Why It Happened: Identifying and Modeling the Reasons of the Happening of Social Events

Yu Rong
The Chinese University of Hong Kong
yrong@se.cuhk.edu.hk

Hong Cheng
The Chinese University of Hong Kong
hcheng@se.cuhk.edu.hk

Zhiyu Mo
The Chinese University of Hong Kong
zymo@se.cuhk.edu.hk

ABSTRACT

In nowadays social networks, a huge volume of content containing rich information, such as reviews, ratings, microblogs, etc., is being generated, consumed and diffused by users all the time. Given the temporal information, we can obtain the event cascade which indicates the time sequence of the arrival of information to users. Many models have been proposed to explain how information diffuses. However, most existing models cannot give a clear explanation why every specific event happens in the event cascade. Such explanation is essential for us to have a deeper understanding of information diffusion as well as a better prediction of future event cascade.

In order to uncover the mechanism of the happening of social events, we analyze the rating event data crawled from Douban.com, a Chinese social network, from year 2006 to 2011. We distinguish three factors: social, external and intrinsic influence which can explain the emergence of every specific event. Then we use the mixed Poisson process to model event cascade generated by different factors respectively and integrate different Poisson processes with shared parameters. The proposed model, called Combinational Mixed Poisson Process (CMPP) model, can explain not only how information diffuses in social networks, but also why a specific event happens. This model can help us to understand information diffusion from both macroscopic and microscopic perspectives. We develop an efficient Classification EM algorithm to infer the model parameters. The explanatory and predictive power of the proposed model has been demonstrated by the experiments on large real data sets.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*

KDD'15, August 10-13, 2015, Sydney, NSW, Australia.
© 2015 ACM. ISBN 978-1-4503-3664-2/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2783258.2783305.

General Terms

Algorithms, theory, experimentation

Keywords

Information diffusion; event cascade; Poisson process; social influence; external influence; intrinsic influence

1. INTRODUCTION

Social networks have become an important medium for the emergence and diffusion of information. A huge volume of content is being generated and transmitted continuously in different social networks. Understanding information diffusion process from the huge volume of user-generated content remains an important and challenging research problem. Generally, a social event is defined as a user adopting a piece of content at a certain time. For example, in a review site, we can observe that a user rates or reviews a product at a certain time. However, why the user rates or reviews this product and how information related to the product diffuses in the network are unknown. Actually, answering the questions of 'why' is very essential to understand the nature of information diffusion and meaningful in real applications. For example, in viral marketing, knowing the concrete reason for the message spreading can help us to evaluate the viral potential of this message.

In the literature, many studies have been proposed to explain the underlying mechanism of information diffusion. Among them, [7, 9, 10] propose to construct an implicit network between users to explain the observed data. [13, 24] use the stochastic process to model the occurrences of events. [2] adopts embedding technique to map the observed data onto a continuous space. However, the existing works focus on constructing a model which can best describe the observed event cascade as a whole. Details related to the emergence and dissemination of every specific event are missing. For example, for a specific user and content, the current models cannot clearly explain why the user adopts the content at that time.

In order to answer the question "why it happened", in this paper, we propose a generative probabilistic model which aims to explain the emergence of every single event and the formation of event cascade simultaneously in social networks. We distinguish three different factors which trigger an event to happen: *social influence*, *external influence* and *intrinsic influence*. Specifically, social influence causes an event to happen via the user's social connections. External influence triggers an event under the influence of an external

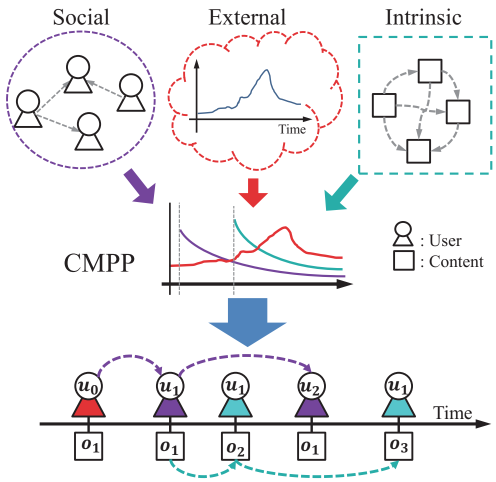

Figure 1: We distinguish three factors to explain the emergence of every social event. In the event cascade above, initially, user u_0 adopts content o_1 under external influence, and then u_1 adopts o_1 under social influence of u_0, and similarly u_2 adopts o_1 under social influence of u_1. After u_1 adopts o_1, u_1 adopts o_2 and o_3 under intrinsic influence.

out-of-network source, such as newspapers, TV stations and online news sites. These two types of influence have been considered in some models [4, 19]. Interestingly, according to our data analysis on the real world social event data, another type of influence can significantly promote the emergence of events as well, but it has not been applied to model information diffusion in the existing literature. This kind of influence reflects user's own preference over time. We call it *intrinsic influence*. Accordingly, an observed event cascade can be partitioned into three sub-event cascades triggered by the above three factors respectively. A sub-event cascade can be treated as a discrete event occurrence sequence over time. The Poisson process is a fundamental stochastic process to describe the occurrences of discrete events in a finite time interval [5]. Hence, we use the mixed Poisson process to model each sub-event cascade separately and propose a combinational stochastic process to model the whole observed event cascade. We call this model **C**ombinational **M**ixed **P**oisson **P**rocess (CMPP). Figure 1 depicts an overview of our model and an example event cascade triggered by the three factors. We also design an efficient Classification EM algorithm to infer the parameters of our model. This inference procedure can provide an explanatory mechanism to the happening of every event. Furthermore, we can apply the model to predict future event cascade.

The main contributions of this work are summarized as follows.

- We distinguish three factors which trigger information diffusion in social networks. The three factors, that is, social, external and intrinsic influence provide a comprehensive mechanism to explain the formation of an event cascade. To the best of our knowledge, this is the first work to take user's intrinsic influence into consideration for modeling information diffusion.
- We propose a combinational model which integrates the three factors based on the mixed Poisson process to model an event cascade, and design a Classification EM algorithm to infer the parameters of the model. This model can not only explain the reason how information diffuses, but also why a specific event happens. Such explanation provides a better understanding of information diffusion process.
- We conduct extensive experiments on large real data sets to demonstrate the explanatory and predictive power of our proposed model. We have the following findings: (1) On average, there are 30.0%, 12.2% and 57.8% events triggered by social, external and intrinsic influence respectively in an event cascade in the Douban data set. (2) Some event cascades are dominated by social influence, while some others are dominated by user's intrinsic influence. These two types of event cascades exhibit very different volume and diffusion patterns. (3) User's intrinsic influence plays a major role in some event cascades, which shows that it is an important factor in information diffusion. (4) Our CMPP model can predict the future event cascade accurately based on the observed cascade.

The rest of the paper is organized as follows. We introduce preliminary concepts in Section 2. Then we discuss how to model the event cascade using a combinational mixed Poisson process in Section 3. In Section 4 we design a Classification EM algorithm to efficiently estimate the parameters of the CMPP model. Extensive experimental results are presented in Section 5. We discuss related work in Section 6 and conclude the paper in Section 7.

