Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

Inferring Motif-Based Diffusion Models for Social Networks

Qing Bao, William K. Cheung, and **Jiming Liu**
Dept. of Computer Science, Hong Kong Baptist University, Hong Kong
{qingbao,william,jiming}@comp.hkbu.edu.hk

Abstract

Existing diffusion models for social networks often assume that the activation of a node depends independently on their parents' activations. Some recent work showed that incorporating the structural and behavioral dependency among the parent nodes allows more accurate diffusion models to be inferred. In this paper, we postulate that the latent temporal activation patterns (or motifs) of nodes of different social roles form the underlying information diffusion mechanisms generating the information cascades observed over a social network. We formulate the inference of the temporal activation motifs and a corresponding motif-based diffusion model under a unified probabilistic framework. A two-level EM algorithm is derived so as to infer the diffusion-specific motifs and the diffusion probabilities simultaneously. We applied the proposed model to several real-world datasets with significant improvement on modelling accuracy. We also illustrate how the inferred motifs can be interpreted as the underlying mechanisms causing the diffusion process to happen in different social networks.

1 Introduction

How information spreads from one node to another over online social and information networks is well known to be highly related to the influence between the nodes. In the literature, different diffusion models have been proposed and the underlying diffusion networks can be inferred to explain the observed information cascades [Goldenberg *et al.*, 2001; Lee *et al.*, 2012; Goyal *et al.*, 2010; Gomez-Rodriguez *et al.*, 2011]. Most of them share the assumption that the activation of a node is caused independently by its parents. This assumption limits our understanding on how a node is influenced by the *interactions* of its parent node activations. For instance, sometimes people may not respond to discussions of their so-so friends until an more influencial friend steps in.

The long-standing framework which associates the probability of adopting a behavior with multiple neighbors [Granovetter, 1978; Jackson and Yariv, 2007] has recently been observed also for the user engagement behavior in Facebook [Ugander *et al.*, 2012]. This notion was later on incorporated into diffusion models for social networks [Bao *et al.*, 2013], where a component-based diffusion model was proposed in which the influence on a node by its parents is exerted independently by the connected components of the parents. In [Bao *et al.*, 2015], a set of inferred co-activation patterns of the parents of each node are assumed to be independently exerting their influence on the node, where the co-activation patterns are assumed to be static in nature.

In this paper, we propose a diffusion model with the dynamic interaction patterns of nodes incorporated. Motifs, often defined as *over-represented* patterns, have widely been applied for characterizing the structural and functional properties of sequential data (e.g., genomic DNA [Kim *et al.*, 2008]) and network data (e.g., [Milo *et al.*, 2002]). We postulate that the information cascades over the whole network are embedded with a common set of latent temporal activation patterns as the underlying interaction mechanisms of the nodes causing the information to diffuse over the social network. We propose a *stochastic temporal activation motif* model to represent the temporal activation patterns, and thus define a novel diffusion model based on the motifs. Also, we assume that the motifs are attributed with social roles (as in [Scripps *et al.*, 2007]). We formulate the motif detection problem and the diffusion network inference problem under a unified probabilistic framework. A two-level EM algorithm is derived so as to infer the diffusion-specific motifs and the diffusion probabilities simultaneously. For performance evaluation, we apply the proposed model to three real-world social network datasets with significant improvement on modelling accuracy compared with some recent work. We also illustrate how the inferred temporal activation motifs can be interpreted as the underlying interaction mechanisms causing the diffusion to happen in different social networks.

To the best of our knowledge, this is the first work where *the identification of temporal activation motifs and the inference of the information diffusion network are solved within a unified framework.* Also, the temporal activation motifs discovered are *diffusion-specific* ones, which also makes this work unique compared to other motif detection work.

2 Related Work

Diffusion models have been studied for the past decade to gain theoretical understanding of how information spreads within social networks [Goldenberg *et al.*, 2001; Lee *et al.*, 2012; Goyal *et al.*, 2010; Gomez-Rodriguez *et al.*, 2011]. Two commonly used models are the Independent Cascade (IC) model [Goldenberg *et al.*, 2001] and the Linear Threshold (LT) model [Kempe *et al.*, 2005], which have also been put under the unified framework of General Threshold and Cascade Models [Kempe *et al.*, 2003]. Point process models such as Hawkes process diffusion models have recently been proposed so as to model the inter-activation time [Yang and Zha, 2013; Zhou *et al.*, 2013; He *et al.*, 2015]. Also, instead of the conventional way to represent information spread as cascade sequences, some recent work assumes that a cascade takes the form of a tree [Sun *et al.*, 2009]. Some also studied how the diffusion processes in social networks are affected by factors like the cascade structure, its size, and the roles of the users involved [Anderson *et al.*, 2015]. In parallel, the basic IC model has also been extended to uncover temporal dynamics [Lee *et al.*, 2012], and to take continuous time [Goyal *et al.*, 2010] and *et al.*

Most of the extensions of the IC model still share the same assumption that a node is influenced *independently* by any of its parents. By exploring the structural and behavioral properties of neighboring nodes and their relationship with information diffusion, a component-based diffusion model was recently proposed [Bao *et al.*, 2013] where the influence of the parent nodes to a node is not exerted individually but by connected components of nodes. A structural diversity factor was applied to each such component to quantify the information redundancy. In [Zhang *et al.*, 2013], a related notion called social influence locality has been studied for modeling retweeting behaviors. In [Bao *et al.*, 2015], a set of inferred co-activation patterns of the parents of each node are assumed to be independently influencing the node. To contrast, our proposed motif-based diffusion model makes contribution to identifying latent temporal interaction patterns of nodes which cause the information diffusion to happen in social networks.

