Scalable and Consistent Estimation in Continuous-time Networks of Relational Events

Makan Arastuie
University of Toledo
Toledo, Ohio, USA
makan.arastuie@rockets.utoledo.edu

Subhadeep Paul
The Ohio State University
Columbus, Ohio, USA
paul.963@osu.edu

Kevin S. Xu
University of Toledo
Toledo, Ohio, USA
kevin.xu@utoledo.edu

ABSTRACT
In many application settings involving networks, such as messages between users of an on-line social network or transactions between traders in financial markets, the observed data consist of timestamped relational events, which form a continuous-time network. We propose the \textit{Community Hawkes Independent Pairs (CHIP)} generative model for such networks. We show that applying spectral clustering to adjacency matrices constructed from relational events generated by the CHIP model provides \textit{consistent community detection} for a growing number of nodes. We also develop consistent and computationally efficient estimators for the model parameters. We demonstrate that our proposed CHIP model and estimation procedure scales to large networks with tens of thousands of nodes and provides superior fits than existing continuous-time network models on several real networks.

This submission is a novel research paper.

CCS CONCEPTS
• Mathematics of computing → Random graphs; Probabilistic representations; Probabilistic reasoning algorithms; Stochastic processes; • Computing methodologies → Learning latent representations; Spectral methods.

KEYWORDS
Community Hawkes Independent Pairs model, event-based network, continuous-time network, timestamped network, relational events, spectral clustering, Hawkes process

ACM Reference Format:

1 INTRODUCTION
A variety of complex systems in the computer, information, biological, and social sciences can be represented as a network, which consists of a set of objects (nodes) and relationships (edges) between the nodes. In many application settings, we observe edges in the form of distinct events occurring between nodes over time. For example, in online social networks, users interact with each other through events that occur at specific time instances such as liking, mentioning, or sharing another user’s content. Such interactions form \textit{timestamped relational events}, where each event is a triplet \((i, j, t)\) denoting events from node \(i\) (sender) to node \(j\) (receiver) at timestamp \(t\). The observation of these triplets defines a dynamic network that continuously evolves over time.

Relational event data are usually modeled by combining a point process model for the event times with a network model for the sender and receiver [10, 13, 14, 33, 42, 43, 53, 57]. We refer to such models as \textit{continuous-time network models} because they provide probabilities of observing events between two nodes during arbitrarily short time intervals. For model-based exploratory analysis and prediction of future events with relational event data, continuous-time network models are often superior to their discrete-time counterparts [41, 54–56, 58], which first aggregate events over time windows to form discrete-time network “snapshots” and thus lose granularity in modeling temporal dynamics. However, theoretical analysis of estimators is significantly more advanced for discrete-time network models [9, 19, 23, 45].

We propose the \textit{Community Hawkes Independent Pairs (CHIP)} model which is inspired by the recently proposed Block Hawkes Model (BHM) [33] for timestamped relational event data. Both CHIP and BHM are based on the Stochastic Block Model (SBM) for static networks [29]. In the BHM, events between different pairs of nodes belonging to the same pair of communities are dependent, which makes it difficult to analyze. In contrast, for CHIP the pairs of nodes in the same community generate events according to \textit{independent} univariate Hawkes processes with shared parameters, so that the number of parameters remains the same as in the BHM. The independence between node pairs enables tractable analysis of the CHIP model and more scalable estimation than the BHM.

Our main contributions are as follows. (1) We demonstrate that spectral clustering provides consistent community detection in the CHIP model for a growing number of nodes. (2) We propose consistent and computationally efficient estimators for the model parameters for a growing number of nodes and time duration. (3) We show that the CHIP model provides better fits to several real datasets and scales to much larger networks than existing models, including a Facebook network with over 40,000 nodes and over 800,000 events. Other point process network models have demonstrated good empirical results, but to the best of our knowledge, this work provides the first theoretical guarantee of estimation accuracy. Our asymptotic analysis also has tremendous practical value given the scalability of our model to large networks with tens of thousands of nodes.
2 BACKGROUND

2.1 Hawkes Processes

The Hawkes process [25] is a counting process designed to model continuous-time arrivals of events that naturally cluster together in time, where the arrival of an event increases the chance of the next event arrival immediately after. They have been used to model earthquakes [39], financial markets [8, 16], and user interactions on social media [13, 62].

