

---

# The Hawkes Edge Partition Model for Continuous-time Event-based Temporal Networks

---

**Sikun Yang**

German Center for Neurodegenerative Diseases  
Bonn, Germany  
sikun.yang@dzne.de

**Heinz Koepl**

Technische Universität Darmstadt  
Darmstadt, Germany  
heinz.koepl@bcs.tu-darmstadt.de

## Abstract

We propose a novel probabilistic framework to model continuous-time interaction events data. Our goal is to infer the *implicit* community structure underlying the temporal interactions among entities, and also to exploit how the community structure influences the interaction dynamics among these nodes. To this end, we model the reciprocating interactions between individuals using mutually-exciting Hawkes processes. The base rate of the Hawkes process for each pair of individuals is built upon the latent representations inferred using the hierarchical gamma process edge partition model (HGaP-EPM). In particular, our model allows the interaction dynamics between each pair of individuals to be modulated by their respective affiliated communities. Moreover, our model can flexibly incorporate the auxiliary individuals' attributes, or covariates associated with interaction events. Efficient Gibbs sampling and Expectation-Maximization algorithms are developed to perform inference via Pólya-Gamma data augmentation strategy. Experimental results on real-world datasets demonstrate that our model not only achieves competitive performance for temporal link prediction compared with state-of-the-art methods, but also discovers interpretable latent structure behind the observed temporal interactions.

## 1 INTRODUCTION

There has been considerable interest in modeling and understanding the information diffusion pathways and interaction dynamics among entities from continuously generated streams of data. These streaming data include the timestamped interaction events among entities (e.g.,

question-answering threads (Mavroforakis et al., 2017), email communications (Yang et al., 2017) and interaction events among nations (Schein et al., 2016; Yang et al., 2018; Schein et al., 2019)), and the auxiliary contents created by these interacting entities. Such temporal interaction data enable us not only to track the *topics* underlying the human-generated contents, but also to understand the network formation and evolving process among these interacting entities.

A fundamental problem in the analysis of continuous-time interaction events is to capture the underlying community structure and reciprocity in these interactions. Reciprocity is a common social norm, in which an individual's actions towards another will increase the likelihood of the same type of action being returned in the near future (Blundell et al., 2012). Specifically, Hawkes processes are well-fitted to model such reciprocating behaviors in temporal interactions. To further capture the underlying community structure, some recent works (Blundell et al., 2012; DuBois et al., 2013; Linderman et al., 2014; Yang et al., 2017; Miscouridou et al., 2018) attempt to hybridize statistical models for *static* networks with Hawkes processes to model both *implicit* social structure and *reciprocity* among entities. The Hawkes stochastic block models (Hawkes-SBMs) (Blundell et al., 2012; Junuthula et al., 2019; Arastuie et al., 2019) characterize the interaction dynamics between groups of individuals using mutually-exciting Hawkes processes. To further capture the reciprocity between each pair of two individuals, Miscouridou et al. (2018) proposes to model pair-wise reciprocating dynamics by letting the base intensities depending on the underlying community structure.

Despite having many attractive properties, the Hawkes-CCRM (Miscouridou et al., 2018) is restrictive in that the reciprocity in all the interactions are captured via the same triggering kernel, and thus cannot interpret the differences in interaction dynamics across individuals. For example, an employee may reply back to the emails from

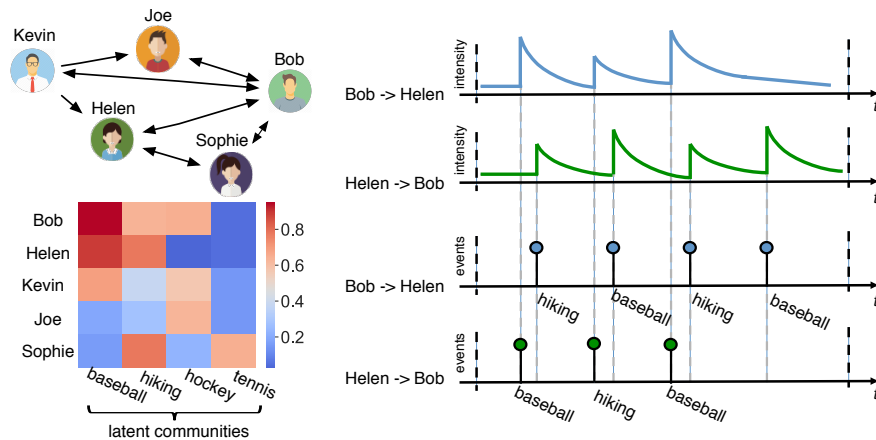


Figure 1: An illustrative example. The top left figure plots the aggregated *directed* networks from the interactions among five nodes. The bottom left graph shows the underlying community structure. We see that both Bob and Helen have interest in “baseball” and “hiking”. The top right graph plots the intensity functions of the interactions from Bob to Helen, and from Helen to Bob, respectively. The bottom right graph plots the interactions from Bob to Helen, and from Helen to Bob, respectively. These interactions may represent the messages between the involved nodes. As in this example, some of their interactions are about “baseball”, and others relate to “hiking”. We assume that behind each interaction, the latent patterns of the involved nodes determines the excitation effects of that event on the opposite direction.

his/her department more quickly than responding to non-urgent emails from outside. A fundamental problem in modeling such temporal dynamics is to infer the *latent structure* behind observed events (Du et al., 2015; Mavrouforakis et al., 2017; Xu et al., 2017; Tan et al., 2018a). To account for heterogeneity both in how two individuals initiate interactions as well as in the dynamics within each specific event, Yang et al. (2017) proposes to modulate both the *base* and *reciprocate* rate with a dual latent space model, instead of exploiting the latent structure underlying observed events.

In this paper, we attempt to develop a new framework, the Hawkes edge partition model (Hawkes-EPM), which hybridizes the recently advanced hierarchical gamma process edge partition model (HGaP-EPM) (Zhou, 2015) with Hawkes processes. More specifically, the base intensity of the Hawkes process is built upon the latent representations inferred by the HGaP-EPM, which enables us to capture the *overlapping* communities, degree *heterogeneity* and *sparsity* underlying the observed interactions. To accurately capture the interaction dynamics between two individuals, our model augments each specific interaction between them with a pair of latent variables, to indicate which of their latent communities (features) leads to the occurring of that interaction. Accordingly, the excitation effect of each interaction on its opposite direction is determined by its latent variables. For instance, as shown in Figure 1, Bob and Helen have many common interests (features), and some of their interactions are due to their common interests in playing

baseball. Moreover, our model can automatically determine the number of the underlying communities via the inherent shrinkage mechanism of the hierarchical gamma process (Zhou and Carin, 2015). Furthermore, our model construction can flexibly incorporate the auxiliary individuals’ attributes, or covariates associated with interaction events.