Recently, network motifs have been applied to characterize online social networks [Liu *et al.*, 2013]. However, to the best of our knowledge, there does not exist work on inferring *application-specific* motifs with social roles integrated. In this paper, we adopt the roles defined in [Scripps *et al.*, 2007]. There also exist other definitions proposed in the literature, e.g., Pagerank Scores and Structural Hole Spanner Scores [Fang and Tang, 2015].

3 A Motif-based Diffusion Model

In this section, we first present some key concepts needed for defining the information diffusion modeling problem. In particular, we make use of also social roles in our formulation. Then, we present a stochastic temporal activation motif model, followed by a corresponding diffusion model and a two-level EM algorithms for the model learning.

3.1 Preliminaries

Social Network and Information Cascade

We represent a *social network* as a directed graph $G = (V, E)$ where V is the set of nodes and E is the set of edges. Let $e = (v, w)$ be an edge from node v to node w, and $f(w)$ and $b(w)$ be the sets of child nodes and parent nodes of node w respectively, given as: $f(w) = \{u : (w, u) \in E\}$ and $b(w) = \{v : (v, w) \in E\}$.

Let $D_s = \{D_s(0), D_s(1) \cdots D_s(T_s)\}$ be the s^{th} observed *information cascade* where $D_s(t)$ is the set of nodes activated at time step t and T_s is the final time step for the cascade D_s.

Social Role

In our model, we assume that each node v carries a *social role* $R(v) := \{r_1, ...r_n\}$ based on [Scripps *et al.*, 2007]. In particular, there are four different social roles, namely *ambassadors*, *big fish*, *bridges* and *loners*, defined based on the local structural information of different nodes. Ambassadors have both high degrees and many connections to different communities, representing the ones with high *global* influence. Big fish (big fish in a small pool) have high degrees but connections to only a limited number of communities, standing for those with high local influence, like leaders in a small research field. Bridges have low degrees but are connected to many communities, playing the roles of bringing knowledge exchange among communities although not influential. Meanwhile, loners have both lower degrees and less connections to different communities, representing the ones being alone. By adopting these roles, we can then try to discover and see if there exist patterns characterizing how discussions are developed via the interactions of the roles. In general, other role definitions can also be used according to the specific need of analysis.

Activation Sequence and Subsequence

Diffusion models for social networks explain how the *activations* of some user nodes makes *influence* on their neighbors so that they will activate as well. Given that in the s^{th} cascade, assume that a node w is activated at time t and $t - L_s(w, t)$ consecutively where $L_s(w, t)$ is defined as the interval between the latest activation of the node w prior to time step t in the s^{th} cascade. To explain the activation of w at t, the set of activations being considered are those found in $b(w)$ (parents of w) and activated in the time interval $[t - L_s(w, t), t]$. The activations are then sorted in ascending order, each tagged with a pre-computed social roles to form a particular *temporal activation sequence* for node w.

In addition, we make the conjectures that (i) there exist temporal subsequence patterns embedded in the set of temporal activation sequences for all nodes in G and that (ii) the parent activations even though sorted in time may not necessarily mean that they happen in response to each other one by one exactly according to the temporal order. We thus sample subsequences from a temporal activation sequence by grouping adjacent activations happening within a pre-set Δt in an overlapping manner. We implement a depth-first search with a First-In First-Out stack. We first push all the individual activations in a temporal activation sequence (each as a subsequence candidate) into the stack. Then, we pop a subsequence candidate X from the stack. Suppose x_i is the last activation in the candidate X. We search for the activation x_j happening later than x_i in the original activation sequence. If such x_j can be found and happens within a Δt time window, we will add x_j to the end of X and push the new subsequence candidate into the stack again. Otherwise, the candidate will be finalized. The process repeats until there are no more candidates in the stack. We can also set a limit on the maximum number of activations allowed for a subsequence as l. In the sequel, we denote the corresponding set of subsequences for node w observed at time t in the s^{th} cascade as $X_w^{(s)}(t)$.

3.2 Formulation

To formulate a motif-based diffusion model under a probabilistic framework, we first define a stochastic temporal activation motif to represent the patterns of the aforementioned activation subsequence.

