

Incorporating Structural Diversity of Neighbors in a Diffusion Model for Social Networks

Qing Bao, William K. Cheung and Yu Zhang
Department of Computer Science
Hong Kong Baptist University
Kowloon Tong, Hong Kong
 {qbao,william,yuzhang}@comp.hkbu.edu.hk

Abstract—Diffusion is known to be an important process governing the behaviours observed in network environments like social networks, contact networks, etc. For modeling the diffusion process, the Independent Cascade Model (IC Model) is commonly adopted and algorithms have been proposed for recovering the hidden diffusion network based on observed cascades. However, the IC Model assumes the effects of multiple neighbors on a node to be independent and does not consider the structural diversity of nodes' neighbourhood. In this paper, we propose an extension of the IC Model with the community structure of node neighbours incorporated. We derive an expectation maximization (EM) algorithm to infer the model parameters. To evaluate the effectiveness and efficiency of the proposed method, we compared it with the IC model and its variants that do not consider the structural properties. Our empirical results based on the MemeTracker dataset, shows that after incorporating the structural diversity, there is a significant improvement in the modelling accuracy, with reasonable increase in run-time.

Keywords-Social networks, structural diversity, diffusion network, Independent Cascade Model

I. INTRODUCTION

Diffusion is the phenomenon that an action or information spreads from one node to another via edges in a network. It is known to be an important process governing the behaviours observed in network environments like social networks, contact networks, etc. The Independent Cascade Model (IC Model) [1] and the Linear Threshold Model (LT Model) [2] are two widely used diffusion models for social networks. The IC Model [1] assumes that a node can be activated by just one of its neighbors in the diffusion process and the effects of multiple neighbors on a node are assumed to be independent. The LT Model [2] assumes that whether a node will be activated depends on the aggregation of its neighbors' activations. In this paper, we focus on the IC model and its extensions for more accurate modeling of diffusion processes in social networks. In the literature, the simple IC model has been extended in various ways. For instance, one can consider the time information and model the influence of a node activation to be discounted by an exponential decay, based on the observation that most people are more interested in recent news [3, 4]. Also, the basic IC model assumes that the transmission rate over

the whole network is the same and constant. By relaxing the assumption, some recent work assumes the transmission rates to be dynamic [5, 6]. However, most of the existing work does not consider the effect of structural diversity of the neighbors on node activation.

All diffusion models come with the notion of neighborhood via which the information is spread. For neighbors with direct connections (also called ties) with an individual, they could exhibit different forms of influence which may depend on their structural properties in the social network. For example, in the literature, there have been studies on the relationship between the presence of different local ties and the overall network properties. The importance of weak ties acting as "local bridges" in facilitating effective information diffusion in social networks has long been recognized [7]. Recently, related perspectives have been taken for online communication and social media analysis [8], [9], [10]. For example, J.-P. Onnela *et al.* [8] studied the presence of strong and weak ties in mobile communication networks and showed that random removal of weak ties could end up with networks falling apart, no longer supporting the communication. In [9], it was demonstrated that considering the connected components of the neighbors correlates well with the probability of forming social coalition. So, it is not how many friends are influencing you that matters but how many loosely coupled "groups" (also referred to as non-redundant contacts in [11]). In [10], online social ties in heterogeneous networks are also investigated.

In this paper, we focus on the diffusion network inference problem using the IC Model and argue that the structural diversity of nodes' neighborhood should be incorporated into the diffusion model. An extended IC model is proposed and the corresponding inference algorithm based on the EM algorithm is derived. Given the inferred structure, a number of network analysis tasks like influence maximization [12, 13] can also be carried out. Note that unlike some related work where the network structure is not known [14], we assume that the network structure is known and infer the diffusion probabilities based for a set of observed cascades. Also, for

the simplicity sake, we consider only static transmission rates and assume that activations only happen at discrete time steps.

The remainder of this paper is organized as follow. Section II presents related work on variants of the IC Model. Section III describes the proposed model as well as the algorithm for inferring the model parameters. The experimental results can be found in Section IV. Section V concludes the paper and provides pointers for future work.

II. RELATED WORK

The following summarizes related work presented in the context of contagion. According to [15], contagion can be categorized as simple and complex ones. With simple contagion, only one source is sufficient for diffusion while complex contagion is related to collective behaviours that require social affirmation from multiple sources. The IC Model [1] models diffusion processes where a node can be activated by just one neighbor and the effect of the neighbors on a node are assumed to be independent. Thus, one can consider the IC Model corresponds to the mechanism for simple contagion. The LT Model [2] assumes a node being activated with respect to the sum of the neighbors' aggregation. Thus, the LT Model is for complex contagion.

In the literature, there exist many variants to the basic IC Model. In [3], [4], the influence was modelled with exponential decay, based on the observation that most people are more interested in recent news. Each node is allowed to be activated multiple times in [16], [17], [18] for the reason that a user can post several posts on the same topic instead of just one. In [3], [4], and [19], the IC model was extended to allow continuous trials by neighbors instead of just once based on the assumption that one could be motivated when he/she revisits the post, even though not motivated before. Besides, the IC Model for modelling the propagation of competing opinions has been proposed in [20], [21]. Also, some recent work assumes the transmission rates of diffusion to be dynamic [5, 6]. Satio *et al.* [22], [23] proposed an asynchronous time model where activations can happen in continuous time.

