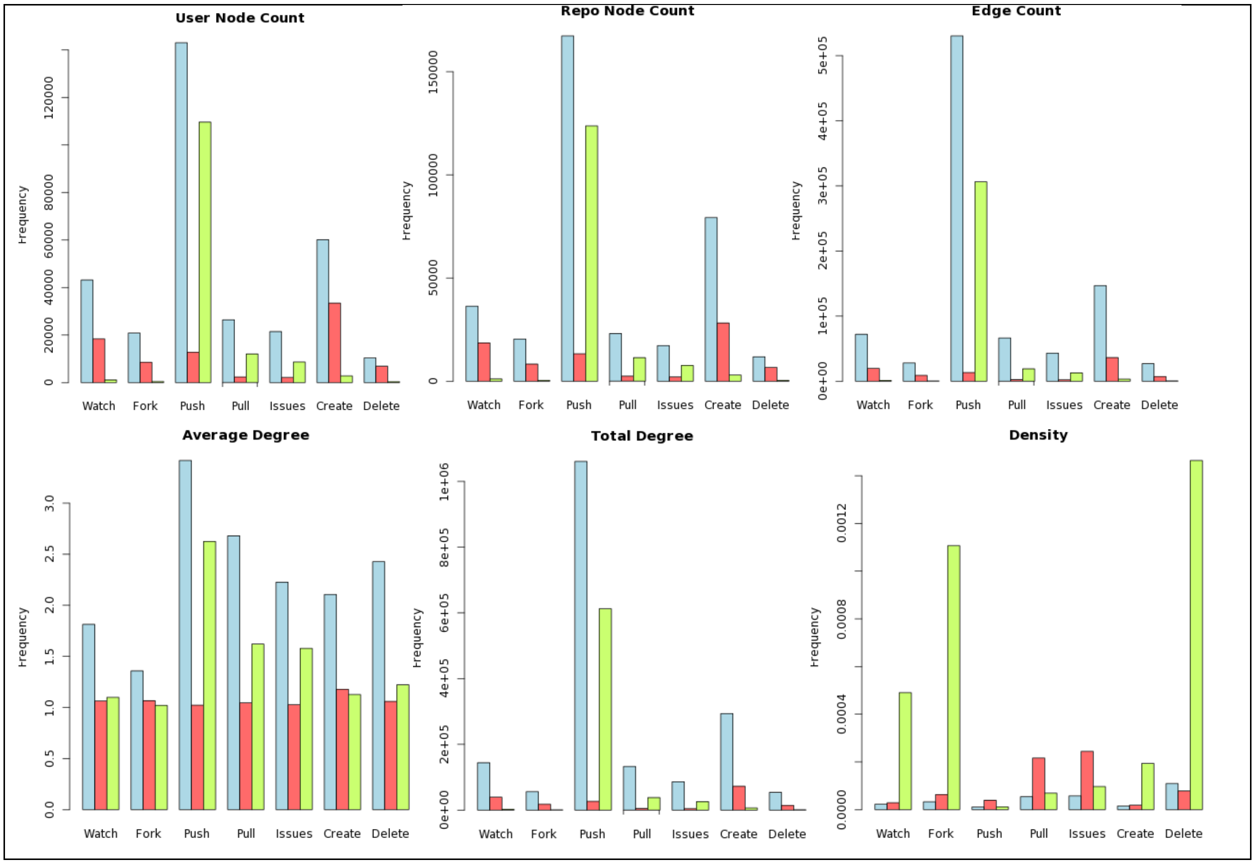

Figure 3. Comparison of benchmark (blue), AAM (red) and EAM (green) for 1/30/2017

