Inferring Latent Co-activation Patterns for Information Diffusion

Qing Bao[1], William K. Cheung[1], Jiming Liu[1], and Yunya Song[2]

[1]Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong
{qbao,william,jiming}@comp.hkbu.edu.hk
[2]Department of Journalism, Hong Kong Baptist University, Kowloon Tong, Hong Kong
yunyasong@hkbu.edu.hk

***Abstract*—Different diffusion models have been proposed in previous literature to model information diffusion, in which each node is often assumed to be independently influenced by its parents. More recently, some have begun to challenge this assumption based on the observation that structural and behavioral dependency among the parent nodes exerts a notable role in diffusion within networks. In this paper, we postulate that a node is independently influenced by a set of latent co-activation patterns of its parents, instead of the parents directly. We integrate the latent class model with the conventional independent cascade model where each latent class corresponds to a particular co-activation pattern of the parent nodes. Each parent activation is essentially first "projected" onto the latent space and then "reconstructed" before exerting its influence onto the child nodes. The co-activation patterns are to be inferred based on the information cascades observed without using the connectivity related cues except the information of direct parents. We formulate the co-activation pattern identification problem and the diffusion network inference problem under a unified probabilistic framework. A two-level EM algorithm is derived for inferring the model parameters. We applied the proposed model to a meme dataset and two social network datasets with promising results obtained. Using the results obtained based on the meme dataset, we also illustrate how the identified co-activation patterns can support the analysis of dependency among online news media.**

I. Introduction

Online social and information networks exploit the influence of neighbors to achieve information spreading. For instance, social networking sites like Google+, Facebook and Twitter allow their connected users to share views and information, which have now become major marketing platforms. Also, for mass media, the mainstream media outlets are increasingly scanning blogs and other online sources for news items, citing the links to weblogs in their websites and even hosting their own blogs.

With the objective to better our understanding on how information diffuses to exert influence in such online social and information networks, various diffusion models have been proposed in the literature [1, 2, 3, 4]. Early studies of diffusion built on the Independent Cascade (IC) model [1] which posits that an infected node infects each of its neighbors independently with some chosen probability. In contrast with the long-standing framework which associates the probability of adopting a behavior with the number of network neighbors already adopting [5] [6], Ugander *et al.* found that the user engagement in Facebook was affected by the connected components of users instead of the individual users in the contact neighborhood [7]. Bao *et al.* [8] proposed a component-based diffusion model which assumes that the influence of the parent nodes to a child node in a social network is not exerted individually but by connected components, and the validity of the assumption was supported by the empirical results when compared with the IC model based on a meme dataset. In short, for a more accurate diffusion model, the dependency of the parent nodes' activations should be considered. In this paper, we consider in particular the *co-activation* patterns of the parents for each node.

To that end, we postulate that the activation of a node is caused by a set of latent co-activation patterns of its parent nodes, and propose to integrate the latent class model (LCM) [9] into the conventional IC model for modeling the co-activation patterns. Under the proposed LCM-IC model, each parent's activation is first "projected" onto the latent space and then "reconstructed" before exerting its influence to the child node. Using latent variable models has been found effective to capture hidden patterns embedded in the data with missing values and noise. Applications include topic modeling [9] and collaborative filtering [10]. We here adopt it for diffusion modeling.

In this paper, we assume only the knowledge of direct parents for each node and the cascade information to infer the LCM-IC diffusion model. We formulate the co-activation pattern identification problem and the diffusion network inference problem under a unified probabilistic framework. The maximum likelihood approach is adopted to infer simultaneously the latent co-activation patterns and the diffusion probabilities. A two-level EM algorithm is derived for the inference. For performance evaluation, we apply the proposed model to both synthetic and a number of real social network datasets. Other than the objective of achieving a more accurate diffusion model, we use online news media as an example to illustrate how the inferred co-activation patterns provides insights in practice to support quantitative analysis of influence among the nodes in an information network.

978-1-4673-9618-9/15 $31.00 © 2015 IEEE
DOI 10.1109/WI-IAT.2015.115

The main contributions of this work are highlighted as follows: (1) To the best of our knowledge, this is the first work where the identification of co-activation patterns of parent nodes and the inference of the overall information diffusion network are solved within a unified framework solely based on information cascades and the knowledge of direct parents. (2) The proposed LCM-IC diffusion model allows the co-activation patterns to be discovered, and the positive empirical results obtained further hints the importance of considering the dependency among the parent nodes for diffusion modeling. (3) We demonstrate in detail how the proposed model with the consideration of the co-activation patterns can be applied to support dependency analysis of different online news media.

The remainder of this paper is organized as follows. Section II presents some related work. Section III gives the detailed formulation of our proposed model, followed by the EM algorithm for the model learning. Experimental results and related discussion can be found in Section IV. Section V concludes the paper and provides pointers for future work.

