

# MMRate: Inferring Multi-aspect Diffusion Networks with Multi-pattern Cascades

Senzhang Wang\*, Xia Hu‡, Philip S. Yu†, Zhoujun Li\*

\* Beihang University, Beijing, 100191, China

‡ Arizona State University, Tempe, AZ 85287, USA

† University of Illinois at Chicago, Chicago, IL 60607, USA

{szwang@cse., lizj@}buaa.edu.cn, psyu@uic.edu, xia.hu@asu.edu

## ABSTRACT

Inferring diffusion networks from traces of cascades has been extensively studied to better understand information diffusion in many domains. A widely used assumption in previous work is that the diffusion network is homogenous and diffusion processes of cascades follow the same pattern. However, in social media, users may have various interests and the connections among them are usually multi-faceted. In addition, different cascades normally diffuse at different speeds and spread to diverse scales, and hence show various diffusion patterns. It is challenging for traditional models to capture the heterogeneous user interactions and diverse patterns of cascades in social media. In this paper, we investigate a novel problem of *inferring multi-aspect diffusion networks with multi-pattern cascades*. In particular, we study the effects of various diffusion patterns on the information diffusion process by analyzing users' retweeting behavior on a microblogging dataset. By incorporating aspect-level user interactions and various diffusion patterns, a new model for inferring Multi-aspect transmission Rates between users using Multi-pattern cascades (MMRate) is proposed. We also provide an Expectation Maximization algorithm to effectively estimate the parameters. Experimental results on both synthetic and microblogging datasets demonstrate the superior performance of our approach over the state-of-the-art methods in inferring multi-aspect diffusion networks.

## Categories and Subject Descriptors

H.2.8 [Database Management]: Database Application: *data mining*

## General Terms

Algorithms; Experimentation

## Keywords

Information Diffusion; Graph Mining; Social Network

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

KDD '14, August 24-27 2014, New York, NY, USA

Copyright 2014 ACM 978-1-4503-2956-9/14/08\$15.00.

<http://dx.doi.org/10.1145/2623330.2623728>.

## 1. INTRODUCTION

The increasing popularity of social networking websites, such as Twitter and Sina Weibo, has greatly promoted the research on social network analysis due to the massive real data available. A ubiquitous problem of social network analysis is how the information, idea, and innovation diffuse in the networks. Studying the process of information dissemination, cascading behavior and influence spreading in social networks is fundamental in many applications, such as technological innovations spreading [12], word of mouse effects in viral marketing [2, 13] and personalized recommendation in E-commerce websites [17].

As a fundamental research issue in information diffusion, inferring the network structure from a collection of diffusion cascades has been studied in many domains, such as the hyperlink network of on-line news articles [6, 7, 8] and the collaboration network of scientists [18]. Existing models can be roughly divided into inferring the network connectivity, such as NetInfer [7] and MulTree [9], and inferring not only the network structure but also the transmission rates over the edges, like NETRATE [6] and InfoPath [8]. These methods normally use a continuous time distribution to model the cascade diffusion process, and try to find the optimal homogeneous network which can maximize the likelihood of all the observed cascades.

Existing studies assume the network connections are homogeneous and have only one aspect [6, 8, 18]. This assumption does not hold in social media. The interactions on social networks are usually heterogeneous and have many aspects. For example, the research on social trust showed that the trust relationship between online users has many facets [22]. One user tends to trust different users in terms of different facets. Similarly, social theories [11, 15] demonstrated that the effect of social influence from different angles such as topics may vary significantly. Users tend to interact with different friends in terms of different angles in society. People's multi-aspect interests and influence of different types suggest people may adopt different types of information from different people in social networks.

Figure 1 shows a toy example of a network with two aspects: research and sports. The entire network can be decomposed into two subgraphs. Each subgraph is associated with one aspect. The width of the edges represents the relationship strength. For different aspects, the corresponding relationship strength is different. For example, user 1 and 2 are researchers working on data mining related topics. They might interact with each other intensively on data mining related cascades, but they do not necessarily share similar

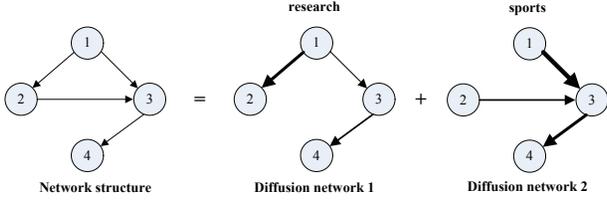


Figure 1: Example of a 2-Aspect Diffusion Network

interests on sports. It is challenging to infer the multi-aspect diffusion network due to the fact that the aspects are usually not explicitly observed. Though the toy example suggests we can consider each topic as an aspect, it is hard to verify whether different topics can be integrated as one aspect due to similar diffusion cascades. Moreover, the topics in different social network websites may differ significantly. For example, topics a user discusses on Twitter could be different from the topics he/she discusses on Epinions. To address this challenge, we need a general model to automatically distinguish the aspects of the cascades and infer the aspect-sensitive diffusion networks.

Meanwhile, the underlying reasons driving a piece of information spreading from node to node is more complex in social networks than those in traditional networks. Social theories showed that information and influence can be transmitted by various mechanisms, such as *conformity*, *social learning*, and *social contagion* [19]. In medical science, it is also well-established that more infectious disease, compared against less infectious ones, can spread to a wider scale in a short period of time [14]. In social networks, cascades usually spread at different speeds and to diverse scales, and thus present various diffusion patterns. Inspired by the sociological and epidemiologic theories, we are particularly interested in investigating the following question: *do various diffusion patterns of the cascades reflect users’ retweeting behavior in social networks?* Distinguishing diverse diffusion patterns of the cascades in social networks may potentially help us better understanding information diffusion and more accurately inferring the diffusion networks.

In this paper, we propose a new model to infer the multi-aspect diffusion networks by taking advantage of both aspect-sensitive user interactions and various diffusion patterns. In particular, we investigate 1) whether different cascades present distinct diffusion patterns in social networks, and 2) how their diffusion patterns reflect users’ behavior in information diffusion. Then we discuss how to incorporate the aspect-sensitive transmission rates and diffusion patterns of cascades into a new model to more effectively infer the multi-aspect diffusion networks. Finally, we conduct extensive experiments to evaluate the proposed model. The main contributions of this paper can be summarized as follows:

- We formally define the problem of inferring the multi-aspect diffusion networks to enable the utilization of multi-aspect user interactions and the various diffusion patterns of cascades in social networks;
- By analyzing 300,000 retweeting cascades in Sina Weibo, we study 1) the existence of various diffusion patterns of the microblogs; and 2) users’ retweeting behaviors when encountering the cascades with remarkably different diffusion patterns;

- By incorporating the aspect-sensitive diffusion cascades and various diffusion patterns, we present a new model MMRate. Moreover, to effectively estimate the parameters, an Expectation Maximization algorithm is given;
- We empirically evaluate MMRate on both synthetic and real-world microblogging datasets and study the application of MMRate to help other data mining tasks, such as aspect-sensitive influential mining.

The remainder of the paper is organized as follows: in Section 2, we formally define the studied problem. Section 3 describes the dataset and analyzes how the diffusion patterns reflect users’ retweeting behavior. Section 4 introduces the proposed model MMRate and gives an EM-algorithm to effectively infer the parameters. In Section 5, we evaluate our approach and report the results. We discuss related work in Section 6, and conclude this paper in Section 7.

## 2. PROBLEM FORMULATION

In this section, we will first introduce some notations, and then we will give a formal definition of the studied problem.

**Cascade:** A cascade  $\mathbf{t}^c$  can be denoted as a  $N$ -dimensional vector  $\mathbf{t}^c = (t_1^c, \dots, t_N^c)^T$ , where  $N$  is the number of nodes. The  $i$ th dimension of  $\mathbf{t}^c$  records the time stamp when information  $c$  infects node  $i$ , and  $t_i^c \in [0, T^c] \cup \{\infty\}$ . The symbol  $\infty$  labels nodes that are not infected during the observation window  $[0, T^c]$ . The time stamp is set to 0 at the start of each cascade.