A univariate Hawkes process is a self-exciting point process where its conditional intensity function given a sequence of event arrival times \( \{t_1, t_2, t_3, ..., t_l\} \) for \( l \) events up to time duration \( T \) takes the general form \( \lambda(t) = \mu + \sum_{t_i < t} \gamma(t - t_i) \), where \( \mu \) is the background intensity and \( \gamma(\cdot) \) is the kernel or the excitation function. A frequent choice of kernel is an exponential kernel, parameterized by \( \alpha, \beta > 0 \) as \( \gamma(t - t_i) = \alpha e^{-\beta(t-t_i)} \) where the arrival of an event instantaneously increases the conditional intensity by the jump size \( \alpha \), after which the intensity decays exponentially back towards \( \alpha \) at rate \( \beta \). Restricting \( \alpha < \beta \) yields a stationary process. We use an exponential kernel for the CHIP model, since it has been shown to provide a good fit for relational events in social media [22, 33, 40, 60].

2.2 The Stochastic Block Model

Statistical models for networks typically consider a static network rather than a network of relational events. Many static network models are discussed in the survey by Goldenberg et al. [20]. A static network with \( n \) nodes can be represented by an \( n \times n \) adjacency matrix \( A \) where \( A_{ij} = 1 \) if there is an edge between nodes \( i \) and \( j \) and \( A_{ij} = 0 \) otherwise. We consider networks with no self-edges, so \( A_{ii} = 0 \) for all \( i \). For a directed network, we let \( A_{ij} = 1 \) if there is an edge from node \( i \) to node \( j \).

One model that has received significant attention is the stochastic block model (SBM), formalized by Holland et al. [29]. In the SBM, every node \( i \) is assigned to one and only one community or block \( c_i \in \{1, \ldots, k\} \), where \( k \) denotes the total number of blocks. Given the block membership vector \( c = [c_i]_{i=1}^n \), all entries of the adjacency matrix \( A_{ij} \) are independent Bernoulli random variables with parameter \( p_{c_i, c_j} \), where \( p \) is a \( k \times k \) matrix of probabilities. Thus the probability of forming an edge between nodes \( i \) and \( j \) depends only on the block memberships \( c_i \) and \( c_j \). There have been significant recent advancements in the analysis of estimators for the SBM. Several variants of spectral clustering [51], including regularized versions [4, 11], have been shown to be consistent estimators of the community assignments for a growing number of nodes in the SBM and various extensions [12, 18, 32, 35, 46–48, 52, 59]. Spectral clustering scales to large networks with tens of thousands of nodes and is generally not sensitive to initialization, so it is also a practically useful estimator.

2.3 Related Work

One approach for modeling continuous-time networks is to treat the edge strength of each node pair as a continuous-time function that increases when an event occurs between the node pair and then decays afterwards [3, 31, 63]. Another approach is to combine a point process model for the event times, typically some type of Hawkes process, with a network model. The conditional intensity functions of the point processes then serve as the time-varying edge strengths. Point process network models are used in two main settings. The first involves estimating the structure of a latent or unobserved network from observed events at the nodes [17, 21, 27, 37, 38, 49]. These models are often used to estimate static networks of diffusion from information cascades.

In the second setting, which we consider in this paper, we directly observe events between pairs of nodes so that events take on the form \((i, j, t)\) denoting an event from node \( i \) to node \( j \) at timestamp \( t \). Our objective is to model the dynamics of such event sequences. In many applications, including messages on online social networks, most pairs of nodes either never interact and thus have no events between them. Thus, most prior work in this setting utilizes low-dimensional latent variable representations of the networks to parameterize the point processes.

The latent variable representations are often inspired by generative models for static networks such as continuous latent space models [28] and stochastic block models [29], resulting in the development of point process network models with continuous latent space representations [57] and latent block or community representations [10, 13, 14, 33, 42, 43, 53]. Point process network models with latent community representations are most closely related to the model we consider in this paper. Exact inference in such models is intractable due to the discrete nature of the community assignments. Approximate inference techniques including Markov Chain Monte Carlo (MCMC) [10, 13, 43] or variational inference [33, 42] have been used in prior work. While such techniques have demonstrated good empirical results, to the best of our knowledge, they come with no theoretical guarantees.