2. PRELIMINARY CONCEPTS

An online social network can be modeled as a directed graph $G(\mathcal{U}, \mathcal{V})$, where $\mathcal{U}$ is the node set representing the users, and $\mathcal{V}$ is the edge set representing the user relationships. In addition, we use $\mathcal{O}$ to denote the set of user-generated content in the social network, such as ratings, tweets, or reviews.

An event is defined as a triple $e\langle u, o, t\rangle$ which indicates that user $u \in \mathcal{U}$ adopts content $o \in \mathcal{O}$ at time t. The adopting behavior depends on the network and the content type, for instance, rating a product, retweeting a tweet, etc. An event cascade is a temporally ordered sequence of events, denoted as $E = (e_1, e_2, \ldots)$, where $\forall e_i, e_j$, $t_i < t_j$ if $i < j$. A subsequence of the event cascade observed before time T is denoted as $E_{t_e<T} = (e_1, e_2, \ldots, e_n)$ where $t_n < T$.

From the event cascade, we can observe 'who', 'what' and 'when' of every event. But two questions remain unknown: (1) why does an event happen? and (2) how does the event cascade form? To answer the first question, we distinguish three factors which can trigger an event to happen, that is, *social influence*, *external influence*, and *intrinsic influence*. The three factors are introduced in details in the remainder of this section. To answer the second question, we will propose a model to explain the formation of an event cascade in Section 3.

2.1 Social Influence

In social networks, individuals can be influenced by the actions taken by others. This phenomenon is known as *social influence* [1], and is defined as follows.

DEFINITION 1. ***(Social Influence)*** *Social influence is a social phenomenon that the action of individuals can induce their connections to act in a similar way.*

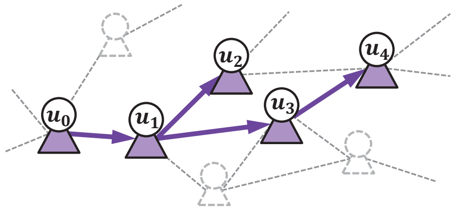

Figure 2: Illustration of the social event cascade for a single piece of content o. In this example, all users who have adopted o (in purple) except user u_0 are influenced by the previous event of the user's friend. The purple arrows indicate the diffusion path. Here, the social event cascade is $E_{soc} = (e_1, e_2, e_3, e_4)$.

A key issue is to quantify the social influence from a user to his followers. For this purpose, we define two variables: *influence rate* and *infection rate* for each user in the social network according to the *social impact theory* [14]. The social impact theory regards an individual in a social community as either the source or target of social influence. In the social network scenario, the user who causes an event to happen is the influence source, while the user who is affected by the event through the social relationship is the influence target.

For a source user, the social influence he brings by adopting an event is quantified by the *influence rate*, which is defined as follows.

DEFINITION 2. ***(Influence Rate)*** *For any user $u \in \mathcal{U}$, the influence rate α_u measures how likely user u's action can induce u's followers to act in a similar way.*

The influence rate α_u depends on the user u who causes the event to happen. For example, by adopting the same event, the influence caused by a famous person will be larger than by an ordinary user.

For a target user, the intention to adopt an event caused by the social influence is measured by the *infection rate*.

DEFINITION 3. ***(Infection Rate)*** *For any user $u \in \mathcal{U}$, the infection rate r_u measures the intention that the user tends to be influenced by the source he follows.*

Similarly, the infection rate also varies from user to user. Some users are more likely to be influenced by others, while others are less likely to be influenced. The social influence diffused over an edge of a social network is determined by both the influence rate of the source and the infection rate of the target.

If we select all the events caused by social influence from an event cascade E and sort them according to the temporal order, we get a subsequence of E. We call it a *social event cascade*, denoted as E_{soc}, where every event in E_{soc} is caused by social influence. Figure 2 illustrates an example of the social event cascade for a single piece of content.

2.2 External Influence

According to a recent study by [19, 4], a user can not only receive information through the links of the social network, but also through the influence of exogenous out-of-network sources, such as newspapers, TV stations, etc. The emergence of information in the social network can be explained by the influence of such external sources. This kind of influence is so called *external influence*.

DEFINITION 4. ***(External Influence)*** *External influence is the impact generated by some out-of-network sources to induce social network users to adopt some content.*

Figure 3 depicts external influence observed from real data. We show the number of ratings from October 1, 2009 to June 1, 2011 for three popular movies: *Avatar*, *Inception* and *Let the Bullets Fly*. We can observe that a sudden peak appears in all three curves. This sudden peak indicates an explosive growth of the number of ratings. An interesting thing is that the corresponding date of each peak is exactly the premiere date of each movie in China. It is obvious that the premiere of a movie is a significant influence on users. Therefore the corresponding relation between the premiere date and the peak in the curve shows the signature that the exogenous sources can induce social network users to adopt some content.

Similar to social influence, we use the *social impact theory* to model external influence. In this case, all the users in social networks are targets. The source can be treated as a hidden node which connects to all the users. For each content o, we can define the *external influence strength* k_o as follows.

DEFINITION 5. ***(External Influence Strength)*** *For any content $o \in \mathcal{O}$, the external influence strength k_o measures how likely an external source related to o can affect all the users in the social network.*

Similarly, we define an *external event cascade*, denoted as E_{ext}, as a subsequence of an event cascade E, where every event in E_{ext} is caused by external influence.

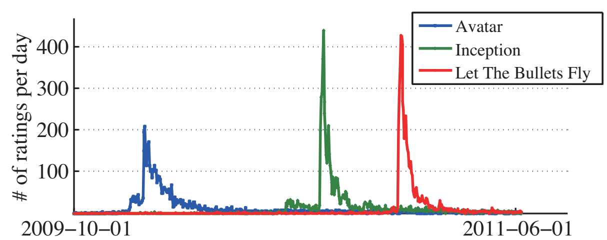

Figure 3: The number of ratings per day for three movies from October 1, 2009 to June 1, 2011. The corresponding date of the peak in each curve is exactly the premiere date of each movie.

2.3 Intrinsic Influence

Besides social influence and external influence which have been studied in the literature, user's behavior may be influenced by their preferences. Users may adopt some content spontaneously based on their current interests or preferences. For example, if a user posts a tweet about an action movie, he may post some related tweets afterwards, such as tweets on other action movies, or movies by the same director or the same leading actors. In this example, the subsequent event cascade of this user is neither triggered by his social connections nor by external influence. Instead, it is influenced by his own preferences. We call this kind of influence *intrinsic influence*.