III. A DIFFUSION MODEL WITH NEIGHBORS STRUCTURAL DIVERSITY INCORPORATED

A. Preliminaries

We represent a given network as a directed graph $G = (V, E)$ with V being the set of nodes and E being the set of edges. Let $e = (v, w)$ denote an edge from node v to node w , and $f(v)$ and $b(v)$ denote the sets of child and parent nodes for node v respectively, given as: $f(v) = \{w : (v, w) \in E\}$ and $b(v) = \{u : (u, v) \in E\}$. For each node w , we define the connected components of its parent nodes as *parent components* $B(w) = \{B_i(w) : i = 1, \dots, N_c(w)\}$ where $N_c(w)$ is the number of components corresponding to $b(w)$. And reversely, for a component c , we define the

set of nodes w which has component c as one of its parent components as $F(c) = \{w : c \in B(w)\}$.

With the assumption that the component structure of the parents of each node is static, we define for each component-node pair $(c \in B(w), w)$ a *component-based diffusion probability* $\tau_{c,w}$ with $0 \leq \tau_{c,w} \leq 1$. We also allow a node to be activated multiple times.

Fig. 1 illustrates a node w and its neighborhood $b(w) = \{v_1, v_2, v_3, v_4, v_5, v_6\}$. The neighborhood forms two components, that is $B(w) = \{B_1(w) = C_1, B_2(w) = C_2\}$. Each component consists of interconnected nodes, *i.e.*, $C_1 = \{v_1, v_2, v_3\}$ and $C_2 = \{v_4, v_5, v_6\}$. Thus, the component-based diffusion probability are $\tau_{C_1,w}$ and $\tau_{C_2,w}$ respectively.

B. Problem Formulation

Let $D_s = \{D_s(0), D_s(1) \dots D_s(T_s)\}$ be the s^{th} information diffusion cascade observed where $D_s(t)$ is the set of nodes activated at time step t and T_s is the final time step for cascade D_s . In our proposed diffusion model, the diffusion process of a particular cascade proceeds as follow. Given the initial set of activated nodes in the s^{th} cascade ($D_s(0)$), we assume that they can cause the nodes with their "parent" components containing at least one of the nodes in $D_s(0)$ to be activated. We define $C_s := \{C_s(t)\}$ where $C_s(t)$ refers to the set of components active at time step t in the s^{th} cascade. A component is active at time step t if it has at least one node in it which is activated on or before time step t . In general, we assume that the parent component of a node w has influence on node w at time step t if the component contains at least one node activated after the latest activation of the node w . In the context of social networks, one can interpret this as the fact that we are only interested in recent news and that the posts before our previous excitement have little influence. Then, based on the diffusion probabilities $\{\tau_{c,w}\}$ associated to the active components, some of the corresponding nodes w will be activated accordingly. The process continues until there are no more nodes being activated and the cascade stops. To

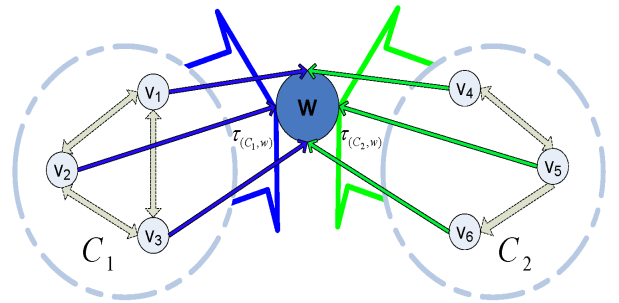


Figure 1. Node-based vs. component-based diffusion.

infer an accurate diffusion model, it is expected that the overall probability of generating all the observed cascades $\{D_s\}$ should be maximized.

In this paper, we investigate factors which may govern the effectiveness of the component's activation caused by the activation of its element nodes. In particular, we postulate that within a parent component, the degree of influence on the node with the component as its parent could be affected by (1) when its element nodes are activated, and (2) the structural diversity of the component. We represent the two factors as $F_a^{(s)}(c, w, t) \in [0, 1]$ and $F_b^{(s)}(c, w, t) \in [1, n_c]$ respectively where n_c is the number of nodes in the component c , and their combined effect as

$$N_s(c, w, t) = F_a^{(s)}(c, w, t)F_b^{(s)}(c, w, t). \quad (1)$$

$N_s(c, w, t)$ takes values between 0 and n_c and is here interpreted as the effective count of an observed component activation. In other words, instead of considering each observed component activation equally in the subsequent model inference algorithm, we allow the counting of each activation to be "weighted" by $N_s(c, w, t)$.

