

BRUNCH: Branching Structure Inference of Hybrid Multivariate Hawkes Processes with Application to Social Media

Hui Li¹, Hui Li², and Sourav S Bhowmick¹

¹ School of Computer Science & Engg, Nanyang Technological University, Singapore

² School of Cyber Engineering, Xidian University, China

HLI019@e.ntu.edu.sg, hli@xidian.edu.cn, assourav@ntu.edu.sg

Abstract. Multivariate Hawkes processes (MHPs) are a class of point processes where an arrival in one dimension can affect the future arrivals in all dimensions. Existing MHPs are associated with *homogeneous link functions*. However, in reality, different dimensions may exhibit different temporal characteristics. In this paper, we augment MHPs by incorporating *heterogeneous link functions*, referred to as *hybrid MHPs*, to capture the temporal characteristics in different dimensions. Since the *branching structure* can be utilized to equivalently represent MHPs, we propose a novel model called BRUNCH via *intensity-driven Chinese Restaurant Processes (intCRP)* to identify the optimal branching structure of hybrid MHPs. Furthermore, we relax the constraint on the shapes of triggering kernels in MHPs. We develop a *Monte Carlo-based inference* algorithm called MEDIA to infer the branching structure. Experiments on real-world datasets demonstrate the superior performance of BRUNCH and its usefulness in social media applications.

Keywords: Branching structure · Hawkes process · Heterogeneous link functions · Social media.

1 Introduction

Multivariate Hawkes processes (MHPs) are a class of point processes with mutually exciting components to model sequences of discrete events in continuous time, where an arrival in one dimension can affect future arrivals in all dimensions [5, 6]. Recently, MHPs have emerged in multiple fields to capture mutual excitation between dimensions, including high frequency trading [1], social influence analysis [16] and computational biology [13]. However, these MHPs are limited to a specific scenario where a past arrival can only excite the occurrence of future arrivals, and the corresponding *link functions*[‡] are linear (*i.e.*, linear MHPs). However, in reality, inhibitory arrivals and non-additive aggregation of effects from past arrivals are present in several application domains [10]. For instance, negative feedbacks of online consumers may inhibit others' purchasing behaviours. Consequently, MHPs associated with nonlinear link functions, namely *nonlinear*

[‡] *Link functions* describe the dynamics of the comprehensive effects from previous events.

MHPs [10], have been proposed where effects from past events encompass *both* excitation and inhibition.

Prior work on MHPs also assume that all dimensions take the *same* link function *i.e.*, either linear or nonlinear MHPs (*homogeneous* MHPs). That is, all dimensions follow roughly the same temporal characteristics. Note that a dimension may record actions of a user (*e.g.*, tweet, retweet, comment, share) in social media, behavior of a customer (*e.g.*, purchase, comment, return) in online shopping websites, and so on. In reality, different dimensions may exhibit different temporal characteristics. For example, in *Twitter*, one individual (*i.e.*, dimension) may be extremely interested in a popular topic and interact with her followers frequently whereas another individual may take little interest in that topic and seldom respond to his followers. In such scenario, the cumulative influence from recent events on the former individual is clearly different from the one on the latter. Hence, *homogeneous MHPs are insufficient to capture such diverse temporal characteristics*. To address this problem, in this paper we augment MHPs by incorporating *heterogeneous* link functions, referred to as *hybrid MHPs*, allowing us to cope with diverse impact of past events on future events in different dimensions.

The *cluster Poisson process interpretation* [7] of MHPs separates the events into two categories, namely, *immigrants* and *offspring*. The offspring events are triggered by past events, while the immigrants arrive independently and thus do not have an existing parent event. Offsprings are structured into *clusters* associated with each immigrant event. This is called the *branching structure* [8], which is an useful representation of MHPs in various applications. For example, in social influence analysis it can construct the narrative of information diffusion to pave the way for strategies to encourage or limit individual behaviors [14]. Additionally, the branching structure is widely utilized as a strategy in the maximum likelihood estimation of MHPs [16]. Unfortunately, such cluster Poisson process representation can only be applied to linear MHPs due to the mutually exciting assumption. Nonlinear MHPs cover both mutual excitation and mutual inhibition stochastically. Consequently, existing approaches based on the cluster Poisson process representation cannot be adopted to infer the branching structure of hybrid MHPs. In this paper, we *infer the branching structure* of hybrid MHPs regardless of the shapes of the *triggering kernel functions*[§].

We propose a novel probabilistic model called BRUNCH (**B**ranching **stR**ucture **iN**feren**Ce** of **H**ybrid multivariate Hawkes processes) to reveal the branching structure of hybrid MHPs *without assuming homogeneity of link functions or shapes of triggering kernel functions* (Section 3). It is important for our probabilistic model to emphasize the following two features mirrored by the event sequences in MHPs: (a) the chronological order of events is nonexchangeable; and (b) the triggering relations could distribute within or across dimensions stochastically. To this end, we propose *intensity-driven Chinese Restaurant Process (intCRP)*, a novel extension of classical CRP [3] in which the random seating assignment of the

[§] *Triggering kernel functions* describe the dynamics of how previous events trigger future events and may vary widely across different applications, *e.g.*, the triggering patterns in social media can be very different from the ones in high frequency trading.

Table 1. Key notations.

