

# Dyadic Event Attribution in Social Networks with Mixtures of Hawkes Processes

Liangda Li and Hongyuan Zha

College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30032  
lqli, zha}@cc.gatech.edu

## ABSTRACT

In many applications in social network analysis, it is important to model the interactions and infer the influence between pairs of actors, leading to the problem of dyadic event modeling which has attracted increasing interests recently. In this paper we focus on the problem of dyadic event attribution, an important missing data problem in dyadic event modeling where one needs to infer the missing actor-pairs of a subset of dyadic events based on their observed timestamps. Existing works either use fixed model parameters and heuristic rules for event attribution, or assume the dyadic events across actor-pairs are independent. To address those shortcomings we propose a probabilistic model based on mixtures of Hawkes processes that simultaneously tackles event attribution and network parameter inference, taking into consideration the dependency among dyadic events that share at least one actor. We also investigate using additive models to incorporate regularization to avoid overfitting. Our experiments on both synthetic and real-world data sets on international armed conflicts suggest that the proposed new method is capable of significantly improve accuracy when compared with the state-of-the-art for dyadic event attribution.

## Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning; J.4 [Computer Applications]: Social and Behavioral Sciences

## General Terms

Algorithms, Experimentation, Performance

## Keywords

Dyadic event, missing data problem, Hawkes process, variational inference, international armed conflicts

## 1. INTRODUCTION

Analyzing dyadic event data using temporal point processes has attracted much of recent research interests in the context of information diffusion and prediction for social networks [6, 17]. Dyadic

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

CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.

Copyright 2013 ACM 978-1-4503-2263-8/13/10 ...\$15.00.

<http://>—enter the whole DOI string from rightsreview form confirmation.

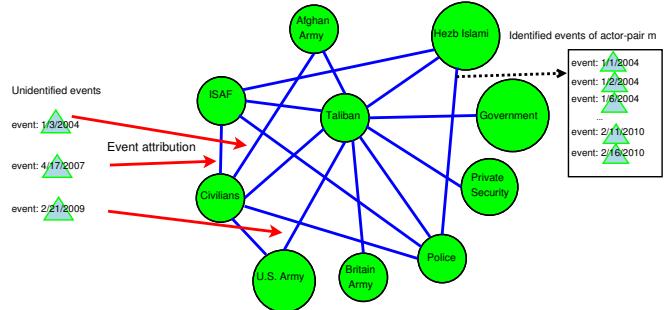


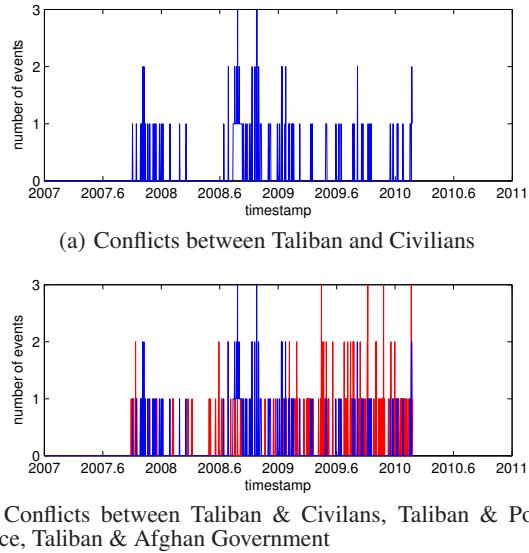
Figure 1: Illustration of the Dyadic Event Attribution Problem

events refer to the timestamped interactions involving pairs of actors, e.g., one user sending an email message to another, one user retweeting a post of a particular celebrity, or the Taliban force attacking civilian. Dyadic event data arise in a wide range of social network applications, such as communication studies, social media, crime prevention, and health care. The analysis of dyadic events is generally much more complicated than the analysis of events with single actors [9]: particular consideration needs to be given to the interdependency among events and to the relationship among actor-pairs [10]. For instance, an attack on the Afghanistan army by the Taliban may incur counter-attacks from Afghanistan army's allies, resulting in conflicts between U.S. army and Taliban or those between British army and Taliban. Such interdependency resulting from the underlying reciprocity or mutuality makes dyadic event data modeling particularly useful and challenging.

The focus of this paper is an important missing data problem in dyadic event data modeling: Because of the difficulty and uncertainty in data collection, the dyadic data collected are usually incomplete, and it is generally the case that we have the timestamps of the events but we do not observe the actors of the events [16]. For example, we may observe an event with civilian casualty but we did not observe who carried out the act. Given a set of dyadic events with some of the events missing one or two actors, we call the inference problem of estimating the missing actors the *dyadic event attribution problem* (DEAP). In this paper, we will develop temporal point process models to tackle DEAP and the key to its solution is the exploration of the clustering and self- and mutual-excitation properties of the events. The solution of DEAP will enhance our understanding of dyadic event dynamics and it is also useful for down-stream applications such as dyadic event classification and visualization.