Temporal Activation Motif

Let $M = \{M_1, \ldots, M_k\}$ be a set of k *stochastic temporal activation motifs*, which are parameterized as a set of probability matrices $\Theta = \{\Theta_1, \ldots, \Theta_k\}$. In particular, $\Theta_m = (\theta_{ij}^m)_{n\times n}$ is defined so that θ_{ij}^m is the probability of transiting from role r_i to role r_j as represented in the m^{th} motif where n is the number of social roles. The probability that a particular subsequence $h \in X_w^{(s)}(t)$ is generated by the m^{th} motif is given as:

$$p(h|\Theta_m) = u_{h_1} \prod_{t=2}^{l_h} \prod_{i=1}^{n} e_i^h(t-1) \prod_{j=1}^{n} (\theta_{ij}^m)^{e_j^h(t)} (1-\theta_{ij}^m)^{(1-e_j^h(t))}$$

where $0 \le \theta_{ij} \le 1$, $1 \le m \le k$, u_{h_1} denotes the initial probability of observing the first element of the subsequence h. $e_i^h(t)$ is an indicator which will be 1 if $h_t = r_i$ where h_t is the t^{th} element in h. Also, a background model M_0 is used to account for subsequences of activations which may happen at random. For simplicity, we compute the probability of generating h by M_0 as $\prod_t p(h_t)$, where $p(h_t)$ is estimated as N_r/N_{total} if $h_t = r$, where N_r and N_{total} denote the total number of activations with role r and the total number of all activations respectively.

Motif-based Information Diffusion

For a node to be influenced by a set of parent nodes causing information diffusion, we postulate that it is the interaction patterns of the parent node activations (i.e., the temporal activation motifs) which form the underlying mechanisms. Therefore, we assume that each subsequence of parent node activations will contribute to the probability that one of the motifs is in effect. And given a particular motif and a particular node, we define the corresponding *motif-based* diffusion probability.

We denote $\tau_{m,w} = p(w = 1|\Theta_m)$ as the motif-based diffusion probabilities between M_m and a node w, and the prior probabilities of the M_m as $\alpha_m = p(\Theta_m)$. The diffusion process of a particular cascade proceeds as follows. Given the initial set of activated nodes in the s^{th} cascade ($D_s(0)$), we assume that each of them tries to activate its child nodes. For each such child node w, the probability that it will be activated by a subsequence $h \in X_w^{(s)}(t)$ is then given as:

$$\begin{aligned} & p(w(s,t+1)|h) \\ = & \sum_{m=0}^{k} p(w|\Theta_m)p(\Theta_m|h) = \sum_{m=0}^{k} \tau_{m,w}\alpha_m p(h|\Theta_m)/p(h). \end{aligned}$$

With the independency assumption that node w will be activated if at least one of those subsequences succeeds, the probability for w to activate at time step $t+1$ in the s^{th} cascade is given as:

$$\begin{aligned} & p(w(s,t+1)|X_w^{(s)}(t)) \\ = & \ 1 - \prod_{h\in X_w^{(s)}(t)} (1 - p(w(s,t+1)|h)). \end{aligned}$$

The diffusion process proceeds until there are no more nodes being activated and the cascade stops.

The likelihood function of the set of observed cascades $\{D_s\}$ is then given as:

$$\begin{aligned} L(\mathbf{\Theta}) = & \sum_{s=1}^{S} logP(D_s|\mathbf{\Theta}, D_s(0)) \\ = & \sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Bigg(\sum_{w\in D_s(t+1)} \log p(w(s,t+1)|X_w^{(s)}(t)) \\ & + \sum_{w\notin D_s(t+1)} \sum_{h\in X_w^{(s)}(t)} \log(1-p(w(s,t+1)|h))\Bigg). \end{aligned}$$

3.3 Learning Algorithms

To infer the latent temporal activation motifs and the diffusion probabilities simultaneously, we propose a two-level EM algorithm to maximize the likelihood function $L(\mathbf{\Theta})$ with respect to the parameters $\mathbf{\Theta} = \{\{\tau_{m,w} = p(w|\Theta_m)\}, \{\alpha_m = p(\Theta_m)\}, \{\Theta_m\}\}$. The inferred temporal motifs will be those making the information diffusion mostly likely to happen.

First level EM

Let $I_{h,m}$ be a latent variable that takes the value of 1 when a parent subsequence h belongs to the latent pattern m, and 0 otherwise, given the constraint $\sum_{m=0}^{k} I_{h,m} = 1$. Let $I = \{I_{h,m}\}$ denote the whole set of the latent variables. If we assume that I is known, the complete likelihood function can be written as:

$$P(D, I|\mathbf{\Theta}) = P(D|I, \mathbf{\Theta})P(I|\mathbf{\Theta})$$

where

$$\begin{aligned} & P(I|\mathbf{\Theta}) \\ = & \prod_{s=1}^{S}\prod_{t=0}^{T_s-1}\Bigg(\prod_{w}\prod_{h\in X_w^{(s)}(t)}\prod_{m=0}^{k}\big(\alpha_m p(h|\Theta_m)/p(h)\big)^{I_{h,m}}\Bigg) \end{aligned}$$