**Contributions.** We make the following contributions: (1) We propose a statistical model for continuous-time dynamic networks by capturing the underlying community structure via the base rate of the mutually-exciting Hawkes process, and estimating the number of communities with the hierarchical gamma process. (2) The proposed model accounts for heterogeneity both in *exogenous* and *endogenous* activities. (3) Efficient approximate inference can be performed with closed-form update equations using data augmentation techniques. (4) The developed model is applied for temporal link prediction using real-world data, and shows competitive performance compared with state-of-the-art models.

The paper is organized as follows. Section 2 shortly reviews the necessary background. Section 3 describes the Hawkes-EPM model. Section 4 discusses how the proposed model relates to previous works. Section 5 describes the developed inference procedure. Section 6 presents the experimental results on real-world interaction event datasets.

## 2 BACKGROUND

The proposed Hawkes edge partition model is built upon the hierarchical gamma process edge partition model, which infers the underlying community structure behind the aggregated temporal events, and also relies on Hawkes processes, which capture the reciprocating behaviors between nodes. Next we shall briefly review the two building components.

### 2.1 HIERARCHICAL GAMMA PROCESS EDGE PARTITION MODELS

The hierarchical gamma process edge partition (HGAP-EPM) model (Zhou, 2015) was recently proposed to detect overlapping community structure in static relational data. Formally, let  $\mathcal{V}$  denotes a set of nodes, and the (static) relationships among  $V \equiv |\mathcal{V}|$  nodes be represented by a binary adjacency matrix  $E \in \{0, 1\}^{V \times V}$ , where  $e_{u,v} = 1$  if there is an (directed) edge from nodes  $u$  to  $v$ , and 0 otherwise. We ignore self-edges  $\{e_{u,u}\}_{u \in \mathcal{V}}$  as a node never interacts with itself. The (truncated) HGAP-EPM is generated as

$$\begin{aligned} \phi_{u,k} &\sim \text{Gamma}(a_u, 1/c_u), & a_u &\sim \text{Gamma}(e_0, 1/f_0), \\ r_k &\sim \text{Gamma}(r_0/K, 1/c_0), \\ \Omega_{k,k'} &\sim \begin{cases} \text{Gamma}(\xi r_k, \chi), & \text{if } k = k' \\ \text{Gamma}(r_k r_{k'}, \chi), & \text{otherwise} \end{cases}, \\ e_{u,v} &\sim \text{Bernoulli} \left[ 1 - \prod_{k,k'=1}^K \exp(-\phi_{u,k} \Omega_{k,k'} \phi_{v,k'}) \right], \end{aligned}$$

where each node  $u \in \mathcal{V}$  is characterized by a *positive* feature vector  $[\phi_{u,1}, \dots, \phi_{u,K}]^T$  with  $\phi_{u,k}$  measuring how strongly node  $u$  is affiliated to each community  $k = 1, \dots, K$ .  $a_u$  captures the *sociability* of node  $u$ , and thus node  $u$  exhibiting a large number of interactions will be characterized by a large  $a_u$ . The prevalence of each community  $k$  is captured by a positive weight  $r_k$ . The HGAP-EPM can infer an appropriate number of communities via its inherent shrinkage mechanism: many communities' weights  $\{r_k\}$  tend to be small as  $K \rightarrow \infty$ , and thus most redundant communities will be shrunk effectively. The parameters  $\Omega_{k,k}$  and  $\Omega_{k,k'}$  capture the intra-community and inter-community interaction weights, respectively. In particular,  $\xi$  prevents overly shrinking  $\Omega_{k,k}$  for small communities. The probability of there being an edge from node  $u$  to node  $v$  is parameterized under the Bernoulli-Poisson link (BPL) function  $\Pr(y = 1 | \zeta) = 1 - e^{-\zeta}$ , where  $\zeta$  defines the positive rate. Following (Zhou, 2015), we impose the Gamma(1, 1) prior over the hyperparameters  $c_u, c_0, e_0, f_0, r_0, \xi, \chi$ , independently. Interestingly, the probability of an edge  $e_{u,v}$  mod-

eled by the BPL can be equivalently generated as

$$\begin{aligned} e_{u,v} &= \mathbf{1}(\tilde{e}_{u,v} \geq 1), \\ \tilde{e}_{u,v} &\sim \text{Poisson} \left( \sum_{k=1}^K \sum_{k'=1}^K \phi_{u,k} \Omega_{k,k'} \phi_{v,k'} \right), \end{aligned}$$

where  $\phi_{u,k} \Omega_{k,k'} \phi_{v,k'}$  capture the connecting strength between nodes  $u$  and  $v$  due to their affiliations to communities  $k, k'$ , respectively. Note that the HGAP-EPM not only captures the *overlapping* community structure, degree *heterogeneity*, but also characterizes structured sparsity patterns in community-community interactions (Zhou, 2018a).

### 2.2 HAWKES PROCESSES

Let  $N(t)$  be a counting process recording the number of events occurring at times  $\{t_i\}$  with  $t_i < t$ . The probability of an event occurring in a small time interval  $[t, t+dt)$  is given by  $\Pr(dN(t) = 1 | \mathcal{H}(t)) = \lambda(t)dt$ , where  $\mathcal{H}(t) \equiv \{t_i | t_i < t\}$  denotes the history of events up to but not including time  $t$ ,  $dN(t)$  is the increment of the process, and  $\lambda(t)$  is the conditional intensity function (intensity, for short) of  $N(t)$ . A Hawkes process is a stochastic point process (Daley and Vere-Jones, 2003) with intensity function defined as

$$\lambda(t) = \mu + \int_0^t \gamma(t-s) dN(s) = \mu + \sum_{j:t_j \in \mathcal{H}(t)} \gamma(t-t_j),$$

where  $\mu \geq 0$  is the base rate capturing the *exogenous* activities, and  $\gamma(t)$  is the nonnegative triggering kernel modelling the *endogenous* activities. Note that this intensity function characterizes the self-excitation effects that past events have on the current event rate. Here, we consider an exponential kernel  $\gamma(t-s) \equiv \alpha \exp[-(t-s)/\delta]$  where  $\alpha \geq 0$  determines the magnitude of excitations, which exponentially decays with a constant rate  $\delta \geq 0$ . The stationary condition for Hawkes processes requires  $\alpha\delta < 1$ . Recent work (Blundell et al., 2012; Yang et al., 2017; Miscouridou et al., 2018; Junuthula et al., 2019; Arastuie et al., 2019) were proposed to capture the reciprocity in communications between a pair of individuals using mutually-exciting Hawkes processes. Formally, for a pair of nodes  $u, v \in V$ , we have the counting processes  $N_{u,v}(t)$ , which defines the number of *directed* interactions from node  $u$  to node  $v$  in the time interval  $[0, t)$ . Let the history of interactions from nodes  $u$  to  $v$  be denoted as  $\mathcal{H}_{u,v}(t)$ . Accordingly,  $N_{u,v}(t)$  and  $N_{v,u}(t)$  are mutually-exciting Hawkes processes if their intensity