II. Related Work

Diffusion modeling has been researched extensively for the past decades, with a view to furthering theoretical understanding of how information spreads and exerts influence within online networks [1, 2, 3, 4]. Two commonly used models are the Independent Cascade (IC) model [1] and the Linear Threshold (LT) model [11]. The IC model [1] allows a node to be activated independently by any of its parents, while the LT model [11] assumes whether a node will be activated depends on the aggregation of the parents' activations. Both models have been further extended since they were first proposed. For instance, the IC model has been extended to uncover temporal dynamics [2], and to take continuous time [3]. Bao *et al.* [8] proposed a *component-based* diffusion model which assumes that the influence of the parent nodes to a child node is not exerted individually but by connected components. A community detection algorithm was applied to the neighborhood of each node to identify the underlying components and a structural diversity factor was also considered. In [12], a related notion called social influence locality has been studied for modeling retweeting behaviors. In the literature, studies on the relationship between the structural and behavioral properties of node neighbors and information diffusion is still rare. There have been works where the information cascades were used to detect *global* communities [13, 14]. However they focus on first grouping all individuals into global communities and then modeling influence among the communities, and thus are different from our work.

Our consideration of the parent nodes' co-activation behaviors has some overlap with the social circle detection problem [15]. Researchers [15, 16] have proposed algorithms to detect social circles by analyzing the users' profiles and generated contents. Under the context of social circle detection, this work aims to infer the "social circles" in the contact neighborhood of each user according to the friends' co-activation patterns.

Our work is also related to the studies analyzing the credibility of web/blog sites in a hyperlink structure. Hyperlinks as connections represent networks among people or organizations, and thus are often interpreted as the social or communication structure among those social actors. In the diffusion network of the new news ecosystem, through a hyperlink, an individual web/blog site plays the role of an actor who could influence other website's perceived credibility [17]. Most of the related works analyze the incoming and outgoing links of the sites in the hyperlink structure [18]. As hyperlinks could be created with different reasons, the use of information cascades observed over the network as presented in this work could provide more evidence to better capture the influence related structure, as detailed in Section IV-E.

III. A Diffusion Model with Co-activation Patterns Incorporated

In this section, we present a novel diffusion model where the co-activation patterns of the parents are incorporated for each node. Our conjecture is that parent nodes which often co-activate before the activation of a node should implicitly hint some underlying reason causing that, and detecting such latent parent co-activation patterns can better our understanding on the hidden reasons causing the underlying diffusion behaviors in social and information networks. For instance, recent postings of some friends with similar political views as yours may cause you to put forward their views via the information network. Also, postings of friends of different nationalities may get your attention on some international news. To formulate the proposed model, we integrate the latent class model and the conventional IC model as a unified one to represent the co-activation patterns and the *pattern-based* information diffusion. In the following, we present the mathematical notations and the formulation of the model, followed by a two-level EM algorithm for the model learning.

A. Preliminaries

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 $b(w) = \{v : (v, w) \in E\}$. For each node w, we assume that its activation depends on $N_z(w)$ different latent co-activation patterns of its parent nodes $b(w)$. We denote by $z^w \in \{1, ..., N_z(w)\}$ the index to the latent patterns. For each parent node $v \in b(w)$, we denote the probabilities that node v belongs to the $N_z(w)$ different latent patterns

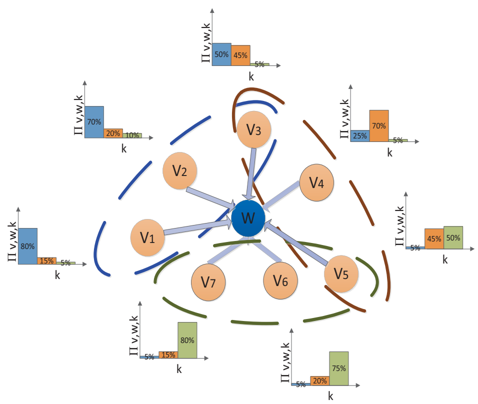


Figure 1: An illustration of the pattern-based diffusion model.

as $\Pi_{v,z^w} = \{\pi_{v,z^w=1}, \ldots, \pi_{v,z^w=N_z(w)}\}$ where π_{v,z^w} is the probability of the parent node v being assigned to the latent co-activation pattern z^w, $\sum_{z^w} \pi_{v,z^w} = 1$ and $\forall z^w(\pi_{v,z^w} \geq 0)$. In this paper, we assume the co-activation patterns to be static, leaving the modeling of evolving co-activation patterns as our future work.

