User modeling plays an important role in delivering customized web services to the users and improving their engagement. However, most user models in the literature do not explicitly consider the temporal behavior of users. More recently, continuous-time user modeling has gained considerable attention and many user behavior models have been proposed based on temporal point processes. However, typical point process-based models often considered the impact of peer influence and content on the user participation and neglected other factors. Gamification elements are among those factors that are neglected, while they have a strong impact on user participation in online services. In this article, we propose interdependent multi-dimensional temporal point processes that capture the impact of badges on user participation besides the peer influence and content factors. We extend the proposed processes to model user actions over the community-based question and answering websites, and propose an inference algorithm based on Variational-Expectation Maximization that can efficiently learn the model parameters. Extensive experiments on both synthetic and real data gathered from Stack Overflow show that our inference algorithm learns the parameters efficiently and the proposed method can better predict the user behavior compared to the alternatives.

CCS Concepts: • Mathematics of computing → Probability and statistics; • Information systems → Collaborative and social computing systems and tools; Web applications; • Computing methodologies → Machine learning approaches;

Additional Key Words and Phrases: User modeling, user profiling, temporal point process, gamification, badge, stack overflow, variational EM

ACM Reference format:
Ali Khodadadi, Seyed Abbas Hosseini, Erfan Tavakoli, and Hamid R. Rabiee. 2018. Continuous-Time User Modeling in Presence of Badges: A Probabilistic Approach. ACM Trans. Knowl. Discov. Data. 12, 3, Article 37 (March 2018), 30 pages.
https://doi.org/10.1145/3162050

1 INTRODUCTION
In recent years, online social media have become more and more popular. Social media websites such as Facebook, Twitter, Stack Overflow, and Wikipedia have millions of users engaged in various activities. In these systems, the value is created from voluntary user contributions and they need to engage users as much as possible. Hence, it is of key importance for these systems to provide customized services to satisfy their users and preventing user churns. Therefore, these
systems need to maintain and analyze user profiles that contain a summary of important personal information. As much as the systems can create more detailed user profiles, they can provide more customized services to the users. Traditionally, the services created profiles based on the user-provided information, that has two main drawbacks. First, many of the users do not create profiles and second, user activities over websites are dynamic over time. Therefore, having automated user profiling systems that consider the temporal dynamics of user activities will result in more realistic and applicable profiles.

The unprecedented traces of user activities over social computing services provides rich information about what is done by which entities, corresponding to their location and time. These information create new opportunities for learning user interests and modeling users behavior over these systems, which can be used for creating automated user profiles. The temporal dynamics of user behavior carry a great deal of information and can help the service providers in many aspects, such as designing efficient marketing strategies, providing customized services, and preventing users’ churn. For example, in an online shopping service, it is important to predict when a user will use an item, or in a community-based question and answering (CQA) site, predicting when a question will be answered is of great importance. Furthermore, in an e-commerce service, analyzing the dynamics of service usage by the users will also help in finding potential churning users and providing them with more incentives to avoid the churn. Hence, it is important having behavior models that not only predict the type of user actions, but also model the time of user activities and their temporal dynamics. The availability of such data over time enable us to learn the key dynamic characteristics of users such as their interests, their loyalty to the provided services, and their satisfaction over time.

There exists a rich literature on modeling user activity that only considers the type of activities but does not pay attention to their timing. Indeed, these models have only concentrated on what is done on the social media, and do not consider when it is done [2]. However, a large and growing interest has been emerged on modeling user behavior over time. Most of these methods do not model the time of user actions and only consider time as a covariate, and hence are unable to predict the time of users future actions [5, 58]. Moreover, most of the existing temporal user modeling methods approach the problem in a discrete time manner. These models suffer from two main drawbacks. First, they assume the process of generating data proceeds by unit time-steps, and hence are unable to capture the heterogeneity of the time to predict the timing of the next event. Second, these models need to choose the length of time-steps to discretize the time, which is a challenging task. Actually, the precise time interval or the exact distance between two actions carries a great deal of information about the dynamics of the underlying systems, and hence the emerging continuous-time models that consider the exact timing of events are more applicable to real-world scenarios [23, 24, 33, 36, 55, 57, 60, 63].

Different factors affect the user participation in a social computing service over time, and considering them in a continuous time user behavior model will result in more realistic models. These factors include the following:

— **Peer influence**: The user engagement may be affected by other users activities in the social media, known as peer influence. For example, in social networks such as Twitter and Foursquare, the user engagement in the service heavily depends on his friends activities.

— **Content**: One of the main factors that affect user participation in online services, is the content being shared over these media. As much as the content is related to the user interests, the user will be more engaged. For example, in knowledge sharing services such as Stack Overflow, the amount of user engagement depends on the content being generated over this media.
Incentives: To engage users in more contributions and steer user behaviors, many social computing services incorporate some techniques known as Gamification. Gamification is defined as using game elements in non-game systems to increase user engagement [20]. One of the most widely used gamification elements are Badges. Badges are some medals that are awarded to users based on some predefined levels of engagement. Recently, many social media sites are using badges to encourage users for more contributions. For example, Stack Overflow, which is a well-known questions answering service, uses badges to encourage users for more contributions. Foursquare, which is a location-based social network awards badges for user check-ins. TripAdvisor awards badges for writing reviews about different places. Recent studies have shown that these gamification elements act as incentive mechanisms and have a significant effect on user participation behavior over social media websites [6].

Continuous-time modeling of user activities over social media has gained considerable attention in recent years. Some preliminary works tried to model the user behavior and predict the future user actions using the time of users’ actions and the impact of activity of friends on each other [57, 63]. More recently, some works tried to jointly model the content and time of actions [23, 24, 36]. The main drawback of the aforementioned methods is that they only consider the impact of friends actions on user activity and ignore other factors such as gamification elements, while considering the ways in which the gamification elements impact user activities and participation in the site is very important in user modeling.

In this article, we propose novel multi-dimensional temporal point processes to jointly model the time and content of user actions by considering the impact of gamification elements, especially badges which is called UMUB (User Modeling Using Badges). We customize the method for modeling user actions over CQA sites. To this end, we propose intertwined point processes to jointly model two main types of user actions over these sites: Asking questions, answering to the questions. Considering all the aforementioned factors in user model, the proposed method is able to infer user interests and predict the time and content of future actions. Modeling when a user will raise a question or answer, can help the site owners in many directions. First, modeling the timing of activities will help them to better predict the site’s traffic, and to provision more suitable load balancing and traffic management strategies. Second, it is of great interest to site owners to predict when a question will be answered in order to inform the asking person about that. Third, modeling when users will ask questions can help site owners to predict the site demand and provide incentives to the users to answer the questions in those times. Another problem is to let the users know when is the best time to ask questions in order to receive prompt answers, in which case we need to model the time that users will answer the questions. The when to post problem is an interesting research problem which is investigated in other domains. Finally, using detailed timing of user questions and answers will help the site owners to better predict the user timely interests, and to provide more intelligent recommendation systems for both asking questions and answering them. In summary, the major contributions of the proposed method are as follows:

— We propose a continuous-time user behavior model that models the user activities in presence of gamification elements, especially badges. The method learns the impact of badges on each user which allows us to differentiate between users.
— We customized the method for modeling the time and content of user activities over CQA sites. We consider two main activities over these sites: asking questions, and answering questions. To do that, we define two interdependent point processes for each user. One for asking questions, and the other for answering the questions.
— We consider the temporal dynamics of user actions in the proposed model. In our model, the time of current user actions depends on the content of previous actions and also the content of current user actions depends on the time of user actions.
— We model the dependency of the processes for asking and answering questions, using interdependent processes. Using this novel framework, beside modeling the dependency between content and time, we also model the dependency between the two different multidimensional processes for asking and answering questions.
— We propose an inference algorithm based on variational expectation maximization (Variational-EM). The proposed inference algorithm efficiently learns the parameters of the model.

We evaluate the proposed method over synthetic and real-world datasets. The real-world dataset is gathered from Stack Overflow, which is a popular question answering website. The results show the efficiency of proposed method in both estimating the time and content of user actions.

The remainder of this article is organized as follows. In Section 2, we briefly review related works in continuous-time modeling of user activities and impact of gamification on user behavior. Details of the proposed method is discussed in Section 3. To demonstrate the effectiveness of the proposed model, extensive experimental results are reported and analyzed in Section 4. Finally, Section 5 concludes this article and discusses the paths for future research.

2 CONTINUOUS-TIME USER MODELING

The abundant available data over social media creates new opportunities for learning models of user behavior. These models can be used to predict future activity, and identify temporal information. There exists a rich literature that use user models over social media ranging from recommender systems [4, 42], to diffusion network analysis [52, 54, 63], and content analysis [2, 3]. Most of the primary works on user modeling have not paid attention to the time of activities. Some have focused on describing aggregate behavior of many people [14, 37, 41], while others have focused on individual behavior models [34, 56]. In this direction, the works have mainly concentrated on the impact of social network [47] and topics of interests on user activities [3, 17, 56]. However, a large and growing literature have been emerged on modeling user behavior over time. Most of these methods do not model the time of user actions and only consider time as a covariate, and hence are unable to predict the time of users future actions [5, 58]. Moreover, most of the existing temporal user modeling methods approach the problem in a discrete time manner, and the classic varying-order Markov models [9, 49] have been used by most of those methods [29, 44, 50]. These models suffer from two main drawbacks. First, they assume the process of generating data proceeds by unit time-steps, and hence are unable to capture the heterogeneity of the time to predict the timing of the next events. Second, these models need to choose the length of time-steps to discretize the time which is a challenging task. Furthermore, in the varying-order Markov models when the number of states is large, due to the exponential growth of the state-space, they cannot capture the long dependency on the history of events.