DEFINITION 6. ***(Intrinsic Influence)*** *Intrinsic influence is the impact generated by the user's preference which induces the user to adopt the content spontaneously.*

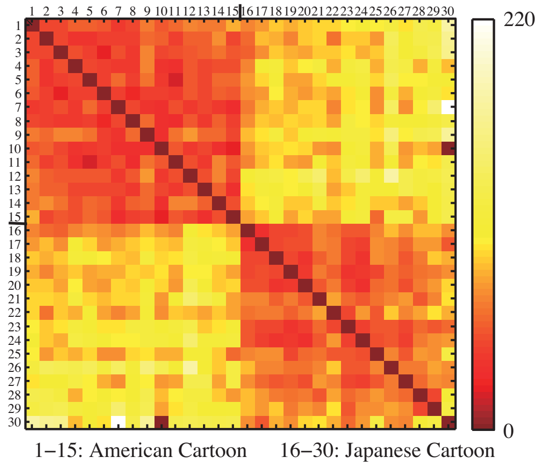

Figure 4: Heat map for the average interval in days between the ratings of two movies made by the same user. id 1-15 denote American cartoon movies, while id 16-30 denote Japanese cartoon movies. The average interval of two movies which belong to the same category is much smaller than that of two movies which belong to different categories.

Intrinsic influence can induce users to adopt content which is similar to his previously adopted content. The effect of intrinsic influence on user behavior can be especially notable within a short time window. To validate this claim, we select all the rating records of 15 movies (which have the largest number of ratings) tagged with "American Cartoon" and "Japanese Cartoon" respectively from Douban during January 10, 2010 and January 1, 2012. Then for any two movies, we find the users who have rated both movies and calculate the average interval between the two ratings in days. Figure 4 shows the heat map in which the color of each grid represents the average interval between two ratings. From Figure 4, we can observe that the area of movies with the same tag is much darker than the movies with different tags. This shows that a previously rated movie would induce users to rate a movie with the same tag in a short period. This observation validates the claim that intrinsic influence can affect user behavior and cause rating events to happen in a short time window.

Intrinsic influence can be measured by the similarity of the two pieces of content. We define an *intrinsic event cascade*, denoted as E_{int}, as a subsequence of an event cascade E, where every event in E_{int} is caused by intrinsic influence.

3. MODELING THE EVENT CASCADE

3.1 Model Overview

According to the three factors, the observed event cascade $E_{t_e<T}$ can be partitioned into three sub-event cascades E_{soc}, E_{ext} and E_{int}. We propose to model these three cascades separately, and then use a combinational model to combine them.

The event cascade can be treated as a discrete event occurrence sequence over time. The Poisson process is a fundamental stochastic process to describe the occurrences of discrete events in a finite time interval [5]. Hence, for any event cascade E, we can model it using the Poisson process with the intensity function $\lambda(t)$. Different from the classical Poisson process whose intensity is constant, the intensity function $\lambda(t)$ is time-dependent since the influence changes with time. For example, the event that happened one day ago would have more influence than a similar event that happened one year ago. Thus, for social influence and intrinsic influence, we add a time decay function $f(\Delta t, \gamma)$ to model the influence decrease over time. There are many parametric models to describe the influence decay [10]. In this paper, we adopt exponential model for social and intrinsic influence decay, i.e., $f(\Delta t, \gamma) = e^{-\gamma \Delta t}$. For external influence, we use a numeric function $\mathcal{M}(t)$ to model the fluctuation of external influence. Since for any timestamp t, the observed event cascade $E_{t_e<t}$ is generated by a stochastic process and the rate of social influence and intrinsic influence depends on historical events, the Poisson processes to model E_{soc} and E_{int} are actually mixed Poisson process [5].

Based on the three separate models, we construct a combinational stochastic process to describe the whole event cascade E with shared parameters. This stochastic process is called **C**ombinational **M**ixed **P**oisson **P**rocess(CMPP). In the following, we first introduce how to model these three sub-event cascades separately. Then we present the combinational model to model the whole event cascade. Table 1 lists the notations used in the paper.

3.2 Modeling Social Influence

For any event $e_i\langle u_i, o_i, t_i\rangle$, we use the intensity function $\lambda_{e_i}(u', t)$ to quantify social influence of e_i to a target user u' at time t as:

$$\lambda_{e_i}(u', t) = \begin{cases} \alpha_{u_i} r_{u'} f(t - t_i, \gamma_\omega) & t > t_i, u' \in \mathcal{U}_{u_i} \\ 0 & \textit{otherwise.} \end{cases} \quad (1)$$

In (1), α_{u_i} is the influence rate of user u_i, $r_{u'}$ is the infection rate of user u', and γ_ω is the parameter to model the influence decay over time.

For the set of followers of u_i, denoted as $\mathcal{U}_{u_i}$, the total social influence caused by e_i at time t is:

$$\lambda_{e_i}(t, \mathcal{U}_{u_i}) = \sum_{u' \in \mathcal{U}_{u_i}} \lambda_{e_i}(u', t). \quad (2)$$

Then for any event $e_i \in E_{t_e<t}$, let E_i be the event cascade caused by e_i under social influence, and $E_{soc} = \bigcup_{i=0}^{|E_{t_e<t}|} E_i$.

Table 1: Notations

Notation	Description
$\mathcal{U}$	user set
$\mathcal{O}$	content set
$E_{t_e<T}$	event cascade before time T
$e_i\langle u_i, o_i, t_i\rangle$	event e_i for user u_i adopting content o_i at time t_i
$\mathcal{U}_{u_i}$	follower set of user u_i
$\lambda_{soc}(t \mid E_{t_e<t})$	intensity of social influence at time t
$\lambda_{int}(t \mid E_{t_e<t})$	intensity of intrinsic influence at time t
$\lambda_{ext}(t)$	intensity of external influence at time t
$\lambda_{all}(u, o, t)$	intensity of user u adopting content o at time t
$f(\Delta t, \gamma)$	decay function
$\mathcal{M}_o(t)$	base external influence function of content o
α_u	influence rate of user u
r_u	infection rate of user u
γ_ω	decay parameter of social influence
γ_ι	decay parameter of intrinsic influence
k_o	external influence strength of content o
$\boldsymbol{z}_j$	latent vector indicating the event trigger of e_j

Each sub-event cascade E_i can be generated by a Poisson process $PP(\lambda_{e_i}(t,\mathcal{U}_{u_i}))$ with $\lambda_{e_i}(t,\mathcal{U}_{u_i})$. Since the event cascades caused by different events are disjoint, these Poisson processes are independent. The superposition of independent Poisson processes is also a Poisson process. Thus the social event cascade E_{soc} can be generated by a Poisson process with the intensity function $\lambda_{soc}(t|E_{t_e<t})$ as:

$$E_{soc} \sim PP(\lambda_{soc}(t|E_{t_e<t})), \tag{3}$$

where $\lambda_{soc}(t|E_{t_e<t})$ is calculated by:

$$\lambda_{soc}(t|E_{t_e<t}) = \sum_{e_i \in E_{t_e<t}} \lambda_{e_i}(t,\mathcal{U}_{u_i}). \tag{4}$$

Hence, from the observed event cascade $E_{t_e<t}$, we can calculate the intensity function $\lambda_{soc}(t|E_{t_e<t})$ at any time t. According to the intensity function, we can generate the social event cascade E_{soc} via the Poisson process.