B. Formulation

Given the latent co-activation patterns as defined for w, we further define for each pattern a *pattern-based diffusion probability* $\tau_{z^w,w}$ with $0 \leq \tau_{z^w,w} \leq 1$. That is when any node belonging to pattern z^w is "activated" (e.g., making a post online) at time t, there will be a probability $\tau_{z^w,w}$ that node w will then be activated by the pattern z_w. In addition, as in [2, 3], we allow an activated parent node to make influence via the latent patterns on node w multiple times within a short period after time t. Figure 1 shows an illustration of the proposed model. In the figure, three co-activation patterns are highlighted, each corresponding to some specific co-activating parent nodes. The probabilities of each node to belong to the three patterns are shown in the bar chart next to each node. Note that we allow the patterns to overlap and that the patterns are node specific.

Using the proposed pattern-based diffusion model, the diffusion process of a particular cascade proceeds as follows. 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. 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. Note that we assume a parent node to be able to activate its child node not just for the next immediate time step but also the subsequent ones up to a limit. To explain that, we define $C_s(w,t)$ as the set of nodes which have at least one activation within the interval between the latest activation of the node w in the s^{th} cascade denoted as $L_w^{(s)}(t+1)$ and the time step t. This assumes that we are only interested in recent news and that the posts earlier than our latest post have little influence on our future posting behaviour. $b(w) \cap C_s(w,t)$ then gives the subset of $C_s(w,t)$ which are parents of w.

Thus, the probability of a parent node v to activate its child node w $p(v|w)$ becomes an expected value of the diffusion probabilities $\{\tau_{z^w,w}\}$ over all the latent patterns based on $\{\pi_{v,z^w}\}$, that is $p(v|w) = \sum_{z^w=1}^{N_z(w)} \pi_{v,z^w}\tau_{z^w,w}$.

Then, the probability that the child node w will be activated at time $t+1$ is given as: $P_w^{(s)}(t+1) = 1 - \prod_{v \in b(w) \cap C_s(w,t)} (1 - \sum_{z^w=1}^{N_z(w)} \pi_{v,z^w}\tau_{z^w,w})$, and whether node w will be activated is determined accordingly. The process proceeds until there is no more node being activated and the cascade will stop.

The likelihood function of the observed cascades D_s can thus be formulated as:

$$L(\boldsymbol{\theta}) = \sum_{s=1}^{S} logP(D_s|\boldsymbol{\theta}, D_0^{(s)})$$
$$= \sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Bigg(\sum_{w \in D_s(t+1)} logP_w^{(s)}(t+1)$$
$$+ \sum_{w \notin D_s(t+1)}\sum_{v \in b(w) \cap C_s(w,t)} log(1 - \sum_{z^w=1}^{N_z(w)} \pi_{v,z^w}\tau_{z^w,w})\Bigg).$$

C. Learning Algorithm

We propose a two-level EM algorithm to maximize the likelihood function $L(\boldsymbol{\theta})$ with respect to the parameters $\boldsymbol{\theta} = \{\{\tau_{z^w,w}\}, \{\pi_{v,z^w}\}\}$ to infer the latent co-activation patterns and the diffusion probabilities.

1) First level EM: Let I_{v,z^w} be a latent variable that takes the value of 1 when a parent node v of a node w belongs to the latent pattern z^w, and 0 otherwise, given the constraint $\sum_{z^w=1}^{N_z(w)} I_{v,z^w} = 1$. Let $I = \{I_{v,z^w}\}$ denote the whole set of the latent variables. If we assume that I is known, the complete likelihood function can be written as:

$$P(D, I|\boldsymbol{\theta}) = P(D|I, \boldsymbol{\theta})P(I|\boldsymbol{\theta})$$

where

$$P(I|\boldsymbol{\theta}) = \prod_{w \in V}\prod_{v \in b(w)}\prod_{z^w=1}^{N_z(w)} \pi_{v,z^w}^{I_{v,z^w}}$$

and

$$P(D|I, \boldsymbol{\theta}) = L(\boldsymbol{\theta}|I)$$
$$= \sum_{s=1}^{S} logP(D_s|\boldsymbol{\theta}, D_0^{(s)}, I)$$
$$= \sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Bigg(\sum_{w \in D_s(t+1)} logP_w^{(s)}(t+1, I)$$
$$+ \sum_{w \notin D_s(t+1)}\sum_{v \in b(w) \cap C_s(w,t)} log(1 - \sum_{z^w=1}^{N_z(w)} I_{v,z^w}\tau_{z^w,w})\Bigg).$$

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 $\hat{\tau}_{z^w,w}$ and $\hat{\pi}_{v,z^w}$, given as

$$\eta_{v,z^w} = P(I_{v,z^w}=1|w,\hat{\theta}) = \frac{\hat{\tau}_{z^w,w}\hat{\pi}_{v,z^w}}{\sum_{z^w=1}^{N_z(w)} \hat{\tau}_{z^w,w}\hat{\pi}_{v,z^w}}.$$

Then, the expected likelihood function can be defined as:

$$\begin{aligned}
&\mathcal{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) \\
=&\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Bigg(\sum_{w\in D_s(t+1)} E_I[\log P_w^{(s)}(t+1,I)] \\
&+\sum_{w\notin D_s(t+1)}\sum_{v\in b(w)\cap C_s(w,t)}\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w}log(1-\tau_{z^w,w})\Bigg) \\
&+\sum_{w\in V}\sum_{v\in b(w)}\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w}log\pi_{v,z^w} \qquad (1)
\end{aligned}$$

For the M-step, we maximize $\mathcal{Q}$ by taking the derivative of $\mathcal{Q}$ with respect to θ to obtain the updating rule of the model parameters.