Actually, the precise time interval or the exact distance between two actions carries significant information about the dynamics of the underlying systems, and hence it is of great importance to have continuous-time models that consider the exact timing of events. Temporal point processes are a general mathematical framework for modeling continuous-time events [40, 51]. Recently, they have attracted considerable attention for modeling user activity over social media. Several models have been proposed in the literature that use temporal point processes to model the information diffusion over networks [57, 63]. The first studies considered the impact of peer influence on user diffusion behavior and used a special point process, called Hawkes [25, 52, 57, 63], to model
the user activities. In this direction, other extensions were considered, that removed the independence assumption between different cascades [55, 60]. Some other works tried to consider the content in the diffusion behavior [23, 24, 33, 36]. More recently, there have been some studies that tried to use temporal point processes in other domains. The authors in [26, 30, 35] incorporated temporal point processes in recommender systems. Linderman and Adams [43] used point processes to infer latent network structure in financial economic interactions and reciprocity in gang violence. Farajtabar et al. [28] tried to use temporal point processes to jointly model the network evolution and diffusion together, and Zarezade et al. [59] tried to model user check-in behavior in location-based social networks using temporal point processes. Du et al. [22] introduced a temporal point process using recurrent neural networks for modeling general marked event data. Their model is capable of predicting the time and mark of future events and has gained a considerable attention recently. Our proposed model is different from the previous literature in different aspects; First, the continuous-time models in the literature mainly consider the impact of social network on the user participation and less attention is paid to the content, and more importantly, no attention is paid to the gamification elements and their impact on user participation. We extend the previous works by considering the impact of gamification elements on user participation. Second, the previous works mainly concentrate on modeling information diffusion and pay little attention to other applications. We applied temporal point processes to modeling user activity over CQA sites, which is substantially different from information diffusion.

Since we consider the impact of gamification elements on user actions in the proposed model, in the remainder of this section, we briefly review the works on user modeling that is related to gamification. The study of gamification and impact of gamification elements on user participation in online media has gained a considerable attention in recent years [21]. Most of the works done in this domain are empirical studies of user participation in social computing systems. Antin and Churchill [8] discuss the various functions and motivations of badges in terms of their psychological incentives. Halavais et al. [31] study the impact of social influence on individual badge earnings on Stack Overflow, and conclude that the influence of friends on badge selection is weak but has some effect. Zhang et al. [62] study an existing badge system in Foursquare and unlike Halavais, they found that users who are friends are more likely to obtain common badges. The authors in [38] analyze how much gamification techniques influence the member response tendencies. The authors in [53] study the impact of reputation on user activities and clustered users based on the reputation trends. Analyses of Stack Exchange reputation schema and its influence on user behavior has been performed by Bosu et al. [13] and Movshovitz-Attias et al. [46]. The authors in [15] study the impact of a hierarchical badge system on user participation and engagement at Stack Overflow. Their initial results present strong empirical evidence that confirms the value of the badges and their effectiveness on stimulating voluntary participation.

Besides the aforementioned works on empirical analysis of user traces of activities, some works have tried to actively study the impact of gamification on real systems. Anderson et al. [7] studied a large-scale deployment of badges as incentives for engagement in a massive open online course system. They found that making badges more salient increases the forum engagement. Hamari [32] actively studied the impact of gamification elements (badges) on an international peer-to-peer trading service. His results show that users in the gamified condition were significantly more likely to use the service in a more active way. While there are many studies about empirical analysis of impact of gamification on user participation, there are a few works on modeling user activities in presence of badges. The main work in this domain is the seminal work of Anderson et al. [6]. They analyzed the impact of badges on user behavior on Stack Overflow. They observed that as users approach the badges boundaries they steer their efforts toward achieving the badges. Using these observations, they proposed a theoretical discrete-time user behavior model and evaluated
it through different experiments. Marder [45] also proposed a discrete-time model of user actions and performed a regression analysis of user activities over Stack Overflow, which conforms with previous empirical observations. In summary, many works have been done on empirical analysis of gamification impact on user behaviors, and little attention is paid to modeling user behavior in presence of gamification elements. Moreover, the existing user activities models approach the problem in a discrete-time manner and do not pay any attention to the content. Our work is the first to offer a continuous-time user model that considers the impact of gamification elements on user activities, besides the content and social influence factors.

3 PROPOSED METHOD

In this section, we introduce the proposed method for continuous-time user modeling in presence of badges. Our model considers all the aforementioned factors in user modeling with more emphasis on badges. The CQA websites, which have gained a considerable attention during recent years, incorporate different elements to increase user engagement. For example, Stack Overflow which is the most popular CQA website on programming questions, uses badges in an effective manner to increase users’ participation. Moreover, user engagement over this site depends on other users participation and the content being shared. These facts make CQA websites good candidates for continuous-time user modeling in presence of badges. Hence, we customize our model for user actions over Stack Overflow.

Asking questions and answering others questions are the two main type of user actions over CQA sites that guarantee the survival of these sites. Therefore, it is of great importance that the proposed method be able to model and predict users’ actions in these two areas. In this section, we propose two interdependent multi-dimensional temporal point processes to model the asking questions and answering behaviors over a CQA site in presence of badges. In order to model the dependency of these two types of activities, we interrelate the processes to each other. In order to model the impact of badges in user participation, we develop a rich set of flexible temporal kernels that accumulate the impact of badges on the intensity function of user activity over time. Moreover, we consider the temporal dynamics in the content of user actions. Finally, we propose an efficient inference algorithm to infer the parameters of the proposed model.

Since our model is based on stochastic temporal point processes, to make the presentation self-sufficient, some theoretical background on these processes is provided in Section 3.1. The proposed generative model is described in Section 3.2, followed by the details of the inference algorithm in Section 3.4.

3.1 Background

A temporal point process is a powerful mathematical tool for modeling random events over time. More formally, a temporal point process is a stochastic process whose realizations consists of a list of time-stamped events \( \{t_1, t_2, \ldots, t_n\} \) with \( t_i \in \mathbb{R}^+ \). Different types of activities over a CQA site, such as asking a question, and answering the questions can be considered as events generated by a point process.

The length of the time interval between successive events is referred to as the inter-event duration. A temporal point process can be completely specified by distribution of its inter-event durations [18]. Let \( \mathcal{H}_t \) denote the history of events up to time \( t \), then by applying the chain rule, we have the following:

\[
 f(t_1, \ldots, t_n) = \prod_{i=1}^{n} f(t_i|t_1, \ldots, t_{i-1}) = \prod_{i=1}^{n} f(t_i|\mathcal{H}_t).
\] (1)
Therefore, to specify a point process, it suffices to define \( f^*(t) = f(t)\mathcal{H}_t \), which is the conditional density function of an event occurring at time \( t \) given the history of events.

A temporal point process can also be defined in terms of counting process \( N(t) \), which denotes the number of events up to time \( t \). The increment of the process, \( dN(t) \), in an infinitesimal window \([t, t + dt)\), is parametrized by the conditional intensity function \( \lambda^*(t) \). The function \( \lambda^*(t) \) is formally defined as the expected rate of events occurring at time \( t \) given the history of events, that is,

\[
\lambda^*(t) dt = \mathbb{E}[dN(t)|H_t].
\]

There is a bijection between the conditional intensity function (intensity for short) and the conditional density function

\[
\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)},
\]

where \( F^*(t) \) is the Cumulative Distribution Function of \( f^*(t) \). Using the definition of \( \lambda^*(t) \) in Equation (3), the likelihood of a list of events \((t_1, \ldots, t_n)\), which is observed during a time window \([0, T] \), can be defined as

\[
\mathcal{L} = \prod_{i=1}^{n} \lambda^*(t_i) \exp \left(- \int_{0}^{T} \lambda^*(s) ds \right),
\]

where \( n \) is the number of observed events and \( T \) is the duration of observation. Intuitively, \( \lambda^*(t) \) is the probability of an event occurring in time interval \([t, t + dt)\) given the history of events up to \( t \), and it is a more intuitive way to characterize a temporal point process [1]. For example, a temporal Poisson process can be characterized as a special case of a temporal point process with a history-independent intensity function which is constant over time, i.e., \( \lambda^*(t) = \lambda \) [40]. Users’ actions usually exhibit complex longitudinal dependencies such as self-excitation, where a user tends to repeat what he has done recently. Such behavioral patterns cannot be characterized by a homogeneous Poisson process, and hence more advanced temporal point processes are needed. Hawkes process is a temporal point process with a particular intensity function, which is able to capture the self-excitation property. The intensity function of a Hawkes process is given by

\[
\lambda^*(t) = \mu + \alpha g_{\omega}(t) \star dN(t) = \mu + \alpha \sum_{t_i < t} g_{\omega}(t - t_i),
\]

where \( \mu \) is a constant base intensity, \( \alpha \) is a weighting parameter that controls the impact of previous events on the current intensity, \( g_{\omega}(t) \) is a kernel that defines the temporal impact of events on the future intensity, and \( \star \) is the convolution operator. In the case that \( g_{\omega}(t) \) is a decreasing function, Hawkes process produces clustered point patterns over time, and hence is able to model the self-excitation property of users events. The right-hand side of Equation (5) comes from the fact that the number of events occurred in a small window \([t, t + dt)\) is \( dN(t) = \sum_{t_i \in H_t} \delta(t - t_i) \), where \( \delta(t) \) is a Dirac delta function.

In many situations, we need to model the events generated by a set of dependent sources. Multi-dimensional point processes are a set of powerful tools for modeling such events. In a multi-dimensional point process, the intensity of a dimension depends on the event history of all dimensions. For example, as it was mentioned before, the users’ actions over social media depends on each other, and hence cannot be modeled independently. In order to model these action, we can use a multi-dimensional temporal point process in which each user corresponds to a dimension and the events of each user impacts the intensity function of other users. For example, the
intensity of a multi-dimensional Hawkes process is given by
\[ \lambda_u(t) = \mu_u + \sum_{v \in U} \alpha_{vu}g_{\omega_v}(t) \star dN_v(t), \]  
where \( \lambda_u \) shows the intensity of user \( u \) to do an action, and \( \mu_u \) is the base intensity of user \( u \), and \( \alpha_{vu} \) shows the influence of user \( v \) on user \( u \). As it is evident in Equation (6), the intensity of user \( u \) depends on history of all users through the second term. Hawkes process models the dependency among different dimensions through the convolution with a temporal kernel which is a linear operator. However, the event in different dimensions may exhibit more complicated dependencies, and hence we need more complex methods for modeling such phenomena. For example, in order to model the effect of badges on user activities, we propose a non-linear multi-dimensional temporal point process in Section 3.2.

Each event can also be associated with some auxiliary information known as the mark of an event. For example, the tags of the questions in a CQA website can be considered as the marks of events. A marked temporal point process is a point process for modeling such events. If \( k \) denotes the mark of the events, then the intensity of the marked temporal point process is given by
\[ \lambda^k(t, k) = \lambda^k(t) f^k(k|t), \]
where \( \lambda^k(t) \) denotes the temporal intensity function, and \( f^k(k|t) \) is the conditional probability density function of observing an event with mark \( k \) at time \( t \). Therefore, in order to determine a temporal point process, we need a temporal intensity that shows the rate of occurring event given the history and a conditional probability density function over marks. We propose a non-linear multi-dimensional marked point process to model user activities in presence of badges, in the next section.