and

$$\begin{aligned} & P(D|I, \mathbf{\Theta}) = L(\mathbf{\Theta}|I) \\ = & \sum_{s=1}^{S} logP(D_s|\mathbf{\Theta}, D_s(0), I) \\ = & \sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Bigg(\sum_{w\in D_s(t+1)} \log p(w(s,t+1), I|X_w^{(s)}(t)) \\ & + \sum_{w\notin D_s(t+1)} \sum_{h\in X_w^{(s)}(t)} log(1 - \sum_{m=0}^{k} I_{h,m}\tau_{m,w})\Bigg) \end{aligned}$$

where

$$p(w(s,t+1), I|X_w^{(s)}(t)) = 1 - \prod_{h\in X_w^{(s)}(t)} (1 - \sum_{m=0}^{k} I_{h,m}\tau_{m,w}).$$

As I is missing in most of the cases, we can do the E-step by first computing the posterior probabilities of I with the current parameter estimates $\hat{\tau}_{m,w} = p(w|\hat{\Theta}_m)$, $\hat{\alpha}_m = p(\hat{\Theta}_m)$ and $\hat{\Theta}_m$, given as

$$\eta_{h,m} = P(I_{h,m}=1|w,h,\hat{\mathbf{\Theta}}) = \frac{p(w|\hat{\Theta}_m)p(\hat{\Theta}_m|h)}{\sum_{m=0}^{k} p(w|\hat{\Theta}_m)p(\hat{\Theta}_m|h)} = \frac{\hat{\tau}_{m,w}\hat{\alpha}_m p(h|\hat{\Theta}_m)}{\sum_{m=0}^{k}\hat{\tau}_{m,w}\hat{\alpha}_m p(h|\hat{\Theta}_m)}.$$

Note that for the case when w is not activated, we substitute $\hat{\tau}_{m,w}$ with $(1-\hat{\tau}_{m,w})$ for $\eta_{h,m}$. Then, the expected likelihood function can be defined as:

$$\begin{aligned}&\mathcal{Q}(\mathbf{\Theta}|\hat{\mathbf{\Theta}})\\ =&\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Big(\sum_{w\in D_s(t+1)} E_I[\log p(w(s,t+1),I)|X_w^{(s)}(t)]\\ &+\sum_{w\notin D_s(t+1)}\sum_{h\in X_w^{(s)}(t)}\sum_{m=0}^{k}\eta_{h,m}\log(1-\tau_{m,w})\\ &+\sum_{w}\sum_{h\in X_w^{(s)}(t)}\sum_{m=0}^{k}\eta_{h,m}\log\big(\alpha_m p(h|\Theta_m)/p(h)\big)\Big). \quad (1)\end{aligned}$$

For the M-step, we maximize $\mathcal{Q}(\mathbf{\Theta}|\hat{\mathbf{\Theta}})$ by taking the derivative of $\mathcal{Q}$ with respect to $\mathbf{\Theta}$ to obtain the updating rule of the model parameters.

To update $\{\alpha_m\}$, according to the Lagrange multiplier method, maximizing $\mathcal{Q}(\mathbf{\Theta}|\hat{\mathbf{\Theta}})$ with constraint $\sum_{m=0}^{k}\alpha_m = 1$ yields

$$\partial\Big(\mathcal{Q}(\mathbf{\Theta}|\hat{\mathbf{\Theta}}) - \lambda(\sum_{m=0}^{k}\alpha_m - 1)\Big)/\partial\alpha_m = 0.$$

Thus, $\forall\alpha_m \ \sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\sum_w\sum_{h\in X_w^{(s)}(t)}\frac{\eta_{h,m}}{\alpha_m} - \lambda = 0$. Then it can be easily shown that

$$\alpha_m = \frac{\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\sum_w\sum_{h\in X_w^{(s)}(t)}\eta_{h,m}}{\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\sum_w\sum_{h\in X_w^{(s)}(t)}1}.$$

To update $\{\Theta_m\}$, we consider $p(h|\Theta_m)$, and

$$\partial\Big(\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\sum_{w}\sum_{h\in X_w^{(s)}(t)}\sum_{m=1}^{k}\eta_{h,m}\sum_{t'=2}^{l_h}\sum_{i=1}^{n}e_i^h(t'-1)\sum_{j=1}^{n}\big(e_j^h(t')\log\theta_{ij}^m + (1-e_j^h(t'))\log(1-\theta_{ij}^m)\big)\Big)/\partial\theta_{ij}^m = 0.$$

Therefore,

$$\theta_{ij}^m = \frac{\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\sum_w\sum_{h\in X_w^{(s)}(t)}\eta_{h,m}\sum_{t'=2}^{l_h}e_i^h(t'-1)e_j^h(t')}{\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\sum_w\sum_{h\in X_w^{(s)}(t)}\eta_{h,m}\sum_{t'=2}^{l_h}e_i^h(t'-1)},$$

where $m \neq 0$. Since $p(h|\Theta_0)$ is estimated as constant, there is no corresponding parameters for $m = 0$.

