MULTIVARIATE SPATIOTEMPORAL HAWKES PROCESSES AND NETWORK RECONSTRUCTION*

BAICHUAN YUAN†, HAO LI†, ANDREA L. BERTOZZI†, P. JEFFREY BRANTINGHAM‡, AND MASON A. PORTER†

Abstract. There is often latent network structure in spatial and temporal data, and the tools of network analysis can yield fascinating insights into such data. In this paper, we develop a non-parametric method for network reconstruction of spatiotemporal data sets using multivariate Hawkes processes. In contrast to prior works on network reconstruction using point-process models, which have often focused on exclusively temporal information, our approach — which does not assume a specific parametric form of network dynamics — uses both temporal and spatial information. This leads to an effective way of recovering an underlying network. We illustrate our approach using both synthetic networks and networks constructed from real-world data sets (a location-based social media network, a narrative of crime events, and violent gang crimes). Our results demonstrate that, in comparison to using only temporal data, our spatiotemporal approach yields improved network reconstruction, providing a basis for meaningful subsequent analysis — such as community structure and motif analysis — of the reconstructed networks.

Key words. Multivariate Hawkes processes, community structure, spatiotemporal data, social networks, motifs

AMS subject classifications. 60G55, 62H11, 91D30

map: a few things:

1. **there are still notational issues to polish**
2. **various things still need to be clarified, especially in the second half of the paper**

BY: also need to add funding GRYD

1. Introduction.

Digital devices — such as smartphones, tablets, and others — generate a massive amount of spatiotemporal data about human activities, providing a wonderful opportunity for researchers to gain insight into human dynamics through our "digital footprints". A broad variety of human activities are analyzed using such data, creating new disciplines [34] such as computational social science and digital humanities. Examples include online check-ins in large cities [12], effects of human mobility [5] and currency flow [9] on the spread of contagious diseases, online communications during Occupy Wall Street [13], crime reports in Los Angeles county [29], and many others.

Network analysis is a powerful approach for representing and analyzing complex systems of interacting components [47], and network-based methods can provide considerable insights into the structure and dynamics of complex spatiotemporal data [7]. It has been valuable for studies of both digital human footprints and human mobility [6]. To give one recent example, Noulas et al. [49] studied geographic online social networks to illustrate similarities and heterogeneities in human mobility patterns.

Suppose that each node in a network represents an entity, and the edges (which can either be undirected or directed, and can either be unweighted or weighted) repre-

*Submitted to the editors DATE.

Funding: BY, HL, ALB, and PJB were funded by NSF grant DMS-1417674, NSF grant DMS-1737770, and DARPA grant FA8750-18-2-0066.

†Department of Mathematics, University of California, Los Angeles, Los Angeles, CA (byuan@math.ucla.edu, lihao0809@math.ucla.edu, bertozzi@math.ucla.edu, mason@math.ucla.edu).

‡Department of Anthropology, University of California, Los Angeles, Los Angeles, CA (branting@ucla.edu).

This manuscript is for review purposes only.

sent spatiotemporal connections between pairs of entities. For instance, in a check-in data set from a social medium, one can model each user as a node, which has associated check-in time and locations. In this case, one can suppose that an edge exists between a pair of users if they follow each other on the social medium. One can use edge weights to quantify the amount of "influence" between users, where a larger weight signifies a larger impact. In our investigation, we assume that the relationships between nodes are time-independent. For other regimes of relative time scales between spatiotemporal processes and network dynamics, it is necessary to consider time-dependent edges [24, 54]. In some cases, the entities and relationships are both known, and one can investigate the structure and dynamics of the associated networks. However, in many situations, network data is incomplete — with potentially a large amount of missing data, in the form of missing entities, interactions, and/or metadata [64] — and the relationships between nodes may not be directly observable [58]. For example, social media companies attempt to infer friendship relationships among their users to provide accurate friendship recommendations for online social networks.

In the last few years, there has been a considerable amount of work on inferring missing data in networks. A basic approach of inferring relationships among entities is to calculate cross-correlations of their associated time series [33]. Another approach is to use coefficients from a generalized linear model (GLM) [45], a generalization of linear regression that allows response variables to have a non-Gaussian error-distribution. Recently, people begin to use point-process methods [61] in network reconstruction. Perry and Wolfe [53] modeled networks as a multivariate point process and then inferred covariate-based edges by estimating a point process. Among point-process models, it is very popular to use Hawkes processes (also known as self-exciting point processes; both are used interchangeably in this paper) for studying human dynamics [20, 36]. Hawkes process models are characterized by mutual "triggering" among events [50], as one event may increase the probability for subsequent events to occur. Such models can capture inhomogeneous inter-event times and causal (temporal) correlations, which have both been observed in human dynamics [28]. These properties make it a useful approach in social-network applications [26]. It thus seems promising to use this type of processes for network inference on dynamic human data such as crime events or stock-market dynamics. For example, Linderman and Adams [36] proposed a network Hawkes model that appears to be more accurate at inferring missing edges than GLMs, cross-correlations, and a simple self-exciting point process with an exponential kernel in the data that they studied. Further self-exciting point process is applied to reconstruct more sophisticated network such as multilayer networks [27, 66]. However, the aforementioned point-process models are not without limitations. For example, most of these models do not use spatial information, even when it plays a significant role in a system's dynamics. Furthermore, many assume an a priori model [36] or a specific parametrization [67] for their point processes.

In the present paper, we propose a nonparametric and multivariate version of the spatiotemporal Hawkes process. Spatiotemporal Hawkes processes have been used to study numerous topics (e.g., crime [42], social media networks [32], and earthquake prediction [19]). In our model, each node in a network is associated with a spatiotemporal Hawkes process. The nodes can "trigger" each other, so events that are associated with one node increase the probability that there will be events associated with the other nodes. We measure the extent of such mutual-triggering effects using a $U \times U$ "triggering matrix" $\mathbf{K}$, where U is the number of nodes. If one considers an exclusively temporal scenario, a point process u does not "cause" (in the Granger sense [21]) a point process v if and only if $\mathbf{K}(u, v) = 0$ [17]. Because triggering be-

tween point processes reflects an underlying connection, one can try to recover latent relationships in a network from $\mathbf{K}$. Such triggering should decrease with both distance and time according to some spatial and temporal kernels. In our work, instead of assuming exponential decay [20] or some other distributions [36, 67], we adopt a nonparametric approach [39] to learn both spatial and temporal kernels from data using an expectation-maximization-type (EM-type) algorithm [70].

We compare our approach with other recent point-process network-reconstruction methods [20, 36] on both synthetic and real-world data sets with spatial information such as a location-based social networking website and crime data sets. We illustrate the importance both of incorporating spatial information and of using nonparametric kernels. Although we assume that the relationships among nodes are time-independent, our model still recovers a causal structure among events in synthetic data sets. We also build event-causality networks on data sets about violent crimes of gangs and examine gang retaliation patterns using motif analysis.

The rest of our paper is organized as follows. In section 2, we review self-exciting point processes and recent point-process methods for network reconstruction. In section 3, we introduce our nonparametric spatiotemporal model and our approaches to model estimation and simulation. In section 4, we compare our model with others on both synthetic and real-world data sets. We conclude in section 5. We give details about our preprocessing for one of our data sets in Appendix A.

2. Self-Exciting Point Processes. A *point process* S is a random measure on a complete separable metric space that takes values on $\mathbb{N} \cup \{\infty\}$ [59]. We first consider a *temporal point process*, which consists of a list of N time points $\{t_1, t_2, \ldots, t_N\}$ with corresponding events $1, 2, \ldots, N$. Let $S[a, b)$ denote the number of points (i.e., events) that occur in a finite time interval $[a, b)$, with $a < b$. One typically models the behavior of a simple temporal point process (multiple events cannot occur at the same time) by specifying its conditional intensity function $\lambda(t)$, which represents the rate at which events are expected to occur around a particular time t, conditional on the prior history of the point process before time t. Specifically, when $H_t = \{t_i | t_i < t\}$ is the history of the process up to time t, one defines the *conditional intensity function*

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{\mathbb{E}[S[t, t + \Delta t) | H_t]}{\Delta t} .$$

One important point-process model is a *Poisson process*, in which the number of points in any time interval follows a Poisson distribution and the number of points in disjoint sets are independent. A Poisson process is called *homogeneous* if $\lambda(t) \equiv$ constant and is thus characterized by a constant rate at which events are expected to occur per unit time. It is called *inhomogeneous* if the conditional intensity function $\lambda(t)$ depends on the time t (e.g., $\lambda(t) = \sin(t)$). In both situations, the numbers of points (i.e., events) in disjoint intervals are independent random variables.

We now discuss self-exciting point processes, which allow one to examine a notion of causality in a point process. If we consider a list $\{t_1, t_2, \ldots, t_N\}$ of time stamps, we say that a point process is *self-exciting* if

$$\mathrm{Cov}\left[S(t_{k-1}, t_k), S(t_k, t_{k+1})\right] > 0 \quad \text{for } k \text{ such that} \quad t_{k-1} < t_k < t_{k+1} .$$

That is, if an event occurs, another event becomes more likely to occur locally in time.

A univariate *Hawkes process* is a self-exciting temporal point process with the

conditional intensity function

$$\lambda(t) = \mu(t) + K \sum_{t_k<t} g(t - t_k)\,, \tag{2.1}$$

where the background rate $\mu(t)$ can either be a constant or a time-dependent function that describes how the likelihood of some process (crimes, emails, tweets, and so on) evolves in time. For example, violent crimes are more likely to happen at night than during the day, and business e-mails are less likely to be sent during the weekend than on a weekday. One can construe the rate $\mu(t)$ as a process that designates the likelihood of an event to occur, independent of the other events. The summation term in Equation (2.1) describes the self-excitation: past events increase the current conditional intensity. The function $g(t)$ is called the *triggering kernel*, and the parameter K denotes the mean number of events that are triggered by an event. One standard example is a Hawkes process with an exponential kernel $g(t) = \omega e^{-\omega t}$, where ω is a constant decay rate for the triggering kernel that controls how fast the rate $\lambda(t)$ returns to its baseline level $\mu(t)$ after an event occurs.