Before we proceed to technical discussions, we use a real-world application, the Armed Conflict Location and Event Data (ACLED),

to illustrate DEAP. ACLED is a very comprehensive public collection of political violence data for developing countries [1, 15]. ACLED contains several data sets categorized by geographic regions, such as Africa and several Asian states. In the Afghanistan data set, for example, 36.3% of the events are without the actor-pair information which will be called unidentified events, while for the African dataset, unidentified events account for 16.9%. DEAP is then a problem of attributing the actor-pairs to those unidentified events (Figure 1), i.e., inferring the actor-pairs of those unidentified events based on their timestamps as well as the timestamps and actor-pairs of those identified events.



**Figure 2: Sequence of dyadic events recoding conflicts during the period 2007-2011: Blue line denotes the conflicts between Taliban and civilians, and red line denotes other conflicts Taliban involved in**

Next we show that the clustering and self- and mutual-excitation properties of the events can play a key role for solving DEAP. Figure 2(a) shows the sequence of 149 dyadic events of conflicts between Taliban and civilians during the years 2007-2011. In this sequence, 70% conflicts happened within 5 days of the previous conflict, i.e., the sequence shows chain reactions where the occurrence of one conflict tends to increase the probability of conflicts in the near future. For instance, Taliban’s attack against civilians can result in civilians’ protest against Taliban. We also find that one conflict not only will trigger future conflicts happen between the same actor-pair, but also future conflicts that share at least one actor. For instance, in a large-scale terrorist attack, Taliban may attack U.S army, Afghanistan army, and civilians in sequence. Taliban’s attack against civilians may cause Afghanistan army to revenge. Figure 2(b) shows the sequence of conflicts happen between Taliban & civilians, Taliban & Afghan Government, Taliban & Police Force. We find that over 93% conflicts between Taliban and civilians happened within 5 days of the previous conflicts that Taliban involved in, i.e. any conflicts Taliban participated in probably caused conflicts between Taliban and civilians in the near future.

Our main contribution is to propose principled probabilistic models based on mixtures of Hawkes processes that simultaneously addresses DEAP and network parameter inference underlying the probabilistic models. We model the interactions between each actor-pair using a Hawkes process. These Hawkes processes mutually

affect each other as we consider the interdependency among the dyadic events shared by at least one actor. In addition, we discuss an additive model to incorporate regularization considerations on model parameters to avoid overfitting.

We conduct experiments on both synthetic and the ACLED data set to evaluate the performance of our proposed methods, and compare them with the state-of-the-art methods. We find that the proposed methods are capable of significantly improve accuracy in DEAP. Results on ACLED provide us a clear view of the trigger and influence of conflict events under different factors like time, regions, and actors.

## 2. RELATED WORK

Recent works usually modeled social networks that vary with time by self-exciting point processes. One important self-exciting process is the Hawkes process, which was first used to analyze earthquakes [13], and later applied to a wide range of tasks such as market modeling [5, 2], crime modeling [16], terrorist [14], conflict [18], and viral videos on the Web [4]. An EM framework was proposed to estimate the maximum likelihood of Hawkes processes [11]. However, most existing works focused on modeling the behaviors of a single user.

Our research is inspired by a recent work on DEAP targeting Los Angeles gang network. Gang activities are successfully modeled using Hawkes processes [16, 8]. However, these existing works employed substantial approximation schemes and failed to optimize the data likelihood. Another recent work on DEAP proposed a spatial-temporal latent point process that independently modeled a number of distinct event-cascades [3], which ignored the influence among dyadic events that share at least one actor.

## 3. HAWKES PROCESSES

Before proposing our model for DEAP, we briefly describe a powerful statistical tool, Hawkes processes, for modeling and analyzing event sequence data. A univariate Hawkes process  $\{N_t\}$  is defined by

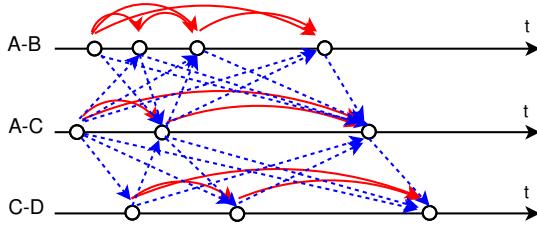
$$\lambda^*(t) = \mu(t) + \int_{-\infty}^t \kappa(t-s)dN(s),$$

where  $\mu : \mathbb{R} \rightarrow \mathbb{R}_+$  is a deterministic base intensity (i.e. how likely an event will occur when no other event triggers it),  $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a kernel function expressing the positive influence of past events on the current value of the intensity process [7]. The process is well known for its *self-exciting* property, which refers to a common social phenomenon that the occurrence of one event increases the probability of related events (events of the same type or share at least one actor) in the near future.

The multivariate Hawkes process  $\{N_m(t)|m = 1, \dots, M\}$ , a multi-dimensional extension to the univariate case, describes the occurrences of  $M$  coupling point series [7, 12]. The intensity function  $\lambda^* = [\lambda_1^*, \dots, \lambda_M^*]^\top$  is defined by

$$\lambda_m^*(t) = \mu_m(t) + \sum_{m'=1}^M \int_{-\infty}^t \kappa_{m'm}(t-s)dN_{m'}(s),$$

where  $\kappa_{m'm}$  is a triggering kernel between a pair of dimensions  $m'$  and  $m$ . This process is also known as linear mutually exciting process since the occurrence of an event in one dimension increases the likelihood of future events in all dimensions. Hence the Hawkes process is suitable for DEAP as it ties unidentified events and identified events together through multiple dependencies.