To update $\{p(w|\Theta_m)\}$, setting to zero the derivative for the first term $E_I[\log p(w(s,t+1), I|X_w^{(s)}(t))]$ in Eq.(1) does not have a simple solution. So, within this M-step, we introduce another level of the EM algorithm.

Second level EM

Let $Y_{h,w}^{(s)}(t)$ denote a latent variable that indicates whether the activation of a node w at time step t in the s^{th} cascade is due to w's parent subsequence h or not. We further define $Y_s = \{Y_s(0), Y_s(1)\cdots Y_s(T_s)\}$ where $Y_s(t) := \{Y_{h,w}^{(s)}(t)\}$ represents the set of latent variables corresponding to the activations at time step t in the s^{th} cascade. Then, we compute the posterior probability of $Y_{h,w}^{(s)}(t)$, given as

$$\gamma_{h,w,s,t} = P(Y_{h,w}^{(s)}(t+1)=1|w,\{\eta_{h,m}\},\hat{\mathbf{\Theta}}) = \frac{\sum_{m=0}^{k}\eta_{h,m}\hat{\tau}_{m,w}}{\hat{p}(w(s,t+1)|X_w^{(s)}(t))}$$

where $\hat{\tau}_{m,w}$ stands for the current estimate of $\tau_{m,w}$, and

$$\hat{p}(w(s,t+1)|X_w^{(s)}(t)) = 1 - \prod_{h\in X_w^{(s)}(t)}(1-\sum_{m=0}^{k}\eta_{h,m}\hat{\tau}_{m,w}).$$

The corresponding $\mathcal{Q}'$ function can then be defined as

$$\begin{aligned}&\mathcal{Q}'(\mathbf{\Theta}|\hat{\mathbf{\Theta}})\\ =&\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Bigg(\sum_{w\in D_s(t+1)}\sum_{h\in X_w^{(s)}(t)}\sum_{m=0}^{k}\eta_{h,m}\\ &\big(\gamma_{h,w,s,t}\log\tau_{m,w} + (1-\gamma_{h,w,s,t})\log(1-\tau_{m,w})\big)\\ &+\sum_{w\notin D_s(t+1)}\sum_{h\in X_w^{(s)}(t)}\sum_{m=0}^{k}\eta_{h,m}\log(1-\tau_{m,w})\\ &+\sum_{w}\sum_{h\in X_w^{(s)}(t)}\sum_{m=0}^{k}\eta_{h,m}\log\big(\alpha_m p(h|\Theta_m)/p(h)\big)\Bigg).\end{aligned}$$

We define $T_{w,s}^{+}$ as the set of time steps $\{t\}$ with reference to the s^{th} cascade satisfying the condition that node w is activated at time step $t+1$ and at least one of its parents have been activated since $t - L_s(w,t)$. Meanwhile, $T_{w,s}^{-}$ is the set of time steps $\{t\}$ where node w is not activated at $t+1$, but at least one of its parents have been activated since $t - L_s(w,t)$. Moreover, we define a set of cascades where $T_{w,s}^{+}$ is not empty as $S_w^{+} = \{D_s : \exists t\big(X_w^{(s)}(t) \neq \emptyset \wedge w \in D_s(t+1)\big)\}$, and a set of cascades where $T_{w,s}^{-}$ is not empty as $S_w^{-} = \{D_s : \exists t\big(X_w^{(s)}(t) \neq \emptyset \wedge w \notin D_s(t+1)\big)\}$.

Then $\partial\mathcal{Q}/\partial\tau_{m,w} = 0$ yields:

$$\tau_{m,w} = \frac{1}{N_{m,w}^{+} + N_{m,w}^{-}}\sum_{s\in S_w^{+}}\sum_{t\in T_{w,s}^{+}}\sum_{h\in X_w^{(s)}(t)}\eta_{h,m}\gamma_{h,w,s,t}$$

$$N_{m,w}^{+} = \sum_{s\in S_w^{+}}\sum_{t\in T_{w,s}^{+}}\sum_{h\in X_w^{(s)}(t)}\eta_{h,m}$$

$$N_{m,w}^{-} = \sum_{s\in S_w^{-}}\sum_{t\in T_{w,s}^{-}}\sum_{h\in X_w^{(s)}(t)}\eta_{h,m}.$$

4 Experiments

We compare our model with some recently proposed diffusion models using three real-world social and information network datasets.

4.1 Experimental Settings

We compare our model (abbreviated as *Motif-IC*) with the basic IC model, a component-based IC model (abbreviated as *COMP-IC* [Bao *et al.*, 2013]) and a co-activation pattern based IC model (abbreviated as *LCM-IC* [Bao *et al.*, 2015]). The two variants of the IC models are chosen for comparison as both considered structural and behavioral dependency of the parent nodes. For all three models, we allow a node to be activated multiple times in a single cascade for fair performance comparison.