3 THE COMMUNITY HAWKES INDEPENDENT PAIRS (CHIP) MODEL

We consider a generative model for timestamped relational event networks that we call the Community Hawkes Independent Pairs (CHIP) model. The CHIP model has parameters \((\pi, \mu, \alpha, \beta)\). Each node is assigned to a community or block \( a \in \{1, \ldots, k\} \) with probability \( p_a \), where each entry of \( \pi \) is non-negative and all entries sum to 1. We represent the block assignments of all nodes either by a length \( n \) vector \( c = [c_i]_{i=1}^n \) or an \( n \times k \) binary matrix \( C \) where \( c_i = q \) is equivalent to \( C_{iq} = 1 \). Let \( C_{ij} = 0 \) for all \( i \neq q \). Each of the parameters \( \mu, \alpha, \beta \) is a \( k \times k \) matrix. Event times between node pairs \((i, j)\) within a block pair \((a, b)\) follow independent exponential Hawkes processes with shared parameters: baseline rate \( \mu_{ab} \), jump size \( \sigma_{ab} \), and decay rate \( \beta_{ab} \). The generative process for our proposed CHIP model is as follows:

\[
\begin{align*}
    c_i & \sim \text{Categorical}(\pi) \quad \text{for all nodes } i \\
    t_{ij} & \sim \text{Hawkes process}(\mu_{c_i c_j}, \alpha_{c_i c_j}, \beta_{c_i c_j}) \quad \text{for all } i \neq j \\
    Y & = \text{Row concatenate } \{(i, j, t_{ij})\} \quad \text{over all } i \neq j
\end{align*}
\]

Let \( T \) denote the end time of the Hawkes process, which would correspond to the duration of the data trace. The column vector of event times \( t_{ij} \) has length \( N_{ij}(T) \), which denotes the number of events from node \( i \) to node \( j \) up to time \( T \). Let \( Y \) denote the event matrix and has dimensions \( l \times 3 \), where \( l = \sum_{ij} N_{ij}(T) \) denotes the total number of observed events over all node pairs.
Algorithm 1 Estimation procedure for Community Hawkes Independent Pairs (CHIP) model

**Input:** Relational event matrix $Y$, number of blocks $k$  
**Result:** Estimated block assignments $\hat{C}$ and CHIP model parameters $(\hat{\pi}, \hat{\mu}, \hat{\alpha}, \hat{\beta})$

1. for all node pairs $i \neq j$ do
2. $N_{ij} =$ number of events from $i$ to $j$ in $Y$
3. $\hat{\mathcal{C}} \leftarrow$ Spectral clustering($N, k$)
4. for all block pairs $(a, b)$ do
5. Compute estimates $(m_{ab}, \hat{\beta}_{ab})$ using (1)
6. $\hat{\alpha}_{ab} \leftarrow$ maximize log-likelihood by line search
7. $\hat{\beta}_{ab} \leftarrow \hat{\beta}_{ab} m_{ab}$
8. $\hat{\pi} \leftarrow$ proportion of nodes in each block
9. return $[\hat{\mathcal{C}}, \hat{\pi}, \hat{\mu}, \hat{\alpha}, \hat{\beta}]$

It is constructed by row concatenating triplets $(i, j, t_{ij}(q))$ over all events $q \in \{1, \ldots, N_{ij}(T)\}$ for all node pairs $i, j \in \{1, \ldots, n\}, i \neq j$.

### 3.1 Relation to Other Models

Our proposed CHIP model has a generative structure inspired by the SBM for static networks. Other point process network models in the literature have also utilized similar block structures, but they have been incorporated in two different approaches.

One approach involves placing point process models at the level of block pairs [10, 33, 42, 53]. For a network with $k$ blocks, $k^2$ different point processes are used to generate events between the $k^2$ block pairs. To generate events between pairs of nodes, rather than pairs of blocks, the point processes are thinned by randomly selecting nodes from the respective blocks so that all nodes in a block are stochastically equivalent, in the spirit of the SBM. Such models have demonstrated good empirical results, but the dependency between node pairs complicates analysis of the models.