2.1. Temporal Multivariate Models. In network reconstruction, one seeks to infer the relationships (i.e., edges) and their strengths (i.e., weights) among a set of entities (i.e., nodes). When modeling the relationships on a network, it is more appropriate to use a multivariate point process than a univariate one. In a temporal multivariate point process, there are U different point processes $(S_u)_{u=1,\dots,U}$, and the corresponding conditional intensity functions are $(\lambda_u(t))_{u=1,\dots,U}$. We seek to infer the intensity functions from observed data $(t_j, u_j)_{j=1,\dots,N}$, where t_j and u_j, respectively, are the time and point-process index of event j. There are numerous applications of temporal multivariate point processes, such as financial markets [4] and neural spike trains [10]. Here we focus on the specific application of network reconstruction.

A trivial example of the multivariate model is the multivariate Poisson process, in which each point process is a univariate Poisson process. Another example is the multivariate Cox process, which consists of doubly stochastic Poisson processes where the conditional intensity itself is a stochastic process. Perry and Wolfe [53] use this to model e-mail interactions (edges) among different e-mail users (nodes). Neither the multivariate Poisson nor the multivariate Cox process are self-exciting.

Instead of modeling edges as Cox processes, Fox et al. [20] used multivariate Hawkes processes to model people (nodes) communicating with each other via e-mail. Their conditional intensity function has an exponential kernel and a nonparametric background function $\mu_u(t)$ for each person (process) u:

$$\lambda_u(t) = \mu_u(t) + \sum_{t_i<t} K_{u_i u} \omega e^{-\omega(t-t_i)}\,, \tag{2.2}$$

where $K_{uv} = \mathbf{K}(u, v)$ is the expected number of events of person v that are triggered by one event of person u. One can estimate the parameters by minimizing the negative log-likelihood function

$$-\log(L(\Omega)) = -\sum_{k=1}^{N} \log(\lambda_{u_k}(t_k)) + \sum_{u=1}^{U} \int_0^T \lambda_u(t)\mathrm{d}t\,. \tag{2.3}$$

There are several variants of the multivariate Hawkes process. One approach is to add regularization terms on Equation (2.3) to improve the accuracy of parameter estimations. Lewis and Mohler [35] used maximum-penalized likelihood estimation, which

enforces some regularity on the model parameters, to infer Hawkes processes. Zhou et al. [71] extended this idea and promoted the low-rank and sparsity properties of $\mathbf{K}$ by adding nuclear and L_1 norms of $\mathbf{K}$ to Equation (2.3) with the conditional intensity function $\lambda_u(t)$ defined in Equation (2.2). Linderman et al. [36] added random-graph priors on $\mathbf{K}$ and developed a fully Bayesian multivariate Hawkes model. See [38] for theoretical guarantees on inferring Hawkes processes with a regularizer. Another direction is to speed up the parameter estimation of point-process models. For example, Hall et al. [23] tried to learn the triggering matrix $\mathbf{K}$ via an online learning framework for streaming data. Instead of using likelihood-based method, Achab et al. [1] developed a fast moment matching method to estimate only the matrix $\mathbf{K}$.

2.2. Spatiotemporal Point Processes. Many real-world data sets include not only timestamps but also accompanying spatial information, which can be particularly important for correctly inferring and understanding the associated dynamics [7]. For example, in earthquakes, most aftershocks usually occur geographically near the main shock [51]. In online social media networks, if two people often check in at the same location at closely proximate times, there is more likely to be a connection between them than if such "joint check-ins" occur rarely [12]. **map: need a reference for the second statement above; it is not obvious that the inequality should go in that direction, and in fact for many applications, I find that counterintuitive; if we are going to use that phrasing as an example, we need a research study that showed evidence of this as a concrete example (and I think even then we need to weaken the statement)** These situations suggest that it is important to examine spatiotemporal point processes rather than just temporal ones. Indeed, there are myriad applications of spatiotemporal Hawkes processes, including crime prediction [42], seismology [51], and Twitter topics [32]. The successful employment of such processes in earthquake prediction and predictive policing [43] have inspired our work, in which we extend these ideas to network reconstruction.

As before, we characterize a spatiotemporal point process $S(t, x, y)$ via its conditional intensity $\lambda(t, x, y)$, which is defined as the expected rate on the accumulation of points around a particular spatiotemporal location. Given the history $\mathcal{H}_t$ of all points up to time t, we write

$$\lambda(t, x, y) = \lim_{\Delta t, \Delta x, \Delta y \downarrow 0} \left(\frac{\mathbb{E}\left[S\{(t, t+\Delta t) \times (x, x+\Delta x) \times (y, y+\Delta y)\} | \mathcal{H}_t \right]}{\Delta t \Delta x \Delta y} \right).$$

For the purpose of modeling earthquakes, [51] used a self-exciting point process with a conditional intensity of the form

$$\lambda(t, x, y) = \mu(x, y) + \sum_{t > t_i} g(x - x_i, y - y_i, t - t_i).$$

In this setting, if an earthquake occurs, aftershocks are more likely to occur locally in time and space. The choice of the triggering kernel $g(t, x, y)$ is inspired by physical properties of earthquakes. For example, [51] used a modified Omori formula (a power law; see [50]) to describe the frequency of aftershocks per unit time. In sociological applications, there is no direct theory to indicate appropriate choices for the kernel function. Some researchers have chosen specific kernels (e.g., exponential kernels) that are easy to compute. For example, Tita et al. [67] used a spatiotemporal point process to infer missing information about event participants. They modeled interactions

between event participants as a combination of a temporal Hawkes process with an exponential kernel and a spatial Gaussian mixture model. A key problem is how to justify kernel choices in specific applications.

3. Spatiotemporal Models for Network Reconstruction. Many network-reconstruction methods using self-exciting point processes, such as [20, 36], have inferred time-independent relationships (i.e., edges) among entities (i.e., nodes) with corresponding (exclusively) temporal point processes. Entity (process) u is adjacent to v if $\mathbf{K}(u, v) > 0$, where one estimates the triggering matrix $\mathbf{K}$ from the data. Entity u is not adjacent to v if entity u's point process does not cause entity v's point process [17] in time. For many problems, it is desirable — or even crucial — to incorporate spatial information [7, 14]. For example, spatial information is an important part of online fingerprints in human activity, and it has a significant impact on most other social networks. In crime modeling, there is a so-called "near repeat" phenomenon indicating the necessity of including spatial information. The spatial neighborhood of an initial burglary has a higher risk of repeat victimization than more-distant locations [60]. In our work, we propose multivariate spatiotemporal Hawkes processes to define relationships and provide a novel approach for analyzing spatiotemporal dynamics.

Another important issue is the assumptions on triggering kernels for a Hawkes process. In seismology, for example, researchers attempt to use an underlying physical model to help determine a good kernel. However, it is much more difficult to validate such models in social networks than for physical or even biological phenomena [55]. The content of social data is often unclear, and there is often minimal understanding of the underlying mechanisms that produce them. With less direct knowledge of possible triggering kernels, it is helpful to employ a data-driven method for kernel selection. Using a kernel with an inappropriate decay rate may lead to either underestimation or overestimation of the elements in the triggering matrix $\mathbf{K}$. This may also include false negatives or positives in the inferred relationships among entities. Therefore, we ultimately use a nonparametric approach to learn triggering kernels in various applications to avoid a priori assumptions about a specific parametrization.

A multivariate spatiotemporal Hawkes process is a sequence $\{(t_i, x_i, y_i, u_i)\}_{i=1}^{N}$ with N events, where t_i and (x_i, y_i) are spatiotemporal stamps, u_i is the point-process index of event i. Each of the U nodes is a *marginal process*. The conditional intensity function for node u is

$$\lambda_u(t, x, y) = \mu_u(x, y) + \sum_{t > t_i} K_{u_i u} g(x - x_i, y - y_i, t - t_i) \,. \tag{3.1}$$

The above Hawkes process assumes that each node u has a background Poisson process that is constant in time but inhomogeneous in space with conditional intensity $\mu_u(x, y)$. There is also self-excitation, as past events increase the likelihood of subsequent events. We quantify the amount of impact that events associated with node u_i have on subsequent events of node u_j with a spatiotemporal kernel and the element $\mathbf{K}(u_i, u_j) = K_{u_i u_j}$ of the triggering matrix.

3.1. A Parametric Model. We first propose a multivariate Hawkes process with a specific parametric form. We use this model to generate spatiotemporal events on synthetic networks and provide a form of "ground truth" that we can use later.

The background rate μ_u and the triggering kernel g for Equation (3.1) are given

by

$$g(x,y,t) = g_1(t) \times g_2(x,y) = \omega \exp(-\omega t) \times \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right),$$

$$\mu_u(x,y) = \sum_{i=1}^{N} \frac{\beta_{u_i u}}{2\pi\eta^2 T} \times \exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{2\eta^2}\right).$$

For simplicity, we use exponential decay in time [50] and a Gaussian kernel in space [41]. We let T denote the time window of a data set; $K_{u_i u}$ denote the mean number of the events in process u that are triggered by each event in the process u_i; the quantity $\beta_{u_i u}$ denote the extent to which events in process u_i contribute to the background rate for events in the process u; σ and η denote the standard deviation of the spatial Gaussian kernel. The value of σ determines the spreading scale of the triggering effect in space.

3.2. A Nonparametric Model. With the conditional intensity given in Equation (3.1), we estimate the triggering kernel $g(x,y,t) = g_1(t) \times g_2(x,y)$ nonparametrically using histogram estimators [39]. We assume that g_2 is isotropic, which entails that $g_2(x,y) = g_2(r)$, where $r = \sqrt{x^2+y^2}$. Then the spatial kernel in the radial coordinate $h(r) = 2\pi g_2(r)$. For the background rate, we extend the one proposed in [19] to multivariate case

$$\mu_u(x,y) = \gamma_u \tau(x,y) = \frac{\gamma_u}{T} \sum_{i=1}^{N} \frac{p_{ii}}{2\pi d_i^2} \exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{2d_i^2}\right), \tag{3.2}$$

where γ_u is the background intensity of process u and p_{ii} is the probability that event i is a background event (i.e., it is not triggered by any event). We compute d_i by determining the radius of the smallest disk centered at (x_i, y_i) that includes at least n_p other events and is at least as large as some small value ϵ that represents the error in location.