3.3 Modeling External Influence

Since the external source which triggers an event is usually unobserved in the social network, we use a numeric function $\mathcal{M}_o(t)$ in $[0,1]$ to measure the base value of external influence over time. Based on this function, we can define the intensity function $\lambda_{ext}(o,u,t)$ for user u to content o at time t as:

$$\lambda_{ext}(o,u,t) = r_u k_o \mathcal{M}_o(t). \tag{5}$$

In (5), r_u is the infection rate which reflects user u's intention to be influenced, and k_o is the external influence strength of content o. The external event cascades triggered by different content are independent of each other. Thus the intensity function $\lambda_{ext}(t)$ of the Poisson process which generates E_{ext} is:

$$\lambda_{ext}(t) = \sum_{o\in\mathcal{O}}\sum_{u\in\mathcal{U}} \lambda_{ext}(o,u,t). \tag{6}$$

3.4 Modeling Intrinsic Influence

Similar to social influence, for a specific content $o \in \mathcal{O}$ and user u_i, the intensity function $\lambda_{e_i}(o,t)$ of the intrinsic influence by a previous event $e_i\langle u_i, o_i, t_i\rangle$ to u_i can be defined as:

$$\lambda_{e_i}(o,t) = \begin{cases} sim(o_i,o) f(t-t_i,\gamma_\iota) & t > t_i \\ 0 & t \le t_i. \end{cases} \tag{7}$$

In (7), γ_ι is the time decay parameter. $sim(o_i,o)$ is the similarity between the two pieces of content o_i and o. A higher similarity indicates that the user is more likely to adopt the content o right after e_i happens.

Similar to E_{soc}, E_{int} can be modeled by a Poisson process with the intesity function $\lambda_{int}(t|E_{t_e<t})$. For all the content in $\mathcal{O}$, the intensity function $\lambda_{e_i}(t,\mathcal{O})$ is:

$$\lambda_{e_i}(t,\mathcal{O}) = \sum_{o\in\mathcal{O}} \lambda_{e_i}(o,t). \tag{8}$$

Therefore, E_{int} can be generated by a Poisson process with the intensity function $\lambda_{int}(t|E_{t_e<t})$:

$$\lambda_{int}(t|E_{t_e<t}) = \sum_{e_i \in E_{t_e<t}} \lambda_{e_i}(t,\mathcal{O}). \tag{9}$$

3.5 Combinational Mixed PP Model

Based on the three separate models corresponding to the three factors, we can construct a combinational model to model the whole event cascade. For the observed event cascade $E_{t_e<T}$, we assume E is generated by a stochastic process $\mathcal{X}_{\{\mathcal{U},\mathcal{O}\}}(t)$ as follows:

$$\begin{aligned} E_{t_e<t} &\sim \mathcal{X}_{\{\mathcal{U},\mathcal{O}\}}(t) \\ \mathcal{X}_{\{\mathcal{U},\mathcal{O}\}}(t+\Delta t) &= PP(\lambda_{soc}(t+\Delta t|E_{t_e<t})) \\ &\quad + PP(\lambda_{ext}(t+\Delta t)) + PP(\lambda_{int}(t+\Delta t|E_{t_e<t})). \end{aligned} \tag{10}$$

In (10), $\mathcal{X}_{\{\mathcal{U},\mathcal{O}\}}(t)$ consists of three Poisson processes $PP(\cdot)$ which use the intensity function $\lambda(t|\cdot)$ to model the event cascades $E_{soc}, E_{ext}, E_{int}$ generated by social influence, external influence and intrinsic influence, respectively. The whole event cascade is the union of the three cascades, i.e., $E_{t_e<T} = E_{soc} \cup E_{ext} \cup E_{int}$.

The three Poisson processes are not independent of each other, since from the definition, $\lambda_{soc}(t|E_{t_e<t})$ and $\lambda_{int}(t|E_{t_e<t})$ are based on the historical events generated by $\mathcal{X}_{\{\mathcal{U},\mathcal{O}\}}(t)$. Thus $\mathcal{X}_{\{\mathcal{U},\mathcal{O}\}}(t)$ is the combinational model to explain how the event cascade forms.

3.6 Likelihood Function

Based on the CMPP model described above, for any user u and content o, the intensity function $\lambda_{all}(u,o,t)$ of user u adopting content o at time t under the influence of the three factors is:

$$\lambda_{all}(u,o,t) = \sum_{e_i\in E_o} \lambda_{e_i}(u,t) + \sum_{e_i\in E_u} \lambda_{e_i}(o,t) + \lambda_{ext}(o,u,t). \tag{11}$$

In (11), E_o is the event cascade related to content o, while E_u is the event cascade generated by u.

From (11), for any event $e_j\langle u_j, o_j, t_j\rangle$, we can write the probability density $p(e_j|E_{t_e<t_j},\Theta)$ as:

$$p(e_j|E_{t_e<t_j},\Theta) = \lambda_{all}(u_j,o_j,t_j) e^{-\int_0^{t_j} \lambda_{all}(u_j,o_j,t)dt}. \tag{12}$$

In (12), $\Theta = \{\boldsymbol{\alpha}, \boldsymbol{R}, \boldsymbol{\gamma}, \boldsymbol{K}\}$ is the parameter set of CMPP. $\boldsymbol{\alpha} = \{\alpha_u\}_{u\in\mathcal{U}}$ is the influence rate set of users. $\boldsymbol{R} = \{r_u\}_{u\in\mathcal{U}}$ is the infection rate set of users. $\boldsymbol{\gamma} = \{\gamma_\omega, \gamma_\iota\}$ is the decay parameter set. $\boldsymbol{K} = \{k_o\}_{o\in\mathcal{O}}$ is the external influence strength set of content.

Therefore, for the observed event cascade $E_{t_e<T}$, the joint probability density $p(E_{t_e<T}|\Theta)$ is:

$$p(E_{t_e<T}|\Theta) = \prod_{e_j\in E_{t_e<T}} p(e_j|E_{t_e<t_j},\Theta). \tag{13}$$

The integral part in (12) can be computed analytically:

$$\begin{aligned} \int_0^{t_j} \lambda_{all}(u_j,o_j,t)dt &= \sum_{e_i\in E_{o_j}} \int_{t_i}^{t_j} \lambda_{e_i}(u_j,t)dt \\ &+ \sum_{e_i\in E_{u_j}} \int_{t_i}^{t_j} \lambda_{e_i}(o_j,t)dt + \int_0^{t_j} \lambda_{ext}(o_j,u_j,t)dt, \end{aligned}$$

where:

$$\int_{t_i}^{t_j} \lambda_{e_i(u_j,t)}dt = \frac{r_{u_j}\alpha_{u_i}}{\gamma_\omega}(1-e^{-\gamma_\omega(t_j-t_i)}), \tag{14}$$

$$\int_{t_i}^{t_j} \lambda_{e_i(o_j,t)}dt = \frac{sim(o_i,o_j)}{\gamma_\iota}(1-e^{-\gamma_\iota(t_j-t_i)}), \tag{15}$$

$$\int_0^{t_j} \lambda_{ext}(o_j,u_j,t)dt = r_{u_j}k_{o_j}\int_0^{t_j} \mathcal{M}_{o_j}(t)dt. \tag{16}$$