### 3.2 Proposed Generative Model

In this section, we propose multi-dimensional marked point processes to model the user activities over a CQA website in the presence of badges. We consider two main type of activities over such websites, i.e., asking questions and answering them. In order to model such actions using temporal point processes, we consider each of these actions as marked events. In the following, we first detail our notations and assumptions, and then introduce the proposed generative model for user actions over CQA sites.

Let \( D^q(t) = \{e^q_i\}_{i=1}^{N^q(t)} \) and \( D^a(t) = \{e^a_i\}_{i=1}^{N^a(t)} \) denote the set of questioning and answering events observed until time \( t \), respectively. We denote the number of questioning and answering events up to time \( t \) by \( N^q(t) \) and \( N^a(t) \), respectively. The event \( e^q_i \) is a triple \((t_i, u_i, z_i)\), which indicates that at time \( t_i \), user \( u_i \) asks a question with tag \( z_i \). The event \( e^a_i \) is also a triple \((t_i, u_i, p_i)\), which indicates that at time \( t_i \), user \( u_i \) answers a question with id \( p_i \). In the following, we propose two dependent multi-dimensional marked temporal point process to model these two dependent set of marked events. It is worth mentioning that in our model, we try to model the mark of events as their content and do not aim at modeling detailed textual content of user events.

We can consider different intents behind user actions over a CQA site. User activities may be due to drivers external to the website which we call exogenous activities or the influences he receives from the media, such as others actions or website incentives like badges which we call endogenous activities. Paying attention to these drivers will result in more realistic user models. Hence, we assume that the user intensities for asking questions and answering questions are as follows:
\[ \lambda^q_u(t) = \mu^q_u + \rho^q_u \sum_{b \in B^q} g^q_u(h_b(D^q_u(t)), \tau_b), \]
where
\[ \text{Exogenous Intensity} \]
\[ \text{Endogenous Intensity} \]
\[
\lambda^a_u(t) = \mu^a_u + \rho^a_u \sum_{b \in B^a} g^a_w(h_b(D^a_u(t)), \tau_b) + \sum_{e_j \in D^a_u(t)} \eta_{uzi,f^a_w(t_i)}. \tag{9}
\]

We consider the gamification elements (more specifically; badges) as the main drivers of endogenous activities. Badges are awarded to the users based on predefined levels of engagement. Recent analysis has shown that the badges increase user participation over CQA sites. It has also been noticed that the amount of increase depends on how much the user is close to the badge; i.e., the impact of badge on users participation increases as they are close to achieving it [29].

The selected kernels in user intensity should be able to reflect the two aforementioned facts. They should be also able to handle the heterogeneity in the criteria for winning badges. For example, some badges are awarded based on the amount of activities of a given type (we call it a badge \( b_1 \)), while others are awarded based on number of days the user performed a given type of activity (we call it \( b_2 \)). To do that, we define the parameter \( \tau_b \) that captures the criteria for winning the badge \( b \). For our examples, \( \tau_{b_1} \) is the total amount of actions that is required to win a badge, while \( \tau_{b_2} \) is the amount of active days required to win a badge. We also define per badge functions \( h_b \) that extracts the required information from the history of user actions related to a specified badge.

Again, for our example badges, \( h_{b_1}(D_u(t)) = N_u(t) \), which is the count of total actions of user \( u \) until time \( t \), while \( h_{b_2}(t) \) is the total number of active days of user \( u \) until time \( t \). Finally, the temporal kernels \( g^a_w(x, y) \) and \( g^a_u(x, y) \), represent the impact of badges on user intensity. To capture the two aforementioned facts, they should have the following features:

- **Non-negativity**: To capture the positive impact of badges on user participation, the kernels should never become negative. Here, we consider only a positive impact for badges on user participation. The negative impact of miss-designed badges on user participation can be a good direction for the future research.

- **Exponential increase**: To capture the observed feature of badges for which the impact of badge on users participation increases as they are close to achieving it, the kernels should be able to capture these phenomenon.

Different kernels can be utilized that have the required features. Gaussian RBF and Exponential kernels are among the most popular choices. We used the Gaussian RBF kernel \( g_{gauss} \) defined as

\[
g_{gauss}^a(x, y) = e^{-\frac{(x - y)^2}{\alpha^2}}. \tag{10}
\]

And, exponential kernel \( g_{exp} \) as

\[
g_{exp}^a(x, y) = e^{-\omega(y-x)}[y \geq x]. \tag{11}
\]

Figure 1 shows the two selected kernels and their features. The second term in user intensities (Equations (9) and (8)) reflects the cumulative impact of badges on users activities. To capture the heterogeneity of users, motivated by badges, we have also incorporated personal parameters \( \rho^a_u \) and \( \rho^q_u \). As mentioned before, asking questions and answering questions activities are interwoven. Since, in a CQA site the asking question activity is mainly derived by exogenous factors, we did not consider any impact from the previous answering activities. However, we considered a positive impact of questioning activities on answering behaviors. This impact is captured through the third part of the user intensity for answering questions in Equation (9). It captures the impact of other users previous questions on the user answering intensity.

Another important factor that impacts the user activities is the content. The users usually have interest in some fields and also have some expertise in other fields. Therefore, they will ask
questions and answer to questions in some limited number of fields. Hence, we considered the temporal effect of content on the intensity of users for answering questions. We modeled the content of user activities as the mark of events in the proposed temporal point processes. The mark for a question is the tag associated to it, and for an answer is the question it belongs to. The proposed mark probabilities for questions and answers are as follows:

\[
P(z_i = k|u_i) = \frac{\alpha_{u_i k}}{\sum_{s=1}^{K} \alpha_{u_i s}}, \tag{12}
\]

\[
P(p_j = i|t_j, u_j) = \frac{\eta_{u_j z_i} f^a_{\omega}(t_i, t_j)}{\sum_{e_r \in D(y(t_j))} \eta_{u_j z_r} f^a_{\omega}(t_r, t_j)}. \tag{13}
\]

Let \( \alpha_u \) be a vector of user interest over tags, and let \( K \) be the total number of tags, then \( \alpha_u \) is a \( K \) dimensional vector, where \( \alpha_{uk} \) represents the interest of user \( u \) in asking questions of domain \( k \). Also, the parameter \( \eta_u \) is a \( K \) dimensional vector showing the expertise of \( u \) in different domains. Another important factor that has a great impact on the content of user actions is time. We also consider the negative impact of time on user answering to capture the fact that users have less interest to answer older questions. The exponentially decaying function \( f^a_{\omega}(t_i, t_j) \) captures the negative effect of time, and for simplicity, we consider the widely used exponential function \( f^a_{\omega}(t_i, t_j) = e^{-\omega(t_j - t_i)} \).

In summary, the proposed process can capture the following desirable properties:

- **Capturing heterogeneous impact of badges**: The proposed intensity functions for both asking questions and answering questions, captures heterogeneous impact of badges on user participation through the functions \( h_b(.) \), and the decaying kernels \( g^q_{\omega}(.) \) and \( g^a_{\omega}(.) \).

- **Interdependency and mutual excitation**: The events of users have an impact on the others events. For example, The existence of more questions will results in more answers (the third part of user answering intensity in Equation (9)).

- **Impact of time on content and impact of content on time**: The proposed model captures the impact of time on content through the proposed mark probability function in Equation (13). We also consider the impact of content on timing of user answers, utilizing the impact of user interests \( \eta_{uk} \) on intensity function in third part of Equation (9).

- **Temporal decays**: Both, the impact of previous content on the time of events and the processes on each other exponentially decays as a function of time difference, through the decaying functions \( f^a_{\omega}(.) \) in Equations (13) and (9). This means the model pays more attention to recent actions.
**ALGORITHM 1: Variational Expectation Maximization UMUB**

1: for each user \( u \in U \) do
2: initialize \( \rho_a^u, \rho_q^u, \mu_a^u, \mu_q^u, \eta_u \) randomly
3: while \( \Delta \log \mathcal{L} > \delta \) do ▷ check for model convergence
4: **E Step:**
5: for each user \( u \in U \) do
6: for each event \( e_i^q \in D_u^q \) do
7: update \( \phi_i^q \) using Equation (30)
8: for each event \( e_i^a \in D_u^a \) do
9: update \( \phi_i^a \) using Equation (31)
10: update \( \zeta_i^a \) using Equation (32)
11: **M Step:**
12: for each user \( u \in U \) do
13: update \( \mu_q^u \) using Equation (33)
14: update \( \mu_a^u \) using Equation (35)
15: update \( \rho_q^u \) using Equation (34)
16: update \( \rho_a^u \) using Equation (36)
17: update \( \alpha_u \) using Equation (37)
18: update \( \eta_u \) using Equation (38)

3.3 Extension to Other Social Media Sites

Although we introduced our model for user behavior over CQA sites, it can be generalized to other social media sites. Due to a lack of available open access datasets about user actions in presence of badges over other social media sites, here, we only discuss some preliminary extensions of the proposed method to other social media sites without providing experimental results.

The first part of the proposed model that describes the intensity of user actions in relations (8) and (9) can be defined in a general form:

\[
\lambda(t) = \mu + \sum_{s \in B} g(h_s(D(t)), \tau_s),
\]

(14)

where \( s \) is a Badge, \( \tau_s \) is the criteria for winning the badge, and \( h_s \) is a function that extracts the required information from the history of user actions related to badge \( s \).

We can also consider the history of previous actions and the effect of different actions on each other in the intensity function:

\[
\lambda^a(t) = \mu + \sum_{s \in B} g(h_s(D^a(t)), \tau_s) + \sum_{a' \in A} \sum_{e_j \in D^{a'}(t)} \kappa_{a'a}(t, t_j),
\]

(15)

where, \( a' \) and \( a \) are the type of actions in the social media site, \( A \) is the set of all user actions over social media site, and \( \kappa_{a'a}() \) is a temporal kernel that determines the impact actions of type \( a' \) on actions of type \( a \).

We only mentioned a simple extension of proposed method to other social media sites. There are many other extensions that one can consider such as considering the influence of users on each other, and the negative impact of badges. The detailed analysis of these extensions could be considered as an interesting future work.
3.4 Inference

In this section, we discuss the details and practical challenges of estimating the model parameters by using the users activity data, and present a Variational-EM algorithm as a scalable solution.