Notation	Definition	Notation	Definition
\mathbf{x}	event sequences	t_{ik}	k -th arrival in i -th dimension
X_i	i -th event sequence	$N_i(t)$	event number until t in i -th dimension
\mathcal{B}	collection of event links	$t_{il} \rightarrow t_{ik}$	event link from event t_{il} to event t_{ik}
\mathcal{C}	collection of cluster links	$s \rightarrow g$	cluster link from cluster s to cluster g
P_{ik}	events triggered by t_{ik}	$\mathcal{I}(\mathcal{B}, \mathcal{C})$	collection of cascades
Z_{ik}	parent event of t_{ik}	\mathcal{Z}_{ik}	the cascade the event t_{ik} belongs to

customers depends on the triggering kernels between them. In particular, intCRP has a nested structure – *inner* intCRP to explore the possible triggering relations among events occurring in one dimension (*i.e.*, *event links*), and *outer* intCRP to identify the collection of triggering relations between all events and their parents (if any) across dimensions (*i.e.*, *cluster links*). Obviously, the changes to the triggering relations within and across dimensions are highly coupled, *i.e.*, the inner intCRP and outer outCRP are strongly interlaced. Since there are countably infinite sets of triggering relations, we propose a novel inference approach called **MEDIA** (**MontE** Carlo-base**D** Inference **A**pproach) that leverages the triggering nature of MHPs to sample event links and cluster links alternatively (Section 4). Finally, we apply BRUNCH on real-world social media datasets, and the experimental study in Section 5 demonstrates its superior performance and usefulness. Formal algorithms and proofs of theorems and lemmas appear in [9]. List of key symbols used in this paper is given in Table 1.

2 Preliminaries

In this section, we introduce relevant concepts for understanding this paper.

2.1 Multivariate Hawkes Processes (MHPs)

The conditional intensity function of the i -th dimension for an M -dimensional MHP takes the following form [10]:

$$\lambda_i(t) = \mathcal{F}_i(\mu_i + \sum_{j=1}^M \sum_{t_{jl} < t} \alpha_{ij} \phi_{ij}(t - t_{jl})) \quad (2.1)$$

where $\mu_i > 0$ is the base intensity capturing the arrival rate of exogenous events independent of historical events. The term $\sum_{j=1}^M \sum_{t_{jl} < t} \alpha_{ij} \phi_{ij}(t - t_{jl})$ represents the accumulation of endogenous intensity caused by history [4]. The coefficient α_{ij} measures the influence from the j -th dimension to the i -th dimension, allowing mutual excitation ($\alpha_{ij} > 0$) and mutual inhibition ($\alpha_{ij} < 0$). The *triggering kernel function* $\phi_{ij}(t - t_{jl})$ quantifies the triggering effect from the event t_{jl} (*i.e.*, t_{jl} denotes the l -th arrival occurring in the j -th dimension) to the occurrence rate of the i -th dimension. Most of the existing work use predefined kernel functions with unknown parameters, such as the exponential kernels [16] and the power-law kernels [15]. The *link function* $\mathcal{F}_i : \mathcal{R} \rightarrow \mathcal{R}^+$, recognizes the triggering pattern of the i -th dimension over the historical events. The linear MHPs [6] is the case

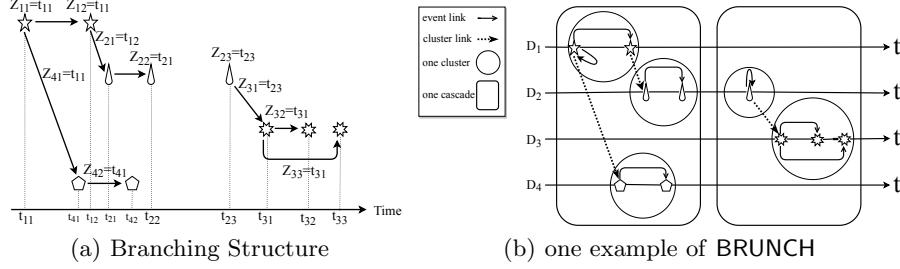


Figure 1. (a) An example of branching structure of MHPs and the events on the time axis; (b) Branching structure construction using BRUNCH.

$\mathcal{F}_i(x) = x$ with nonnegative α_{ij} for each dimension, while nonlinear MHPs apply various $\mathcal{F}_i(x)$ to guarantee the positive intensities. In *hybrid* MHPs, we allow each dimension to take a *personalized* link function to capture diverse temporal characteristics in real-world scenarios.

2.2 Branching Structure

Recall that the events in MHPs are classified as either *immigrants* or *offsprings*. An immigrant event arrives independently of other events, while an offspring event is triggered by a previous event. In the sequel, we refer to an immigrant together with its offsprings as a *cascade*. The collection of triggering relations in cascades is called the *branching structure* [8]. The cluster Poisson processes [7] could equivalently represent the branching structure of linear MHPs. Briefly, each immigrant starts one cascade, which consists of offspring events of the 1st, 2nd, 3rd, \dots generations, controlled by the endogenous intensity in Eq. 2.1 [11]. Due to the non-linear link functions, the above branching structure representation based on the cluster Poisson processes is inapplicable for nonlinear MHPs and hybrid MHPs.

Suppose that $X_i = \{t_{ik}\}_{k=1}^{N_i(t)}$ denotes the i -th event sequence during a time window $[0, t]$. Then $\mathbf{X} = \{X_i\}_{i \in [M]}$ is the collection of events from M dimensions during $[0, t]$. While modeling the sequences \mathbf{X} via hybrid MHPs, the corresponding *branching structure* could be represented by the variable set $\mathcal{K} = \{ \{ (Z_{ik}, P_{ik}) \}_{k=1}^{N_i(t)} \}_{i \in [M]}$ mathematically, such that

- $Z_{ik} = t_{ik}$ if t_{ik} is an immigrant; and $Z_{ik} = t_{jl}$ if event t_{jl} triggers t_{ik} ; and
- $P_{ik} = \{t_{hm}, t_{rs}, \dots\}$ if there are some events $\{t_{hm}, t_{rs}\}$ triggered by t_{ik} ; otherwise, $P_{ik} = \emptyset$.