For the factor $F_a^{(s)}(c, w, t)$, with reference to node w , we argue that an activated parent component with its first node activation happening more recently will be more influential compared to the components with their first node activations happening earlier. In the context of social networks, we assume that ones are more interested in recent news items than older ones. Using an exponential decay, we define for a node w which may get activated at time step $t + 1$ $F_a^{(s)}(c, w, t)$ as

$$F_a^{(s)}(c, w, t) = e^{-(t - T_c^{(s)}(w, t))/\alpha}. \quad (2)$$

where $T_c^{(s)}(w, t)$ is the time of first node activation in component c which falls in the interval $[L_w^{(s)}(t+1), t]$. Here $L_w^{(s)}(t+1)$ defines the latest activation time of node w before time $t + 1$. If none of the activations fall into the interval, we consider that the component c has no influence on node w and $F_a^{(s)}(c, w, t)$ equals to 0.

The examples illustrated by Figs. 2 and 3 show how to calculate $T_c^{(s)}(w, t)$. As shown in Fig. 2, the activations of the component $C_1 = \{v_1, v_2, v_3\}$ are enumerated as the union of activations of the three nodes. v_1 is activated at time steps $\{1, 3\}$, v_2 at time steps $\{2, 5\}$, v_3 at time step $\{6\}$. Hence the activations of C_1 are considered to happen at $\{1, 3\} \cup \{2, 5\} \cup \{6\} = \{1, 2, 3, 5, 6\}$. Then, given that the activations of the component's child node w happen at time steps $\{2, 7\}$, as shown in Fig. 3, we can get $L_w^{(s)}(2) = 0$ (If there are no previous activations for node w , we set the value as 0.), and $L_w^{(s)}(7) = 2$. At time step 2, the first activation of component C_1 within the interval $[L_w^{(s)}(2), 2 - 1] = [0, 1]$ (which may cause the activation of w at time step 2), i.e., $T_c^{(s)}(w, 1)$, is 1. At time step 7, the first activation of

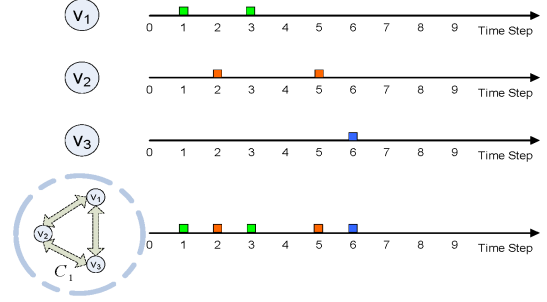


Figure 2. Node and component activations.

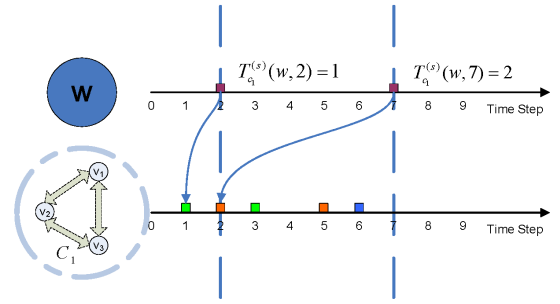


Figure 3. Illustrating how node w is affected by recent component activations.

component C_1 within the interval $[L_w^{(s)}(7), 7 - 1] = [2, 6]$, i.e., $T_c^{(s)}(w, 6)$, is 2.

For the factor $F_b^{(s)}(c, w, t)$, we argue that the information provided by subsequent activations of nodes within a more tightly connected component will be more redundant. In social networks, we assume that the news provided by similar websites are relatively redundant. We use modularity which is well-known for its use in community detection [24, 25] to measure the tightness of connections within one component. The modularity Q is formally defined as the fraction of edges that fall within communities subtracted from it the expected value of the same quantity in terms of randomized network with the same node degree distribution [24]: $Q = \frac{1}{2m} \sum_{vw} [Adj_{vw} - \frac{k_v k_w}{2m}] \delta(c_v, c_w)$ where Adj_{vw} equals to 1 if v and w are connected, and 0 otherwise, $\delta(i, j)$ is a δ -function which gives 1 if $i = j$ and 0 otherwise, m is the number of edges in the network, k_v (k_w) is the out-degree of node v (w) and c_v (c_w) is the community of node v (w). This definition can be rewritten as

$$Q = \sum_c Q_c = \sum_c \frac{1}{2m} \left(\sum_{v \in c} \sum_{w \in c} Adj_{vw} - \frac{(\sum_{v \in c} k_v)^2}{2m} \right). \quad (3)$$

Q_c will reach the maximum value when component c is fully connected, which gives $Q_c^{max} = \frac{1}{2m}(n_c(n_c - 1) - \frac{(\sum_{v \in c} k_v)^2}{2m})$ where n_c denotes the number of nodes in community c . And Q_c will reach the minimum value when there is no connection in a component, which gives $Q_c^{min} = -\frac{1}{4m^2}(\sum_{v \in c} k_v)^2$.

We use $-Q_c$ as structural metric for component c to model that more tightly connected component will have discounted influence due to the redundancy. To compute $F_b^{(s)}(c, w, t)$, instead of normalizing the value of $-Q_c$ to $[-Q_c^{max}, -Q_c^{min}]$, we normalize them to the range $[1, n_c]$, which gives

$$F_b^{(s)}(c, w, t) = 1 + \frac{(n_c - 1)(Q_c^{max} - Q_c)}{Q_c^{max} - Q_c^{min}} = n_c - \frac{2|E_c|}{n_c}. \quad (4)$$

where $|E_c|$ is the number of edges in component c .