Given the users’ activities in a time window \([0,T]\), and the data \(D = D^q \cup D^a\), we would like to infer the model parameters \(\Theta = \{\mu^q_u, \mu^a_u, \rho^q_u, \rho^a_u, \eta_u, \alpha_u\}_{u \in U}\). Based on the proposed generative model, the log-likelihood of observed data \(\log p(D|\Theta)\) is given by

\[
\log p(D|\Theta) = \sum_{u \in U} \log p(D_u|\theta_u),
\]

where \(\theta_u\) is the parameter set for user \(u\), i.e., \(\{\mu^q_u, \mu^a_u, \rho^q_u, \rho^a_u, \eta_u, \alpha_u\}\). Using the chain rule, we can further split the likelihood for each user as

\[
\log p(D_u|\theta_u) = \log p(D_u^q|\theta_u) + \log p(D_u^a|D_u^q, \theta_u),
\]

where, the first term is the log-likelihood of user \(u\)’s asking question activities, and the second term is the log-likelihood of answering activities of user \(u\). Using the theory of point processes, i.e., Equations (4) and (7), we have

\[
\log p(D|\Theta) = \sum_{e_j \in D^q} \log(\lambda^q_{u_j}(t_j)) + \sum_{e_j \in D^q} \log(f^q_{u_j}(z_j|t_j)) - \sum_{u} \int_0^T \lambda^q_u(\tau)d\tau
\]

\[
+ \sum_{e_j \in D^a} \log(\lambda^q_{u_j}(t_j)) + \sum_{e_j \in D^a} \log(f^a_{u_j}(p_j|t_j)) - \sum_{u} \int_0^T \lambda^a_u(\tau)d\tau.
\]

We would like to find the parameter set \(\Theta\) that maximizes this log-likelihood function. However, maximizing this log-likelihood function turns out to be a complex problem. The difficulty arises from the summation over different components of the intensity functions that appears inside the logarithm in Equation (18):

\[
\log(\lambda^q_{u_j}(t_j)) = \log\left(\mu^q_{u_j} + \rho^q_{u_j} \sum_{b \in B^q} g^q_w(h_b(D^q_{u_j}(t_j)), \tau_b)\right)
\]

\[
\log(\lambda^a_{u_j}(t_j)) = \log\left(\mu^a_{u_j} + \rho^a_{u_j} \sum_{b \in B^a} g^a_w(h_b(D^a_{u_j}(t_j)), \tau_b) + \sum_{e_j \in D^q(t_j)} \eta_{u_j} f^q_w(t_j, t_i)\right).
\]

Therefore, if we set the derivatives of the log-likelihood to zero, we would not obtain a closed-form solution. In order to resolve this issue, we first give an alternative formulation of the model in which we add an additional layer of latent variables. These auxiliary variables facilitate the inference algorithm without changing the model [30]. As we mentioned in Section 3.2, there are different factors that trigger the users’ actions. Hence, for each activity \(e_j\) of user \(u\), we introduce a latent variable \(s_i\), which denotes the latent factor that trigger the action \(e_i\). For questions in \(D^q\), \(s_i\) is a binary random variable that denotes whether the badges are the driving factor of action \(e_i\) or the driving factor is exogenous

\[
\lambda^q_{t_j}(t_j, s_i) = \begin{cases} \mu^q_{u_j} + \rho^q_{u_j} \sum_{b \in B^q} g^q_w(h_b(D^q_{u_j}(t_j)), \tau_b) & s_i = 1 \\
\mu^q_{u_j} & s_i = 0. \end{cases}
\]

For each answering activity \(e_j\) in \(D^a\), let \(s_j \in \{-1, 0\} \cup \{1, 2, \ldots, |D^a(t_j)|\}\) be the latent factor, where \(s_j = 0\) indicates that the badge incentives drove the user to answer the questions, \(s_j = r > 0\) indicates that the correspondence between the topic of question \(r\) with user \(u\)’s expertise triggered him to answer the question, and \(s_j = -1\) shows that an external factor triggered the answer event.
By using the latent variable $s_j$, the conditional intensity of user $u$ for answering questions can be written as

$$
\lambda_u^q(t_j, s_j) = \begin{cases} 
\mu_u^q & s_j = -1 \\
\rho_u^q \sum_{b \in B^u} g^q_w(h_b(D^q_u(t_j)), \tau_b) & s_j = 0 \\
\eta_{u z_j} f^q_w(t_j - t_s) & s_j \in \{1, 2, \ldots, |D^q(t_j)|\} 
\end{cases}
$$

(21)

It is known that the sum of Poisson processes is itself a Poisson process with rate equal to the sum of all individual rates. Thus, these new latent variables preserve the marginal distribution of the observation. Combining Equations (20) and (21) with Equation (18), the log-likelihood of observations $D$ and auxiliary latent variables $S$ is given by

$$
\log p(D, S|\Theta) = \sum_{e_i \in D^q} \log(\lambda_{u i}^q(t_i, s_i)) + \sum_{e_j \in D^a} \log(f^a_{u j}(p_j|t_j)) - \sum_u \int_0^T \lambda^q_u(\tau) d\tau 
$$

(22)

which can be written as

$$
\sum_{e_j \in D^a} \log(\lambda_{u j}^a(t_j, s_j)) = \sum_{u \in U} C^q_{u 1} \log \mu_u^q + C^q_{u 0} \log \rho_u^q 
$$

(23)

where $C^q_{u 1}$ and $C^q_{u 0}$ also denote the weighted number of times that the user actions are triggered by the badge incentives. $C^q_{u z}$ also denotes the weighted number of times that user $u$’s answers triggered by questions in topic $z$:

$$
C^q_{u 1} = \sum_{e_i \in D^q_u} [s_i = 1] 
$$

(24)

$$
C^q_{u 0} = \sum_{e_i \in D^q_u} [s_i = 0] \times \rho_u^q \sum_{b \in B^q} g^q_w(h_b(D^q_u(t_i)), \tau_b) 
$$

$$
C^{a, -1}_{u} = \sum_{e_j \in D^a_u} [s_j = -1] 
$$

$$
C^{a, 0}_{u} = \sum_{e_j \in D^a_u} [s_j = 0] \times \rho_u^a \sum_{b \in B^a} g^a_w(h_b(D^a_u(t_j)), \tau_b) 
$$

$$
C^{e, z}_{u} = \sum_{e_j \in D^e_u} \sum_{e_t \in D^e_u} [z_i = z][s_j = i] \times \eta_{u z} f^a_w(t_j - t_i). 
$$

As it is mentioned before, the direct optimization of $\log p(D|\Theta)$ is difficult, but the optimization of complete-data log-likelihood $\log p(D, S|\Theta)$ is significantly easier. For any distribution $q(S)$ over the auxiliary latent variables, the following decomposition holds [10]:

$$
\log p(D|\Theta) = {\mathcal L}(q(S), \Theta) + \text{KL} (q(S)||p(S|D, \Theta)) ,
$$

(25)

where the functions $\mathcal{L}(q(S), \Theta)$, and $\text{KL} (q(S)||p(S|D))$ are defined as follows:

$$
\mathcal{L}(q(S), \Theta) = \mathbb{E}_q \left[ \frac{p(D, S|\Theta)}{q(S)} \right].
$$

(26)
\[
\text{KL}(q(S)||p(S|D, \Theta)) = -E_q \left[ \frac{p(S|D, \Theta)}{q(S)} \right].
\]

\text{KL}(q(S)||p(S|D, \Theta)) is the Kullback–Leibler divergence between \(q(S)\) and the posterior distribution \(p(S|D, \Theta)\), and satisfies \(\text{KL} \geq 0\) with equality, if and only if, \(q(S) = p(S|D, \Theta)\). Therefore, \(\mathcal{L}(q, \theta)\) is a lower bound on the log-likelihood function. In order to find the value of \(\Theta\) that maximizes the log-likelihood function, we use the Variational-EM algorithm. Variational-EM is an iterative optimization algorithm that has two main steps in each iteration. In the E-step, the lower bound \(\mathcal{L}(q(S), \Theta)\) is maximized with respect to q(S) while holding \(\Theta\) fixed. In the subsequent M-step, the distribution q(S) is held fixed, and the lower bound \(\mathcal{L}(q(S), \Theta)\) is maximized with respect to \(\Theta\). In the E-step, we should find the posterior of the latent variables S in order to maximize the \(\mathcal{L}(q(S), \Theta)\). However, since finding the joint posterior of all latent variables is not computationally tractable, we use the mean-filed approximation assumption over q:

\[
q(S) = \prod_{e_i \in D^q} q(s_i|\phi_i^q) \prod_{e_j \in D^a} q(s_j|\phi_j^a).
\]

In other words, we find the best factorized \(q(S)\) distribution that is most similar to the posterior \(p(S|D, \Theta)\) in the KL-divergence sense. In the M-Step, we optimize the \(\mathcal{L}(q(S), \Theta)\) with respect to \(\theta\).

Although introducing the auxiliary latent variables simplify the inference algorithm, we have another challenge in inferring the answer parameters. While computing the \(\log f_w^a(p_i|t_i)\), the \(\log \{\sum_{e_x \in D^y(t_i)} \eta u_x f_w^a(t_i - t_x)\}\) that exists in the denominator of \(f_w^a(p_i|t_i)\) makes the inference of user expertise challenging. Using the approach used in [11, 12], we define a new variational parameters \(\zeta_j\) for each answer event

\[
\log \left\{ \sum_{e_x \in D^y(t_i)} \eta u_x f_w^a(t_j - t_x) \right\} \leq \zeta_j \left( \sum_{e_x \in D^y(t_i)} \eta u_x f_w^a(t_j - t_x) \right) - 1 - \log(\zeta_j).
\]

In summary, we should tighten the lower bound \(\mathcal{L}(q(S), \Theta)\) with respect to q(S) and \(\zeta\) in the E-step, and maximize it with respect to \(\Theta\) in the M-Step.