Our objective is to infer (Z_{ik}, P_{ik}) for each event. Consider Figure 1(a). Directed links sketch an example of branching structure. $Z_{11} = t_{11}$ indicates the event t_{11} is an immigrant, and $Z_{12} = t_{11}$ shows t_{11} triggers t_{12} . $Z_{32} = t_{31}$ and $Z_{33} = t_{31}$ leads to $P_{31} = \{t_{32}, t_{33}\}$, denoting that t_{31} triggers t_{32} and t_{33} successively.

3 The BRUNCH Model

While modeling the asynchronous time-stamped event sequences via hybrid MHPs, we aim to reveal the underlying branching structure. Although the traditional

Chinese Restaurant Process (CRP) [3] provides a flexible class of distributions that is amenable for modeling dependencies between elements, the exchangeability assumption here is problematic for elements with temporal dependencies. This is because events in MHPs occur at different time points, and are nonexchangeable. In addition, in our problem setting the influence that determines the pairwise dependencies between events is not homogeneous within and across different sequences. Intuitively, it is natural to quantify the influence via the triggering kernels in Eq. 2.1. In order to tackle these issues, we present the BRUNCH model, which is based on *intensity-driven Chinese restaurant processes* (intCRP), a new variant of CRP that allows a number of intensity-driven distributions as priors on triggering relations between events.

The events in one cascade could stem from different dimensions. So, beyond the triggering relations obtained from single-sequence intCRP, we need to capture the cross-sequence triggering relations among events. To this end, BRUNCH allows us to identify the possible dependencies among multi-dimensional event sequences in hybrid MHPs. Specifically, it presents a class of prior distributions over branching structure according to intCRP, which has nested structure, *inner intCRP* and *outer intCRP*. Briefly, the inner intCRP identifies the possible triggering relations in each sequence independently, which are referred to as **event links**. Linked events in each sequence form one cluster. Subsequently, *outer intCRP* captures the potential cross-sequence triggering relations, referred to as **cluster links**, which connects parent events with children from cross-sequence *clusters*.

We resort to the Chinese Restaurant metaphor to describe the generative process of event links and cluster links in BRUNCH. Imagine a collection of event sequences as a collection of restaurants, and the events in each sequence as customers entering a restaurant. The linked events in each sequence compose one cluster, and such clusters correspond to tables. Figure 1(b) illustrates the process. Note that in traditional CRP, the probability of a customer sitting at a table is computed from the number of other customers already sitting at that table.

3.1 Event Link Construction

In each sequence, if one event is an immigrant, there exists one self-link with itself; otherwise, there exists a triggering dependency for the event. That is, if event t_{il} generates event t_{ik} , the link from t_{il} to t_{ik} is created spontaneously. The inner intCRP assigns event link $t_{il} \rightarrow t_{ik}$ in a biased way, according to the following cluster-specific distribution:

$$\Pr(t_{il} \rightarrow t_{ik} | \rho_i, A^i) \propto \begin{cases} A_{ik}^i & t_{ik} > t_{il} \\ \rho_i & t_{ik} = t_{il} \end{cases} \quad (3.1)$$

where self-affinity $\rho_i = \mu_i$ yields new immigrant events, and larger self-affinity favors more cascades. A_{ik}^i describes how the affinity between a pair of events affects the probability for t_{il} triggering t_{ik} . In accordance with the propagation characteristics described by hybrid MHPs, child events tend to be generated by preceding events with stronger triggering effect. We define A_{ik}^i as the product of two parts: $f^w(t_{il}, t_{ik})$ and $f^d(t_{il}, t_{ik})$. For a window size W , we set the decay function $f^w(t_{il}, t_{ik})$ such that for $t_{ik} - t_{il} < W$, $f^w(t_{il}, t_{ik}) = 1$, and zero otherwise. It determines the probability to link with events that are at most W timespan

away, and disregards the historical events as time progresses. Intuitively, the possible links existing between events that are far away from each other are negligibly rare. For $f^d(t_{il}, t_{ik})$, we apply the self-triggering kernel $f^d(t_{il}, t_{ik}) = \alpha_{ii}\phi_{ii}(t_{ik} - t_{il})$ which decays the probability of connecting events along with the timespan to the current one. Consequently, the event links $B_i = \{t_{il} \rightarrow t_{ik} | k = 1, 2, \dots, N_i(t), l \in \{1, 2, \dots, k\}\}$ in i -th dimensional sequence assign events into clusters, where two events are assigned to the same cluster if one is reachable from the other by traversing the directed links. Once the event link $t_{il} \rightarrow t_{ik}$ is confirmed, we can obtain the branching structure $Z_{ik} = t_{il}$ within sequences.

3.2 Cluster Link Construction

Accordingly, the collection of event links $\mathcal{B} = \{B_1, B_2, \dots, B_M\}$ will divide the event sequences \mathbf{X} into clusters, denoted by $\mathcal{C}(\mathcal{B})$. By involving the mutual-triggering kernels, we could measure the pairwise affinity between cross-sequence events. Hence, given two clusters, s and g , the outer intCRP assigns the cluster link $s \rightarrow g$ according to the following cascade-specific distribution:

$$\Pr(s \rightarrow g | \mathcal{B}) \propto \max_{t_{ik} \in s, t_{ik} < t_{je}} (|\alpha_{ij} \phi_{ij}(t_{je} - t_{ik})|) \quad g \neq s \quad (3.2)$$

where $t_{je} = \min\{t_{jl} | t_{jl} \in g\}$. Only if one event (*i.e.*, t_{ik}) in cluster s generates the earliest event in cluster g , there exists a cluster link $s \rightarrow g$. In particular, self-loop cluster link is non-existent. Once the cluster link $s \rightarrow g$ is determined, we could construct the equivalent branching structure $P_{ik} = P_{ik} \cup \{t_{je}\}$. In summary, we construct the complete branching structure $\mathcal{K} = \{(Z_{ik}, P_{ik})\}$ via scanning the obtained event links \mathcal{B} and cluster links \mathcal{C} .