**Figure 3: Illustration of the dependency among dyadic event in our mixture model.** Existing works only model the dependency shown by red line, while our work models the dependency shown by both red and blue line. A-B, A-C, C-D are actor-pairs.

## 4. MIXTURES OF HAWKES PROCESSES

This section introduces a novel probabilistic model based on mixtures of Hawkes processes (MHP) that simultaneously addresses DEAP and network parameter inference. DEAP can be formulated as follows: our observation is a sequence of  $N$  dyadic events,  $t_n, n = 1, \dots, N$ , where  $t_n$  is the time-stamp of the  $n$ -th event. These  $N$  dyadic events belong to  $M$  actor-pairs, for each  $n$ , we associate a  $M$ -dimensional binary vector  $Z_n = [Z_{n,1}, \dots, Z_{n,M}]^T$ , and  $Z_{n,m} = 1$  if and only if the  $n$ -th event happens on actor-pair  $m$ . Here only part of actor-pair  $Z_n$ 's are observed, while other  $Z_n$ 's are unknown. Based on  $t_n$ 's and observed  $Z_n$ 's, DEAP predicts unknown  $Z_n$ 's, in order to maximize the likelihood on the complete data  $(Z, t) = \{(Z_n, t_n)\}_{n=1}^N$ .

We predict unknown  $Z_n$ 's of dyadic events based on their interactions with related events that share at least one actor with them. And we take such interactions as exciting processes where each event will raise the probability of all the related events in the near future. We turn to the Hawkes process to model such interactions.

We model the interactions between the two actors in each actor-pair as a Hawkes process. That is to say, for each  $m, m = 1, \dots, M$ , the intensity of the model on actor-pair  $m$  is defined as:

$$\lambda_m^*(t) = \mu_m(t) + \sum_{t_l < t} \kappa_{m'm}(t - t_l) \sum_{m' \in S_m} Z_{l,m'},$$

where  $S_m$  is the set of actor-pairs that share at least one actor with actor-pair  $m$ . The above definition illustrates that the probability of a dyadic event at timestamp  $t$  is influenced by all dyadic events that share at least one actor with it<sup>1</sup>. Figure 3 compares the dependency among dyadic events modeled in our work with that modeled in state-of-the-art methods.

In DEAP for conflict data, the baseline intensity  $\mu_m(t)$  captures how often actor-pair  $m$  starts a conflict spontaneously (i.e., not triggered by any other conflicts). For simplicity, we assume this cascade-birth process is a homogeneous Poisson process with  $\mu_m(t) = \mu_m$ . Kernel  $\kappa_{m'm}(t - t_l)$  captures the influence between sequential conflicts. We propose to decompose this pairwise triggering kernel into

$$\kappa_{m'm}(t - t_l) = \hat{\alpha}_{m'm} \kappa(t - t_l) = \alpha_m \kappa(t - t_l),$$

here for simplicity, we assume the degrees of influence from all historical events to the current event from actor-pair  $m$  are similar and enforce that  $\forall m', \hat{\alpha}_{m'm} = \alpha_m$ . Thus  $\alpha_m$  models the degree of

<sup>1</sup>For each event that shares at least one actor with actor-pair  $m$ , obviously we have  $\sum_{m' \in S_m} Z_{l,m'} = 1$ .

influence from any historical event to the current event from actor-pair  $m$ . And  $\kappa(t - t_l)^2$  captures the time-decay effect only.

Suppose we have observations  $(Z, t) = \{(Z_n, t_n)\}_{n=1}^N$  over the observation window  $[0, T]$ , the likelihood for the complete data is

$$\begin{aligned} L(Z, t) &= \prod_{m=1}^M \left( \prod_{n: Z_{n,m}=1} \lambda_m(t_n) \exp \left( - \int_0^T \lambda_m(s) ds \right) \right) \\ &= \prod_{n=1}^N \prod_{m=1}^M \lambda_m(t_n)^{Z_{n,m}} \exp \left( - \sum_{m=1}^M \int_0^T \lambda_m(s) ds \right). \end{aligned}$$

## 5. VARIATIONAL INFERENCE

In this section, we derive a mean-field variational Bayesian inference algorithm for our proposed MHP model.

