Title

Spatially embedded social networks: dynamic models and data reconstruction

Permalink

Author

Hegemann, Rachel Anne

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Spatially embedded social networks: dynamic models and data reconstruction

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

Rachel Anne Hegemann

2012

Abstract of the Dissertation

Spatially embedded social networks: dynamic models and data reconstruction

by

Rachel Anne Hegemann

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2012

Professor Andrea L. Bertozzi, Chair

"Bottom-up" and "top-down" identify two fundamental approaches to modeling complex systems. As the name suggests, a bottom-up approach analyzes how elements on a micro scale affect observations on the macro scale. On the other hand, top-down approaches use macro scale data to identify patterns evolved from the micro scale. This thesis details two models, agent-based and data driven, designed for complex systems. These models are applied to the complex system of street gang violence.

The first method employed is agent-based and is used to explain potential geographic influences in the formation of street gang rivalries. In this framework each agent possesses certain properties and movement rules. The agents then move and interact with the simulated environment and other agents. From these simple rules the emergent behavior of the location of interactions and the rivalry network are observed.

The second method addresses the need to infer process parameters and identify gangs involved in violent crimes in the presence of incomplete data. The violent events among gangs can be viewed as realizations of a self-exciting point process on the rivalry network. For many of the events in the data, the time of the event is known, but the rivalry association is not. Using the structure of the point process a method is proposed that simultaneously estimates the process and infers the rivalry affiliation for the unknown events. We call this

method the Estimate & Score algorithm.

There are two major findings in this dissertation. The first is that the proposed agent-based model used to simulate the gang rivalry network observed in Hollenbeck provided better results than a simplified model and a Geographic Threshold graph. The second major finding is that the Estimate & Score algorithm is a computationally efficient method that produces comparable results to previous work and is better than chance. It also successfully approximates the process parameters in the presence of incomplete data.

The dissertation of Rachel Anne Hegemann is approved.

P. Jeffrey Brantingham

Joseph M. Teran

Christopher Anderson

Andrea L. Bertozzi, Committee Chair

University of California, Los Angeles

2012

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Acknowledgments

It is a pleasure to thank those who made this thesis possible including the graduate students, departmental staff, professors, committee members, family, and friends. There have been many on the path that believed in me.

I am indebted to many of my colleagues and friends that have supported me through graduate school, particularly Nancy Rodríguez, Chris Nelsen, Kristin Becker, Laura Smith, Hayden Schaefer, James von Brecht, and Alethea Barbaro. I am eternally grateful for them cheering me on during the the various hurdles of the graduate program.

I am honored to have had the opportunity of working with such high caliber researchers at UCLA. I am grateful to my advisor Andrea Bertozzi for giving me the opportunity to work on interesting projects and for her helpful guidance. I am thankful for all that I have learned under her guidance and owe her a debt of gratitude. I would also like to thank my committee Jeff Brantingham, Joseph Teran, and Chris Anderson for their input and letters of recommendations. Further, I would like to thank Allon Percus, Martin Short, and George Mohler for their helpful comments and conversations through the duration of my time at UCLA.

There are many people at my undergraduate institution at the University of Colorado, Boulder who were instrumental in my path to the Ph.D. I am particularly grateful to my

advisors James Curry and Anne Dougherty. It was because of their encouragement I decided to pursue a Ph.D. I would also like to thank Adam Norris for persuading me to add applied math as a major.

I greatly appreciate the love and motivation I have been given by my family throughout the years. To Michelle Danson and David Plume, thank you for always having a place where I can land and for believing in me. To the belated David Danson, thank you for paving the way and for all of the valuable advice you provided when I was young. I would also like to thank my siblings and their significant others for their understanding through the past few years. Finally, I owe my deepest gratitude to my husband Jan for his infinite support, love, understanding, and encouragement.

Vita

2003-2005   Resident Adviser, University of Colorado, Boulder, Colorado.

2007   B.A. (Psychology), University of Colorado, Boulder, Colorado.

2007   B.S. (Applied Mathematics), University of Colorado, Boulder, Colorado.

2007   M.S. (Applied Mathematics), University of Colorado, Boulder, Colorado.

2007–2011   Teaching Assistant and VIGRE Fellowship, Mathematics Department, UCLA.

2011   TA Consultant, Mathematics Department, UCLA.

2012-2013   Fulbright Scholar, Germany

Publications and Presentations

Hegemann, Rachel, "Data Assimilation: Particle Filters" MSRI/NCAR Summer 2010 Graduate Workshop Mathematics of Climate Change, Workshop Project Presentation, July 12-23, 2010

Hegemann, Rachel,"Geographical Influences of an Emerging Network of Gang Rivalries" SAMSI Dynamics Of Networks Workshop Poster Presentation, January 10-12, 2011

Hegemann, Rachel, "Geographical Influences of an Emerging Network of Gang Rivalries"

IPAM Women in Mathematics Symposium Lightning Round and Poster Presentation, February 24-26, 2011

Hegemann, Rachel, "Constructing Gang Rivalry Networks: An Agent-Based Modeling Approach" Invited Speaker at CSU Channel Islands Graduate Seminar, April 6, 2011

Hegemann, Rachel, "Robust Optimal Design of Heliostat Arrays for Concentrating Solar Power Plants" Industrial Math and Stat. Modeling Workshop project presentation, Raleigh, North Carolina, July 15, 2011

Hegemann, Rachel, "Constructing Gang Rivalry Networks: An Agent-Based Modeling Approach" Invited Speaker at Cypress College AMC^2 Undergraduate Math Club, Cypress, California, Aug. 24, 2011

Hegemann, Rachel, "Constructing Gang Rivalry Networks: An Agent-Based Modeling Approach" Poster presentation for ARO MURI, Marina del Rey, California, Sept. 15 2011

Hegemann, Rachel, "Inferring gang affiliation for violent events with incomplete data" Contributed presentation and poster presentation at the MAA-AMS Joint Math Meetings, Jan. 4 2012

Rachel A. Hegemann, Laura M. Smith, Alethea B.T. Barbaro, Andrea L. Bertozzi, Shannon E. Reid, George E. Tita, "Geographical Influences of an Emerging Network of Gang Rivalries" *Physica A: Statistical Mechanics and its Applications* 390 (21-22) 2011.

Rachel A. Hegemann, Erik A. Lewis, Andrea L. Bertozzi, "An "Estimate & Score Algorithm" for Simultaneous Parameter Estimation and Reconstruction of Missing Data on Social Networks" Security Informatics Special issue Computational Criminology, *in review* 2012.

CHAPTER 1

Introduction and Preliminaries

Complex systems arise in many natural and human systems including biology, ecology, criminology and finance to name a few [Chu11]. Depending on access to data and research questions, one might take on one of two modeling approaches. The first, bottom-up modeling, determines how elements on the small scale of a complex system can affect behavior on the macro level. In cases where partial data is available on the macro and micro scales, but the relationships between these scales are not well understood, such a bottom-up approach may be useful. For example in the arena of collective behavior, the interactions on the individual level have a large impact on the observed pattern [BUK12]. The second common approach is loosely named top-down modeling. Here the underlying mechanisms of the system are approximated by using a model on the macro scale. In this modeling paradigm the focus is on factors and mechanisms affecting the system at the macro scale. Many density estimation and PDE models often fall under this category. One example of this is the estimation of gang territories via a PDE model [SBB12].

In this dissertation we narrow our focus to modeling complex networks by utilizing two broad methodologies: agent-based and data-driven models. Agent-based models provide a flexible framework to test theories on the potential genesis of emergent behavior. This bottom-up approach constructs entities, called agents, that are instilled with basic, pre-described movement and interaction behaviors. These agents are placed in a simulated environment, and then allowed to interact with one another and the environment. From these basic elements, the emergent behavior of interest is observed. This provides a simulation which can explore theories on how the actions and behaviors on microscopic level among

actors affect the emergent, often complicated, phenomena of interest.

The flexibility of agent-based modeling is a great strength of this methodology, however, a few issues arise with such a framework. Due to the ease of incorporating new elements into the model, it is tempting to include more elements than needed, thus making the model overly complex and in some sense over fitting the data. To avoid this Occam's razor should be observed.

Data-driven models describe a large class of models where the modeler incorporates data at a given level to describe the complex system. The use of data can lead to novel extensions to classical models, and can steer the model closer to observed behavior. This modeling approach is used widely by statisticians and can often lead to a deeper understanding of the system. As with agent-based models, one must find a balance between sufficiently capturing the phenomenon of interest and maintaining a simple model.

With either modeling framework, it is standard procedure to describe a system with the simplest model possible and analyze the results. Then, with each addition to the model, determine a measure to assess the improvement of the complex model. When employing a statistical model, one can consider the Akaike information criterion (AIC) to find an appropriate balance between a goodness of fit and the complexity of the model [Aka73]. The AIC is defined as

$$AIC = 2k - 2\ln(L). \tag{1.1}$$

The number of parameters is defined by k and L is the maximized value of the likelihood function.

For analysis we use the system of street gang violence as a test case, but the techniques and models described in this thesis could be applied to a larger class of problems. Street gangs are a natural choice to showcase our methods for numerous reasons. These types of social systems, though extraordinarily complex, have a number of well understood underlying mechanisms and concrete data with which to model. Mathematically modeling street gangs and their violence is a relatively new field with many opportunities to develop new methods.

Further, in areas where there are multiple gangs, it is common that rivalries will form among them, thus introducing a coupling between the activity of the gang entities with the rivalry network. Finally one cannot ignore the social impact of this topic. Street gang violence is a persistent problem in urban areas that drains police resources. Finding ways to allocate these resources intelligently can be of great help mitigating the impact of street gang violence.

This dissertation uses mathematical methodology to understand two aspects of street gangs: the formation of gang rivalries and inferring gang rivalry affiliation given an incomplete data set. Each aspect requires different mathematical techniques. When considering the formation of street gang rivalries, an agent-based model is employed. This allows for a flexible framework in which environmental factors are included in the model. The details are described in Chapter 2. This work was done in collaboration with Laura M. Smith, Alethea B.T. Barbaro, Andrea L. Bertozzi, Shannon E. Reid, and George E. Tita and was published in Physica A in [HSB11].

Once the network is established, we investigate events occurring on the network. We assume that these are realizations of a self-exciting point process and use data to estimate parameters of the process. Using Hollenbeck as case study, we relate the events with violent occurrences between rival gangs. Due to the retaliatory behavior observed, we justify the assumption that the data as coming from a self-exciting point process (for mathematical background see Section 1.3.2). As with many other sociological data sets, the data is often incomplete in that the gangs involved in these events are not always known. In this case, it is beneficial for law enforcement officials and sociological researchers to have a method to infer the appropriate parties involved in such events. This work was conducted in collaboration with Erik Lewis and Andrea Bertozzi in [HLB12].

1.1 Sociological Aspects of Modeling

A large factor in the success of modeling a social system is in finding a way to translate sociological theories into a mathematical framework. Much like in physical modeling, models

will do poorly if the fundamental sociological mechanics of the system are not incorporated.

Criminal activity lends itself to using both bottom-up and data-driven modeling approaches due to the general, well established Routine Activity Theory coined by Clarke and Felson [CF79]. Routine Activity Theory hypothesizes that crimes occur when there is 1) a motivated offender, 2) a suitable target, and 3) an absence of capable guardian against a violation [CF79, Gro07, BB93b, BB93c, Fel02]. In this framework, most crimes are thought to be the result of individuals living their daily life and committing crimes as opportunities present themselves. This is in contrast to previous theories that emphasized the motives of an individual when committing a crime [CF04]. This theory further emphasizes the role of environment in criminal activity [BB93b, BB93c].

From a modeling perspective, Routine Activity Theory enables the researcher to consider tangible data such and locations and times of criminal activity, and not focus on intangible, unquantifiable information such as motivation or "mind set" of an individual criminal. The complex system of criminal activity is contextualized in a way that lends itself to mathematical modeling. One key aspect of this theory with regards to modeling criminal phenomenon is the understanding of routine activity. By most definitions this relates the activities and locations that the potential criminal engages in on a regular basis, such as their place of residence, place of work, or even a coffee shop that they frequent. To formulate this into a mathematical model, one must be able to model normal human activity and movements at the necessary scale. Further this theory underlines the critical influence of both time and space in the production of criminal activity. These theories outline the way that circumstances on individual scale can interact to create crime on a macro scale [SW87]. For this thesis, the models created are approached from a Routine Activity Theory perspective.

Further elements of street gang activities are known and well documented. The identity of gang members is not a secret. Through tattoos, clothing, colors, and gang signs people inside and outside of a gang can identify membership. Street gangs tend to partition space into different territories. Gang members know where these territories are and tend to avoid the territories of rival gangs [And00]. Also, it has been documented that violence among

gangs tends to cluster in time. In other words if an act, or perceived act, of violence occurs between two gangs, the likelihood of other retaliatory events increases. This is thought to occur due to the nature of the violence. Gangs often use violence to gain respect and street reputation [Thr27, DW96].

1.2 Agent-Based Models

This is a broad methodology utilized by a diverse spattering of fields. Due to the variations of agent-based models, presenting them to other researchers has been a stumbling block for the agent-based modeling community. To mitigate this issue, the ODD (Overview, Design concepts, and Details) protocol has been established [GBD10]. The purpose of this protocol is to communicate agent-based models in such a way that they can be replicated and compared. It also provides a checklist for modelers to ensure all of the important elements and details are discussed. The agent-based model described in Chapter 2 utilizes the ODD protocol. In this section we review three agent-based models.

1.2.1 Modeling Capelin Fish

Particle models have been successfully applied in ecological settings. The authors of [BEB09] utilize an agent-based model to predict the spawning migration of Capelin fish in the Iceland Sea. Each fish, k, is seen as a particle with a speed $v_k(t)$ and position, $q_k(t) = (x_k(t), y_k(t))^T$ that depends on the location of the other migrating fish and the temperature of the water. So that fish are not swimming too close to one another, each fish has a region of repulsion, R_k. Each fish has a zone of attraction, $A_k(t)$, to maintain the coherence of the migration. Finally, each fish has a zone of orientation, $O_k(t)$.

One element of the environment that influences the migration is the temperature of the water. Capelin fish have a preferential temperature range, $[T_1, T_2]$, and display a stronger aversion to lower temperatures than higher temperatures. This range in encoded in the

model by

$$
r(T) = \begin{cases} -(T - T_1)^4 & \text{if} \quad T \leq T_1 \\ 0 & \text{if} \quad T_1 \leq T \leq T_2 \\ -(T - T_2)^2 & \text{if} \quad T_2 \leq T \end{cases} . \tag{1.2}
$$

To reach the preferred temperature, each fish moves up gradients of $r(T)$.

Taking these forces into consideration, the position and velocities of the particles are updated via

$$
q_k(t + \Delta t) = q_k(t) + \Delta t v_k(t + \Delta t) \frac{D_k(t + \Delta t)}{||D_k(t + \Delta t)||} + C(q_k(t)) \tag{1.3}
$$

and

$$
v_k(t + \Delta t) = \frac{1}{|O_k|} \sum_{j \in O_k} v_j(t), \tag{1.4}
$$

respectively. Here $D_k(t + \Delta t)$ is defined by

$$
D_k(t + \Delta t) := (1 - \beta) \frac{d_k(t + \Delta t)}{||d_k(t + \Delta t)||} + \beta \frac{\nabla r(T(q_k(t)))}{||\nabla r(T(q_k(t)))||}, \tag{1.5}
$$

where

$$
d_k(t + \Delta t) := \frac{1}{|R_k| + |O_k| + |A_k|} \left[\sum_{f \in R_k} \frac{q_k(t) - q_f(t)}{||q_k(t) - q_f(t)||} \right. \tag{1.6}
$$

$$
\left. + \sum_{f \in O_k} \begin{pmatrix} \cos(\phi_0(t)) \\ \sin(\phi_0(t)) \end{pmatrix} + \sum_{a \in A_k} \frac{q_a(t) - q_k(t)}{||q_a(t) - q_k(t)||} \right] . \tag{1.7}
$$

In this equation $\beta \in [0, 1]$, $C(x, y)$ is the current.

With the model, the researchers were able to match the migration of Capelin fish for three years. Further, modeling this way allowed for the inclusion of a large number of particles and an understanding of the long-term behavior of the system. Finally, a sensitivity analysis of the system gained insight into the system. As the value of the temperature dependence, β, changed so did the shape and location of the migration. Also, as the ratio between the radii of influence changes so does size of the school. These models and their corresponding analysis have a large impact on the Icelandic fishing industry. Capelin are a feeder fish for the larger, more economically important fish. Knowing the size and location of the migrations can help maintain the health of these schools.

1.2.2 Modeling Self Emerging Networks in Ants

Due to their flexibility, agent-based models provide an ideal framework to explore the local interaction between individuals to form networks in a complex system. For example, foraging ants leave chemical trails to communicate to one another about the location of food. As more ants follow the trail, a network is formed, reinforcing the ants behavior. In their work [ST02], the authors develop such a self assembling network using positional information by chemical gradients and a non-trivial local interactions with existing chemical trails.

The mechanics of the model are fairly straight forward. Each agent maintains a position, r_i, velocity, v_i, and internal state θ_i. The position of each agent is updated via

$$\frac{dr_i}{dt} = a_i \frac{\partial h^\epsilon(r,t)}{\partial r}\big|_{r_i,\theta_i} + \sqrt{2\epsilon_i}\xi_i(t). \tag{1.8}$$

The gradient of chemical trail is incorporated in the $h^\epsilon(r,t)$ term. The attraction to the chemical gradient is determined by the coefficient a_i. A positive coefficient corresponds to an attraction to the field, whereas a negative coefficient implies repulsion. The random component, $\xi_i(t)$, is white noise mitigated by the strength ϵ_i.

In this scheme the agents follow a biased random walk in an environment scattered with nodes with positive and negative potentials. The agents' "goal" is to link the nodes with opposite potentials. Once an agent discovers a node, θ_i is updated via

$$\Delta\theta_i(t) = \int_{Aj=1}^{z} (V_j - \theta_i)\frac{1}{A}\delta(r_j^z - r_i(t))dr. \tag{1.9}$$

The agents produce their chemical trail, $s_i(\theta_i,t)$, dependent on the internal state, θ_i, of the agent, see Equation 1.10.

$$\begin{aligned} s_i(\theta_i,t) &= \frac{\theta_i}{2}\left[(1+\theta_i)s_{+1}^0 \exp\{-\beta_{+1}(t-t_{n+}^i)\}\right. \\ &\quad - \left.(1-\theta_i)s_{-1}^0 \exp\{-\beta_{-1}(t-t_{n-}^i)\}\right]. \end{aligned} \tag{1.10}$$

The concentration of chemicals persist in the environment by the equation

$$\frac{\partial h_\theta(r,t)}{\partial t} = -k_\theta h_\theta(r,t) + \sum_{i=1}^{N} s_i(\theta_i,t)\delta_{\theta;\theta_i}\delta(r-r_i(t)). \tag{1.11}$$

An agent interacts with the chemicals according to their current state through the effective chemical trail,

$$\frac{\partial h^{\epsilon}(r,t)}{\partial r} = \frac{\theta_i}{2}\left[(1+\theta_i)\frac{\partial h_{-1}(r,t)}{\partial r} - (1-\theta_i)\frac{\partial h_{+1}(r,t)}{\partial r}\right]. \qquad (1.12)$$

Simulations of this model show an initialization phase where the agents are moving according to the random walk. Once nodes of the opposite label have been discovered networks begin to form. These network connections gradually become stable and persist for the remainder of the simulation. In this way the authors were able to investigate the formation of ant networks, but the applications of such an exploration is not limited to networks. Similar methods could be used in engineering self-repairing networks in the field of electronics and understanding the formation of neural networks in a developing brain.

1.2.3 Modeling Burglaries

In their work [SDP08], Short et al. produced similar macro behavior observed by criminologists by using a simplified model on the individual scale. Their model hinged on following principals:

1. Along the lines of routine activity theory, potential burglars are primarily going about their normal life. Due to their criminal tendencies, however, these criminals will also position themselves where there are attractive opportunities to burgle a house.

2. If a house is burgled, then the attractiveness of the house and its neighbors also increases. This is known as repeat-near repeat victimization.

The model reformulates these two sociological principals observed in the burglary data into a mathematical framework, using them to dictate the movement rules of the criminal agents. The simulation is run on a 2 dimensional lattice representing a geographic location. For each lattice point there is an attractiveness score, $A_{i,j}(t)$, and is determined by

$$A_{i,j}(t) = A_{i,j}^0 + B_{i,j}(t), \qquad (1.13)$$

8

where $A_{i,j}^0$ is the background attractiveness of that location and $B_{i,j}(t)$ denotes the influence of the history of the activity at that, and nearby locations. The response to previous activity is updates via

$$B_{i,j}(t + \delta t) = \left[(1 - \eta) B_{i,j}(t) + \frac{\eta}{z} \sum_{(k,l) \in N/(i,j)} B_{k,l} \right] (1 - \omega \delta t) + \theta E_s(t). \qquad (1.14)$$

The first term encapsulates the effect of burglaries in the neighborhood, N, of (i, j). The weighting factor, $\eta \in [0, 1]$ defines the strength of the effect of neighboring lattice locations. The factor z is a normalization equal to the number of neighboring lattice locations included in the summation. The number of burglary events is represented by the $E_s(t)$ term.

The burglars move on this lattice and are allowed to either burgle a house or move to a neighboring lattice location. The criminal will burgle a house with probability

$$p_s(t) = 1 - e^{-A_{i,j}(t)\delta t}. \qquad (1.15)$$

If a criminal does burgle a house, that criminal is removed from the system. To maintain a criminal presence in the model, criminals are creates at each lattice location with probability Γ. If an agent does not burgle a house, then they move according to a biased random walk. Specifically, the agent moves to a neighboring location, $(\hat{i}, \hat{j})$, with probability

$$q_{\hat{i},\hat{j}} = \frac{A_{\hat{i},\hat{j}}(t)}{\sum_{(k,l) \in N/(i,j)} A_{\hat{i},\hat{j}}(t)}, \qquad (1.16)$$

where $\sum_{(k,l) \in N/(i,j)}$ is the sum of all the neighboring locations of $(\hat{i}, \hat{j})$, excluding (i, j).

From this simple model three distinct emergent behaviors were observed.

1. Spatial Homogeneity: Here the value of $A_{i,j}(t)$ is essentially the same at all lattice locations.

2. Dynamic Hotspots: In this regime there are local peaks in the attractiveness at certain locations. These peaks may persist, move, or decay after a long period of time.

3. Stationary Hotspots: These peaks in the attractiveness field are stable and are surrounded by a areas of characteristically low attractiveness.

Each of these emergent behaviors have been observed in the real world, adding validation to the model. Further gains were made when the authors took the continuum limit of the particle model, yielding a PDE. With this new model the authors were able to utilize the machinery of PDE theory to understand the role of the parameters in the model. As an extension to the model, the authors in [MS12] were able to use a Bayesian framework to infer the anchor location of a repeat offender.

1.3 Data-Driven Models

One example of a data-driven method is that of density estimation. Many systems can be viewed as stochastic processes where the data collected is just one realization. In many cases, one would like to estimate the underlying process given the data. Classically this has been accomplished via parametric estimation. Here, the underlying functional form, $f(\cdot|\vec{\theta})$, with parameters $\vec{\theta}$ is assumed. If the data collected are independently and identically distributed (iid), then the joint distribution of observing the data $\vec{x}$ is

$$f(x_1, x_2, \cdots x_n|\hat{\theta}) = \prod f(x_i|\hat{\theta}). \tag{1.17}$$

The maximum likelihood estimation of the parameters $\hat{\theta}$ are determined by

$$\sup_{\hat{\theta}} \left\{ \prod f(x_i|\hat{\theta}) \right\}, \tag{1.18}$$

or equivalently

$$\sup_{\hat{\theta}} \left\{ \sum \ln \left(f(x_i|\hat{\theta}) \right) \right\} \tag{1.19}$$

[Cam90]. For an interesting history of this method see [Ald97].

1.3.1 Non-parametric Density Estimation

For many applications assuming a functional form of the underlying density is not plausible. For example, consider a density on a two dimensional random variable where it is known

10

that

$$P\{X = x, Y = y\} = 0 \tag{1.20}$$

for (x, y) in some, possibly irregular, region A. In these cases alternative methods that do not rely on a closed form of the density function are needed. Such methods typically fall under the classification of non-parametric methods.