Intuitively, a collection of cluster links will divide clusters into cascades. We represent one cascade as one set of events. Let \mathcal{Z}_{ik} denote the cascade associated with event t_{ik} , and $\mathcal{Z} = \{\mathcal{Z}_{ik} | i = 1, 2, \dots, M; k = 1, 2, \dots, N_i\}$ records the final cascade assignments. Notice that events t_{ik} and t_{jl} belong to one cascade (*i.e.*, $\mathcal{Z}_{ik} = \mathcal{Z}_{jl}$) if and only if they are reachable via combinations of event links and cluster links. Given the collection of event links \mathcal{B} and cluster links \mathcal{C} , we denote $\mathcal{I}(\mathcal{B}, \mathcal{C})$ as the final collection of cascades. We initialize the Hawkes likelihood parameters $\Theta^{1:M} = \{\mu_i, \alpha_{ij}, \phi_{ij}\}_{i,j \in [M]} \sim \pi(\gamma)$ where γ are the hyper-parameters, and $\pi(\gamma)$ is the collection of distributions.

The central goal of BRUNCH is to infer the posterior distribution of the latent links $(\mathcal{B}, \mathcal{C})$, given a collection of time-stamped events. It places a prior distribution over a combinatorial number of possible event links and cluster links, according to inner intCRP (Eq. 3.1) and outer intCRP (Eq. 3.2), respectively. Intuitively, applying Bayes's Theorem, the posterior distribution takes the form $\Pr(\mathcal{B}, \mathcal{C} | \mathbf{X}) = \frac{\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C})}{\Pr(\mathbf{X})}$. Unfortunately, we cannot compute $\Pr(\mathbf{X})$. Hence, the posterior inference of links is intractable. To address this, in the following section we present a strategy that approximately infers the posterior distribution.

4 Model Inference

As the number of event links and cluster links varies with the observed events, we need to undertake Bayesian inference over a link set of unknown cardinality.

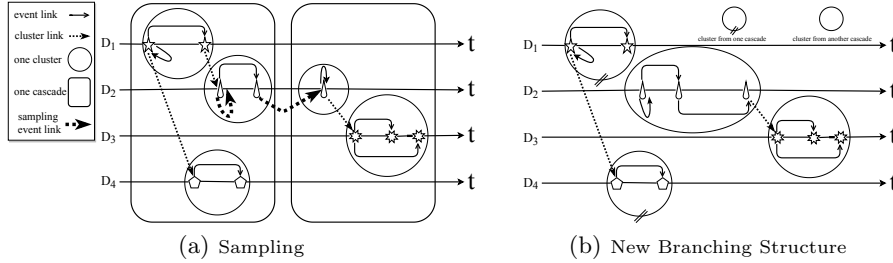


Figure 2. Sampling event links in Panel (a) leads a chain of changes to current branching structure, and constructs new one in Panel (b).

Moreover, changes over event links may induce subsequent changes to other event links and current cluster links. To this end, we propose the Monte Carlo-based inference approach (MEDIA) that leverages the triggering nature of hybrid MHPs to sample event links and cluster links.

We aim to construct a Markov chain whose stationary distribution is the target posterior distribution $\Pr(\mathcal{B}, \mathcal{C} | \mathbf{X})$. The state of the chain is represented by $(\mathcal{B}, \mathcal{C})$, a collection of links over event sequences. Furthermore, event links \mathcal{B} and cluster links \mathcal{C} are strongly coupled, that is, sampling event links could trigger a chain of merges and splits to the current structure, as shown in Figure 2. In view of this, we design the Monte Carlo-based inference approach, which involves two key phases: (a) sampling event links \mathcal{B} via a Metropolis-Hastings rule, which could possibly bring changes to current cluster links \mathcal{C} , and then (b) update cluster links \mathcal{C} via a Gibbs sampler. We elaborate on them in turn.

Sampling event links. Let current links be $(\mathcal{B}, \mathcal{C})$. After sampling one event link, the reconstructed links are denoted as $(\tilde{\mathcal{B}}, \tilde{\mathcal{C}})$. The iterative procedure runs until approaching the stationary distribution as follows:

1. Using the current links \mathcal{B} , sample a candidate link set $\tilde{\mathcal{B}}$ from the transition probability $q(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})$;
2. Sampled $\tilde{\mathcal{B}}$ leads to cluster links $\tilde{\mathcal{C}}$. Calculate the acceptance probability $\eta(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})$ for the candidate set $\tilde{\mathcal{B}}$,

$$\eta(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C}) = \min \left\{ \frac{\Pr(\mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) q(\mathcal{B}, \mathcal{C} | \mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})}{\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C}) q(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})}, 1 \right\}$$

Notice that we are considering the ratio of $\Pr(\mathcal{B}, \mathcal{C} | \mathbf{X})$ under two different structures, so the denominator $\Pr(\mathbf{X})$ is eliminated.