Once we fit the model to spatiotemporal data, the triggering matrix $\mathbf{K}$ gives our inferences for the underlying relationships between entities. For two entities u and v, the matrix element $\mathbf{K}(u,v)$ indicates a mixture of temporal causality and spatial dependence between them. In inferring latent relationships in a network, we assume that entity u is not related to v if $\mathbf{K}(u,v) = 0$. We threshold the matrix $\mathbf{K}$ at a certain level: we set elements that are smaller than the threshold value to zero and either maintain the values of larger elements to obtain a weighted network or set them to one to produce an unweighted network. We use $\tilde{\mathbf{K}}$ to denote the thresholded matrix $\mathbf{K}$. We interpret that there is no relation between two nodes u and v if $\tilde{\mathbf{K}}(u,v) = \tilde{\mathbf{K}}(v,u) = 0$.

3.3. Model Estimation. We use an EM-type algorithm [70] to estimate the parameters and kernel functions of our model. The EM-type algorithm is an iterative method to find maximum likelihood estimates of parameters. We now assume that the original model depends on unobservable latent variables. Suppose that we have data X and want to estimate a parameter Ω. One can then view the likelihood function $L(\Omega; X)$ as the marginal likelihood function of $L(\Omega; Y, X)$, where Y is a latent variable. We call $L(\Omega; Y, X)$ the "complete-data likelihood function" and $L(\Omega; X)$ the "incomplete-data likelihood function". Because both Y and $L(\Omega; Y, X)$ are random variables, we cannot estimate them directly. Therefore, we introduce the following

expectation function:

$$\begin{aligned} Q(\Omega, \Omega^{i-1}) &= \mathbb{E}\left[\log(L(\Omega; Y, X))|X, \Omega^{i-1}\right] \\ &= \int \log(L(\Omega; Y, X)) f(Y|X, \Omega^{i-1}) \mathrm{d}Y \,, \end{aligned} \tag{3.3}$$

where $f(Y|X, \Omega^{i-1})$ is the probability density function of Y, given the data X and Ω^{i-1}. We then update parameters by solving the following equation:

$$\hat{\Omega}^i = \arg\max_{\Omega} Q(\Omega, \Omega^{i-1}) \,.$$

3.3.1. Parametric Model. The log-likelihood for the parametric model defined in Equation (3.1) within a spatial region R and time window $[0, T]$ is

$$\log(L(\Omega; X)) = \sum_{k=1}^{N} \log(\lambda_{u_k}(t_k)) - \sum_{u=1}^{U} \iint_R \int_0^T \lambda_u(t) \mathrm{d}t \, \mathrm{d}x \, \mathrm{d}y \,. \tag{3.4}$$

We define random variables Y_{ij} and Y_{ij}^b using the approach from [41]. If event j triggers event i via the kernel g, then $Y_{ij} = 1$; otherwise, $Y_{ij} = 0$. The equality $Y_{ij}^b = 1$ indicates that event i is triggered by event j at a background rate μ. We write $p_{ij} = \mathbb{E}[Y_{ij}]$ and $p_{ij}^b = \mathbb{E}[Y_{ij}^b]$. We convert the incomplete-data log-likelihood function in (3.4) into the following complete-data log-likelihood function:

$$\begin{aligned} \log(L(\Omega; X, Y)) = &\sum_{j<i} Y_{ij} \log\left(K_{u_i u_j} g(t_i - t_j, x_i - x_j, y_i - y_j)\right) - \sum_{u=1}^{U}\sum_{i=1}^{N} \beta_{u u_i} \\ &- \sum_{u=1}^{U}\sum_{i=1}^{N} K_{u_i u}\left(1 - e^{-w(T-t_i)}\right) + \sum_{i=1}^{N}\sum_{j=1}^{N} Y_{ij}^b \log(\mu_{u_i}) \,. \end{aligned}$$

We then calculate the expectation function using (3.3) to obtain

$$\begin{aligned} Q(\Omega) = &\sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}^b \log\left(\frac{\beta_{u_j u_i}}{2\pi\eta^2 T} \exp\left(-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\eta^2}\right)\right) - \sum_{u=1}^{U}\sum_{i=1}^{N} \beta_{u_i u} \\ &+ \sum_{j<i} p_{ij} \log\left(\omega K_{u_j u_i} e^{-\omega(t_i - t_j)} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma^2}\right)\right) \\ &- \sum_{u=1}^{U}\sum_{i=1}^{N} K_{u_i u}\left(1 - e^{-w(T-t_i)}\right) \,. \end{aligned}$$

The expectation function is concave, so we can obtain closed-form maximizers directly by taking derivatives with respect to the parameters and setting them to zero. This is the maximization step of the EM-type algorithm. For the expectation step, we use the "optimal" parameter values from the prior maximization step to update the probabilities p_{ij} and p_{ij}^b. By iterating between these two steps, we obtain Algorithm 3.1 for the parametric model. For initialization, we usually sample Ω^0, p_{ij}, and p_{ij}^b uniformly at random. We also note that $p_{ij} = 0$ for $i > j$.

Algorithm 3.1 EM-type Algorithm for the Parametric Model

Inputs: point process: $\{(u_i, t_i, x_i, y_i)\}_{i=1}^N$; initial guesses for parameters: $\Omega^{(0)} = \left(\{K_{uv}^{(0)}\}_{u,v=1}^U, \{\beta_{uv}^{(0)}\}_{u,v=1}^U, \sigma^{(0)}, \omega^{(0)}\right)$ and $\{p_{ij}^{(0)}\}_{i,j=1}^N, \{p_{ij}^{b(0)}\}_{i,j=1}^N$; termination threshold: ϵ.

Outputs: model parameters $\Omega = \left(\{K_{uv}\}_{u,v=1}^U, \{\beta_{uv}\}_{u,v=1}^U, \sigma, \omega\right)$.

Initialize $\delta = 1$ and $k = 0$.

while $\delta > \epsilon$ **do**

Expectation step: for $i, j \in \{1, 2, \cdots, N\}$,

$$p_{ij}^{(k)} = \frac{K_{u_j u_i} g(t_i - t_j, x_i - x_j, y_i - y_j)}{\lambda(x_i, y_i, t_i)}, \quad p_{ij}^{b(k)} = \frac{\beta_{u_j u_i}^{(k)} \exp\left(-\frac{(x_j - x_i)^2 + (y_j - y_i)^2}{2\eta^{(k)2}}\right)}{2\pi\eta^{(k)2} T \lambda(x_i, y_i, t_i)}.$$

Maximization step: for $u, \hat{u} \in \{1, 2, \cdots, U\}$,

$$\omega^{(k+1)} = \frac{\sum_{j<i} p_{ij}^{(k)}}{\sum_{j<i} p_{ij}^{(k)}(t_i - t_j) + \sum_{u=1}^U \sum_{i=1}^N K_{u_i u}(T - t_i) e^{-\omega(T - t_i)}},$$

$$K_{\hat{u}u}^{(k+1)} = \frac{\sum_{l=1}^{n_u} \sum_{t_{i_{\hat{l}}^{\hat{u}}} < t_{i_l^u}} p_{i_l^u i_{\hat{l}}^{\hat{u}}}^{(k)}}{\sum_{l=1}^{n_{\hat{u}}} \left(1 - \exp\left(-w(T - t_{i_{\hat{l}}^{\hat{u}}})\right)\right)}, \quad \beta_{\hat{u}u}^{(k+1)} = \frac{\sum_{i=1}^{n_u} \sum_{j=1}^{n_{\hat{u}}} p_{i_l^u i_{\hat{l}}^{\hat{u}}}^{b(k)}}{n_{\hat{u}}}.$$

Let n_u denote the number of events in point process u and i_l^u, with $l \in \{1, \ldots, n_u\}$, index the events for the process u.

$$\sigma^{(k+1)} = \eta^{(k+1)} = \left(\frac{\sum_{i,j=1}^N \left(p_{ij}^{b(k)} + p_{ij}^{(k)}\right)\left((x_i - x_j)^2 + (y_i - y_j)^2\right)}{2\sum_{i,j=1}^N \left(p_{ij}^{b(k+1)} + p_{ij}^{(k)}\right)}\right)^{1/2}.$$

$\delta = \left\|\Omega^{(k)} - \Omega^{(k+1)}\right\|$, $k = k + 1$.

end while

3.3.2. Nonparametric Model. The log-likelihood function of the nonparametric model is the same as the parametric model in Equation (3.4). We use a similar approach as before to derive an EM-type algorithm for the nonparametric model. The main differences are that (1) only Y_{ij} are latent variables and $Y_{ii} = 1$ means that event i is a background event, whereas $Y_{ji} = 1$ means that event i is triggered by event j; and (2) we assume the triggering kernels $g_1(t)$ and $g_2(r)$ are piece-wise constant functions. We discretize space and time into n_t^{bins} temporal bins and n_r^{bins} spatial bins, on each of which the respective kernel is constant.

To formally present the EM-type algorithm Algorithm 3.2, we borrow notations from [19]. Let C_k denote the set of event pairs (i, j) for which $t_j - t_i$ belongs to the k^{th} temporal bin, D_k denote the set of event pairs (i, j) for which r_{ij} (the distance between nodes i and j) belongs to the k^{th} spatial bin, N_u denote the number of events that include node u, the parameter δt_k denote the size of the k^{th} temporal bin, and δr_k denote the size of the k^{th} spatial bin.

3.4. Simulations. To generate synthetic data for model comparison, we need to simulate self-exciting point processes with the conditional intensity Equation (3.1) for each process u. We use the branching structures [73] of self-exciting point processes to develop Algorithm 3.3 for simulation.