Such a situation is consider in [SKW10]. Here, Smith et al. develop the Weighted H_1 Maximum Penalized Maximum Likelihood Method, as an alternative to TV Maximum Penalized Maximum Likelihood or Gaussian kernel density estimation. The Weighted H_1 Maximum Penalized Maximum Likelihood Method seeks to minimize

$$\hat{u}_H(x) = \inf_u \left\{ \frac{1}{2} \int_\Omega z_\epsilon^2 |\nabla u|^2 dx - \mu \sum_{i=1}^n \log(u(x_i)) + \frac{\gamma}{2} \left(\int_\Omega u(x)dx - 1 \right)^2 \right\}, \tag{1.21}$$

subject to $u(\vec{x}) \geq 0$ and $\int_\Omega u(\vec{x})d\vec{x} = 1$. Here z_ϵ is a continuous function such that,

$$z_\epsilon = \begin{cases} 1 & : \text{if } d(x, \partial\Omega) > \epsilon \\ 0 & : \text{if } x \in \partial\Omega \end{cases}.$$

In the presence of sparse data, the authors were able to improve the reconstruction of the density by incorporating information of valid and invalid regions.

1.3.2 Self-exciting Point Processes

Point process models are a versatile tool widely used to analyze earthquakes [VS08, Oga98, Oga88, ZOV02], model financial contagion in credit markets [EGG10, ACL10], viral videos on the web [CS08], terrorist activity in Indonesia [PW11], and the spread of infectious disease [MEH11]. The authors in [MSB11] and [EFL10] have successfully modeled the pairwise gang violence as a Hawkes process [HO74, Haw71b, Haw71a].

A point process in time is a random process defined by a history of events, $H_t = \{t_1, t_2, \cdots t_N, s.t \ t \geq t_N \ \& \ t_{N+1} > t\}$. Each t_i denotes the ordered times of when an event occurred. This process can be naturally extended to include additional information, for example location and magnitude of an earthquake event are often incorporated in earthquake

modeling. Each point process can be alternatively formulated as a counting process, $N(t)$, defined as the number of events that have occurred at time t.

One example of a simple point process is a Poisson process with rate λ. Formally, a Poisson process must satisfy

1. $N(0) = 0$

2. $N(s+t) - N(t) = \text{Poisson}(\lambda s)$

3. $N(t)$ has independent increments [Dur99].

For reference, X = Poisson(μ), is

$$P(X = n) = e^{-\mu}\frac{\mu^n}{n!} \ , \quad \text{for } n = 0, 1, 2, 3, \cdots \tag{1.22}$$

In the case of gang data, it has been shown that certain rivalries in Hollenbeck can be modeled with a Poisson process [EFL10], however, for other rivalries the third constraint it too strong. For these rivalries an event can spark a series of retaliations between the two gangs, producing events more clustered in time than would be expected from a Poisson process [Dec96, MSB11, EFL10]. For these rivalries a more sophisticated model must be employed to capture the features observed in the system.

To generalize our model of point processes we allow the rate $\lambda(t|H_t)$ to depend on time and the history of the process. In this dissertation we consider the self-exciting point process proposed by Hawkes such that

$$\text{P}\{N_{t+\Delta t} - N_t = 1|H_t\} = \lambda(t|H_t)\Delta t + o(\Delta t) \tag{1.23}$$

and

$$\text{P}\{N_{t+\Delta t} - N_t \geq 2|H_t\} = o(\Delta t). \tag{1.24}$$

Intuitively this means that the probability of having exactly one event in a small window of time can be approximated by the function $\lambda(t|H_t)$. In this dissertation we assume that

12

$\lambda(t|H_k)$ is of the form

$$\lambda(t|H_t) = \mu + \int_0^t g(t - u)dN(u), \text{ where } \int_0^\infty g(t)dt < 1 \qquad (1.25)$$

[Haw71a, Oga88, OA82, Haw71b]. Let $\Delta_t N = N_{t+\Delta t} - N_t$, and note that

$$E\{\Delta_t N|H_t\} = 1 \cdot P\{\Delta_t N = 1|H_t\} + \sum_{i=2} i \cdot P\{\Delta_t N = i|H_t\} \qquad (1.26)$$

$$E\{\Delta_t N\} = \lambda(t|H_t)\Delta t + o(\Delta t). \qquad (1.27)$$

A direct calculation leads to

$$\lambda(t|H_t) = \frac{E\{\Delta_t N|H_t\}}{\Delta t} \qquad (1.28)$$

$$= \lim_{h \to 0} \frac{P\{\Delta_t N = 1|H_t\}}{h} \qquad (1.29)$$

$$= \lambda(t|H_t) \qquad (1.30)$$

$$= \mu + \sum_{t > t_j} g(t - t_j) \qquad (1.31)$$

[Haw71b, Oga88]. $\lambda(t|H_t)$ can be thought of as the instantaneous rate of the process, much like the constant rate of the Poisson process. In this way the process is defined by its intensity $\lambda(t|H_t)$.

Intensity functions of the form described by Equation 1.31 lead themselves to an intuitive interpretation. The rate of background activity is controlled by the parameter μ. In the case where there is no response to previous events, i.e. $g(t) = 0$, the process is Poisson with rate μ. The response to the history of events is defined by the function $g(t)$. It is common to write this function in the form $g(t) = \alpha \cdot h(t)$, where $\alpha \in [0, 1)$ and $\int_0^\infty h(t)dt \leq 1$. The constant α determines the expected number of offspring for each event in the process. The constraint on this parameter ensures that there are a finite number of events in the process. For this dissertation we restrict our attention to response functions of the form $h(t) = \omega e^{-\omega t}$. An example of a simulated process with the intensity $\lambda(t|H_t)$, is plotted in Figure 1.1. When given a realization of a process determined by the intensity $\lambda(t|H_t)$, one can determine the parameters of the underlying process via the log likelihood function

$$\hat{\ell}_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k) = \sum_{i=1}^{M_k} \lambda_k(t_i|H_{\tau,k}) - \int_0^T \lambda_k(t|H_{\tau,k})dt. \qquad (1.32)$$
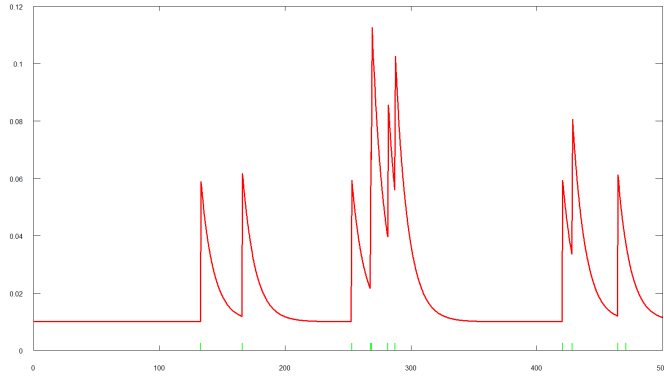
Figure 1.1: Plot of the intensity, $\lambda(t|H_t)$ for one realization of a self-exciting point process with $\mu = .01$, $\alpha = .1$, $\omega = .5$.

[DV03].

Alternative methods to determine the process parameters are discussed in Section 3.2.2 in Chapter 3 of this dissertation.

CHAPTER 2

Long-term Rivalry Genesis

In this chapter we propose an agent-based model that incorporates geographical features and is coupled to a dynamically evolving network to infer geographical influences in the formation of street gang rivalry networks. This model simulates the mobility of gang members and the resulting interactions. We compare the simulated network to the gang rivalry network observed in the eastern Los Angeles division of Hollenbeck [TRR03, RFT10]. In Section 2.2, we outline the proposed model.

In Section 2.3, we describe two baseline models, one instance of a Geographical Threshold Graph and a network derived from Brownian Motion, to which we compare our model. In Section 2.4, we describe a series of metrics from network theory, examine long term behavior of the model, and compare the networks against the metrics. Section 2.5 provides a sensitivity analysis of our model. We conclude and give future directions in Section 2.6. This work was done in collaboration with Laura M. Smith, Alethea B.T. Barbaro, Andrea L. Bertozzi, Shannon E. Reid, and George E. Tita and was published in Physica A in [HSB11].

2.1 Introduction

Street gangs are a growing problem around the world [Cov10, KWT06, KKM01]. In fact, recent statistics from The National Gang Intelligence Center estimate there are 1 million active gang members in the United States alone [Pro09]. Violence is intrinsic to street gangs, and rival gangs battle to gain respect and street reputation [Thr27, DW96]. Criminal activities perpetrated by gang members, including armed robbery, homicide, drug dealing,

15

and auto theft, drain cities and governments of tight resources and also pose safety threats to community members. Much of the research on street gangs has been conducted within the United States, though there have been some efforts to understand the phenomenon in Europe and other parts of the world [Cov10, KWT06, KKM01].

Violence perpetrated by gang members is frequently against members of a different gang. In areas with numerous gangs, it is common for gangs to have multiple violent interactions with many of the other gangs. Further, street gang members typically have locations, known as set spaces, where they spend large quantities of time [TCE05, Pap09]. It is therefore reasonable to think of each gang as a node embedded in Euclidean space [RFT10, TRR03]. Within this framework, the existence of persistent violence between two gangs becomes an edge connecting two nodes. From this construction, one can view a collection of gangs as a spatially embedded network [TR10]. The Hollenbeck policing division of eastern Los Angeles is marked by a particularly high degree of violent crimes involving gang members, including homicides and aggravated assaults [RFT10, Inf08]. It is for this reason and others listed in Section 2.1.4 we consider Hollenbeck as a test case for our model.

2.1.1 Rivalry Models

General network models and the corresponding analysis are useful for describing the behavior of complex systems and have played an increasingly active role [New01a, New01b, New03]. One way networks are treated in the literature is by analyzing the statistical properties of a given network. Another approach is to consider the construction of a network. There are many instances where the network of interest is not known, but there is some knowledge of the processes by which the network is formed. One popular method to construct a network is to view it as a random graph. Each edge is added with a predetermined probability, often dependent on the weight of the nodes [AB02, NSJ01, NJS02].

In some applications, including gang rivalry networks, the geographic location of the nodes influences the structure of the network. In such cases, geographic features should be

considered as part of the random network model. For example, interstate highways have been shown to be structurally different from scale-free networks such as Internet and airline flight networks [GN06]. The importance of geography is also seen in friendship networks [Wel96]. In their paper, Liben-Nowell et al. use the publicly accessible location of 495,836 bloggers in the Live Journal to investigate effects of spatial proximity on friendships [LNK05]. The study found that an estimated 69% of a person's friends can be described by geography.

One method for incorporating geographical information into the random graph construction is by using a Geographical Threshold Graph [MMK05, BHP07, BHP09]. This is a random graph on a set of randomly weighted nodes, where the nodes are located in a metric space and the connections are determined by thresholding a function of the distance and the weights. This provides a computationally efficient way to construct a rivalry network while incorporating some geographic information. We use an instance of a Geographical Threshold Graph as a baseline against which we compare our model.

2.1.2 Agent-Based Models

Though using a randomly constructed network may produce a reasonable simulation of an observed network, other phenomena of interest beyond the structure of the network are not obtained from this type of model. An alternative is to use an agent-based approach. For a detailed explaination of this modeling framework see Section 1.2. This has become a widely used tool in the area of complex systems [TLL10, Tes06, Wil06, FF09, EGK04, MW02, SSH09]. For example, agent-based models have been used for modeling many types of cooperative behavior [LST10, SDP08, DCB06, CCR10, HH04]. In a network context, this approach enables exploration of how changing dynamics of individual agents can affect the evolution of the network, providing control parameters which would be inaccessible in a graph-based model. This method can easily incorporate environmental and spatial information inherent to the system, e.g. in [BEB09], which uses environmental cues to reproduce and predict fish migration.

In particular, we are interested in the coupling between the network and the underlying system. There has been some exploration of this in the literature. For example, Schweitzer and Tilch provide one example of model that uses an agent-based approach to form an emerging network [Sch03, ST02]. They model the chemical trail formed by ants searching for food at an unknown location. As the ants search their environment, networks of chemical trails form with which the ants interact. Another example is that of the EpiSims model [TG07, MDS08]. Here, the contact networks of the populations are evolving over time and depend on the internal attributes of the people in the population. In turn, as a disease is spread through the contact network, the movements of the people change in response to the disease, producing a non-trivial interaction between the system and the network.

2.1.3 Previous Work on Crime Modeling

Various models have been created to address criminal activity [Gor10, Pit10, OL09, BT08, SDP08]. One such work uses an agent-based model to understand the formation of crime hot spots [SDP08]. In the model proposed by Egesdal et al., an agent-based approach was used to simulate the location of violent interactions and gang retaliations in Hollenbeck [EFL10]. Embedded in the model was a rivalry network. Though the model recreated similar features to the violence data, the model did not incorporate geographic features. In addition, agents targeted specific gangs based on probabilities corresponding to the current rivalry network. However, according to the criminology literature, this retaliatory behavior is only seen on short time scales [MSB11]. Gang members tend to avoid the territory of rival gangs [And00, LC74].

Although not an agent-based model, Mohler examines the short term retaliatory behavior of the rivalries based on between gang violence data from the LAPD [MSB11]. Each violent event between two gangs is considered an instance of a point process associated with that pair of gangs. The intensity of the rivalry depends directly on the network of unidirectional violent interactions. This provides a top-down approach to understanding immediate consequences of violence among gangs within a system. In our work, we wish to understand factors

associated with the long term gang rivalry structure.

2.1.4 Hollenbeck

Hollenbeck is a policing division located in eastern Los Angeles, surrounded by downtown Los Angeles to the west, Pasadena to the northeast, Vernon to the south, and to the east the unincorporated area of East Los Angeles, see Figure 2.1. Hollenbeck provides a diverse geography with many highways cutting through the region and is bounded by the Los Angeles River. It encompasses an area of roughly 39.4 km^2. Hollenbeck is home to approximately twenty-nine active gangs with sixty-nine rivalries among them [RFT10, TRR03]. The set spaces for the gangs and the corresponding observed rivalry network are displayed in Figure 2.1, as given in [RFT10].

Certain properties of Hollenbeck make it accessible to modeling the gang rivalry networks outlined in [TRR03, RFT10]. First, it is a closed system in that the gang activity within Hollenbeck is generally isolated from gang activity outside of Hollenbeck. Further, the motivation for violence between gangs is largely characterized by disputes over geographical gang territories, as opposed to drug and racially motivated violence. Data on the geography of Hollenbeck is easily accessible, and there has been explicit documentation of the observed rivalry network.

We propose an agent-based model that incorporates geographical features and is coupled to a dynamically evolving network. This model simulates the mobility of gang members and the resulting interactions. We compare the resulting simulated network to the gang rivalry network observed in the eastern Los Angeles division of Hollenbeck [TRR03, RFT10]. In Section 2.2, we outline the proposed model. In Section 2.3, we describe two baseline models, one instance of a Geographical Threshold Graph and a network derived from Brownian Motion, to which we compare our model. In Section 2.4, we describe a series of metrics from network theory, examine long term behavior of the model, and compare the networks against the metrics. Section 2.5 provides a sensitivity analysis of our model. We conclude and give

19
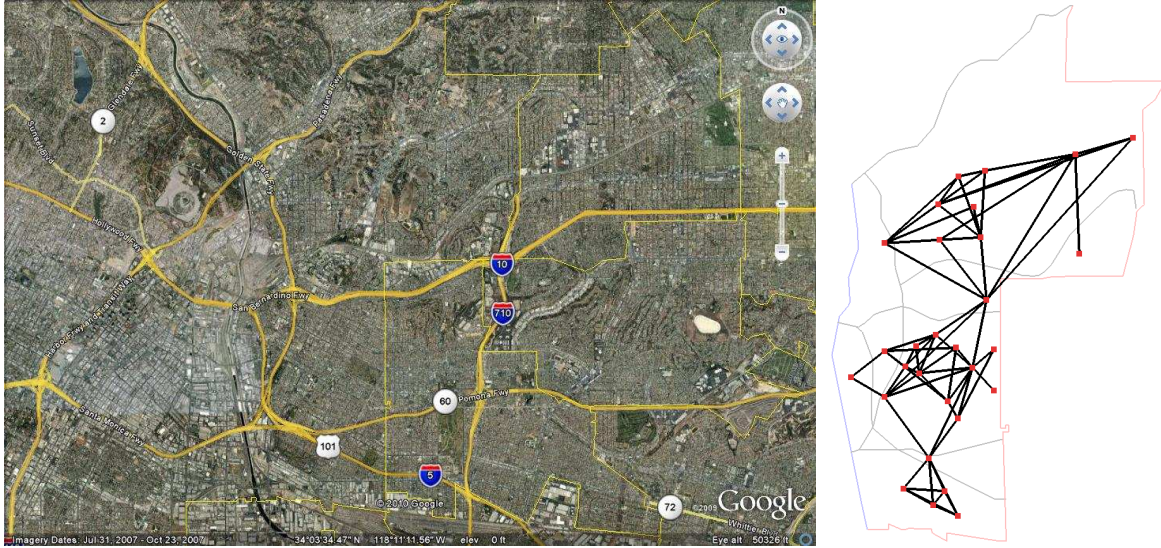
Figure 2.1: Google Earth^{TM} Image of the Hollenbeck area (left). Map of the Hollenbeck area with the location of the gang set spaces and the corresponding rivalry network approximated by [RFT10] (right). Major roads, highways, the Los Angeles river, and division lines are also seen in both images.

future directions in Section 2.6.

2.2 The Simulated Biased Lévy Walk Model (SBLN)

The purpose of this model is to understand the extent to which simple behavioral rules and geographical factors, such as road density, highways, and locations of gangs' set spaces, could influence the structure of gang rivalry networks.

2.2.1 Motivation for Model Construction

The intent of this model is to capture the broad statistical features of human mobility with an emphasis on gang members' movements. Empirical data on the location and individual movements of each gang member is inaccessible, so we characterize the movements of the individual gang members in a statistical sense based on the literature on human mobility.

20

Several studies give compelling evidence that when people move in an unconstrained environment, the jump lengths between movements is distributed like a power law [BHG06, RSH08]. Further, in the presence of obstacles such as roads and buildings, the jump lengths more accurately follow a bounded power law distribution [GHB08].

However, determining the statistical properties of the jump length is only one aspect of movement dynamics. In their paper, Rhee et al. discuss the need to incorporate geographical features and the tendency for people to go home [RSH08]. Gonzalez et al. confirmed in their data that humans do tend to frequent a small number of locations often [GHB08]. For these reasons, the agents in our model pick their jump length from a Bounded Pareto distribution and have a directional choice in movement.

In the case of gangs in Hollenbeck, it is reasonable to assume that the gang members have a clear sense of the location of their home territory, or set space, as well as the location of their rival gangs' set spaces [TCE05]. Literature on gang activity suggests that, in general, gang members tend to stay away from their rival gangs' set spaces [LC74]. Unlike other criminal groups, such as organized crime syndicates and insurgency groups that strive for secrecy, street gangs are social organizations that proudly demarcate their territory and announce their enemies through the use of graffiti. Gangs create social boundaries and therefore areas of avoidance [And00]. Our model incorporates this social geography into agents' movement dynamics.

One aspect of modeling human mobility that was touched on, but not fully explored, by the previous literature is the role of physical features specific to urban areas that may constrain agents' movement. The first consideration is the ease with which an agent can move through a city. We posit that in areas where there is a dense street network, the likelihood of an agent to move long distances is small due to such obstacles as the high density of people and cars, as well as traffic lights. On the other hand, areas where the road density is lower, agents should be able to move longer distances. A second physical consideration that affects human mobility in a city are the highway systems and rivers that can cut across the region. These features are not impassible, in that there are underpasses and bridges. However, they

do provide an obstacle that may make it difficult to cross. Therefore, in our model, these are seen as semi-permeable boundaries that effect the agents' movements.

2.2.2 Model Summary

Agents in the model move based on their location with respect to their and other gangs' set spaces and interact with agents of different gang affiliations. We count the number of interactions between gangs, and when agents of different gangs move within a certain distance of each other, the number of interactions between those gangs increases by one. As the simulation progresses, a network structure emerges. The weighted network of interactions in turn influences the directional decisions of the agents.

2.2.3 Entities, State Variables, and Scales

2.2.3.1 Agents

The agents of this model are gang members in a city. Each agent is associated with exactly one gang. For simplicity we assume agents' directional choice is dictated only by the location of the gang set spaces. All agents know the location of their home and rivals' set spaces. We divide the city into regions based on geographical boundaries, such as rivers and highways. An agent knows which region it is currently in as well as the region of any prospective new locations. When two agents are within interaction range, we consider them to have interacted and the corresponding element of the rivalry matrix, R, is updated. There are no immediate changes to the location of the gang members. Refer to Section 2.2.3.3 for details on R.

2.2.3.2 Environment

The environment of interest is on the scale of a small city. Agents and gang set spaces in the environment have a coordinate location in Euclidean space. Further, the set spaces provide the spatially embedded nodes of the gang rivalry network. The physical geographical

features of the city are approximated by an NxM grid. Agents move in Euclidean space, and each point in the region is identified with the nearest grid element. The size and number of grid elements are constant throughout the simulation and will be limited by the available data and the memory of the computer.

Two features encoded in this NxM grid are the road density and semi-permeable boundaries represented by a region map. The road density is estimated for this project using the Weighted H^1 Maximum Penalized Likelihood Estimation method with a road map as the initial data, as in [SKW10]. Other methods for density estimation, such as kernel density estimation or other Maximum Penalized Likelihood Estimation (MPLE) methods, could also be used to construct the road density [Sil86, EL01, MBG09, GG71]. Each element of the NxM density map contains a number between 0 and 1. A value of 0 implies a low road density whereas a value of 1 implies high road density. The semi-permeable boundaries, corresponding to such objects as highways and rivers, are assumed to split the environment into distinct regions. Therefore, each element of the region grid corresponds to a specific region. Paired with this region grid is a matrix storing the associated probability of an agent to cross from one region to another. This is implemented to discourage agents from crossing freeway boundaries.

2.2.3.3   Rivalries

The network structure of the rivalries is encoded in a weighted adjacency matrix, R. Each element R_{ij} contains the current history of interactions between gang i and gang j. At the end of a simulation, we construct a thresholded rivalry graph where an edge between gang i and j exists if either $\rho_i(j)$ or $\rho_j(i)$ is larger than a given threshold T, where

$$\rho_i(j) = \frac{R_{ij}}{\sum_{k=1}^{N} R_{ik}} \qquad \text{and} \qquad \rho_j(i) = \frac{R_{ji}}{\sum_{k=1}^{N} R_{jk}}. \tag{2.1}$$

The quantity, $\rho_i(j)$, represents the proportion of gang i's interactions which have occurred with gang j. Note that $\rho_i(j)$ is not necessarily equal to $\rho_j(i)$; however, this thresholding yields a bidirectional network or, equivalently, a symmetric adjacency matrix.

2.2.4 Process Overview and Scheduling

At each iteration an agent is chosen from the set of all agents with equal probability. The agent then preforms one step of a biased truncated Lévy walk. In particular, the jump length is chosen from the Bounded Pareto probability distribution,

$$P(x; k, x_m, x_M) = \frac{k x_m^k x^{-k-1}}{1 - \left(\frac{x_m}{x_M}\right)^k} \quad k > 0, \quad x_M \geq x \geq x_m > 0. \tag{2.2}$$

For all agents the minimum jump length, x_m, and scale, k, are fixed. To determine the maximum jump length, x_M, the agent uses the approximated road density of the agent's corresponding location from the environment grid. The road density at this location, δ, is between 0 and 1. The maximum jump length is then calculated via

$$x_M = (1 - \delta) \cdot A + a, \tag{2.3}$$

where A is the largest maximum jump length and a is the smallest maximum jump length.