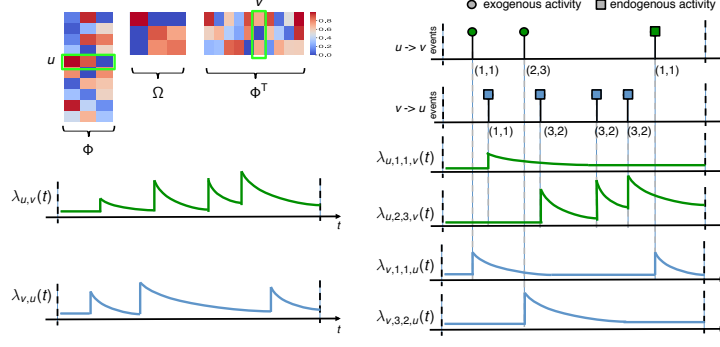


Figure 2: A simple example for the Hawkes-EPM model. The top left figure shows the inferred matrix of node features  $\Phi$ , and the community-community interaction strength  $\Omega$ . Here, node  $u$  connects to node  $v$  through the intra-community interaction (1, 1) and inter-community interaction (2, 3). The top right figure plots the interaction events between  $u$  and  $v$ . Each event is denoted by a bar, under which we use  $(a, b)$  to indicate the latent variables  $a, b$  of nodes  $u, v$  in that event, e.g.,  $z_1^u = 1, z_1^v = 1$  for 1-st event. The bottom left figure plots the intensities of the interactions from  $u$  to  $v$ , and from  $v$  to  $u$ , respectively. Equivalently,  $\lambda_{u,v}(t)$  can be represented by the summation of  $\{\lambda_{u,k,k',v}(t)\}_{k,k'}$ , where  $\lambda_{u,k,k',v}(t)$  denotes the interaction intensity from  $u$  to  $v$  via the inter-community  $(k, k')$ .

functions take the forms

$$\lambda_{u,v}(t) = \mu_{u,v} + \sum_{t_j \in \mathcal{H}_{v,u}(t)} \gamma(t - t_j),$$

$$\lambda_{v,u}(t) = \mu_{v,u} + \sum_{t_i \in \mathcal{H}_{u,v}(t)} \gamma(t - t_i),$$

respectively. Note that mutually-exciting Hawkes processes capture the reciprocating interactions from node  $u$  to node  $v$  at time  $t$  as a response to the past interactions from  $v$  to  $u$ .

### 3 THE HAWKES EDGE PARTITION MODEL

Let  $\{(t_i, s_i, d_i)\}_{i=1}^N$  be a sequence of temporal interaction events, where  $(t_i, s_i, d_i)$  is a *directed* interaction from node  $s_i$  (sender) to node  $d_i$  (receiver) at time  $t_i$ . To capture reciprocity in interactions, mutually exciting Hawkes processes (MHPs) assume that a specific event  $(t_i, s_i, d_i)$  is either an exogenous event triggered by the base rate  $\mu_{s_i, d_i}$ , or is an endogenous one, responding to a past event.

To further capture the underlying community structure, we augment each event  $(t_i, s_i, d_i)$  with two auxiliary variables  $z_i^s$  and  $z_i^d$ , which refer to the latent communities affiliated with respectively the sender and receiver. Hence, for the event sequence from node  $u$  to node  $v$ , the first event is driven by one of the sub-rates,  $\{\mu_{u,k,k',v}\}_{k,k'}$  where  $\mu_{u,k,k',v}$  denotes the sub-rate accounting for the exogenous interactions from  $u$  to  $v$  due to their respective affiliations to  $k, k'$ . Accordingly, each subsequent event from  $u$  to  $v$  is either driven by one of its corresponding sub-rates, or driven by a past event of the

opposite direction. Figure 2 presents a simple illustrative example for the Hawkes Edge Partition Model (Hawkes-EPM).

Formally, for a pair of two nodes  $u$  and  $v$ , the base rate  $\mu_{u,v}$  is built upon the latent parameters  $\{\phi_{u,k}\}_{u,k}$  and  $\{\Omega_{k,k'}\}_k$  inferred using the HGaP-EPM. More specifically, we define the intensity function for nodes  $u$  and  $v$  as

$$\lambda_{u,v}(t) = \sum_{k,k'} \lambda_{u,k,k',v}(t),$$

$$\lambda_{u,k,k',v}(t) = \mu_{u,k,k',v} + \sum_{j: t_j \in \mathcal{H}_{v,k',k,u}(t)} \gamma_{k,k'}(t - t_j), \quad (1)$$

$$\gamma_{k,k'}(t - s) = \alpha_{k,k'} \exp[-(t - s)/\delta],$$

where  $\lambda_{u,v}(t)$  factorizes into the summation of the sub-intensities  $\{\lambda_{u,k,k',v}(t)\}_{k,k'}$ . We set the base rate  $\mu_{u,v} = \sum_{k,k'} \mu_{u,k,k',v}$ , where  $\mu_{u,k,k',v} = \phi_{u,k} \Omega_{k,k'} \phi_{v,k'}$ .  $\phi_{u,k}$  captures the affiliation of node  $u$  to community  $k$ , and  $\Omega_{k,k'}$  the inter-community interaction strength between  $k$  and  $k'$ . Hence, the base rate  $\mu_{u,v}$  naturally models that two nodes sharing more features are more likely to interact with each other.