3.4.1 E-Step. Substituting Equations (29), (24), and (23) in Equation (24), and considering a multinomial distribution for each \(q(s_i)\) with parameter \(\phi_i\), given the current state of all parameters, we find \(q(s_i)\) that maximizes the lower bound in the E-step, which is equivalent to updating the \(\phi_i\) parameters. Using some straightforward calculations, the update equations for the question latent variables is given by

\[
\phi_{ik}^q = q(s_i = k) = \begin{cases} 
\frac{\rho_i^q \sum_{b c} g_w^q \eta u_x f_w^a(t_j - t_i)}{\rho_i^q + \sum_{c b} g_w^q \eta u_x f_w^a(t_j - t_i)} & k = 0 \\
\frac{\rho_i^q \sum_{b c} g_w^q \eta u_x f_w^a(t_j - t_i)}{\rho_i^q + \sum_{c b} g_w^q \eta u_x f_w^a(t_j - t_i)} & k = 1,
\end{cases}
\]

and the update equations for the answer latent variables is given by

\[
\phi_{jk}^a = q(s_j = k) = \begin{cases} 
\frac{\rho_i^a \sum_{b c} g_w^a \eta u_x f_w^a(t_j - t_i)}{\rho_i^a + \sum_{c b} g_w^a \eta u_x f_w^a(t_j - t_i)} & k = -1 \\
\frac{\rho_i^a \sum_{b c} g_w^a \eta u_x f_w^a(t_j - t_i)}{\rho_i^a + \sum_{c b} g_w^a \eta u_x f_w^a(t_j - t_i)} & k = 0 \\
\frac{\rho_i^a \sum_{b c} g_w^a \eta u_x f_w^a(t_j - t_i)}{\rho_i^a + \sum_{c b} g_w^a \eta u_x f_w^a(t_j - t_i)} & k \in \{1, 2, \ldots, |D^q(t_j)|\}.
\end{cases}
\]
Finally, the update equations for variational parameters $\zeta_j$ is given by

$$
\zeta_j = \frac{1}{\sum_{e_i \in D^a_q(t_j)} \eta_{uz} f_w^a(t_j - t_r)}.
$$

(32)

3.4.2 M-Step. In the subsequent M-step, the distribution $q(S)$ is held fixed and the lower bound $\mathcal{L}(q(S), \Theta)$ is maximized with respect to $\Theta$, where $\Theta = \{\mu^q_u, \mu^a_u, \rho^q_u, \rho^a_u, \eta_u, \alpha_u\}_{u \in U}$. To find the point estimations for different parameters, we should maximize the lower bound with respect to different parameters. Maximizing the lower bound with respect to $\mu^a_u$ and $\rho^a_u$ will result in the following closed-form solutions:

$$
\mu^a_u = \frac{\sum_{e_i \in D^a_q} \phi^a_{j,0}}{T},
$$

(33)

$$
\rho^a_u = \frac{\sum_{e_i \in D^a_q} \phi^a_{j,1}}{\sum_{b \in B^a} G^a(u, b, T)},
$$

(34)

where $G^a(u, b, T) = \int_0^T g^a_w(h_b(D^a_u(s)), \tau_b) ds$, and maximizing the lower bound with respect to $\mu^q_u$ and $\rho^q_u$ will result in

$$
\mu^q_u = \frac{\sum_{e_i \in D^q_u} \phi^q_{j,1}}{T},
$$

(35)

$$
\rho^q_u = \frac{\sum_{e_i \in D^q_u} \phi^q_{j,0}}{\sum_{b \in B^a} G^q(u, b, T)}.
$$

(36)

Maximizing the lower bound with respect to $\alpha_u$ will result in the following simple closed-form solution:

$$
\hat{\alpha}_{uk} = \frac{\sum_{e_i \in D^q_u} \mathbb{I}[z_i = k]}{|D^q_u|}.
$$

(37)

Finally, to find $\eta_u$, we should maximize the lower bound with respect to $\eta_u$. Unfortunately, there is no closed-form solution for $\eta_u$, and we should solve the following optimization to estimate the vector $\eta_u$:

$$
\hat{\eta}_u = \arg \max_{\eta_u} \log \eta_u^T F_u - \eta_u^T H_u
$$

s.t. $\sum_{k=1}^K \eta_u = 1,

(38)

where $F_u$ and $H_u$ are $K$ dimensional vectors defined over different tags, and for each tag $k$, we have

$$
F_u(k) = \sum_{e_j \in D^q_u} \mathbb{I}[z_{p_j} = k] + \sum_{e_i \in D^q(u, z_i = k)} \phi^q_{ji}.
$$

(39)

$$
H_u(k) = \sum_{e_i \in D^q, z_i = k} \frac{1}{\beta} \left[ 1 - e^{-\beta(T - t_i)} \right] + \sum_{e_j \in D^a_u} \zeta_j \sum_{e_r \in D^a_q(t_j), z_r = k} f_w^a(t_j - t_r).
$$

(40)

Since, Equation (38) is convex in $\eta_u$, we can find the optimal solution by using different convex optimization tools such as CVX. The overall steps of inference algorithm is depicted in Algorithm 1.
### Table 1. Selected Badges Information

| Name | Description |
|------|-------------|
| Curious | Asked a well-received question on 5 separate days |
| Inquisitive | Asked a well-received question on 30 separate days |
| Socratic | Asked a well-received question on 100 separate days |
| Explainer | Edited and answered 1 question (answer score > 0) |
| Refiner | Edited and answered 50 question (answer score > 0) |
| Refiner | Edited and answered 500 question (answer score > 0) |

### 4 EXPERIMENTS

#### 4.1 Dataset Description

In this section, we evaluate the proposed inference algorithm, and the effectiveness of the proposed method by performing several experiments on synthetic and real datasets. To validate the proposed inference algorithm, we generate a set of events by using the proposed method, and validate the estimated parameters by using different criteria. Moreover, to evaluate the performance of our model, we use a real dataset that has been collected by crawling the Stack Overflow. The datasets are explained in more details in the following paragraphs.

— **Synthetic data**: We considered $U = 100$ users and $K = 50$ different tags as the indicators of different topics. We also considered $|\mathcal{B}^q| = 5$ badges as threshold badges for asking questions, and $|\mathcal{B}^a| = 5$ badges as threshold badges for answering questions. The thresholds for badges are sampled from a uniform distribution $r_b \sim U(0, 500)$. Then, we generated random vectors of user interests ($\alpha_u$) and user expertise ($\eta_u$) over different tags by using a Dirichlet distribution, i.e., $\alpha_u, \eta_u \sim \text{DIR}(50, 0.1)$. The exogenous intensities ($\mu^q_u, \mu^a_u$), and badge impact parameters ($\rho^q_u, \rho^a_u$) are drawn from uniform distributions, $\mu^q_u \sim U(0, 0.01)$, $\mu^a_u \sim U(0, 0.05)$, and $(\rho^q_u, \rho^a_u) \sim U(0, 1)$. Then, by using the Ogata’s thinning method [48], we generated the events of our model. It is worth mentioning that we used both exponential ($g^\text{exp}(x, y)$) and Gaussian ($g^\text{gauss}(x, y)$) kernels for badge impacts. Using each kernel type, we generated up to 1,000 train events and 400 test events per user. To eliminate the randomness in the results, we repeated the procedure for 10 different times and aggregated the results.

— **Real data**: For the real dataset, we used the user questioning and answering data of Stack Overflow, which is the most popular CQA website. To this end, we crawled the user data by using an API.\(^2\) We selected the 2,000 top rank users of this website based on their reputations, and collected all their questions and answers. We selected three threshold badges for question activities, namely; *Curious, Inquisitive, and Socratic*. These badges are threshold badges that are awarded to the users based on the total active days for questioning. We also selected three threshold badges for answering activities; *Explainer, Refiner, and Illuminator*. These badges are threshold badges that are awarded to the users based on total amount of their answers. The criteria for getting different badges are depicted in Table 1. These two types of badges have different criteria. The question badges are awarded based on active days, while the answer ones are awarded based on total amount of participation, which helped us to evaluate the generality of the proposed kernels for capturing the effect of badges. We considered the first tag of each question as its mark. The word cloud of

---

1. A question is well-received if it is open and has a score greater than 0. [http://meta.stackexchange.com/questions/234259/asking-days-badges](http://meta.stackexchange.com/questions/234259/asking-days-badges)
2. [http://data.stackexchange.com/](http://data.stackexchange.com/).
Continuous-Time User Modeling in Presence of Badges: A Probabilistic Approach

Fig. 2. Visualization of tags of questions in our Stack Overflow dataset. (a) Word cloud of tags, (b) Histogram of total number of tags used by users. A big portion of users only used fewer than 20 tags.

questions tags, and the histogram of total number of different tags used by different users is plotted in Figure 2. As it can be seen, a big portion of users used fewer than 20 different tags that is an indicator of limited interest of users in different domains. It is worth mentioning that we only selected the events that participate in winning the selected badges. For the selected users and for each type of activities, we considered 80% of their actions as train data, and the remaining 20% as the test data. For each type of activity, the train and test selection was based on the time of actions, i.e., we sorted the activities based on their times and selected the first 80% as the train data and the remaining as the test data.

4.2 Synthetic Results

We evaluated the performance of proposed inference algorithm using synthetic data. In this case, we are seeking to answer the following questions: (1) Does our simulator truly generate the events from the proposed model?; (2) How does the proposed inference algorithm converges?; (3) Can the proposed inference algorithm accurately infer the model parameters?; and (4) What is the predictive performance of the proposed algorithm? We also compared the performance of inference algorithm for both Exponential and Gaussian kernels to capture the badge impacts.

4.2.1 Data Generation. We used the Ogata’s thinning method [48] to simulate the events from our model. In this method, we simulate the time of events by utilizing a homogeneous Poisson process with an intensity greater than the intensity of target point process, then reject some events. To check the conformance of generated event times with the target process, we used the time-change theorem of temporal point processes. According to the time-change theorem [19], given all $t_i$ and $t_{i+1}$ subsequent event times of a particular point process, if the samples are truly sampled from the process $\lambda(t)$, then the set of intensity integrals $\int_{t_i}^{t_{i+1}} \lambda(t) dt$ should conform to the unit-rate exponential distribution. Hence, we compared the theoretical quantiles from the exponential distribution with the ones computed from the intensity integrals of generated events of our model in Figure 3. The closer the slope is to one, the better the model matches the generated events. The results clearly show that the points approximately lie on the same line, providing empirical evidence that generated events conform with the real intensities, in our model.

4.2.2 Convergence. To see how the proposed variational-EM inference algorithm converges, we plotted average log-likelihood per event after the E-step vs. the iteration number for both

---

3 All codes and implementations are available at: https://github.com/alikhodadadi/UMUB.
Fig. 3. Quantile–quantile plots of generated data of our proposed model. (a) Questions. (b) Answers.

Fig. 4. The performance of proposed inference algorithm over different kernel types. The results are the average over question and answer data over 10 different runs. (a) Per event log-likelihood over different iterations of EM algorithm. (b) Average run-time over different sample sizes. (c) Per event log-likelihood of test data over different fractions of train data.

Exponential and Gaussian kernels for badge impacts in Figure 4(a). It is noticed that only after a few steps the log-likelihood converges for both kernels which is an indicator of fast convergence of the proposed inference algorithm. The results also show that the inference algorithm performs slightly better for the Exponential kernel against the Gaussian kernel. We also plotted the run-time of inference algorithm over different sample sizes in Figure 4(b).