For the likelihood  $L(Z, t)$  discussed above, the latent variables  $Z$ 's are inter-dependent, i.e., the actor assignment at current step  $Z_n$  depends on all the past actor assignments  $\{Z_1, \dots, Z_{n-1}\}$  as well as all the future ones  $\{Z_{n+1}, \dots\}$ . Marginalizing over such inter-connected series is intractable. We use mean-field variational inference by assuming a fully-factorizable variational distribution  $q$  for  $Z$ 's, which is parametrized by free variables  $\phi$ 's as

$$q(\{Z_n\}) = \prod_{n=1}^N \text{Multinomial}(Z_n | \phi_n).$$

With the help of  $q$ , we lower-bound the log-likelihood:

$$\mathcal{L}(t) = \log(\int_{\{Z\}} L(Z, t) d\{Z\}) \geq \mathbb{E}_q[\mathcal{L}(Z, t)] + \mathcal{E}[q], \quad (1)$$

where  $\mathcal{L}(Z, t) = \log(L(Z, t))$ , and  $\mathcal{E}[q]$  denotes the Shannon entropy of  $Z$ 's under  $q$ .  $\mathcal{L}$  denotes the right-hand side of Eqn (1), known as the **evidence lower-bound** (ELBO), which is the alternative of the true log-likelihood we are to optimize. We have

$$\mathcal{L} = \sum_{n=1}^N \sum_{m=1}^M \phi_{nm} (\mathbb{E}_q[\log \lambda_m(t_n)]) - \sum_{m=1}^M \int_0^T \mathbb{E}_q[\lambda_m(s)] ds + \mathcal{E}[q],$$

where the second term reduces to

$$T \sum_{m=1}^M \mu_m + \sum_{m=1}^M \sum_{n=1}^N K(T - t_n) \sum_{m' \in S_m} \phi_{nm'} \alpha_m,$$

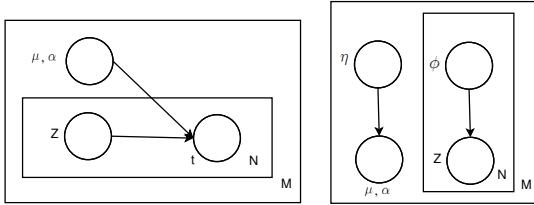
$$\text{where } K(t) = \int_0^t \kappa(s) ds.$$

To break down the log-sum in  $\mathbb{E}_q[\log(\lambda_m(t_n))]$ , we again apply Jensen's inequality

$$\begin{aligned} \mathbb{E}_q[\log(\lambda_m(t_n))] &\geq \sum_{l=1}^{n-1} \sum_{m' \in S_m} \phi_{lm'} \eta_{ln}^{(m)} \log(\alpha_m \kappa(t_n - t_l)) \\ &\quad - \sum_{l=1}^{n-1} \sum_{m' \in S_m} \phi_{lm'} \eta_{ln}^{(m)} \log(\eta_{ln}^{(m)}) + \eta_{nn}^{(m)} \log(\mu_m) - \eta_{nn}^{(m)} \log(\eta_{nn}^{(m)}), \end{aligned}$$

where we introduce a set of branching variables  $\{\eta^{(m)}\}_{m=1}^M$ . Each  $\eta^{(m)}$  is a  $N \times N$  lower-triangular matrix with  $n$ -th row  $\eta_{\cdot n}^{(m)} =$

<sup>2</sup>Our paper uses the exponential kernel in experiments, i.e.,  $\kappa(\Delta t) = \omega e^{-\omega \Delta t}$  if  $\Delta t \geq 0$  or 0 otherwise. However, the model development and inference is independent of kernel choice and extensions to other kernels such as power-law, Rayleigh, non-parametric kernels are straightforward.



**Figure 4: Graphical model representation of MHP and the variational distribution that approximates the likelihood**

$[\eta_{1,n}^{(m)}, \dots, \eta_{n,n}^{(m)}]^T$  satisfying two set of constraints list in the following:

$$\begin{aligned} \eta_{ln}^{(m)} &\geq 0, \quad l = 1, \dots, n \\ \eta_{nn}^{(m)} + \sum_{l=1}^{n-1} \eta_{ln}^{(m)} \sum_{m' \in S_m} \phi_{l,m'} &= 1, \forall n, m. \end{aligned} \quad (2)$$

Figure 4 concludes the graphical model representation of our MHP model and the variational distribution used to approximate the data likelihood.

Optimizing the Lagrangian of  $\mathcal{L}'$ , we have the following inference rule for  $\phi$ 's:

$$\begin{aligned} \phi_{nm} &\propto (\mu_m)^{\eta_{nn}^{(m)}} : \text{self triggering} \\ &\times \prod_{l=1}^{n-1} \left( \frac{\alpha_m \kappa(t_n - t_l)}{\eta_{ln}^{(m)}} \right)^{\eta_{ln}^{(m)} \sum_{m' \in S_m} \phi_{lm'}} : \text{influences from past} \\ &\times \prod_{l=n+1}^N (\alpha_m \kappa(t_l - t_n))^{\eta_{nl}^{(m)} \sum_{m' \in S_m} \phi_{lm'}} : \text{influences to future} \\ &\times e^{-K(T-t_n) \sum_{m' \in S_m} \alpha_{m'}} : \text{time-decay effect.} \end{aligned}$$