Theorem 4.1. *The above Metropolis-Hastings rule satisfies detailed balance.*

Based on Theorem 4.1, we can guarantee the resulting Markov chain converges to a stationary distribution uniquely [2]. BRUNCH provides a joint distribution for a collection of events and current links as following:

$$\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C}) = \Pr(\mathcal{B}) \Pr(\mathcal{C} | \mathcal{B}) \Pr(\mathbf{X} | \mathcal{B}, \mathcal{C}) \quad (4.1)$$

If one event is an offspring, it has only one parent event. So once the affinity functions in inner intCRP are predefined, the generated event links are conditionally independent. Furthermore, event links divide all events into clusters. Hence, once the event links and the affinity functions in outer intCRP are predefined, the generated cluster links are conditionally independent. That is to say, event

links \mathcal{B} are conditionally independent given $(\rho_{1:M}, A^{1:M})$, and cluster links \mathcal{C} are conditionally independent given \mathcal{B} , thereby causing the independent cascades. As a consequence, the joint distribution on events and links equals to:

$$\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C} | \rho_{1:M}, A^{1:M}, \gamma) = \prod_{I \in \mathcal{I}(\mathcal{B}, \mathcal{C})} \Pr(t_{Z=I} | \gamma) \prod_{i=1}^M \prod_{k=1}^{N_i} \prod_{l=1}^k \Pr(t_{il} \rightarrow t_{ik} | \rho_i, A^i) \prod_{t_{ik} \in \mathcal{C}(\mathcal{B})} \Pr(s \rightarrow g | \mathcal{B}) \quad (4.2)$$

The activities belonging to cascade I are represented as $t_{Z=I}$, hence,

$$\Pr(t_{Z=I} | \gamma) = \int \prod_{\mathcal{Z}_{ik}=I} \Pr(t_{ik} | \Theta^{1:M}) d\pi(\Theta^{1:M} | \gamma) \quad (4.3)$$

where the parameter set Θ could be drawn from pre-determined distributions associated with the hyper-parameters γ . Hence, the above integral is tractable. Based on the hybrid MHPs associated with conditional intensity (Eq. 2.1), we derive the conditional probability density [8] that an event occurs at time t_{ik} is:

$$\Pr(t_{ik} | \mathcal{Z}_{ik} = I, \Theta^{1:M}) = \lambda_i(t_{ik} | \mathcal{Z}_{ik} = I, \Theta^{1:M}) \cdot \exp\left(-\int_0^{t_{ik}} \lambda_i(s | \Theta^{1:M}) ds\right) \quad (4.4)$$

where the integral $\int_0^{t_{ik}} \lambda_i(s | \Theta^{1:M}) ds$ is not always analytically integrable w.r.t. various link functions of hybrid MHPs. To address the issue, we propose the approximation: $\int_0^t \lambda_i(s | \Theta^{1:M}) ds = \sum_{m=1}^n (t_m - t_{m-1}) \lambda_i(t_m)$.

We describe how replacing an event link affects the current links $(\mathcal{B}, \mathcal{C})$, and there are four cases:

1. Split. If adding event link $t_{ik} \rightarrow t_{ik}$ breaks the existing link $t_{il} \rightarrow t_{ik}$ and divides t_{ik} and t_{il} into two clusters, it shows that event t_{ik} is an offspring. Moreover, there are no existing clusters to be merged after sampling event link $t_{ik} \rightarrow t_{il'}$. Thus, the new cluster including t_{il} has the same incoming cluster links as the previous cluster containing both t_{ik} and t_{il} . We assume that the incoming links are independently linked to the new cluster with equal probability. New cluster s including t_{ik} collects its outgoing cluster links \mathcal{C}_s according to the distribution $\Pr(\mathcal{C}_s | \mathbf{X}, \tilde{\mathcal{B}}, \mathcal{C}_{-s})$ where \mathcal{C}_{-s} represents the set of cluster links excluding the ones $s \rightarrow g, g \in \mathcal{C}(\tilde{\mathcal{B}})$. Then, the transition probability is:

$$q_1(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C}) = \Pr(t_{ik} \rightarrow t_{il'} | \rho^i, A^i) \cdot 0.5^{|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il}}|} \Pr(\mathcal{C}_s | \mathbf{X}, \tilde{\mathcal{B}}, \mathcal{C}_{-s}) \quad (4.5)$$

where $|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il}}|$ records the number of incoming cluster links for the old cluster containing both t_{ik} and t_{il} . Calculate $\Pr(t_{ik} \rightarrow t_{il'} | \rho^i, A^i)$ according to Eq. 3.1.

2. Split and Merge. Adding event link $t_{ik} \rightarrow t_{ik}$ breaks existing link $t_{il} \rightarrow t_{ik}$. Moreover, the cluster including event t_{ik} and the cluster including event $t_{il'}$ are merged after sampling event link $t_{ik} \rightarrow t_{il'}$. Thus, the outgoing cluster links of new cluster retain the outgoing cluster links of the old cluster including $t_{il'}$, and the incoming cluster links of new cluster combine the incoming cluster links connecting to the cluster including t_{ik} and the cluster including $t_{il'}$. The transition probability is:

$$q_2(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C}) = q(t_{ik} \rightarrow t_{il'}) \cdot 0.5^{|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il}}|} \quad (4.6)$$