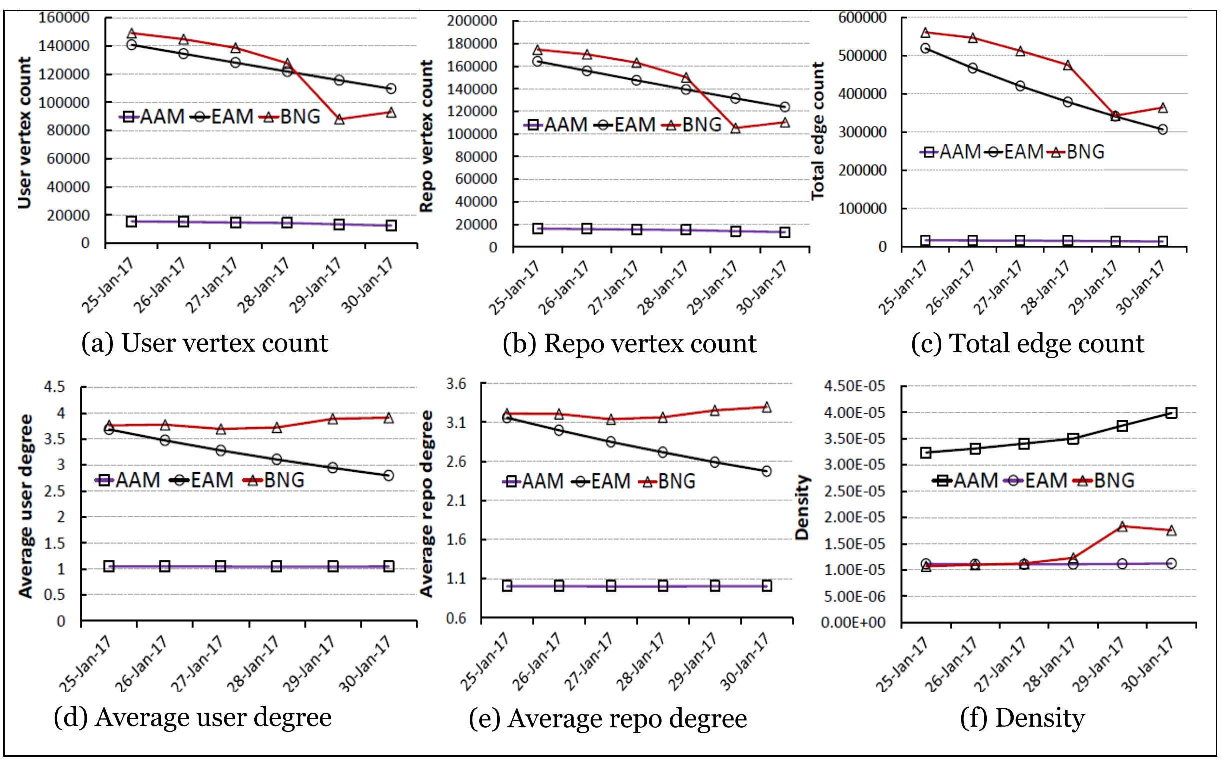

Figure 4. Comparison of AAM and EAM with benchmark for PUSH events

(a) Event Counts						
Model	Jan. 25	Jan. 26	Jan. 27	Jan. 28	Jan. 29	Jan. 30
AAM	89,194	89,222	89,129	89,087	89,422	89,365
EAM	789,257	631,523	525,820	450,022	391,417	343,772
BNG	1,116,735	1,096,130	1,023,371	943,596	647,561	672,218

(b) Efficiency	
Model	Running Time
AAM	31 min.
EAM	1 hr. 20 min.
BNG	64 sec.

Table 2. Comparison of Model Performance on Scalability

Implication

Several implications are observed from the results. *First*, EAM was found to achieve higher accuracy than AAM in predicting multi-time events, such as Push and Pull. This is possibly because EAM has the capability of modeling recency and primacy effects that serve an important role in human cognitive processing of repeatable events. By contrast, single-time events (such as Fork and Create) do not require repetition for the same repository, allowing AAM that models conforming behavior (herd effect) to achieve a higher accuracy. *Second*, the values of transmission parameter in EAM is critical as they dictate the use of historical information for predicting future behavior. Finding their optimal values requires further experimentations and evaluation. *Third*, AAM was found to perform well in simulating single-time events, possibly due to the relatively lower fluctuation in these activities (e.g., each repo is created only one time). However, the simple herd effect that AAM models may not be a dominant factor in shaping online social networks as collaborative behavior increases. Other factors may need to be considered in such situations. *Fourth*, the empirical studies cover a period of one month, with training phase being 24 days and testing phase being 6 days. This time span allows us to observe changing emergent norms in fast-evolving OSNs of GitHub. Different lengths of this period may allow different observation and learning of historical behavior in other online platforms (e.g., Twitter, Facebook).

Conclusions

Online social networks (OSNs) are becoming increasingly prevalent in human life. However, the factors that influence human behavior in OSNs are not well understood. Prior research assumes nodes in OSNs to be static and unchanged over time. Unfortunately, existing methods are limited in addressing the characteristics of temporal OSNs. This research developed a new theory-driven framework for modeling temporal OSNs. The framework considers different factors and allow flexible development of models to account for online social phenomena. Three models were built based on the framework to represent the effects that may possibly shape temporal online social networks. The random network model (RNM) represents randomness in the interaction among networks nodes and serves as a baseline model. Two window-based models, namely Average Aggregation Model (AAM) and Exponential Aggregation Model (EAM), represent respectively the herd effect, and the recency and primacy effects in human social behavior. They use a sliding window of most recent time frame to compute dynamic emergence probabilities of network links.