To update Π_{v,z^w}, according to the Lagrange multiplier method, maximizing $\mathcal{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ with the constraint $\sum_{z^w=1}^{N_z(w)}\pi_{v,z^w}=1$ yields

$$\frac{\partial\left(\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w}log\pi_{v,z^w}-\lambda(\sum_{z^w=1}^{N_z(w)}\pi_{v,z^w}-1)\right)}{\partial\pi_{v,z^w}}=0.$$

Then, it can be easily shown that $\forall z^w \quad \pi_{v,z^w}=\eta_{v,z^w}$.

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

2) Second level EM: Let $Y_{v,w}^{(s)}(t)$ 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 node v or not. We further define $Y_s=\{Y_s(0),Y_s(1)\cdots Y_s(T_s)\}$ where $Y_s(t):=\{Y_{v,w}^{(s)}(t)\}$ 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 $Y_{v,w}^{(s)}(t)$, given as

$$\begin{aligned}
\gamma_{v,w,s,t} =&P(Y_{v,w}^{(s)}(t+1)=1|w,\{\eta_{v,z^w}\},\hat{\theta}) \\
=&\frac{\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w}\hat{\tau}_{z^w,w}}{\hat{P}_w^{(s)}(t+1)}
\end{aligned}$$

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

$$\hat{P}_w^{(s)}(t+1)=1-\prod_{v\in b(w)\cap C_s(w,t)}(1-\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w}\hat{\tau}_{z^w,w})$$

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

$$\begin{aligned}
&\mathcal{Q}'(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) \\
=&\sum_{s=1}^{S}\sum_{t=0}^{T_s-1}\Bigg(\sum_{w\in D_s(t+1)}\sum_{v\in b(w)\cap C_s(w,t)}\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w} \\
&\big(\gamma_{v,w,s,t}log\tau_{z^w,w}+(1-\gamma_{v,w,s,t})log(1-\tau_{z^w,w})\big) \\
&+\sum_{w\notin D_s(t+1)}\sum_{v\in b(w)\cap C_s(w,t)}\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w}log(1-\tau_{z^w,w})\Bigg) \\
&+\sum_{w\in V}\sum_{v\in b(w)}\sum_{z^w=1}^{N_z(w)}\eta_{v,z^w}log\pi_{v,z^w}
\end{aligned}$$

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 $L_w^{(s)}(t+1)$. 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 $L_w^{(s)}(t+1)$. Moreover, we define a set of cascades where $T_{w,s}^{+}$ is not empty as $S_w^{+}=\{D_s:\exists v\big(v\in b(w)\wedge\exists t\big(v\in C_s(w,t)\wedge w\in D_s(t+1)\big)\big)\}$ and a set of cascades where $T_{w,s}^{-}$ is not empty as $S_w^{-}=\{D_s:\exists v\big(v\in b(w)\wedge\exists t\big(v\in C_s(w,t)\wedge w\notin D_s(t+1)\big)\big)\}$. Then $\partial\mathcal{Q}/\partial\tau_{z^w,w}=0$ yields:

$$\tau_{z^w,w}=\frac{1}{N_{z^w,w}^{+}+N_{z^w,w}^{-}}\sum_{s\in S_w^{+}}\sum_{t\in T_{w,s}^{+}}\sum_{v\in C_s(w,t)\cap b(w)}\eta_{v,z^w}\gamma_{v,w,s,t}$$

$$N_{z^w,w}^{+}=\sum_{s\in S_w^{+}}\sum_{t\in T_{w,s}^{+}}\sum_{v\in C_s(w,t)\cap b(w)}\eta_{v,z^w}$$

$$N_{z^w,w}^{-}=\sum_{s\in S_w^{-}}\sum_{t\in T_{w,s}^{-}}\sum_{v\in C_s(w,t)\cap b(w)}\eta_{v,z^w}.$$

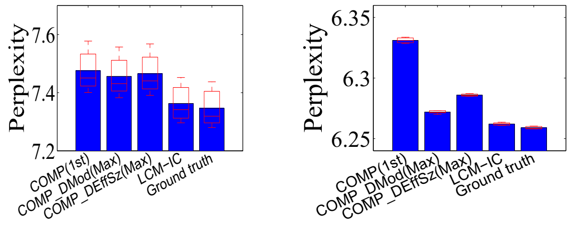
IV. Experiments

For performance comparison, we implement the proposed pattern-based diffusion model (*LCM-IC*), the basic IC model and three variants of a component-based diffusion model proposed in [8] where the parents are grouped based on their structural relations to exert influence. The three variants (*COMP(1st)*, *COMP_DMod(Max)*, *COMP_DEffSz(Max)*) differ in term of implementation details related to a structural diversity factor and a decay factor, which are not to be detailed here. Both synthetic and real social and information network data sets are used in our evaluation. We also visualize the results obtained based on a meme data set and illustrate how the co-activation patterns obtained can provide insights on the dependency of different news media in the news ecosystem regarding the news being released.