4.2.3 Parameter Estimation. To evaluate whether the proposed inference algorithm is able to accurately estimate the model parameters, we plotted the average MSE and Rank for the temporal and content parameters. The MSE metric, measures the mean square error between the true and estimated parameters, i.e., $\frac{1}{n} \sum_{i=1}^{n} |\theta_i - \hat{\theta}_i|^2$. We also evaluated how well the order of parameter values is preserved. To this end, we used the Kendall rank coefficient [39]. We averaged the metrics over all temporal parameters $\Theta_T = \{\mu_u^q, \mu_u^a, \rho_u^q, \rho_u^a \}_{u=1}^{|U|}$ and content parameters $\{\alpha_u, \eta_u \}_{u=1}^{|U|}$. To obtain more intuition about the inferred parameters, we generated the scatter plot of real and inferred parameters for Gaussian kernel. The more the points are close to the $y = x$ line, the more the inferred and true values are similar. The average results over temporal and content parameters for different kernels are depicted in Figure 5. The first row (Figure 5(a)–(c)) shows the average recovering error of temporal parameters for different train sizes, and the second row (Figure 5(d)–(f)) shows the average errors for content parameters. The results show that the performance of the proposed method in all metrics for both temporal and content parameters improves as the amount of train data increases. The MSE decreases as the number of train data increases, and the Rank Correlation increases as the number of train data increases. It was also noticed that the inference algorithm performs slightly better for the exponential kernel against the Gaussian kernel.
Continuous-Time User Modeling in Presence of Badges: A Probabilistic Approach

4.2.4 Predictive Performance. To evaluate the predictive performance of proposed method using the inferred parameters, we plotted the average log-likelihood per event for the test data vs. different train data sizes. Figure 4(c) shows the results. As illustrated, when the size of train data increases, the log-likelihood of test events increases for both kernels. It is an indicator that with the increase in the train data the model can better learn the temporal dynamics of data, and hence better predict the future. It is again noticed that the model performs better for Exponential kernel against the Gaussian kernel, which conforms with the previous results. Since the inference algorithm can better learn the parameters for exponential kernel, it can better predict the future.

4.3 Real-Data Results

We also evaluated the performance of proposed method on real data gathered from Stack Overflow. We used the following criteria to evaluate the predictive performance of methods for time and mark predictions.

— Average test log-likelihood: We calculated the average temporal log-likelihood for test events of each user, and reported the average results over all users. The higher log-likelihood means better performance.
— **Time prediction**: For each user we predicted when the test events will occur using the density of next event times \( f(t) = \lambda(t) \exp - \int_0^t \lambda(s)ds \). To achieve this, we computed the expected time of next event by sampling the future events. We reported the Mean Absolute Error (MAE) between the predicted time and the true time.

— **Mark prediction**: We also evaluated the predictive performance of the proposed method for predicting the marks of events. As we mentioned before, the mark of a question is its tag, and the mark of an answer is the question it belongs to. For each test event, given the time of event, we evaluated the probability of its mark (tag/parent). We rank all the tags/parents in the descending order of their probability, and create a recommendation list. We reported average Precision@k and NDCG@k [16] for different test events over all users. We also reported the Error in predicting the true mark for different users.

We used two types of baselines for temporal and content evaluations. For temporal evaluations, to see whether considering the effect of badges would help to obtain better predictions, we used the following baselines.

— **RMTPP [22]**: We compared the proposed model with RMTPP which is one of the most recent and successful works on modeling marked events using temporal point processes. The RMTPP combines point processes and neural networks. It utilizes neural networks to model the intensity function of point processes. The RMTPP independently models both mark and time information and does not differentiate between different users.

— **Multi-dimensional Hawkes process**: Hawkes process is the most widely used continuous-time model in recent user behavior studies. These works mainly concentrate on the self and mutual excitations among the events. Since there is no friendship network in Stack Overflow, we did not use the mutually exciting models. To explore whether there is a self-excitation behavior among the events, we fitted a self-exciting Hawkes process to the events of each user with intensity \( \lambda_u^{a/q}(t) = \mu_u^{a/q} + \rho_u^{a/q} \sum_{\tau \in D_u^{a/q}} \kappa(t, \tau) \), where \( \kappa(t, \tau) = \exp(-(t - \tau)) \).

— **Homogeneous Poisson process**: We fitted a Poisson process with a constant intensity function for each user and type (question/answer), \( \lambda_u^{a/q}(t) = \lambda_u^{a/q} \).

For mark evaluations, we used the following baselines.

— **MC0**: In majority prediction, which is known as zero-order Markov Chain (MC0), at each time step regardless of the time, we predict the most popular marker for each user. Usually, predicting the most popular marker is a strong heuristic [26]. Using this baseline, we can evaluate the impact of utilizing customization and time on the mark prediction results.

— **MC1**: We also compared the results with the Markov models. In the first-order Markov Chain, at each time step, we predict the most recent used marker for each user. Using this baseline, we can evaluate impact of utilizing the customization and preferences in our model on the mark prediction results.

— **MF**: For answer parent prediction, we also used a heuristic baseline. We selected a window of 50 recent questions and selected the question to answer based on the total number of answers each question has received. We call this baseline Most Favorite (MF). By using this baseline, we can evaluate the impact of paying attention to the user’s interests and the time on the mark prediction results.

— **RMTPP**: We also compared the results with the RMTPP model in predicting the mark of events.
Continuous-Time User Modeling in Presence of Badges: A Probabilistic Approach

4.3.1 Temporal Results. In Figure 6, we have plotted the performance of different methods in predicting the time of next events for both question and answer events. As illustrated, for both log-likelihood metric (Figure 6(d) and (a)), and MAE metric (Figure 6(e) and (b)), the proposed method outperforms the competitors. It was also noticed that the RMTTP and Hawkes models perform comparably. It may be due to the fact that we define Hawkes processes which per user learn the parameters, while the RMTTP does not differentiate between users. RMTTP can learn complex temporal dynamics, while Hawkes cannot do that. Since the UMUB model considers both the impact of badges and per user parameters, it can better model the temporal dynamics of events in presence of badges. It is also noticed that the overall performance of all methods degrades from answers to questions. It may be due to the fact that questions mainly are derived by external factors, and hence modeling them are a harder task than the answering events.

According to the time-change theorem [19], given all $t_i$ and $t_{i+1}$ subsequent event times of a particular point process, the set of intensity integrals $\int_{t_i}^{t_{i+1}} \lambda(t) dt$ should conform to the unit-rate exponential distribution, if the samples are truly sampled from the process $\lambda(t)$. Hence, we compare the theoretical quantiles from the exponential distribution with the ones from different models to the real sequence of events. The closer the slope is to one, the better the model matches the event patterns. Figure 6(c) and (f) shows the results for different models for answers and questions.

It is worth mentioning that for each question and answer, we only model its mark and do not model its latent content which is the subject of topic modeling studies. Therefore, for mark prediction, we only compared our model with mark prediction methods. Utilizing the complex latent content of activities is an interesting venue for future research. Moreover, to improve the efficiency in our experiments, we only considered the last 200 preceding questions for each answer in the MC0 method and the Equations (9) and (13) of our proposed model, and the time unit in experiments is 1 day. The last but not the least, we call our model UMUB in the experiments, which is an abbreviation for User Modeling Using Badges.
respectively. The results show that our UMUB model can better explain the observed data compared to the other models. It worth mentioning that since the available implementation of RMTPP model do not provide any intensity for test events and only produces the MAEs, we do not reported it in the quantile–quantile results. In summary, the temporal results confirm the superior performance of proposed method in predicting the time of next actions against the baseline competitors.

4.3.2 Mark Prediction Results. We plotted the results of different methods in predicting the mark of test events in Figure 7. Figure 7(a) and (c) shows the precision@k and NDCG@k for different methods in predicting the parent of answers (i.e., the question it belongs to) vs. different k. In the same way, Figure 7(b) and (d) shows the precision@k and NDCG@k for different methods in predicting the tag of questions vs. k. As it can be seen, by increasing k, the performance of all methods improves. The proposed method outperforms the competitors in predicting both the parents and tags. For predicting the tag of questions, the MC1 method performs close to the proposed method (Figure 7(b) and (d)). This is because the asking questions is a sequential task, and usually a user asks questions in a domain, sequentially. In addition, by the fact that users usually ask questions in a limited number of topics (Figure 2(b)), the MC0 is also a strong predictor for the tags of questions. The high difference between the proposed method and the competitors for predicting the parent of answers (Figure 7(a) and (c)) is because, although the popularity (MC0) and recency (MC1) heuristics are strong predictors, but for predicting the questions that the user will answer they cannot work well. This is due to the huge amount of questions that arrive in every moment
and the user must select one of them. Therefore, paying attention only to time or popularity will result in poor performance. The MF method for answer parent prediction, which only considers the recent questions based on the amount of attention they have received, performs poorly on both metrics. It is because of the fact that only the attention received by a question is not a good indicator of relatedness, and the user’s interests and content of question play an important role in predictions. The proposed method considers both the personal preferences and the temporal impacts in predictions, and hence it can better predict the question that user will answer. We should note that, since the RMTPP method do not provide any recommendation list and only reports the prediction error for marks, we did not report it in the results presented in Figure 7.

To compare the results with RMTPP method, we also reported the mark prediction error of different methods in Figure 8. Since each user answer a question once and these events are not recurrent, the RMTPP method cannot predict which question the user will answer. Instead, we defined another prediction problem in which the mark of an answer is the tag of its parent question. Indeed, we predict the tag in which the user will answer the question which is an interesting research question. Figure 8(a) presents the mark prediction error for different methods for this prediction task. As it can be seen for answer mark prediction the UMUB outperforms other methods, then the RMTPP method performs better than the other two baselines. We also reported the mark prediction error for question marks in Figure 8(b). For question mark prediction error, the MC1 performs the best. It is because of the fact that users ask questions sequentially in different tags that results the MC1 outperforms other methods. This is shown previously in Figure 7(b) that MC1 performs the best for small $k$’s. After that, the UMUB method performs better than the RMTPP and MC0 baselines.

In summary, the UMUB method, which consider both the user interests and temporal dynamics in tag predictions performs reasonably over different $k$’s and over both questions and answers, while the MC0 and MC1 methods do not work very well for answer mark prediction.