A cascades dataset  $\mathbb{C}$  consists of a collection of cascades, i.e.  $\mathbb{C} = \{\mathbf{t}^1, \dots, \mathbf{t}^M\}$ , where  $M$  is the number of cascades. To simplify the problem formulation, we make the following two assumptions. First, within a cascade  $\mathbf{t}^c$ , a node  $i$  with a time stamp  $t_i^c$  can only be infected by the node  $j$  with an earlier time stamp, i.e.  $t_j^c < t_i^c$ . In a cascade  $\mathbf{t}^c$ , if  $t_j^c > t_i^c$ , we refer node  $j$  as one of node  $i$ ’s child node and node  $i$  as one of node  $j$ ’s parent node, though there may not exist an edge between them on the true network. Second, similar to the Independent Cascade model [5], we assume the child node is infected once its first parent succeeds in infecting it, and it can not be infected again. All the parents try to infect their child nodes independently.

**Aspect distribution:** Assuming there are  $K$  aspects in  $\mathbb{C}$ , the aspect distribution of  $\mathbb{C}$  can be represented as  $\Pi = \{\pi_1, \dots, \pi_K\}$ , and  $\sum_{i=1}^K \pi_i = 1$ , where  $\pi_k$  is the probability of the  $k$ th aspect. Each cascade  $c_i$  is associated with a  $K$ -dimensional vector  $\mathbf{z}_i = (z_{i1}, \dots, z_{iK})$  and  $\sum_{k=1}^K z_{ik} = 1$ , with the  $k$ th element  $z_{ik}$  denoting the probability of the cascade  $\mathbf{t}^{c_i}$  belonging to the  $k$ th aspect.

**Diffusion patterns:** Each cascade  $\mathbf{t}^c$  presents a unique diffusion pattern which is independent of the network structure. To model the diffusion pattern of  $\mathbf{t}^c$ , we use a function  $g(\Delta t; \delta^c)$  to represent the probability distribution of the decaying time  $\Delta t$  between two successively infected node ( $t_i^c, t_j^c$ ) in cascade  $\mathbf{t}^c$ , where  $\delta^c$  is the distribution parameter. The parameter  $\delta^c$  determines the spreading speed of the cascade. We will describe the function  $g(\Delta t; \delta^c)$  in detail later.

**Aspect-sensitive information diffusion graph:** An information diffusion graph  $\mathbb{G} = (\mathbb{N}, \mathbb{E})$  can be decomposed as  $K$  aspect-sensitive subgraphs  $\mathbb{G} = \cup_{k=1:K} G_k$  with  $G_k = (\mathbb{N}, \mathbb{E}_k)$ .  $\mathbb{N}$  is the node set,  $\mathbb{E}$  is the edge set, and  $\mathbb{E}_k$  is the edge set associated with aspect  $k$ .  $\mathbb{E}_k$  can be further denoted as  $\mathbb{E}_k = \cup_{i,j=1:N} \{\alpha_{ij}^k\}$ , where  $\alpha_{ij}^k$  is the  $k$ -aspect

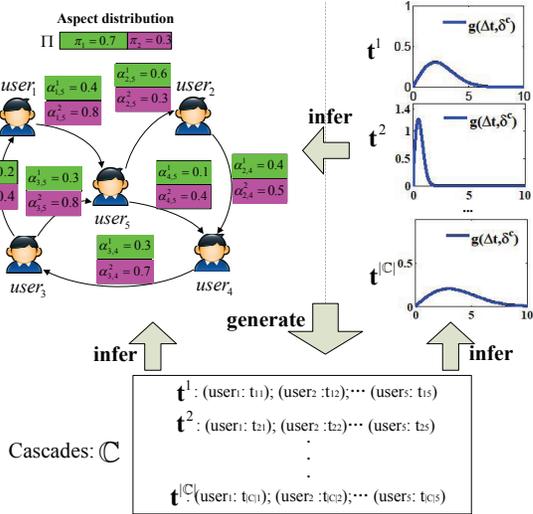


Figure 2: An Illustration of the Studied Problem

transmission rate from node  $i$  to  $j$ . In most cases, the aspect-sensitive pairwise transmission probability is asymmetric, i.e.,  $\alpha_{ij}^k \neq \alpha_{ji}^k$ .

We provide an illustration of the studied problem in Figure 2. The upper left part of Figure 2 shows a 2-aspect network, the upper right part presents various diffusion patterns of the cascades, and the lower part is the collection of cascades  $\mathbb{C}$ . In the network, the green and red bars denotes the two aspects with probabilities  $\pi_1 = 0.7$  and  $\pi_2 = 0.3$ , respectively. Each connected pair of users have two transmission rates  $\alpha_{ij}^1$  and  $\alpha_{ij}^2$  with each one associated with an aspect. The diffusion pattern of cascade  $\mathbf{t}^c$  is modeled by the distribution function  $g(\Delta t, \delta^c)$  shown in the upper right part. The x-axis is the decaying time of the cascades during spreading, and the y-axis is the probability. Each cascade  $\mathbf{t}^c$  in  $\mathbb{C}$  is composed by the tuples  $(user_i : t_{ic})$ , where  $user_i$  is the node and  $t_{ic}$  is the time when  $user_i$  is infected by the cascade  $\mathbf{t}^c$ . Given the multi-aspect network  $\mathbb{G}$  with heterogeneous relationships between users, the aspect distribution  $\Pi$  and the diffusion patterns  $\delta^c$ , a set of corresponding cascades  $\mathbb{C}$  can be generated. Conversely, given the cascades data  $\mathbb{C}$ , the task here is inferring the aspects distribution  $\Pi$ , the diffusion patterns  $\delta^c$  and the multi-aspect diffusion network  $\mathbb{G}$ .

Based on the above definitions, we define the task of *inferring the multi-aspect diffusion networks with multi-pattern cascades* as follow.

**PROBLEM 1.** Given 1) a network  $\mathbb{G} = \{\mathbb{N}, \mathbb{E}\}$  with hidden edges, where  $\mathbb{N}$  represents nodes in the graph and  $\mathbb{E}$  are the edges which are unknown, and 2) a collection of observed cascades  $\mathbb{C} = \{\mathbf{t}^1, \dots, \mathbf{t}^M\}$  spreading with various diffusion patterns, our goal is to accurately 1) infer the edges  $\mathbb{E}$  of  $\mathbb{G}$ , and 2) estimate the aspect-sensitive information transmission rates  $(\alpha_{ij}^1, \dots, \alpha_{ij}^K)^T$  over each edge  $e_{ij}$ .

### 3. DATASET AND DATA ANALYSIS

In this section we will first introduce the dataset studied in this paper. Then we will study the difference of the diffusion

	DPU	DUNF
# of nodes	1,038	750
# of edges	11,385	2,974
#of cascades	5,124	4,206
average # of cascades per edge	1.9	6.8

patterns and investigate the effect of different patterns on information diffusion via real data.

### 3.1 Datasets

We study the large microblogging network dataset crawled from Sina Weibo<sup>1</sup> [25]. This dataset is crawled to study the retweeting behaviors in the microblogging network. It contains 1,776,950 users, 308,489,739 following relationships, 300,000 original microblogs, and 23,755,810 retweets. On average, each microblog has been retweeted for about 80 times in the dataset. For the purpose of this study, we build the following two datasets from the initial dataset.

The first dataset is built as follows. As more popular Sina Weibo users can transmit information to a wider scale, we first select the users with over 1,500 followers as nodes and extract corresponding relationships among them as edges. In this way, we obtain a network with 1,624 nodes and 23,097 corresponding directed connections. Then for each microblog, we extract its diffusion path in the extracted network. As a result, 5,124 microblogs are extracted as diffusion cascades. Some nodes or edges have not participated in any of the selected diffusion cascades are then eliminated. Finally, we obtain the first diffusion network with 1,038 users and 11,385 following relationships. For simplicity, we name this dataset *DPU* (Dataset of Popular Users).