We thus calculate the probability that the node w becomes active at time $t + 1$ as

$$P_w^{(s)}(t + 1) = 1 - \prod_{c \in B(w) \cap C_s(t)} (1 - \tau_{c,w})^{N_s(c,w,t)}. \quad (5)$$

where $N_s(c, w, t)$ is the effective number of activations due to component c influencing node w at time t , and $C_s(t)$ is the set of components active at time step t in the s^{th} cascade.

Given $D = \{D_s : s = 1, \dots, S\}$ be the set of independent information diffusion cascades, the log-likelihood function with respect to θ can be defined as

$$\begin{aligned} L(\theta) &= \sum_{s=1}^S \log P(D_s | \theta, D_0^{(s)}) \sum_{s=1}^S \log L(\theta; D_s) \\ &= \sum_{s=1}^S \sum_{t=0}^{T_s-1} \left(\sum_{w \in D_s(t+1)} \log P_w^{(s)}(t+1) \right. \\ &\quad \left. + \sum_{c \in C_s(t)} \sum_{w \in F(c) \setminus D_s(t+1)} N_s(c, w, t) \log(1 - \tau_{c,w}) \right) \end{aligned} \quad (6)$$

where $D_0^{(s)}$ are the nodes which are activated with absence of parent components' activations, *i.e.*, activated by itself and serve as the sources in the s^{th} cascade. Since there could be multiple paths from multiple sources in a cascade, D_s^0 could consist of more than one node.

Then, the remaining problem is to estimate $\theta = \{\tau_{v,w}\}$ so as to maximize (6).

C. Learning Algorithm

We first find the parent components for each node according to the neighbourhood network structure. Here we use the community detection method "Clauset-Newman-Moore" [24], which has been applied to large networks

and the detected communities have considerable number of nodes. We then apply the Expectation-Maximization (EM) [26] algorithm to estimate the model parameters. We define latent variables $Y_{c,w}^{(s)}(t)$ to indicate whether the activation of a node w at time step t in the s^{th} cascade is caused by w 's parent component c or not. Similar to the definition for D_s , we represent the corresponding set of latent variables as $Y_s = \{Y_s(0), Y_s(1) \dots Y_s(T_s)\}$ where $Y_s(t) = \{Y_{c,w}^{(s)}(t)\}$ is the set of latent variables for all the activations at time step t for the s^{th} cascade, which is used to represent whether or not c activated w by setting $Y_{c,w}^{(s)}(t)$ as 1 or 0.

Based on the latent variables defined, we derive the Q -function and estimate the model parameters via the EM algorithm which consists of an E-step and an M-step.

1) *E-step*: Take expectation of all possible assignments of Y which can explain the observed cascades.

Given that for a node $w \in D_s(t + 1)$ and its parent component $c \in C_s(t)$, the probability of successful activation is $(1 - (1 - \tau_{c,w})^{N_s(c,w,t)})$ given $Y_{c,w}^{(s)}(t + 1) = 1$ and $(1 - \tau_{c,w})^{N_s(c,w,t)}$ given $Y_{c,w}^{(s)}(t + 1) = 0$. The Q -function can be formulated as:

$$\begin{aligned} Q(\theta | \hat{\theta}) &= \sum_{s=1}^S \sum_{t=0}^{T_s-1} E_Y[\log P(D_s(t+1), Y_s(t+1) | \theta) | D, \hat{\theta}] \\ &= \sum_{s=1}^S \sum_{t=0}^{T_s-1} \sum_{c \in C_s(t)} \left(\sum_{w \in F(c) \cap D_s(t+1)} E_{Y_{c,w}^{(s)}(t+1)} \right. \\ &\quad \left[\log \left((1 - (1 - \tau_{c,w})^{N_s(c,w,t)}) Y_{c,w}^{(s)}(t+1) \right. \right. \\ &\quad \left. \left. + (1 - \tau_{c,w})^{N_s(c,w,t)} (1 - Y_{c,w}^{(s)}(t+1)) \right) \right] \\ &\quad \left. + \sum_{w \in F(c) \setminus D_s(t+1)} N_s(c, w, t) \log(1 - \tau_{c,w}) \right). \end{aligned} \quad (7)$$

Also, the probability for the activated parent component c of the node w to succeed in activating w at time step $t + 1$, *i.e.*, $P(Y_{c,w}^{(s)}(t + 1) = 1)$, can be calculated as $\frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t)}}{\hat{P}_w^{(s)}(t+1)}$ where $\hat{\tau}_{c,w}$ stands for the current estimate of $\tau_{c,w}$ and $\hat{P}_w^{(s)}(t + 1)$ can be calculated using (5). Therefore,

$$\begin{aligned} Q(\theta | \hat{\theta}) &= \sum_{s=1}^S \sum_{t=0}^{T_s-1} \sum_{c \in C_s(t)} \left(\sum_{w \in F(c) \cap D_s(t+1)} \right. \\ &\quad \left(\frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t)}}{\hat{P}_w^{(s)}(t+1)} \log(1 - (1 - \tau_{c,w})^{N_s(c,w,t)}) \right. \\ &\quad \left. \left. + \left(1 - \frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t)}}{\hat{P}_w^{(s)}(t+1)} \right) N_s(c, w, t) \log(1 - \tau_{c,w}) \right) \right) \end{aligned}$$

$$+ \sum_{w \in F(c) \setminus D_s(t+1)} N_s(c, w, t) \log(1 - \tau_{c,w}) \Bigg). \quad (8)$$

2) *M-step*: Solve the optimality condition $\partial \mathcal{Q} / \partial \tau_{c,w} = 0$ for the new estimate of $\tau_{c,w}$.