Also, the incoming cluster links are assumed to be assigned to the new merged cluster equally. When the sampling link $t_{ik} \rightarrow t_{il'}$ merges two clusters from one cascade, $q(t_{ik} \rightarrow t_{il'}) = \Pr(t_{ik} \rightarrow t_{il'} | \rho^i, A^i)$. Otherwise, when $t_{ik} \rightarrow t_{il'}$ merges two clusters from different cascades, it combines the two cascades. Hence,

$$q(t_{ik} \rightarrow t_{il'}) = \Pr(t_{ik} \rightarrow t_{il'} | \rho^i, A^i) \cdot \frac{\Pr(\mathbf{X}_{\mathcal{Z}^1=\mathcal{Z}_{ik}} \cup \mathbf{X}_{\mathcal{Z}^2=\mathcal{Z}_{il'}} | \mathcal{Z}_{ik} \neq \mathcal{Z}_{il'}, \gamma)}{W_1 \cdot W_2} \quad (4.7)$$

wherein $W_1 = \Pr(\mathbf{X}_{\mathcal{Z}^1=\mathcal{Z}_{ik}} | \mathcal{Z}_{ik} \neq \mathcal{Z}_{il'}, \gamma)$, and $\mathbf{X}_{\mathcal{Z}^1=\mathcal{Z}_{ik}}$ represents the events in the cascade \mathcal{Z}_{ik} excluding the events in the cascade $\mathcal{Z}_{il'}$. Similarly, $W_2 = \Pr(\mathbf{X}_{\mathcal{Z}^2=\mathcal{Z}_{il'}} | \mathcal{Z}_{ik} \neq \mathcal{Z}_{il'}, \gamma)$, and $\mathbf{X}_{\mathcal{Z}^2=\mathcal{Z}_{il'}}$ denotes the events in the cascade $\mathcal{Z}_{il'}$ excluding the events in the cascade \mathcal{Z}_{ik} .

3. Merge. After adding event link $t_{ik} \rightarrow t_{il'}$, if there is no new cluster to appear, it shows that event t_{ik} is an immigrant. Moreover, sampling event link $t_{ik} \rightarrow t_{il'}$ causes two existing clusters to be merged. Hence, the transition probability is:

$$q_3(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C}) = q(t_{ik} \rightarrow t_{il'}) \quad (4.8)$$

Considering the two merged clusters come from one cascade or two cascades, $q(t_{ik} \rightarrow t_{il'})$ is analogous to the analysis of $q_2(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})$.

4. No Change. There is no new cluster after setting $t_{ik} \rightarrow t_{il'}$, and also no merge after sampling event link $t_{ik} \rightarrow t_{il'}$. In this case, the corresponding transition probability is: $q_4(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C}) = \Pr(t_{ik} \rightarrow t_{il'} | \rho^i, A^i)$.

Obviously, after sampling each event link, the resulting split (*i.e.*, case 1) and merge (*i.e.*, case 3) are inverse of each other, meanwhile, the other two cases (*i.e.*, split and merge vs. no change) are inverse transform. Accordingly, by taking the inverse pairs, we derive the corresponding acceptance ratios $\eta(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C}) = \min\{\tau, 1\}$, where the Hastings ratio $\tau = \frac{\Pr(\mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) q(\mathcal{B}, \mathcal{C} | \mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})}{\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C}) q(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})}$.

As aforementioned, sampling an event link may lead four possible changes to current links $(\mathcal{B}, \mathcal{C})$. Hence, the corresponding Hastings ratio is:

* A single offspring event becomes an immigrant and the previous cluster is split into two clusters. Thus the candidate partition structure $(\tilde{\mathcal{B}}, \tilde{\mathcal{C}})$ is generated. The transitions corresponding to q_1 and q_3 are the inverse of each other. Substituting Eq. 4.1, 4.5 and 4.8, the Hastings ratio works out to be

$$\tau_1 = \frac{\Pr(\mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) q_3(\mathcal{B}, \mathcal{C} | \mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})}{\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C}) q_1(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})} = \frac{1}{0.5^{|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il'}}|}} \frac{\Pr(\mathcal{C}_{-s} | \tilde{\mathcal{B}})}{\Pr(\mathcal{C} | \mathcal{B})} \frac{\Pr(\mathbf{X} | \tilde{\mathcal{B}}, \mathcal{C}_{-s})}{\Pr(\mathbf{X} | \mathcal{B}, \mathcal{C})} \quad (4.9)$$

When the offspring event changes its parent event from one cluster to another cluster under the condition that the two clusters locate in different cascades, the reverse transition q_3 leads to the merger of two cascades. Consequently, the transition probability is:

$$q_3(\mathcal{B}, \mathcal{C} | \mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) = \Pr(t_{ik} \rightarrow t_{il'} | \rho^i, A^i) \cdot \frac{\Pr(\mathbf{X}_{\mathcal{Z}^1=\mathcal{Z}_{ik}} \cup \mathbf{X}_{\mathcal{Z}^2=\mathcal{Z}_{il'}} | \mathcal{C}_{t_{ik}} \neq \mathcal{C}_{t_{il'}}, \gamma)}{W_1 \cdot W_2}$$

$$W_1 = \Pr(\mathbf{X}_{\mathcal{Z}^1=\mathcal{Z}_{ik}} | \mathcal{Z}_{ik} \neq \mathcal{Z}_{il'}, \gamma) \quad W_2 = \Pr(\mathbf{X}_{\mathcal{Z}^2=\mathcal{Z}_{il'}} | \mathcal{Z}_{ik} \neq \mathcal{Z}_{il'}, \gamma)$$

Similarly, $q_1(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C}) = \Pr(t_{ik} \rightarrow t_{il'} | \rho^i, A^i) \cdot 0.5^{|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il'}}|} \Pr(\mathcal{C}_s | \mathbf{X}, \tilde{\mathcal{B}}, \mathcal{C}_{-s}, \rho)$.