4. Numerical Experiments and Results. We apply our algorithm to both synthetic and real-world data sets to demonstrate the usefulness of incorporating spatial information and our nonparametric approach. We consider a synthetic data set in subsection 4.1, a Gowalla data set in subsection 4.2, and a crime-topic network data set in subsection 4.3. On these three data sets, we compare our nonparametric model

Algorithm 3.2 EM-type Algorithm for our Nonparametric Model

Inputs: point process: $\{(u_i, t_i, x_i, y_i)\}_{i=1}^N$; initial guesses of parameters: $\{K_{uv}^{(0)}\}_{u,v=1}^U$ and $\{p_{ij}^{(0)}\}_{i,j=1}^N$; termination threshold: ϵ.
Outputs: model parameters: $\{K_{uv}\}_{u,v=1}^U$; triggering probability between events: $\{p_{ij}\}_{i,j=1}^N$; temporal triggering kernel: g_1; spatial triggering kernel: g_2.
Initialize $\delta = 1$ and $\eta = 0$.
while $\delta > \epsilon$ **do**
  Update background kernel $\tau^\eta(x, y)$ (see (3.2))
  $\gamma_u^{(\eta)} = \frac{\sum_{u_i=u} p_{ii}^{(\eta)}}{Z^{(\eta)}}$, where $Z^{(\eta)}$ satisfies $\frac{1}{Z^{(\eta)}} \int_0^T \iint_S \tau^\eta(x, y) ds\, dt = 1$ for a bounded spatial domain S and for $u \in \{1, \dots, U\}$.
  $K_{uv}^{(\eta)} = \frac{\sum_{u_i=u} \sum_{u_j=v} p_{ij}^{(\eta)}}{N_u}$ for $u, v \in \{1, \dots, U\}$.
  $g_1^{(\eta)}(t) = \frac{\sum_{i,j \in C_k} p_{ij}^{(\eta)}}{\delta t_k \sum_{i<j} p_{ij}^{(\eta)}}$ for t in the k^{th} temporal bin.
  $h^{(\eta)}(r) = \frac{\sum_{i,j \in D_k} p_{ij}^{(\eta)}}{\delta r_k \sum_{i<j} p_{ij}^{(\eta)}}$ for r in the k^{th} spatial bin.
  $p_{ij}^{(\eta+1)} = K_{u_i u_j}^{(\eta)} g_1^{(\eta)}(t_j - t_i) g_2^{(\eta)}(r_{ij})$ for $i < j$ and $p_{jj}^{(\eta+1)} = \mu_{u_j}^{(\eta)}(x_j, y_j)$.
  Normalize $p_{ij}^{(\eta+1)}$ so that $\sum_{i=1}^N p_{ij}^{(\eta+1)} = 1$ for any j.
  $\delta = \max_{i,j} \|p_{ij}^{(\eta+1)} - p_{ij}^{(\eta)}\|$ and $\eta = \eta + 1$.
end while

Algorithm 3.3 Multivariate Hawkes Process Simulation

Inputs: time-window size: T; spatial region: $S \subset \mathbb{R}^2$; background rate: $\{\gamma_u\}_{u=1}^U$; triggering matrix: $\{K_{uv}\}_{u,v=1}^U$; temporal and spatial triggering kernels: $g_1(t)$, $g_2(x, y)$.
Output: point process: $\mathbf{C} = \{(u_i, t_i, x_i, y_i)\}_{i=1}^N$.
Initialize an empty set $\mathbf{C}$ and an empty stack $\mathbf{Q}$.
Generate background events:
  Draw N_u, the number of background events of type u, from a Poisson distribution with parameter $\lambda = \gamma_u T$ for each $u \leq U$.
  Add each background event $i \leq \sum_{u=1}^U N_u$ — i.e., (x_i, y_i, t_i, u_i) — to the set $\mathbf{C}$ and the stack $\mathbf{Q}$, where (x_i, y_i, t_i) is drawn from the uniform spatiotemporal distribution over the time interval $[0, T]$ and a bounded spatial region S.
Generate triggered events:
  while $\mathbf{Q}$ is not empty **do**
    Remove the most recently added element (x_i, y_i, t_i, u_i) from the stack $\mathbf{Q}$.
    Draw N_i, the number of events triggered by event i, from a Poisson distribution with parameter $\lambda_i = \sum_{u'=1}^U K_{u_i u'}$.
    Generate events (x_k, y_k, t_k, u_k) for each $k \leq N_i$ as follows:
      Sample t_k, (x_k, y_k) and u_k according to $g_1(t - t_i)$, $g_2(x - x_i, y - y_i)$, and $P(u_k = \tilde{u}) = \frac{K_{u_i \tilde{u}}}{\sum_{v=1}^U K_{u_i v}}$, respectively.
      Add (x_k, y_k, t_k, u_k) to the set $\mathbf{C}$.
      if $t_k \leq T$ **then**
        Add the element (x_k, y_k, t_k, u_k) to the stack $\mathbf{Q}$.
      end if
  end while

(Nonparametric) with the network Hawkes model in [36][1] (Network Hawkes), the exclusively temporal Hawkes model with kernel $g(t) = \omega \exp(-\omega t)$ from [20] (Temporal), and the parametric spatiotemporal model detailed in subsection 3.1 (Parametric). We make comparisons by examining how well the following properties are recovered in the inferred triggering matrix: (1) the symmetry and reciprocity; (2) existence of edges; and (3) community structures. We also demonstrate the ability of our algorithm to infer the triggering kernel g. In subsection 4.4, we investigate the network of crime events using a violent gang-crime data set to demonstrate an additional utility of our model.

4.1. Synthetic Data. We first generate synthetic triggering matrices $\mathbf{K}$ using a weighted stochastic block model (WSBM) [2,52]. We assign a network's nodes to four sets and assign edges to adjacency-matrix blocks based on the set memberships of the nodes. Two of the communities, i.e. sets, consist of ten nodes each, and the other two sets consist of five nodes each. For each edge, we first draw a Bernoulli random variable to determine whether it exists, and we then draw an exponential random variable to determine the weight of the edge (if it exists). The parameter of the Bernoulli random variable is 0.68 for an edge between nodes from the same community and 0.2 for an edge between nodes from different communities. The parameter for the exponential random variable in these two situations is 0.1 and 0.01, respectively. We design the triggering matrices to be symmetric.

The triggering matrices that we generate in this way are not guaranteed to satisfy the stability condition for Hawkes processes, $|\lambda_{\max}(\mathbf{K})| < 1$ [15]. Here $\lambda_{\max}(\mathbf{K})$ is the largest-in-magnitude eigenvalue of $\mathbf{K}$. When this condition is satisfied, each event has finitely many subsequent events as "offsprings" almost surely. In our work, we discard any simulated adjacency matrix that does not satisfy the stability condition, and we generate a new one to replace it. With these choices of the parameters, we discard about 65% of the generated adjacency matrices.

map: 65% seems like a lot

With such triggering matrix $\mathbf{K}$, we use Algorithm 3.3 to simulate multivariate spatiotemporal Hawkes processes with our parametric model in subsection 3.1 with $\omega = 0.6$, $\sigma^2 = 0.3$, $T = 250$, $S = [0,1] \times [0,1]$, and a uniform $\gamma_u = 0.2$ for all nodes u. We then reconstruct the underlying networks and the triggering kernels from the simulated data.

4.1.1. Symmetry and Reciprocity. As we noted in subsection 4.1, our simulated triggering matrices are symmetric, but our estimations generally are not symmetric. Measuring deviation from symmetry gives one way to evaluate the performance of our inference methods. We use various reciprocity measures to quantify such deviation.

We conduct two sets of experiments. In the first one, we fix a single synthetic triggering matrix and simulate ten point processes. We then estimate the triggering matrix $\mathbf{K}$ from each point process using various methods, which we thereby compare with each other. In the second set of experiments, instead of fixing a single triggering matrix, we generate ten different triggering matrices using the same WSBM model and parameters, and we simulate one point process for each triggering matrix.

There is no standard way of measuring reciprocity in a weighted network. In

[1] We use the code provided by the authors of [36]; it is available at https://github.com/slinderman/pyhawkes. In all of our experiments, we use the default hyperparameters that come with the published code unless otherwise specified.

TABLE 1

Reciprocity of the triggering matrices estimated with different methods. We report the mean and standard deviation (in parentheses) over ten simulations with the same ground-truth triggering matrix).

	Nonparametric	Temporal	Parametric	Network Hawkes
R_1	0.59 (0.05)	0.29 (0.06)	0.54 (0.03)	0.36 (0.03)
Correlation	0.84 (0.05)	0.36 (0.16)	0.79 (0.05)	0.30 (0.14)
Ratio	0.55 (0.02)	0.37 (0.11)	0.58 (0.02)	0.32 (0.02)
Coherence	0.75 (0.01)	0.63 (0.03)	0.71 (0.02)	0.68 (0.02)
Entropy	0.71 (0.01)	0.59 (0.03)	0.68 (0.02)	0.60 (0.02)

TABLE 2

Reciprocity of the triggering matrices estimated with different methods. We report the mean and standard deviation (in parentheses) over ten simulations, each with a different triggering matrix.