We define $T_{c,w,s}^+$ as the set of time steps $\{t\}$ with reference to the s_{th} cascade satisfying the condition that node w is activated at t and its parent component c has been activated since $L_w^{(s)}(t)$. $T_{c,w,s}^-$ is the set of time steps $\{t\}$ where node w is not activated at t , but its parent component c has been activated since $L_w^{(s)}(t)$. Moreover, we define a set of cascades where $T_{c,w,s}^+$ is not empty as $S_{c,w}^+ = \{D_s : \exists t(c \in C_s(t) \wedge w \in D_s(t+1))\}$ and a set of cascades where $T_{c,w,s}^-$ is not empty as $S_{c,w}^- = \{D_s : \exists t(c \in C_s(t) \wedge w \notin D_s(t+1))\}$. Thus the derivation is given as

$$\begin{aligned} & \partial \mathcal{Q} / \partial \tau_{c,w} = 0 \\ \Rightarrow & \sum_{s \in S_{c,w}^+} \sum_{t \in T_{c,w,s}^+} \left(\frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t-1)}}{\hat{P}_w^{(s)}(t)} \right. \\ & \frac{N_s(c,w,t-1)(1 - \tau_{c,w})^{N_s(c,w,t-1)-1}}{1 - (1 - \tau_{c,w})^{N_s(c,w,t-1)}} \\ & \left. + \left(1 - \frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t-1)}}{\hat{P}_w^{(s)}(t)} \right) N_s(c,w,t-1) \frac{1}{\tau_{c,w} - 1} \right) \\ & + \sum_{s \in S_{c,w}^-} \sum_{t \in T_{c,w,s}^-} N_s(c,w,t-1) \frac{1}{\tau_{c,w} - 1} = 0 \\ \Rightarrow & \sum_{s \in S_{c,w}^+} \sum_{t \in T_{c,w,s}^+} \left(\frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t-1)}}{\hat{P}_w^{(s)}(t)} \right. \\ & \left. N_s(c,w,t-1) \frac{1}{1 - (1 - \tau_{c,w})^{N_s(c,w,t-1)}} \right) = N_{c,w} \\ & N_{c,w} = N_{c,w}^+ + N_{c,w}^- \\ & N_{c,w}^+ = \sum_{s \in S_{c,w}^+} \sum_{t \in T_{c,w,s}^+} N_s(c,w,t-1) \\ & N_{c,w}^- = \sum_{s \in S_{c,w}^-} \sum_{t \in T_{c,w,s}^-} N_s(c,w,t-1). \end{aligned} \quad (9)$$

Because the function

$$f(\tau_{c,w}) = \sum_{s \in S_{c,w}^+} \sum_{t \in T_{c,w,s}^+} \left(\frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t-1)}}{\hat{P}_w^{(s)}(t)} \right. \\ \left. N_s(c,w,t-1) \frac{1}{1 - (1 - \tau_{c,w})^{N_s(c,w,t-1)}} \right) - N_{c,w}$$

is monotonic¹, we can use the bisection method to get the solution $\tau_{c,w}$ for $f(\tau_{c,w}) = 0$. For our case, the starting interval to solve for $\tau_{c,w}$ is set to $[0, 1]$ satisfying the condition $f(0)f(1) \leq 0$. Since

$$\begin{aligned} f(1) &= \sum_{s \in S_{c,w}^+} \sum_{t \in T_{c,w,s}^+} \left(\frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c,w,t-1)}}{\hat{P}_w^{(s)}(t)} \right. \\ & \left. N_s(c,w,t-1) \right) - N_{c,w}^+ - N_{c,w}^- \\ &\leq N_{c,w}^+ - N_{c,w}^+ - N_{c,w}^- \\ &\leq 0. \\ \lim_{x \rightarrow 0} f(x) &= \infty, \end{aligned}$$

the condition is thus satisfied.

The E-step and M-step repeat until convergence. The detailed steps for learning the model is summarized in Algorithm 1.

IV. EXPERIMENTS

A. Data Set

To evaluate the effectiveness of the proposed model, we made use of the MemeTracker [27] dataset which can give both the network structure for websites of news articles and blog posts (red and blue points in Fig. 4) and cascades. The data covers a period from August 1 2008 to April 30 2009.