Results of the empirical study of the models using the GitHub software social network show that both EAM and AAM achieved superior performance compared with RNM. Compared with AAM, EAM achieved higher performance in predicting contribution events, generated closer distributions of network metrics to actual networks, and was able to accurately simulate relative differences in these events. Future research may consider additional factors (e.g., user interaction) that may shape online social behavior. New metrics that can measure both network-level and node-level performance would provide new guidance on model design and development. Multi-resolution and multi-environment simulations of temporal OSNs would offer new insights to studying other online social behavior. The research makes three contributions: (1) developing a new theory-driven framework for characterizing online social behavior, (2) developing and validating three models for simulating temporal OSNs using social and cognitive theories, and (3) providing empirical findings of using the models to simulate a large online social community of importance to business and software industries. These contributions should be relevant to both academic researchers and industries affected by collaborative software development.

Acknowledgements

This research was supported in part by funding from U.S. Defense Advanced Research Project Agency (contract no.: FA8650-18-C-7824), University of Central Florida's Preeminent Postdoctoral-Scholar Program, and Fulbright Core US Scholar Fellowship (2017-19). We thank all project members, the editor and reviewers for their contributions.

REFERENCES

Ardon, S., Bagchi, A., Mahanti, A., Ruhela, A., Seth, A., Tripathy, R. M., and Triukose, S. 2013. "Spatio-Temporal and Events Based Analysis of Topic Popularity in Twitter," *CIKM*.

Bampis, C. G., Li, Z., Moorthy, A. K., Katsavounidis, I., Aaron, A., and Bovik, A. C. 2017. "Study of Temporal Effects on Subjective Video Quality of Experience," *IEEE TRANSACTIONS ON IMAGE PROCESSING* (26:11), pp. 5217-5231.

Barbieri, N., Bonchi, F., and Manco, G. 2014. "Who to Follow and Why: Link Prediction with Explanations," *KDD*.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., and Settle, J. E. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature* (489:7415), pp. 295-298.

Chung, W., and Zeng, D. 2018. "Social-Media-Based Public Policy Informatics: Cyber-Surveillance for Homeland Security and Public Health Informatics," in *Policy Analytics, Modelling, and Informatics,* G.-G. J., P. T. and L.-R. L. (eds.). Springer, Cham.

Erdős, P., and Rényi, A. 1959. "On Random Graphs," *Publicationes Mathematicae* (6), pp. 290-297.

Gino, F., Ayal, S., and Ariely, D. 2009. "Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel," *Psychological Science* (20:3), pp. 393-398.

Hanneke, S., Fu, W., and Xing, E. P. 2010. "Discrete Temporal Models of Social Networks," *Electronic Journal of Statistics* (4:0), pp. 585-605.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *Management Information Systems Quarterly* (28:1), pp. 75-105.

Holme, P., and Saramäki, J. 2012. "Temporal Networks," *Physics Reports* (519:3), pp. 97-125.

Hovland, C. I. 1957. *The Order of Presentation in Persuasion*. Yale University Press, Inc.

Jamali, M., Haffari, G., and Ester, M. 2011. "Modeling the Temporal Dynamics of Social Rating Networks Using Bidirectional Effects of Social Relations and Rating Patterns," *WWW*.

Jiang, J., Zhang, L., and Li, L. 2013. "Understanding Project Dissemination on a Social Coding Site," *2013 20th Working Conference on Reverse Engineering (Wcre)*), pp. 132-141.

Lazarsfeld, P. F., and Merton, R. K. 1954. "Friendship as a Social Process: A Substantive and Methodological Analysis," *Freedom and control in modern society* (18), pp. 18-66.

Le Bon, G. 1895. *The Crowd: A Study of the Popular Mind*. New York, NY: The MacMillan Co.

Li, Q., Nourbakhsh, A., Shah, S., and Liu, X. 2017a. "Real-Time Novel Event Detection from Social Media," *IEEE 33rd International Conference on Data Engineering*, San Diego, CA: IEEE Computer Society, pp. 1129-1140.

Li, Z., Fang, X., and Sheng, O. R. L. 2017b. "A Survey of Link Recommendation for Social Networks: Methods, Theoretical Foundations, and Future Research Directions," *ACM Transactions on Management Information Systems* (9:1), pp. Article 1: 1-26.