Based on the joint probability density function (13), given the observed event cascade $E_{t_e<T}$, our target is to find the parameters set Θ to maximize the likelihood function $\mathcal{L}(E_{t_e<T}|\Theta) = \log p(E_{t_e<T}|\Theta)$, i.e.,

$$\Theta^* = \arg\max_{\Theta} \mathcal{L}(E_{t_e<T}|\Theta). \quad (17)$$

4. INFERENCE

We use the Classification EM (CEM) algorithm [3] to infer the parameter set in (17). Since it is very hard to infer the parameters from $\mathcal{L}(E_{t_e<T}|\Theta)$ directly, for any event e_j, we assign a latent $(n+1)\times 1$ vector $\boldsymbol{z}_j$ to indicate which event has triggered e_j, $\sum_{i=0}^{n} \boldsymbol{z}_j(i) = 1$. $\boldsymbol{z}_j(i) = 1$ indicates that event e_j is triggered by event e_i. Here i is the index of event in $E_{t_e<T}$. $\boldsymbol{z}_j(0) = 1$ means that event e_j is triggered by external influence. Thus we can rewrite the intensity function $\lambda_{all}(u_j, o_j, t_j, \boldsymbol{z}_j)$ for event e_j with $\boldsymbol{z}_j$ as:

$$\lambda_{all}(u_j, o_j, t_j, \boldsymbol{z}_j) = \sum_{e_i \in E_{o_j}} z_j(i)\lambda_{e_i}(u_j, t_j) + \sum_{e_i \in E_{u_j}} z_j(i)\lambda_{e_i}(o_j, t_j) + z_j(0)\lambda_{ext}(o_j, u_j, t_j), \quad (18)$$

and the joint log-likelihood function $\mathcal{L}(E_{t_e<T}, \mathcal{Z}|\Theta)$ for the event cascade $E_{t_e<T}$ and the latent variable set $\mathcal{Z} = \{\boldsymbol{z}_i\}_{e_i \in E_{t_e<T}}$ as:

$$\mathcal{L}(E_{t_e<T}, \mathcal{Z}|\Theta) = \sum_{e_j \in E_{t_e<T}} \log(p(e_j, \boldsymbol{z}_j | E_{t_e<t_j}, \Theta)) = \sum_{e_j \in E_{t_e<T}} \log \lambda_{all}(u_j, o_j, t_j, \boldsymbol{z}_j) - \int_0^{t_j} \lambda_{all}(u_j, o_j, t, \boldsymbol{z}_j)dt. \quad (19)$$

By adopting the CEM algorithm [3], we can infer the parameters and the latent variables based on (19).

E-step. Given the current parameters set Θ^{k-1}, we compute the posterior distribution of $\boldsymbol{z}_j$:

$$p(\boldsymbol{z}_j(i) = 1|e_j, E_{t_e<t_j}, \Theta^{k-1}) = \frac{p(e_j, \boldsymbol{z}_j(i) = 1|E_{t_e<t_j}, \Theta^{k-1})}{p(e_j|E_{t_e<t_j}, \Theta^{k-1})}.$$

C-step. In order to find the major factor that causes an event to happen, for any $\boldsymbol{z}_j$, we assign $\boldsymbol{z}_j(i) = 1$ which provides the maximum posterior probability of (20). That is:

$$\boldsymbol{z}_j = \arg\max_{\boldsymbol{z}_j} p(\boldsymbol{z}_j(i) = 1|e_j, E_{t_e<t_j}, \Theta^{k-1}). \quad (20)$$

M-step. We estimate the parameter set Θ in the M-step. Indeed, we use the well-known coordinate descent method to infer the influence rate $\boldsymbol{\alpha}$, the infection rate $\boldsymbol{R}$ and the external influence strength $\boldsymbol{K}$. We find the partial derivative equation $\frac{\partial \mathcal{L}}{\partial \alpha_u} = 0$, $\frac{\partial \mathcal{L}}{\partial r_u} = 0$ and $\frac{\partial \mathcal{L}}{\partial k_o} = 0$ can be computed analytically. Thus the updating rules for parameters are as follows:

$$\alpha_u = \frac{\left|\sum_{e_i \in E_u} \sum_{e_j \in E_{o_i}} z_j(i)\right|}{\sum_{e_i \in E_u} \sum_{e_j \in E_{o_i}} z_j(i) r_{u_j} \frac{(1-e^{-\gamma_\omega(t_j - t_i)})}{\gamma_\omega}}, \quad (21)$$

$$r_u = \frac{\left|\sum_{e_j \in E_u} \sum_{e_i \in E_{o_j}} z_j(i)\right| + \left|\sum_{e_j \in E_u} z_j(0)\right|}{\sum_{e_j \in E_u} \sum_{e_i \in E_{o_j}} z_j(i)A' + \sum_{e_j \in E_u} z_j(0)A''}, \quad (22)$$

$$k_o = \frac{\left|\sum_{e_j \in E_o} z_j(0)\right|}{\sum_{e_j \in E_o} z_j(0) r_{u_j} \int_0^{t_j} \mathcal{M}_o(t)dt}, \quad (23)$$

where $A' = \alpha_{u_i} \frac{(1-e^{-\gamma_\omega(t_j - t_i)})}{\gamma_\omega}$, $A'' = (k_{o_j} \int_0^{t_j} \mathcal{M}_{o_j}(t)dt)$.

For the decay parameter $\boldsymbol{\gamma}$, $\frac{\partial \mathcal{L}}{\partial \gamma} = 0$ cannot be computed analytically. We adopt the Newton's method to calculate γ_ω and γ_ι [13].

We present the procedure of model parameter inference in Algorithm 1.

Algorithm 1 Model Parameter Inference

Input: $E_{t_e<T}$
Output: $\Theta = \{\boldsymbol{\alpha}, \boldsymbol{R}, \boldsymbol{\gamma}, \boldsymbol{K}\}$, $\mathcal{Z}$
  Initialize $\mathcal{Z}^{(0)}$, $\Theta^{(0)}$
  while not converged **do**
    for $e_j \in E_{t_e<T}$ **do**
      $\boldsymbol{z}_j = \arg\max_{\boldsymbol{z}_j} p(\boldsymbol{z}_j(i) = 1|e_j, E_{t_e<t_j}, \Theta^{k-1})$
    end for
    $\boldsymbol{\alpha}^k \leftarrow$ updated by (21)
    $\boldsymbol{R}^k \leftarrow$ updated by (22)
    $\boldsymbol{K}^k \leftarrow$ updated by (23)
    $\boldsymbol{\gamma}^k \leftarrow$ updated by the Newton's method.
  end while

5. EXPERIMENTS

5.1 Data Set Description

Our experiment is conducted on two real world data sets: Douban[1] [25] and Epinions[2].