Websites denoted by nodes are connected by directed edges. A website A is assumed to have influence on another site B if a post in B has mentioned a post in A . The dataset contains 4 million nodes and 13 million edges. The parent nodes of a node form multiple meaningful components. For example, the website “ksat.com”, a local news website for San Antonio, has parent websites grouped by topics such as general news (“news.bbc.co.uk”, “cnn.com”), business news (“economist.com”, “forbes.com”), sports (“sports.espn.go.com”) and so on. Each group of websites represents an information source. A website can be influenced by different sources. To gain further insight about the dataset, we randomly sampled 300,000 websites, about 10 percent of the complete dataset. Among these websites, 93,675 (31.2%) websites have at least one parent. A parent component is different from a parent node only if it contains more than one nodes. And out of the 93,675 websites, 49,764 (53.1%) websites have at least two parent components containing more than one node, which shows that modelling structural diversity is meaningful for the MemeTracker dataset.

¹Note that $(1 - \tau_{c,w})^{N_s(c,w,t-1)}$ is monotonic because $N_s(c,w,t-1) \log(1 - \tau_{c,w})$ is monotonic.

Algorithm 1 Incorporating Structural Diversity in ICM

Input: network $G = (V, E)$, cascades $D = \{D_1, \dots, D_S\}$
Output: parent components $\{c\}$ for each node $w \in V$,
component-based diffusion probabilities $\theta = \{\tau_{c,w}\}$

Detect parent components $\{c\}$ for each node $w \in V$
Assign initial values to $\hat{\theta} = \{\hat{\tau}_{c,w}\}$

for all (c, w) pairs **do**

$$N_{c,w}^+ \leftarrow \sum_{s \in S_{c,w}^+} \sum_{t \in T_{c,w,s}^+} N_s(c, w, t - 1)$$

$$N_{c,w}^- \leftarrow \sum_{s \in S_{c,w}^-} \sum_{t \in T_{c,w,s}^-} N_s(c, w, t - 1)$$

if $N_{c,w}^+ = 0$ **and** $N_{c,w}^- \neq 0$ **then**

$$\tau_{c,w} \leftarrow 0$$

end if //special cases for diffusion probabilities

end for

while not convergence **do**

E-step:

for all $\hat{P}_w^{(s)}(t)$ **do**

$$\hat{P}_w^{(s)}(t) \leftarrow 1 - \prod_{c \in B(w) \cap D_s(t-1)} (1 - \hat{\tau}_{c,w})^{N_s(c, w, t-1)}$$

end for

M-step:

for all $(c, w) : S_{c,w}^+ \neq \emptyset$ **do**

calculate $\tau_{c,w}$ using the bisection method for function

$$\sum_{s \in S_{c,w}^+} \sum_{t \in T_{c,w,s}^+} \left(\frac{1 - (1 - \hat{\tau}_{c,w})^{N_s(c, w, t-1)}}{\hat{P}_w^{(s)}(t)} \right) - N_{c,w} = 0.$$

end for

$$\hat{\theta} \leftarrow \theta$$

end while

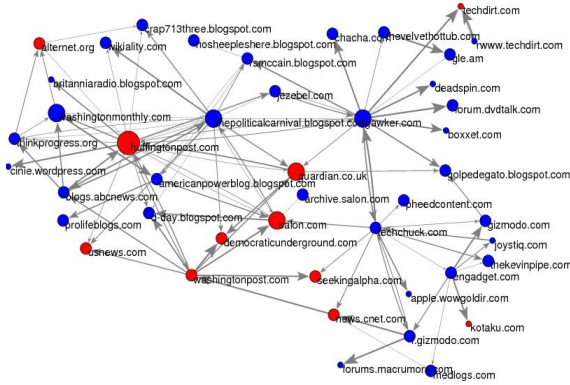


Figure 4. A small part of the MemeTracker network [27].

B. Experimental Settings

While we considered all 4 million nodes in computing the parent components, we evaluated only the diffusion probabilities of 1,000 randomly chosen nodes with non-zero in-degree, and applied the algorithm to infer the diffusion probabilities for their parents. Since the estimation

of the diffusion probabilities for each node is independent given (9), we anticipate that similar results can be obtained for the whole set of nodes, which will be evaluated in the future. Also, we considered different number of cascades (17500, 35000, 52500, 71568). For each setting, we randomly divided the cascades into five folds and reported the average performance based on five-fold cross-validation.

We compare our designed model with three different versions of IC models, given as

- *Model 1:* The IC Model,
- *Model 2:* The IC Model with decay effect,
- *Model 3:* The IC model with decay effect and un-weighted structural diversity, and
- *Model 4:* Our proposed model.

Model 1 is a baseline model which simply extends the IC Model by considering parent nodes which are activated after the latest activation of the child node instead of only the last time step based on the intuition that readers are not only interested in news just happened, but also those happened some time ago. The number of parameters for each node in *Model 1* equals to the number of its parent nodes.

Model 2 extends *Model 1* by considering decay of influence. One reason to consider the decay is that an activated parent node with its first activation happening more recently will be more influential compared to the parent nodes with their first activations happening earlier. Using the same exponential decay, we define $F_a^{(s)}(v, w, t)$ for a parent node v and its child node w in the s^{th} cascade at time step $t + 1$ as: $F_a^{(s)}(v, w, t) = e^{-(t - T_v^{(s)}(w, t))/\alpha}$ where $T_v^{(s)}(w, t)$ is the time of first activation of node v which falls in the interval $[L_w^{(s)}(t+1), t]$. And thus we can define the weighted counting $N_s(v, w, t) = F_a^{(s)}(v, w, t)$. The number of parameters for each node in *Model 2* again equals to the number of its parent nodes.