Liang, Y., Zheng, X., Zeng, D. D., Zhou, X., Leischow, S. J., and Chung, W. 2015. "Characterizing Social Interaction in Tobacco-Oriented Social Networks: An Empirical Analysis," *Nature Scientific Reports* (5).

Liao, H., Mariani, M. S., Medo, M., Zhang, Y. C., and Zhou, M. Y. 2017. "Ranking in Evolving Complex Networks," *Physics Reports-Review Section of Physics Letters* (689), pp. 1-54.

Liben-Nowell, D., and Kleinberg, J. M. 2007. "The Link-Prediction Problem for Social Networks," *Journal of the American society for information science and technology (JASIST)* (58), pp. 1019-1031.

Lim, K., Lee, K., Kendal, D., Rashidi, L., Naghizade, E., Winter, S., and Vasardani, M. 2018. "The Grass Is Greener on the Other Side: Understanding the Effects of Green Spaces on Twitter User Sentiments," *WWW*, pp. 275-282.

Marble, J. L., Lawless, W. F., Mittu, R., Coyne, J., Abramson, M., and Sibley, C. 2015. "The Human Factor in Cybersecurity: Robust & Intelligent Defense," in *Cyber Warfare: Building the Scientific*

Foundation, S. Jajodia, P. Shakarian, V.S. Subrahmanian, V. Swarup and C. Wang (eds.). Switzerland: Springer.

Marlow, J., Dabbish, L., and Herbsleb, J. 2013. "Impression Formation in Online Peer Production: Activity Traces and Personal Profiles in Github," *CSCW*, San Antonio: ACM, pp. 117-128.

Miller, N., and Campbell, D. T. 1959. "Recency and Primacy in Persuasion as a Function of the Timing of Speeches and Measurements," *The Journal of Abnormal and Social Psychology* (59:1), pp. 1-9.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. 2007. "Measurement and Analysis of Online Social Networks," *IMC*.

Ngai, E. W. T., Tao, S. S. C., and Moon, K. K. L. 2015. "Social Media Research: Theories, Constructs, and Conceptual Frameworks," *International Journal of Information Management* (35:1), pp. 33-44.

Pelled, A., Zilberstein, T., Tsirulnikov, A., Pick, E., Patkin, Y., and Tal-Or, N. 2017. "Textual Primacy Online: Impression Formation Based on Textual and Visual Cues in Facebook Profiles," *American Behavioral Scientist* (61:7), pp. 672-687.

Quan-Haase, A. 2016. *Technology and Society*, (2 ed.). Oxford, UK: Oxford University Press.

Reicher, S. 2000. "Collective Behavior," in: *Encyclopedia of Psychology,* A.E. Kazdin (ed.). Washington, DC: American Psychological Association, pp. 374-377.

Scellato, S., Noulas, A., and Mascolo, C. 2011. "Exploiting Place Features in Link Prediction on Location-Based Social Networks," *KDD*.

Snijders, T. A. B., van de Bunt, G. G., and Steglich, C. E. G. 2010. "Introduction to Stochastic Actor-Based Models for Network Dynamics," *Social Networks* (32:1), pp. 44-60.

Turner, R., and Killian, L. M. 1957. *Collective Behavior*. Englewood Cliffs, NJ: Prentice Hall.

Ugander, J., Karrer, B., Backstrom, L., and Marlow, C. 2011. "The Anatomy of the Facebook Social Graph," *CoRR* (abs/1111.4503).

Wang, P., Xu, B. W., Wu, Y. R., and Zhou, X. Y. 2015. "Link Prediction in Social Networks: The State-of-the-Art," *Science China-Information Sciences* (58:1).

Wei, W., and Carley, K. M. 2015. "Measuring Temporal Patterns in Dynamic Social Networks," *ACM Transantions on Knowledge Discovery from Data* (10:1), pp. 1-27.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K., and Zhao, B. 2009. "User Interactions in Social Networks and Their Implications," *EuroSys*, Nuremberg.

Yang, C., Zhong, L., Li, L., and Jie, L. 2017. "Bi-Directional Joint Inference for User Links and Attributes on Large Social Graphs," *WWW*.

Yu, Y., Yin, G., Wang, H., and Wang, T. 2014. "Exploring the Patterns of Social Behavior in Github," *ACM CrowdSoft*, Hong Kong.

Zeng, D. 2015. "Policy Informatics for Smart Policy-Making," *IEEE Intelligent Systems* (30:6), pp. 2-3.