In this work, we assume that if an occurring event is driven by a past event, the latent pattern of the occurring event is also determined by that past event. To this end, the rate  $\lambda_{u,k,k',v}(t)$  from  $u$  to  $v$  under the pattern  $(k, k')$ , is only allowed to be influenced by the past opposite interactions under the pattern  $(k', k)$ ,  $\{(t_j, s_j, d_j) \mid t_j < t, s_j = v, d_j = u, z_j^s = k', z_j^d = k\}$ , which we denote by  $\{t_j \in \mathcal{H}_{v,k',k,u}(t)\}$ . Therefore, in Eq. (2), we define a nonnegative kernel function  $\gamma_{k,k'}$ , which captures the decaying influence of past events under the pattern  $(k', k)$  on the current intensity. More specifically,  $\alpha_{k,k'}$

controls the excitatory effect under the pattern  $(k', k)$ , and we impose a gamma prior over  $\alpha_{k,k'}$ , i.e.,  $\alpha_{k,k'} \sim \text{Gamma}(1, 1)$ . As reported in related works (Yang et al., 2017; Mavroforakis et al., 2017), we find that inferring time scale  $\delta$  suffers from identifiability issue. Instead of modeling temporal dynamics via weighted combinations of basis kernels, we allow  $\alpha_{k,k'}$  to be varying between different patterns but fix  $\delta$  as a constant. Putting all this together, the conditional intensity function of the Hawkes-EPM, for the directed events from  $u$  to  $v$ , is

$$\begin{aligned} \lambda_{u,v}(t) &= \mu_{u,v} + \sum_{j:t_j \in \mathcal{H}_{v,k',k,u}(t)} \gamma_{k,k'}(t - t_j) \\ &= \sum_{k,k'} \left\{ \mu_{u,k,k',v} + \sum_{j:t_j \in \mathcal{H}_{v,k',k,u}(t)} \alpha_{k,k'} \exp[-(t - t_j)/\delta] \right\}, \end{aligned} \quad (2)$$

The latent patterns associated with each interaction is sampled as follows. If  $(t_i, s_i, d_i)$  is an *exogenous* event induced by  $\mu_{s_i, d_i}$ , the latent patterns  $z_i^s, z_i^d$  for  $s_i, d_i$  are determined by their affiliated communities via  $\phi_{s_i}, \phi_{d_i}$ , respectively. In case that  $(t_i, s_i, d_i)$  is an *endogenous* event,  $z_i^s, z_i^d$  are determined by the past opposite interactions from  $d_i$  to  $s_i$ . More specifically, the latent patterns associated to  $i$ -th event can be generated as

$$\begin{aligned} \Pr(z_i^s = k, z_i^d = k' \mid t_i, s_i = u, d_i = v) \\ = \left( \mu_{u,k,k',v} + \sum_{j:t_j \in \mathcal{H}_{v,k',k,u}(t_i)} \alpha_{k,k'} \exp[-(t_i - t_j)/\delta] \right) / \lambda_{u,v}(t_i), \end{aligned} \quad (3)$$

In real temporal interactions, some additional information such as auxiliary node attributes, explicitly declared relationships among entities, and communicating contents are also available for accurately modelling temporal interaction dynamics when interaction events are incomplete (say, due to the privacy issues of individuals). Formally, let  $\mathbf{x}_{u,v} \equiv [x_{u,v}^1, \dots, x_{u,v}^D]^T$  denotes the covariates of  $D$ -dimension associated with a pair of nodes  $u$  and  $v$ . For example, the covariates  $\mathbf{x}_{u,v}$  may represent the common attributes shared by  $u$  and  $v$ , or the word embeddings inferred from the communicating contents between  $u$  and  $v$ . We generalize the Hawkes-EPM model by letting

$$\mu_{u,k,k',v} \sim \text{Gamma}(\tilde{\mu}_{u,k,k',v}, 1/(\exp[-\mathbf{x}_{u,v}^T \boldsymbol{\beta}_{k,k'}])), \quad (4)$$

where  $\tilde{\mu}_{u,k,k',v} \equiv \phi_{u,k} \Omega_{k,k'} \phi_{v,k'}$ , and  $\boldsymbol{\beta}_{k,k'} \equiv (\beta_{k,k'}^1, \dots, \beta_{k,k'}^D)^T$  is the regression coefficient vector of latent pattern  $(k, k')$ . The base intensity in (4) is drawn from a gamma prior where the shape parameter incorporates the underlying community structure information via  $\tilde{\mu}_{u,k,k',v}$ , and the scale parameter is a function of the

input auxiliary covariates. To our knowledge, the regression component in (4) closely relates to (Rai et al., 2015; Zhou, 2018b), but firstly applied in this context and the inference derivation is non-trivial.

**Remarks.** Note that the proposed model allows an unbounded number of latent patterns to be shared across all pairs of interacting nodes via the hierarchical gamma process (HGAP) (Zhou and Carin, 2015). As shown in Eq. (3), the sub-rate  $\mu_{u,k,k',v}$  of the latent pattern  $(k, k')$  is non-negligible over the whole time period, and thus our model allows the events widely separated in time but with similar dynamics to be parameterized under the same pattern, to avoid *vanishing* prior issue (Mavroforakis et al., 2017; Kapoor et al., 2018).

## 4 RELATED WORK

The proposed model closely relates to the Hawkes process-based interaction models and the Bayesian nonparametric prior-based Hawkes process models.

**Hawkes Processes-based Interaction Models.** Blundell et al. (2012) describes the Hawkes stochastic block model (Hawkes-SBM), in which each node is allowed to be affiliated with only one community (non-overlapping), and the interaction dynamics between two nodes are determined by their respective community-specific intensities. The recent extensions (Junuthula et al., 2019; Arastuie et al., 2019) can be seen as the variants of Hawkes-SBMs.

Tan et al. (2018a) describes an Indian buffet Hawkes process model, which assumes that each event can be simultaneously driven by multiple evolving factors shared among events. In contrast, the Hawkes-EPM relies on a *clustering* structure, where each interaction is categorized as one subtype, while the multiple evolving subtypes are shared among behind the events.

Miscouridou et al. (2018) describes an unified framework, which captures the overlapping community structure, graph sparsity and degree heterogeneity using compound completely random measure model, and models reciprocity between each pair of nodes via mutually exciting Hawkes processes. However, Hawkes-CCRM cannot capture the differences in temporal dynamics of individuals by using the same triggering kernel for all the entities. Our proposed model not only models the interpretable latent structure underlying observed interactions as in (Miscouridou et al., 2018), but also captures the latent pattern behind each event using community-specific triggering kernels.

**Bayesian Nonparametric Hawkes Processes (BNHPs).** Recently, Bayesian nonparametric priors (BNPs) (Ferguson, 1973) are introduced to capture the

$$\begin{aligned} \mathcal{L}(\Theta) = & \sum_i \log \left\{ \mu_{s_i, d_i} + \sum_{k, k'} \sum_{j: t_j \in \mathcal{H}_{d_i, k', k, s_i}(t_i)} \alpha_{kk'} \exp[-(t_i - t_j)/\delta] \right\} \\ & - \sum_i \left\{ \mu_{s_i, d_i} T + \sum_{k, k'} \sum_{j: t_j \in \mathcal{H}_{d_i, k', k, s_i}(t_i)} \alpha_{kk'} \delta (1 - \exp[-(t_i - t_j)/\delta]) \right\} + \log \Pr(\Theta). \end{aligned} \quad (5)$$

latent structure underlying the observed event sequence. The Dirichlet-Hawkes process (DHP) (Du et al., 2015) models the latent clustering structure underlying the observed events using the Dirichlet process.