For the second dataset, we select the users with a large number of friends as the nodes, because such users can usually receive more posts and hence have more chances to participate in more cascades. Similarly, we extract corresponding relationships among them and obtain a network with 1,345 nodes and 5,874 edges. In the similar way as *DPU*, we extract 4,206 cascades, After eliminating some isolate nodes and edges, we obtain the second dataset with 750 nodes and 2,974 edges. We name this dataset *DUNF* (Dataset of Users with Numerous Friends). The statistics of the two datasets are given in Table 1.

### 3.2 Study the Diffusion Patterns in Microblogging Network

In this subsection, we will study the diffusion patterns of the cascades and investigate how these patterns reflect users' retweeting behavior in microblogging network. More specifically, we want to answer the following two questions: *Are the diffusion patterns of some cascades significantly different from others in microblogging networks? How does the various diffusion patterns of the cascades reflect users' retweeting behavior during information diffusion?* To answer the two questions, we conduct the following two groups of  $t$ -tests.

To answer the first question, we conduct a group of two-sample  $t$ -tests to verify whether more popular posts spread faster than ordinary ones. We first rank the microblogs according to their numbers of retweets, then we select the top

<sup>1</sup><http://arnetminer.org/Influencelocality>.

$k\%$  microblogs from the ranking results as the first group of samples. The entire dataset are considered as the second group of samples. From each group, we randomly select 1000 posts composing two test samples  $S_1$  and  $S_2$ . Then we compute the average spreading speed  $sc_t, sc_r$  of  $S_1$  and  $S_2$ , respectively.  $sc_t$  can be computed by  $sc_t = \frac{1}{1000} \sum_{i=1}^{1000} \frac{n_i}{t_i}$ , where  $n_i$  denotes the repost number of the post  $p_i$  and  $t_i$  is the time span of  $p_i$ .  $sc_r$  can be computed in the similar way. The null hypothesis is that there is no significant difference between the two spreading speeds,  $H_0 : sc_t = sc_r$ ; while the alternative hypothesis is the average spreading speed of  $S_1$  is faster than that of  $S_2$ ,  $H_1 : sc_t > sc_r$ . The null hypothesis is rejected at the significant level  $\alpha = 0.01$ . The  $p$ -values of the  $t$ -test with various  $k$  are shown in Table 2. The star next to the  $p$ -value means that there is statistical evidence to reject the null hypothesis. One can see that the null hypothesis can be rejected when  $k$  is 5, 10 and 20. In others words, the top 5%, 10% and 20% popular posts spread significantly faster than the random selected posts.

**Table 2: Hypothesis Testing Results (P-Values) to Verify More Popular Posts Spread Faster.**

$p(k=5)$	$p(k=10)$	$p(k=20)$	$p(k=40)$
3.1259e-10*	1.9653e-4*	6.8347e-4*	1.39e-2

To answer the second question, we conduct the second groups of two-sample  $t$ -tests to investigate 1) whether the spreading of popular posts counts on the closeness of the senders and recipients, and 2) whether the user’s retweeting behavior is related to the diffusion patterns of the posts. Before describing the detail of the  $t$ -test, we first give a definition as follow.

**Definition 1. Closeness:** the closeness from  $user_i$  to  $user_j$  is defined as  $closeness_{ij} = \frac{t_j}{t_{all}}$ , where  $t_j$  denotes the number of posts  $user_i$  retweeting from  $user_j$ , and  $t_{all}$  denotes the total number of retweeted posts by  $user_i$ .

To obtain enough data for the  $t$ -test, we tend to choose the users with larger number of friends and posts. Hence we first randomly select 100 users who posted more than 1000 microblogs and have more than 200 friends. Then for each selected user  $user_i$ , we rank all his/her friends according to the closeness to them. We select top  $k\%$  friends from the ranking as the first group of samples  $S_1$ , and bottom  $k\%$  friends as the second group of samples  $S_2$ . Then we compute the average number of retweet  $sp_t, sp_l$  of the posts  $user_i$  retweeted from the friends in  $S_1$  and  $S_2$ , respectively. The null hypothesis is that there is no significant difference between  $sp_t$  and  $sp_l$ ,  $H_0 : sp_t = sp_l$ ; while the alternative hypothesis is  $sp_t$  is significantly smaller than  $sp_l$ ,  $H_1 : sp_t < sp_l$ . The null hypothesis is rejected at the significant level  $\alpha = 0.01$ . The  $p$ -values of the  $t$ -test with various  $k$  are shown Table 3. The  $t$ -test result reveals that there is statistical evidence to reject the null hypothesis when  $k$  is 2, 5 and 8. That is, the average popularity of posts retweeted from the top 2%, 5% and 8% close friends is significantly smaller than that from those less close friends. Note that the  $t$ -test result does not mean popular tweets spread mostly through less close friends. It just means that if a user retweets a post

**Table 3: Hypothesis Testing Results (P-Values) to Verify Diffusion Patterns Reflect Users Behavior**

$p(k=2)$	$p(k=5)$	$p(k=8)$	$p(k=50)$
8.2723e-8*	2.9124e-8*	7.4367e-6*	3.65e-2

from a less close friend, the retweeted post is more likely to be popular.

In summary, the evidence from the two groups of  $t$ -test results suggests the following answers to the questions: 1) *The more popular microblogs are more likely to spread faster, and this infers that different cascades may present significantly different diffusion patterns.* 2) *Users’ retweeting behavior is related to the diffusion patterns of the posts. With a higher probability, a post spreading through two users who are not closed is more popular than a post spreading through two very close users.* Based on these findings, we next will introduce how to exploit the diffusion patterns of cascades to help us better infer diffusion networks.

## 4. MMRATE: INFERRING MULTI-ASPECT DIFFUSION NETWORKS

In this section, we will start by briefly describing some preliminaries. Then we will give the multi-aspect information diffusion model and present how to compute the likelihood of a collection of cascades using the proposed model. To estimate the parameters, we next will give a EM-algorithm to effectively infer both the aspect-sensitive diffusion networks and the aspect of each cascade belonging to.

We first revisit some basic concepts and introduce some standard notations. We define a nonnegative random variable  $T$  to be the time when an event happens, such as  $user_i$  adopting a piece of information. Let  $f(t)$  be the probability density function of  $T$ , then the cumulative density function can be denoted as  $F(t) = P(T \leq t) = \int_0^t f(x)dx$ .

**Survival function.** The survival function  $S(t)$  is the probability that an event does not happen to a node by time  $t$ :

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^\infty f(x)dx.$$

**Hazard function.** Given functions  $f(t)$  and  $S(t)$ , we can further define the *hazard function*  $H(t)$ , which means the instantaneous rate that an event will happen to a particular node within a small interval just after time  $t$  if it has not happened to the node yet by time  $t$ .

$$H(t) = \lim_{\Delta t \rightarrow 0} \frac{p(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

### 4.1 Incorporating Patterns and Aspects

In this subsection, we will introduce the proposed model MMRate. For each cascade, we use the decaying time distribution function to model its diffusion pattern. We consider that each cascade is also associated with a hidden aspect. By a liner combination of the diffusion pattern and aspect, the MMRate is proposed.