4.3.3 Impact of Kernels and Different Components of Model. We also studied the impact of different configurations on the performance of the proposed method. First, we studied the impact of different kernels on the performance. Figure 9(a) shows the impact of Gaussian and Exponential kernels, which previously we depicted in Figure 1, on the log-likelihood of test events. The results shows that the Gaussian kernel works better than exponential for both type of actions. This means that for both questioning and answering actions, the impact of badges on user actions do not drop suddenly after obtaining the badge. In addition, the badges have an impact on user activities even after obtaining them.
We also studied the impact of different components of proposed method on its performance. We compared different variants of proposed method for modeling the answering process. The simplest version is the one that do not consider both the impact of badges and history. We call it \textit{NB-NH}, which stands for No Badge No History. Another variant is the one that only considers the impact of history of previous questions on answering process, but do not consider the impact of badges, and we call it \textit{NB-H}. In the same way, we define a variant of the proposed method that only considers the impact of badges and does not consider the impact of history. We call this method \textit{B-NH}, and finally we define the full variant which considers both the impact of badges and history of previous questions (Intertwined). We call this variant \textit{B-H}. Figure 9(b) shows the performance of these variants based on the log-likelihood of test events. As it can be seen, the \textit{NB-H} version that only considers the impact previous questions, performs slightly better than the simple version. Considering the impact of badges (\textit{B-NH}) has a more significant impact on the performance, and finally considering both the impact of badges and the history (\textit{B-H}), improve the results. We also compared the impact of different components on the proposed model for asking questions. The simple variant which do not consider the impact of badges is the Poisson process. The results show that utilizing the badges significantly improves the performance. To summarize, the results show that considering both the impact of badges and history of questions (Intertwining) have an impact on the results, and compared to the history, only considering the badges has a more significant impact on the results.

4.3.4 Qualitative Analysis of Parameters. To make some intuitions about the learned parameters, we also analyzed the learned parameters for different users of Stack Overflow. In Figure 10(a), we plotted the learned base intensities of different users for asking questions and answering the questions. We can see two interesting patterns in the results. First, the intensity for answering the questions is more than the one for asking questions in general. It may be due to the fact that answering questions needs less effort than asking good questions. Second, most of the users are not active in both tasks, i.e., they do not ask many questions and answer many of them simultaneously. But, for those who are so active, we see that they are single tasks, i.e., they ask many questions or they answer many questions. It may be due to the limited time and attention of users that makes them focus on single tasks. We also plotted the learned badge interests for different users in Figure 10(b). It is interesting that we can see the patterns in base intensities again in badge interests. In general, the badge impacts for answering badges are greater than those for questions.
Fig. 10. The learned parameters for users of Stack Overflow. (a) Learned base intensities for asking questions and answering them for different users. For each user, we plotted a single point \((\mu^q_u, \mu^a_u)\). (b) Learned badge interests for different users. For each user, we plotted a single point \((\rho^q_u, \rho^a_u)\). (c) The learned badge interests for different membership ages. (d) The histogram of correlation coefficient of interests \(\alpha_u\) and expertises \(\eta_u\) for the users.

The fewer impact for question badges may be due the fact that asking questions mainly is derived by external factors, i.e., user needs while the answering is derived by the site incentives. It is also interesting that those who are mostly impacted by the badges are single task users, i.e., those who have a large-badge impact for asking questions, are less impacted by the answering badges and vice versa.

We also studied that whether the badge impacts differ between different membership ages. To this end, we segmented the users based on the years from their membership in the site and presented the average badge impact for different bins in Figure 10(c). As it can be seen from the results, there are two different patterns for answer and question badges. We see an increase and decrease for the impact of answer badges in user lifecycle. We see that with the increase in the age the impact of answer badges on users increases till fourth year of membership, but after that we see a decrease in the impact. Many parameters may affect this pattern such as the badge placement strategy of website. But, we think one factor may be the boredom in user activities. They keep getting badges till some point of their life but after that they do not pursue to get the badges and are less affected. We see a different pattern for the impact of question badges. As the age increases, the impact of badges increases. There are many factors behind this phenomenon. It may be due to
the fact that asking good questions is a hard task, and so the question badges have a less impact on users and only the aged people approach the badges and get impacted by them. Another reason may be that, users pay more attention to answer badges in the earlier staged of their lifecycle and as they get aged, they tend to question badges. The reasons behind the different patterns of badges impacts is an interesting question, which is behind this study and is a venue for future research works. The last but not the least, we also analyzed the content parameters of users. We considered two different distribution vectors over tags for each user to separate their interests for asking questions and expertise for answering the questions. To test whether this separation is true, we calculated the correlation between the learned interest and expertise vectors for each user and plotted the histogram of this correlations. The results are depicted in Figure 10(d). As the results show, a big portion of users have a correlation lower than zero, which means that their interests and expertises differs significantly.

4.4 Badge Placement

An interesting aspect of a badge system is where to place the threshold badges. It is important for the site owners to place the badges such that it results in the most users’ engagement. The problem of shaping user activities has gained a considerable attention during recent years [27, 61]. Studying how we can use the badges to shape user activities is an interesting research problem. Here, we briefly review how different badge placements impacts the total generated events by our model and leave the detailed analysis of badge placement to future research works.

We considered five users and three question badges and five answer badges. We designed three different badge placement scenarios: Early assignment, in which we reward the badges early till 50 events, i.e., all badge thresholds are lower than 50. Late assignment, in which we reward the badges for around 200 events, and Uniform assignment, in which we uniformly distribute the badges from 0 to 200 events. To be more precise, we rewarded the five answer badges for 30, 60, 90, 120, and 150 answers and also rewarded the three question badges for 30, 50, and 100 questions in the uniform assignment, respectively. To test how different placements, impact the total engagement, we generated 1,000 events by each configuration and studied the time it takes that these events be generated by each configuration. The lower it takes that the events are generated, the configuration will result in more engagement. Figure 11 shows the generated questions and answers for different configurations. It is interesting that the most engagement is provided by the uniform badge placement, which conforms with previous studies over badge placement [6]. For uniform assignment, only 1,000 time steps needed that the events are generated, while for early and late assignments it took about 7,800 and 6,500 time steps to reach the goal. It is also interesting that the early assignment performs better than the late assignment at the beginning but as the
time pasts the late assignment performs better. We should note that the badge placement is a complex problem and the total amount of events is only one aspect of it. Considering other aspects will help to better understand the problem. For example, the early badge assignment will result in the more answers generated by newbie users that may result in poor quality answers, or the late assignment may result in user boredom since most of them cannot win the badges and increase the churn rate. Considering these different aspects in the proposed model is an interesting venue for our future research works, which will produce a tool that help the site owners to explore different badge placement strategies and design the badges in a way that results in more user engagements.

5 CONCLUSION

In this article, we proposed new continuous-time user models by using a powerful mathematical framework; Temporal Point Processes. Unlike previous works on continuous-time user modeling that mainly focus on the impact of peer influence on user actions, the proposed method also considers the impact of content and gamification elements, especially badges on user actions. We extended the proposed method for modeling user questioning and answering activities over CQA sites and proposed an inference algorithm based on Variational-EM that can efficiently learn the model parameters. The learnt parameters can help in categorizing users and understanding their preferences. These information will help the social media owners in creating implicit user profiles and delivering customized services. The proposed model will also help social media owners to study the different badge placements and their impacts on user engagement. The empirical evaluations on both synthetic and real datasets demonstrate the superior predictive performance of the proposed method.

There are many interesting lines for future works. In this work, we only considered the single activity threshold badges, incorporating other hybrid badges, and also other gamification elements like reputation systems is a new line of research. On the content side, we used simple models, considering more complex models such as Bayesian non-parametric ones will help to improve the model in content side. Activity shaping over social media has gained a considerable attention in recent years. One means to shape user activities is the badges. Studying the continuous-time activity shaping problem [27, 61] by means of badges is another venue for future works. Modeling the quality of user actions and incorporating the impact of badges on the quality in the proposed model is another interesting line of future investigation.

REFERENCES

[1] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. 2008. Survival and Event History Analysis: A Process Point of View. Springer Science & Business Media.
[2] Ahmad Abdel-Hafez and Yue Xu. 2013. A survey of user modelling in social media websites. Computer and Information Science 6, 4 (2013), 59.
[3] Amr Ahmed. 2012. Modeling content and users: Structured probabilistic representation and scalable inference algorithms. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 1.
[4] Amr Ahmed, Yucheng Low, Mohamed Aly, Vanja Josifovski, and Alexander J. Smola. 2011. Scalable distributed inference of dynamic user interests for behavioral targeting. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 114–122.
[5] Amr Ahmed and Eric P. Xing. 2013. Scalable dynamic nonparametric Bayesian models of content and users. In Proceedings of the 23rd International Joint Conference on Artificial Intelligence.
[6] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2013. Steering user behavior with badges. In Proceedings of the 22nd International Conference on World Wide Web (WWW’13). ACM, New York, NY, 95–106. DOI: http://dx.doi.org/10.1145/2488388.2488398
[7] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. Engaging with massive online courses. In Proceedings of the 23rd International Conference on the World Wide Web.
[8] Judd Antin and Elizabeth F. Churchill. 2011. Badges in social media: A social psychological perspective. In Proceedings of the CHI Workshop on Gamification, Vancouver, BC, Canada.

[9] Ron Begleiter, Ran El-Yaniv, and Golan Yona. 2004. On prediction using variable order Markov models. Journal of Artificial Intelligence Research 22 (2004), 385–421.

[10] Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Secaucus, NJ.

[11] David M. Blei and John D. Lafferty. 2006. Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning. ACM, 113–120.

[12] David M. Blei and John D. Lafferty. 2007. A correlated topic model of science. The Annals of Applied Statistics 1, 1 (2007), 17–35.

[13] Amiangshu Bosu, Christopher S. Corley, Dustin Heaton, Debarsi Chatterji, Jeffrey C. Carver, and Nicholas A. Kraft. 2013. Building reputation in stackoverflow: An empirical investigation. In Proceedings of the 10th Working Conference on Mining Software Repositories. IEEE, 89–92.

[14] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. 2009. Statistical physics of social dynamics. Reviews of Modern Physics 81, 2 (2009), 591.

[15] Huseyin Cevusoglu, Zhuolun Li, and Ke-Wei Huang. 2015. Can gamification motivate voluntary contributions?: The case of stackoverflow Q&A community. In Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing, ACM, 171–174.

[16] Laurent Charlin, Rajesh Ranganath, James McInerney, and David M. Blei. 2015. Dynamic poisson factorization. In Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 155–162.