Given an agent in gang i, the bias direction, μ_i, incorporates the agent's location with respect to its home set space and the location of its rival gangs' set spaces via

$$\langle x, y \rangle = H_i(||\overrightarrow{G_i}||_2) \; \frac{\overrightarrow{G_i}}{||\overrightarrow{G_i}||_2} + \sum_{j \neq i} D_{ij}(||\overrightarrow{G_j}||_2) \; \frac{\overrightarrow{G_j}}{||\overrightarrow{G_j}||_2} \tag{2.4}$$

$$\mu_i = \tan^{-1}\left(\frac{y}{x}\right).$$

Here, $\overrightarrow{G_l}$ is the vector that points to the set space of gang l from the location of the agent. When $l = i$, this vector points towards the agent's home set space, and when $l \neq i$, it points towards a different gang's set space. This concept is shown in the cartoon example in Figure 2.2.

In Equation 2.4, H_i gives the rules for weighting towards a gang member's own home set space. The weightings toward or away from different gangs' set spaces are determined by D_{ij}. Our H_i and D_{ij} take the following form:

$$H_i(||\overrightarrow{G_i}||_2) = h_i \, ||\overrightarrow{G_i}||_2, \tag{2.5}$$

$$D_{ij}(||\overrightarrow{G_j}||_2) = w_{ij}(R) \; \frac{1}{||\overrightarrow{G_j}||_2}. \tag{2.6}$$

One notable feature about these equations is that $H_i(\cdot)$ is large when an agent in gang i is far from his or her gang's set space, but the $D_{ij}(\cdot)$ function is large when the agent is close to a rival gang j's set space. The factors h_i and $w_{ij}(R)$ of the weighting functions are chosen according to the rules for agent movement. In our implementation, the factor $w_{ij}(R)$ depends on the current state of the rivalry network. Negative values of these functions result in repulsion and positive values result in attraction.
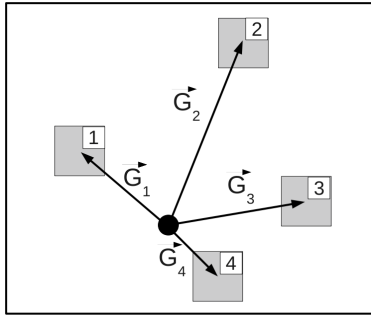
Figure 2.2: Cartoon example of the direction vectors incorporated in the direction of bias formula, Equation 2.4. The agent in this example is located at the dot. Here G_1, G_2, G_3, and G_4 show the vectors pointing toward the set spaces of gangs 1 through 4, respectively. Depending on the choices of H_i and D_{ij}, different movement dynamics are possible.

After determining the direction of bias from Equation 2.4, we must choose in which direction the agent will move. The direction, θ, is drawn from a von Mises distribution (also known as the Circular Normal distribution) [MJ00, JS01, BF79]. For $\theta \in [-\pi, \; \pi]$, the von Mises distribution is given by

$$f(\theta|\mu,\kappa) = \frac{\exp\left(\kappa\cos(\theta - \mu)\right)}{2\pi I_0(\kappa)}.$$

Here I_0 is a modified Bessel function of order zero. The von Mises distribution requires two parameters, one for the angle of bias, μ, and one for the strength of the bias, κ. We can think of μ as being the mean of the distribution, and $\frac{1}{\kappa}$ as being comparable to the variance. The

larger κ is, the stronger the bias is for the direction μ. If $\kappa = 0$, this is a uniform distribution on a circle.

From the direction and jump length, a prospective location is calculated. The new location is then checked to see if the result would move the agent into a different region. If it does not, the agent moves, meaning movement within a region is not restricted. However, if its next move would result in a region change, i.e. it is crossing a semi-permeable boundary, it has a given probability of crossing into that region. If the agent moves, it searches the other agents to see if it is close enough to interact with agents of other gangs. When an interaction does occur, the rivalry matrix, R, is updated.

The final network is observed after 20,000,000 iterations and then thresholded to ignore infrequent interactions. The location of interactions is also recorded and could be of interest to other applications, see discussion in Section 2.6 and Figure 2.14.

2.2.5 Initialization and Input Data

Before the simulation begins, the region map and an estimated density of the road networks must be provided in matrix form on the same grid. The probability of crossing each boundary must also be provided. Additionally, parameter values must be specified. Table 2.1 describes the full list of parameters needed for implementation. At the start of the simulation all of the agents are located at their gang's set space. The size of each gang must also be specified.

2.2.6 Hollenbeck Parameters

The grid of environment features of Hollenbeck was approximated from the Google Earth^{TM} image in Figure 2.1. Hollenbeck is about 39.4 km^2 [RFT10, TRR03]. In our implementation, one Hollenbeck city block corresponds to approximately six grid elements. The interaction radius between agents is 3 units, or roughly half a city block. The approximated road density and region grids are show in Figure 2.3. The boundaries of the Hollenbeck region were approximated using points from the geographic features visible from Google Earth^{TM}.

Params	Acceptable Values	Hollenbeck Values	Tested Range	Description
x_m	$0 < x_m < a$	0.1		Minimum jump length
k	$0 < k$	1.1	$[1, 1.9]$	Bounded Pareto scaling
κ	$0 \leq \kappa$	3.5	$[1.5, 5]$	Von Mises scaling
h_i	$h_i \in \mathbb{R}$	1		Home weighting
$w_{ij}(R)$	$w_{ij}(R) \in \mathbb{R}$	$-\rho_i(j)$		Rival gang weighting
N_i	$N_i \in \mathbb{Z}^+$	$14 \leq N_i \leq 598$		Size of gang i
S_i	$S_i \in \mathbb{R}^2$	see Figure 2.1		Location of gang i set space
A	$a < A$	200	$[100, 400]$	Largest max jump length
a	$x_m < a$	100	$[100, 200]$	Smallest max jump length
B	$0 \leq B \leq 1$	0.2	$[0, .5]$	Permeability of boundaries
T	$0 \leq T$	0.04	$[0, 0.6]$	Threshold for existence of an edge

Table 2.1: Parameters needed for model implementation are listed in the first column. The second column lists theoretically acceptable parameter values. The values corresponding to the SBLN are displayed in the Hollenbeck Values Column. The Tested Range column provides the range for each variable for simulations run. The last column provides a description of each of the parameter values.

These boundaries were used to construct the region grid. To approximate the road density of Hollenbeck we used a Weighted H^1 Maximum Penalized Likelihood Estimation method with a road map as the initial data [SKW10]. To extend the approximated road density to the same sized grid as the region grid, the average value of the density over Hollenbeck was computed and used for the extended regions. The number of agents in each gang reflects historical information obtained from the LAPD.

The boundary crossing probability between the regions was calculated by the minimum
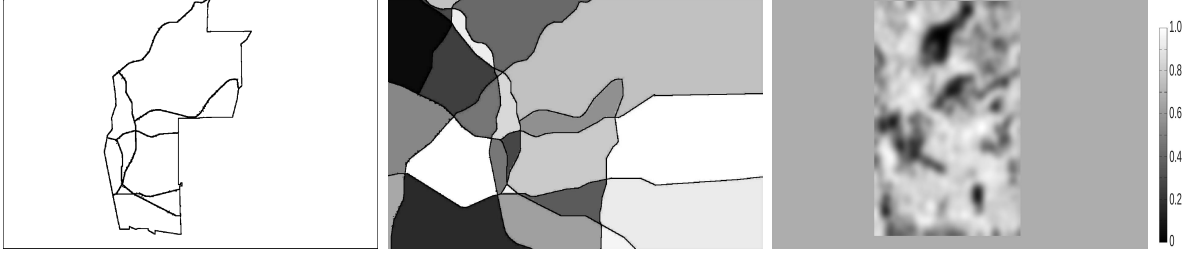
Figure 2.3: The image on the left shows the location of Hollenbeck in the N × M environment grid. The semi-permeable boundaries encoded in the model are displayed in the center image. The shades of gray of this image are used to distinguish among regions. On the right, we used a Weighted H^1 Maximum Penalized Likelihood Estimation method with a road map as the initial data to approximate the road density of Hollenbeck [SKW10]. The scale, seen on the far right, gives the approximated road density intensity. Light shades of gray correspond to high density values near one and dark shades correspond to low densities near zero.

number of boundaries one must cross to get from one region to the next. For instance, if region 1 and region 2 were separated by one boundary, the agent would have a probability, B, of accepting a move from region 1 to region 2. If region 1 and 2 were separated by α boundaries, then the agent would have a B^α probability accepting the move.

2.3 Baseline Comparison Models

2.3.1 Geographical Threshold Graphs (GTG)

For comparison to the networks produced by our simulations, we constructed an instance of a Geographical Threshold Graph (GTG). Geographical Threshold Graphs are random graphs that use spatial proximity to assist in determining whether or not two nodes are connected with an edge [MMK05, BHP07, BHP09]. Geographical Threshold Graphs randomly assign weights η_i to the N nodes. Then, using an interaction function $F(\eta_i, \eta_j)$, an edge between

nodes n_i and n_j exists only if

$$\frac{F(\eta_i, \eta_j)}{d(n_i, n_j)^\beta} \geq \text{Threshold},$$

where $d(n_i, n_j)$ is the distance between nodes n_i and n_j. Constructing an instance of this graph is fast and computationally inexpensive. In our case, we take the multiplicative weight function $F(\eta_i, \eta_j) = \eta_i \cdot \eta_j$, since this is the number of possible pairings between members of gang i and gang j. We use Euclidean distance for the $d(n_i, n_j)$ function. The weights η_i are taken to be the size of each gang, and we choose $\beta = 2$. The threshold was chosen to give the same number of rivalries as the observed rivalry network.

2.3.2 Brownian Motion Network (BMN)

Another model we use to compare with the simulated network is a simplified version of the proposed model using Brownian Motion and unbiased movement rules. The semi-permeable boundaries of the model are incorporated also in this model. Specifically, each agent chooses the next prospective location from a standard normal distribution, ignoring any directional decisions. These simplifications reduce the number of variables to the threshold, T, and the permeability, B, while still incorporating the geographic boundaries. The parameter space around the Hollenbeck values was explored and run for $2 \cdot 10^7$ iterations. *A priori*, it was unclear how many iterations to run the simulation. We observed that the accuracy of the Brownian Motion networks peaked around $1.2 \cdot 10^7$ iterations and then decreased as the simulations progressed. The parameters and number of iterations that produced the highest accuracy were used for analysis. We will refer to the resulting network as the Brownian Motion Network (BMN).

Inherent in the BMN is a level of stochasticity. To understand how this stochasticity influences the final rivalry network and the resulting metrics, the BMN simulation was run for 100 different seed values. The resulting collection of final networks will be called the Ensemble BMN.

2.3.3 Baseline Network Graphs

Figure 2.4 displays the resulting GTG and BMN as compared to the observed rivalry network. The lower portion of the GTG graph has similar shape to the observed network, but contains more connections. The GTG does not make long connections. This is particularly evident in the upper half of Hollenbeck. The BMN picks up many of the longer connections, but includes far too many connections.

Figure 2.4: A visual comparison of the observed rivalry network (left), GTG (center), and BMN (right).

2.4 Results

The results of our network were obtained by searching the parameter space within the ranges specified in the fourth column of Table 2.1, allowing for dependencies between parameters. The 34,128 simulated networks were then sorted according to accuracy, defined in Equation 2.7. Because each of the gangs in Hollenbeck are active, the graph with the highest accuracy with all non-zero degree nodes was chosen as a showcase of the model. The param-

eter values for the optimal run are found in the third column of Table 2.1. We will to refer to this as Simulated Biased Lévy walk Network (SBLN). Figure 2.5 displays the network with our optimal parameters. The SBLN has a shape and structure similar to the observed network, but does not capture all of the longer edges. We also verified that all of the metrics we use to evaluate our model have reached a statistical equilibrium for the SBLN.
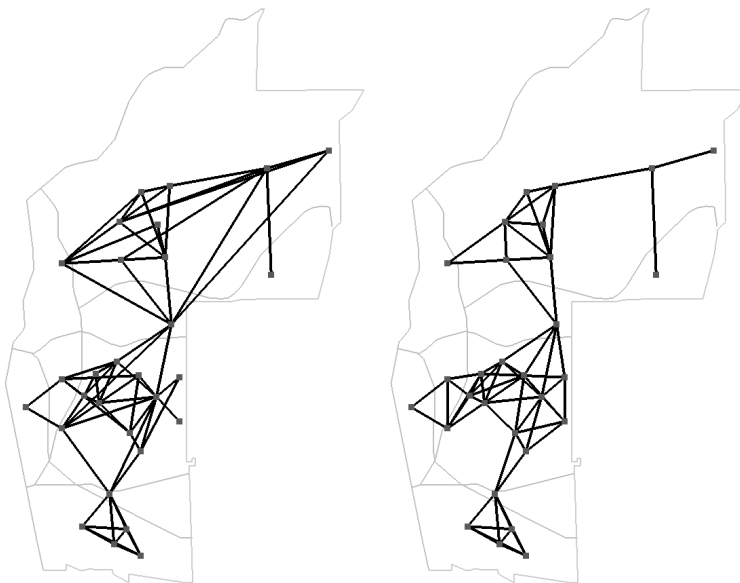
Figure 2.5: Comparison of the observed rivalry network (left) and the SBLN (right). The SBLN has a shape and structure similar to the observed network, but does not capture many of the longer edges.

2.4.1 Stochastic Effects Observed in the Simulated Biased Lévy Walk Network (SBLN)

Implicit in the model is a degree of stochasticity intended to capture the gross features of human movement. In particular, the jump length and direction choice are sampled from probability distributions, and the directional bias is determined by the (inherently stochastic) current rivalry structure. These elements affect the inclusion and exclusion of rivalry network edges. To understand the effect of stochasticity on the network produced by the model,

each simulation was run 100 times with different random seed values with the same SBLN parameter values. We refer to the collection of runs as the Ensemble SBLN. Each simulation was run independently and evaluated with several metrics. The resulting metrics were then averaged for analysis.

We also recorded the persistence of each edge in the ensemble of networks, and this is denoted as the *percent edge agreement*. For example, an ensemble network with 10% edge agreement refers to a network consisting of all edges that appear in at least 10% of the runs. Figure 2.6 displays the Ensemble SBLN with 100%, 50%, and 1% edge agreement next to the observed rivalry network. As expected, increasing the percent edge agreement decreases the number of edges present in the network. The network constructed with 100% edge agreement does not give a close representation of the observed network, because there are too few edges. However, allowing for 50% edge agreement produces a similar shape to the observed network. The Ensemble SBLN 1% edge agreement network shows all possible edges observed in the ensemble of simulation runs. Taken together, these images demonstrate the stochastic effects inherent in the model.

Figure 2.6: Percent edge agreement for the ensemble of runs for the SBLN parameter values. These four images give a comparison of, from left to right, the observed rivalry network, the Ensemble SBLN 1% edge agreement, the Ensemble SBLN 50% edge agreement, and the Ensemble SBLN 100% edge agreement.

For comparison, we simulated a random model that incorporates only the distance between nodes. In particular, we constructed a collection of randomly weighted Geographical Threshold Graphs by fixing the locations of the nodes and sampling the weights, η_i, independently from a uniform distribution. We selected a threshold to yield a median of 69 edges. Figure 2.7 displays the percent agreement of each possible edge for the Ensemble SBLN, a collection of randomly weighted Geographical Threshold Graphs, and the Ensemble BMN. For visualization, the edges for each ensemble were sorted separately in descending order based on percent edge agreement. In the Ensemble SBLN, there is 100% edge agreement for the existence of 39 of the edges (corresponding to the first 39 edges of the Ensemble SBLN along the horizontal axis in Figure 2.7). The 100% edge agreement network in Figure 2.6 shows these edges. All runs in the Ensemble SBLN consistently agree on the nonexistence of 309 edges (corresponding to the last 309 edges of the Ensemble SBLN in Figure 2.7). These are the edges not appearing in the 1% edge agreement network in Figure 2.6.

The transition between edge existence and nonexistence in the Ensemble SBLN is marked by a steep drop over 58 edges. The collection of randomly weighted Geographical Threshold Graphs displays a large degree of stochasticity indicated by fewer edges with 100% edge agreement and the more gradual decline of edge agreement. The Ensemble BMN appears to have a smaller degree of stochasticity with more edges with 100% edge agreement and a steeper decline than the Ensemble SBLN and the collection of randomly weighted Geographical Threshold Graphs. Despite the stochasticity observed in these models, there is agreement among the edges of the Ensemble BMN and Ensemble SBLN, maintaining some structure within the simulated networks.

2.4.2 Long Term Behavior of the SBLN

The simulated network, through the movements of each of the agents, evolves as the simulation progresses. Because of this evolution, it is natural to ask if any sort of steady state is achieved. Keeping in mind the stochasticity of the model and the interaction between the network and the agents' movements, an equilibrium in the strictest sense cannot be obtained.
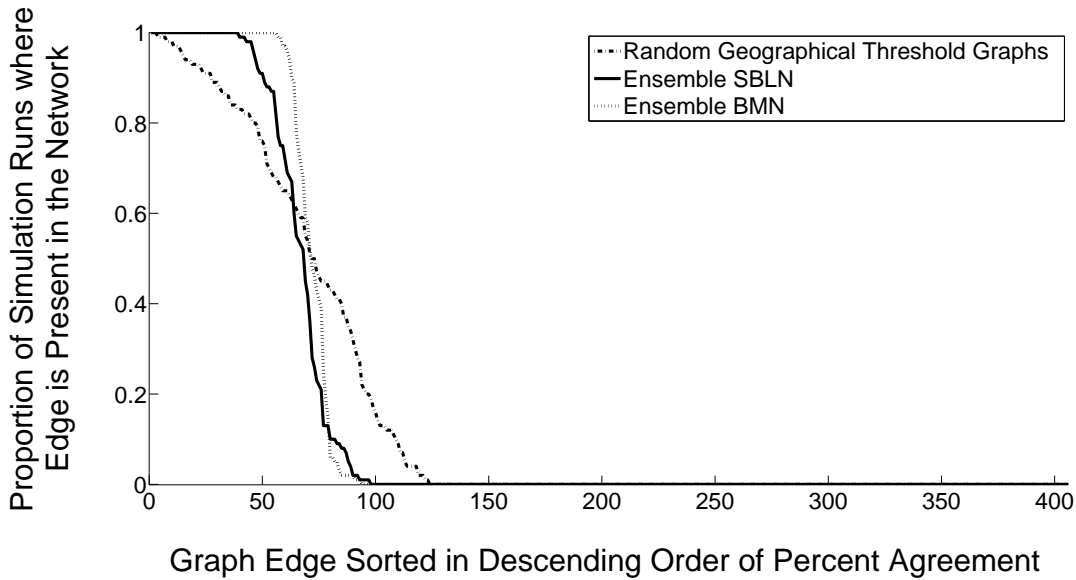
Figure 2.7: Plot of the edge persistence for the Ensemble SBLN (solid), Ensemble BMN (thin-dash), and an ensemble of random Geographical Threshold Graphs (thick-dash). The randomly weighted Geographical Threshold Graphs were constructed with random weights and have a median of 69 edges present. The edges were sorted in descending order according to the proportion of simulation runs where the edge is present in the network. Each ensemble of runs were sorted separately, yielding different edge numbers among ensembles.

Despite this, the results indicate there is limiting behavior of the observed metrics as the simulation progresses. Figure 2.8 displays the density and accuracy over the progression of the simulations for the Ensemble SBLN; for the definition of these metrics, refer to Section 2.4.3. Each run is observed every 1,000 iterations and the results of each simulation are shown as a thin line. The average metric value at each iteration is calculated and plotted as the thick line. For visual investigation the vertical axis on the accuracy plot has been refined to include only the area of interest. Accuracy values can range from 0 to 1. Both of these plots suggest that after a short phase of initialization, the metrics of each run appear to stabilize. For the average values of the density and accuracy of the last iteration, refer to Table 2.2 and 2.3.
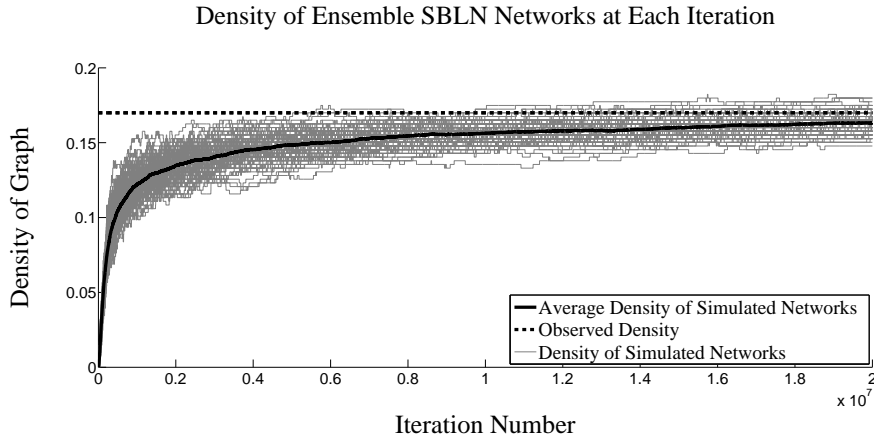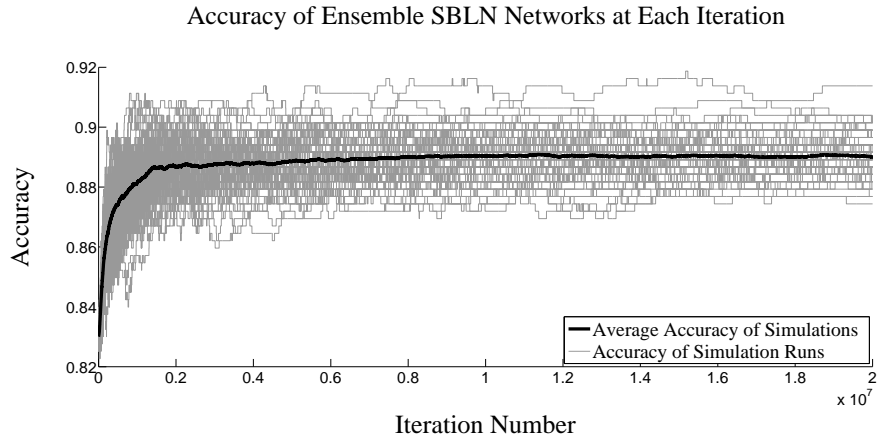
Figure 2.8: Plots of the accuracy (top) and the density (bottom) of the SBLN over the $2 \cdot 10^7$ iterations. Each of the 100 Ensemble SBLN runs are plotted by thin lines. The average over all the runs at each sampled iteration is shown with the solid, thick line. The density of the observed network is shown in the thick, dashed line. For visual investigation the vertical axis on the accuracy plot has been refined to include the area of interest. Accuracy values can range from 0 to 1.

The Ensemble SBLN is shown to exhibit stable long term behavior the simulated rivalry network, with some variation due to stochasticity. Despite this variation, the network emerging from the model results in metrics with a small deviation from the average. Further, the stochasticity observed may provide a more realistic model of the true rivalry structure.

Research has demonstrated that the rivalry networks that link gangs tend to be stable over time [TR10, RFT10, Pap09], and that the activity spaces of gangs are anchored to specific places [TCE05, MP93]. However, over longer periods of times, the membership ranks of gangs may ebb and flow due to incarceration, individuals "aging out" of active status, or other forms of incapacitation [TRR03]. Thus, gangs may lay dormant and, though identified in the rivalry network, not actually participate in violence. In extreme cases, either through high levels of victimization at the hands of rival gangs or through the focused enforcement of law enforcement agencies, a gang may simply disappear altogether. As more data become available, inherent stochasticity in the model may allow for further understanding of the rivalry structure.

2.4.3   Metrics Used for Analysis

We analyze our model according to several common metrics of accuracy, shape, and community structure. These statistics are compared to the observed rivalry network in [RFT10], in which there are 69 rivalries among 29 active gangs in the Hollenbeck policing precinct.

2.4.3.1   Accuracy Metrics

The first measures of interest are the raw values for the number of correct and incorrect edges. These values provide a means for evaluating the performance of the model. However, when comparing the observed network with the constructed network, each edge can be correct in two ways and incorrect in two ways. First, the constructed network can correctly identify an edge, *true positive* (TP), and correctly identify the lack of an edge, *true negative* (TN). The constructed network can also be wrong in two different ways. It can place an edge where there is none, *false positive* (FP), and also fail to place an edge where there is one, *false negative* (FN).