	Nonparametric	Temporal	Parametric	Network Hawkes
R_1	0.61 (0.12)	0.36 (0.12)	0.55 (0.10)	0.40 (0.05)
Correlation	0.81 (0.16)	0.48 (0.27)	0.76 (0.15)	0.23 (0.14)
Ratio	0.63 (0.04)	0.43 (0.06)	0.62 (0.03)	0.33 (0.03)
Coherence	0.78 (0.04)	0.62 (0.03)	0.72 (0.03)	0.70 (0.03)
Entropy	0.75 (0.05)	0.58 (0.03)	0.69 (0.03)	0.62 (0.04)

our calculations, we use scores that were proposed in [63] and [3]. First, as in [63], we compute the reciprocated edge weight $K_{uv}^{\leftrightarrow} = \min\{K_{uv}, K_{vu}\}$, and we then calculate a network-level reciprocity score R_1 as the ratio between the total reciprocated weight $W^{\leftrightarrow} = \sum_{u \neq v} K_{uv}^{\leftrightarrow}$ and the total weight $W = \sum_{u \neq v} K_{uv}$. That is, the "reciprocity" is $R_1 := W^{\leftrightarrow}/W$. Second, Akoglu et al. [3] proposed three node-level measures of reciprocity: (1) the "ratio" $R_{\text{ratio}} := \min\{K_{uv}, K_{vu}\}/\max\{K_{uv}, K_{vu}\}$; (2) "coherence" $R_{\text{coher}} = 2\sqrt{K_{uv}K_{vu}}/(K_{uv} + K_{vu})$; and (3) "entropy" $R_{\text{entropy}} := -r_{uv}\log_2(r_{uv}) - r_{vu}\log_2(r_{vu})$, where $r_{uv} = K_{uv}/(K_{uv} + K_{vu})$. These last three measures of reciprocity are measured at a node level, whereas R_1 is a network-level measure. For the other measures, we obtain a network-level measure by calculating those scores for each pair of nodes and then taking a mean over all pairs of nodes. Each of the above quantities gives a score between zero and one. In a perfectly symmetric and reciprocated network, each of the four methods gives a value one, where a larger value indicates a stronger tendency for the nodes in a network to reciprocate.

In Table 1, we report the mean reciprocity and the standard deviation over ten simulations with the same triggering matrix. In Table 2, we report the mean results from ten different triggering matrices. Both spatiotemporal models give higher scores than the exclusively temporal models, which is what we expected, as the temporal models discard spatial information. According to these measures of success, the nonparametric model has the best performance.

4.1.2. Edge Reconstruction. We also evaluate the reconstruction methods based on the ability to recover the existence of edges. This is particularly relevant if we want to know whether there is a connection between two entities.We will discuss this application on the Gowalla data set in detail (see subsection 4.2).

In our model, we consider an edge exists if the corresponding weighted entry in the estimated triggering matrix exceeds a certain threshold. For different threshold levels, we compute the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for a given ground-truth triggering matrix. We summarize our results in a receiver operating characteristic (ROC) plot (see Figure 1),

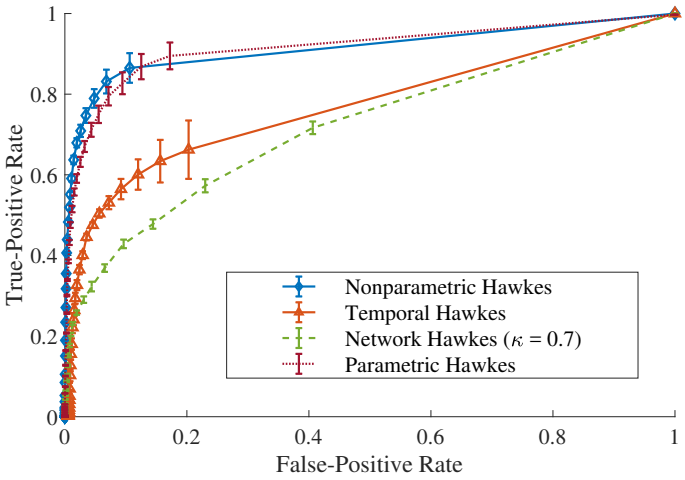


FIG. 1. *Model comparison using synthetic networks. We show the mean ROC curves with error bars (averaged over ten simulations, each with a different triggering matrix) on edge reconstruction. The ROC curve of a better reconstruction should be closer to one for a larger range of horizontal-axis values, such that it has a larger area under the curve (AUC), which is equal to the probability that a uniformly-randomly chosen existing edge in a ground-truth network has a larger weight than a uniformly-randomly chosen missing edge in the estimated network.*

in which we plot the true-positive rate (TPR) (where $\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$) versus the false-positive rate (FPR) (where $\mathrm{FPR} = \mathrm{FP}/(\mathrm{FP}+\mathrm{TN})$). A better estimation of the triggering matrix should give a higher TPR at the same FPR.

Based on the ROC plot in Figure 1, we conclude that the spatiotemporal models (both parametric and nonparametric Hawkes models proposed in section 3) outperform the exclusively temporal ones. Therefore, incorporating spatial information improves the quality of the reconstructed binary networks, at least according to this measure of success. The best results are from our parametric model, which is not surprising, given that we use the same model to simulate the data. The performance of our nonparametric model is very close to the parametric model, confirming its effectiveness at determining edge existence.

4.1.3. Estimated Kernels. We report the estimated kernels of the different models in Figure 2. Recall that the ground-truth kernels that we use to simulate point processes are $g_1(t) = \omega \exp(-\omega t)$ and $h(r) = 2\pi r g_2(x,y) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)$, where $r^2 = x^2 + y^2$, $\omega = 0.6$ and $\sigma^2 = 0.3$. If we let $\hat{g}_1$ and $\hat{h}$ denote the estimated temporal and spatial kernel, respectively, we can measure the L_1 errors: $\int |g_1(t) - \hat{g}_1(t)|\,\mathrm{d}t$ and $\int |h(r) - \hat{h}(r)|\,\mathrm{d}r$. We report L_1 errors in Table 3 and present visualizations of the kernels in Figure 2. As expected, both spatiotemporal models give more accurate kernel estimation than the exclusively temporal model. The nonparametric model does not use any information about the ground-truth kernels. Superisingly, it still manages to estimate the spatial triggering kernel more accurately, in terms of L_1 error, than the parametric model whose kernel shares the same parametric form as the ground-truth kernel.

4.1.4. Community-Structure Recovery. Network community structure, in which dense sets of nodes are connected sparsely to other dense sets of nodes, is a

TABLE 3

L_1 errors of estimated spatial and temporal kernels. We simulate ten point processes with the same triggering matrix and triggering kernel. We report the mean and standard deviation (in parentheses) of the L_1 errors averaged over the ten simulations with the same triggering kernel and matrix. Note that the exclusively temporal model does not estimate a spatial kernel.

	Nonparametric	Temporal	Parametric
Temporal kernel	0.07 (0.02)	0.20 (0.06)	0.02 (0.02)
Spatial kernel	0.06 (0.02)	-	0.12 (0.02)

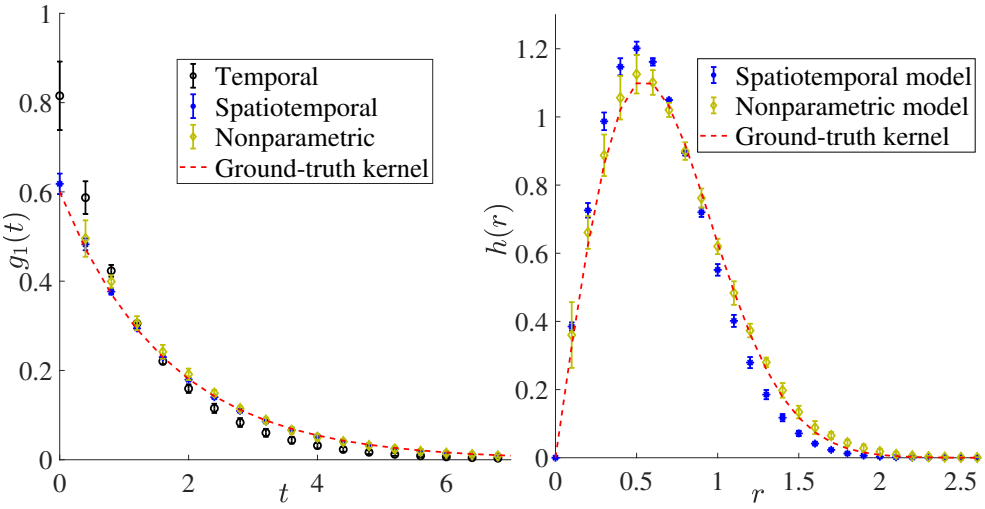


FIG. 2. *Model comparison on synthetic networks: Estimated (left) temporal and (right) spatial kernels.*

very popular topic in network science [18, 56]. We also evaluate the quality of the inferred network by its community structure. We apply the community-detection methods from [2] (a WSBM used to generate the synthetic adjacency matrices), [31] (symmetric non-negative matrix factorization; NMF), and [25, 44, 46, 48] (modularity maximization[2]) on the estimated weighted network to infer a community structure. To evaluate our inferred community structure, we use the square-root variant of the *normalized mutual information* (NMI) [65] between the inferred community assignment and the ground-truth community labels. Specifically, Let S_1 and S_2 be community membership assignments that assign the U nodes into one of C_1 and C_2 communities, respectively; and let $S_{\ell k}$, with $\ell \in \{1, 2\}$ and $k \in \{1, 2, \cdots, C_\ell\}$, denote the set of nodes in the kth community in assignment S_ℓ. The NMI between S_1 and S_2 is

$$\mathrm{NMI}(S_1, S_2) = \frac{I(S_1, S_2)}{\sqrt{H(S_1)H(S_2)}} \in [0, 1]\,,$$

where $I(S_1, S_2) = \sum_{i=1}^{C_1} \sum_{j=1}^{C_2} \frac{|S_{1i} \cap S_{2j}|}{U} \log \frac{|S_{1i} \cap S_{2j}|/U}{|S_{1i}||S_{2j}|/U^2}$ (where $|\cdot|$ of a set denotes its cardinality) and the entropy is $H(S_\ell) = -\sum_{i=1}^{N_\ell} \frac{|S_{\ell i}|}{N} \log \frac{|S_{\ell i}|}{N}$ (with $\ell \in \{1, 2\}$). Intuitively, NMI measures the amount of information shared by two community as-

[2]For modularity maximization, we use the implementation of a (locally greedy) Louvain-like [8] method (called GENLOUVAIN) from [25] with the default resolution parameter one.

TABLE 4

Normalized mutual information (NMI) of outputs of different community-detection methods applied on the estimated triggering matrices (averaged over ten simulations, each with a different triggering matrix).