Douban data set. Douban is a Chinese SNS website allowing users to generate and share information such as ratings and reviews related to movies, books, etc. This data set contains three components: (1) social network structure including users and their follow relationship, (2) user rating records with timestamp which is measured by days from year 2006 to 2011, and (3) movie information such as directors and categories. We extract the tag set of each movie and use the Jaccard Coefficient on the tag set to measure the similarity between movies. Meanwhile, we use the movie name as the keyword to query the Google Trends data of the target period as the reference of external influence. The intuition is that the strength of external influence can be reflected by the search volume in the Internet. As preprocessing, we filter users who adopted less than 30 events and movies which appeared in less than 1000 events.

Epinions data set. Epinions is a consumer review site which helps users make the purchase decisions. This data set contains three components: (1) trust/distrust relationship between users, (2) user rating records for the articles with timestamp which is measured by days from year 2001 to 2003, and (3) author and subject information of the articles. We extract the social network structure from the trust statement made by users. If user u_i trusts user u_j, we consider there exists an edge from u_i to u_j in the social network. Then we utilize the Jaccard Coefficient to measure the similarity between articles based on the author and

[1]http://www.douban.com
[2]http://www.trustlet.org/wiki/Extended_Epinions_dataset

subject information. Since the textual content of articles is not available in this data set, we cannot search any external source to evaluate external influence. In order to address this problem, we randomly divide the data set into two parts: the experimental part (80%) and the reference part (20%). We use the experimental part to conduct our experiment and the reference part as the reference of external influence. For any article o at time period $[t, t+\Delta t]$, we use the ratio of users who have rated article o during $[t, t+\Delta t]$ versus users who have not rated o at time t to represent external influence during $[t, t+\Delta t]$ for article o. As preprocessing, we filter users who adopted less than 10 events and articles which were rated less than 200 times.

Table 2 lists the statistics of these two data sets.

Table 2: Statistics of the Data Sets

	#users	#edges	#events	#days
Douban	7,892	195,457	1,018,567	2,155
Epinions	2,939	1,734	109,320	928

5.2 Explanatory Experiment

We conduct the experiment on the Douban data set to demonstrate the explanatory power of our model.

5.2.1 *Event Type Breakdown*

We apply our CMPP model on the event cascades in Douban, which gives a label to every event indicating the influence factor that triggers the event. For each movie, we identify the specific event cascade in which all events are related to this movie from the whole event cascade. Then we calculate the percentage of events triggered by external influence, intrinsic influence and social influence for each movie related event cascade, respectively. The mean and variance of percentage of the three event types over all event cascades in Douban are plotted in Figure 5. We can make the following observations:

- On average in an event cascade there are 12.2%, 57.8%, and 30.0% events triggered by external influence, intrinsic influence and social influence, respectively.
- The variance of intrinsic and social influence is much larger than external influence.
- The variance of external influence is very small. It indicates that the influence from exogenous sources to this social network is stable for all content.

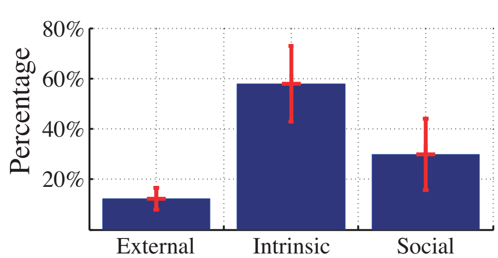

Figure 5: The mean and variance of percentage of each event type over all event cascades in Douban.

5.2.2 *Comparison of Four Event Cascades*

Since the proportion of intrinsic and social influence varies greatly from content to content, different diffusion patterns may exist for different types of content. In order to inspect the details of the diffusion pattern, firstly we conduct a case study. We compare the event cascades of four movies observed during March 2010 - September 2010: *12 Angry Men(12ANG)* (1957), *Mulholland Drive(MD)* (2001), *How to Train Your Dragon(HTYD)* (2010) and *Echoes of The Rainbow(ETR)* (2010). The first two movies are old ones, while the other two are new in the observed period as they were released in March 2010. The event type breakdown is shown in Figure 6. We have the following observations.

- User's intrinsic influence dominates in the diffusion process of the two old movies, while social influence dominates in the diffusion process of the two newly-released movies. This shows that people are more likely to be influenced by their social connections in adopting new movies than old ones. User's behavior of adopting old movies is spontaneously and largely induced by their intrinsic influence.
- The number of events related to the two old movies is much smaller than that of the two newly-released movies. This volume difference can be explained by both strong external influence and social influence in the diffusion process of the two new movies.

We plot the detailed event cascade structure of these four movies in Figure 7. Here, a point represents an event, i.e., a user rating/reviewing a movie, and the color of the point indicates the type of influence that triggers this event. If an event influences another through the social connection, we add an arrow pointing from the source to the target. Figure 8 gives a zoom-in plot of the social diffusion path between events for the movie *HTYD*. From these figures, we can clearly observe the different diffusion processes of the two old movies and two new movies as discussed above.

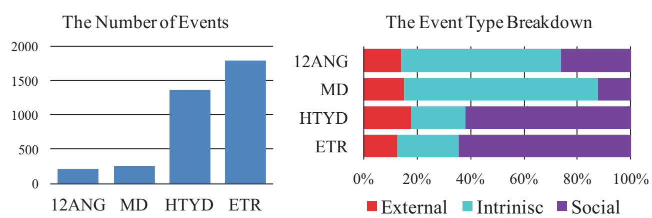

Figure 6: The event type breakdown of four movies.

5.2.3 *Social Dominated versus Intrinsic Dominated Event Cascades*

From the previous case study we find that some event cascades are dominated by social influence, while some others are dominated by intrinsic influence. We call them *social dominated event cascade* and *intrinsic dominated event cascade*, respectively. Indeed, these two types of event cascades exhibit very different diffusion patterns. In this experiment, these two types of event cascades are compared analytically.

For an event cascade, we define the *i/s ratio* as the ratio between the number of intrinsic influence triggered events and the number of social influence triggered events. Then we pick 50 event cascades with the smallest i/s ratio, and another 50 event cascades with the largest i/s ratio. The first set with i/s ratio < 1 corresponds to the social dominated event cascades, and the second set with i/s ratio > 1 corresponds to the intrinsic dominated event cascades. We plot the normalized cumulative frequency of events over time in an event cascade in Figure 9(a) and (b) for the social dominated and intrinsic dominated event cascades respectively.

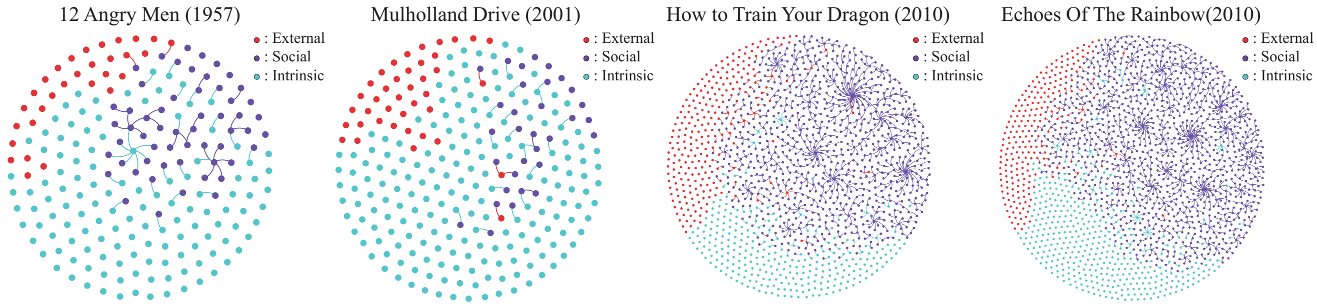

Figure 7: The event cascade structure of four movies. For the two old movies, intrinsic influence is the major factor of the emergence of events, while for the two new movies, social influence is the major factor of the emergence of events. The diffusion pattern is different for these two types of movies.