There are three quantities that are of particular interest that summarize the TP, TN, FP, and FN values. First is the *accuracy* of the model. The accuracy in the context of edges

on a graph is defined by

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{2.7}$$

The ACC ranges between 0 and 1, with 1 being a perfect reproduction of the observed network. This measure is proportional to the Q_α measure discussed in [BBC00]. The *F1 score* provides another measure to analyze the accuracy of the predicted network, [Seb02, YL99], and is defined as

$$\text{F1} = \frac{2TP}{2TP + FP + FN}. \tag{2.8}$$

An exact replication of the network would have an F1 score of 1. The other summary statistic for the raw closeness to the network is the *Matthews Correlation Coefficient (MCC)* [Mat75, BBC00]. This measurement varies between -1 and 1, where a value of 1 is a perfect prediction. The MCC is defined as follows:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \tag{2.9}$$

The measurements of the TP, TN, FP, and FN provide one means by which to determine the success of the model. However, they do not describe how these correct or incorrect measurements affect the overall network structure. A strong model would create a network that is not only accurate but also, maintains the same network structure, even in the event that the individual connections are the not same.

2.4.3.2 Shape Metrics

We would like to verify that the simulated network has a similar shape to that of the true network. To do this, we calculate the graph *density*, standard *variance* of nodal degree and Freeman's *centrality measure* of the graph. For definition of these metrics, see [WF09, Fre79]. The density of a network provides a normalized average of the degrees of the network. Networks with the same number of edges and nodes have the same density measure. The

centrality measure of the graph is defined to be

$$\sum_{i=1}^{N} \frac{\text{maxDegree} - \text{degree}(i)}{(N-1)(N-2)}. \tag{2.10}$$

The centrality measure and the variance of the nodal degree provide measures for the spread of the degrees.

2.4.3.3 Metrics of Community Structure

Another class of measurements which are of interest to observers of social networks are those that describe a network's community structure. Here, we use the idea of community structure strictly as a network property. The *degree distribution* has been widely used to understand the overall network structure [New01c, NSJ01, New03, AB02]. We compare the nodal degree cumulative distribution function (CDF) of our simulations with the observed network. The nodal *clustering coefficient* is another popular metric to analyze the community structure [AB02, New01c, WF09]. Intuitively, this is the proportion of a node's neighbors that are also neighbors with one another to the total possible connections of this type. This measure is calculated for each node yielding a distribution of clustering coefficients. From this distribution, the mean clustering coefficient over all nodes is computed.

2.4.4 Evaluating Models Using Graph Metrics

2.4.4.1 Accuracy Metric Results

Table 2.2 provides the accuracy measures for the GTG, BMN, Ensemble BMN, SBLN, and Ensemble SBLN. The SBLN outperforms all of the other networks on all of the accuracy metrics. Observe that the GTG also performs well on these metrics. The Ensemble SBLN metrics are comparable to the GTG and BMN metrics. In particular the average number of true negatives (TN) and false positives (FP) perform slightly better for the Ensemble SBLN than for the GTG, BMN, and Ensemble BMN. The Ensemble SBLN average of the true positives (TP) and false negatives (FN) performs slightly worse than the GTG and

BMN. Only the GTG and SBLN have higher accuracy, F1 Score, and MCC values than the Ensemble SBLN average.

	SBLN	Ensemble SBLN Average $\pm\ \sigma$	GTG	BMN	Ensemble BMN Average $\pm\ \sigma$
TP	50	45.50 ± 1.269	48	47	43.61 ± 1.380
TN	320	316.1 ± 2.424	316	313	309.2 ± 1.390
FP	17	20.90 ± 2.424	21	24	27.76 ± 1.39
FN	19	23.50 ± 1.269	21	22	25.39 ± 1.380
ACC	0.9113	0.906 ± 0.007	0.8966	0.8867	0.8691 ± 0.0051
F1 Score	0.7353	0.6722 ± 0.020	0.6957	0.6714	0.6213 ± 0.016
MCC	0.6822	0.6069 ± 0.025	0.6333	0.6031	0.5424 ± 0.019

Table 2.2: Accuracy measures for the SBLN, Ensemble SBLN, GTG, BMN, and Ensemble BMN. The σ denotes the standard deviation of the ensemble metric values.

2.4.4.2 Shape Metric Results

Table 2.3 provides the shape measures for the observed network, GTG, BMN, Ensemble BMN, SBLN, and Ensemble SBLN. Note that the density of the GTG is exactly the same as the observed rivalry network by construction, but it does not perform well for the nodal degree variance. The density for the BMN, Ensemble BMN, SBLN, and Ensemble SBLN are all close to the observed network. The BMN and the Ensemble BMN average have the closest nodal degree variance to the observed network's nodal degree variance. The centrality measure for the SBLN is the closest to that of the observed network.

	Density	Variance of Nodal Degree	Centrality	Mean Clustering Coefficient
Observed	0.16995	4.32105	0.20106	0.4921
SBLN	0.16503	3.54578	0.16799	0.6325
Ensemble Average SBLN $\pm \sigma$	0.16355 $\pm$ 0.005593	3.66423 $\pm$ 0.48395	0.15040 $\pm$ 0.01883	0.6364 $\pm$ 0.02718
GTG	0.16995	9.97622	0.27778	0.6719
BMN	0.17488	3.88585	0.15741	0.7540
Ensemble Average BMN $\pm \sigma$	0.17579 $\pm$ 0.004546	3.93926 $\pm$ 0.41351	0.16065 $\pm$ 0.02635	0.7009 $\pm$ 0.02681

Table 2.3: This table provides the shape measures for the observed network, SBLN, Ensemble SBLN, GTG, BMN, and Ensemble BMN. The σ denotes the standard deviation of the ensemble metric values. Note that the density of the GTG is exactly the same as the observed rivalry network by construction.

2.4.4.3 Community Structure Results

The cumulative distribution function (CDF) of nodal degree for the observed network, GTG, BMN, normalized Ensemble BMN, SBLN, and the normalized Ensemble SBLN are shown in Figure 2.9. A normalized ensemble CDF shows the CDF of the degree distribution of all runs divided by the number of runs. The SBLN and the normalized Ensemble BMN have the most similar distributions as the observed network. The normalized Ensemble SBLN performs better than the GTG and the BMN. In the same figure, the normalized Ensemble BMN and SBLN are plotted with two standard deviations above and below together with the observed network distribution. Here we see that there is a smaller standard deviation for the normalized BMN than the normalized SBLN. Even with the standard deviations, the degree distributions of both classes of networks are close to that of the observed degree

distribution.

The mean clustering coefficient for each of the networks is seen in the last column of Table 2.3. On this measure of community structure, the SBLN and Ensemble SBLN average outperform all other networks. The Ensemble BMN average has the farthest mean clustering coefficient from the observed network.
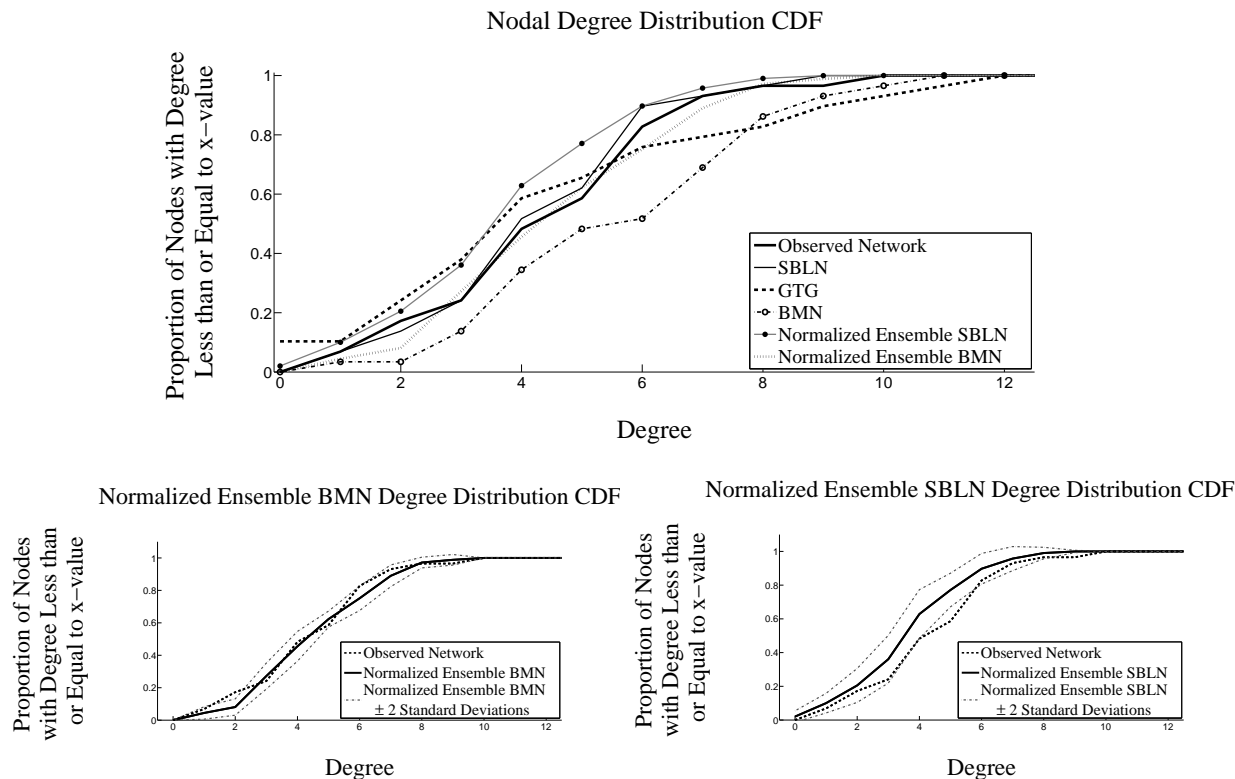
Figure 2.9: The top figure plots together the cumulative distribution functions of the degree distribution for the observed network (thick-solid), GTG (thick-dashed), BMN (dot-dash), normalized Ensemble BMN (thin-dash), SBLN (thin-solid), and normalized Ensemble SBLN (dot-solid). A normalized ensemble CDF shows the CDF of the degree distribution of all runs divided by the number of runs. The normalized Ensemble BMN (bottom left) and SBLN (bottom right) are plotted with two standard deviation above and below (thin-dash) with the observed network distribution (thick-dash).

41

2.4.5 Summary of Results

In all metrics except the density, the SBLN performs better than the GTG (note that the density measure of the GTG is exactly the same as the observed network by construction). Although the GTG is unable to closely replicate the standard shape measures, it has fairly high accuracy values. The Ensemble SBLN average performs similarly to the GTG in the accuracy, but performs better with shape measures, even with the stochastic considerations. On average, the Ensemble SBLN produces a slightly more accurate degree distribution than the GTG. The BMN is able to reproduce the degree distribution fairly well, however, the BMN and Ensemble BMN average have lower values for the accuracy (ACC), Matthews Correlation Coefficient (MCC), and F1 Score when compared to the other models. Our analysis demonstrates that the SBLN is the strongest model in reproducing the observed rivalry network.

2.5 Sensitivity Analysis

Our objective in this section is to understand the effects of the input parameters on the system by comparing the different metrics of the resulting networks as the parameters change. Due to computational constraints, we perform a local analysis of the parameter space around the SBLN parameters specified in column 3 of Table 2.1.

In particular, we perturb one parameter at a time by 30% from the SBLN parameter values in 10% increments. To account for the stochasticity inherent in the model, each perturbation was run using the same 25 seed values for the random number generator. The range of each parameter examined is listed in Table 2.4.

For each simulation run, we compute the accuracy, Matthews Correlation Coefficient, F1 score, centrality measure, variance of nodal degree, density, and mean clustering coefficient for the resulting network. Plots of each combination of metric versus parameter values were created for the general analysis. Three examples of parameter and metric combinations with

Bounded Pareto Scaling Parameter	k	$\in$	$[0.77 \, , \, 1.43]$
Von Mises Scaling Parameter	κ	$\in$	$[2.45 \, , \, 4.55]$
Largest Maximum Jump Length	A	$\in$	$[140 \, , \, 260]$
Smallest Maximum Jump Length	a	$\in$	$[70 \, , \, 130]$
Boundary Permeability	B	$\in$	$[0.14 \, , \, 0.26]$
Network Threshold	T	$\in$	$[0.028 \, , \, 0.052]$

Table 2.4:   Ranges of the parameters used in the sensitivity analysis. Each parameter was changed 30% from the SBLN parameters in 10% increments. For SBLN parameter values refer to the Hollenbeck column of Table 2.1

more dramatic results are plotted in Figure 2.10. In this figure, we display the variance of nodal degree versus the smallest maximum jump length, a, and the network threshold, T. We also display the density versus the Bounded Pareto scaling parameter, k, where the vertical axis has been rescaled for visualization. The dots represent the metric values of the simulation run at the specified parameter. The solid curve indicates the average metric value over all runs at each parameter value.

As seen in Figure 2.10, the plots varying the network threshold and Bounded Pareto scaling parameters have a negative trend on average. The smallest maximum jump length, however, shows a positive trend. The stochastic effects can also be observed by the range of metric values associated with each parameter input, as illustrated by the dots in Figure 2.10. These plots suggests that stochasticity may influence the metric values for a particular run, and on average the resulting metric output is sensitive with respect to these parameters.

These plots give a view of how the particular metric and parameter value interact. We changed all of the parameter values by the same 30% from the SBLN parameters, and so we can compare plots with the same metric. For example in Figure 2.10, we can see that in general nodal degree variance for the smallest maximum jump length has a steeper trend than the nodal degree variance for the threshold, but we can not compare the trend of the
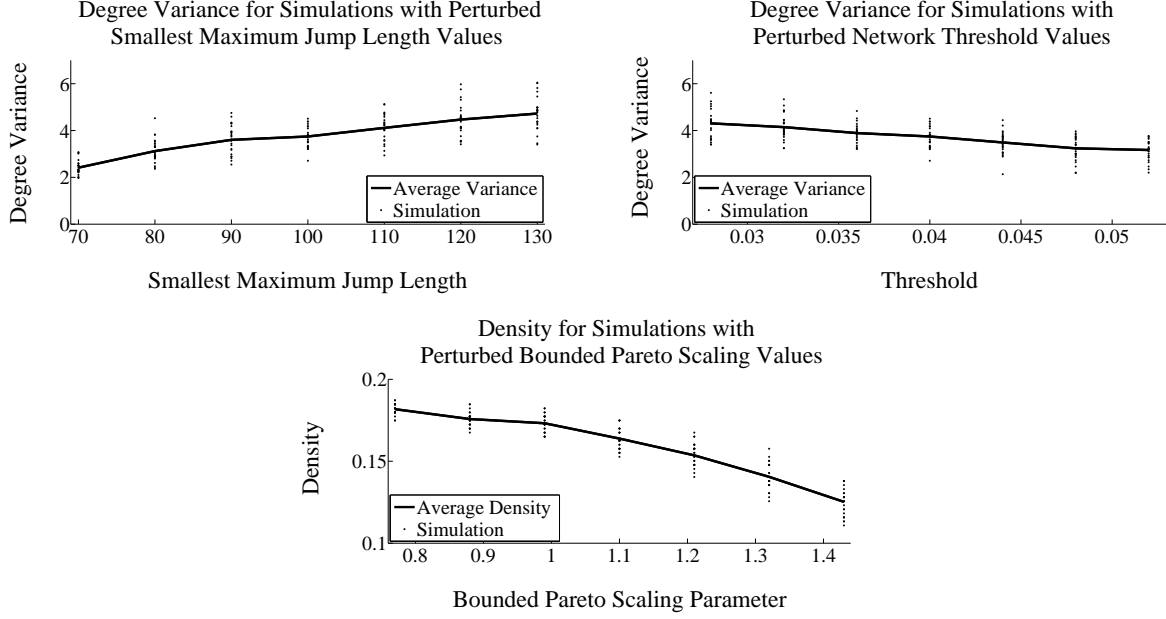
Figure 2.10: Plots of the nodal degree variance versus the smallest maximum jump length (top left), and the network threshold (top right). We also display the density versus the Bounded Pareto scaling parameter (bottom), where the vertical axis has been rescaled for visualization. The solid curve indicates the average metric value over all runs at each parameter value. The dots represent the metric values of the simulation run at the specified parameter.

nodal degree variance plots directly to that of the density plot.

To compare the effects of all the parameters on all metrics, we rescale the data points to percent deviation from the SBLN parameter values. For example, when considering the affects of the Bounded Pareto scaling parameter, k, on the density metric, we rescaled the observed data points

$$(k_i, \text{density}_i) \mapsto \left(\frac{k_i - k_{\text{SBLN}}}{k_{\text{SBLN}}} \ , \ \frac{\text{density}_i - \overline{\text{density}}_{\text{SBLN}}}{\overline{\text{density}}_{\text{SBLN}}} \right),$$

where k_{SBLN} is the SBLN Bounded Pareto scaling parameter. Here, $\overline{\text{density}}_{\text{SBLN}}$ is the

average density at the k_{SBLN} value for all 25 runs. A line was fitted to the rescaled data points, and the slope of this line was recorded. This process was repeated for each parameter and metric value combination.

The results are recorded in Table 2.5 and visualized in Figure 2.11. In Table 2.5, negative values indicate a negative slope of the best fit line to the scaled data, and positive values indicate a positive slope. Slopes with a greater magnitude indicate a stronger correlation between the metric and parameter. To get a clearer impression of overall sensitivity of the system, this information is displayed in Figure 2.11. The dark, and light, intensities of the color map represent large positive, and negative, values of the best fit line slope.

	k	κ	A	a	B	T
Accuracy	-0.0120	-0.0031	0.0011	0.0001	0.0031	0.0000
MCC	-0.2066	-0.0161	0.0023	0.1458	0.0293	-0.0000
F1 Score	-0.2149	-0.0146	0.0018	0.1562	0.0278	-0.0000
Centrality	-0.1705	-0.0100	-0.0131	0.7119	0.0195	-0.1751
Nodal Degree Variance	-0.4385	-0.0489	-0.0146	0.9456	0.0154	-0.5412
Density	-0.7410	-0.0080	-0.0114	0.6131	0.0640	-0.2460
Mean Clustering Coefficient	-0.3812	-0.0011	0.0005	0.1484	0.0084	-0.0614

Table 2.5: Slope of the best fit to the rescaled data for each metric and parameter combination. For reference, coefficients that correspond to the images in Figure 2.10 are highlighted in bold font. Figure 2.11 displays this information in a color map.

In general, the metrics are not very sensitive to the von Mises parameter, κ, the largest maximum jump length, A, and the boundary permeability, B, within the parameter space investigated. On the other hand, the Bounded Pareto scaling parameter, k, the smallest maximum jump length, a, and the network threshold, T, have the most influence on the metrics. As seen in the table and figure, the accuracy measures are fairly robust to changes
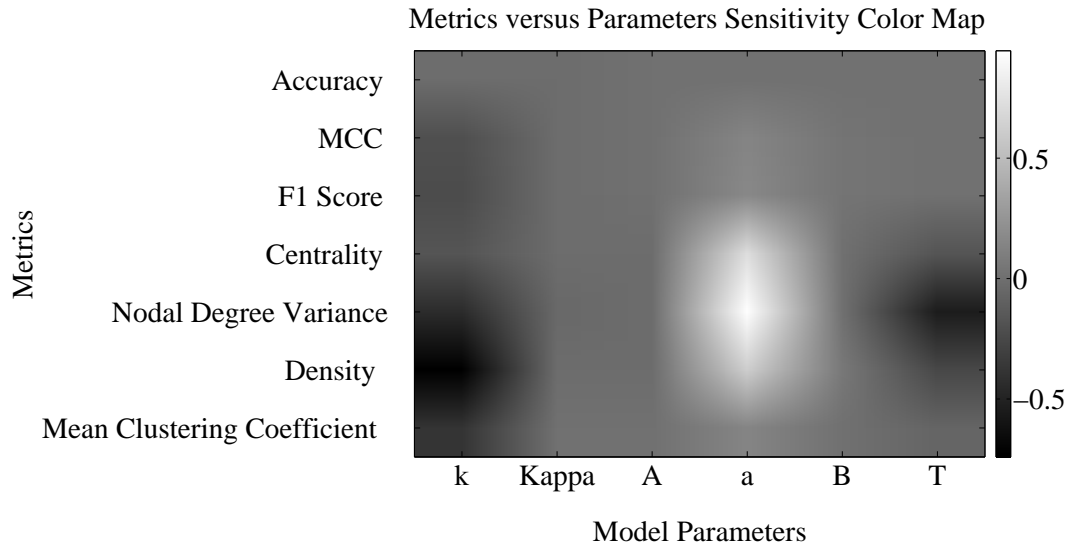
Figure 2.11: Slopes of the best fit line to the rescaled data for each parameter and metric combinations depicted in a color map. The parameters varied include the Bounded Pareto scaling parameter, k, the von Mises scaling parameter, Kappa, the largest maximum jump length, A, the smallest maximum jump length, a, the boundary permeability, B, and the network threshold, T. The scale to the right of the image gives the slope values. Tones close to the center of the scale represent combinations where the metrics are not very sensitive to the respective parameter. Combinations with tones at the ends of the spectrum (black and white) represent metrics that are sensitive to the respective parameter. The numerical values are also stored in Table 2.5.

in all parameter values. Further, note that nodal degree variance and density measures appear to be the most affected by the changes in these parameters.

The Bounded Pareto scaling parameter values result in negative slopes for all metrics. This is to be expected because an increase in the Bounded Pareto scaling parameter will decrease the likelihood of larger jumps and result in fewer edges. This phenomena is particularly evident in the density metric. Also this parameter appears to have the most effect on the accuracy measures, in particular the MCC and F1 score.

Increasing the network threshold parameter also has a negative effect on the shape and community metrics. By increasing the network threshold, the number of connections decreases. This in turn decreases the density, nodal degree variance, centrality, and mean clustering coefficient. On the other hand, increasing the smallest maximum jump length increases the connectivity of the network by allowing for larger jumps in areas of high road density. The effect of changing this parameter is more significant than changing the largest maximum jump length. Interestingly, as the largest maximum jump length increases, the connectivity decreases. This could be where attempts to cross boundaries are more likely to occur. The lower portion of Hollenbeck is approximately 300 units wide and has many boundaries. When varying the largest maximum jump length between 140 to 260, it becomes very probable that at least one boundary cross would be attempted. At this point, the boundary permeability is expected to play a stronger role in the simulation.

Depending on the network, changes in the number of connections could be more or less beneficial in terms of accuracy. Further, small changes in the connectivity, i.e. the existence or non-existence of an edge, could have small effects on the accuracy measures and large effects on the shape measures, as seen for our simulations in the case of the network threshold parameter.

2.6   Discussion

Using biased truncated Lévy walks with semi-permeable boundaries, we have designed an agent-based model for gang members in Hollenbeck that incorporates quasi-realistic movement rules as well as physical geographic features existing in Hollenbeck. We have shown that it is able to simulate a gang rivalry network similar to the one observed in [TRR03, RFT10]. The Simulated Biased Lévy walk Network (SBLN), the Brownian Motion Network (BMN), and an instance of a Geographical Threshold Graph (GTG) were compared to the observed rivalry network using measures of accuracy, shape, and community structure.

The GTG method performs well on the accuracy metrics and provides an alternative,

computationally inexpensive method to construct the rivalry network. One could extend this model to incorporate boundary information by increasing the distance function $d(n_i, n_j)$ if n_i and n_j are in distinct regions, see Section 2.3.1. However, the GTG is limited to reproducing only the rivalry network and does not lend itself to understanding other phenomena, such as the gang member mobility and the locations of interactions between gang members. On the other hand, agent-based models such as the BMN and SBLN provide a reasonable approximation to the observed network and can be easily extended to include policing strategies, the location of violence, retaliatory behavior, and effects of injunctions.

Although the BMN accuracy results were not as strong as the GTG and SBLN results, this method was able to reproduce a similar shape and community structure as the observed rivalry network. This model was able to incorporate geographical features, but ignored directional decisions of the agents. One major problem with this model is that the stopping iteration for the model was artificial, in that we chose to stop it at the observed peak in accuracy. In general, there may not be an observed network, and so it would be difficult to determine stopping criteria. However, our proposed SBLN model exhibits long term stabilization of the accuracy and density metrics.