The other approach involves modeling pairs of nodes with independent point processes that share parameters among nodes in the same block pair [13, 14]. By having node pairs in the same block pair share parameters, the number of parameters is the same as for the models with block pair-level point processes. However, by using independent point processes for all node pairs, there is no dependency between node pairs, which simplifies analysis of the model. We exploit this independence to perform the theoretical analysis of our estimator in Section 4.

### 3.2 Estimation Procedure

As with many other block models, the maximum-likelihood estimator for the discrete community assignments $\mathcal{C}$ is intractable except for extremely small networks (e.g. 10 nodes). We propose a scalable estimation procedure for the CHIP model that has two components as shown in Algorithm 1: a community detection component and a parameter estimation component. For the community detection component, we use spectral clustering on the weighted adjacency or count matrix $N(T)$ or simply $N$ with entries $N_{ij}(T)$. Since this is a directed adjacency matrix, we use singular vectors rather than eigenvectors for spectral clustering.

For the parameter estimation component, we first consider estimating the Hawkes process parameters $(\mu_{ab}, \alpha_{ab}, \beta_{ab})$ for each block pair $(a, b)$ using only the count matrix $N$, which discards event timestamps. Even without access to the event timestamps, we are able to estimate $\mu_{ab}$ and the ratio $m_{ab} = \alpha_{ab} / \beta_{ab}$, but not the parameters $\alpha_{ab}$ and $\beta_{ab}$ separately. Let $n_{ab}$ denote the number of node pairs in block pair $(a, b)$, where $n_{ab} = |a| |b|$ for $a \neq b$ and $n_{ab} = |a||a - 1|$ for $a = b$, with $|a|$ denoting the number of nodes in block $a$. Let $N_{ab}$ and $S_{ab}^2$ denote the sample mean and (unbiased) sample variance, respectively, of the number of events between all node pairs $(i, j)$ in block pair $(a, b)$. Using $N_{ab}$ and $S_{ab}^2$, we propose the following method of moments estimators (conditioned on the estimated blocks) for $m_{ab}$ and $\mu_{ab}$ from the count matrix $N$:

$$m_{ab} = 1 - \frac{N_{ab}}{S_{ab}^2}, \quad \hat{\mu}_{ab} = \frac{1}{T} \frac{(N_{ab})^3}{S_{ab}^2}.$$  

Finally, the vector of block assignment probabilities $\pi$ can be easily estimated using the proportion of nodes in each block, i.e. $\hat{\pi}_a = \frac{1}{n} \sum_{i=1}^n \mathcal{C}_{ia}$ for all $a = 1, \ldots, k$.

In some prior work, exponential Hawkes processes are parameterized only in terms of $m$ and $\mu$, with $\beta$ treated as a known parameter that is not estimated [6, 7, 61]. In this case, the estimation procedure is complete. On the other hand, if we want to estimate the values of both $\alpha$ and $\beta$ rather than just their ratio, we have to use the actual event matrix $Y$ with the event timestamps. To separately estimate the $\alpha_{ab}$ and $\beta_{ab}$ parameters, we replace $\alpha_{ab} = \beta_{ab} m_{ab}$ in the exponential Hawkes log-likelihood for block pair $(a, b)$ then plug in our estimate $\hat{m}_{ab}$ for $m_{ab}$. Then the log-likelihood is purely a function of $\beta_{ab}$ and can be maximized using a standard scalar optimization or line search method.

### 4 THEORETICAL ANALYSIS OF ESTIMATORS

We first derive non-asymptotic bounds on the misclustering error in a simplified setting typically employed in the literature. We then derive consistency and asymptotic normality properties of the estimators for the Hawkes process parameters. In an extended version of the paper on arXiv [5], we provide an analogous theorem for the general CHIP model, theorems for spectral clustering on an unweighted adjacency matrix, comparisons between the weighted (count) and unweighted adjacency matrices, and proofs of all theorems.