To model information diffusion, first we need to consider how to compute the diffusion probability between two nodes in a specific cascade. In our model, the pairwise diffusion probability is aspect-sensitive and related to the diffusion

**Table 4: Notations**

SYMBOL	DESCRIPTION
$N$	the number of nodes in the network
$M$	the number of cascades
$K$	the number of aspects
$\mathbf{t}^c$	a cascade, $\mathbf{t}^c = (t_1^c, \dots, t_N^c)^T$
$\mathbb{C}$	the collection of cascades, $\mathbb{C} = \{\mathbf{t}^1, \dots, \mathbf{t}^M\}$
$\mathbb{G}$	the hidden multi-aspects diffusion network
$\Pi$	the aspects distribution, $\Pi = \{\pi_1, \dots, \pi_K\}$
$\mathbb{Z}$	the aspect matrix of the cascades, $\mathbb{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_M)$
$\Theta$	the variable set, $\Theta = \{\mathbb{G}, \mathbb{Z}\}$
$G^k$	the diffusion network associated with aspect $k$
$\alpha_{ij}^k$	the transmission rate from node $i$ to $j$ in aspect $k$
$\mathbf{z}_i$	the aspects distribution of cascade $c_i$
$z_{ik}$	if $c_i$ belongs to aspect $k$ , $z_{ik} = 1$ ; otherwise, $z_{ik} = 0$
$\pi_k$	the probability of the $k_{th}$ aspect
$\delta^c$	the diffusion pattern parameters, $\delta^c = \{\delta^1, \dots, \delta^M\}$

patterns. Hence we decompose the probability of one node  $i$  activating one of its neighbors  $j$  in the cascade  $\mathbf{t}^c$  into two parts: the aspect-sensitive transmission rate  $\alpha_{ij}^k$ , and the probability that cascade  $\mathbf{t}^c$  with a pattern parameter  $\delta^c$  takes time  $\Delta t_{ij}^c$  to infect a new node. Before describing MMRate in detail, we first give some notations in Table 4.

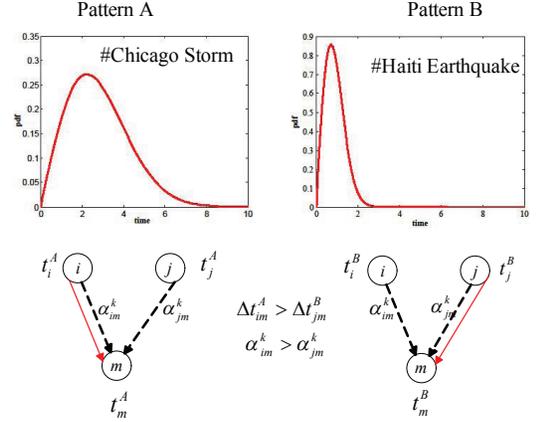
**Probability of node  $i$  infecting node  $j$ .** The probability of node  $i$  infecting node  $j$  in the cascade  $\mathbf{t}^c$  with aspect  $k$  and pattern  $\delta^c$  can be represented as :

$$f(\Delta t_{ij}^c; k, \delta^c) = g_1(\Delta t_{ij}^c; g_2(\alpha_{ij}^k, \delta^c)). \quad (1)$$

Here,  $g_1(\cdot)$  is the distribution to model the conditional likelihood of node  $i$  infecting  $j$  given the aspect-sensitive transmission rate  $\alpha_{ij}^k$ , diffusion pattern parameter  $\delta^c$  and the time gap  $\Delta t_{ij}^c$ .  $g_2(\cdot)$  is a mapping function. Given  $\alpha_{ij}^k$  and  $\delta^c$ ,  $g_2(\cdot)$  combines the two values. Here we simply sum the two values, i.e.  $g_2(\alpha_{ij}^k, \delta^c) = \alpha_{ij}^k + \delta^c$ .

In this paper, we apply the following three distributions to model the diffusion probability function  $g_1(\cdot)$ : exponential, power-law and Rayleigh. The exponential and power-law distributions are widely used to model information diffusion in social networks [7, 18]. Rayleigh distribution is a non-monotonic parametric model used in epidemiology [24]. Recently the Rayleigh model is also adopted to model information diffusion [6, 8]. For simplicity, we assume the diffusion pattern function  $g_2(\cdot)$  of the cascade  $\mathbf{t}^c$  also follows one of the three distributions with a distribution parameter  $\delta^c$ , and the maximum likelihood estimation (MLE) of  $\delta^c$  can be derived from the distribution function. The mathematical expressions of the three models and the corresponding MLE of  $\delta^c$  are given in Table 5.

**Further explanation of MMRate.** Intuitively, the probability that node  $i$  transmits the information  $c$  to node  $j$  depends on both their aspect-sensitive transmission rate  $\alpha_{ij}^k$  and the diffusion pattern with the parameter  $\delta^c$ . As we studied previously, the diffusion patterns of cascades can reflect users’ retweeting behavior. The propagation of microblogs are also related to their diffusion patterns, and those popular ones can spread between two users independent of their *closeness*. Figure 3 gives an example to show how different diffusion patterns reflect information propagation. The upper two figures show the diffusion patterns of the two hashtags in Twitter: “#Chicago Storm” and “#Haiti Earth-



**Figure 3: An Illustration to Show How the Diffusion Patterns Reflect Information Propagation**

quake”, and the lower two figures show the corresponding diffusion paths. The dashed black lines in the lower two figures represent the underlying possible diffusion paths, and the red solid ones are the real diffusion paths. The transmission rate from  $i$  to  $m$  is larger than that from  $j$  to  $m$  ( $\alpha_{im}^k > \alpha_{jm}^k$ ), but the time interval between  $j$  and  $m$  is smaller than that between  $i$  and  $m$  ( $\Delta t_{im}^A > \Delta t_{jm}^B$ ). One can see that the diffusion paths of the two hashtags are different. For the hashtag “#Chicago Storm”, the transmission rate contributes more to its diffusion, but for the spreading of “#Haiti Earthquake”, time plays a more important role.

**Likelihood of node  $i$  infecting  $j$  in cascade  $\mathbf{t}^c$ .** In a cascade, we assume 1) one node gets infected once the first parent infects it, and 2) all the parents infect their child nodes independently. Based on the two assumptions, the likelihood of the parent node  $i$  infecting the child node  $j$  in the cascade  $\mathbf{t}^c$  with aspect  $k$  and pattern  $\delta^c$  can be computed by

$$f(\Delta t_{ij}^c; \alpha_{ij}^k, \delta^c, G^k) \times \prod_{i \neq u, t_u^c < t_j^c} S(\Delta t_{uj}^c; \alpha_{uj}^k, \delta^c, G^k). \quad (2)$$

In the cascade  $\mathbf{t}^c$ , the node  $j$  could possibly be infected by any one of its parent nodes. Hence the likelihood of  $j$  getting infected in the cascade  $\mathbf{t}^c$  can be calculated by summing over the likelihoods of each potential parent being the first one to infect it.

$$\Gamma_j^+(\mathbf{t}^c; k, \mathbb{G}, \delta^c) = \sum_{i: t_i^c < t_j^c} f(\Delta t_{ij}^c; \alpha_{ij}^k, \delta^c, G^k) \times \prod_{i \neq k, t_u^c < t_j^c} S(\Delta t_{uj}^c; \alpha_{uj}^k, \delta^c, G^k). \quad (3)$$

**Likelihood of node survived from the cascade  $\mathbf{t}^c$ .** If node  $j$  is survived from all the parents by the time  $T^c$ , we say the node is survived from the cascade  $\mathbf{t}^c$ . The likelihood that node  $j$  is survived from the cascade  $\mathbf{t}^c$  can be represented by the following product of survival function

$$\Gamma_j^-(\mathbf{t}^c; k, \mathbb{G}, \delta^c) = \prod_{t_i^c < T^c} S(T^c - t_j^c; \alpha_{ij}^k, \delta^c, G^k). \quad (4)$$

