Co-evolutionary Dynamics of Information Diffusion and Network Structure

Mehrdad Farajtabar
GeorgiaTech
mehrdad@gatech.edu

Manuel Gomez-Rodriguez
MPI for Software Systems
manuelgr@mpi-sws.org

Yichen Wang
GeorgiaTech
yichen.wang@gatech.edu

Shuang Li
GeorgiaTech
sli370@gatech.edu

Hongyuan Zha
GeorgiaTech
zha@cc.gatech.edu

Le Song
GeorgiaTech
lsong@cc.gatech.edu

ABSTRACT

Information diffusion in online social networks is obviously affected by the underlying network topology, but it also has the power to change that topology. Online users are constantly creating new links when exposed to new information sources, and in turn these links are alternating the route of information spread. However, these two highly intertwined stochastic processes, information diffusion and network evolution, have been predominantly studied *separately*, ignoring their co-evolutionary dynamics.

In this project, we propose a probabilistic generative model, COEVOLVE, for the joint dynamics of these two processes, allowing the intensity of one process to be modulated by that of the other. This model allows us to efficiently simulate diffusion and network events from the co-evolutionary dynamics, and generate traces obeying common diffusion and network patterns observed in real-world networks. Furthermore, we also develop a convex optimization framework to learn the parameters of the model from historical diffusion and network evolution traces. We experimented with both synthetic data and data gathered from Twitter, and show that our model provides a good fit to the data as well as more accurate predictions than alternatives.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

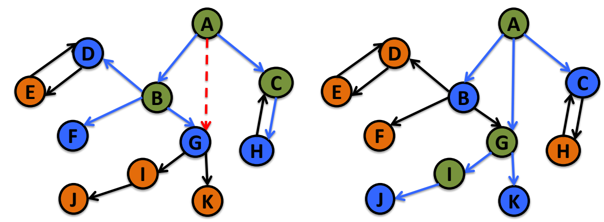
Keywords

Network Co-evolution; Hawkes Process; Survival Analysis

1. INTRODUCTION

Users in social networks often forward to their *followers* information they are exposed to via their *followees*, triggering the emergence of information cascades that travel through the network. Besides these dynamic processes *on* the network, the network topology itself often undergoes dynamic changes, since online users are constantly creating new links to information sources.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2744105.



(a) Information diffusion creates new links (b) New links create new paths for information diffusion

Figure 1: Joint Dynamics. Blue links form the diffusion paths. Green and blue circles are nodes which see the information. Green circles are those who re-share and propagate the information. Red circles are nodes unaware of A's post.

While there have been many recent works on modeling information diffusion and network evolution, most of them treat these two stochastic processes independently and separately, ignoring the influence one may have on the other over time. However, recent empirical studies have been able to overlay these two sources of data, and show that both processes are indeed coupled and network changes are often triggered by information diffusion [5, 1, 4].

We propose a probabilistic generative model for the joint dynamics of information diffusion and network evolution. The model consists of two interwoven components, illustrated in Figure 1.

I. **Information diffusion model.** We design an "identity revealing" multivariate Hawkes process [3] to capture the mutual excitation behavior of retweeting events, where the intensity of such event in a user is boosted by an aggregation of events from her followees. Although Hawkes processes have been used for information diffusion before [2], a major improvement of our approach is that we explicitly model the excitation due to a particular source node.

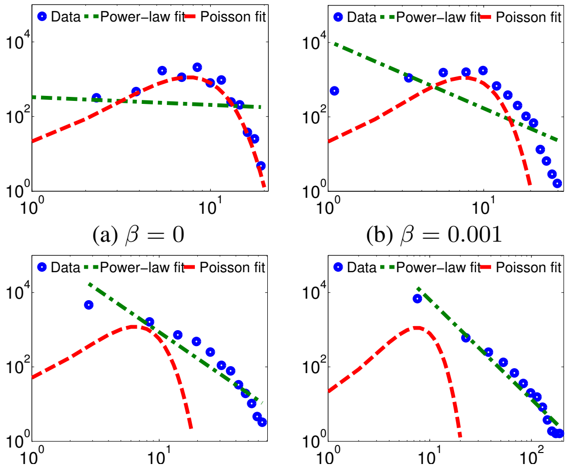
II. **Network evolution model.** We model link creation as an "information driven" survival process, and couple the intensity of this process with retweeting events.