[17] Freddy Chong Tat Chua, Hady W. Lauw, and Ee-Peng Lim. 2013. Generative models for item adoptions using social correlation. IEEE Transactions on Knowledge and Data Engineering 25, 9 (2013), 2036–2048.

[18] D. J. Daley and D. Vere-Jones. 2002. An Introduction to the Theory of Point Processes. Springer Series in Statistics, vol. I, Springer, New York.

[19] D. J. Daley and D. Vere-Jones. 2008. An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure (2nd ed.). Springer, New York. xviii+573 pages.

[20] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: Defining gamification. In Proceedings of the 15th International Academic Mindtrek Conference: Envisioning Future Media Environments. ACM, 9–15.

[21] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O’ Hara, and Dan Dixon. 2011. Gamification. using game-design elements in non-gaming contexts. In Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI EA’11). ACM, 2425–2428.

[22] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 1555–1564.

[23] Nan Du, Mehrdad Farajtabar, Amr Ahmed, Alexander J. Smola, and Le Song. 2015. Dirichlet-Hawkes processes with applications to clustering continuous-time document streams. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 219–228.

[24] Nan Du, Le Song, Hyenkyun Woo, and Hongyuan Zha. 2013. Uncover topic-sensitive information diffusion networks. In Proceedings of the 16th International Conference on Artificial Intelligence and Statistics. 229–237.

[25] Nan Du, Le Song, Ming Yuan, and Alex J. Smola. 2012. Learning networks of heterogeneous influence. In Proceedings of the Advances in Neural Information Processing Systems. 2780–2788.

[26] Nan Du, Yichen Wang, Niao He, Jimeng Sun, and Le Song. 2015. Time-sensitive recommendation from recurrent user activities. In Proceedings of the Advances in Neural Information Processing Systems. 3492–3500.

[27] Mehrdad Farajtabar, Nan Du, Manuel Gomez Rodriguez, Isabel Valera, Hongyuan Zha, and Le Song. 2014. Shaping social activity by incentivizing users. In Proceedings of the Advances in Neural Information Processing Systems. 2474–2482.

[28] Mehrdad Farajtabar, Yichen Wang, Manuel Gomez Rodriguez, Shuang Li, Hongyuan Zha, and Le Song. 2015. Co-evolve: A joint point process model for information diffusion and network co-evolution. In Proceedings of the Advances in Neural Information Processing Systems. 1954–1962.

[29] Yang Gao, Yan Chen, and K. J. Ray Liu. 2015. Understanding sequential user behavior in social computing: To answer or to vote? IEEE Transactions on Network Science and Engineering 2, 3 (July 2015), 112–126. DOI: http://dx.doi.org/10.1109/TNSE.2015.2470542.

[30] Prem Gopalan, Jake M. Hofman, and David M. Blei. 2015. Scalable recommendation with hierarchical poisson factorization. In Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence (UAI’15). 326–335.

[31] Alexander Halavais, K. Hazel Kwon, Shannon Havener, and Jason Striker. 2014. Badges of friendship: Social influence and badge acquisition on Stack Overflow. In Proceedings of the 47th Hawaii International Conference on System Sciences. IEEE, 1607–1615.

ACM Transactions on Knowledge Discovery from Data, Vol. 12, No. 3, Article 37. Publication date: March 2018.
Continuous-Time User Modeling in Presence of Badges: A Probabilistic Approach

[32] Juho Hamari. 2017. Do badges increase user activity? A field experiment on the effects of gamification. Computers in Human Behavior 71 (2017), 469–478.

[33] Xinran He, Theodoros Rekatsinas, James R. Foulds, Lise Getoor, and Yan Liu. 2015. Hawkes Topic: A joint model for network inference and topic modeling from text-based cascades. In Proceedings of the International Conference on Machine Learning (ICML’15). 871–880.

[34] Tad Hogg, Kristina Lerman, and Laura M. Smith. 2013. Stochastic models predict user behavior in social media. In Proceedings of 5th IEEE International Conference on Social Computing (SocialCom). IEEE, 63–68.

[35] Seyed Abbas Hosseini, Keivan Alizadeh, Ali Khodadadi, Ali Arabzadeh, Mehrdad Farajtabar, Hongyuan Zha, and Hamid R. Rabiee. 2017. Recurrent poisson factorization for temporal recommendation. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM.

[36] Seyed Abbas Hosseini, Ali Khodadadi, Soheil Arabzade, and Hamid R. Rabiee. 2016. HNP3: A hierarchical nonparametric point process for modeling content diffusion over social media. In Proceedings of the IEEE International Conference on Data Mining (ICDM’16). IEEE.

[37] José Luis Iribarren and Esteban Moro. 2009. Impact of human activity patterns on the dynamics of information diffusion. Physical Review Letters 103, 3 (2009), 038702.

[38] Yong Jin, Xin Yang, Raula Gaikovina Kula, Eunjong Choi, Katsuro Inoue, and Hajimu Iida. 2015. Quick trigger on stack overflow: A study of gamification-influenced member tendencies. In Proceedings of the IEEE/ACM 12th Working Conference on Mining Software Repositories. IEEE, 434–437.

[39] M. G. Kendall. 1938. A new measure of rank correlation. Biometrika 30, 1–2 (1938), 81.

[40] John Frank Charles Kingman. 1992. Poisson Processes. Oxford University Press.

[41] Kristina Lerman and Tad Hogg. 2012. Using stochastic models to describe and predict social dynamics of web users. ACM Transactions on Intelligent Systems and Technology 3, 4 (2012), 62.

[42] Dawen Liang, Laurent Charlin, James McInerney, and David M. Blei. 2016. Modeling user exposure in recommendation. In Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 951–961.

[43] Scott W. Linderman and Ryan P. Adams. 2014. Discovering latent network structure in point process data. In Proceedings of the International Conference on Machine Learning (ICML’14). 1413–1421.

[44] Eren Manavoglu, Dmitry Pavlov, and C. L. Giles. 2003. Probabilistic user behavior models. In Proceedings of the 3rd IEEE International Conference on Data Mining. IEEE Comput. Soc, 203–210. DOI: http://dx.doi.org/10.1109/ICDM.2003.1250921

[45] Andrew Marder. 2015. Stack overflow badges and user behavior: An econometric approach. In Proceedings of the 12th Working Conference on Mining Software Repositories. IEEE, 450–453.

[46] Dana Movshovitz-Attias, Yair Movshovitz-Attias, Peter Steenkiste, and Christos Faloutsos. 2013. Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM’13). IEEE, 886–893.

[47] Seth A. Myers, Chenguang Zhu, and Jure Leskovec. 2012. Information diffusion and external influence in networks. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 33–41.

[48] Yoshihiko Ogata. 1981. On Lewis’ simulation method for point processes. IEEE Transactions on Information Theory 27, 1 (1981), 23–31.

[49] Lawrence R. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77, 2 (1989), 257–286.

[50] V. Raghavan, G. Ver Steeg, A. Galstyan, and A. G. Tartakovsky. 2014. Modeling temporal activity patterns in dynamic social networks. IEEE Transactions on Computational Social Systems 1, 1 (March 2014), 89–107. DOI: http://dx.doi.org/10.1109/TCSS.2014.2307453

[51] Jakob Gulddahl Rasmussen. 2011. Temporal point processes the conditional intensity function (Lecture Notes). Aalborg University, Denmark. Retrieved from http://people.math.aau.dk/~jgr

[52] Manuel Gomez Rodriguez, David Balduzzi, and Bernhard Schölkopf. 2011. Uncovering the temporal dynamics of diffusion networks. In Proceedings of 28th International Conference on Machine Learning (ICML ’11). IMLS, 561–568.

[53] Tanmay Sinha, Wei Wei, and Kathleen Carley. 2015. Modeling similarity in incentivized interaction: A longitudinal case study of stackoverflow. In Proceedings of the Workshop on Social and Information Networks at 29th Annual International Conference on Neural Information and Processing Systems, (NIPS’15).

[54] Maryam Tahani, Ali M. A. Hemmatyar, Hamid R. Rabiee, and Maryam Ramezani. 2016. Inferring dynamic diffusion networks in online media. ACM Transactions on Knowledge Discovery from Data 10, 4 (June 2016), Article 44, 22 pages. DOI: http://dx.doi.org/10.1145/2882968

[55] Isabel Valera and Manuel Gomez-Rodriguez. 2015. Modeling adoption and usage of competing products. In Proceedings of the International Conference on Data Mining (ICDM’15). IEEE, 409–418.
[56] Dingqi Yang, Daqing Zhang, Vincent W. Zheng, and Zhiyong Yu. 2015. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. IEEE Transactions on Systems, Man, and Cybernetics: Systems 45, 1 (2015), 129–142.

[57] Shuang-Hong Yang and Hongyuan Zha. 2013. Mixture of mutually exciting processes for viral diffusion. In Proceedings of the International Conference on Data Mining (ICML’13), vol. 28, II-1–II-9.

[58] Hongzhi Yin, Bin Cui, Ling Chen, Zhitong Hu, and Xiaofang Zhou. 2015. Dynamic user modeling in social media systems. ACM Transactions on Information Systems 33, 3 (March 2015), 1–44. DOI: http://dx.doi.org/10.1145/2699670

[59] Ali Zarezade, Sina Jafarzadeh, and Hamid R. Rabiee. 2016. Spatio-temporal modeling of check-ins in location-based social networks. arXiv:1611.07710v2.

[60] Ali Zarezade, Ali Khodadadi, Mehrdad Farajtabar, Hamid R. Rabiee, and Hongyuan Zha. 2015. Correlated cascades: Compete or cooperate. In Proceedings of 31st AAAI Conference on Artificial Intelligence (AAAI’17). AAAI, 238–244.

[61] Ali Zarezade, Utkarsh Upadhyay, Hamid R. Rabiee, and Manuel Gomez-Rodriguez. 2017. RedQueen: An online algorithm for smart broadcasting in social networks. In Proceedings of the 10th ACM International Conference on Web Search and Data Mining (WSDM’17). Cambridge, United Kingdom, (February 6–10). 51–60. http://dl.acm.org/citation.cfm?id=3018684.

[62] Jiawei Zhang, Xiangnan Kong, and S. Yu Philip. 2016. Social badge system analysis. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM’16). IEEE, 453–460.

[63] Ke Zhou, Hongyuan Zha, and Le Song. 2013. Learning social infectivity in sparse low-rank networks using multidimensional Hawkes processes. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS’13), vol. 31, 641–649.

Received June 2017; revised November 2017; accepted November 2017