The Indian buffet Hawkes process (Tan et al., 2018a) and the nested Chinese restaurant process-Hawkes process (NCRP-HP) (Tan et al., 2018b) have been developed to capture the rich factor-structured and hierarchically-structured temporal dynamics, respectively.

Mavroforakis et al. (2017) points out that most previous BNHP models suffer from the vanishing prior problem as the instantiated patterns in these models are only captured via the endogenous intensity. Hence, an already used pattern will vanish if its intensity tend to be zero. As a consequence, these BNHP methods unavoidably generate many redundant patterns for the events widely separated in time but sharing similar dynamics. Mavroforakis et al. (2017) resolved this issue using the hierarchical Dirichlet process (Teh et al., 2006) framework, where the top-layer Dirichlet process defines the distribution over latent patterns, and the bottom-layer Hawkes processes capture the temporal dynamics across multiple event sequences. Nevertheless, it is unclear how to generalize the Hierarchical Dirichlet Hawkes Process (HDHP) to model temporal interaction events. Our proposed model infers the appropriate number of communities (patterns) using the hierarchical gamma process prior (Zhou and Carin, 2015). In the Hawkes-EPM, each latent pattern is modelled by a community-specific intensity function, which is non-negligible over time, and thus effectively prevents from the vanishing prior issue.

## 5 INFERENCE

The proposed model admits efficient approximate inference as the posteriors of all the model parameters are available in closed-form using Pólya-Gamma data augmentation strategy. Let  $\mathcal{D}$  denote the whole events data,  $E$  the binary adjacency matrix aggregated from  $\mathcal{D}$ , i.e.,  $e_{u,v} = 1$  for  $u, v \in \mathcal{V}$  if there being at least one interaction observed in the time interval  $[0, T]$ ,  $\Xi$  the parameters of the HGaP-EPM, and  $\Theta$  the parameters of the Hawkes-EPM. The model parameters of the Hawkes-EPM consist of

$\{\phi_{u,k}, \Omega_{k,k'}, \mu_{u,k,k',v}, \alpha_{k,k'}, \beta_{k,k'}, \psi_{k,k'}, \omega_{u,k,k',v}\}$ . We use the “ $\hat{x}$ ” to denote the maximum a posteriori (MAP) estimate of  $x$ . A two-step inference procedure is developed to perform maximum-a-posteriori (MAP) estimate: (i) Approximate  $\Pr(\Xi | \mathcal{D}, E)$  by  $\Pr(\Xi | E)$ , and obtain a MAP estimate  $\hat{\Xi}$ , and then (ii) Approximate  $\Pr(\Theta | \Xi, \mathcal{D})$  by  $\Pr(\Theta | \hat{\Xi}, \mathcal{D})$ . The full posterior is approximated by  $\Pr(\Theta, \Xi) = \Pr(\Xi | E) \Pr(\Theta | \hat{\Xi}, \mathcal{D})$ . The posterior inference for  $\hat{\Xi}$  is performed using the Gibbs sampling procedure described in (Zhou, 2015). Next we shall explain the Expectation-Maximization algorithms to perform MAP estimation following (Lewis et al., 2011; Zhou et al., 2013; Xu et al., 2016). We also use index summation shorthands:  $\cdot$  sum out that index, e.g.  $x_{\cdot j} = \sum_i x_{ij}$ . The log-posterior of the observed temporal events  $\mathcal{D} \equiv \{(t_i, s_i, d_i)\}_{i=1}^N$  is shown in Eq. (5). More specifically, let  $\Theta^{(l)}$  denote the current model parameters, we construct a tight upper-bound of log-posterior in (5) via the Jensen’s inequality as

$$\begin{aligned} \mathcal{Q}(\Theta | \Theta^{(l)}) = & \quad (6) \\ & - \sum_i \left\{ \mu_{s_i, d_i} T + \sum_{k, k'} \sum_{j: t_j \in \mathcal{H}_{d_i, k', k, s_i}(t_i)} \gamma_{kk'}(t_i - t_j) \right\} \\ & + \sum_i \sum_{k, k'} \hat{p}_{i,k,k'} \log \left[ \frac{\mu_{s_i, k, k', d_i}}{\hat{p}_{i,k,k'}} \right] \\ & + \sum_{k, k'} \sum_{j: t_j \in \mathcal{H}_{d_i, k', k, s_i}(t_i)} \check{p}_{i,k,k'} \left[ \frac{\gamma_{kk'}(t_i - t_j)}{\check{p}_{i,k,k'}} \right] \\ & + \log \Pr(\Theta), \end{aligned}$$

where we define  $\hat{p}_{i,k,k'}$  and  $\check{p}_{i,k,k'}$  as

$$\begin{aligned} \hat{p}_{i,k,k'} = & \frac{\mu_{s_i, k, k', d_i}^{(l)}}{\mu_{s_i, d_i}^{(l)} + \sum_{k, k'} \sum_{j: t_j \in \mathcal{H}_{d_i, k', k, s_i}(t_i)} \gamma_{kk'}^{(l)}(t_i - t_j)}, \\ \check{p}_{i,k,k'} = & \frac{\sum_{j: t_j \in \mathcal{H}_{d_i, k', k, s_i}(t_i)} \gamma_{kk'}^{(l)}(t_i - t_j)}{\mu_{s_i, d_i}^{(l)} + \sum_{k, k'} \sum_{j: t_j \in \mathcal{H}_{d_i, k', k, s_i}(t_i)} \gamma_{kk'}^{(l)}(t_i - t_j)}. \end{aligned} \quad (7)$$

in which  $\hat{p}_{i,k,k'}$  can be interpreted as the probability that  $i$ -th event is drawn from the base rate under the latent pattern  $(k, k')$ .  $\check{p}_{i,k,k'}$  is the probability that  $i$ -th

event is triggered by the opposite interaction events under the pattern  $(k', k)$ . Accordingly, we update the sufficient statistics as

$$\begin{aligned}\hat{m}_{u,k,k',v} &\equiv \sum_{i:s_i=u, d_i=v} \hat{p}_{i,k,k'}, \\ \check{m}_{u,k,k',v} &\equiv \sum_{i:s_i=u, d_i=v} \check{p}_{i,k,k'}.\end{aligned}\quad (8)$$

Expectations of Pólya-Gamma random variables are available in closed-form (Scott et al., 2013), and given by

$$\begin{aligned}\mathbb{E}\left[\omega_{u,k,k',v}^{(l+1)}\right] &= \\ &\left(\frac{\tilde{\mu}_{u,k,k',v} + \hat{m}_{u,k,k',v}}{2\psi_{u,k,k',v}^{(l)}}\right) \tanh\left(\frac{\psi_{u,k,k',v}^{(l)}}{2}\right)\end{aligned}\quad (9)$$