The corresponding transition probability becomes: $\tau_1 = \frac{1}{0.5^{|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il'}}|}} \frac{\Pr(\mathcal{C}_{-s} | \tilde{\mathcal{B}})}{\Pr(\mathcal{C} | \mathcal{B})}$

* A single offspring event switches to an immigrant, and sampling a new event link leads to the combination of two local clusters. In this case, the transitions corresponding to q_2 and q_4 are the inverse of each other. If the two merged clusters come from one cascade,

$$\tau_2 = \frac{\Pr(\mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) q_4(\mathcal{B}, \mathcal{C} | \mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})}{\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C}) q_2(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})} = \frac{1}{0.5^{|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il}}|}} \frac{\Pr(\tilde{\mathcal{C}} | \tilde{\mathcal{B}}) \Pr(\mathbf{X} | \tilde{\mathcal{B}}, \tilde{\mathcal{C}})}{\Pr(\mathcal{C} | \mathcal{B}) \Pr(\mathbf{X} | \mathcal{B}, \mathcal{C})} \quad (4.10)$$

If the sampling new event link causes two cascades to be merged, we obtain

$$\tau_2 = \frac{1}{0.5^{|\mathcal{C}_{t_{ik}} \cup \mathcal{C}_{t_{il}}|}} \frac{\Pr(\tilde{\mathcal{C}} | \tilde{\mathcal{B}})}{\Pr(\mathcal{C} | \mathcal{B})}.$$

* Sampling $\mathcal{C}_{t_{il'}} = t_{ik}$ leads the immigrant event t_{ik} to trigger new offspring.

Similar to τ_1 , we can calculate the Hastings ratios: $\tau_3 = \frac{\Pr(\mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) q_1(\mathcal{B}, \mathcal{C} | \mathbf{X}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})}{\Pr(\mathbf{X}, \mathcal{B}, \mathcal{C}) q_3(\tilde{\mathcal{B}}, \tilde{\mathcal{C}} | \mathbf{X}, \mathcal{B}, \mathcal{C})}$.

* Samplings which change event links but cause no change to the cluster links always have $\tau_4 = 1$.

Sampling cluster links. Once the event links \mathcal{B} are sampled via the aforementioned Metropolis-Hasting rule, we update the cluster links \mathcal{C} again via Gibbs sampler as follows: i) Scan each cluster $s \in \mathbf{C}(\mathcal{B})$; ii) Draw $\mathcal{C}_s \sim \Pr(\mathcal{C}_s | \mathbf{X}, \mathcal{B}, \mathcal{C}_{-s})$. Repeat the above steps until convergence. \mathcal{C}_s represents the outgoing cluster links for cluster s , and \mathcal{C}_{-s} records the set of cluster links excluding the ones $s \rightarrow g, g \in \mathcal{C}(\mathcal{B})$. We calculate $\Pr(\mathcal{C}_s | \mathbf{X}, \mathcal{B}, \mathcal{C}_{-s})$ according to Eq. 3.2.

We keep all event links in an adjacency matrix $S \in R^{n \times n}$ ($n = \sum_{i=1}^M N_i(t)$), wherein rows and columns are indexed by ordered events from all sequences, value 1 or 0 is recorded in entry (t_{ik}, t_{jl}) according to whether t_{ik} triggers t_{jl} or not. The formal description of MEDIA, is given in [9]. The time complexity of MEDIA is $\mathcal{O}(LM^3 n_a^2 n_c)$ [9].

5 Experiments

In this section, we investigate the performance of our model and inference algorithm and report the key results. All experiments are performed on a machine with 16GB RAM with Intel(R) Core(TM) E5-1620V2 CPU@3.70 GHz processor running on Windows 8.1 Pro.

Competitors. Recall that there is no existing work that infers the branching structure of nonlinear MHPs. Hence, we are confined to compare our proposed frameworks to techniques that infer the branching structure of linear MHPs in [12, 17]. In particular, we consider the following strategies for our study. (a) Cluster-L: Based on the alternative representation of the linear Hawkes process in terms of cluster Poisson processes, [12, 17] propose the cluster-based method. (b) MEDIA-L: Use MEDIA to infer the branching structure of linear MHPs. (c) MEDIA-E: Apply MEDIA to infer the branching structure of exponential MHPs (*i.e.*, the corresponding link function $\mathcal{F}_i(x) = e^x$). (d) MEDIA-H: The odd dimensions of MHPs take linear link function, and the even dimensions adopt exponential one. Then we use MEDIA to infer the branching structure of hybrid MHPs.

As mentioned earlier, BRUNCH is not sensitive to shapes of the triggering kernel functions in hybrid MHPs. Hence, we adopt an exponential kernel function $\phi_{ij}(t) = \exp(-\beta_{ij}t)$ ($\beta_{ij} > 0$) in the experiments (see [9] for the performance associated with other kernel functions). For each dimension, the hyper-parameter γ is sampled by a uniform distribution $\mathcal{U}(0, 10)$. The base intensity μ is set varying over dimensions and is sampled from $\mathcal{U}(0, \gamma)$, then the coefficient α_{ij} is sampled from $\mathcal{N}(0, \gamma^2)$, and the decay parameter β_{ij} has the form of $\beta_{ij} = c * \alpha_{ij}$ where c is sampled from $\mathcal{U}(0, \gamma)$. The initial parameters $\Theta^{1:M} = (\mu, \alpha, \phi)$ need

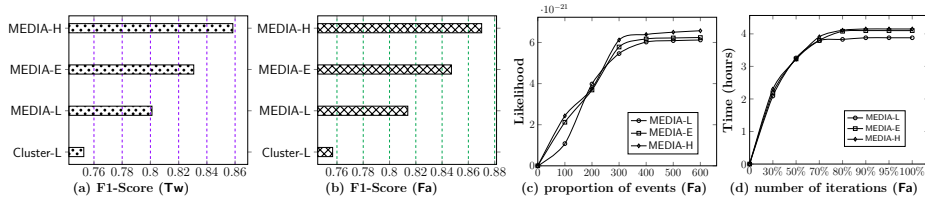


Figure 3. Experimental results.

satisfy the stability and uniqueness conditions (see details in [9]). Additionally, we set the time window size to $W = 12$ hours.