A. Experimental Settings

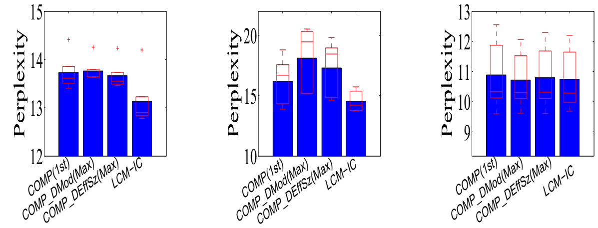
For all the experiments performed, the initial values of $\{\hat{\tau}_{z^w,w}\}$ are within $[0,0.1]$ as the diffusion probabilities in real data are known to be very small (*e.g.*, with a mean value of 0.04 and standard deviation of 0.07 [19]). And the initial values of $\{\hat{\pi}_{v,z^w}\}$ are generated within $[0,1]$ satisfying $\sum_{z^w}\hat{\pi}_{v,z^w}=1$. In this work, we obtain the optimal number of latent patterns per node with best performance using the cross-validation method.

As the ground-truth is unknown for real data, we use perplexity as the performance evaluation metric. Perplexity is widely used for evaluating language models [20], which calculates the average probability for each word to be



(a) $5,000$ edges (b) $10,000$ edges

Figure 2: Model comparison on synthetic data.



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

Figure 3: Performance comparison on three real data sets.

generated by the trained model. For our case, the *perplexity* over the cascades is defined as

$$Perplexity = \frac{-\sum_{s=1}^{S} \ln P(D_s)}{W}. \tag{2}$$

where $P(D_s)$ is the probability for the s^{th} cascade to be generated, and the normalization term 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, we divide the cascades into five folds and obtain the average performance using cross-validation.

B. Experiments on Synthetic Data

We first generate two scale-free networks with $1,000$ nodes using the snap platform [21], with $5,000$ and $10,000$ edges respectively. For each network, 100 cascades are generated based on our proposed model with $N_z(w) = 20$. For model initialization, all the parameters are randomly assigned under the constraints. Note that the network with $10,000$ edges is denser and thus there are more activations generated in the cascades available for inferring the model parameters. We apply our proposed LCM-IC model and the baseline models *COMP_DMod(Max)* and *COMP_DEffSz(Max)* to the synthetic networks. The performance of the basic IC model is much worse and thus its performance is not further reported. According to Figure 2, all the models perform better for the network with $10,000$ edges when compared to that with $5,000$ edges as anticipated due to the increased size of the training set. Also, the performance of our proposed LCM-IC model can approach the ground truth and is apparently better than the other baseline models. The performance ranking among all the models is consistent for both data sets.

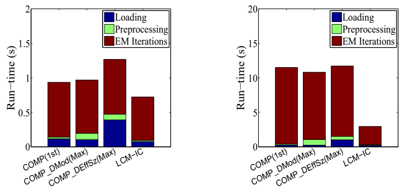
C. Experiments on Real Data

To validate that the proposed model is in fact modeling what is happening in the real diffusion processes, we apply the model to three real data sets. We use three real datasets MemeTracker [22], Digg [23] and Flixster [24] where both (1) the network structure and (2) information cascades are available. (i) The MemeTracker dataset covers a period of 9 months 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. A website A is assumed to have influence on another site B if a post in website B has referred to a post in A. Then, there will be a corresponding edge from A to B. The MemeTracker dataset contains 4 million nodes, 13 million edges, and $71,568$ cascades. (ii) The Digg dataset records the story voting process under a directed friendship network of users over one month in 2009. Users are modeled as nodes. A user A has influence on a user B if B is A's follower, modeled as an edge from user A to user B. The Digg dataset contains 280 thousand 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. In Flixster, if users A and B are friends, there is an undirected edge for nodes A and B. The Flixster dataset contains 787 thousand nodes and 5.9 million edges. We select $5,318$ cascades which correspond to the frequently rated movies in the dataset.

We apply again the proposed model and the baseline models to the three data sets. Figure 3 shows the performance comparison results. The optimal numbers of latent co-activation pattern $N_z(w)$ are evaluated by cross-validation separately for each node. The proposed LCM-IC model outperforms all the baseline models with a decrease in perplexity value of at least 0.54 and 1.66 for the MemeTracker and Digg datasets respectively. For the Flixster dataset, the proposed model gives comparable performance to the baseline models we tested. We find this encouraging given that the structural information for modeling detailed relation of parent nodes is not used at all in the proposed model.