For all the experiments performed, the initial values of $\{\hat{\tau}_{m,w}\}$ are within $[0, 0.1]$ as the diffusion probabilities in real data are known to be very small (*e.g.*, with a mean value of 0.04 and standard deviation of 0.07 [Gruhl *et al.*, 2004]). The initial values of $\{\hat{\theta}_{ij}^m\}$ are generated within $[0, 1]$. And the initial values of α_m are generated within $[0, 1]$ satisfying $\sum_m \hat{\alpha}_m = 1$. Also, for *COMP-IC*, *LCM-IC* and *Motif-IC*, we obtain the optimal number of model components using the cross-validation method.

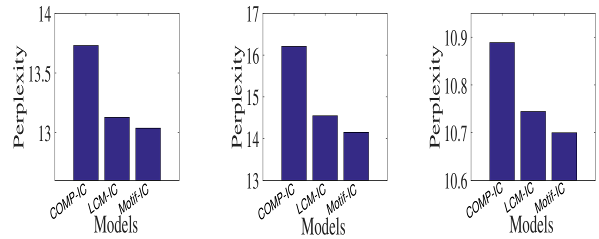
As the ground-truth is unknown for real data, we use perplexity as the metric for performance evaluation. The *perplexity* over a set of observed cascades is defined as

$$Perplexity = \frac{-\sum_{s=1}^{S} \ln P(D_s)}{W},$$

where $P(D_s)$ is the probability for the s^{th} cascade to be generated, and W is the number of activations due to the influence of the corresponding nodes' parents. A smaller perplexity value indicates the inferred model to be more accurate, and thus better performance. Also, five-fold cross-validation is adopted to avoid experimental bias.

4.2 Performance Evaluation

Three real datasets are used for the evaluation, namely MemeTracker [Leskovec *et al.*,], Digg [Lerman and Ghosh, 2010] and Flixster [Jamali and Ester, 2010] where both the network structure and the information cascades are available. (i) The MemeTracker dataset records the posts of mass media and weblogs from August 1 2008 to April 30 2009. Websites with news articles and blog posts are modeled as nodes which are further connected by directed edges (hyperlinks). It contains 4 million nodes, 13 million edges, and $71,568$ cascades. Each cascade is defined based on a set of posts reporting the same event. (ii) The Digg dataset records the story voting process under a directed friendship network over one month in 2009. Users are modeled as nodes and following relations are modeled as edges. It contains $280k$ nodes, 2.6 million edges and $3,553$ cascades. Each cascade is defined based on a particular frequently voted story. (iii) The Flixster dataset records the movie rating process under an undirected friendship network of users over a period from November 2005 to November 2009. Users are modeled as nodes and friend relations are modeled as edges. It contains $787k$ nodes and 5.9 million edges. We select $5,318$ cascades for the frequently rated movies in the dataset.



(a) MemeTracker (b) Digg (c) Flixster

Figure 1: Performance comparison on three real datasets.

Figure 1 shows the experimental results. For all the three datasets, the proposed *Motif-IC* model achieves the best performance consistently. Compared with *COMP-IC* and *LCM-IC*, *Motif-IC* results in decreases in perplexity by 0.69 and 0.09 respectively for MemeTracker, by 2.06 and 0.40 for Digg, and by 0.19 and 0.04 for Flixster. The optimal number of motifs, k, is found to be 5 for MemeTracker, 5 for Flixster, and 15 for Digg. The basic IC Model has much worse performance and thus the result is not shown in the figure.

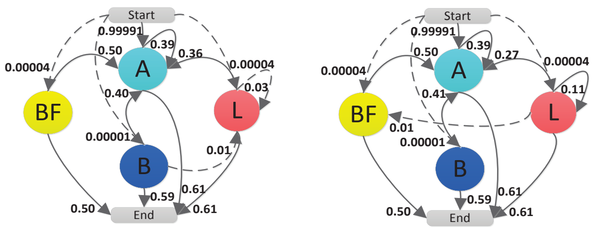
4.3 Discussion on the Inferred Motifs

The temporal activation motifs of parent nodes Θ_m inferred from the three datasets are reported in Figures 2, 3 and 4. The roles of *Ambassador*, *Big Fish*, *Bridge*, and *Loner* are indicated as "A", "BF", "B" and "L" respectively. We show only two motifs with the highest value for their mixing portions due to the page limit.

Table 1: Representative websites for each role in MemeTracker.

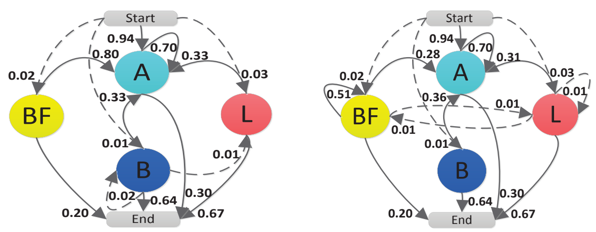
Role	Representative websites
Ambassador (covering various topics)	news.bbc.co.uk
	galvestondailynews.com
	articlesbase.com
Big Fish (featuring specific domains)	climateark.org
	waugh.standard.co.uk
	funkmysoul.gr
Bridge (news of several domains)	ronsen.org
	threadden.com
	merchantlaw.com
Loner (weblog sites)	humppazoid.blogspot.com
	matthewkeegan.com
	dpwriters.wordpress.com

For the Digg and Flixter datasets, the interpretation of the social roles is straight-forward based on the role definitions. For the MemeTracker dataset, Ambassadors are referring to the websites covering various topics (*e.g.*, *news.bbc.co.uk* a major news media). Big Fish are referring to the websites featuring specific domains (*e.g.*, *climateark.org* that publishes



(a) MemeTracker-Motif 1 (b) MemeTracker-Motif 2

Figure 2: Examples of temporal activation motifs for MemeTracker dataset.