Maximizing  $\mathcal{Q}(\Theta)$  with respect to each of the model parameters  $\{\mu_{u,k,k',v}\}$ ,  $\{\alpha_{k,k'}\}$ ,  $\{\beta_{k,k'}\}$ ,  $\{\psi_{k,k'}\}$  fixing the rest, leads to closed-form updates for each of these:

$$\mu_{u,k,k',v}^{(l+1)} = \frac{\tilde{\mu}_{u,k,k',v} + \hat{m}_{u,k,k',v}}{T + \exp[-\mathbf{x}_{u,v}^T \boldsymbol{\beta}_{kk'}^{(l)}]}.\quad (10)$$

Via the gamma conjugacy, we update  $\alpha_{kk'}$  as

$$\alpha_{k,k'}^{(l+1)} = \frac{e_0 + \sum_{u,v} \check{m}_{u,k,k',v}}{f_0 + \sum_i \sum_j t_j: t_j \in \mathcal{H}_{d_i, s_i}(t_i) \delta \left(1 - \exp\left[-\frac{(T-t_j)}{\delta}\right]\right)}.\quad (11)$$

Given the expectations of Pólya-Gamma random variables  $\mathbb{E}[\omega_{u,k,k',v}]$ , we update  $\psi_{k,k'}$  and  $\beta_{k,k'}$  as

$$\begin{aligned}\boldsymbol{\psi}_{k,k'}^{(l+1)} &= \left[\text{diag}(\mathbb{E}[\boldsymbol{\omega}_{k,k'}^{(l)}]) + \tau \mathbf{I}\right]^{-1} \\ &\times \left[\frac{\tilde{m}_{k,k'} - \boldsymbol{\mu}_{k,k'}^{(l)}}{2} + \tau(\mathbf{X}^T \boldsymbol{\beta}_{k,k'}^{(l)} + \log(T))\right], \\ \boldsymbol{\beta}_{k,k'}^{(l+1)} &= (\mathbf{X}^T \mathbf{X} + \tau^{-1} \mathbf{A})^{-1} \mathbf{X}^T \left(\boldsymbol{\psi}_{k,k'}^{(l)} - \log(T)\right),\end{aligned}\quad (13)$$

in which for clarity we define the following notations

$$\begin{aligned}\boldsymbol{\omega}_{k,k'}^{(l)} &\equiv [\omega_{1,k,k',1}^{(l)}, \dots, \omega_{U,k,k',V}^{(l)}], \\ \tilde{m}_{k,k'} &\equiv [\tilde{m}_{1,k,k',1}, \dots, \tilde{m}_{U,k,k',V}]^T, \\ \mathbf{A} &\equiv \text{diag}[\nu_1^{-1}, \dots, \nu_D^{-1}], \\ \boldsymbol{\mu}_{k,k'}^{(l)} &\equiv [\mu_{1,k,k',1}^{(l)}, \dots, \mu_{U,k,k',V}^{(l)}]^T, \\ \mathbf{X} &\equiv [\mathbf{x}_{1,1}, \dots, \mathbf{x}_{U,V}]^T.\end{aligned}$$

The full procedure of our EM algorithm is summarized in Algorithm 1. We also develop a simple-to-implement

---

**Algorithm 1** Expectation-Maximization algorithm for the Hawkes Edge Partition Model

---

**Input:** events data  $\mathcal{D} = \{(t_i, s_i, d_i)\}_{i=1}^N$ ,  $\{\Phi, \Omega\}$  inferred by the HGaP-EPM, time scale  $\delta$

**Output:**  $\{\mu_{u,k,k',v}\}$ ,  $\{\alpha_{kk'}\}$

- 1: **repeat**
- 2:   **for**  $n = 1:N$  **do**
- 3:     Update  $(\hat{p}_{i,k,k'}, \check{p}_{i,k,k'})$  (Eq. 7)
- 4:     Update the intensity function  $\lambda_{s_i, d_i}(t_i)$  (Eq. 2)
- 5:   **end for**
- 6:   Update  $\hat{m}_{u,k,k',v}$  and  $\check{m}_{u,k,k',v}$  (Eq. 8)
- 7:   Update the base intensities  $\{\mu_{u,k,k',v}\}$  (Eq. 10)
- 8:   Update the parameters  $\{\beta_{kk'}\}$ ,  $\{\omega_{u,k,k',v}\}$ ,  $\{\psi_{k,k'}\}$  (Eqs. 13; 9; 12)
- 9:   Update the kernel parameters  $\{\alpha_{k,k'}\}$  (Eq. 11)
- 10: **until** convergence

---

Gibbs sampling algorithm, and present its full procedure in the supplement.

**Computational Cost.** For the second inference step, computing the latent variables  $\{z_i^s, z_i^d\}$  and updating the intensities for all the given events takes  $\mathcal{O}(NK^2)$  time, where  $K$  is the estimated number of communities by HGaP-EPM. Estimating  $\{\alpha_{k,k'}\}$  and  $\{\mu_{u,k,k',v}\}$  requires  $\mathcal{O}(K^2)$  and  $\mathcal{O}(K^2V^2)$  time, respectively. Estimating  $\{\beta_{k,k'}\}$  and  $\{\psi_{u,k,k',v}\}$  requires solving a linear system, and takes  $\mathcal{O}(K^2D^3)$  and  $\mathcal{O}(K^2\bar{N})$  time, where  $\bar{N}$  denotes the number of node pairs with at least one interaction in  $[0, T]$ . To sample the Pólya-Gamma variables  $\{\omega_{u,k,k',v}\}$ , we employed a fast and accurate approximate sampler of Zhou (2016), which matches the first two moments of the original distribution. Using the EM algorithm, the Pólya-Gamma variables are updated in closed-form (as a hyperbolic function) (Scott et al., 2013).

## 6 EXPERIMENTS

We evaluate the proposed Hawkes-EPM model on three benchmark datasets: (1) **Bosnia**. This dataset<sup>1</sup> consists of interaction events among 159 nations over 1,819 days (17/01/1991-31/12/1995). There are 1,918 edges, and 34,014 interactions. (2) **Gulf**. This dataset<sup>2</sup> contains 304,401 interaction events among 202 nations over 7,291 days (15/04/1979-31/03/1999). There are 7,184 edges. (3) **EU-email**. This dataset<sup>3</sup> consists of 332,334 email communications among 1,005 individuals over 526 days. There are 24,929 edges. We generated the covariate data between each pair of nodes using their common at-

<sup>1</sup><http://eventdata.parusanalytics.com/data.dir/pevehouse.html>.

<sup>2</sup><http://eventdata.parusanalytics.com/data.dir/gulf.html>.