Notice that the above inference rule only applies to those  $\phi_{nm}$ 's whose corresponding  $Z_{n,m}$ 's are unknown. For those  $\phi_{nm}$ 's whose corresponding  $Z_{n,m}$ 's are already known, we just set  $\phi_{nm} = Z_{n,m}$ . In a similar way, we obtain the following update rules for  $\eta$ :

$$\begin{aligned} \eta_{nn}^{(m)} &= \frac{\mu_m}{\mu_m + \sum_{l=1}^{n-1} \sum_{m' \in S_m} \phi_{lm'} \alpha_m \kappa(t_n - t_l)}, \\ \eta_{ln}^{(m)} &= \frac{\alpha_m \kappa(t_n - t_l)}{\mu_m + \sum_{l=1}^{n-1} \sum_{m' \in S_m} \phi_{lm'} \alpha_m \kappa(t_n - t_l)}. \end{aligned}$$

Next we derive maximum likelihood estimation of our proposed MHP model. This model involves two parameters, i.e., the self-instantaneous rate  $\mu \in \mathbb{R}_+^M$  and the infectivity vector  $\alpha \in \mathbb{R}_+^M$ , where  $\mathbb{R}_+$  denotes the nonnegative real domain. Setting the derivative of  $\mathcal{L}'$  with respect to  $\alpha_m$  and  $\mu_m$  to zero, we can obtain:

$$\begin{aligned} \alpha_m &= \frac{\sum_{n=1}^N \phi_{nm} \sum_{l=1}^{n-1} \sum_{m' \in S_m} \phi_{lm'} \eta_{ln}^{(m)}}{\sum_{n=1}^N \phi_{nm} K(T - t_n)}, \\ \mu_m &= \frac{1}{T} \sum_{n=1}^N \phi_{nm} \eta_{nn}^{(m)}. \end{aligned}$$

Based on above discussions, we obtain a mean-field variation inference algorithm, named MHP, iteratively updates  $\phi$ 's, and  $\eta$ 's to attribute unidentified dyadic events, and updates  $\mu$ 's, and  $\alpha$ 's to infer the network diffusion. The computation complexity of MHP is  $O(M * \hat{N}^2)$ , where  $\hat{N}$  is the average number of events one actor involved in. Thus MHP is particularly efficient if all events happen between limited actor-pairs.

## 6. ADDITIVE MODEL

This section describes an additive model that regularizes model parameters  $\mu$  and  $\alpha$  to avoid overfitting. Instead of modeling each actor-pair  $m$  with parameters  $\mu_m$  and  $\alpha_m$ , we model each actor  $i$  with parameters  $\mu'_i$  and  $\alpha'_i$ . These new parameters  $\{\mu'_i\}$ ,  $\{\alpha'_i\}$ ,  $i = 1, \dots, I$  ( $I$  is the number of actors) replace  $\mu_m$ 's and  $\alpha_m$ 's through the following rules:

$$\mu_m = \mu'_{m1} + \mu'_{m2}, \quad \alpha_m = \alpha'_{m1} + \alpha'_{m2}, \quad (3)$$

where  $m1$  and  $m2$  are the two actors in the actor-pair  $m$ . Eqn (3) emphasizes that 1) each actor has his/her own behavior pattern  $(\mu'_i, \alpha'_i)$  in conflicts; 2) the behavior pattern of one actor-pair  $(\mu_m, \alpha_m)$  depends on the behavior patterns of both actors. We named the new algorithm solving this additive model Additive Mixture Hawkes Process (AMHP) algorithm.

## 7. EXPERIMENTS

We conducted experiments on synthetic and real-world data sets, and compared the performance of our MHP/AMHP algorithm with state-of-the-art DEAP methods including: Parameter-Fixed Hawkes Process (PFHP) proposed in [16], Estimate & Score Algorithm (ESA) proposed in [8] and Latent Point Process Model (LPPM) proposed in [3].

### 7.1 Data set

We also test our methods on Synthetic data and its variations listed as bellow:

**Synthetic data:** Under a setting  $(n, N, I, M, \hat{\mu}, \hat{\alpha})$ , the synthetic data is generated by Hawkes processes using a group of fixed parameters  $\{\mu_m\}$  and  $\{\alpha_m\}$ , where each  $\mu_m$  and  $\alpha_m$  are randomly generated in  $[0.5\hat{\mu}, 1.5\hat{\mu}]$  and  $[0.5\hat{\alpha}, 1.5\hat{\alpha}]$  respectively before simulations. Notice that for different compared approaches, the intensity function employed for simulation can be different from each other, which depends on each approach's assumption on the dependency among dyadic events from different actor-pairs. Among the generated sequence of events, we randomly select some events for prediction, i.e. the corresponding  $Z_n$ 's are unknown. Here  $n$  refers to the number of events with actors unknown.

**Structured data:** To test the effectiveness of AMHP, we simulate with a group of  $\mu_m$ 's and  $\alpha_m$ 's that satisfy the regularization considerations defined in Eqn (3).

**Event Noisy:** We generate additional  $10\% * N$  dyadic events randomly in the time window  $[0, T]$ , and add them to the original event sequence. These events are assigned to all actor-pairs with equal probability.