2. MODEL

We will model the generation of two types of events: tweet/retweet events, e^r, and link creation events, e^l. We represent the events as

$$e^r \text{ or } e^l := (\underset{\substack{\uparrow \\ \text{destination}}}{u}, \overset{\substack{\text{source} \\ \downarrow}}{s}, \underset{\substack{\uparrow \\ \text{time}}}{t}). \tag{1}$$

For retweet event, the triplet means that the destination node u retweets at time t a tweet originally posted by source node s. This event can happen when u is retweeting a message by another node u' where the original information source s is acknowledged.



(a) $\beta = 0$ (b) $\beta = 0.001$

(c) $\beta = 0.1$ (d) $\beta = 0.8$

Figure 2: Degree distributions when for different β.

Given a list of retweet events $\{e_1^r = (u_1, s_1, t_1), \ldots\}$ up to time t, the history $\mathcal{H}_{us}^r(t)$ of retweets by u due to source s is $\mathcal{H}_{us}^r(t) = \{e_i^r = (u_i, s_i, t_i) | u_i = u \text{ and } s_i = s\}$.

For link creation event, the triplet means that destination node u creates at time t a link to source node s, *i.e.*, from time t on, node u starts following node s. We restrict ourselves to the case where each (directed) link is created only once. We denote the link creation history as $\mathcal{H}^l(t)$. Given m users, we will use two sets of counting processes to record the generated events.

- **Counting processes for retweets** are denoted as a matrix $\boldsymbol{N}(t)$ of size $m \times m$ for each fixed time point t. The (u, s)-th entry in the matrix, $N_{us}(t) \in \{0\} \cup \mathbb{Z}^+$, counts the number of retweets of u due to source s up to time t. These counting processes are "identity revealing".
- **Survival processes for links** are denoted also as a matrix $\boldsymbol{A}(t)$ of size $m \times m$ for each fixed time point t. The (u, s)-th entry in the matrix, $A_{us}(t) \in \{0, 1\}$, indicates whether u is directly following s. That is $A_{us}(t) = 1$ means the directed link has been created before t. We do not allow self-links.

We will denote matrices $d\boldsymbol{N}(t) := (\, dN_{us}(t) \,)_{u,s \in [m]}$ and similarly $d\boldsymbol{A}(t) := (\, dA_{us}(t) \,)_{u,s \in [m]}$.

$$\kappa_\omega(t) \star d\boldsymbol{N}(t) := (\, \kappa_\omega(t) \star dN_{us}(t) \,)_{u,s \in [m]}$$

$$\boldsymbol{A}(t) \circ d\boldsymbol{N}(t) := (\, A_{us}(t)\, dN_{us}(t) \,)_{u,s \in [m]}$$

These operate elementwisely. Furthermore, we can also carry out matrix multiplication operation on $\boldsymbol{A}(t)$ and $d\boldsymbol{N}(t)$. For instance, $\boldsymbol{A}(t)\, d\boldsymbol{N}(t) := \left(\sum_{w \in [m]} A_{uw}(t)\, dN_{ws}(t) \right)_{u,s \in [m]}$. Then the interwoven information diffusion and network evolution processes can be characterized using their respective intensities

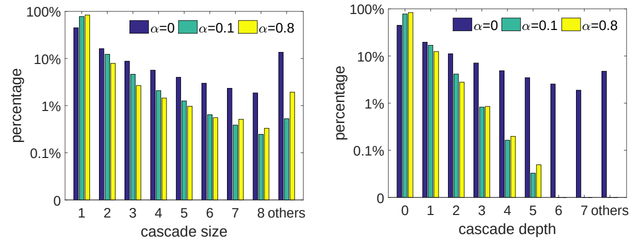
$$\mathbb{E}[d\boldsymbol{N}(t) \,|\, \mathcal{H}^r(t) \cup \mathcal{H}^l(t)] = \boldsymbol{\Gamma}^*(t)\, dt, \quad \text{and} \tag{2}$$