Model 3 extends *Model 2* by grouping parent nodes based on their community structure. *Model 4* extends *Model 3* by considering the modularity measure as weighting for each component. Thus, *Model 3* differs from *Model 4* by setting $F_b^{(s)}(c, w, t) = 1$ instead of (4). The number of parameters for each node in *Model 3* and *Model 4* equals to the number of its parent components.

Here we used $\alpha = 1,500$ for all related models. Also, according to [28], the diffusion probabilities found in real-world datasets are quite small with mean 0.04 and standard deviation 0.07. Thus we set initial assignments of $\theta = \{\hat{\tau}_{c,w}\}$ within $[0, 0.1]$ and we used the same set of values for all models. To more accurately detect the neighbour community structure, we considered nodes within two hops instead of just the immediate neighbours.

We used perplexity to evaluate different models as the ground-truth is unknown. Perplexity is commonly used in the evaluation of language models [29]. It measures the average probability of each word based on the trained

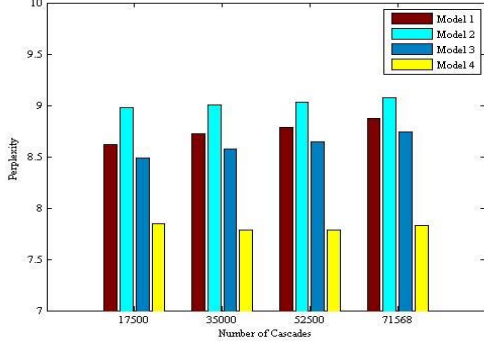


Figure 5. Comparison of model accuracy in terms of perplexity.

model. For our case, the *perplexity* over the cascades is here defined as

$$Perplexity = \frac{-\sum_{s=1}^S \log P(D_s)}{W}. \quad (10)$$

$P(D_s)$ is the probability for the s^{th} cascade to be generated, S is the total number of cascades, and the normalization term W is the number of activations with the existence of activations from parent components. A smaller perplexity value indicates a higher average probability under the model, and thus better performance.

C. Results

1) *Generative Ability*: Fig. 5 shows the average perplexity under five-fold cross validation for the four models we previously described. For each model, we found that the performance remains stable given the different numbers of cascades we tested. This shows that the difference caused by different numbers of activations does not lead to apparent difference in performance. Also, we found that after considering decay, component structure and component modularity measure, there was apparent improvement on the performance. The comparable performance among *Model 1*, *Model 2* and *Model 3* shows that considering only decay and component structure alone cannot accurately model diffusion. After adding component modularity which models the structural properties in components, the performance was significantly improved by a decrease in perplexity of 0.94. This demonstrates that the structure diversity, modelled by community structure, plays an important role in the diffusion process and the community structure cannot be characterized by the properties of individual neighbours alone. And after adding modularity as weighting to each parent component, our proposed model mimicked the phenomenon that parent nodes within a more tightly connected component share more information and thus provide more redundant information. While the model with unweighted structural diversity always regards the influence under all structures equally.

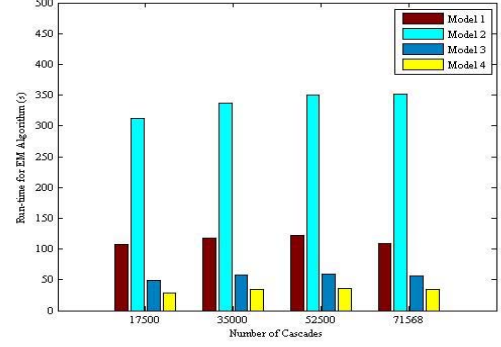


Figure 6. Comparison of run-time for the EM iterations.

2) *Run-time*: The run-time for the EM iterations measured in seconds under each model is shown in Fig. 6. The maximum run-time for the four models were 121.64s, 351.59s, 58.96s and 36.40s respectively. After considering decay, the run-time increased apparently because the bisection method needs a number of iterations to converge. After further considering the parent components, the run-time significantly decreased again because the number of parent components is much smaller than the number of parent nodes.

The total run-time performance for the four models is shown in Fig. 7 where all the pre-computing steps are included. The run-time was comparably longer because the calculation of the weighted counting $N_s(c, w, t)$ from the activation logs is time-consuming. For the models with parent components among larger numbers of cascades, although the number of parent components is much smaller than the number of parent nodes which reduces the run-time, we observed yet some increase in run-time due to the calculation of $N_s(c, w, t)$ which involves integration of nodes' activations in each parent component. Note that as mentioned above, the estimation of diffusion probabilities $\theta = \{\tau_{c,w}\}$ is independent given (9). Therefore, we can estimate the diffusion probabilities in parallel, and the run-time can then be scaled down by the number of processes running at the same time.