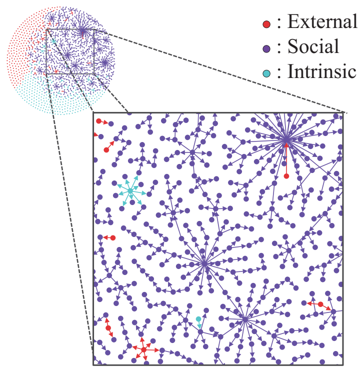

Figure 8: Zoom-in plot of the social diffusion paths for *HTYD*. We can observe many DAGs as the diffusion paths of this content in the original social graph.

Clearly these two figures exhibit very different event growth patterns. For the social dominated event cascades, there is a sharp increase in the number of events, or burst, at a certain time point. The explosive growth can be explained by the strong and widespread social influence. In contrast, for the intrinsic dominated event cascades, the growth of events over time is steady and smooth. This is because user behavior is induced by user's intrinsic influence. It is not likely to cause a burst in the social network.

Identifying these two types of event cascades is very essential for real applications, such as long tail recommendation, popular event detection, viral marketing, etc. Our CMPP model can naturally distinguish these two types of event cascades and give a sound explanation.

5.2.4 Remarks on Intrinsic Influence

According to Figure 5, it is interesting to see that intrinsic influence is the major factor of the event emergence in Douban data. This result proves that intrinsic influence is very important and cannot be ignored in information diffusion. On the other hand, such high proportion of intrinsic influence triggered events may be partially attributed to a feature in Douban. That is, users can easily jump from one movie to another through links such as 'movie by the same director', or 'people who like this movie also like'.

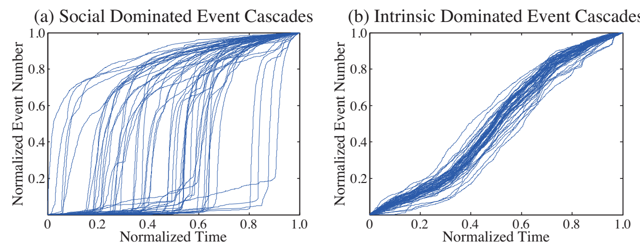

Figure 9: The normalized cumulative frequency curve. (a) The social dominated event cascades. (b) The intrinsic dominated event cascades. In (a) we can observe an explosive growth of the number of events while the growth in (b) is steady and smooth.

Table 3 lists four popular diffusion path examples triggered by intrinsic influence identified by our model. By 'popular' we mean each transition from one movie to another in these paths is adopted by at least 45 users in the Douban data set.

5.2.5 Summary

In this experiment, we show the event type breakdown in the event cascades in Douban, and compare the different diffusion patterns of the event cascades of four movies, which are dominated by intrinsic influence and social influence respectively. Through the explanatory experiment, we demonstrate that our CMPP model can provide accurate explanation to two important questions: (1) how information diffuses in the social network; and (2) which factor triggers a single event to happen. Such concrete explanation reveals the subtle structure and the growth pattern of an event cascade, leading to a deeper understanding of the information diffusion process.

5.3 Predictive Experiment

In this experiment, we apply our CMPP model for future event cascade prediction on the Douban and Epinions data

Table 3: Four Example Diffusion Paths by Intrinsic Influence

Diffusion Path	Remark
Waterloo Bridge→Gone with the Wind→Big Fish→Scent of a Woman	Drama movies
Pulp Fiction → Leon → The Shawshank Redemption	Douban top 100 movies
Chungking Express→Flirting Scholar →A Chinese Ghost Story→Initial D	Hong Kong movies
The Godfather → The Godfather: Part II → The Godfather: Part III	Movies of the same series

sets to demonstrate the predictive power of our model. The event cascade prediction problem is defined as:

Definition 7. ***(Event Cascade Prediction)*** *Event cascade prediction is to predict the future event cascade* $E_{T \le t_e < T+\Delta t}$ *in time* $[T, T+\Delta t)$*, based on the observed event cascade* $E_{t_e<T}$.

More specifically, in our experiment, we aim to predict the content set $\mathcal{O}^u_{[T,T+\Delta t)}$ that user u adopts in $[T, T+\Delta t)$. The prediction can be regarded as a retrieval task. For each user u, our model returns a content list L_u sorted in the descending order of probability which indicates how likely user u will adopt this content in $[T, T+\Delta t)$. The probability is calculated according to the parameter set Θ which is learned from $E_{t_e<T}$.

We use Mean Average Precision (MAP) to evaluate the prediction performance similar to [2]. We define $\mathcal{P}_{u,k}$ as the precision at rank k in L_u, i.e., the percentage of true positive content among the top k pieces of content in L_u. Thus, the MAP is defined as:

$$MAP = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\sum_{o \in \mathcal{O}^u_{[T,T+\Delta t)}} \mathcal{P}_{u,\sigma(o,L_u)}}{\left|\mathcal{O}^u_{[T,T+\Delta t)}\right|}. \tag{24}$$

In (24), $\sigma(o, L_u)$ indicates the rank of content o in L_u, $|\cdot|$ indicates the number of elements in the target set.

From each data set, we extract the event cascade $E_{T-\Delta\xi \le t_e < T}$ in $[T-\Delta\xi, T)$ as the training event cascade, and the event cascade $E_{T \le t_e < T+\Delta t}$ in $[T, T+\Delta t)$ as the test event cascade. Here $\Delta\xi$ is the training interval and Δt is the test interval. The experimental configuration for the two data sets is shown in Table 4.

Table 4: Experimental Configuration

	Training Interval (days)	Test Interval (days)
Douban	600	3
Epinions	300	1

5.3.1 Candidate Methods for Comparison

We compare our CMPP model with the following information diffusion models.

- **INFOPATH**: INFOPATH [10] is an online algorithm that relies on stochastic convex optimization to solve the dynamic network inference problem. This method can infer the transmission rate which models how frequently information spreads between two users.
- **ProfileRank**: ProfileRank (P-RANK) [23] is a PageRank-based model to identify the influential users and relevant content in the event cascade data. This method can calculate the likelihood that a user will adopt the content in the future.

5.3.2 Experimental Results

In order to reduce the influence of sampling, we select 5 different timestamps for the parameter T to extract the training and test event cascades. Table 5 reports the MAP results for each test case on the two data sets. The row 'Average' reports the average MAP performance for all the cases on each data set.