#### 4.1 Analysis of Estimated Community Assignments

We define the error of community detection as the misclustering error rate $r = \inf \Pi \frac{1}{n} \sum_{i=1}^n \chi(\hat{\mathcal{C}}_i \neq \Pi(\mathcal{C}_i))$, where $\Pi(\cdot)$ denotes the set of all permutations of the community labels. Our proposed CHIP model considers directed events; however, we analyze community detection on undirected networks to better match up with the literature on analysis of spectral clustering for the SBM. The bounds and consistency properties we derive still apply to the directed case with only a change in the constants. We assume that $T \to \infty$, which can be achieved by rescaling the time unit for event times. Under this assumption, the mean and variance of the number of events between nodes $(i, j)$ are [24, 26, 36]

$$v_{ab} = \frac{\mu_{ab} T}{1 - \alpha_{ab}/\beta_{ab}}, \quad \sigma_{ab}^2 = \frac{\mu_{ab} T}{(1 - \alpha_{ab}/\beta_{ab})^2}.$$  

Finally, the vector of block assignment probabilities $\pi$ can be easily estimated using the proportion of nodes in each block, i.e. $\hat{\pi}_a = \frac{1}{n} \sum_{i=1}^n \mathcal{C}_{ia}$ for all $a = 1, \ldots, k$.
We analyze community detection error in a simplified special case of our CHIP model which is in similar spirit to a commonly-employed case in the stochastic block models literature [12, 18, 35, 44, 47]. In this special case, all communities have roughly equal number of elements $|a| \approx n/k$, all intra-community processes (diagonal block pairs) have the same set of parameters $\mu_1, \alpha_1, \beta_1$ and all inter-community processes (off-diagonal block pairs) have the same set of parameters $\mu_2, \alpha_2, \beta_2$. We use the notation $Y \sim \text{CHIP}(C, n, k, \mu_1, \alpha_1, \beta_1, \mu_2, \alpha_2, \beta_2)$ to denote a relational event matrix $Y$ generated from this simplified model. Define $m_1 = \alpha_1/\beta_1$ and $m_2 = \alpha_2/\beta_2$. Let $v_1 = \mu_1/(1 - m_1)$ and $v_2 = \mu_2/(1 - m_2)$, while $\sigma_1^2 = \mu_1/(1 - m_1)^2$ and $\sigma_2^2 = \mu_2/(1 - m_2)^2$. Assume $v_1 \geq v_2$, $v_1 \times v_2$, and $\sigma_1 \times \sigma_2$, where the asymptotic equivalence is with respect to both $n$ and $T$. These assumptions imply that the expected number of events are higher between two nodes in the same community compared to two nodes in different communities and that the asymptotic dependence on $n$ and $T$ are same for both set of parameters. This setting is useful to understand detectability limits and has been widely employed in the literature on stochastic block models [1, 2, 12, 18, 44, 52]. In this setting, we have the following upper bound on the misclustering error rate.

**Theorem 1.** Let $Y \sim \text{CHIP}(C, n, k, \mu_1, \alpha_1, \beta_1, \mu_2, \alpha_2, \beta_2)$. The misclustering error rate for spectral clustering on the weighted adjacency matrix $N$ at time $T \rightarrow \infty$ is

$$r \leq \frac{n}{(n/k)^2(v_2 - v_1)^2T^2} \times \frac{k^2}{nT(v_1 - v_2)^2}.$$

We note that if the set of parameters $\mu, \alpha, \beta$ remain constant as a function of $n$ and $T$ then the misclustering error rate decreases as $1/T$ with increasing $T$, decreases as $1/n$ with increasing $n$, and increases as $k^2$ with increasing $k$. Hence, as we observe the process for more time, spectral clustering on $N$ has lower error rate. The rate of convergence with increasing $T$ is the same as one would obtain for detecting an average community structure if discrete snapshots of the network were available over time [9, 44, 45]. The dependence of the misclustering error rate on $n$ and $k$ is what one would expect from the SBM literature.