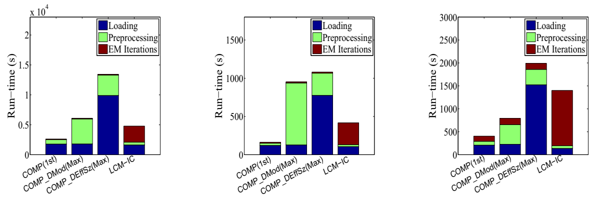
D. Run-time

To facilitate run-time comparison, we record the time for (1) loading the network and the cascades, (2) preprocessing the cascades and (3) running the EM algorithm, for each of the models as shown in Figures 4 and 5. The LCM-IC model takes shorter time in the first two steps for both synthetic and real data sets as the component-based models require the component information to be pre-computed. But it takes more in the third step as the co-activation patterns are to be inferred at the same time. For the overall run-time, the LCM-IC model is still more efficient than *COMP_DEffSz(Max)* and *COMP_DMod(Max)* when applied to Meme and Digg. However, the run-time for the EM iterations on Flixster is relatively long. Also, it is worth mentioning that the parameter estimation for each node is independent of each other, and thus can always be easily parallelized.



(a) $5,000$ edges (b) $10,000$ edges

Figure 4: Run-time comparison on synthetic data.



(a) Memes (b) Digg (c) Flixster

Figure 5: Run-time comparison on three real data sets.

E. Analysis of Dependency among News Media

As a case study, we apply the LCM-IC model to the online version of the New York Times (NYTimes) and demonstrate how the latent parent co-activation patterns identified can help understand the effectiveness of different news sources on NYTimes. To ease the result interpretation, given w to be the node corresponding to the NYTimes and v be one of its parents, we plot the values of $p(v|w)$ as shown in Figure 6 where the effect of the latent patterns are aggregated so as to compare the *overall* importance among the news sources (parent nodes). In addition, we plot the values of $p(v|z)$ over $\{v\}$ given different latent patterns so as to identify the pattern-specific influential news sources, and the values of $p(z|v)$ to identify the news sources unique to different co-activation patterns, as shown in Figure 7. In particular, we look into the details of only five patterns due to the page limit. Also, before the discussion, it is worth pointing out that an overall observation is that Boston.com makes significant contribution to quite some of the patterns even though it is in fact not as famous as other news sources. According to [25], Boston Globe was purchased by the NYTimes during the time period covered by the data set (the year 2008-2009). We believe that this accounts for its strong presence as one of the new sources of the NYTimes. In the sequel, we will exclude the discussion of the effect of Boston.com in the results. By referring to the plots shown in Figures 6 and 7, we made the following observations:

O1: The NYTimes is biased towards liberal news sources. It is well-known that the selective use of news sources in America is determined according to the newspapers' liberal or conservative [26]. The news sources detected with significant influence to the NYTimes are found to include MSNBC (msnbc.msn.com), the Washington Post (washingtonpost.com) and BBC (news.bbc.co.uk). They all have liberal bias, and so is the NYTimes.

Table I: Semantic contexts of parent co-activation patterns.

Pattern #	semantic context	major source(s)
Pattern 1	national news	CNN
Pattern 3	big events	both liberals and conservatives
Pattern 5	international events	IHT
Pattern 6	national political news	Washington Post
Pattern 10	national news	MSNBC, BBC

O2: The major news sources can be more reliably detected by referring to the latent co-activation patterns inferred. It is interesting to observe that the latent co-activation patterns inferred (as shown in Figure 7) match well with the sources preferred by liberals (other than the New York Times) including the Washington Post (Pattern 6, washingtonpost.com), the International Herald Tribune (Pattern 5, iht.com), BBC (Pattern 10, news.bbc.co.uk), CNN (Pattern 1, cnn.com) and MSNBC (Pattern 10, msnbc.msn.com), as stated in [26]. To better understand the importance of considering the co-activation patterns, one may refer to the plot of $p(v|w)$ in Figure 6. The $p(v|w)$ values of Free Republic (freerepublic.com), Reuters (reuters.com) and the Huffington Post (huffingtonpost.com) are found to be comparable to that of CNN. However, they are not commonly considered as major liberal news sources in the literature. Based on our experimental results, they are also not the dominating nodes in the co-activation patterns. We further examine the operation of the three news media. Free Republic is an *online forum* with frequent discussions among users. According to our results, it in fact does not characterize a unique co-activation pattern but co-activates with most of the major news sources, which accounts for the high values of $p(v|w)$. Reuters is a well-known news agency for *international news*. It co-activates specifically with IHT to cause the activations of the NYTimes. The Huffington Post is a *news aggregator and blog*, and it frequently co-activates with the Washington Post.