(a) Digg-Motif 1 (b) Digg-Motif 2

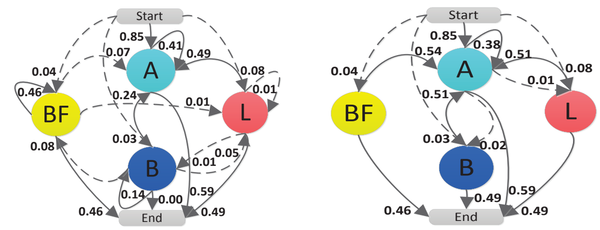
Figure 3: Examples of temporal activation motifs for Digg dataset.

news about climate). Bridges are referring to the websites covering news of several domains (*e.g.*, *merchantlaw.com* containing business laws in several practice areas like agriculture law and so on). The weblog sites often turn out to be Loners (*e.g.*, *humppazoid.blogspot.com*). Some examples can be found in Table 1.

In the following, we provide interpretations on the inferred motifs as *diffusion mechanisms*:

MemeTracker-Motifs: Referring to Figure 2, we can interpret MemeTracker-Motif 1 as the pattern of first starting with a posting at some famous websites (e.g., major news media) and then following up again by famous websites, with occasional starts at some domain specific websites and then picked up by famous websites afterwards. MemeTracker-Motif 2 is another pattern with some posting interactions among weblogs before it is being picked up by some famous websites. There is the possibility that MemeTracker-Motif 1 accounts for the diffusion of major news, while MemeTracker-Motif 2 accounts for the diffusion of news of some more specific topics which gain momentum of diffusion within the weblogs at the earlier stage.

Digg-Motifs: According to Figure 3, Digg-Motif 1 reveals a pattern that the globally reputable users starting the voting and then followed by also other reputable users, again with occasional starting votes from users of other types and then being followed by reputable users. Digg-Motif 2 refers to a pattern with occasional starts by big fish users and the voting process is retained among the big fish before the reputable users react. So, the first motif may refer to story votings among the circle of reputable users, while the second one may refer to story votings first among big fish users and got



(a) Flixster-Motif 1 (b) Flixster-Motif 2

Figure 4: Examples of temporal activation motifs for Flixster dataset.

spread also to reputable users. A user in principle can respond differently to these different mechanisms.

Flixster-Motifs: According to Figure 4, Flixster-Motifs 1 and 2 are more different from those inferred from the other two datasets. In particular, there is a much higher chance of transiting from globally reputable movie lovers to other types of users, and that there is a much higher chance that the movies are kept being evaluated within big fish, bridge, or loner circles. This could be because views on movies are more personal, resulting in more diverse parent activation patterns (instead of globally dominated by some users).

Note that in order to further validate the insights gained from the inferred motifs as discussed, more in-depth analysis and evaluation efforts will be needed. However, we believe that the inferred patterns form new hypotheses which are data-driven and worth further investigation for deepening our understanding on user behaviors and information diffusion in large-scale social networks.

5 Conclusion

In this paper, we proposed a novel motif-based diffusion model for social networks and a corresponding learning algorithm to infer both the latent temporal activation motifs as well as the diffusion model under a unified framework. The empirical results carried out on a number of real-world datasets consistently show that the proposed model achieves a higher modeling accuracy compared with a number of recently proposed models. Also, we have shown how the inferred temporal activation motifs can be interpreted as the underlying interaction mechanisms for the diffusion to happen in different social and information networks.

Possible extensions of this work may include fusing multimodal information in the cascades to infer more informative "colored" motifs. Also, the stochastic motif models can be further extended to increase its representation power. In general, we believe that this work provides a bridge between work of diffusion modeling and motif-based network characterization. Both have been used to support different social science related studies. Bridging them implies that it opens up a number of new opportunities. For instance, it will be interesting to perform careful studies to evaluate how the inferred diffusion specific motifs are related to the different types of events happening in social networks under different topics of concern. In addition, how each user responds to different activation motifs can also be studied to achieve better user characterization and influence maximization.

Acknowledgments

This work was partially supported by Hong Kong Baptist University Strategic Development Fund.

References

[Anderson *et al.*, 2015] A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec, and M. Tiwari. Global diffusion via cascading invitations: Structure, growth, and homophily. In *Proceedings of the 24th International Conference on World Wide Web*, pages 66–76, Florence, Italy, 2015.