The SBLN is the best model in replicating the observed network. Further, it allows for easy incorporation of geographic features and alternate movement dynamics, while maintaining a high level of accuracy and allowing for evolution in the observed system. The success of this model and the flexibility of the method leads us to believe that the SBLN could also facilitate in understanding other social phenomena of interest related to gang violence. In fact, this model is able track the location of the agents' interactions during the simulation. This can be compared to violence data for the Hollenbeck area, and preliminary work has been done in this direction. Figure 2.14 shows the locations of the interactions among agents for one of the Ensemble SBLN simulation runs and the density of gang-related violent crimes in Hollenbeck from 1998 through 2000. The juxtaposition of these two plots emphasizes the similarities between the two and illustrates the potential predictive capabilities of this kind of approach. Though movement and interaction rules may need to be slightly altered to

provide a closer match to the data, the current model provides a baseline model for further analysis and investigation of the gang rivalry violence in Hollenbeck. Another potential avenue for future work is to use this model to investigate territoriality and respect as a key elements in the motivation for violence [TRR03].

The current model does not account for the difference between positive, negative, and neutral interactions. Instead the SBLN records interactions between agents with the implicit assumption that these are negative interactions. It is possible that there exists an alliance or truce between two gangs, and such phenomena has been observed in street gangs in Chicago [BB93a]. The proposed SBLN model could be extended to include this.

Now that some of the influences of the geography and the interactions between the agents and the network are better understood, it could be beneficial to reformulate the agent-based model as a PDE. This alternative approach may allow for a deeper understanding of the model and may provide a rigorous analysis of the network dynamics.

Pursuing a model that accurately describes the violent behavior in Hollenbeck is of great value, since Hollenbeck is one of the most violent areas in Los Angeles [TRR03, Inf08]. The advantage of approaching this serious problem using a computer simulation is twofold. First, these simulations may help us understand the underlying mechanisms that are involved in producing violent behavior among the gangs in Hollenbeck. Second, if the simulation can accurately model the social phenomena of interest, then we might gain some insight into how intervention strategies could alter the existing gang rivalry system. The costs of implementing these changes in the simulation are relatively small compared to those costs of public funds needed to implement experimental interventions. If the Hollenbeck area can be well understood by this approach, there may be hope in understanding, and potentially mitigating, other areas of intense violent behavior.
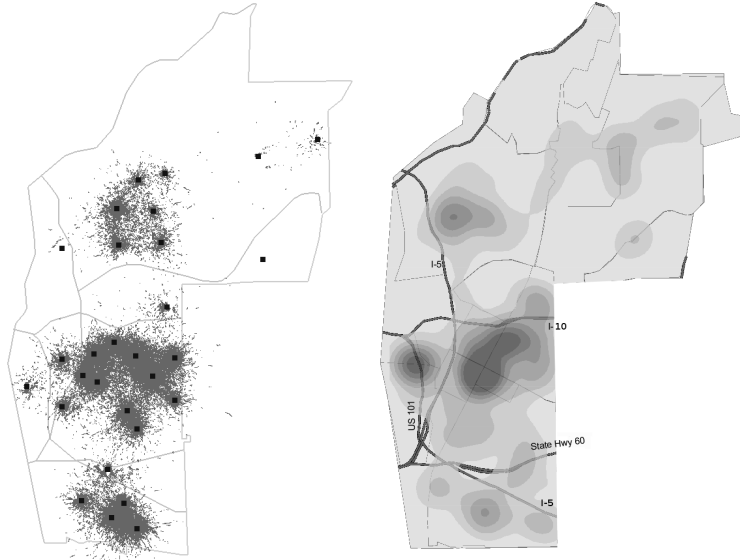
Figure 2.12: Locations of all the interactions between agents for one of the Ensemble SBLN runs (left). Density map of gang-related violent crimes in Hollenbeck between 1998 and 2000 (right).

2.7 Comparing Simulated Violence Locations Observed Violence Locations

We wish to compare the observed data of Hollenbeck with the locations of interactions produced by the simulation. The observed and simulated data are point process data. This means that each point is located in (x,y) space and the existence and non-existence of a point is of interest.

The first test conducted is that of complete spatial randomness, or CSR, to determine if the data could have been obtained from an unstructured homogeneous Poisson process. To test for CSR a quadrant count was conducted. This amounts to breaking the valid region A into equal area parts. A Poisson process will have on average the same number of points in each region, regardless of the shape of the region. Deviations from the null hypothesis of a Poisson process are determined using a χ^2 test.

Before initiating a descriptive statistic, one must make some assumption on the form of data. The first of which is isotropy, i.e. that one unit of distance in the Y direction has the same effect as one unit of distance in the X direction. If the degree of anisotropy is known, the definition of the distance h can be adjusted. The data is also assumed to be an observation of a stationary point process. This means that the underlying stochastic process is unchanged if the origin of the index set is translated [Rip81].

Determining the degree of clustering observed in either data set can be approximated by the K-function, $K(h)$, defined by

$$K(h) = \frac{E[\text{number of events within distance } h \text{ of an arbitrary event}]}{\lambda}. \tag{2.11}$$

Here λ is the number of points per unit area in the domain A. In practice, λ is often approximated by $N/|A|$ where N is the number of points in the observed domain and $|A|$ is the area of the domain [Dig83, DMC00].

When edge effects are ignored, $K(h)$ is approximated by

$$\hat{K(h)} = N^{-1} \sum_i \sum_{i \neq j} I(d_{ij} < t), \tag{2.12}$$

where d_{ij} is the distance between points i and j. Here $I(\cdot)$ is the indicator function. Edge effects bias the estimator for large values of h. As a rule of thumb, h should be smaller than half the distance of the most narrow portion of the domain.

To determine the existence of clustering or inhibition, the approximate $\hat{K}(h)$ function can be compared to the theoretical value of $K(h)$ for a Poisson process. In this case, $K(h) = \pi h^2$.

Finally, another method of comparing point processes is through density estimation. For this analysis a Gaussian is placed on each of the data points. The underlying density is estimated by summing over the Gaussian functions and then normalizing so that the density integrates to one.

2.7.1 Hollenbeck and Simulated Data

The observed data contains the time and location of each violent event, as well as the type of event, suspect gang and victim gang from 1999-2002. In all there are 1207 events with two outliers outside of the region of Hollenbeck. For analysis these two events are ignored.

The simulated data was obtained from the Hollenbeck parameters in Table 2.1. So that the initialization phase of the simulation would not influence the results, the simulation was run for 10^7 iterations, and the next 30,000 interaction locations were considered. For comparison, one representative run of the SBLN model was used. Note that in the simulation each of the locations represents an interaction between two gangs, and not necessarily a violent event. Both datasets rejected the null hypothesis of the data coming from a Poisson process. The Gaussian density estimation for the observed data and the simulated data are shown in Figure 2.13. From these plots, one can see that the general location of the simulated events lines up fairly well with those of the observed events.

The density of the simulated location data appears to be more clustered than that of the observed. This observation is further corroborated when one considers the clustering of these two data-sets seen in Figure 2.14. Both of the data sets show more clustering than what would be expected if the process were Poisson. However, the observed data (left) is not as clustered for small distances than that of the simulated data. This may imply that there are some additional movement rules that need to be incorporated. Further, it could be that a more sophisticated interpretation of "interaction" needs to be employed. For instance, once an agent has interacted with an agent of a rival gang, they could be placed back at their respective set-space after the interactions. It is more plausible that after a series of violent interactions with other gangs, the gang member would go back to their home set-space, and not remain located "at the scene of the crime".
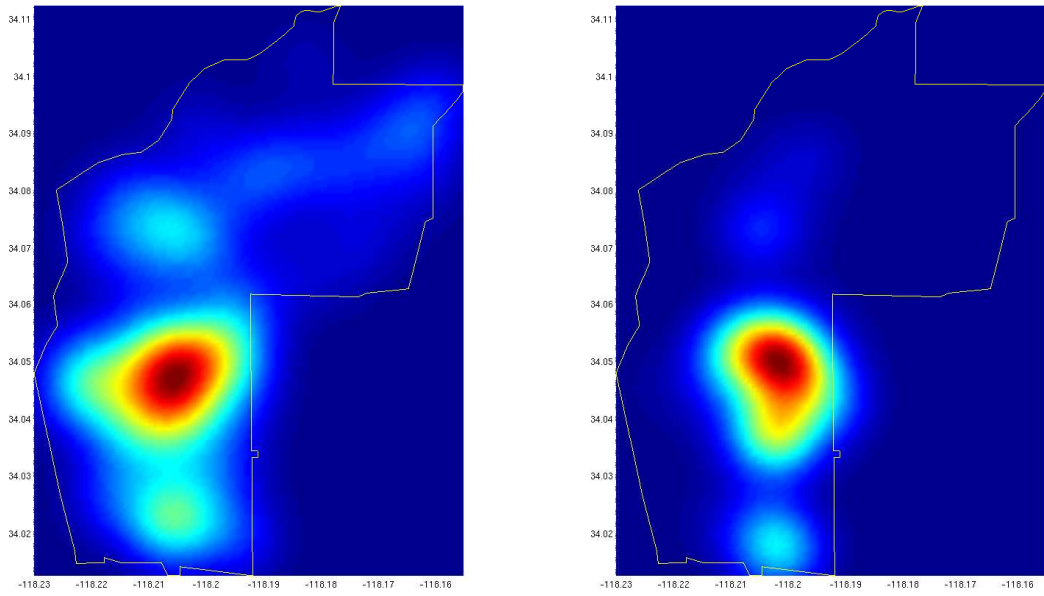
Figure 2.13: Comparison of Gaussian density estimates between the observed (left) and simulated (right) data, with $\sigma = 25$.
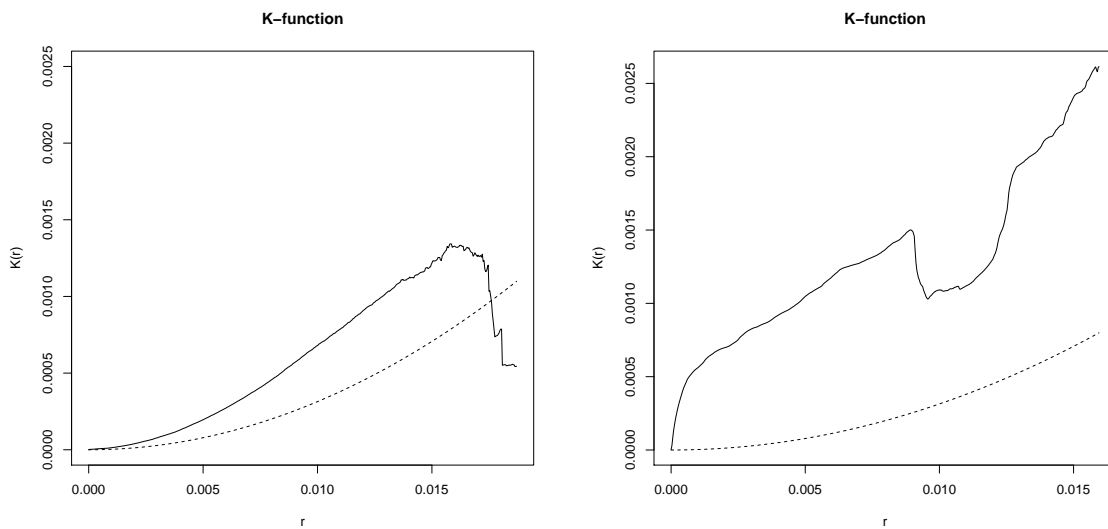
Figure 2.14: Comparison between observed (left) and simulated (right) $K(h)$ functions. The $K(h)$ function associated with a Poisson process is displayed in the dotted line in both images.

CHAPTER 3

Inferring gang rivalry association

3.1 Introduction

Gang violence data sets are a valuable source of information on gang systems. Unfortunately, field data sets are often incomplete. For instance, in the field the physical details such as the time and location of a violent event can be recorded with reasonable accuracy, however, the parties involved are not always identified. Not surprisingly, methods that can extract information from these data sets, in the presence of partial data, is of great interest to law enforcement and sociologists.

To make gains in extracting crucial information from the data and infer gang involvement for these data sets the nature of the gang activity is exploited. These violent events tend to be dyadic between gangs, therefore it is natural to formulate these events as a realization of a stochastic process occurring on the edges of the rivalry network. In theory for each edge in the network there exists a different stochastic process. In our analysis we use identical parameters to generate synthetic data, but the method does not assume that the underlying parameters generating each process are identical. All events in the data occurs at a specified time and involves a pair of rival gangs, however, a subset of these events are unsolved crimes in which one or both of the rival gangs is not known.

It is important to note that the method developed in this chapter could be broadly applied to any social network involving activities in time between pairs of nodes on the network. The interest in the problem was inspired by examining data from the Hollenbeck Division of the Los Angeles Police Department, home to 29 street gangs with a well-known rivalry network

[RFT10, TR11, TRR03]. Possible geographic influences in the emergence of gang rivalry networks are discussed in Chapter 2 in [HSB11]. This work in Chapter 3 was conducted in collaboration with Erik A. Lewis and Andrea L. Bertozzi in [HLB12].

3.1.1 Formulation

Unlike other methods used to address missing data relating to social networks [Hof09, KRP10], the question at hand is not if a rivalry exists, but rather to which rivalry a violent event belongs. To do this, one must first understand the underlying stochastic process. This requires us to capture the behavior of criminal activity through computational means, much like in [BGJ09, BGK05, BB04]. Recently methods have been proposed in the literature to mathematically model gang violence. The authors in [HSB11] employ an agent-based model to investigate the geographic influences in the formation of the gang rivalry structure observed in Hollenbeck. These authors consider the long-term structure of the rivalry network embedded in space. In terms of the rivalry violence, a shorter time scale must be considered.

Violence among gangs exhibits retaliatory behavior [Dec96]. In other words, given an event has happened between two gangs, the likelihood that another event will happen shortly after is increased. A problem such as this is modeled naturally by a self-exciting point process. It is interesting to note that these models were first used to analyze earthquakes [VS08, Oga98, Oga88, ZOV02]. Since then, they have been used to model financial contagion in credit markets [EGG10, ACL10], viral videos on the web [CS08], terrorist activity in Indonesia [PW11], and the spread of infectious disease [MEH11].

The authors in [MSB11] and [EFL10] have successfully modeled the pairwise gang violence as a Hawkes process [Haw71b]. All of the events are associated with exactly one rivalry, or edge of a social network. The violence on each edge, k, is assumed to have the conditional intensity

$$\lambda_k(t|H_{\tau,k}) = \mu_k + \alpha_k \sum_{t > t_j} \omega_k e^{-\omega(t - t_j)}. \tag{3.1}$$

In this Hawkes process, the intensity $\lambda_k(t|H_{\tau,k})$ depends on the history of the process $H_{\tau,k} = \{t_1, t_2, \cdots t_{M_k}\}$, where M_k is the number of events for process, k. In this framework, the window of time, $[0, T]$, observed for each process in the network is the same. However, the number of events in each process, M_k, is stochastic, and therefore varies from process to process. The background rate of the process is defined by the constant μ_k. The expected number of offspring for any event is determined by the constant α_k, and the decay of the intensity back to the background rate is ω_k. Larger values for μ_k and α_k produce more background and offspring events respectively. Larger values of ω_k do not influence the total number of events, but rather the amount of clustering in time.

The authors of [SSB11] produce a mathematical framework to solve the incomplete data problem observed in gang violence data sets. In their work they use an optimization strategy that computes the weights to infer the rivalry affiliation of the incomplete data. In this formulation the authors prove that their optimization has a unique solution under mild constraints. This is a substantial contribution in inferring the affiliation of the unknown violent events. However, the authors of [SSB11] assume that the process parameters are known, an assumption that is often not feasible in practice. Further, finding the weights requires solving a computationally expensive optimization problem.

We propose an iterative method that (A) estimates the process parameters assuming the data is generated by the process defined by Equation 3.1 and (B) infers the process affiliation of simulated data via a direct method of computation. We iterate between (A) and (B) until the estimates for the unknown events converge. We call this the Estimate & Score Algorithm (ESA). The details of the ESA are described in Section 3.2. The ESA is tested on simulated data in Section 3.3, with analysis of the estimation of the parameters in the presence of missing data (see Subsection 3.3.1) and comparison of the proposed score functions with that of the Stomakhin-Short-Bertozzi (SSB) method in [SSB11] (see Subsection 3.3.2). In Subsection 3.3.3 there is an analysis of the runtime between the Stomakhin-Short-Bertozzi and the Forward Backward score functions used to update the weights (see Subsection 3.3.3). Subsection 3.3.4 contains an analysis of the convergence of the Estimation & Score Algorithm. This

method solves the more realistic problem of estimating the process and the weights. Further, the computation for the weight updates is more direct and therefore avoids performing the costly optimization scheme used in [SSB11]. This is a novel piece of work with many exciting extensions. A final discussion of the results and future work is presented in Section 3.4. As in [SSB11] we do not use field data in this dissertation, rather we generated point process data using similar parameters as observed in the field data for Hollenbeck [EFL10]. By using simulated data to test the algorithms we have actual ground truth evaluate the performance of the method.

3.1.2 Problem Formulation

The data is assumed to lie on a known social network containing K processes, where each of the K processes is a pairwise rivalry between two gangs. From this set of events, there are a total of N events where the time is known, but the processes affiliation is not known. Each of the N unknown events are placed into each of the K processes. Since the process affiliation is not known for all of the events in the network, each event is given an associated weight, $S_{i,k}$. Here $S_{i,k}$ is the ith element of the kth process. If the event is known $S_{i,k} = 1$. If $S_{i,k}$ is unknown, then it is assigned a number between 0 and 1 by our algorithm. We enforce the constraint that $\sum_{k=1}^{K} S_{i,k} = 1$.

A simplified representation of our problem formulation can be found in Figure 3.1. The known events are represented by circles and the unknown event is represented by a triangle. Here we can see that since we do not known the affiliation of the triangle event, it is placed in all of the other processes. We emphasize that this represents our lack of information about which rivalry it belongs to.

As indicated in Figure 3.1, for each process in the network events $e_{i,k}$ are indexed by increasing time, $t_1 \leq t_2 \leq t_3 \cdots \leq t_{M_k}$. Ordering the events in such a way has the consequence that the first missing element in time, for example, may have different indexes for different processes. In Figure 3.1 the triangle index in first process is the third event, $e_{1,3}$. However

58

the triangle in the Kth process is the second event, $e_{K,2}$. One can easily keep track of the local index of a unknown event for each process.
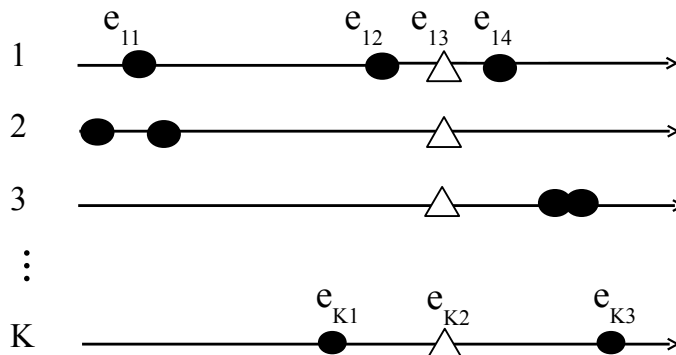
Figure 3.1: Simplified representation of the rivalry network with known (circle) and unknown (triangle) events. Note that since we do not know the affiliated process of this event, we place it in all processes. Associated with this event is a weights $S_{i,k} \in [0, 1]$ such that $\sum_{k=1}^{K} S_{i,k} = 1$.

3.2 The Estimation & Score Algorithm (ESA)

The proposed Estimation & Score Algorithm can be broken into three basic stages: initialization, parameter estimation, and updating the weights. This method is succinctly described in Figure 3.2.

3.2.1 Initialization

For this dissertation, there were two ways of initializing the Estimate & Score Algorithm. The first is used to infer rivalry affiliation given field data. After importing the data, the unknown events are identified and placed into each of the of the K processes. The weights,
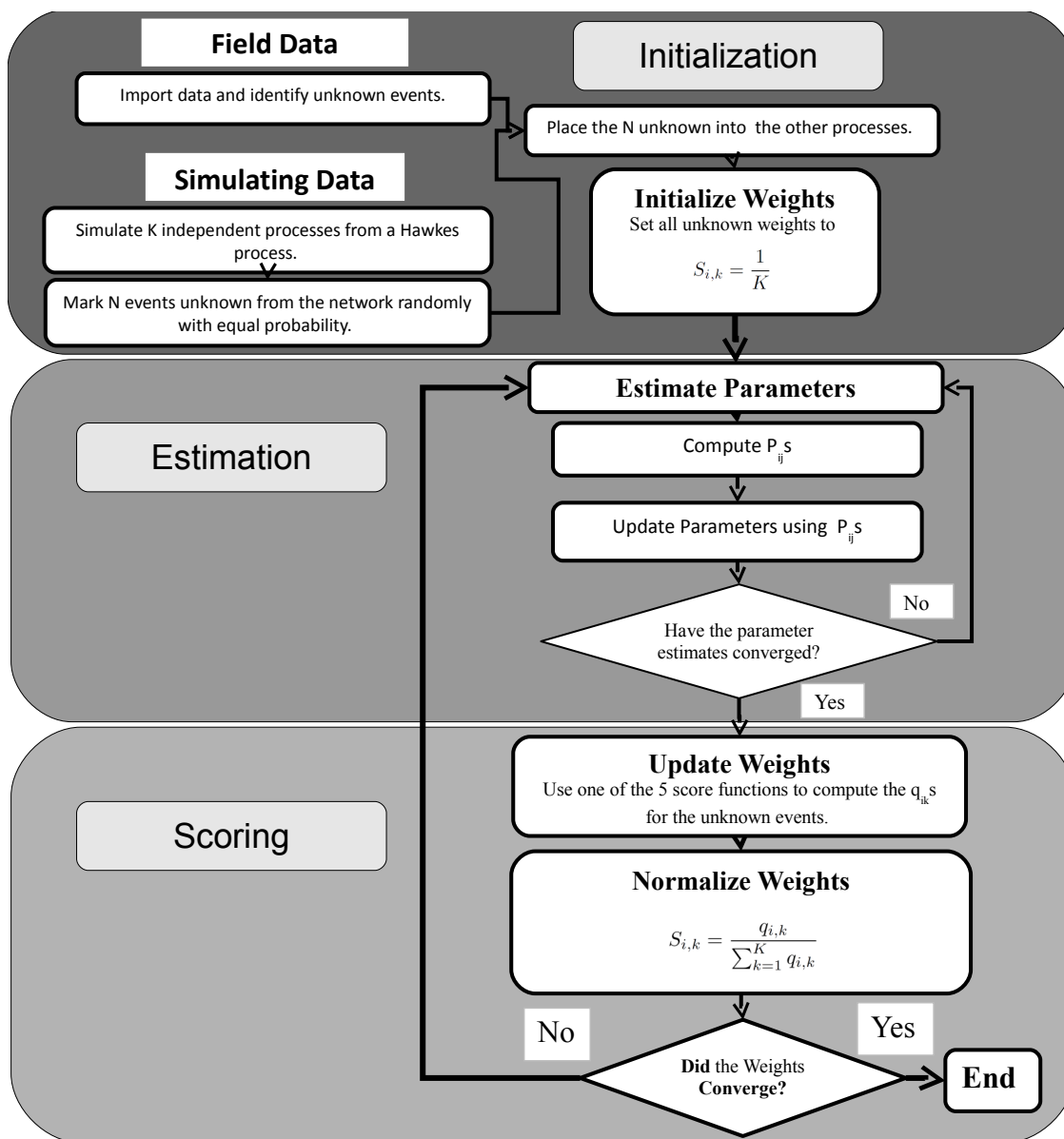
Figure 3.2: Flow chart of the Estimation & Score Algorithm. There are two ways to implement this method. The first, (left of initialization), is the algorithm used when given an incomplete data set. The second, (right of initialization), is the algorithm used in this dissertation to simulate the data and test the components of the ESA. The two main phases of the algorithm are the Estimation phase (see Section 3.3.1) and the Update Weights phase (see Section 3.2.3).