### 4.2 Analysis of Estimated Hawkes Process Parameters

As discussed in Section 3.2, we are able to estimate $m = \alpha/\beta$ and $\mu$ from the count matrix $N$ using (1). We analyze these estimators assuming a growing number of nodes $n$ and time duration $T$. We do not put any assumption on the distribution of the counts; we only require that $T$ is large enough such that the asymptotic mean and variance equations in (2) hold. The sample mean $\hat{N}_{ab}$ and sample variance $\hat{S}_{ab}^2$ of the counts are unbiased estimators of $v_{ab}$ and $\sigma_{ab}^2$, respectively. The following theorem shows that these estimators are consistent and asymptotically normal.

**Theorem 2.** Define $n_{\text{min}} = \min_{a,b} n_{ab}$. The estimators for $n_{ab}$ and $\mu_{ab}$ have the following asymptotic distributions as $n_{\text{min}} \rightarrow \infty$ and $T \rightarrow \infty$:

$$\sqrt{n_{ab}} \left( \hat{n}_{ab} - \left(1 - \frac{v_{ab}}{\sigma_{ab}^2} \right) \right) \xrightarrow{d} N \left(0, \frac{1}{4v_{ab}} \right),$$

$$\sqrt{n_{ab}} \left( \hat{\mu}_{ab} - \left(\frac{v_{ab}^{3/2}}{\sigma_{ab}} \right) \right) \xrightarrow{d} N \left(0, \frac{9}{4v_{ab}} \right).$$

Using Theorem 2, we can obtain confidence intervals for $\mu$ and $m$, which we derive in the extended version of the paper [5]. In the simplified special case of Theorem 1, we have equal community sizes so $n_{ab} \times (n/k)^2$. Therefore, the condition $n_{\text{min}} \rightarrow \infty$ boils down to $(n/k)^2 \rightarrow \infty$, which is a reasonable assumption. Theorem 2 guarantees convergence of our estimators for $\mu$ and $m$ with the asymptotic mean-squared errors (MSEs) decreasing as $1/n_{ab} \times (n/k)^2$ under the assumption that the community structure is correctly estimated. Next, we provide an “end-to-end” guarantee for the convergence of the asymptotic MTE to 0 for estimating the mean number of events in each block pair $v_{ab}$ using the sample mean $\hat{N}_{ab}$ over the estimated communities using spectral clustering.

**Theorem 3.** Assume $n_{ab} \times (n/k)^2$. The weighted average of asymptotic MSEs in estimating $v_{ab}$ using the estimator $\hat{N}_{ab}$ with communities estimated by spectral clustering is

$$\frac{\sum_{ab} n_{ab} \text{E}[(\hat{N}_{ab} - v_{ab})^2]}{\sum_{ab} n_{ab}} \leq \frac{kT}{n} \max \left\{ \frac{\sigma_1^2}{2}, \frac{k^2\sigma_1^2}{(v_1 - v_2)^2} \right\}.$$

For comparison, under the assumption that the community structure is correctly estimated, the weighted average of asymptotic MSEs in estimating $v_{ab}$ using the estimator $\hat{N}_{ab}$ is

$$\frac{\sum_{ab} n_{ab} \text{E}[(\hat{N}_{ab} - v_{ab})^2]}{\sum_{ab} n_{ab}} = \frac{k^2T\sigma_1^2}{n^2}.$$

Theorem 3 guarantees that the MSE for estimating Hawkes process parameters decreases at least at a linear rate with increasing $(n/k)$ when the error from community detection is taken into account instead of the quadratic rate when the error is not taken into account.

### 5 EXPERIMENTS

We begin with a set of simulation experiments to assess the accuracy of our proposed estimation procedure and verify our theoretical analysis. We then present several experiments on real data involving both prediction and model-based exploratory analysis. Additional experiments and code to replicate our experiments are provided in the extended version of the paper [5].