$$\mathbb{E}[d\boldsymbol{A}(t) \,|\, \mathcal{H}^r(t) \cup \mathcal{H}^l(t)] = \boldsymbol{\Lambda}^*(t)\, dt, \tag{3}$$

where $\boldsymbol{\Gamma}^*(t) = (\, \gamma_{us}^*(t) \,)_{u,s \in [m]}$ and $\boldsymbol{\Lambda}^*(t) = (\, \lambda_{us}^*(t) \,)_{u,s \in [m]}$. We model the intensities, $\boldsymbol{\Gamma}^*(t)$, for retweeting events and $\boldsymbol{\Lambda}^*(t)$, for link creation as

$$\begin{aligned} \boldsymbol{\Gamma}^*(t) &:= \left(\eta + \beta \sum_{w \in [m]} A_{uw}(t)\, (\kappa_\omega(t) \star dN_{ws}(t)) \right)_{u,s \in [m]} \\ &= \eta + \beta\, \boldsymbol{A}(t)\, (\kappa_\omega(t) \,\star\, d\boldsymbol{N}(t))\,. \end{aligned} \tag{4}$$

$$\begin{aligned} \boldsymbol{\Lambda}^*(t) &:= \left((1 - A_{us}(t))(\mu + \alpha\, \kappa_\omega(t) \star dN_{us}(t)) \right)_{u,s \in [m]} \\ &= (1 - \boldsymbol{A}(t)) \circ (\mu + \alpha\, \kappa_\omega(t) \star d\boldsymbol{N}(t)). \end{aligned} \tag{5}$$



(a) (b)

Figure 3: Cascade size and depth distributions for different α

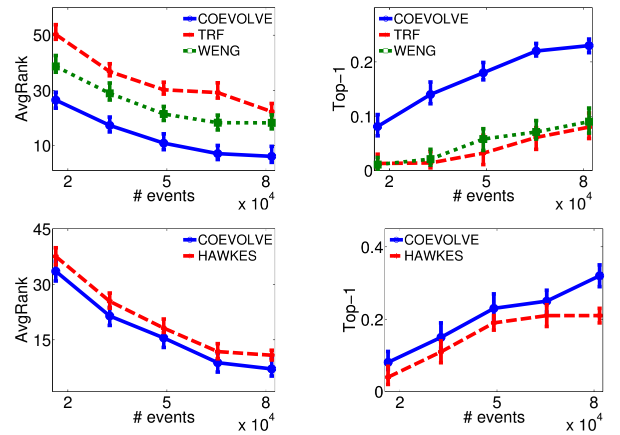


Figure 4: Prediction performance in the Twitter dataset by means of average rank (AR) and success probability that the true (test) events rank among the top-1 events (Top-1).

3. EXPERIMENTS

Degree Distribution. Empirical studies have shown that the degree distribution of social networks and microblogging sites follow a power law, and argued that it is a consequence of the rich get richer phenomena. Intuitively, the higher the values of the parameters α and β, the closer the resulting degree distribution follows a power-law; the lower their values, the closer the distribution to an Erdos-Renyi graph. Figure 2 confirms this intuition via the degree distribution for different values of β.

Cascade Patterns. Figure 3 summarize the cascade size and depth with varying α. The higher the α value, the shallower and wider the cascades.

Link and Activity Prediction. We summarize the results in Fig. 4, where we consider an increasing number of training retweet/tweet for training. Our model outperforms all other link prediction and activity prediction methods consistently.

References

[1] D. Antoniades and C. Dovrolis. Co-evolutionary dynamics in social networks: A case study of twitter. *preprint arXiv:1309.6001*, 2013.

[2] M. Farajtabar, N. Du, M. Gomez-Rodriguez, I. Valera, H. Zha, and L. Song. Shaping social activity by incentivizing users. In *Advances in Neural Information Processing Systems*, 2014.

[3] T. Liniger. *Multivariate Hawkes Processes*. PhD thesis, Swiss Federal Institute of Technology Zurich, 2009.

[4] S. A. Myers and J. Leskovec. The bursty dynamics of the twitter information network. In *WWW* , 2014.

[5] L. Weng, J. Ratkiewicz, N. Perra, B. Gonçalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, and A. Flammini. The role of information diffusion in the evolution of social networks. In *ACM SIGKDD*, 2013.