$S_{i,k}$, must also be initialized. If the event is known, then $S_{i,k}=1$. If the event is unknown then $S_{i,k} = \frac{1}{K}$.

An alternate initialization is used simulate data in order to test the components of the Estimate & Score Algorithm. In this case, data is generated from K independent Hawkes processes with given μ_k, α_k, and ω_k. From these data, choose N events at random from the network to mark as unknown. Place these N unknown events into each of the other processes. Initialize the weights such that for known events $S_{i,k} = 1$ and for unknown events $S_{i,k} = 1/K$. This initialization process is used in this dissertation to test the method and produce the results in Section 3.3.

3.2.2 Parameter Estimation

In the presence of no unknown events, there are both parametric [VS08] and nonparametric [ML08, ML10, SU09, LM11] ways to model the underlying stochastic process on each edge of the social network. For this work, we chose a parametric form for the triggering density to validate the model but the results could easily be extended to the nonparametric case. We note that, as is usual with nonparametric estimates, speed would be compromised for the sake of flexibility.

For this thesis, the data is assumed to be a realization of Equation 3.1, where the parameters are estimated using a method similar to the Expectation Maximization (EM) algorithm [DLR77]. An EM-like approach is taken because of the branching structure present in a Hawkes process. In such a process each event can be associated with a background or response event. However, given a realization from this process it is not immediately obvious whether an event is a background or response event. We can view this information as a hidden variable that we must estimate. In this way, every event in each of the K processes is assigned a probability $P_{i,j}^k$. The probability that event i is a background event is denoted $P_{i,i}^k$, and probability that event i caused event j is denoted $P_{i,j}^k$. This assumes that $t_i < t_j$. From this EM estimation, the approximation for each of the variables is altered to include

the weights for the unknown events. In fact, in the case where all the events are known, the estimation formulas are the same. This section derives the EM estimates when in the presence of missing data.

The classical log-likelihood function $\hat{\ell}_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k)$ for a general point process is

$$\hat{\ell}_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k) = \sum_{i=1}^{M_k} \lambda_k(t_i|H_{\tau,k}) - \int_0^T \lambda_k(t|H_{\tau,k})dt. \tag{3.2}$$

Incorporating the branching structure into the log-likelihood function, the event association is added as a random variable, $\chi_{i,j}$ such that

$$\chi_{i,j} = \begin{cases} 1 & \text{if event } i \text{ caused event } j \text{ and } i \neq j \\ 1 & \text{if event } i \text{ is a background event and } i = j \\ 0 & \text{else} \end{cases} . \tag{3.3}$$

This branching allows us to separate those events associated with the background μ_k and the response $g(t) = \alpha_k \omega_k e^{-\omega_k t}$. This leads to the altered log-likelihood function

$$\ell_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k) = \sum_{i=1}^{M_k} \chi_{i,i} \log(\mu_k) - \int_0^T \mu_k dt \tag{3.4}$$

$$+ \sum_{i=1}^{M_k} \left\{ \sum_{j=i+1}^{M_k} \chi_{i,j} \log\left(\alpha_k \omega_k e^{-\omega_k(t_j - t_i)}\right) \right\} \tag{3.5}$$

$$- \sum_{i=1}^{M_k} \left\{ \sum_{j=i+1}^{M_k} \int_0^{T-t_i} \alpha_k \omega_k e^{-\omega_k(s)} ds \right\} .$$

Taking the expectation of $\ell_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k)$ with respect to $\chi_{i,j}$ results in

$$E_\chi[\ell_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k)] = \sum_{i=1}^{M_k} P_{i,i}^k \log(\mu_k) - \int_0^T \mu_k dt \tag{3.6}$$

$$+ \sum_{i=1}^{M_k} \left\{ \sum_{j=i+1}^{M_k} P_{i,j}^k \log\left(\alpha_k \omega_k e^{-\omega_k(t_j - t_i)}\right) \right\}$$

$$- \sum_{i=1}^{M_k} \left\{ \sum_{j=i+1}^{M_k} P_{i,j}^k \int_0^{T-t_i} \alpha_k \omega_k e^{-\omega_k(s)} ds \right\} .$$

In the EM algorithm, the quantity $E_\chi[\ell_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k)]$ is maximized with respect to each

of the variables $\mu_k, \alpha_k, \omega_k$ given the data $H_{\tau,k}$. This leads to the EM estimates

$$\mu_k = \frac{\sum_{i=1}^{M_k} P_{i,i}^k}{T} \quad, \quad \alpha_k = \frac{\sum_{i<j}^{M_k} P_{i,j}^k}{M_k - \sum_{i=1}^{M_k} e^{-\omega_k(T-t_i)}} \tag{3.7}$$

$$\omega_k = \frac{\sum_{i<j}^{M_k} P_{i,j}^k}{\sum_{i<j}(t_j - t_i)P_{i,j}^k + \alpha_k \sum_{i=1}^{M_k}(T-t_i)e^{-\omega_k(T-t_i)}}. \tag{3.8}$$

Where $P_{i,j}^k$ is defined by

$$P_{i,j}^k = \frac{\alpha_k \omega_k e^{-\omega_k(t_j-t_i)}}{\lambda_k(t_i|H_{\tau,k})} \quad, \quad P_{i,i}^k = \frac{\mu_k}{\lambda_k(t_i|H_{\tau,k})}, \tag{3.9}$$

for $t_i < t_j$. The EM algorithm then becomes a matter of iterating between estimating the probabilities and the parameters. It has been proven that this algorithm will converge under mild assumptions [DLR77].

In the presence of events with unknown process affiliation in the network, we assign weights to the contribution of each event to the log-likelihood function. Specifically, each of the unknown events in process k have a weight $S_{i,k}$, such that $\sum_k S_{i,k} = 1$. For the known events $S_{i,k} = 1$. These weights are incorporated for each process via

$$\begin{aligned} L_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k) &= \sum_{i=1}^{M_k} P_{i,i}^k S_{i,k} \log(\mu_k) - \int_0^T \mu_k dt \\ &+ \sum_{i=1}^{M_k-1} \sum_{j=i+1}^{M_k} S_{i,k} S_{j,k} P_{i,j}^k \log\left(\alpha_k \omega_k e^{-\omega_k(t_j-t_i)}\right) \\ &- \sum_{i=1}^{M_k} S_{i,k} \int_0^{T-t_i} \alpha_k \omega_k e^{-\omega_k(s)} ds. \end{aligned} \tag{3.10}$$

Note that $L_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k)$ is no longer an EM log likelihood in the presence of unknown data. Maximizing $L_k(H_{\tau,k}|\mu_k, \alpha_k, \omega_k)$ with respect to each of the parameters the estimates become

$$\mu_k = \frac{\sum_{i=1}^{M_k} P_{i,i}^k S_{i,k}}{T} \quad, \quad \alpha_k = \frac{\sum_{i<j}^{M_k} P_{i,j}^k S_{i,k} S_{j,k}}{\sum_{i=1}^{M_k} S_{i,k} - \sum_{i=1}^{M_k} S_{i,k} e^{-\omega_k(T-t_i)}} \tag{3.11}$$

$$\omega_k = \frac{\sum_{i<j}^{M_k} P_{i,j}^k S_{i,k} S_{j,k}}{\sum_{i<j}(t_j - t_i)P_{i,j}^k S_{i,k} S_{j,k} + \alpha_k \sum_{i=1}^{M_k} S_{i,k}(T-t_i)e^{-\omega_k(T-t_i)}}. \tag{3.12}$$

When all of the events are known, i.e. $S_{i,k} = 1$ when unknown event i, k belongs to process k and is zero otherwise, these estimates become identical to the EM parameter estimates.

3.2.3 Updating Weights

At the start of the Estimation & Score algorithm all of the weights for the unknown events are $S_{i,k} = 1/K$. Once the parameters are estimated using the altered EM algorithm described in Equation 3.12, the weights, $S_{i,k}$, are updated, see Figure 3.2. Here we present four different score functions and the Stomakhin-Short-Bertozzi method [SSB11], used to define, $q_{i,k}$, the intermediate process affiliation. Each of these score functions synthesize information from different portions of the data set. Given an event early in the data set, a score function that uses future events would be ideal. On the other hand, for later events a score function using previous events is desired. Similar considerations should be made if there are portion of the data with more incomplete data. After all of these intermediate weights, $q_{i,k}$, have been calculated, they are re-normalized as a probability via $S_{i,k} = \frac{q_{i,k}}{\sum_k q_{i,k}}$. For simplicity we consider a response function of the form, $g_k(t) = \alpha_k \omega_k e^{-\omega_k(t)}$.

3.2.3.1 Ratio Score Function

The *Ratio* score function considers the ratio of the background rate μ_k and the sum of all the future events, $\sum_{i<j} g_k(t_j - t_i)$. Mathematically the score is determined by

$$q_{i,k}^{Ratio} = \frac{\sum_{i<j} g_k(t_j - t_i)}{\mu_k(t_i)}. \tag{3.13}$$

3.2.3.2 Lambda Score Function

The *Lambda* score function uses only previous information by taking the ratio of the intensities evaluated at the unknown event time t_i.

$$q_{i,k}^{Lambda} = \frac{\lambda_k(t_i|H_{\tau,k})}{\sum_{m=1}^{K} \lambda_m(t_i|H_{\tau,k})} \tag{3.14}$$

3.2.3.3 Stomakhin-Short-Bertozzi (SSB) method

The method defined in [SSB11] is summarized by

$$\max\left\{\sum_k \sum_{ij} \delta_{i,j}\mu_k q_{i,k}^{SSB} + \frac{1}{2}(1-\delta_{ij})\alpha_k\omega_k e^{-\omega_k|t_i^k-t_j^k|} q_{i,k}^{SSB} q_{j,k}^{SSB}\right\}, \tag{3.15}$$

subject to

$$\sum_{k=1}^{K}\left(q_{i,k}^{SSB}\right)^2 = 1. \tag{3.16}$$

This method is motivated by the Hawkes process defined in Equation 3.1.

3.2.3.4 Probability Score Function

The *Probability* score function uses the approximation of the branching structure of the underlying process. The idea behind this method is events that are background events with no corresponding response events should not belong in the process. An event that is a background with many response events or an event that is a response to another event should be part of that process.

$$q_{i,k}^{Prob} = \frac{\sum_{t_j>t_i} P_{i,j}^k}{P_{i,i}^k} \tag{3.17}$$

$$P_{i,i}^k = \frac{\mu_k(t_i)}{\lambda_k(t_i|H_{\tau,k})} \qquad P_{i,j}^k = \frac{g_k(t_j-t_i)}{\lambda_k(t_j|H_{\tau,k})} \tag{3.18}$$

3.2.3.5 Forward Backward Response Score Function

This method is the ratio of the summation of the response for the events in the future and the past, $\sum_{i\neq j} g_k(|t_i-t_j|)$ over the background rate μ_k.

$$q_{i,k}^{FB} = \frac{\sum_{i\neq j} g_k(|t_i-t_j|)}{\mu_k} \tag{3.19}$$

3.3 Results

The Estimation & Score algorithm is tested for accuracy on simulated data from the Hawkes process defined in Equation 3.1. An analysis of the parameter estimation method outlined in Subsection 3.2.2 is conducted in Subsection 3.3.1. A comparison of the score functions when assuming the true parameters is found in Subsection 3.3.2. Subsection 3.3.3 provides a comparison of the runtime between the Forward Backward score function and the Stomakhin-Short-Bertozzi method. A example of convergence of the Estimate & Score Algorithm is provided in Subsection 3.3.4.

3.3.1 Estimation Analysis

There are many ways we could allow the unknown events to influence our estimates of the underlying parameters for each process. There are two extremes. On the one hand, we could exclude all of the unknown events from the parameter estimation. This would be equivalent to setting the $S_{i,k} = 0$ for all unknown events. On the other hand, we could include all of the unknown events in the estimation of the parameters for each process. This would be equivalent to letting $S_{i,k} = 1$ for all i and k. Another possible estimation method is some combination of these two. We propose this as a way of allowing the unaffiliated events to play some role in the estimation process. The naive choice is allowing each event to play the same role in each process. This amounts to setting $S_{i,k} = 1/K$ for the unknown events. We compare these three choices to the estimations obtained by the Estimate & Score Algorithm (ESA) using the Forward Backward score function. Finally, we want to compare all four of these possible estimation techniques to the best we could possibly do. In this case, that would mean we knew all the affiliations for the events (i.e. there are no unknown events).

Figures 3.3-3.5 displays the results for the μ_k, α_k, and ω_k estimates for the five cases: $S_{i,k} = 0$ for unknown events (dash-triangle), $S_{i,k} = 1$ for unknown events (dash-square), $S_{i,k} = 1/K$ for unknown events (dash-x), the results using ESA (dash-circle), and the estimates you get when you know all the affiliations for the unknown events (solid). These

results with standard deviations are displayed in Tables 3.1-3.3. In each of the three figures, the estimates are plotted vs the number of missing events. Each network has five processes with the true parameters $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$. Different networks are created with 15, 30, 45, 60, and 75 unknown events. Then we estimate the parameters using each of the five methods explained above. This procedure is repeated 100 times with different random seed values and then the average estimate is calculated.

Notice in the estimates for μ_k in Figure 3.3 and Table 3.1, the ESA performs the best compared to the true value and has only a slight reduction in accuracy as the number of missing events increases. On average the other three estimates seem to degrade more rapidly as the number of missing events increases. When $S_{i,k} = 1$, the estimates for μ_k are far above the true value and growing as the number of unknown events increases. This follows from the fact that letting $S_{i,k} = 1$ means we are effectively adding events to the network. Take the case when $K = 2$. Assume that each process has 1000 events, and there are 100 missing from each process. When we do our estimation for the first process, we will use the 900 events we know plus the 200 unknown events from the network. We will get the identical number of events in our estimation for process two. This creates 200 new events and thus biases the estimates for μ_k. This motivated the idea of equal weighting for each unknown event, and that choice is validated by the estimates for μ_k. A similar argument shows why $S_{i,k} = 0$ (i.e. ignoring all the unknown events) has the lowest estimate for μ_k at each level of missing data.

In the estimates for the branching ratio α_k, the ESA on average yields the best estimates and maintains its accuracy in the presence of more incomplete information. It is interesting to note that equal weighting performs worse here than if we let $S_{i,k} = 0$ for all unknown events. Using the ESA overcomes this drawback. Again, setting $S_{i,k} = 1$ for all unknown events performs the worst. This could stem from the fact that most of the unknown events are being labeled background and thus this estimation technique underestimates the branching ratio because fewer events are considered offspring. Notice that the estimate for ESA tracks the best possible estimate (dash-circle) well while the other three start to trail off as more

Figure 3.3: Plots of the estimates for μ_k for the Unknown Not Included (dash-triangle), Unknown Included (dash-square), Equal Weights (dash-x), ESA (dash-circle), and No Unknown (solid). In each of the three figures, the estimates are plotted vs the number of missing events. Each network has five processes with the true parameters $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$. Each data point presented is the average of the results from 100 simulated networks.

and more information is labeled as missing.

Finally, in Figure 3.5 and Table 3.1, it is shown that the ESA estimate (dash-circle) for ω_k tracks the behavior of the best estimate (solid) closer than the other methods. Including all of the unknown events (dash-square) provides the poorest estimate for ω_k. For the other three estimation techniques we see that they are all comparable.

68
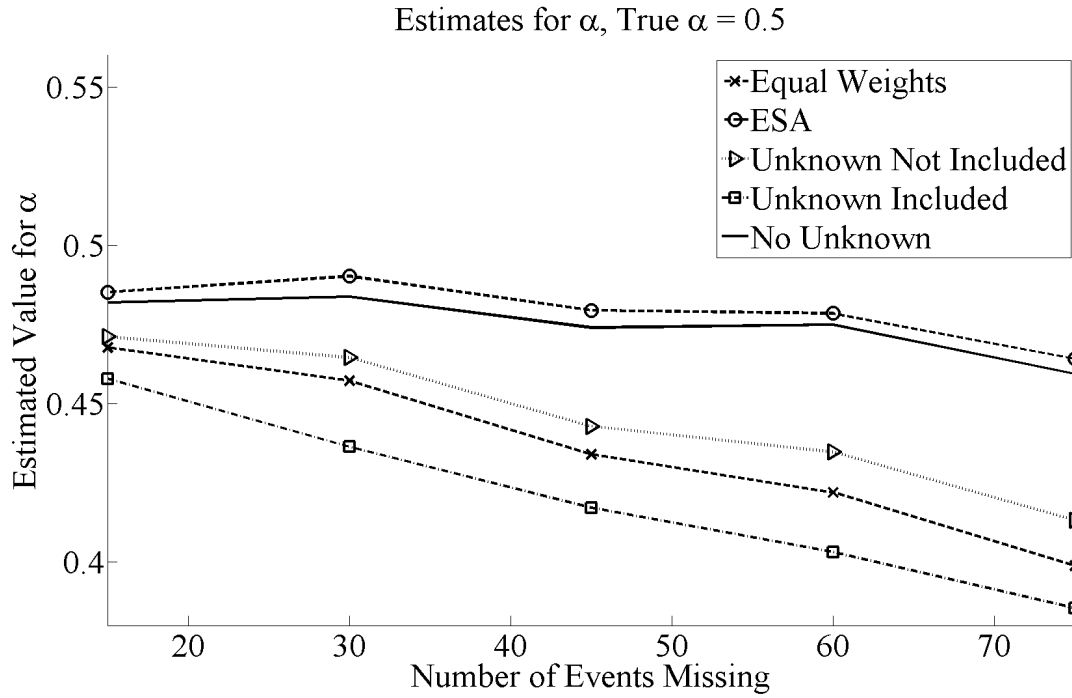
Figure 3.4: Plots of the estimates for α_k for the Unknown Not Included (dash-triangle), Unknown Included (dash-square), Equal Weights (dash-x), ESA (dash-circle), and No Unknown (solid). In each of the three figures, the estimates are plotted vs the number of missing events. Each network has five processes with the true parameters $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$. Each data point presented is the average of the results from 100 simulated networks.
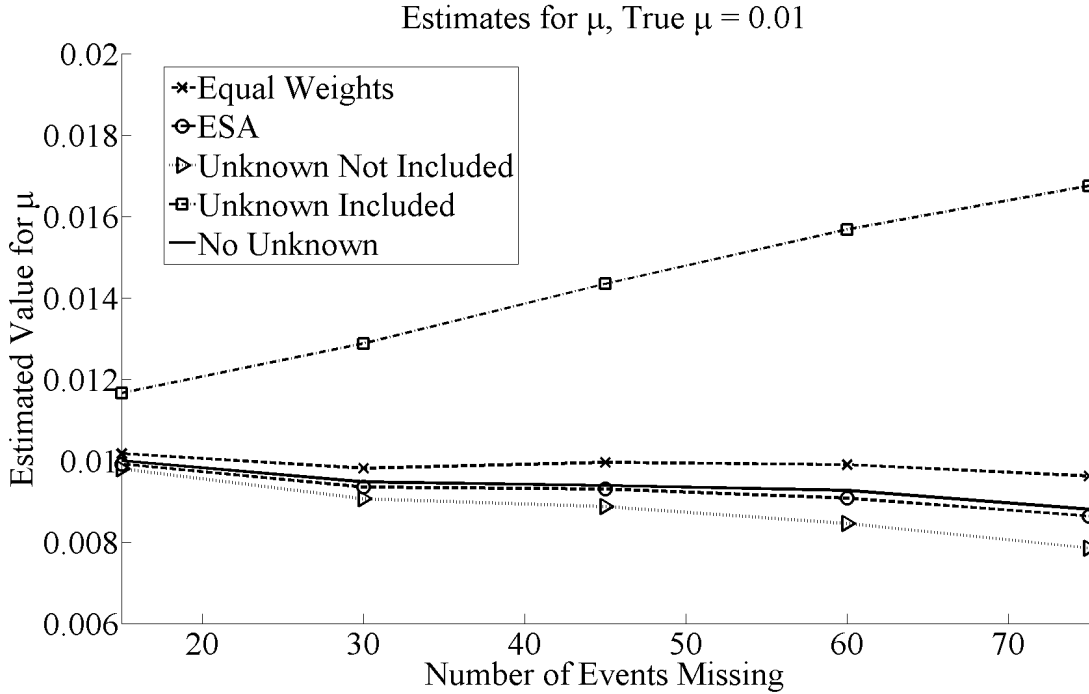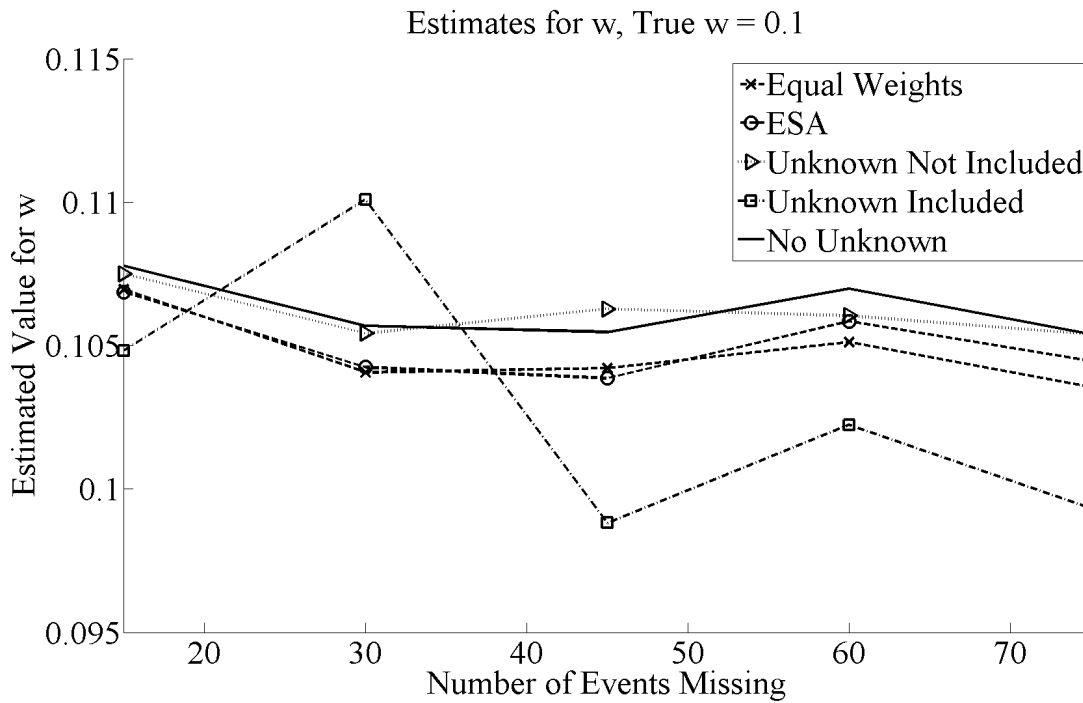
Figure 3.5: Plots of the estimates for ω_k for the Unknown Not Included (dash-triangle), Unknown Included (dash-square), Equal Weights (dash-x), ESA (dash-circle), and No Unknown (solid). In each of the three figures, the estimates are plotted vs the number of missing events. Each network has five processes with the true parameters $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$. Each data point presented is the average of the results from 100 simulated networks.

Table 3.1: Average and standard deviations for μ_k on 100 networks, true value is $\mu_k = 0.01$.

|		15	30	45	60	75
Equal	(Ave)	0.0102	0.0098	0.0100	0.0099	0.0096
Weights	(StDev)	±0.0014	±0.0014	±0.0017	±0.0015	±0.0015
ESA	(Ave)	0.0099	0.0093	0.0093	0.0091	0.0086
	(StDev)	±0.0014	±0.0014	±0.0017	±0.0014	±0.0014
Unknown	(Ave)	0.0098	0.0091	0.0089	0.0085	0.0079
Not Included	(StDev)	±0.0014	±0.0014	±0.0017	±0.0015	±0.0015
Unknown	(Ave)	0.0117	0.0129	0.0143	0.0157	0.0167
Included	(StDev)	±0.0014	±0.0017	±0.0019	±0.0016	±0.0019
No Unknown	(Ave)	0.0100	0.0095	0.0094	0.0093	0.0088
	(StDev)	±0.0014	±0.0014	±0.0017	±0.0015	±0.0015

Table 3.2: Average and standard deviations for α_k on 100 networks, true value is $\alpha_k = 0.5$.