[Bao *et al.*, 2013] Q. Bao, W.K. Cheung, and Y. Zhang. Incorporating structural diversity of neighbors in a diffusion model for social networks. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conference on Web Intelligence*, pages 431–438, Atlanta, GA, USA, 2013.

[Bao *et al.*, 2015] Q. Bao, W.K. Cheung, J. Liu, and Y. Song. Inferring latent co-activation patterns for information diffusion. In *Proceedings of the 2015 IEEE/WIC/ACM International Joint Conference on Web Intelligence*, pages 485–492, Singapore, 2015.

[Fang and Tang, 2015] Z. Fang and J. Tang. Uncovering the formation of triadic closure in social networks. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 2062–2068, Buenos Aires, Argentina, 2015.

[Goldenberg *et al.*, 2001] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review*, 9:1–18, 2001.

[Gomez-Rodriguez *et al.*, 2011] M. Gomez-Rodriguez, David Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *Proceedings of the 28th International Conference on Machine Learning*, pages 561–568, Bellevue, WA, USA, 2011.

[Goyal *et al.*, 2010] A. Goyal, F. Bonchi, and L. V.S. Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, pages 241–250, New York, NY, USA, 2010.

[Granovetter, 1978] M. Granovetter. Threshold models of collective behavior. *The American Journal of Sociology*, 83:1420–1443, 1978.

[Gruhl *et al.*, 2004] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th International Conference on World Wide Web*, pages 491–501, New York, NY, USA, 2004.

[He *et al.*, 2015] X. He, T. Rekatsinas, J. Foulds, L. Getoor, and Y. Liu. Hawkestopic: A joint model for network inference and topic modeling from text-based cascades. In *Proceedings of the 32th International Conference on Machine Learning*, pages 871–880, Lille, France, 2015.

[Jackson and Yariv, 2007] M. O Jackson and L. Yariv. Diffusion of behavior and equilibrium properties in network games. *The American Economic Review*, 97:92–98, 2007.

[Jamali and Ester, 2010] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the 4th ACM International Conference on Recommender Systems*, pages 135–142, New York, NY, USA, 2010.

[Kempe *et al.*, 2003] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, New York, NY, USA, 2003.

[Kempe *et al.*, 2005] D. Kempe, J. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. In *Proceedings of the 32nd International Conference on Automata, Languages and Programming*, pages 1127–1138, Berlin, Heidelberg, 2005.

[Kim *et al.*, 2008] N. Kim, K. Tharakaraman, L. Mariño-Ramírez, and J. L. Spouge. Finding sequence motifs with Bayesian models incorporating positional information: an application to transcription factor binding sites. *BMC Bioinformatics*, 9:1–11, 2008.

[Lee *et al.*, 2012] W. Lee, J. Kim, and H. Yu. CT-IC: Continuously activated and time-restricted Independent Cascade Model for viral marketing. In *Proceedings of the 12th International Conference on Data Mining*, pages 960–965, Brussels, Belgium, 2012.

[Lerman and Ghosh, 2010] K. Lerman and R. Ghosh. Information contagion: an empirical study of the spread of news on Digg and Twitter social networks. In *Proceedings of the 4th AAAI International Conference on Weblogs and Social Media*, pages 90–97, Washington, DC, USA, 2010.

[Leskovec *et al.*,] J. Leskovec, L. Backstrom, and J. Kleinberg. MemeTracker: Download MemeTracker data. [Online]. Available: http://www.memetracker.org/data.html.

[Liu *et al.*, 2013] K. Liu, W.K. Cheung, and J. Liu. Detecting stochastic temporal network motifs for human communication patterns analysis. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 533–540, Niagara, Ontario, Canada, 2013.

[Milo *et al.*, 2002] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[Scripps *et al.*, 2007] J. Scripps, P. Tan, and A. Esfahanian. Exploration of link structure and community-based node roles in network analysis. In *Proceedings of the 7th International Conference on Data Mining*, pages 649–654, Omaha NE, USA, 2007.

[Sun *et al.*, 2009] E. Sun, I. Rosenn, C. Marlow, and T. M. Lento. Gesundheit! Modeling contagion through Facebook news feed. In *Proceedings of the 3rd AAAI Conference on Weblogs and Social Media*, pages 146–153, San Jose, California, 2009.

[Ugander *et al.*, 2012] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109:5962–5966, 2012.

[Yang and Zha, 2013] S. Yang and H. Zha. Mixture of mutually exciting processes for viral diffusion. In *Proceedings of the 30th International Conference on Machine Learning*, pages 1–9, Atlanta, GA, USA, 2013.

[Zhang *et al.*, 2013] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li. Social influence locality for modeling retweeting behaviors. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 2761–2767, Beijing, China, 2013.

[Zhou *et al.*, 2013] K. Zhou, H. Zha, and L. Song. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Proceedings of the 6th International Conference on Artificial Intelligence and Statistics*, pages 641–649, Scottsdale, AZ, USA, 2013.