**Table 5: Three Distributions to Model the Edge Transmission Likelihoods**

Model	Propagation likelihood	Log survival function	Hazard function	MLE of $\delta^c$
	$f(\Delta t_{ij}^c; k, \delta^c, \mathbb{G})$	$\log S_{ij}(\Delta t_{ij}^c; k, \delta^c, \mathbb{G})$	$H_{ij}(\Delta t_{ij}^c; k, \delta^c, \mathbb{G})$	$\delta^c$
EXP	$\begin{cases} (\alpha_{ij}^k + \delta^c) \cdot e^{-(\alpha_{ij}^k + \delta^c)\Delta t_{ij}^c} & \text{if } t_i < t_j \\ 0 & \text{otherwise} \end{cases}$	$-(\alpha_{ij}^k + \delta^c) \cdot \Delta t_{ij}^c$	$\alpha_{ij}^k + \delta^c$	$\frac{1}{\Delta t}$
POW	$\begin{cases} \frac{(\alpha_{ij}^k + \delta^c)}{\sigma} \cdot \left(\frac{\Delta t_{ij}^c}{\sigma}\right)^{-1 - (\alpha_{ij}^k + \delta^c)} & \text{if } t_i + \sigma < t_j \\ 0 & \text{otherwise} \end{cases}$	$-(\alpha_{ij}^k + \delta^c) \cdot \log\left(\frac{\Delta t_{ij}^c}{\sigma}\right)$	$(\alpha_{ij}^k + \delta^c) \cdot \frac{1}{\Delta t_{ij}^c}$	$1 + n \left[ \sum_{i=1}^n \ln \frac{\Delta t_i}{\sigma} \right]^{-1}$
RAY	$\begin{cases} (\alpha_{ij}^k + \delta^c) \cdot \Delta t_{ij}^c \cdot e^{-\frac{1}{2}(\alpha_{ij}^k + \delta^c)(\Delta t_{ij}^c)^2} & \text{if } t_i < t_j \\ 0 & \text{otherwise} \end{cases}$	$-(\alpha_{ij}^k + \delta^c) \cdot \frac{\Delta t_{ij}^2}{2}$	$(\alpha_{ij}^k + \delta^c) \cdot \Delta t_{ij}^c$	$\frac{\Delta t \sqrt{\frac{2}{\pi}}}{\Delta t}$

**Likelihood of the cascade  $\mathbf{t}^c$ .** Given a cascade  $\mathbf{t}^c := (t_1^c, \dots, t_N^c)$ , its likelihood can be computed by multiplying the likelihoods of all the infected and survived nodes in the cascade. By Eq. (3) and Eq. (4), the likelihood can be represented as

$$\begin{aligned}
 f(\mathbf{t}^c; k, \mathbb{G}, \delta^c) &= \prod_{t_j^c < T^c} \Gamma_j^+(\mathbf{t}^c) \times \prod_{t_j^c < T^c} \Gamma_j^-(\mathbf{t}^c) \\
 &= \prod_{t_j^c < T^c} \prod_{t_m^c > T^c} S(T^c - t_j^c; \alpha_{jm}^k, \delta^c, G^k) \times \\
 &\quad \prod_{u: t_u^c < t_j^c} S(\Delta t_{uj}^c; \alpha_{uj}^k, \delta^c, G^k) \sum_{i: t_i^c < t_j^c} H(\Delta t_{ij}^c; \alpha_{ij}^k, \delta^c, G^k). \quad (5)
 \end{aligned}$$

Assuming the cascades spread independently in the network, the likelihood of a set of cascades  $\mathbb{C} = \{\mathbf{t}^1, \dots, \mathbf{t}^M\}$  is the product of the likelihoods of all the individual cascades.

$$\prod_{\mathbf{t}^c \in \mathbb{C}} f(\mathbf{t}^c; \mathbb{Z}, \delta^c, \mathbb{G}). \quad (6)$$

**Inferring the network.** Given a set of cascades  $\mathbb{C} = \{\mathbf{t}^1, \dots, \mathbf{t}^M\}$  with estimated diffusion patterns  $\delta^c$ , our goal here is to find the aspect matrix of the cascades  $\mathbb{Z} = (\mathbf{z}^1, \dots, \mathbf{z}^M)$  and the graph  $\mathbb{G}$  to maximize the likelihood of generating the cascades collection  $\mathbb{C}$ . This problem can be reduced to solve the following convex optimization problem

$$\min_{\mathbb{G}, \mathbb{Z}} - \sum_{c \in \mathbb{C}} \log f(\mathbf{t}^c; \mathbb{Z}, \delta^c, \mathbb{G}) \quad (7)$$

$$\text{s. t. } \alpha_{ij}^k \geq 0, z_{ik} \in \{0, 1\} \text{ and } \sum_{k=1:K} z_{ik} = 1$$

$$i, j = 1, \dots, N, i \neq j; k = 1, \dots, K$$

where  $\alpha_{ij}^k$  is the aspect-sensitive transmission rate variable in the multi-aspect graph  $\mathbb{G}$ ;  $\mathbb{Z} = \{z_{ik} | i, k = 1, \dots, N\}$  is the matrix of the aspect variables. The edges of the  $k$ -aspect network are those pairs of nodes with transmission rate  $\alpha_{ij}^k > 0$ .

## 4.2 Parameter Estimation

To solve Eq. (7), we have to infer the aspect matrix  $\mathbb{Z}$  and the graph  $\mathbb{G}$ . In this subsection, we propose an Expectation Maximization method to infer the two groups of parameters.

Similar to the Gaussian mixture model, we first consider the aspect matrix  $\mathbb{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_M)$  as hidden variables, where  $\mathbf{z}_i$  denotes the aspect vector of cascade  $c_i$ . If cascade  $c_i$  belongs to the  $k_{th}$  aspect, only the  $k_{th}$  element  $z_{ik}$  of  $\mathbf{z}_i$  is 1 and all the other elements  $z_{im}$  ( $m \neq k$ ) are 0. In addition, we use  $\pi_k$  to denote the probability of the  $k_{th}$  aspect in  $\mathbb{C}$ .

The complete likelihood of all the cascades can be represented as

$$\begin{aligned}
 &\log \mathbb{L}(\mathbb{G}, \Pi | \mathbb{C}, \mathbb{Z}, \delta^c) \\
 &= \log p(\mathbb{C}, \mathbb{Z} | \mathbb{G}, \Pi, \delta^c) \\
 &= \log \prod_{i=1}^M p(\mathbf{t}^{c_i}, \mathbf{z}_i | \mathbb{G}, \Pi, \delta^{c_i}) \quad (8) \\
 &= \sum_{i=1}^M \log \sum_{k=1}^K p(\mathbf{t}^{c_i} | z_{ik} = 1, G^k, \delta^{c_i}) \cdot p(z_{ik} = 1 | \mathbb{G}) \\
 &= \sum_{i=1}^M \log \sum_{k=1}^K p(\mathbf{t}^{c_i} | z_{ik} = 1, G^k, \delta^{c_i}) \cdot \pi_k
 \end{aligned}$$

**The E-Step:** For simplicity, we use  $\Theta$  to denote the variables  $\mathbb{G}$  and  $\mathbb{Z}$ . That is,  $\Theta = \{\mathbb{G}, \mathbb{Z}\}$ . Then the E-step can be formulated as

$$\begin{aligned}
 &Q(\Theta, \Theta^{(t-1)}) \\
 &= Q[(\mathbb{G}; \mathbb{Z}), (\mathbb{G}; \mathbb{Z})^{(t-1)}] \\
 &= E_{f(\mathbb{Z}) | \mathbb{C}, \Theta^{(t-1)}} [\mathbb{L}(\mathbb{G}, \Pi | \mathbb{C}, \mathbb{Z}, \delta^c) | \mathbb{C}, \Theta^{(t-1)}] \\
 &= \int_{\mathbf{z}} \log [\mathbb{L}(\mathbb{G}, \pi | \mathbb{C}, \mathbf{z})] p(\mathbf{z} | \mathbb{C}, \Theta^{(t-1)}, \delta^c) d_{\mathbf{z}} \\
 &= \int_{\mathbf{z}} \sum_{i=1}^M \log [p(c_i | z_{ik} = 1, \delta^{c_i}) \cdot \pi_k] \prod_{j=1}^N p(z_{jk} = 1 | c_j, \Theta^{(t-1)}) d_{\mathbf{z}} \\
 &= \sum_{k=1}^K \sum_{i=1}^M \log [p(c_i | G_k, \delta^{c_i}) \cdot \pi_k] p(z_{ik} = 1 | c_i, \Theta^{(t-1)}) \quad (9) \\
 &= \sum_{k=1}^K \sum_{i=1}^M \log (\pi_k) \cdot p(z_{ik} = 1 | c_i, \Theta^{(t-1)}) + \\
 &\quad \sum_{k=1}^K \sum_{i=1}^M \log [p(c_i | G_k, \delta^{c_i})] \cdot p(z_{ik} = 1 | c_i, \Theta^{(t-1)}).
 \end{aligned}$$