V. CONCLUSION

In this paper, we proposed an extension of IC Model which incorporates the community structure of nodes' neighbours for diffusion modeling. The EM algorithm is used to estimate the model parameters. We compared the proposed model with the benchmark models that do not consider network structural properties. Our empirical result on a real dataset shows that after incorporating structural diversity, there is a significant improvement in the generative ability of the proposed model. The new model can more accurately model diffusion processes in social networks. Our work has

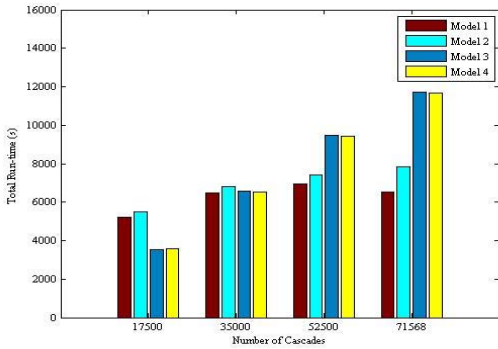


Figure 7. Comparison of overall run-time.

a number of ways for enhancement. First, more structural properties can be utilized and combined with the activation logs. Second, it can be applied to enhance some network analysis tasks (e.g., influence maximization) via the incorporation of the structural diversity.

ACKNOWLEDGMENT

This work is partially supported by HKBU Strategic Development Fund SDF10-0526-P08.

REFERENCES

- [1] J. Goldenberg, B. Libai, and Muller, "Using complex systems analysis to advance marketing theory development," *Academy of Marketing Science Review*, 2001.
- [2] 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 (ICALP'05)*, Berlin, Heidelberg, 2005, pp. 1127–1138.
- [3] 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 (ICDM'12)*, Brussels, Belgium, 2012, pp. 960–965.
- [4] 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 (WSDM'10)*, New York, NY, USA, 2010, pp. 241–250.
- [5] 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 (ICML'11)*, Bellevue, WA, USA, 2011, pp. 561–568.
- [6] M. Gomez-Rodriguez and B. Schölkopf, "Influence maximization in continuous time diffusion networks (ICML'12)," in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012, pp. 313–320.
- [7] M. S. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology*, vol. 78, pp. 1360–1380, 1973.
- [8] J. Saramäki and J. pekka Onnela, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, pp. 7332–7336, 2007.
- [9] 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.
- [10] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogeneous networks," in *Proceedings of the 15th ACM International Conference on Web Search and Data Mining (WSDM'12)*, New York, NY, USA, 2012, pp. 743–752.
- [11] R. S. Burt, *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press, 1992.

- [12] D. Kempe, J. Kleinberg, and E. 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 (KDD'03)*, New York, NY, USA, 2003, pp. 137–146.
- [13] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, New York, NY, USA, 2009, pp. 199–208.
- [14] M. Gomez Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10)*, New York, NY, USA, 2010, pp. 1019–1028.
- [15] D. Centola and M. Macy, "Complex Contagions and the Weakness of Long Ties," *American Journal of Sociology*, vol. 113, pp. 702–734, Nov. 2007.
- [16] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Efficient estimation of cumulative influence for multiple activation information diffusion model with continuous time delay," in *Proceedings of the 11th Pacific Rim International Conference on Trends in Artificial Intelligence (PRICAI'10)*, Berlin, Heidelberg, 2010, pp. 244–255.
- [17] M. Kimura, K. Saito, and H. Motoda, "Efficient estimation of influence functions for sis model on social networks," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, San Francisco, CA, USA, 2009, pp. 2046–2051.
- [18] K. Saito, M. Kimura, and H. Motoda, "Discovering influential nodes for sis models in social networks," in *Proceedings of the 12th International Conference on Discovery Science (DS'09)*, Berlin, Heidelberg, 2009, pp. 302–316.
- [19] S. Foster, W. Potter, J. Wu, B. Hu, and Y. Zhang, "A history sensitive cascade model in diffusion networks," in *Proceedings of the 2009 Spring Simulation Multiconference (SpringSim'09)*, San Diego, CA, USA, 2009.
- [20] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate," in *SIAM International Conference on Data Mining (SDM'11)*, 2011.
- [21] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proceedings of the 20th International Conference on World Wide Web (WWW'11)*, New York, NY, USA, 2011, pp. 665–674.
- [22] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Learning continuous-time information diffusion model for social behavioral data analysis," in *Proceedings of the 1st Asian Conference on Machine Learning: Advances in Machine Learning (ACML'09)*, Berlin, Heidelberg, 2009, pp. 322–337.
- [23] K. Saito, K. Ohara, Y. Yamagishi, M. Kimura, and H. Motoda, "Learning diffusion probability based on node attributes in social networks," in *Proceedings of the 19th International Conference on Foundations of Intelligent Systems (ISMIS'11)*, Berlin, Heidelberg, 2011, pp. 153–162.
- [24] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, Dec. 2004.
- [25] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, Aug. 2003.
- [26] J. Bilmes, "A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," International Computer Science Institute, University of Berkeley, Tech. Rep. TR-97-021, 1997.
- [27] J. Leskovec, L. Backstrom, and J. Kleinberg. (2007) Memetracker: Download memetracker data. [Online]. Available: <http://memetracker.org/data.html>
- [28] 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 (WWW'04)*, New York, NY, USA, 2004, pp. 491–501.
- [29] 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," *Comput. Linguist.*, vol. 18, pp. 31–40, Mar. 1992.