As shown in Table 5, our model CMPP consistently and significantly outperforms the other models on all the cases of these two data sets. The superior performance demonstrates the predictive power of our model. The poor predictive performance of the other two models can be explained by the following reasons.

Table 5: MAP Results (Higher is better)

Epinions	CMPP	INFOPATH	P-RANK
Case 1	**0.0526**	0.0463	0.0304
Case 2	**0.0681**	0.0283	0.0381
Case 3	**0.0643**	0.0270	0.0334
Case 4	**0.0541**	0.0165	0.0164
Case 5	**0.1004**	0.0178	0.0117
Average	**0.0679**	0.0272	0.0260
Douban	CMPP	INFOPATH	P-RANK
Case 1	**0.2816**	0.0219	0.0445
Case 2	**0.1147**	0.0162	0.0642
Case 3	**0.1128**	0.0194	0.0567
Case 4	**0.4456**	0.0142	0.0306
Case 5	**0.3826**	0.0121	0.0303
Average	**0.2675**	0.0168	0.0453

- The INFOPATH model ignores the social network structure and tries to build a transmission network to explain the formation of the event cascade. However, this transmission network allows information diffusion between any two users. Thus the INFOPATH model cannot avoid the overfitting problem which degrades the prediction performance.
- The ProfileRank model only considers the relationship between users and content, but ignores the influence from the social network and the external source. That is why the ProfileRank model achieves a poor prediction performance.

6. RELATED WORK

There are many existing studies related to information diffusion in online social networks, which can be classified into two main categories: (1) graph-explanatory model and (2) graph-predictive model.

Graph-explanatory model. The graph-explanatory model treats the social network structure as a strong prior for information diffusion. *Independent Cascades (IC)* [6, 21] and *Linear Threshold (LT)* [15] are two seminal models. The two models view the happening of events in different ways, i.e., the IC model is sender-dominated, while the LT model is receiver-dominated. Various extensions to these two models have been proposed, such as *AsIC* and *AsLT* [20], which incorporate temporal information. [17] considers both positive and negative relationships in a social network to solve the influence diffusion and maximization problems. [12] models the characteristics of sender and receiver in the diffusion simultaneously. [22] integrates the node attributes to estimate the diffusion probability. [11, 16] consider the event content to extract more information to infer the diffusion patterns. [19] considers both the endogenous and exogenous influence to model the diffusions.

Graph-predictive model. The graph-predictive model ignores the social network structure and infers the implicit connections between users only based on the observed event cascade. [7, 9] focus on estimating the transmission rate between users to best explain the observed data. [10] extends the model in [7] and estimates the structure and the temporal dynamics of the network. [8] aims to infer the individual transmissions to trace the paths of news diffusion among news media sites and blogs. Some works [2, 18] try to map the temporal information onto a continuous space and learn a diffusion kernel to explain the observed data. Another approach [24] tries to utilize Poisson process to model the event cascade related to a single piece of content. [13] extends [24]'s model to describe the multiple event cascades by sharing the parameters.

The most related works to ours are [19, 24]. [24] also adopts Poisson process to construct the model, but it only considers social influence and does not utilize social connections. [19] studies the effect of external influence and social influence. However it does not consider the user preference.

7. CONCLUSION

In this paper, we investigate the reasons why an event happens in social networks. We distinguish three factors as the event triggers to explain the formation of an event cascade. To integrate these factors, we propose a combinational model based on the mixed Poisson process to model the event cascade. Our model can not only explain how information diffuses, but also obtain the concrete reasons why a specific event happens. Furthermore, we develop an efficient Classification EM algorithm to infer the model parameters. Extensive experiments and analysis on large data sets demonstrate that our model can (1) provide sound explanation to the observed event cascade and reveal the subtle structure of the diffusion process, and (2) make accurate prediction of the future event cascade.

8. ACKNOWLEDGMENTS

This work is supported by the Hong Kong Research Grants Council (RGC) General Research Fund (GRF) Project No. CUHK 411211, and the Chinese University of Hong Kong Direct Grant No. 4055015 and 4055048.

9. REFERENCES

[1] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *KDD*, pages 7–15, 2008.

[2] S. Bourigault, C. Lagnier, S. Lamprier, L. Denoyer, and P. Gallinari. Learning social network embeddings for predicting information diffusion. In *WSDM*, pages 393–402, 2014.

[3] G. Celeux and G. Govaert. A classification em algorithm for clustering and two stochastic versions. *Computational Statistics and Data Analysis*, 14(3):315–332, 1992.

[4] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS*, 105(41):15649–15653, 2008.

[5] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume I.* Springer, New York, 2003.

[6] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.

[7] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *ICML*, pages 561–568, 2011.

[8] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *KDD*, pages 1019–1028, 2010.

[9] M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf. Modeling information propagation with survival theory. In *ICML*, pages 666–674, 2013.

[10] M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf. Structure and dynamics of information pathways in online media. In *WSDM*, pages 23–32, 2013.

[11] A. Guille and H. Hacid. A predictive model for the temporal dynamics of information diffusion in online social networks. In *WWW*, pages 1145–1152, 2012.

[12] T.-A. Hoang and E.-P. Lim. Retweeting: An act of viral users, susceptible users, or viral topics? In *SDM*, pages 569–577, 2013.

[13] T. Iwata, A. Shah, and Z. Ghahramani. Discovering latent influence in online social activities via shared cascade poisson processes. In *KDD*, pages 266–274, 2013.

[14] S. J. Karau and K. D. Williams. Social loafing: Research findings, implications, and future directions. *Current Directions in Psychological Science*, 4(5):134–140, 1995.

[15] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.

[16] C. Lagnier, L. Denoyer, E. Gaussier, and P. Gallinari. Predicting information diffusion in social networks using content and user's profiles. In *ECIR*, pages 74–85, 2013.

[17] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *WSDM*, pages 657–666, 2013.

[18] H. Ma, H. Yang, M. R. Lyu, and I. King. Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM*, pages 233–242, 2008.

[19] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *KDD*, pages 33–41, 2012.

[20] K. Saito, M. Kimura, K. Ohara, and H. Motoda. Generative models of information diffusion with asynchronous time-delay. In *ACML*, pages 193–208, 2010.

[21] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *KES*, pages 67–75. Springer, 2008.

[22] K. Saito, K. Ohara, Y. Yamagishi, M. Kimura, and H. Motoda. Learning diffusion probability based on node attributes in social networks. In *FIS*, pages 153–162, 2011.

[23] A. Silva, S. Guimarães, W. Meira Jr, and M. Zaki. Profilerank: finding relevant content and influential users based on information diffusion. In *SNA-KDD*, page 2, 2013.

[24] A. Simma and M. I. Jordan. Modeling events with cascades of poisson processes. In *UAI*, pages 546–555, 2010.

[25] E. Zhong, W. Fan, J. Wang, L. Xiao, and Y. Li. Comsoc: Adaptive transfer of user behaviors over composite social network. In *KDD*, pages 696–704, 2012.