Using Bayes's rule, we can compute

$$p(z_{ik} = 1 | c_i, \Theta^{(t-1)}) = \frac{\pi_k^{(t-1)} p(c_i | G_k^{(t-1)}, \delta^{c_i})}{\sum_{m=1}^K \pi_m^{(t-1)} p(c_i | G_m^{(t-1)}, \delta^{c_i})} \quad (10)$$

and

$$p(\mathbb{Z} | \mathbb{C}, \Theta^{(t-1)}) = \prod_{i=1}^M p(\mathbf{z}_i | c_i, \Theta^{(t-1)}). \quad (11)$$

In order to get the estimation for  $\pi_k$ , we need to maximize

$\mathbb{Q}^{(\Theta, \Theta^{(t-1)})}$ . Introducing Lagrange multiplier  $\lambda$ , we get:

$$\frac{\partial}{\partial \pi_k} [\mathbb{Q}^{(\Theta, \Theta^{(t-1)})} + \lambda (\sum_{k=1}^K \pi_k - 1)] = 0. \quad (12)$$

That is

$$\sum_{i=1}^M \frac{1}{\pi_k} p(z_{ik} = 1 | c_i, \Theta^{(t-1)}) + \lambda = 0. \quad (13)$$

Thus, combing Equ. (10) and (13). We get the new estimation for  $\pi_k$  given previous  $\Theta^{(t-1)}$

$$\pi_k = \frac{\sum_{i=1}^M \pi_k^{(t-1)} p(c_i | G_k^{(t-1)}, \delta^{c_i})}{\sum_{i=1}^M \sum_{m=1}^K \pi_k^{(t-1)} p(c_i | G_m^{(t-1)}, \delta^{c_i})}. \quad (14)$$

**M-Step:** In the M-Step, we maximize the expectation of the parameter  $\Theta$

$$\text{maximize } Q(\Theta, \Theta^{(t-1)}). \quad (15)$$

## 5. EXPERIMENTAL RESULTS

In this section, we will conduct extensive experiments to evaluate the proposed model MMRate. More specifically, we want to evaluate: 1) the effectiveness of MMRate in recovering multi-aspect network structure; 2) the effectiveness of MMRate in estimating the aspect-sensitive transmission rates between each pair of connected nodes; and 3) the application of MMRate to other data mining tasks.

We begin with introducing the experimental setup, and then compare the performance of MMRate against baselines on both synthetic and real-world datasets. Finally, we give a case study to show how MMRate can be applied to help us rank aspect-sensitive influential user in microblogging.

### 5.1 Experimental Setup

**Dataset.** Besides the real social network datasets described in Section 3, we also use various synthetic datasets to evaluate our approach. We use the popular Kronecker Graph model [16] to generate the following three types of graphs with different topological structures: random graph, hierarchical graph, and core-periphery graph. For each type of graph, we first set the aspect size  $K$ . For each aspect  $k$ , we generate an aspect-sensitive graph  $G_k$  by drawing the pairwise transmission rate  $\alpha_{ij}^k$  for edge  $e_{ij}$  from a uniform distribution. In this way, we obtain a  $K$ -aspect graph  $\mathbb{G}$ . Next, over each aspect-sensitive graph  $G_k$  we generate cascades over it. We first generate the aspect distribution  $\Pi = \{\pi_1, \dots, \pi_K\}$ . Each time we randomly draw an aspect  $k$  from the distribution, and then over the corresponding graph  $G_k$ , we generate a cascade. Root nodes of cascades are chosen uniformly at random. Once a node is infected, the propagation probabilities of outgoing edges of the node determine the infection of its neighbors. The diffusion process continues until some terminating conditions are satisfied, such as reaching the maximum time window  $T$  or no more nodes getting infected.

**Baselines.** As MMRate infers both the connectivity of the network and the transmission rates over each edge, we compare MMRate with the following two state-of-the-art baselines which also infer both the network structure and pairwise transmission rates between nodes.

- **NETRATE**<sup>2</sup> [6]. This is an effective and representative model for inferring the hidden diffusion networks.
- **InfoPath**<sup>3</sup> [8]. InfoPath is an approach proposed recently to infer the dynamic diffusion networks. It can also work on static networks as long as the networks at each time step are the same.
- **Multi-aspect Model** This method is included to examine whether incorporating the diffusion patterns can improve the performance. Multi-aspect model only incorporates the multi-aspects, but ignores the diffusion patterns. That is, the parameter  $\delta^c$  is set to 0.

**Evaluation Metrics.** We evaluate the performance via the following two metrics,

- **$F_1$ -measure.**  $F_1$ -measure is defined as  $F_1\text{-measure} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ , where precision is the fraction of edges in the inferred network  $\hat{\mathbb{G}}$  present in the true network  $\mathbb{G}$ , and recall is the fraction of edges of the true network  $\mathbb{G}$  present in the inferred network  $\hat{\mathbb{G}}$ ;
- **Mean Absolute Error.** To access the accuracy of the estimated transmission rates  $\alpha_{ij}$ , we compute the mean absolute error as follows:  $MAE = E(\frac{|\alpha_{ij} - \hat{\alpha}_{ij}|}{\alpha_{ij}})$ , where  $\alpha_{ij} = \sum_{k=1}^K \alpha_{ij}^k$  is the true transmission rate and  $\hat{\alpha}_{ij} = \sum_{k=1}^K \hat{\alpha}_{ij}^k$  is the estimated transmission rate.

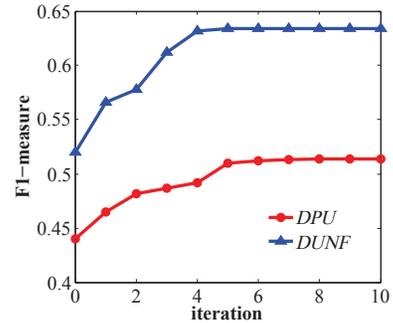


Figure 4: Convergence Curves

**Convergence analysis.** Before evaluation, we first examine the convergence of MMRate. Figure 4 shows the convergence curves of MMRate on the datasets *DPU* and *DUNF*. One can see that the  $F_1$ -measure on both datasets first keeps increasing, and then stabilizes after 5 or 6 iterations, which means MMRate can effectively converge. Through the iterations, the  $F_1$ -measure improvement for *DPU* is about 0.08 (from 0.44 to 0.52), and about 0.12 (from 0.52 to 0.64) for *DUNF*.