**Intensity Noisy:** Instead of using  $\lambda^*(t)$  to simulate the dyadic event generation at time  $t$ , we use a noisy value  $\lambda'(t)$ , which is obtained by adding a Gaussian noise on  $\lambda^*(t)$ :

$$\lambda'(t) = \max(0.1 * e + 1, 0) * \lambda^*(t), \quad e \sim \mathcal{N}(0, \sigma). \quad (4)$$

The default value of  $\sigma$  is set to be 1.

Our experiments also used two real-world data sets both coming from the ACLED data set [1]: Afghanistan Conflict, which contains  $N = 3,384$  dyadic events and 68 actors.  $M = 1,010$  actor-pairs are involved in these events; Africa Conflict, which contains  $N = 52,605$  dyadic events and 3,537 actors.  $M = 1,007$  actor-pairs are involved in these events. We randomly select 10% events from each data set as unidentified events, and make prediction using our proposed methods.

## 7.2 Performance on synthetic data

This series of experiments is to test the accuracy of DEAP for unidentified dyadic events on synthetic data. Table 1 shows our test on a relatively small system with  $N = 120$ ,  $n = 8$ ,  $I = 4$ ,  $M = 6$ ,  $\hat{\mu} = 0.01$ ,  $\hat{\alpha} = 0.5$ , where simulations were run 10,000 times using the pre-generated parameters  $\{\mu_m\}$ ,  $\{\alpha_m\}$ . Table 2 shows results on a larger system with  $N = 10000$ ,  $n = 1000$ ,  $I = 50$ ,  $M = 200$ ,  $\hat{\mu} = 0.01$ ,  $\hat{\alpha} = 0.5$ , where simulations were run 20 times. The accuracy of DEAP is measured by the percentage of unknown events whose ground-truth actor-pairs appear in the predicted top k most likely actor-pairs. Results in Table 1 and 2 show that LPPM, MHP and AMHP are much more accurate than PFHP and ESA in DEAP on both data sets, which demonstrates the importance of using dynamic Hawkes parameters in modeling diffusion processes of dyadic events, and reasonably optimizing the data likelihood. Our proposed MHP and AMHP are more accurate than LPPM, which shows the importance of modeling dependency between dyadic events from different actor-pairs. At last, AMHP performs better than MHP, which indicates that appropriate constraints on Hawkes parameters can benefit DEAP. On Structured data, the advantage of AMHP over other methods becomes greater, which demonstrates the effectiveness of AMHP on dyadic data where the dependency network of actor-pairs has special structures like low-rank or sparsity. On both noisy data sets, the performance of all four compared methods become worse. However, the degradations of the performance of PFHP and ESA are greater than those of LPPM, MHP and AMHP, which shows the robustness of our proposed methods facing noise.

**Table 1: Overall Performance on Small-scale Synthetic Data**

$N = 120$ ,  $n = 8$ ,  $I = 4$ ,  $M = 6$ ,  $\hat{\mu} = 0.01$ ,  $\hat{\alpha} = 0.5$ .

Data set	Method	Top 1	Top 2	Top 3	Top 4	Top 5
Synthetic data	PFHP	50.4%	69.7%	80.2%	88.9%	95.8%
	ESA	52.2%	71.1%	81.6%	90.4%	96.9%
	LPPM	55.8%	75.2%	87.0%	93.6%	97.4%
	MHP	57.3%	77.9%	88.5%	95.0%	97.9%
	AMHP	58.9%	79.3%	89.9%	96.6%	98.3%
Random Guess		16.7%	33.3%	50.0%	66.7%	83.3%

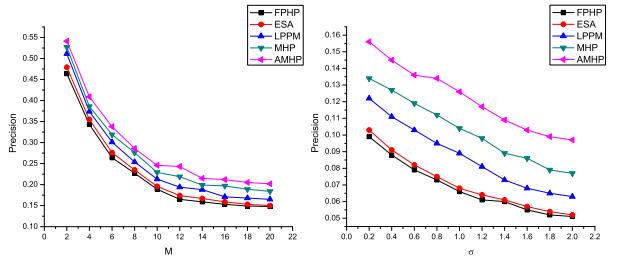
The following series of experiments study how parameter variation affects the accuracy of DEAP. In experiments, we varies the parameter chosen to test, with all other experiment settings fixed. Figure 5(a) shows how the precision of Top 1 inference varies with respect to the increase of the number of actor-pairs  $M$ . We can find that all compared methods experience significant decreases when the number of actor-pairs  $M$  increases. The degradations of MHP and AMHP are smaller than other methods, which shows that appropriately modeling the dependency among actor-pairs becomes more important as the number of actor-pairs increases. Figure 5(b) shows that the accuracy of Top 1 inference decreases wrt. the increase of standard variance  $\sigma$  in Eqn (4). The degradations of LPPM, MHP, and AMHP are smaller than PFHP and ESA, which from another prospect demonstrates the robustness of our methods facing noise.