O3: Each parent co-activation pattern corresponds to a specific semantic context of the news sources. Each co-activation pattern corresponds primarily to a major liberal news source as the influential source. Also, most of the influential news sources found in each pattern also appear as the specific sources of the pattern. Together with the other less influential sources, they co-activate to define different parent co-activation patterns. Via careful examination of the cascades leading to the patterns, we found that each parent co-activation pattern corresponds to a specific semantic context as revealed by the associated news sources. Due to the page limit, we do not include the detailed discussion and instead list only the semantic contexts in Table I.

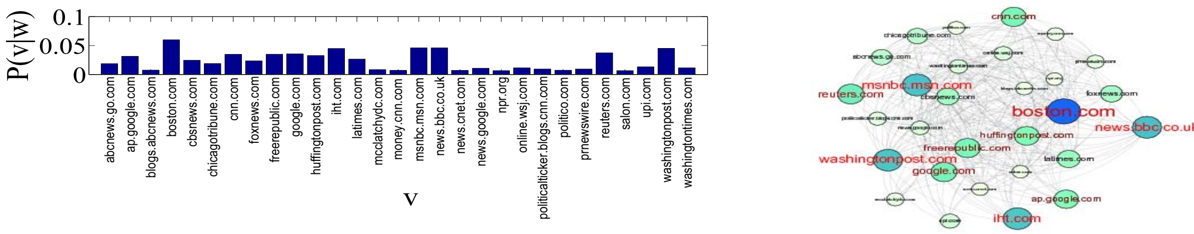


Figure 6: Visualization of the most influential parent nodes identified in the neighborhood of *nytimes.com*.

V. Conclusion

In this paper, we proposed a novel pattern-based information diffusion model for social networks where the latent co-activation patterns of parents for each node are inferred together with the pattern-based diffusion probabilities using the maximum likelihood approach. Results show that the diffusion model proposed in the current study achieves more accuracy than a number of variants of the IC model. Also, we have shown how the inferred co-activation patterns can be used to estimate the dependency among different online news media in terms of news diffusion. Possible extensions for future work include at least (1) incorporation of the cascades' context to infer context-aware parent co-activation patterns and diffusion networks, and (2) modeling the evolution of the latent patterns in the diffusion model.

Acknowledgment

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

References

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

[2] 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*, Brussels, Belgium, 2012, pp. 960–965.

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

[4] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf, "Uncovering the temporal dynamics of diffusion networks," in *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, WA, USA, 2011, pp. 561–568.

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

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

[7] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg, "Structural diversity in social contagion," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 5962–5966, Apr. 2012.

[8] Q. Bao, W. 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*, Atlanta, GA, USA, 2013, pp. 431–438.

[9] T. Hofmann and J. Puzicha, "Latent class models for collaborative filtering," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, San Francisco, CA, USA, 1999, pp. 688–693.

[10] K.-W. Cheung, K.-C. Tsui, and J. Liu, "Extended latent class models for collaborative recommendation," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 34, pp. 143–148, 2004.

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

[12] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li, "Social influence locality for modeling retweeting behaviors," in *Proceedings of the 23th International Joint Conference on Artificial Intelligence*, Beijing, China, 2013, pp. 2761–2767.

[13] N. Barbieri, F. Bonchi, and G. Manco, "Cascade-based community detection," in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, New York, NY, USA, 2013, pp. 33–42.

[14] ——, "Influence-based network-oblivious community detection," in *Proceedings of the 13th IEEE International Conference on Data Mining*, Dallas, TX, USA, 2013, pp. 955–960.

[15] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," in *Proceedings of the 26th Annual Conference on Neural Information Processing*, Lake Tahoe, Nevada, USA, 2012, pp. 548–556.

[16] H. Qin, T. Liu, and Y. Ma, "Mining user's real social circle in microblog," in *Proceedings of the 4th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Istanbul, Turkey, 2012, pp. 348–352.

[17] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, pp. 604–632, 1999.

[18] S. Chakrabarti, B. Dom, S. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, and J. Kleinberg, "Mining the web's link structure," *Computer*, vol. 32, pp. 60–67, 1999.

[19] 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*, New York, NY, USA, 2004, pp. 491–501.

[20] P. F. Brown, V. J. D. Pietra, R. L. Mercer, S. A. D. Pietra, and J. C. Lai, "An estimate of an upper bound for the entropy of English," *Computational Linguistics*, vol. 18, pp. 31–40, 1992.