### 5.2 Performance Evaluation on Synthetic Data

To evaluate the effectiveness of MMRate on recovering the network structure as well as estimating the transmission rates of each aspect, we compare it with NETRATE and InfoPath on various synthetic data. Since the latent network and corresponding transmission rates for generating

<sup>2</sup><http://people.tuebingen.mpg.de/manuelgr/netrate/>

<sup>3</sup><http://snap.stanford.edu/infopath/index.html>

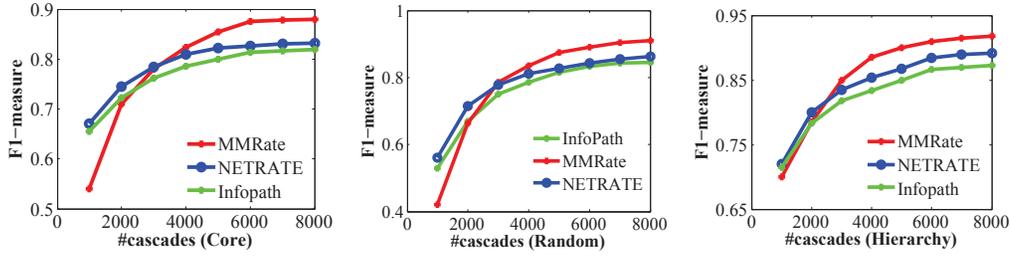


Figure 5:  $F_1$ -measure of the Three Methods on 3-Aspect Networks with 512 Nodes and 1024 Edges

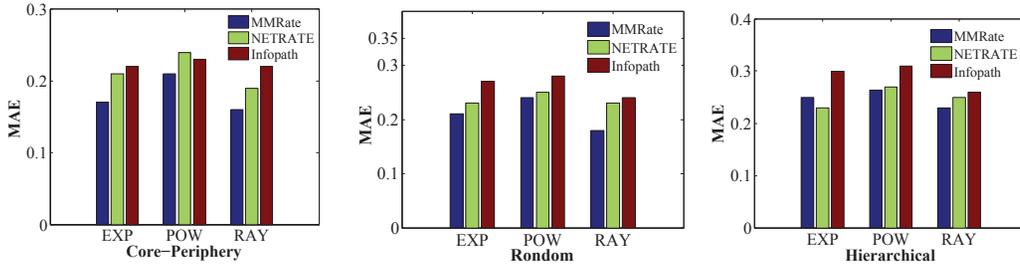


Figure 6: Mean Absolute Error of the Three Methods on 3-Aspect Networks with 512 Nodes and 1024 Edges

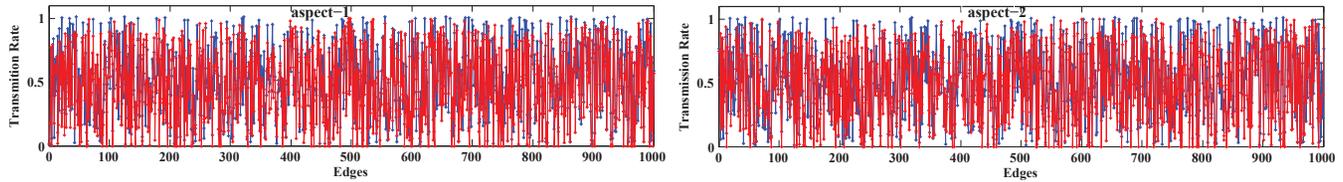


Figure 7: Visualization of the True Transmission Rates (Red Lines) vs the Estimated Transmission Rates (Blue Lines) Over a 2-Aspect Diffusion Network

cascades are known in advance, we can perform quantitative comparisons with the two baselines.

Figure 5 shows the  $F_1$ -measure of the three methods over the three types of generated networks with 512 nodes and 1024 edges. We can observe that MMRate performs inferior to the baselines when the number of cascade is less than 3,000, but with the increase of the number of cascades, MMRate outperforms NETRATE and InfoPath by up to 8%. This is because our model aims to infer the multi-aspect network, and hence each aspect of the network needs enough data to be inferred. One can also see that the  $F_1$ -measure steadily increases as we increase the number of cascades, and stabilizes with around 6,000 cascades.

Figure 6 shows the mean absolute error (MAE) between the estimated networks and the true networks with 5,000 cascades. One can see that in almost all the cases (with only one exception), the MAE of MMRate is smaller than the two baselines. It means that our method can not only more accurately infer the network structure, but also more precisely estimate the transmission rates over the edges.

To further test the effectiveness of MMRate on estimating the transmission rates of each aspect on the multi-aspect diffusion networks, we visualize the estimated and the true transmission rates over the edges of a 2-aspect network in Figure 7. We can observe that MMRate first recovers most edges for each aspect network. MMRate only fails to infer the edges whose estimated transmission rates are zero.

Second, the transmission rates of each aspect estimated by MMRate seem good. The mean absolute error of estimated aspect-sensitive transmission rates are mostly less than 0.3. It indicates that MMRate can effectively estimate the transmission rates over the edges for each aspect network.

### 5.3 Performance Evaluation on Real Data

To evaluate the performance of MMRate on real data, we compare the  $F_1$ -measure achieved by MMRate and three baselines on *DPU* and *DUNF* datasets.

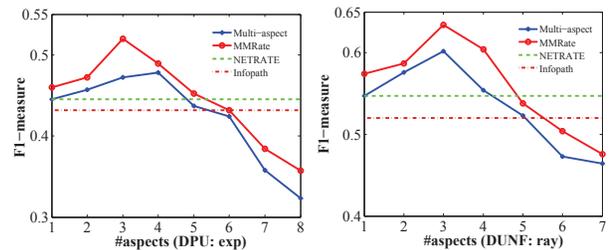


Figure 8: Comparison of  $F_1$ -measure on two Datasets

Figure 8 shows the best experimental results of the three diffusion models on the two datasets. For the dataset *DPU*, the exponential model achieves the best result, and for the dataset *DUNF*, the best result is achieved by the Rayleigh

**Table 6: Aspect-Sensitive Top-5 Users of the 3-Aspect DPU Dataset**

aspect	User ID	# of followers	# of posts	description (occupation or tags)
k=1	1830438495	529,837	10,551	<b>Occupation:</b> a famous writer
	1496850204	3,647,584	13,307	<b>Occupation:</b> the chief editor of <i>New Weekly</i>
	1733596871	198,192	6,381	<b>Occupation:</b> a famous columnist
	1737694433	828,899	10,529	<b>Occupation:</b> a columnist of <i>The Wall Street Journal</i> (Chinese version)
	1641428154	175,533	10,698	<b>Occupation:</b> a famous columnist
k=2	1266321801	29,382,466	7,043	<b>Occupation:</b> a famous actress
	1656809190	27,000,777	2,911	<b>Occupation:</b> a famous actress
	1704116960	24,796,098	1,069	<b>Occupation:</b> a famous TV show host
	1645826702	1323179	7,985	<b>Occupation:</b> writer, screenwriter and director
	1282005885	25,020,327	1,650	<b>Tags:</b> entertainment, art, literature and media
k=3	1813080181	11,709,072	85,012	<b>Tags:</b> mobile internet, Start Your Business (SYB)
	1558226504	1,819,724	8,877	<b>Tags:</b> CEO, design, Start Your Business (SYB)
	1737694433	828,899	10,529	<b>Occupation:</b> a columnist of <i>The Wall Street Journal</i> (Chinese version)
	1653689003	8,350,338	17,520	<b>Occupation:</b> <i>New Weekly</i> (the microblog site of the <i>New Weekly</i> journal)
	1641983193	204,420	11,095	<b>Tags:</b> Internet, E-Commerce and social media

model. The first observation from Figure 8 is that with the increase of cascade number, the  $F_1$ -measure achieved by MMRate first increases, and then gradually decreases. The best performance is achieved with around 3 or 4 aspects. More aspects make the performance worse due to over-fitting. Second, by differentiating the cascades into a small number of aspect groups, MMRate can significantly outperform the baselines. For *DPU*, MMRate outperforms baselines by 15% with  $K$  equals to 3; for *DUNF*, the improvement can be up to 20%. Moreover, the performance of MMRate is always better than the multi-aspect model. It indicates that considering the diffusion patterns of the cascades can better capture the information diffusion process in social networks; and therefore, help us more accurately recover the diffusion networks.

### 5.4 Application of MMRate: A Case Study

MMRate can also be applied to help other data mining tasks, such as aspect-sensitive influential mining and aspect-level influence maximization. Here we give a case study to show how MMRate can be used to help us mining aspect-sensitive influential users in social networks.

Given a diffusion graph  $G^k$  with the aspect  $k$ , the influence of its users can be ranked by a weighted PageRank algorithm as follows

$$\mathbf{R}^k = (1 - d) \cdot \frac{\mathbf{E}}{N^k} + d \cdot \mathbf{A}^k \cdot \mathbf{R}^k \quad (16)$$

where the vector  $\mathbf{R}^k$  denotes the authorities of the nodes,  $\mathbf{E}$  is a column vector containing only ones,  $\mathbf{A}^k$  is the weighted adjacency matrix with each entry  $A_{ij}^k = \alpha_{ij}^k$ ,  $d$  is the damping factor, and  $N^k$  is the number of nodes.