## 7.3 Performance on real-world data

Next we apply the proposed model to two real-world data sets as shown in Table 3. On both real-world data sets, our proposed MHP and AMHP again perform better than LPPM, PFHP and ESA. We also find that AMHP outperforms MHP, which illustrates that the underlying dependency network of actor-pairs in real-world data has some special structures.

**Table 2: Overall Performance on Large-Scale Synthetic Data**  
 $N = 10000$ ,  $n = 1000$ ,  $I = 50$ ,  $M = 200$ ,  $\hat{\mu} = 0.01$ ,  $\hat{\alpha} = 0.5$ .

Data set	Method	Top 1	Top 2	Top 3	Top 4	Top 5
Synthetic data	PFHP	10.0%	15.9%	18.5%	21.2%	23.1%
	ESA	10.3%	16.3%	19.2%	21.9%	23.7%
	LPPM	12.2%	18.7%	21.5%	23.2%	24.4%
	MHP	13.8%	20.2%	23.3%	24.9%	26.4%
	AMHP	14.7%	21.0%	24.4%	26.2%	27.8%
Structured data	PFHP	11.3%	16.6%	19.4%	22.7%	25.4%
	ESA	11.8%	17.1%	20.0%	23.2%	25.8%
	LPPM	13.3%	19.7%	22.1%	24.8%	26.3%
	MHP	17.9%	23.3%	26.9%	28.4%	30.1%
	AMHP	19.1%	24.2%	28.0%	29.8%	31.3%
Event Noisy	PFHP	6.8%	13.2%	15.4%	18.0%	20.1%
	ESA	7.1%	13.5%	15.7%	18.4%	20.4%
	LPPM	10.6%	16.4%	18.8%	20.7%	22.2%
	MHP	12.9%	18.5%	20.8%	23.0%	24.5%
	AMHP	13.7%	19.4%	21.6%	23.9%	25.1%
Intensity Noisy	PFHP	6.6%	12.3%	14.9%	17.5%	19.4%
	ESA	6.8%	12.6%	15.1%	17.7%	19.8%
	LPPM	8.9%	14.7%	16.5%	19.2%	20.6%
	MHP	11.4%	16.6%	19.6%	21.1%	22.8%
	AMHP	12.1%	17.4%	20.4%	22.0%	23.4%
Random Guess		0.5%	1.0%	1.5%	2.0%	2.5%



(a) Accuracy of Top 1 wrt.  $M$  (b) Accuracy of Top 1 wrt.  $\sigma$

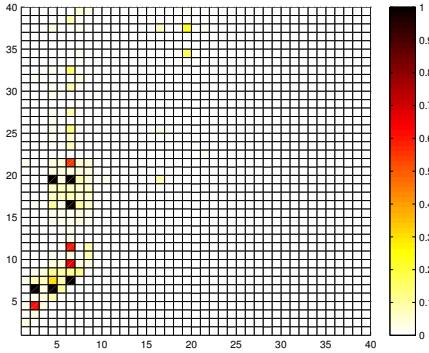
**Figure 5: Parameter Variation**

**Table 3: Overall Performance on Real-world Data**

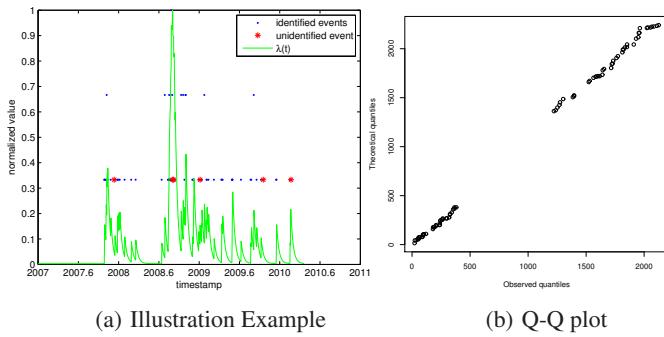
Data set	Method	Top 1	Top 2	Top 3	Top 4	Top 5
Afghanistan	PFHP	11.8%	17.0%	20.2%	21.8%	24.2%
	ESA	12.6%	18.1%	21.3%	23.0%	25.5%
	LPPM	13.4%	20.9%	23.8%	25.4%	26.5%
	MHP	14.6%	23.3%	26.8%	28.6%	30.1%
	AMHP	15.5%	24.0%	27.7%	29.3%	30.8%
Random Guess		0.1%	0.2%	0.3%	0.4%	0.5%
Africa	PFHP	9.0%	14.6%	18.2%	20.0%	22.3%
	ESA	9.9%	15.7%	19.5%	21.3%	23.7%
	LPPM	11.2%	18.7%	21.6%	23.2%	24.4%
	MHP	12.4%	20.9%	24.7%	26.1%	27.5%
	AMHP	13.1%	21.5%	25.4%	26.9%	28.1%
Random Guess		0.1%	0.2%	0.3%	0.4%	0.5%

Figure 6 shows a relational graph among actors in Afghanistan Conflict data based on learned  $\alpha_m$ 's. A large  $\alpha_m$  indicates that one conflict between actor-pair  $m$  is quite likely to be triggered by another conflict sharing at least one actor with  $m$  in the recent past. Most areas in the figure are blank, which illustrates that most sequential conflicts in Afghanistan happened between limited actor-pairs. Analyzing shades of color in each grid, we can find out unique behavior patterns of different actor-pairs. For exam-

ple, comparing  $\alpha_m$  of pair **Taliban–U.S. Army** and that of pair **Taliban–Afghan Government**, we can tell that one conflict event involving Taliban or U.S. army will quickly cause a chain reaction between the **Taliban–U.S. Army** pair, while it takes much longer time for a conflict involving Taliban or Afghan government to trigger consequential conflicts. In other words, when suffering an attack or protest, Taliban's favorite retaliation target is U.S. Army rather than Afghan government.