<sup>3</sup><http://snap.stanford.edu/data/email-EuAll.html>.

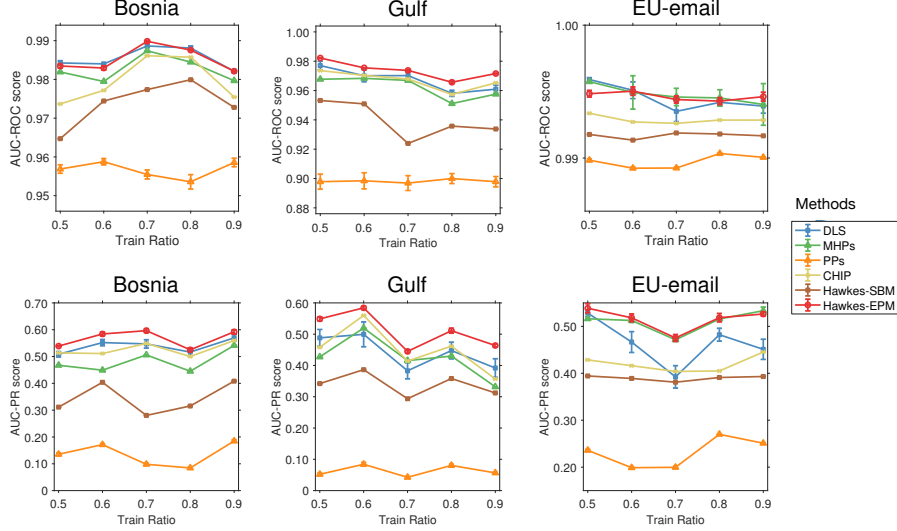


Figure 3: AUC-ROC and PR scores for the temporal link prediction.

tributes.

We compared our model to two basic models: (1) a Poisson process (PPs) model, which independently models the interaction dynamics between each pair of nodes by a constant event rate, (2) a mutually exciting Hawkes process (MHPs) model, in which we assume the same base rate and kernel parameters for each pair of nodes. Following (Yang et al., 2017), we utilized four basis kernels—three exponential kernels with time decaying scale: one hour, one day, one week respectively:  $\gamma^1(t) \equiv \exp(-24t)$ ,  $\gamma^2(t) \equiv \exp(-t)$ ,  $\gamma^3(t) \equiv \exp(-t/7)$ , and a periodic kernel  $\gamma^4(t) \equiv \exp(-t/7) \sin^2(\pi t/7)$ , and also to three state-of-the-art Hawkes interaction models: (3) the Hawkes Dual Latent Space (DLS) model (Yang et al., 2017), which explicitly captures the community structure via the base rate with the Latent space model (Hoff et al., 2001), and models the reciprocating dynamics between each pair of nodes via reciprocal latent space model, (4) the Hawkes stochastic block model (Hawkes-SBM), which captures the interaction dynamics using Hawkes process for each community independently, (5) the community Hawkes independent pairs model (CHIP), which models each node pair with a Hawkes process. For DLS, we set the latent dimensions  $d = 500$  according to the default setting of Yang et al. (2017) in our experiments. We demonstrate that the Hawkes-EPM achieves competitive performance but utilizes much fewer latent dimensions ( $K_{max} = 100$ ) compared to DLS. All the baseline models are detailed in the supplementary material. Due to lack of available code, we are not able to compare (Tan et al., 2018a,b). Given the aggregated graph, we estimated the parameters  $\{\Phi, \Omega\}$  of the HGaP-EPM with the truncation level  $K_{max} = 100$ . We ran the Gibbs sampler detailed

in (Zhou, 2015) for 10,000 MCMC iterations, and used the maximum a posterior estimate  $\{\hat{\Phi}, \hat{\Omega}\}$  in the second step. For the Hawkes-EPM, we choose a kernel decay of  $\delta = 1/10$  for the time scale of 10 days.

**Temporal link prediction.** To evaluate the predictive performance of all the methods, we sorted the interaction events according to the corresponding timestamps, and made a train-test split so that the training dataset consists of  $p$ -percent of the whole events with  $p$  varying between 50% and 90%. We trained all the methods using the training datasets. In this task, we let all the models to predict the probability that an edge appears (at least one interaction occurs) between each pair of nodes in the time interval  $[t, t + \hat{\pi})$  with  $t$  being the end time of the training events. We empirically set  $\hat{\pi}$  to be 50 days for all the datasets because it took one or two months for a nation to respond to actions from the other nations on average. We calculate the probability of there being at least one interaction in  $[t, t + \hat{\pi})$  as  $1 - \exp\{-\int_t^{t+\hat{\pi}} \lambda_{u,v}(s) dN_{v,u}(s)\}$ . Finally, we compute the average area under the curve (AUC) of both the receiver operating characteristic (ROC) and precision-recall (PR) to evaluate the predictive performance. Although AUC-ROC score is widely used in evaluating link prediction performance, we found that the interactions of the temporal events are quite sparse. Hence, one method can obtain high AUC-ROC score even if it accurately predicts zero-links but shows poor performance in predicting non-zero links. In contrast, AUC-PR score mainly reflects the method’s ability to predict non-zero links. As shown in Figure 3, the Hawkes process based models (MHPs, DLS, Hawkes-SBM, CHIP, Hawkes-EPM) capture the reciprocating dynamics of the interactions among nodes, and thus significantly outper-



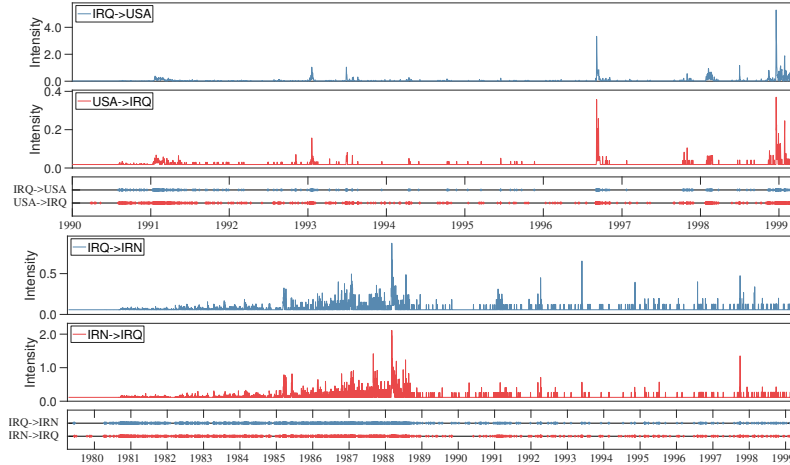


Figure 4: The top and bottom plot show the intensity of interaction events between USA-Iraq (IRQ), and Iran (IRN)-Iraq (IRQ) inferred by the Hawkes-EPM in the Gulf dataset, respectively.