[21] J. Leskovec. SNAP: Stanford network analysis platform. [Online]. Available: http://snap.stanford.edu/snap/index.html

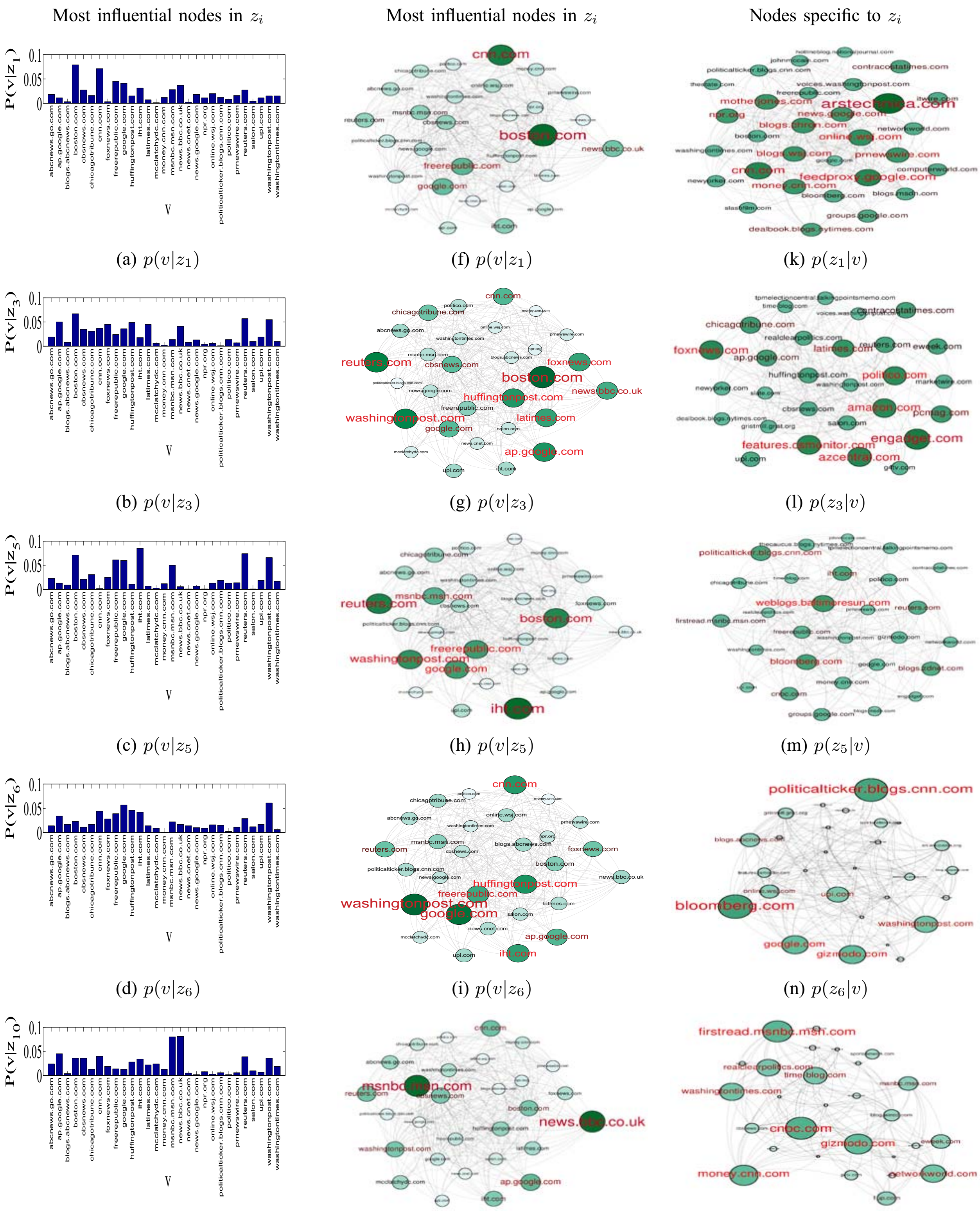
[22] J. Leskovec, L. Backstrom, and J. Kleinberg. MemeTracker: Download MemeTracker data. [Online]. Available: http://memetracker.org/data.html

[23] 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*, Washington, DC, USA, 2010, pp. 90–97.

[24] 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*, New York, NY, USA, 2010, pp. 135–142.

[25] The boston globe. [Online]. Available: https://en.wikipedia.org/wiki/The_Boston_Globe

[26] R. H. Davis, *Typing Politics : The Role of Blogs in American Politics*. Oxford University Press, USA, 2009.



(a) $p(v|z_1)$ (f) $p(v|z_1)$ (k) $p(z_1|v)$

(b) $p(v|z_3)$ (g) $p(v|z_3)$ (l) $p(z_3|v)$

(c) $p(v|z_5)$ (h) $p(v|z_5)$ (m) $p(z_5|v)$

(d) $p(v|z_6)$ (i) $p(v|z_6)$ (n) $p(z_6|v)$

(e) $p(v|z_{10})$ (j) $p(v|z_{10})$ (o) $p(z_{10}|v)$

Figure 7: Visualization of the patterns identified in the neighborhood of *nytimes.com.*