**Figure 6: Relational Graph among Actors in Afghanistan Conflict Based on Learned  $\alpha_m$ 's.** Indice of important actors: 2-U.S. Army, 4-Civilians, 6-Taliban, 7-Afghanistan Army, 9-Britain Army, 11-Afghan Government, 16-Police Force, 19-ISAF, 21-Private Security.



**Figure 7: How Well Our MHP Model Fits Conflict Data.** Blue dot denotes the number of identified events at each timestamp, red star denotes the number of unidentified event. Both the number of events and the value of  $\lambda_m(t)$  are scaled to  $[0, 1]$ .

Next we study how well our MHP model fits conflict data. On a certain actor-pair  $m$  in Afghanistan Conflict, Figure 7(a) compares the number of conflict events at each timestamp with the curve of  $\lambda_m(t)$  using learned  $\mu_m$  and  $\alpha_m$ , while Figure 7(b) shows the Q-Q plot of the real conflict sequence versus the sample event sequence simulated by  $\lambda_m(t)$ . We can find that although the  $\mu_m$  and  $\alpha_m$  are inferred with part of  $Z_n$ 's unknown, they fit both identified conflicts and unidentified conflicts very well.

## 8. CONCLUSION AND FUTURE WORK

This paper proposed a novel mixture Hawkes model for tackling DEAP and network inference simultaneously. This mixture

Hawkes model captures the dependency among dyadic events from actor-pairs that share at least one actor. Our fast inference algorithm based on mean-field methods iteratively predicts actor-pairs and infers network diffusion. Experiments on both synthetic and real-world data sets on international armed conflicts demonstrated the effectiveness of the proposed methods.

In the future, we will look for alternative solutions that capture the dependency among dyadic events from different actor-pairs, and propose novel mixture Hawkes models. Instead of using a human-defined regularization for model parameters, we will explore how to infer structures, such as low-rank or sparsity, from data. We will also study more efficient algorithms that optimize our proposed likelihood.

## 9. ACKNOWLEDGMENTS

Part of this work was supported by NSF grant IIS-1116886, NIH 1R01GM108341, and a DARPA Xdata grant.

## 10. REFERENCES

- [1] ACLED. <http://www.acleddata.com>, 2010.
- [2] Y. Ait-Sahalia, J. Cacho-Diaz, and R. Laeven. Modeling financial contagion using mutually exciting jump processes. *Tech. rep.*, 2010.
- [3] Y.-S. Cho, A. Galstyan, J. Brantingham, and G. Tita. Latent point process models for spatial-temporal networks. *In Manuscript*, 2013.
- [4] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41):15649–15653, 2008.
- [5] E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM J. Fin. Math.*, 1(1):642–665, Sep 2010.
- [6] M. G.-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *ICML*, pages 561–568, 2011.
- [7] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.
- [8] R. A. Hegemann, E. A. Lewis, and A. L. Bertozzi. An "estimate & score algorithm" for simultaneous parameter estimation and reconstruction of missing data on social networks. *accepted in Security Informatics*, 2012.
- [9] D. A. Kashy and D. A. Kenny. The analysis of data from dyads and groups. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology*, page 451477, 2000.
- [10] D. A. Kenny. Dyadic analysis. *The SAGE Encyclopedia of Social Science Research Methods*, 2004.
- [11] E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes. *Journal of Nonparametric Statistics*, 1, 2011.
- [12] T. J. Liniger. Multivariate hawkes processes. *ETH Doctoral Dissertation*, (18403), 2009.
- [13] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.
- [14] M. D. Porter and G. White. Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6(1):106–124, 2011.
- [15] C. Raleigh, A. Linke, H. Hegre, and J. Karlsen. Introducing acled: An armed conflict location and event dataset special data feature. *Journal of Peace Research*, 47(5):651–660, July 2010.
- [16] A. Stomakhin, M. B. Short, and A. L. Bertozzi. Reconstruction of missing data in social networks based on temporal patterns of interactions. *Inverse Problems*, 27(11), Nov 2011.
- [17] D. Q. Vu, A. U. Asuncion, D. R. Hunter, and P. Smyth. Dynamic egocentric models for citation networks. In *ICML*, pages 857–864, 2011.
- [18] A. Z.-Mangion, M. Dewarc, V. Kadirkamanathan, and G. Sanguinetti. Point process modelling of the afghan war diary. *PNAS*, 109(31):12414–12419, July 2012.