		15	30	45	60	75
Equal	(Ave)	0.4678	0.4573	0.4340	0.4220	0.3989
Weights	(StDev)	±0.0636	±0.0759	±0.0686	±0.0726	±0.0699
ESA	(Ave)	0.4853	0.4903	0.4795	0.4786	0.4642
	(StDev)	±0.0640	±0.0767	±0.0712	±0.0741	±0.0719
Unknown	(Ave)	0.4712	0.4646	0.4429	0.4348	0.4132
Not Included	(StDev)	±0.0638	±0.0779	±0.0700	±0.0737	±0.0702
Unknown	(Ave)	0.4580	0.4364	0.4172	0.4032	0.3855
Included	(StDev)	±0.0668	±0.0822	±0.0705	±0.0799	±0.0818
No Unknown	(Ave)	0.4820	0.4838	0.4741	0.4750	0.4595
	(StDev)	±0.0647	±0.0759	±0.0726	±0.0748	±0.0689

Table 3.3: Average and standard deviations for ω_k on 100 networks, True Estimate is $\omega_k = 0.1$.

Number Missing		15	30	45	60	75
Equal	(Ave)	0.1070	0.1041	0.1042	0.1051	0.10364
Weights	(StDev)	±0.0264	±0.0274	±0.0262	±0.0248	±0.0255
ESA	(Ave)	0.1069	0.1042	0.1039	0.1059	0.1045
	(StDev)	±0.0263	±0.0273	±0.0264	±0.0255	±0.0240
Unknown	(Ave)	0.1075	0.1054	0.1063	0.1060	0.1054
Not Included	(StDev)	±0.0264	±0.0286	±0.0269	±0.0246	±0.0269
Unknown	(Ave)	0.1048	0.1101	0.0988	0.1022	0.0993
Included	(StDev)	±0.0275	±0.1035	±0.0273	±0.0301	±0.0285
No Unknown	(Ave)	0.1078	0.1057	0.1055	0.1070	0.1054
	(StDev)	±0.0265	±0.0277	±0.0256	±0.0241	±0.0230

3.3.2 Updating Weights Analysis

To understand the strengths and weaknesses of each of the five score functions, defined in Subsection 3.2.3, the score functions were evaluated for 100 missing events using the true values for $\mu_k, \alpha_k,$ and ω_k when taking the Top 1, Top 2, and Top 3 best inferences. For comparison to [SSB11], the true parameters were taken to be $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$. Due to the stochastic nature of the processes, for each level of process number 100 random networks were tested. The average results of this analysis are found in Figures 3.6, 3.7, and 3.8. The number correctly identified by the each of the score functions is on the vertical axis. The horizontal axis displays the number of processes in the network.

From Figures 3.6, 3.7, and 3.8 it is clear that the Stomakhin-Short-Bertozzi score function in solid dark blue, and the Forward Backward score function (cyan dashed diamond) perform nearly identically when looking at the Top 1, Top 2, and Top 3 inferences. These functions look both forward and backward in time from the missing event, and are therefore able to identify clusters of events in time. The Probability (black dashed asterix) and Ratio (solid

green square) score functions do not preform nearly as well the Stomakhin-Short-Bertozzi and Forward Backward score functions, but better than the Lambda (magenta dashed circle) score function. The Lambda score function appears to perform close to chance (dark green solid plus) for the Top 1, Top 2, and Top 3 inferred process affiliation. Due to the success of the Forward Backward score function and the Stomakhin-Short-Bertozzi score functions only these functions are used for analysis.

Figure 3.6: Display of the number of correctly identified missing events when the Top 1 inference are taken into consideration. For all score functions, the parameters are $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$, and assumed to be known. The Stomakhin-Short-Bertozzi score function (solid dark blue x) and the Forward Backward score (cyan dashed diamond), the Probability (black dashed asterix) and Ratio (solid green square) score functions, and the Lambda (magenta dashed circle) score function and chance (solid dark green plus) produce comparable results with these parameters.
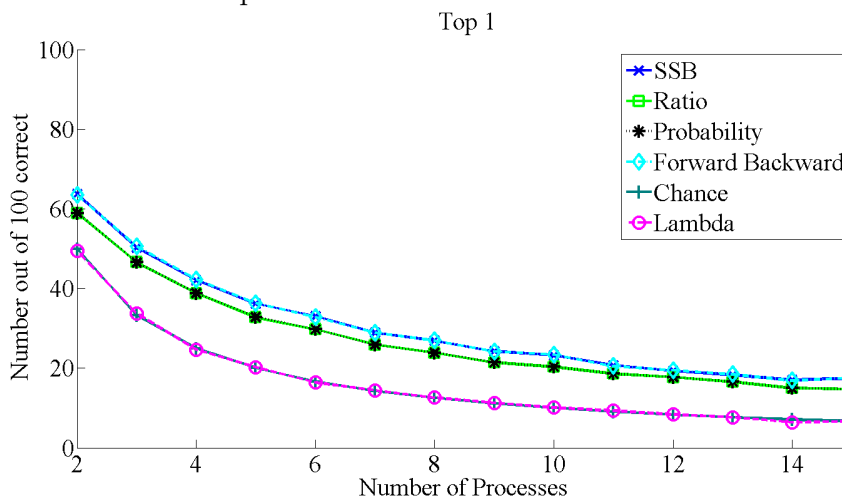
The analysis comparing the score functions assumed that the true parameters were known. However, when applying this method in practice there will be error in the estimated parameters. This estimation error will propagate through to the score functions. To understand how deviations of the estimated parameters influence the score functions pairwise combinations of the parameters were increased and decreased by 90% from the target values $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$ in 10% increments. In particular the Forward

Figure 3.7: Display of the number of correctly identified missing events when the Top 2 inferences are taken into consideration. For all score functions, the parameters are $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$, and assumed to be known. The Stomakhin-Short-Bertozzi score function (solid dark blue x) and the Forward Backward score (cyan dashed diamond), the Probability (black dashed asterix) and Ratio (solid green square) score functions, and the Lambda (magenta dashed circle) score function and chance (solid dark green plus) produce comparable results with these parameters.
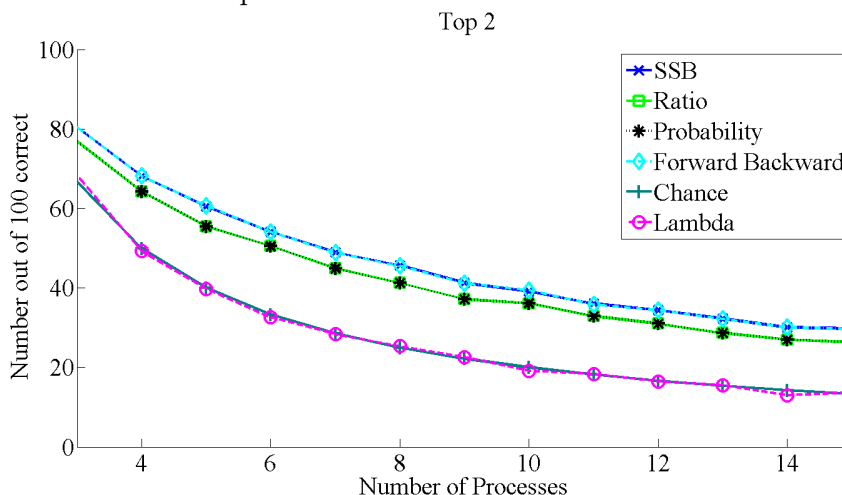
Backward and SSB score functions are computed for pairwise combinations of μ in the range of $[0.001, 0.019]$, ω in the range of $[0.01, 0.19]$, and α in the range of $[0.05, 0.95]$. Further, in these pairwise combinations, the third parameter is kept at the target value. Notice that a 90% change is larger than the errors observed in the parameter estimates in Subsection 3.3.1.

To examine the propagation of errors of the parameters to the score functions one event from a network with 10 processes is chosen to be missing. The score function $S_{1,true}$ with the target parameters, $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$ for the true process is calculated. Then, on the same network, the parameters are offset by

$$\widehat{\text{parameter}} = \text{parameter} \pm \%\text{change} \cdot \text{parameter}, \tag{3.20}$$

and the offset score function $\hat{S}_{1,true}$ is calculated. The difference between $S_{1,true} - \hat{S}_{1,true}$ is taken for each pairwise combination of parameters. Again, due to the stochastic nature of the

Figure 3.8: Display of the number of correctly identified missing events when the Top 3 inferences are taken into consideration. For all score functions, the parameters are $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$, and assumed to be known. The Stomakhin-Short-Bertozzi score function (solid dark blue x) and the Forward Backward score (cyan dashed diamond), the Probability (black dashed asterix) and Ratio (solid green square) score functions, and the Lambda (magenta dashed circle) score function and chance (solid dark green plus) produce comparable results with these parameters.
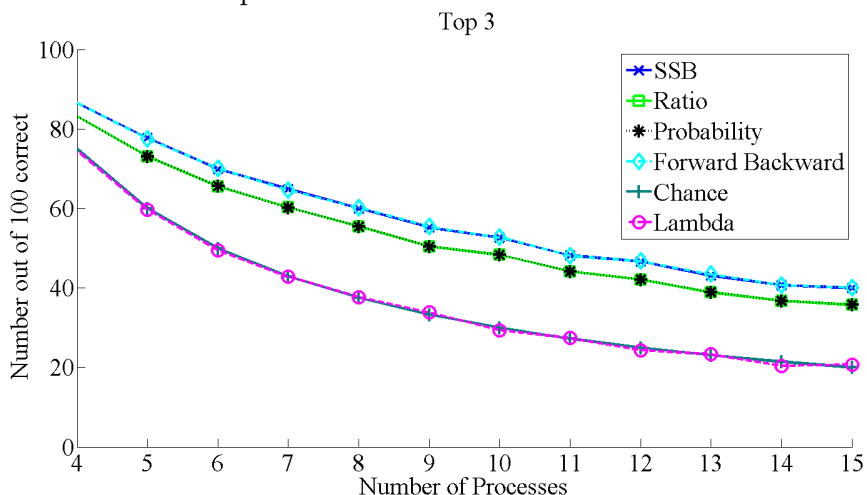

processes, each analysis was done for 100 runs and the average difference in score functions is recorded. The results of this analysis are displayed in Figure 3.9 with those of the Forward Backward score function (left), and those for the Stomakhin-Short-Bertozzi score function (right). In general the Stomakhin-Short-Bertozzi score function is more sensitive to the changes than the Forward Backward score functions for the μ_k and α_k parameters. Changes in the Forward Backward score functions are minimal for most changes of parameters except for small values of ω_k. As ω_k decreases then the approximated Forward Backward score function decreases, causing a positive difference. As seen in Subsection 3.3.1, Figure 3.5, when estimating ω_k, there is a tendency to over, not under estimate the parameter, and so this does not appear to occur within these parameters. The changes in the Stomakhin-Short-Bertozzi score function depend on all of the pairwise changes of the parameters. As μ_k increases the computed Stomakhin-Short-Bertozzi decreases. On the other hand, as ω_k

or α_k increase the score function increases. This analysis shows that though the Stomakhin-Short-Bertozzi method and the Forward Backward score functions preform similarly when the parameters are known exactly, under the influence of estimation error the Stomakhin-Short-Bertozzi score function varies more than the Forward Backward score function.

3.3.3 Runtime Analysis

Though the Forward Backward score function and the SSB method produce comparable results in terms of accuracy, there are fundamental differences in the way these two methods are computed. The Forward Backward score function is designed to be direct, meaning the weights are calculated using available information without need for iteration. The Stomakhin-Short-Bertozzi method, however, determines the weight by solving a optimization problem. A closed form solution for the maximized weights is not known to these authors, so the weights are found by numerically approximating the weights that maximize Equation 3.15. In the implementation of the Stomakhin-Short-Bertozzi we employ a gradient ascent method which requires 4-11 iterations to reach convergence with a tolerance of 0.001.

The direct methods, Forward Backward, Probability, Ratio, and Lambda score functions, are on the same order of operations as one iteration of the gradient ascent used to solve Equation 3.15. Specifically, one iteration of the gradient ascent method and calculating the direct score functions are $O(N \cdot K \cdot M)$ where N is the number of missing events, K is the number of processes and M is the expected number of events in process k. One of the strengths of computing the weights using a direct method is that it reduces the computational cost.

For each stage of the algorithm, the computational cost is estimated. The algorithm requires that we estimated μ_k, α_k, and ω_k for each process, k. The EM algorithm, described in Section 3.3.1, these estimates are iterated until convergence. Using the EM algorithm one uses the intermediate estimated values of $P_{i,j}^k$ and $\lambda_k(t|H_{\tau,k})$. Each of these estimates depends on the number of points in the process M_k. The process estimated is stochastic

76

Figure 3.9: The average difference of the Forward Backward and Stomakhin-Short-Bertozzi score functions with the parameters varied by $\pm 90\%$ of the target values, $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$.
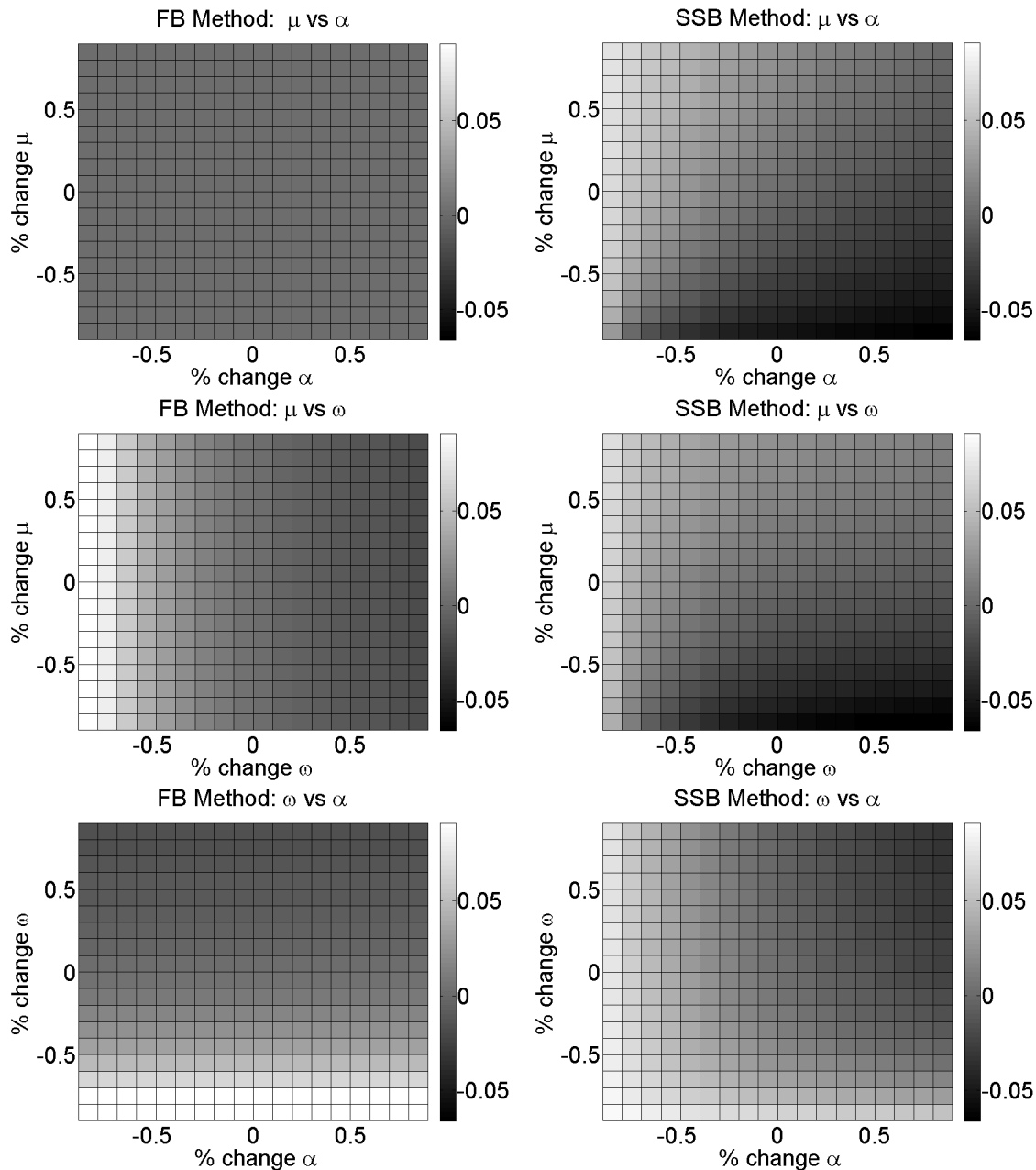
resulting in a varying number of events in each process, M_k. The number of events in a process depends on the number of missing events (N), the final time (T), the background rate (μ_k), and excitation parameter (α_k). An upper bound for the number of events in any

process is

$$M = E[M_k] = \mu \cdot T \cdot \frac{1}{1-\alpha} + \frac{K-1}{K}N. \tag{3.21}$$

In Equation 3.21 $\mu \cdot T$ is the estimated number of background events. Each event is expected to have α offspring, leading to the geometric sum $\sum_{k=0}^{\infty} \left(\frac{1}{\alpha}\right)^k$. The term $\frac{K-1}{K}N$ is the expected number of missing events that will be added to the process. For reference, the computations for each object in the algorithm are summarized in Table 3.4.

3.3.3.1 Computing Lambda and Probability Matrix

The computational cost in computing $\lambda_k(t|H_{\tau,k})$, as defined in Equation 3.1, depends on the number of points before time, t, notated η. In the naive implementation computing $\lambda_k(t|H_{\tau,k})$ takes $O(\eta)$ for one t. This is quite costly, when one considers that lambda must be computed often for all events in the process k. For this reason the implementation of this algorithm calculates $\lambda_k(t|H_{\tau,k})$ at each event of process k, and then stores it for further computations. With this naive implementation, calculating $\lambda_k(t|H_{\tau,k})$ at all of the events of process k has an expected computational cost on the order of $O(M^2)$. In the special case of an exponential response function, the computations can be greatly reduced by using the recursive formula

$$
\begin{aligned}
G(t_{m+1}) &= e^{\omega_k(t_{m+1}-t_m)}\left[G(t_m)+1\right] \\
G(t_1) &= 0.
\end{aligned}
\tag{3.22}
$$

The intensity $\lambda_k(t_m|H_{\tau,k})$ can be written as

$$\lambda_k(t_m|H_{\tau,k}) = \mu_k + \alpha_k\omega_k G(t_m). \tag{3.23}$$

Using the sophisticated method of computation, λ can be computed for each event with approximately $O(M)$ computations, a great improvement. Note that, if a general response function were given to describe the process, then this recursive formula would not apply. Once $\lambda_k(t_m|H_{\tau,k})$ is computed at all events, computing the probability matrix P^k takes $O(M^2)$ operations.

3.3.3.2 Estimation Parameters

To compute the estimated order of operations it is assumed that the weights $S_{i,k}$ and the intensity $\lambda_k(t_m|H_{\tau,k})$ have been computed and stored for all other events. Directly from the equations the major influencing factor in the runtime of the estimation is the number of points in the process. Specifically, μ_k takes $O(M)$, α_k takes $O(M^2)$, and ω_k takes $O(M^2)$ for one iteration of the algorithm. The number of iterations needed depends on the initial guess of the parameters and the true process parameter values.

3.3.3.3 Score Functions

Estimating the operation count for the score functions is not as straight forward as the other parameters of the algorithm. The score functions are evaluated only at the missing events for each process. However, the computational cost of evaluating the scores for some of the functions depends on the number of events in the process k preceding and following the event i denoted η. This is random and would be impossible to determine in general. In this computational analysis the worst case scenario is taken and it is assumed that $\eta = M$. Further, with an intelligent implementation of the Lambda Score function, the order of operations can be reduced from $O(N^2 \cdot K)$ to $O(N \cdot K)$ by calculating lambda using the relationship

$$\lambda_k(t_i|H_{\tau,k}) = \frac{\mu_k(t_i)}{P_{i,i}^k}. \tag{3.24}$$

The Stomakhin method described in [SSB11] relies on a gradient ascent algorithm to determine the weights. To compare with the score functions proposed in this dissertation, the operation count for one iteration of the gradient ascent method. It it is expected that, in practice, many iterations would be needed to reach convergence. From this analysis it is clear that the computational cost of the direct score functions is comparable to that of only one iteration of the Stomakhin method.

    The run time of both the Forward Backward function and the Stomakhin-Short-Bertozzi method are empirically examined in Figure 3.10. Both score functions were calculated with

Table 3.4: Order of operations count for estimating parameters and score functions. All estimates are a function of the number of events in process k, M, the expected number of missing events (see Equation 3.21), N, and the number of processes, K. The operation counts for the parameter estimates assume that $P_{i,j}$ and $S_{i,k}$ are calculated, and $\lambda_k(t|H_{\tau,k})$ has been evaluated and stored for each event in process k.

One EM Iteration

μ_k	O(M)
α_k	O(M^2)
ω_k	O(M^2)

Calculating and Storing

$\lambda_k(t|H_{\tau,k})$ (exponential time)	O(M)
$\lambda_k(t|H_{\tau,k})$ (general)	O(M^2)
P^k	O(M^2)

Score Functions

Ratio Score Function	O($N \cdot K \cdot M$)
Probability Score Function	O($N \cdot K \cdot M$)
Lambda Score Function (Intellegent)	O($N \cdot K$)
Lambda Score Function (Direct)	O($N \cdot K \cdot M$)
Forward Backard Score Function	O($N \cdot K \cdot M$)
Stomakhin Method One Iteration	O($N \cdot K \cdot M$)

20 networks for each level of number missing and number of processes with the known parameter values of $\mu_k = 0.01$, $\omega_k = 0.1$, and $\alpha_k = 0.5$. All of the run times are calculated in milliseconds. It can be seen that the average run time needed to compute the Forward Backward function at every level of N and K is substantially less than that of the Stomakhin-Short-Bertozzi method. Also, it is clear from this figure that the time needed to calculate both of these methods increases as N and K increase.

Figure 3.10: Display of the run time of the Forward Backward function and the Stomakhin-Short-Bertozzi method.
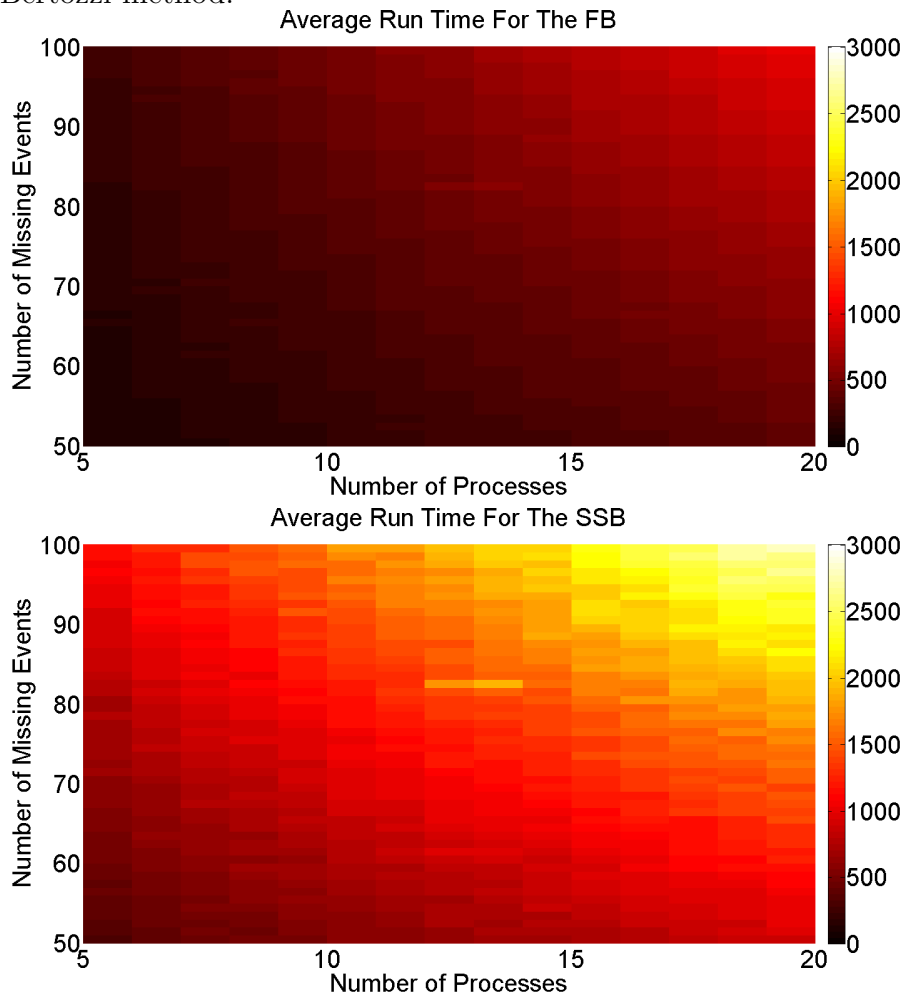
3.3.4 Convergence Results

The Estimation & Score Algorithm converges quickly when either the Forward Backward score function or Stomakhin-Short-Bertozzi method are used. Figure 3.12 displays the parameter estimates for a typical run of the Estimation & Score Algorithm for both the Forward Backward (left) and Stomakhin-Short-Bertozzi (right). Both score functions produce qualitatively similar results, and it appears that the rate of convergence is comparable for both cases. The estimated weights for one missing data event for this typical run versus the iteration for each process are plotted in Figure 3.12. Here the Forward Backward weight

approximations are seen on the left and the Stomakhin-Short-Bertozzi approximations are on the right. It is interesting to note that both methods of weighting choose the same process affiliation. Further tests were conducted with a variable initial weighting. These runs showed similar behavior as initializing the Estimate & Score Algorithm with $S_{i,k} = 1/K$, implying that the Estimate & Score Algorithm is robust to small perturbations of the initial weighting.