Datasets and ground truth. We fit the aforementioned models on two real-world social media datasets: (1) *Facebook* (Fa): 43,679,231 events from 109,211 individuals during March 2018 to May 2018; (2) *Twitter* (Tw): 51,622,139 events from 123,972 individuals from March 2018 to May 2018. While crawling the triggering relations of events (*i.e.*, which event triggers which events) as ground truth via Facebook Graph API and Twitter Streaming API, we crawl the social network structure in advance. For each immigrant event in each individual’s sequence, we grab the triggered events (*i.e.*, offsprings) starting from the individual’s followers via a depth-first search algorithm. Then, while modeling the observed timestamped events via MHPs, we aim to infer the branching structure *without* the knowledge of social network structure.

Inferring branching structure. We convert the branching structure to a binary matrix S as mentioned in Section 4. Given the estimated links $(\tilde{\mathcal{B}}, \tilde{\mathcal{C}})$, we update \tilde{S} according to $\tilde{\mathcal{B}}$, and then derive the across-dimensional branching structure \mathcal{K} from $\tilde{\mathcal{C}}$ before filling in the final \tilde{S} . Afterwards, comparing the estimated matrix \tilde{S} with the ground truth S , we evaluate the effectiveness for all the aforementioned strategies in terms of *F1-Score*.

Figures 3(a) and 3(b) plot the results. Clearly, our proposed model BRUNCH with MEDIA outperforms the baseline Cluster-L, obtaining higher inference *F1-Score*. That is, the triggering relations among events (*i.e.*, branching structure of MHPs) identified by the MEDIA approach are more reliable. While applying the same inferring procedure MEDIA, hybrid MHPs (mixing exponential MHPs and linear MHPs) show superior inference performance compared to other alternatives. In particular, MEDIA-H is superior to MEDIA-L and MEDIA-E. This further verifies the justifiability of our proposed hybrid MHPs.

Convergence. We compare the convergence rate of our proposed techniques in Figure 3(c) on *Facebook* data. The results on *Twitter* are qualitatively similar (see [9]). Clearly, MEDIA-H is of higher likelihood than MEDIA-L and MEDIA-E. This further validates the usefulness of our hybrid MHPs.

Scalability. Figures 3(d) plots the scalability of our algorithms with increasing events on *Facebook* data. The results are qualitatively similar on *Twitter* (see [9]). We run the inference methods on different sizes of datasets (*i.e.*, slice different percentages of events in datasets as input data for BRUNCH). Observe that the average inference time of MEDIA stabilizes with increasing number of events. Since the number of event links and cluster links grows significantly as events increases, we expect the average runtime of MEDIA becomes relatively stable when more than 70% input data are utilized.

6 Conclusions

In this paper, we propose a novel probabilistic model called BRUNCH to infer the branching structure of hybrid MHPs. It bridges a significant chasm between hybrid MHPs (nonlinear MHPs as well) and branching structure inference. We handle the inferencing procedure via the MEDIA method, which provides a heuristic to make coordinated changes to both event links within clusters and cluster links within cascades. Empirically, our model demonstrates good performance and application potential in the real world.

Acknowledgments. This work is partly supported by the National Natural Science Foundation of China (No. 61672408, 61972309) and National Engineering Laboratory of China for Public Safety Risk Perception and Control by Big Data (PSRPC).

References

1. Bacry, E., Muzy, J.F.: Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance* (2014)
2. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer (2006)
3. Blei, D.M., Frazier, P.I.: Distance dependent chinese restaurant processes. *JMLR* (2011)
4. Farajtabar, M., Du, N., Rodriguez, M.G., Valera, I., Zha, H., Song, L.: Shaping social activity by incentivizing users. In: *NIPS* (2014)
5. Hawkes, A.G.: Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)* (1971)
6. Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. *Biometrika* (1971)
7. Hawkes, A.G., Oakes, D.: A cluster process representation of a self-exciting process. *Journal of Applied Probability* (1974)
8. Lee, Y., Lim, K.W., Ong, C.S.: Hawkes processes with stochastic excitations. In: *ICML* (2016)
9. Li, H., Li, H., Bhowmick, S.S.: Brunch: Branching structure inference of hybrid multivariate hawkes processes with application to social media. Technical report, <http://www.ntu.edu.sg/home/assourav/TechReports/BRUNCH-TR.pdf>
10. Mei, H., Eisner, J.M.: The neural hawkes process: A neurally self-modulating multivariate point process. In: *NIPS* (2017)
11. Møller, J., Rasmussen, J.G.: Perfect simulation of hawkes processes. *Advances in applied probability* (2005)
12. Rasmussen, J.G.: Bayesian inference for hawkes processes. *Methodology and Computing in Applied Probability* (2013)
13. Reynaud-Bouret, P., Rivoirard, V., Grammont, F., Tuleau-Malot, C.: Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience* (2014)
14. Wang, Y., Theodorou, E., Verma, A., Song, L.: A stochastic differential equation framework for guiding online user activities in closed loop. In: *AISTATS* (2018)
15. Zhang, C.: Modeling high frequency data using hawkes processes with power-law kernels. *Procedia Computer Science* (2016)
16. Zhou, K., Zha, H., Song, L.: Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In: *AISTATS* (2013)
17. Zhou, K., Zha, H., Song, L.: Learning triggering kernels for multi-dimensional hawkes processes. In: *ICML* (2013)