#### 5.1 Community Detection on Simulated Networks with Varying $T$, $n$, and $k$

We simulate networks from the simplified CHIP model while varying two of $T$, $n$, and $k$ simultaneously. We choose parameters $\mu_1 = 0.085$, $\mu_2 = 0.065$, $\alpha_1 = \alpha_2 = 0.06$, and $\beta_1 = \beta_2 = 0.08$. The upper bounds on the error rates in Theorem 1 involve all three parameters $n, k, T$ simultaneously, making it difficult to interpret the result. To better observe the effects of $n, k, T$ and their relationship with respect to each other, we perform three separate
simulations each time varying two and fixing the other one. The community detection accuracy averaged over 30 simulations using the weighted adjacency matrix \( N \) as two of \( T \), \( n \), and \( k \) are varied is shown in Figure 1. Since the estimated community assignments will be permuted compared to the actual community labels, we evaluate the community detection accuracy using the adjusted Rand score [30], which is 1 for perfect community detection and has an expectation of 0 for a random assignment.

Note that Theorem 1 predicts that the misclustering error rate varies as \( k^2/(nT) \) if all three parameters are varied. Figure 1(a) shows the accuracy to be low for small \( T \) and large \( k \). As we simultaneously increase \( T \) and decrease \( k \), the accuracy improves until the adjusted Rand score reaches 1. We also note that it is possible to obtain high accuracy either with increasing \( T \) or decreasing \( k \) or with both even when \( n \) is fixed. This is in line with the prediction from Theorem 1 that the misclustering error rate varies as \( k^2/T \) if \( n \) remains fixed. We observe a similar effect of increasing accuracy with increasing \( n \) and decreasing \( k \) when \( T \) is kept fixed in Figure 1(b). Finally, Figure 1(c) verifies the prediction that accuracy increases with both increasing \( n \) and \( T \) for a fixed \( k \).

5.2 Hawkes Process Parameter Estimation on Simulated Networks
Next, we examine the estimation accuracy of the Hawkes process parameter estimates as described in Section 4.2. We simulate networks from the simplified CHIP model with \( k = 4 \) blocks, duration \( T = 10,000 \) and parameters \( \mu_1 = 0.0011, \mu_2 = 0.0010, \alpha_1 = 0.11, \alpha_2 = 0.09, \beta_1 = 0.14, \) and \( \beta_2 = 0.16 \) so that each parameter is different between block pairs. We then run the CHIP estimation procedure: spectral clustering followed by Hawkes process parameter estimation.

Figure 2 shows the mean-squared errors (MSEs) of all four estimators decay quadratically as \( n \) increases. Theorem 2 states that \( \hat{m} \) and \( \hat{\mu} \) are consistent estimators with MSE decreasing at a quadratic rate for growing \( n \) with known communities. Here, we observe the quadratic decay even with communities estimated by spectral clustering, where the mean adjusted Rand score is increasing from 0.6 to 1 as \( n \) grows. We observe that \( \alpha \) and \( \beta \) are also accurately estimated for growing \( n \) even though \( \beta \) is estimated using a line search for which we have no guarantees.

5.3 Comparison with Other Models on Real Networks
We perform experiments on three real network datasets. Each dataset consists of a set of events where each event is denoted by a sender, a receiver, and a timestamp. The MIT Reality Mining [15] and Enron [34] datasets were loaded and preprocessed identically to DuBois et al. [13] to allow for a fair comparison with their reported values. On the Facebook wall posts dataset [50], we use the largest connected component of the network excluding self loops (43,953 nodes).

We fit our proposed Community Hawkes Independent Pairs (CHIP) model as well as the Block Hawkes Model (BHM) [33] to all three real datasets and evaluate their fit. We also compare against a simpler baseline: spectral clustering with a homogeneous Poisson process for each node pair. For each model, we also compare against the case \( k = 1 \), where no community detection is being performed. We do not have ground truth community labels for these real datasets so we cannot evaluate community detection accuracy. Instead, we use the mean test log-likelihood per event as the evaluation metric, which allows us to compare against the reported results in DuBois et al. [13] for the relational event model (REM). Since the log-likelihood is computed on the test data, this is a measure of the model’s ability to forecast future events rather than detect communities.

As shown in Table 1, CHIP outperforms all other models in all three datasets. Note that test log-likelihood is maximized for CHIP at relatively small values of \( k \) compared to the BHM. This is because CHIP assumes independent node pairs whereas the BHM assumes all node pairs in a block pair are dependent. Thus, the BHM needs a higher value for \( k \) in order to model independence. This difference is particularly visible for the Reality Mining data, where CHIP with \( k = 1 \) is the best predictor of the test data, while the best BHM has \( k = 50 \) on a network with only 70 nodes! These both suggest a weak community structure that is not predictive of future events in the Reality Mining data, whereas community structure does appear to be predictive in the Enron and Facebook data.