	Nonparametric	Temporal	Parametric	Network Hawkes
Weighted SBM	0.80	0.38	0.83	0.36
Symmetric NMF	0.62	0.31	0.66	0.19
Modularity Maximization	0.64	0.47	0.71	0.28

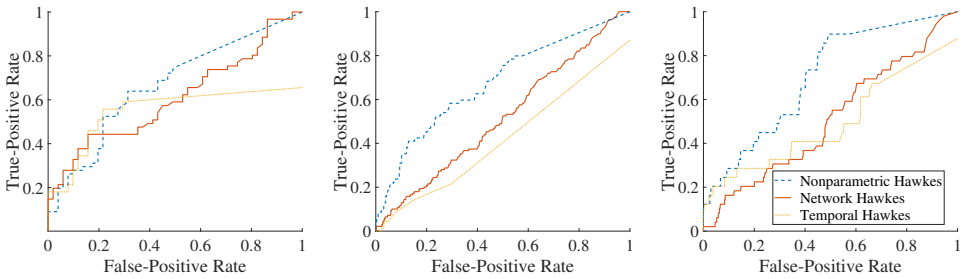
signments. If they are the same after permuting community labels, the NMI is one. A larger NMI score implies that the inferred community assignment shares more information with the ground-truth labels. See [68] for a discussion of other approaches for comparing different community assignments in networks.

There are numerous approaches for detecting communities in networks [18, 52, 56], and we choose methods with readily-available code. According to Table 4, all community-detection methods that we examine perform better with triggering matrices inferred via incorporating both spatial and temporal information than with those inferred with exclusively temporal information. One can, of course, repeat our experiments using other methods.

4.2. Gowalla Friendship Network. Gowalla is a location-based social-media website in which users share their locations by checking in. We use a Gowalla data set — collected in [12] using Gowalla's public API — of a "friendship" network with 196,591 users, 950,327 edges, and a total of 6,442,890 check-ins of these users between February 2009 and October 2010. The data set also includes the latitude and longitude coordinates and the time (with a precision of a second) of each check-in. Similar to a Facebook "friendship", the Gowalla friendship network is undirected. The mean number of friends for each user is ten, the median is three, and the maximum is $14,730$. We study several subnetworks within the Gowalla data set; see Appendix A for details. We view the spatiotemporal check-ins of users within each subnetwork as events in a multivariate point process and infer relationships between the Gowalla users.

We compare our nonparametric model with the network Hawkes model (with $\kappa = 0.9$) and the exclusively temporal Hawkes model in terms of how well the reconstructed edges match the Gowalla friendships. Because a Gowalla friendship network is undirected, we first symmetrize the inferred triggering matrix (via $\left(\mathbf{K} + \mathbf{K}^T\right)/2$) to obtain an undirected network. We then calculate FPRs and TPRs in the same fashion as subsection 4.1.2 given its associated Gowalla friendship network and generate the corresponding ROC curves. From the ROC curves of three different cities in Figure 3, the best results are from our nonparametric model that incorporates spatial information. Moreover, we obtain mean AUCs of 0.4277 (with a standard deviation of 0.1042) for the temporal Hawkes model; 0.5301 (with a standard deviation of 0.0585 for the network Hawkes model; and 0.6692 (with a standard deviation of 0.0421) for our nonparametric model over the subnetworks in Appendix A.

4.3. Crime-Topic Network. In a recent work on crime classification, Kuang et al. [29] performed topic modeling (see [30] for a review) on short narrative (i.e., text) descriptions of all crimes reported to the Los Angeles Police Department (LAPD) between 1 January 2009 and 19 July 2014 with spatial coordinates and time stamps (with a precision of a minute). The premise in their work was that crime topics better reflect the ecological circumstances of crime than standard crime classifications based



(a) San Fransisco (b) New York City (c) Los Angeles

FIG. 3. *ROC curves of different methods on reconstructing three Gowalla friendship networks. (See Appendix A for details about the networks.)*

on legal codes. Targeting discovery of up to 20 topics, they found six topics related to violent crime, eight topics related to property crime, and six topics that seem to be related to deception-based crime. This classifies the 20 crime topics into three classes. They also found overlap in the words shared between crime topics, allowing them to quantify static relationships among crime types using cosine similarity and average linkage clustering.

In the present case study, we extend this work by modeling the above data set as a crime-topic network. Each crime topic is a node, and we infer edges based on whether crime events of one topic trigger events of other topics. That is, we discover latent relationships between different crime topics based on associated crime events. Inspired by previous research on point process models of crime events [42], we model crime events of different topics via a multivariate point process and infer connections between the crime topics using our nonparametric model. To evaluate our method, we compare the communities detected in the reconstructed network with the three crime classes found in [29].

4.3.1. Community Detection. We infer crime-topic networks directly from crime events within each of the LA neighborhoods[3] using our nonparametric model, the parametric model, and the exclusively temporal model. We investigate the 100 neighborhoods with the most crime events among all 296 neighborhoods of LA. On average, there are 4,140 crime events in the top 100 neighborhood and 8,750 in the top-10. We then apply the community-detection methods mentioned in subsection 4.1.4 to the reconstructed networks assigning crime topics to communities. We quantify the difference between these community assignments and the crime topic classifications from [29] with NMI scores. We also visualize the crime-topic networks of Westwood and Wingfoot neighborhood in Figure 4, located in West and South LA respectively. From Table 5, we see that using spatial information combined with nonparametric kernel leads to the best mean NMI score among the methods that we examine.

4.4. Network of Crime Events. In section 3, we constructed a matrix $\mathbf{P}$, where $\mathbf{P}(i,j) = p_{ij}$ is the probability that event j is triggered by event i and $\mathbf{P}(i,i) = p_{ii}$ is the probability that event i is a background event. We now define an event

[3]We use the Zillow neighborhood boundaries from https://www.zillow.com/howto/api/neighborhood-boundaries.htm.

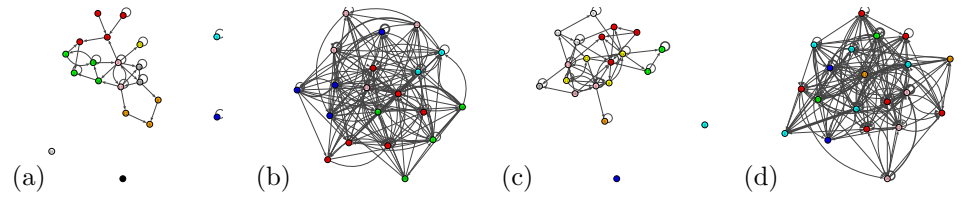


FIG. 4. *Network generated by nonparametric and temporal Hawkes models colored by community-detection results from modularity maximization: (a) nonparametric model in Westwood, (b) temporal model in Westwood, (c) nonparametric model in Wingfoot, and (d) temporal model in Wingfoot.*

TABLE 5

Mean NMI (with one standard deviation reported in parantheses) between community assignments from several community-detection methods and the classifications from [29] over the 100 neighborhoods in Los Angeles with the most recorded crime events between 1 January 2009 and 19 July 2014.

	Nonparametric	Temporal	Parametric
Symmetric NMF	0.25 (0.11)	0.12 (0.084)	0.084 (0.12)
Weighted SBM	0.24 (0.12)	0.085 (0.086)	0.078 (0.079)

network, which is both weighted and directed, in which each event is a node and $\mathbf{P}$ is the adjacency matrix. The weight of an edge reflects the triggering effects between two events, and the direction points from the earlier event to the later one.

4.4.1. Stochastic Declustering. With an event network, a natural question is whether one can differentiate between "true" background events and triggered events. Such differentiation using the probability p_{ii} is called *stochastic declustering* [72]. To determine whether event i is a background event, we compare p_{ii} with a uniformly random sample from the interval $(0, 1)$. If p_{ii} is greater than the random number, we consider this event to be from the background; otherwise, we consider it to be triggered by other events.

We perform the declustering experiment on the synthetic data set; we simulate ten synthetic point processes using a fixed triggering matrix (see subsection 4.1 for details). Recall from Algorithm 3.3 that we retain the causality information in the simulations (i.e., which event causes which and which events are from the background), giving a notion of "ground truth" about ancestors of each event. The quality of declustering is often measured by comparing the estimated branching ratio [62] with the ground-truth one. The branching ratio is defined as $1-\frac{N_b}{N}$, where N_b is the number of background events. However, the difference in branching ratios itself typically does not completely reflect reconstruction errors. For example, in an extreme case, stochastic declustering can erroneously misclassify some number of background events as triggered and the same number of triggered events erroneously as background, although the branching ratio is the same as the true branching ratio in this scenario. To resolve this problem, we view declustering as a binary classification problem that assigns events to be either background or triggered events. We use measurements such as recall and precision to evaluate our declustering results. Recall that "recall" is the ratio between the number of background events that are correctly recovered by the declustering methods (i.e., the true positives) to the total number of background events, and "precision" is the ratio between the number of true positives to the number of events that are labeled as background events by stochastic declustering. From the

TABLE 6

Comparison of stochastic declustering results. We report the mean and the standard deviation (in parentheses) of branching-ratio error, precision, and recall over ten simulations, i.e, ten point processes with the same triggering kernels and matrix. For each simulation, each measure is the mean over 20 *runs of stochastic declustering.*

	Nonparametric	Parametric	Temporal
Branching-ratio error	0.039 (0.0050)	0.01 (0.011)	0.022 (0.019)
Recall	0.75 (0.0098)	0.65 (0.027)	0.60 (0.035)
Precision	0.70 (0.0082)	0.64 (0.0093)	0.59 (0.0086)

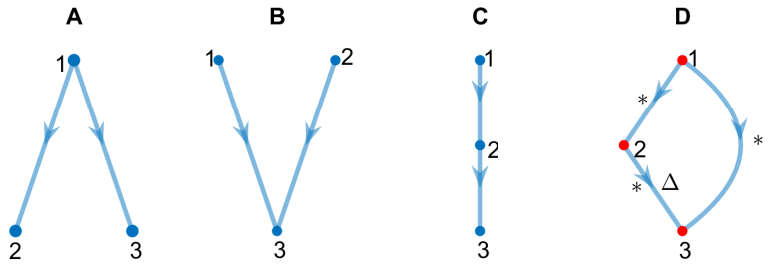


FIG. 5. *All possible three-node motifs for a DAG. We highlight the nodes in the feedforward-loop motif (D) in red.*

results in Table 6, we see the exclusively temporal model has the worst performance among the methods that we consider. Our nonparametric model has the best recall and precision (with the smallest variations as well), whereas the parametric model has the smallest branching-ratio error.