Due to space limitation, we only report the ranking result of the dataset *DPU*. Table 6 lists the top-5 users of each aspect in the 3-aspect network of *DPU*. For each user, we show the user ID, the number of followers, the number of posts and the descriptions for this user. We use the tags or occupations of the users to describe them. As not all the users make their occupations public available, for such users we give their tags. We can see that the ranking result seems reasonable. First, most of the top ranked users have a large number of followers, which implies they are very popular. Some top uses in aspect 2 even have tens of millions followers. Second, most of these users have published more than 6,000 posts. This means these users are also very ac-

tive and have more chances to interact with their friends or followers. Therefore, we can infer more connections to these users and rank them high. Moreover, top users in different aspects are mostly different. The top-5 users of aspect-1 are famous writers, columnists, or editors; the users of aspect-2 are mostly famous actors, TV show hosts, or directors; and users of aspect-3 are mostly interested in commerce, Internet, or new media. Some users are famous in multiple aspects. For example, the user with ID 1737694433 ranks in the top-5 of both aspect-1 and aspect-2. For most top ranking users, however, they are only very influential in just one aspect. It implies that MMRate can effectively categorize different types of users into different diffusion networks.

## 6. RELATED WORK

The problem of inferring the diffusion networks and estimating the diffusion probabilities has been intensively studied in social network analysis [6, 7, 8, 9, 10, 20, 21]. Earlier work in [10, 20, 21] considers the network structure is known and as an input to estimate the diffusion probabilities on the edges through the spreading events. Kazumi et al. proposed a likelihood maximization approach for the learning of the information diffusion probabilities for the Independent Cascade Model [21]. To estimate the diffusion probabilities for a continuous time delay independent cascade (CTIC) model, Saito et al. also proposed an approach to maximize the likelihood of the observed diffusion cascades [20].

Recently, some work has focused on inferring the diffusion networks with hidden structures from the time logs of a collection of cascades [6, 7, 8, 9, 18]. These work can be roughly divided into inferring the network connectivity [7, 9] and inferring both the network structure and the transmission rates between nodes [6, 8, 18]. The representative work on inferring the network structure is NetInfer [7], which formulates this problem as a submodular function maximization problem. A variant of NetInfer to tradeoff between accuracy and scalability is MulTree [9]. NETRATE is a representative approach to infer not only the connectivity of the network, but also to estimate the transmission rates of infections between two connected nodes [6]. This method tries to find the optimal network that maximizes the likelihood of the observed cascades by using convex optimization. A further work based on NETRATE is InfoPath [8], which aims to infer the dynamic diffusion networks.

It has long been studied both in social theories [11, 15] and computer society [1, 22, 23] that influence diffusion and information propagation are related to the types of relationships between users. Tang et al. [22] discerned multi-faceted trust on social media, and proposed a model mTrust to estimate the trust strengths of different facets. To differentiate the social influences from different topics, Tang et al [23] proposed Topic Affinity Propagation (TAP) to model the topic-level social influence on large networks. However, distinguishing different types of cascades for inferring diffusion networks with multiple aspects is less studied. An effort devoted to differentiate cascades with different topics recently is TopicCascade [3]. However, TopicCascade uses topic model to find the topics each cascade related to instead of inferring them from the diffusion data.

## 7. CONCLUSIONS AND FUTURE WORK

Different from tradition network data which are mostly homogeneous, the user interactions on social networks are usually multi-aspected and the cascades show various diffusion patterns, which present challenges to infer the heterogeneous diffusion networks in social media. In this paper, we proposed a novel model MMRate to handle multi-aspect and multi-pattern cascades for inferring multi-aspect diffusion networks. In particular, we studied the existence of various diffusion patterns of cascades and the effects of the different patterns on information diffusion. By incorporating the aspects and patterns of the cascades, the fine-grained model MMRate was proposed. To effectively estimate the parameters, an EM-algorithm was also provided. Experimental results on both synthetic and real social network datasets showed the effectiveness of MMRate compared against state-of-the-art approaches.

For future work it would be interesting to study whether MMRate can be applied to other domains, such as the virus propagation networks. Similar to information diffusion in social networks, the virus propagation networks are also multi-aspected due to the facts that people's immunity differs for different viruses [4]. Different viruses may tend to propagate among different groups of people. We are also interested in applying our model to help other social media applications, such as predicting trending topics with different aspects in social media. In the future, we will investigate whether the aspect-level transmission rates between users estimated by MMRate can help us more accurately predict the future popularity of the topics with different aspects.

## 8. ACKNOWLEDGMENTS

This work is supported by NSFC (Grand Nos. 61170189, 61370126, 61202239), the Research Fund for the Doctoral Program of Higher Education (Grand No. 20111102130003), the Innovation Foundation of BUAA for PhD Graduates (Grand No. YWF-14-YJSY-021), Microsoft Research Asia Fund (Grand No. FY14-RES-OPP-105), NSF through grants CNS-1115234, DBI-0960443, and OISE-1129076, and the Pinnacle Lab at Singapore Management University.

## 9. REFERENCES

- [1] N. Barbieri, F. Bonchi, and G. Manco. Topic-aware social influence propagation models. *Knowledge and information systems*, 37(3):555–584, 2013.
- [2] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, 2010.
- [3] N. Du, L. Song, H. Woo, and H. Zha. Uncover topic-sensitive information diffusion networks. In *AISTATS*, 2013.
- [4] D. D. Erdman. Propagation and identification of viruses. *Topley and Wilson's Microbiology and Microbial Infections*, 2010.
- [5] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 2001.
- [6] M. Gomez-Rodriguez, D. Balduzzi, and B. Scholkopf. Uncovering the temporal dynamics of diffusion networks. In *ICML*, 2011.
- [7] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *KDD*, 2010.
- [8] M. Gomez-Rodriguez, J. Leskovec, and B. Scholkopf. Structure and dynamics of information pathways in online media. In *WSDM*, 2013.
- [9] M. Gomez-Rodriguez and B. Scholkopf. Submodular inference of diffusion networks from multiple trees. In *ICML*, 2012.
- [10] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. Learning influence probabilities in social networks. In *WSDM*, 2010.
- [11] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, May 1973.
- [12] M. Herlihy. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, 24:265–290, August 1998.
- [13] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.
- [14] A. Klondahl, J. J. Potterat, D. E. Woodhouse, J. B. Muth, S. Q. Muth, and W. W. Darrow. Social networks and infectious disease: The colorado springs study. *Social Science and Medicine*, 38(1):79–88, January 1993.
- [15] D. Krackhardt. *The Strength of Strong ties: the Importance of Philos in Organization*. Harvard Business School Press, 1992.
- [16] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. Kronecker graphs: An approach to modeling networks. *Journals of Machine Learning Research*, 11:985–1042, 2010.
- [17] J. Leskovec, A. Singh, and J. Kleinberg. Patterns of influence in a recommendation network. In *PAKDD*, 2006.
- [18] S. A. Myers and J. Leskovec. On the convexity of latent social network inference. In *NIPS*, 2010.
- [19] J. Oh, A. Suasarla, and Y. Tan. Examining the diffusion of user-generated content in online social networks. *Social Science Research Network*, 2008.
- [20] K. Saito, M. Kimura, K. Ohara, and H. Motoda. Learning continuous-time information diffusion model for social behavioral data analysis. In *ACML*, 2009.
- [21] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *KES*, 2008.
- [22] J. Tang, H. Gao, and H. Liu. mtrust: Discerning multi-faceted trust in a connected world. In *WSDM*, 2012.
- [23] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD*, 2009.
- [24] J. Wallinga and P. Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 15(6):509–516, Sep 2004.
- [25] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li. Social influence locality for modeling retweeting behaviors. In *IJCAI*, 2013.