In addition to the improved predictive ability of CHIP compared to the BHM, the computational demand is also significantly decreased. Fitting the CHIP model for each value of \( k \) took on average
We use CHIP to perform model-based exploratory analysis to understand the behavior of different groups of users in the Facebook wall post network. We consider all 852,833 events and choose \( k = 10 \) blocks using the eigengap heuristic [51], which required 141 s to fit. Note that the CHIP estimation procedure can scale up to a much higher number of communities also—fitting CHIP to the Facebook data with \( k = 1,000 \) communities took just under 50 minutes! The adjacency matrix permuted by the block structure is shown in Figure 3(a), and heatmaps of the fitted CHIP parameters are shown in Figures 3(b) and 3(c). Diagonal block pairs on average have a base intensity \( \mu \) of \( 2.8 \times 10^{-7} \), which is higher compared to \( 9.5 \times 10^{-8} \) for off-diagonal block pairs, indicating an underlying assortative community structure. However, not all blocks have higher rates of within-block posts, e.g. \( \mu_{s, s} > \mu_{s, s'} \) and \( \mu_{s', s} > \mu_{s', s'} \), as shown in red boxes in Figure 3(b), which illustrates that the CHIP model does not discourage inter-block events. These patterns often occur in social networks, for instance, if there are communities with opposite views on a particular subject.

While the structure of \( \mu \) reveals insights on the baseline rates of events between block pairs, the structure of the \( \alpha/\beta \) ratio \( m \) shown in Figure 3(c) reveals insights on the burstiness of events between block pairs. For some block pairs, such as \((3, 10)\), there are very low values of \( \alpha \) and \( \beta \) indicating the events are closely approximated by a homogeneous Poisson process, while some block pairs such as \((2, 8)\) are extremely bursty despite low baseline rates. Both block pairs are shown in blue dashed boxes. The different levels of burstiness of block pairs cannot be seen from aggregate statistics such as the the count matrix \( N \) or even the mean number of events per node pair in the block pair, as shown in Figure 3(d).

### 6 Conclusion
We introduced the Community Hawkes Independent Pairs (CHIP) model for timestamped relational event data. The CHIP model has...
Figure 3: Inferred CHIP parameters on the largest connected component of the Facebook Wall Posts dataset with $k = 10$. Axis labels denote block numbers. Each tile corresponds to a block pair where $(a, b)$ denotes row $a$ and column $b$. Boxed block pairs in the heatmap are discussed in the body text.

...many similarities with the Block Hawkes Model (BHM) [33]; however, in the CHIP model, events among any two node pairs are independent which enables both tractable theoretical analysis and scalable estimation. We demonstrated that an estimation procedure using spectral clustering followed by Hawkes process parameter estimation provides consistent estimates of the communities and Hawkes process parameters for a growing number of nodes. Lastly, we showed that CHIP also provides better fits to several real networks compared to the Relational Event Model [13] and the BHM. It also scales to considerably larger data sets, including a Facebook wall post network with over 40,000 nodes and 800,000 events.

There are several limitations to the CHIP model and our proposed estimation procedure. Assuming all node pairs to have independent Hawkes processes simplifies analysis and increases scalability but also reduces the flexibility of the model compared to multivariate Hawkes process-based models that specifically encourage reciprocity [10, 43]. Additionally, our estimation procedure uses unregularized spectral clustering to match our theoretical analysis in Section 4. We note that regularized versions of spectral clustering [4, 11, 32, 46, 59] have been found to perform better empirically and would likely improve the community detection accuracy in the CHIP model. Additionally, methods that jointly estimate the community structure and Hawkes process parameters, such as the local search and variational inference approaches explored in Junuthula et al. [33] for the Block Hawkes Model could also improve estimation accuracy of both.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation grants IIS-1755824 and DMS-1830412.

REFERENCES