4.4.2. Motif Analysis. Declustering methods differentiate between background and triggered events in an event network. To further examine spatiotemporal dynamics, we consider causality information among events. Similar to a relational-event model [11], one can obtain causality information from the matrix $\mathbf{P}$, because p_{ij} is the probability that event j is triggered by event i. We focus on repeated patterns — from a network-science perspective, network motifs [40], which are recurrent and statistically signification patterns in a network — to obtain information about local causality structure. Note that all event networks are directed acyclic graphs (DAGs) because of their orientation in time.

Motif analysis is particularly interesting in gang-crime event networks. Gang crimes are often characterized by retaliations (triggered crime events) among rivalry gangs, which can lead to a series of tit-for-tat reciprocal crimes. To find significant gang retaliation patterns, we use a gang-crime data set provided by the LAPD from 2014–2015 with 4,158 events in Los Angeles. We generate an event network with our nonparametric model. We find that a three-node feedforward-loop motif [37] is significant in the city-wide data set with a z-score 5.1 ± 2.4 and in the South LA area-only data set (1,912 events) with z-score 5.5 ± 2.5. Davies and Marchione [16] found the same three-node motif significant using different methods of network construction in data sets from maritime piracy and residential burglaries.

We focus on the South LA area, because it is the center of a gang intervention program. **map: is there a REF to a newspaper article or something about this program that we can include for the sentence above?** Establishing which

causal structures are statistically significant has important implications for countering gang violence. Fast response to a gang crime may reduce the potential that it triggers a future retaliation. Knowing that feedforward-loop network motifs occur at rates that differ from chance suggests that disrupting retaliation may require assessment of trade-offs in how to allocate intervention resources. For example, in a simple triggering chain (see Figure 5C), one can expect that intervention following an initial triggering event in time will have a direct effect on the second event in time and an indirect effect on the third event, although the effect on the third event may be attenuated by the intervening event. By contrast, we expect that intervention following the first event in the feedforward-loop motif (see Figure 5D) will have a direct effect on the second event in time and both a direct and indirect effect on the third event. It is possible that the third event is more likely to be disrupted given the feedforward structure and intervention following the first event than would be the case with direct intervention following only the second event (see the triangle in Figure 5D).

5. Conclusions and Discussion. In this paper, we used point-process models to infer latent networks from synthetic and real-world spatiotemporal data sets. We then applied tools from network analysis to examine the inferred networks. We studied the role of spatial information and nonparametric techniques in network reconstruction.

Our proposed model incorporates spatial information. However, using such information requires making a good choice of spatiotemporal triggering kernels. We achieved this using a nonparametric approach. Through experiments on synthetic data sets, we showed that our nonparametric model is capable of recovering the spatial and temporal triggering kernels. Moreover, our approach is able to infer a network structure that better recovers — compared to other network reconstruction methods that we studied — symmetry and reciprocity, edge reconstruction, and community structures. Through experiments on real-world data sets, we illustrated that the inferred networks of our approach are meaningful, in the sense that they have large positive correlations with some metadata. For instance, the inferred networks from the Gowalla data set are more similar to the Gowalla friendship network than those inferred with exclusively temporal information when the similarity is measured by ROC curves. The inferred crime-topic networks has a community structure similar to that discovered in [29].

This paper fills the current gaps [57] on incorporating spatial information in multivariate self-exciting point processes and can be applied to other fields such as seismology. Moreover, it is not limited to the Euclidean distance (for the spatial variables) that is used commonly in seismology [51] and crime applications [42]. In other words, the spatial triggering kernel $g_2(r)$ is currently a function of Euclidean distance, but one can potentially use any notion of "distance" between two entities. For example, in a network, one can measure a distance between two entities by the length of shortest paths between them. In a recent paper, Green et al. [22] proposed a social-contagion model in which they assumed, using a parametric form, that the strength of triggering in a Hawkes-process model depends on the shortest-path distance. With our nonparametric approach, we can nonparametrically estimate such dependence. To give another example, consider a point process in which each event is associated with textual information. For instance, in a Twitter data set, one can consider each tweet (a time-stamped body of text) as an event in a point process. One can measure a distance between two tweets based on their text.

Our network reconstruction method has its limitations. It uses $O(U^2)$ parameters

for U nodes. To avoid underfitting, it needs a large number of observed events. The computational complexity and memory requirement scale at least quadratically with the number of events, so the current EM-type algorithm is not ideal for analyzing large data sets. It will thus be important to improve our inference method for network reconstruction.

Appendix A. Preprocessing of the Gowalla Data.

In this section, we detail how we preprocess the Gowalla data that were collected and studied in [12]. We examine data from three cities: New York City, Los Angeles, and San Fransisco. We visualize the networks used in this paper in Figure 6.

A.1. New York City (NYC). We study check-ins in New York City (NYC) during the period April–October 2010. We use a bounding box (with a north latitude of 40.92, a south latitude of 40.48, an east longitude of -73.70, and a west longitude of -74.26[4] to locate check-ins that occur in NYC. We consider "active" users, who have at least 100 check-ins during the period. To alleviate the computational burden, we also only consider users who have at most 500 check-ins during the period to reduce the number of users and the total number of check-ins. Our inference process requires computing a triggering probability for each pair of events/check-ins, which results in a full upper-triangular matrix, with the number of entries given by the square of the total number of events (i.e., check-ins), so the memory requirement scales quadratically with the number of events. We only perform experiments for cases in which the total number of events is at most 10,000, so that we are able to store triggering probabilities for all pairs of events in 4-gigabyte memory. There are 5,801 unique users with at least one check-in in NYC during the period, and there are 101,329 check-ins in total. After removing "inactive" users (i.e., those with strictly fewer than 100 check-ins) and overly active users (i.e., those with strictly more than 500 check-ins), we are left with 160 users and a total of 29,118 check-ins. We also restrict ourselves to users in the largest connected component (LCC) of the network. This yields 46 users and 8495 check-ins, on which we apply our inference methodology.

A.2. Los Angeles (LA). We apply the same procedure as in Appendix A.1 on the check-in data for Los Angeles (LA). The bounding box used for LA has a north latitude of 34.34, a south latitude of 33.70 , an east longitude of -188.16, and a west longitude of -188.67. We restrict the area of LA to be the same as that of NYC, although LA's geographic area is much larger than that of NYC. After selecting only users in the LCC of the Gowalla network among users who are active (with at least 150 check-ins) but not overly active (with at most 1000 check-ins) users, we are left with 23 users and 6,203 check-ins.

A.3. San Francisco (SF). To look at a different type of example, we also examine the 1-ego network of the most popular user (with 14 friends) in San Francisco (SF). (A 1-ego network [69] of a node is an induced subgraph that includes a focal node — the ego — and its direct neighbors.) The bounding box used for SF has a north latitude of 37.93, a south latitude of 37.64, an east longitude of -122.28, and a west longitude of -123.17. In this 1-ego network, there are $9,887$ check-ins.

Acknowledgements. We thank Chandan Dhal and Jialin Liu for helpful discussions and preliminary work on community detection.

[4]We obtain latitude and longitude coordinates from http://www.mapdevelopers.com/geocode_bounding_box.php.

(a) 1-Ego network of a user of Gowalla in SF.

(b) Largest connected component of the Gowalla network in NYC.

(c) Largest connected component of the Gowalla network in LA.

FIG. 6. *Three different friendships networks in the Gowalla data set. We compare different network reconstruction methods on these networks.*

REFERENCES

[1] M. ACHAB, E. BACRY, S. GAÏFFAS, I. MASTROMATTEO, AND J.-F. MUZY, *Uncovering causality from multivariate hawkes integrated cumulants*, Journal of Machine Learning Research, 18 (2018), pp. 1–28.

[2] C. AICHER, A. Z. JACOBS, AND A. CLAUSET, *Learning latent block structure in weighted networks*, Journal of Complex Networks, 3 (2014), pp. 221–248.

[3] L. AKOGLU, P. O. V. DE MELO, AND C. FALOUTSOS, *Quantifying reciprocity in large weighted communication networks*, in Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2012, pp. 85–96.

[4] E. BACRY, I. MASTROMATTEO, AND J.-F. MUZY, *Hawkes processes in finance*, Market Microstructure and Liquidity, 1 (2015), p. 1550005.

[5] D. BALCAN, V. COLIZZA, B. GONÇALVES, H. HU, J. J. RAMASCO, AND A. VESPIGNANI, *Multiscale mobility networks and the spatial spreading of infectious diseases*, Proceedings of the National Academy of Sciences of the United States of America, 106 (2009), pp. 21484–21489.

[6] H. BARBOSA, M. BARTHELEMY, G. GHOSHAL, C. R. JAMES, M. LENORMAND, T. LOUAIL, R. MENEZES, J. J. RAMASCO, F. SIMINI, AND M. TOMASINI, *Human mobility: Models and applications*, Physics Reports, (2018).

[7] M. BARTHELEMY, *Morphogenesis of Spatial Networks*, Springer-Verlag, 2018.

[8] V. D. BLONDEL, J.-L. GUILLAUME, R. LAMBIOTTE, AND E. LEFEBVRE, *Fast unfolding of communities in large networks*, Journal of Statistical Mechanics: Theory and Experiment, 2008 (2008), p. P10008.

[9] D. BROCKMANN, L. HUFNAGEL, AND T. GEISEL, *The scaling laws of human travel*, Nature, 439 (2006), pp. 462–465.

[10] E. N. BROWN, R. E. KASS, AND P. P. MITRA, *Multiple neural spike train data analysis: state-of-the-art and future challenges*, Nature neuroscience, 7 (2004), p. 456.

[11] C. T. BUTTS, *A relational event framework for social action*, Sociological Methodology, 38 (2008), pp. 155–200.