form the Poisson process model. We noticed that most node pairs exhibit no edges in the time interval  $[t, t + \hat{\pi})$  (All the methods accurately predict zero-links and thus achieve high AUC-ROC scores). The Hawkes-SBM captures the interaction dynamics of each node pair within the same community only using a single point process, and thus achieves lower predictive scores. It is not surprising that mutually exciting Hawkes processes (MHPs) achieve higher scores as MHPs model each node pair with three periodic kernels, that sufficiently capture the interaction dynamics between each node pairs. A closer looking into the AUC-PR scores, shows that the Hawkes-EPM performs better than HPs, CHIP and DLS when the training ratio is low. This is because the Hawkes-EPM shares the kernel parameters among node pairs, and thus performs well even if most node pairs exhibit few interactions. Although both the Hawkes-EPM and DLS can capture the heterogeneity in base and reciprocal rate, the Hawkes-EPM effectively exploits the latent structure behind events and thus consistently outperforms DLS.

**Exploratory analysis.** We also used the Gulf dataset to explore the latent structure estimated by the Hawkes-EPM. We found that  $K = 12$  latent communities, and most of those communities correspond to international military conflicts among nations. Figure 4 shows the inferred intensities of the interaction between USA-Iraq (IRQ), and Iraq (IRQ)-Iran (IRN), respectively. We found that the peaks of the intensities correspond to events surrounding the Gulf War (1990-1991), the Cruise missile attack on Iraq on 1993 and 1996, the Bombing of Iraq in 1998. In addition, we also plot the intensities of interaction events between Iran (IRN)-Iraq(IRQ). The intensities of the interaction events between these two nodes are gradually increasing from 1980, and reach the peak at 1988. To interpret the inferred interaction dynam-

ics between these two nodes, we performed a web search, and found that the Iran-Iraq War started on September, 1980 and ended on August, 1988. Most of the inferred intensities between each pair of nations in the Gulf dataset confirm our knowledge of international affairs. We also provide the additional plots of the intensities between the other nations in the supplementary material.

## 7 CONCLUSIONS

We have presented a probabilistic framework, the Hawkes edge partition model (Hawkes-EPM) for inferring the implicit community structure and reciprocating dynamics among entities from their temporal interactions. The Hawkes-EPM not only models the inherent overlapping communities, sparsity and degree heterogeneity behind interactions, but also captures how the latent communities influence the interaction dynamics among their involved entities. Experimental results demonstrate the interpretability and competitive predictive performance of our model in temporal link prediction for several real-world datasets. Our strategy to cluster events into a set of latent patterns using the gamma process prior (Zhou and Carin, 2015) combined with Hawkes processes, can be readily generalized to all the closely-related applications, such as continuous-time topic models (Mavroforakis et al., 2017) and event-based tensor decomposition (Zhe et al., 2018).

### Acknowledgements

We thank the anonymous reviewers for the many useful comments that improved this manuscript. This research is funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement 668858.

## References

- Arastuie, M. et al. (2019). Consistent community detection in continuous-time networks of relational events. *ArXiv*, abs/1908.06940.
- Blundell, C. et al. (2012). Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems (NIPS) 25*, pages 2600–2608.
- Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*. Springer-Verlag, New-York, second edition.
- Du, N. et al. (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*, pages 219–228.
- DuBois, C. et al. (2013). Stochastic blockmodeling of relational event dynamics. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 238–246.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.
- Hoff, P. D. et al. (2001). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098.
- Junuthula, R. et al. (2019). The block point process model for continuous-time event-based dynamic networks. In *The World Wide Web Conference*, pages 829–839.
- Kapoor, J. et al. (2018). Bayesian nonparametric Hawkes processes. In *Bayesian Nonparametrics workshop at the 32nd Conference on Neural Information Processing Systems (NIPS)*.
- Lewis, E. A. et al. (2011). A nonparametric EM algorithm for multiscale Hawkes processes. *Journal of Nonparametric Statistics*.
- Linderman, S. et al. (2014). Discovering latent network structure in point process data. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 1413–1421.
- Mavroforakis, C. et al. (2017). Modeling the dynamics of online learning activity. In *26th International World Wide Web Conference*.
- Miscouridou, X. et al. (2018). Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2343–2352.
- Rai, P. et al. (2015). Large-scale Bayesian multi-label learning via topic-based label embeddings. In *Advances in Neural Information Processing Systems (NIPS) 28*, pages 3222–3230.
- Schein, A. et al. (2016). Bayesian Poisson Tucker decomposition for learning the structure of international relations. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 2810–2819.
- Schein, A. et al. (2019). Poisson-randomized gamma dynamical systems. In *Advances in Neural Information Processing Systems (NIPS) 32*, pages 782–793.
- Scott, J. G. et al. (2013). Expectation-Maximization for logistic regression. *arXiv preprint arXiv:1306.0040*.
- Tan, X. et al. (2018a). The Indian buffet Hawkes process to model evolving latent influences. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 795–804.
- Tan, X. et al. (2018b). Nested CRP with Hawkes-Gaussian processes. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1289–1298.
- Teh, Y. W. et al. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.
- Xu, H. et al. (2016). Learning Granger causality for Hawkes processes. In *Proceedings of The 33rd International Conference on Machine Learning (ICML)*, pages 1717–1726.
- Xu, H. et al. (2017). A Dirichlet mixture model of Hawkes processes for event sequence clustering. In *Advances in Neural Information Processing Systems (NIPS) 30*, pages 1354–1363.
- Yang, J. et al. (2017). Decoupling homophily and reciprocity with latent space network models. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence (UAI)*.
- Yang, S. et al. (2018). Dependent relational gamma process models for longitudinal networks. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 5551–5560.
- Zhe, S. et al. (2018). Stochastic nonparametric event-tensor decomposition. In *Advances in Neural Information Processing Systems (NIPS) 31*, pages 6856–6866.
- Zhou, K. et al. (2013). Learning triggering kernels for multi-dimensional Hawkes processes. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, pages 1301–1309.
- Zhou, M. (2015). Infinite edge partition models for overlapping community detection and link prediction. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1135–1143.
- Zhou, M. (2016). Softplus regressions and convex polytopes. *arXiv preprint arXiv:1608.06383*.
- Zhou, M. (2018a). Discussion on “Sparse graphs using exchangeable random measure” by Francois Caron and Emily B. Fox. *arXiv preprint arXiv:1802.07721*.
- Zhou, M. (2018b). Parsimonious Bayesian deep networks. In *Advances in Neural Information Processing Systems (NIPS) 31*, pages 3190–3200.
- Zhou, M. and Carin, L. (2015). Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):307–320.