[12] E. CHO, S. A. MYERS, AND J. LESKOVEC, *Friendship and mobility: user movement in location-based social networks*, in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2011, pp. 1082–1090.

[13] M. D. CONOVER, C. DAVIS, E. FERRARA, K. MCKELVEY, F. MENCZER, AND A. FLAMMINI, *The geospatial characteristics of a social movement communication network*, PloS ONE, 8 (2013), p. e55957.

[14] N. CRESSIE, *Statistics for Spatial Data*, John Wiley & Sons, 2015.

[15] D. J. DALEY AND D. VERE-JONES, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, Springer Science & Business Media, 2007.

[16] T. DAVIES AND E. MARCHIONE, *Event networks and the identification of crime pattern motifs*, PloS ONE, 10 (2015), p. e0143638.

[17] M. EICHLER, R. DAHLHAUS, AND J. DUECK, *Graphical modeling for multivariate Hawkes processes with nonparametric link functions*, Journal of Time Series Analysis, 38 (2017), pp. 225–242.

[18] S. FORTUNATO AND D. HRIC, *Community detection in networks: A user guide*, Physics Reports,

659 (2016), pp. 1–44.
[19] E. W. Fox, F. P. Schoenberg, and J. S. Gordon, *Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences*, The Annals of Applied Statistics, 10 (2016), pp. 1725–1756.
[20] E. W. Fox, M. B. Short, F. P. Schoenberg, K. D. Coronges, and A. L. Bertozzi, *Modeling e-mail networks and inferring leadership using self-exciting point processes*, Journal of the American Statistical Association, 111 (2016), pp. 564–584.
[21] C. W. J. Granger, *Investigating causal relations by econometric models and cross-spectral methods*, Econometrica: Journal of the Econometric Society, (1969), pp. 424–438.
[22] B. Green, T. Horel, and A. V. Papachristos, *Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014*, JAMA Internal Medicine, 177 (2017), pp. 326–333.
[23] E. C. Hall and R. M. Willett, *Tracking dynamic point processes on networks*, IEEE Transactions on Information Theory, 62 (2016), pp. 4327–4346.
[24] P. Holme, *Modern temporal network theory: A colloquium*, Eur. Phys. J. B, 88 (2015), 234.
[25] L. G. Jeub, M. Bazzi, I. S. Jutla, and P. J. Mucha, *A generalized Louvain method for community detection implemented in* matlab, (2011–2017; Version 2.1.1), http://netwiki.amath.unc.edu/GenLouvain.
[26] M. Karsai, H.-H. Jo, and K. Kaski, *Bursty Human Dynamics*, Briefs in Complexity, Springer-Verlag, 2018.
[27] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *Multilayer networks*, Journal of Complex Networks, 2 (2014), pp. 203–271.
[28] M. Kivelä and M. A. Porter, *Estimating interevent time distributions from finite observation periods in communication networks*, Physical Review E, 92 (2015), p. 052813.
[29] D. Kuang, P. J. Brantingham, and A. L. Bertozzi, *Crime topic modeling*, Crime Science, 6 (2017), p. 12.
[30] D. Kuang, J. Choo, and H. Park, *Nonnegative matrix factorization for interactive topic modeling and document clustering*, in Partitional Clustering Algorithms, Springer, 2015, pp. 215–243.
[31] D. Kuang, C. Ding, and H. Park, *Symmetric nonnegative matrix factorization for graph clustering*, in Proceedings of the 2012 SIAM International Conference on Data Mining, SIAM, 2012, pp. 106–117.
[32] E. L. Lai, D. Moyer, B. Yuan, E. Fox, B. Hunter, A. L. Bertozzi, and P. J. Brantingham, *Topic time series analysis of microblogs*, IMA Journal of Applied Mathematics, 81 (2016), pp. 409–431.
[33] S. L. Lauritzen, *Graphical Models*, vol. 17, Clarendon Press, 1996.
[34] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, *Computational social science*, Science, 323 (2009), pp. 721–723.
[35] E. Lewis and G. O. Mohler, *A nonparametric EM algorithm for multiscale Hawkes processes*, Journal of Nonparametric Statistics, 1 (2011), pp. 1–20.
[36] S. Linderman and R. Adams, *Discovering latent network structure in point process data*, in International Conference on Machine Learning, 2014, pp. 1413–1421.
[37] S. Mangan and U. Alon, *Structure and function of the feed-forward loop network motif*, Proceedings of the National Academy of Sciences of the United States of America, 100 (2003), pp. 11980–11985.
[38] B. Mark, G. Raskutti, and R. Willett, *Network estimation from point process data*, arXiv preprint arXiv:1802.04838, (2018).
[39] D. Marsan and O. Lengline, *Extending earthquakes' reach through cascading*, Science, 319 (2008), pp. 1076–1079.
[40] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Network motifs: simple building blocks of complex networks*, Science, 298 (2002), pp. 824–827.
[41] G. O. Mohler, *Marked point process hotspot maps for homicide and gun crime prediction in Chicago*, International Journal of Forecasting, 30 (2014), pp. 491–497.
[42] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, *Self-exciting point process modeling of crime*, Journal of the American Statistical Association, 106 (2011), pp. 100–108.
[43] G. O. Mohler, M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. Brantingham, *Randomized controlled field trials of predictive policing*, Journal of the American Statistical Association, 110 (2015), pp. 1399–1411.
[44] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community structure in time-dependent, multiscale, and multiplex networks*, Science, 328 (2010),

pp. 876–878.
[45] J. A. Nelder and R. J. Baker, *Generalized linear models*, Wiley Online Library, 1972.
[46] M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices*, Physical Review E, 74 (2006), p. 036104.
[47] M. E. J. Newman, *Networks*, Oxford University Press, 2018.
[48] M. E. J. Newman and M. Girvan, *Finding and evaluating community structure in networks*, Physical Review E, 69 (2004), p. 026113.
[49] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, *A tale of many cities: Universal patterns in human urban mobility*, PloS ONE, 7 (2012), p. e37027.
[50] Y. Ogata, *Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes*, Journal of the American Statistical Association, 83 (1988), pp. 9–27.
[51] Y. Ogata, *Space-time point-process models for earthquake occurrences*, Annals of the Institute of Statistical Mathematics, 50 (1998), pp. 379–402.
[52] T. P. Peixoto, *Bayesian stochastic blockmodeling*, arXiv:1705.10225, (2018). Chapter in "Advances in Network Clustering and Blockmodeling", edited by P. Doreian, V. Batagelj, A. Ferligoj, (John Wiley & Sons, New York City, USA [forthcoming]).
[53] P. O. Perry and P. J. Wolfe, *Point process modelling for directed interaction networks*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75 (2013), pp. 821–849.
[54] M. Porter and J. Gleeson, *Dynamical Systems on Networks: A tutorial*, vol. 4 of Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Springer International Publishing, 2016.
[55] M. A. Porter and S. D. Howison, *The role of network analysis in industrial and applied mathematics*, arXiv preprint arXiv:1703.06843, (2017).
[56] M. A. Porter, J.-P. Onnela, and P. J. Mucha, *Communities in networks*, Notices of the AMS, 56 (2009), pp. 1082–1097, 1164–1166.
[57] A. Reinhart et al., *Rejoinder: A review of self-exciting spatio-temporal point processes and their applications*, Statistical Science, 33 (2018), pp. 330–333.
[58] S. Scellato, A. Noulas, and C. Mascolo, *Exploiting place features in link prediction on location-based social networks*, in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2011, pp. 1046–1054.
[59] F. P. Schoenberg, D. R. Brillinger, and P. Guttorp, *Point processes, spatial-temporal*, Encyclopedia of environmetrics, (2002).
[60] M. B. Short, P. J. Brantingham, A. L. Bertozzi, and G. E. Tita, *Dissipation and displacement of hotspots in reaction–diffusion models of crime*, Proceedings of the National Academy of Sciences of the United States of America, 107 (2010), pp. 3961–3965.
[61] A. Simma and M. I. Jordan, *Modeling events with cascades of Poisson processes*, in Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2010, pp. 546–555.
[62] D. Sornette and S. Utkin, *Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks, and aftershocks*, Physical Review E, 79 (2009), p. 061110.
[63] T. Squartini, F. Picciolo, F. Ruzzenenti, and D. Garlaschelli, *Reciprocity of weighted networks*, Scientific Reports, 3 (2013), p. 2729.
[64] 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 (2011), p. 115013.
[65] A. Strehl and J. Ghosh, *Cluster ensembles—a knowledge reuse framework for combining multiple partitions*, Journal of Machine Learning Research, 3 (2002), pp. 583–617.
[66] P. Suny, J. Li, Y. Mao, R. Zhang, and L. Wang, *Inferring multiplex diffusion network via multivariate marked hawkes process*, arXiv preprint arXiv:1809.07688, (2018).
[67] G. Tita, P. J. Brantingham, A. Galstyan, and Y.-S. Cho, *Latent self-exciting point process model for spatial-temporal networks*, Discrete and Continuous Dynamical Systems - Series B, 19 (2014), pp. 1335–1354.
[68] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *Comparing community structure to characteristics in online collegiate social networks*, SIAM Review, 53 (2011), pp. 526–543.
[69] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, *The anatomy of the Facebook social graph*, arXiv preprint arXiv:1111.4503, (2011).
[70] A. Veen and F. P. Schoenberg, *Estimation of space–time branching process models in seismology using an EM-type algorithm*, Journal of the American Statistical Association, 103 (2008), pp. 614–624.

862 [71] K. Zhou, H. Zha, and L. Song, *Learning social infectivity in sparse low-rank networks using*
863 *multi-dimensional Hawkes processes*, in Artificial Intelligence and Statistics, 2013, pp. 641–
864 649.
865 [72] J. Zhuang, Y. Ogata, and D. Vere-Jones, *Stochastic declustering of space-time earthquake*
866 *occurrences*, Journal of the American Statistical Association, 97 (2002), pp. 369–380.
867 [73] J. Zhuang, Y. Ogata, and D. Vere-Jones, *Analyzing earthquake clustering features by using*
868 *stochastic reconstruction*, Journal of Geophysical Research: Solid Earth, 109 (2004).