Figure 3.11: Parameter estimates for the Forward Backward (left) and Stomakhin-Short--Bertozzi (right) methods vs the iteration number. Plots of the parameter estimates for a typical run of the Estimate & Score Algorithm using the Forward Backward (left) and Stomakhin-Short-Bertozzi (right) methods. Both methods compute nearly identical estimates of the parameters for each of the ten processes. The choice for plotting event 99 was random.
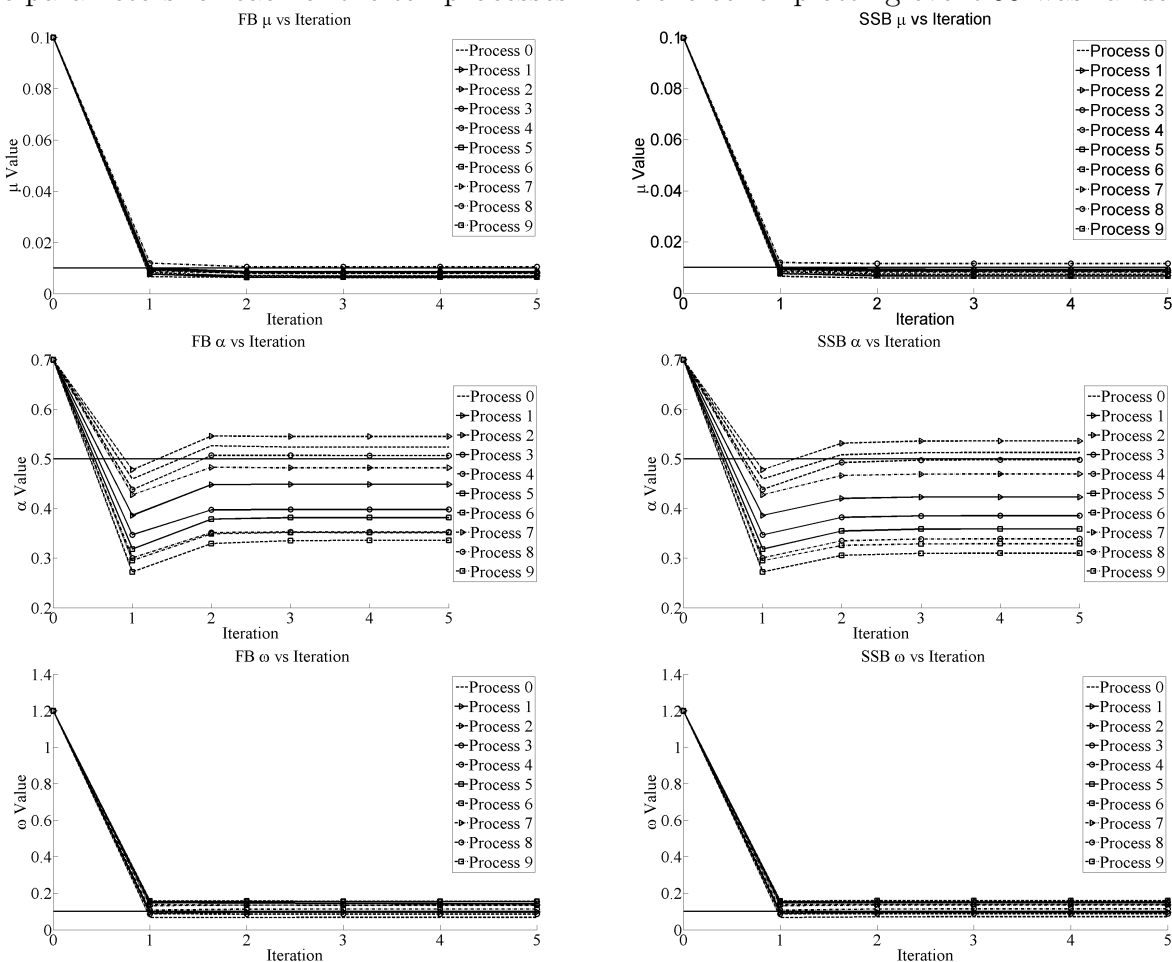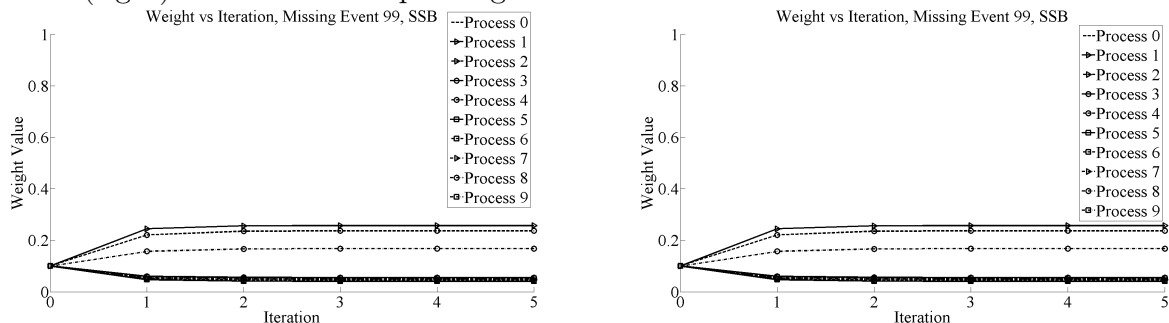
Figure 3.12:   Plots of the weights for one missing event computed by a typical run of the Estimate & Score Algorithm using the Forward Backward (left) and Stomakhin-Short-Bertozzi method (right). The choice for plotting event 99 was random.

3.4   Discussion and Future Work

In this dissertation we propose an effective method for simultaneously estimating the parameters and assigning process affiliation in case of incomplete field data from self-exciting point processes on a network. This problem comes from the demand for law enforcement agencies to identify gang affiliation in the case of unsolved crimes in an area of highly complex gang rivalry activity. We present a new 'Estimate & Score' algorithm for possible application to field data. By testing the method on simulated datasets we can understand its performance features and liabilities. The method is an iterative procedure in which process parameters are estimated alternately with the calculation of network affiliation probabilities. We identify several useful 'score functions' for calculating the network affiliations. We also compare the use of unknown events in the parameter estimation. One upshot of our analysis is that the inclusion of unknown events may increase the accuracy of the parameter estimation. Several score functions are considered and the Forward Backward score function shows the most promise with comparable results to that of the Stomakhin-Short-Bertozzi method of [SSB11] in the parameter regime tested. The score function calculation is a direct method that does not rely on solving a variational problem, and thus is more computationally efficient than [SSB11].

For future work, space often plays a role in understanding criminal activity [BGK05, BB81, Her82, ECL05, BB95]. Further, criminal behavior has non-random structure and can often be framed in terms of routine activity theory [BB93b, CF79]. In the case of gang violence, there is a strong spatial component [Blo00, RFT10, HSB11]. One can easily extend the Estimate & Score Algorithm to include space. There is a precedence in the earthquake literature of adding space to self-exciting point processes [Oga98, ZOV02], however, in the case of gang violence, the spatial response may be different. Instead of retaliatory events clustering around prior events, it appears that the data is clustered around regions in space. A spatial model similar to that of [OL11] could be employed, where the triggering density in space is related to their respective gang set-space, or center of activity [TCE05]. Statistically when modeling spatial point processes one needs to tease out the difference between a hot spots due to risk heterogeneity versus event dependence. The data given will be one realization of the underlying process, however using techniques such as prototyping, [NSK11], one could potentially reformulate the data into multiple realization of the same process and distinguish between these two phenomena.

There are other factors in the data that can be fused into the model. For example, in earthquake modeling the magnitude of the earthquake is often included. Such a factor could be further added to the intensity $\lambda_k(t|H_{\tau,k})$ to better infer the gang affiliation. For example tagging, or other low level gang crimes, could be a precursor to more extreme violent interactions between gangs. Including this data could enrich the overall data set allowing for better analysis.

It is important to note that there are other methods to approximate the underlying form of the self exciting process. For example the authors in [LM11] consider the general form of the intensity function $\lambda_k(t|H_{\tau,k})$ to be

$$\lambda(t|H_\tau) = \mu(t) + \alpha \sum_{t > t_j} g(t - t_j). \tag{3.25}$$

Using a non-parametric method, they are able to approximated the background function $\mu(t)$ and the response function $g(t)$ for a broader class of functions. In this dissertation,

84

the data was assumed to come from a Hawkes process with constant background rate and an exponential response to previous events. There are cases where the background rate is not constant [LMB11]. Further it is conceivable that the response function could be of a form other than an exponential decay. In this circumstances, the model for $\lambda(t|H_\tau)$ in Equation 3.1 would not be appropriate.

Finally, this method has a great potential in the field of policing. Once such a model has been calibrated correctly, the Estimation & Score Algorithm using the quicker Forward Backward score function can be used to infer the gang association in real time, while the investigation is on going. Given an accurate model of the underlying process, such a method could identify rivalries that have heightened activity.

References

[AB02]   Réka Albert and Albert-László Barabási. "Statistical mechanics of complex networks." *Rev. Mod. Phys.*, **74**(1):47–97, January 2002.

[ACL10]   Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger J.A. Laeven. "Modeling financial contagion using mutually exciting jump processes." Technical report, National Bureau of Economic Research, 2010.

[Aka73]   Hirotsugu Akaike. "Information theory and an extension of the maximum likelihood principle.", 1973.

[Ald97]   John Aldrich. "R. A. Fisher and the Making of Maximum Likelihood 1912-1922." *Statistical Science*, **12**(3):162–176, 1997.

[And00]   Elijah Anderson. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City.* 2000.

[BB81]   Paul J. Brantingham and Patricia L. Brantingham. *Environmental Criminology.* Sage Publications, Inc, 1981.

[BB93a]   Carolyn Rebecca Block and Richard Block. "Street Gang Crime in Chicago." Technical report, December 1993.

[BB93b]   Patricia L. Brantingham and Paul J. Brantingham. "Environment, routine and situation: Toward a pattern theory of crime." *Advances in Criminological Theory*, **5**:259–294, 1993.

[BB93c]   Patricia L. Brantingham and Paul J. Brantingham. "Nodes, paths and edges: Considerations on the complexity of crime and the physical environment." *Journal of Environmental Psychology*, **13**(1):3 – 28, 1993.

[BB95]   Patricia L. Brantingham and Paul J. Brantingham. "Criminality of place." *European Journal on Criminal Policy and Research*, **3**:5–26, 1995. 10.1007/BF02242925.

[BB04]   Patricia L. Brantingham and Paul J. Brantingham. "Computer simulation as a tool for environmental criminologists." *Security Journal*, **17**(1):21–30, 2004.

[BBC00]   Pierre Baldi, Soren Brunak, Yves Chauvin, Claus A. Andersen, and Henrik Nielsen. "Assessing the accuracy of prediction algorithms for classification: an overview." *Bioinform. Rev.*, **16**(5):412–424, May 2000.

[BEB09]   Alethea Barbaro, Baldvin Einarsson, Bjorn Birnir, Sven Sigurdsson, Heoinn Valdimarsson, Olafur K. Palsson, Sveinn Sveinbjornsson, and Torsteinn Sigurdsson. "Modelling and simulations of the migration of pelagic fish." *ICES J. Mar. Sci.*, **66**(5):826–838, 2009.

[BF79]   D. J. Best and Nicholas I. Fisher. "Efficient Simulation of the von Mises Distribution." *J. R. Stat. Soc. Ser. C (Appl. Stat.)*, **28**(2):152–157, 1979.

[BGJ09]   Patricia L. Brantingham, Uwe Glasser, Piper Jackson, and Mona Vajihollahi. "Modeling Criminal Activity in Urban Landscapes." In Nasrullah Memon, Jonathan David Farley, David L. Hicks, and Torben Rosenorn, editors, *Mathematical Methods in Counterterrorism*, pp. 9–31. Springer Vienna, 2009. 10.1007/978-3-211-09442-6_2.

[BGK05]   Patricia L. Brantingham, Uwe Glasser, Bryan Kinney, Komal Singh, and Mona Vajihollahi. "A computational model for simulating spatial aspects of crime in urban environments." In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 4, pp. 3667 – 3674 Vol. 4, oct. 2005.

[BHG06]   Dirk Brockmann, Lars Hufnagel, and Theo Geisel. "The Scaling Laws of Human Travel." *Nature*, **439**:462–465, 2006.

[BHP07]   Milan Bradonjić, Aric Hagberg, and Allon G. Percus. "Giant component and connectivity in geographical threshold graphs." In *Algorithms and Models for the Web-Graph (WAW 2007)*, pp. 209–216, San Diego, CA USA, 2007.

[BHP09]   Milan Bradonjić, Aric Hagberg, and Allon G. Percus. "The structure of geographical threshold graphs." *Internet Math.*, **5**:113–140, 2009.

[Blo00]   Richard Block. "Gang Activity and Overall Levels of Crime: A New Mapping Tool for Defining Areas of Gang Activity Using Police Records." *Journal of Quantitative Criminology*, **16**:369–383, 2000. 10.1023/A:1007579007011.

[BT08]   P. Jeffrey Brantingham and George Tita. "Offender Mobility and Crime Pattern Formation from First Principles." In L. Liu and J. Eck, editors, *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*, pp. 193–208. Idea Press, 2008.

[BUK12]   James H. von Brecht, David Uminsky, Theodore Kolokolnikov, and Andrea L. Bertozzi. " Predicting Pattern Formation in Particle Interactions ." *Mathematical Models and Methods in Applied Sciences*, **22**, 2012.

[Cam90]   L. Le Cam. "Maximum Likelihood: An Introduction." *International Statistical Review / Revue Internationale de Statistique*, **58**(2):pp. 153–171, 1990.

[CCR10]   Jose A. Canizo, Jose A. Carrillo, and Jesus Rosado. "A well-posedness theory in measures for some kinetic models of collective motion." 2010.

[CF79]   Lawrence E. Cohen and Marcus Felson. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review*, **44**(4):pp. 588–608, 1979.

[CF04]   Ronald V. Clarke and Marcus Felson. *Routine Activity and Rational Choice*, volume 5. Transaction Publishers, 2004.

[Chu11] Dominique Chu. "Complexity: against systems." *Theory in Biosciences*, **130**:229–245, 2011. 10.1007/s12064-011-0121-4.

[Cov10] Herbert C. Covey. *Street Gangs Throughout the World.* Charles C Thomas, Publisher, 2010.

[CS08]   Riley Crane and Didier Sornette. "Robust dynamic classes revealed by measuring the response function of a social system." *Proceedings of the National Academy of Sciences*, **105**(41):15649–15653, 2008.

[DCB06] Maria R. D'Orsogna, Yao-Li Chuang, Andrea L. Bertozzi, and Lincoln Chayes. "Self-propelled particles with soft-core interactions: patterns, stability, and collapse." *Phys. Rev. Lett.*, **96**, 2006. 104302.

[Dec96] Scott Decker. "Collective and normative features of gang violence*." *Justice Quarterly*, **13**(2):243–264, 1996.

[Dig83] Peter J. Diggle. *Statistical Analysis of Spatial Point Patterns.* Academic Press, 1983.

[DLR77] Arthur P. Dempster, Nan Laird, and Donald Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *J. Roy. Statist. Soc. Ser. B*, **39**(1):1–38, 1977. With discussion.

[DMC00] Peter J. Diggle, Jorge Mateu, and Helen E. Clough. "A Comparison between Parametric and Non-Parametric Approaches to the Analysis of Replicated Spatial Point Patterns." *Advances in Applied Probability*, **32**(2):331–343, June 2000.

[Dur99] Rick Durrett. *Essentials of Stochastic Processes.* Springer Texts in Statistics, 1999.

[DV03]   Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003.

[DW96]   Scott H. Decker and Barrik Van Winkle. *Life in the Gang: Family, Friends, and Violence.* Cambridge University Press, 1996.

[ECL05] John E. Eck, Spencer Chainey, James G. Cameron Michael Leitner, and Ronald E. Wilson. *Mapping Crime: Understanding Hot Spots.* National Institute of Justice, 2005.

[EFL10] Mike Egesdal, Chris Fathauer, Kym Louie, and Jeremy Neuman. "Statistical Modeling of Gang Violence in Los Angeles." *SIAM Undergrad. Res. Online*, **3**, 2010.

[EGG10] Eymen Errais, Kay Giesecke, and Lisa R. Goldberg. "Affine Point Processes and Portfolio Credit Risk." *SIAM Journal on Financial Mathematics*, **1**:642–665, 2010.

[EGK04] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltán Toroczkai, and Nan Wang. "Modelling disease outbreaks in realistic urban social networks." *Nature*, **429**:180–184, 2004.

[EL01] Paul P. Eggermont and Vincent N. LaRiccia. *Maximum Penalized Likelihood Estimation*. Springer, 2001.

[Fel02] Marcus Felson. *Of Crime and Criminality: the use of theory in everyday life*, chapter 11. Pine Forge Press, 2002.

[FF09] J. Doyne Farmer and Duncan Foley. "The economy needs agent-based modelling." *Nature*, **460**:685–686, August 2009.

[Fre79] Linton C. Freeman. "Centrality in social networks conceptual clarification." *Soc. Netw.*, **1**(3):215 – 239, 1978-1979.

[GBD10] Volker Grimm, Uta Berger, Donald L. DeAngelis, J. Gary Polhill, Jarl Giske, and Steven F. Railsback. "The ODD protocol: A review and first update." *Ecological Modelling*, **221**(23):2760 – 2768, 2010.

[GG71] Irving J. Good and Ray A. Gaskins. "Nonparametric Roughness Penalties for Probability Densities." *Biometrika*, **58**(2):255–277, 1971.

[GHB08] Marta González, César A. Hildalgo, and Albert-László Barabási. "Understanding individual human mobility patterns." *Nature Lett.*, **453**:779–782, 2008.

[GN06] Michael T. Gastner and Mark E. J. Newman. "The spatial structure of networks." *Eur. Phys. J. B*, **49**:247–252, 2006. 10.1140/epjb/e2006-00046-8.

[Gor10] Mirta B. Gordon. "A random walk in the literature on criminality: A partial and critical view on some statistical analyses and modelling approaches." *Eur. J. Appl. Math.*, **21**(Special Double Issue 4-5):283–306, 2010.

[Gro07] Elizabeth Groff. "Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery." *Journal of Quantitative Criminology*, **23**:75–103, 2007. 10.1007/s10940-006-9021-z.

[Haw71a] Alan G. Hawkes. "Point Spectra of Some Mutually Exciting Point Processes." *Journal of the Royal Statistical Society. Series B*, **33**(3):438–443, 1971.

[Haw71b] Alan G. Hawkes. "Spectra of some self-exciting and mutually exciting point processes." *Biometrika*, **58**:83–90, 1971.

[Her82] David T. Herbert. *Geography of Urban Crime*. Longman Inc, 1982.

[HH04]   Charlotte K. Hemelrijk and Kunz Hanspeter. "Density distribution and size sort-
           ing in fish schools: an individual-based model." *Behavioral Ecology*, **16**(1):178–
           187, 2004.

[HLB12]   Rachel A. Hegemann, Erik A. Lewis, and Andrea L. Bertozzi. "An "Estimate
           & Score Algorithm" for Simultaneous Parameter Estimation and Reconstruction
           of Missing Data on Socal Networks." *submitted to Security Informatics: Special
           issue Computational Criminology*, 2012.

[HO74]   Alan G. Hawkes and David Oakes. "A cluster process representation of a self-
           exciting process." *J. Appl. Probability*, **11**:493–503, 1974.

[Hof09]   Peter Hoff. "Multiplicative latent factor models for description and prediction of
           social networks." *Computational & Mathematical Organization Theory*, **15**:261–
           272, 2009. 10.1007/s10588-008-9040-4.

[HSB11]   Rachel A. Hegemann, Laura M. Smith, Alethea B.T. Barbaro, Andrea L. Bertozzi,
           Shannon E. Reid, and George E. Tita. "Geographical influences of an emerging
           network of gang rivalries." *Physica A: Statistical Mechanics and its Applications*,
           390(21-22):3894 – 3914, 2011.

[Inf08]   Information Technology Division Management Report Unit. "Statistical Digest."
           Los Angeles Police Department, 2008.

[JS01]   S. Rao Jammalamadaka and Ambar SenGupta. *Topics in Circular Statistics.*
           World Scientific Publishing Co. Pte. Ltd., 2001.

[KKM01]   Malcolm W. Klein, Hans-Jürgen Kerner, Cheryl L. Maxson, and Elmar G. M.
           Weitekamp. *Euro Gang Paradox: Street Gangs and Youth Groups in the U.S.
           and Europe.* Kluwer Academic Publisher, 2001.

[KRP10]   Johan H. Koskinen, Garry L. Robins, and Philippa E. Pattison. "Analysing ex-
           ponential random graph (p-star) models with missing data using Bayesian data
           augmentation." *Statistical Methodology*, **7**(3):366 – 384, 2010.

[KWT06]   Malcolm W. Klein, Frank M. Weerman, and Terence P. Thornberry. "Street Gang
           Violence in Europe." *Eur. J. Criminol.*, **3**(4):413–437, 2006.

[LC74]   David Ley and Roman Cybriwsky. "Urban Graffiti as Territorial Markers." *Ann.
           Assoc. Am. Geogr.*, **64**(4):491–505, 1974.

[LM11]   Erik Lewis and George Mohler. "A Nonparametric EM Algorithm for Multiscale
           Hawkes Processes." *Preprint*, 2011.

[LMB11]   Erik Lewis, George Mohler, P. Jeffrey Brantingham, and Andrea L. Bertozzi.
           "Self-exciting point process of Insurgency in Iraq." *Security Journal*, September
           2011.

[LNK05] D. Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. "Geographic routing in social networks." *Proc. Natl. Acad. Sci.*, **102**(33):1162311628, August 2005.

[LST10] Wangyi Liu, Martin B. Short, Yasser E. Taima, and Andrea L. Bertozzi. "Multiscale Collaborative Searching Through Swarming." *Proc. 7th Int. Conf. on Inform. Control, Autom., Robot. (ICINCO)*, June 2010.

[Mat75] Brian W. Matthews. "Comparison of the predicted and observed secondary structure of T4 phage lysozyme." *Biochim. Biophys. Acta*, **405**(2):442–451, 1975.

[MBG09] George O. Mohler, Andrea L. Bertozzi, Tom A. Goldstein, and Stanley J. Osher. "Fast TV Regularization for 2D Maximum Penalized Likelihood Estimation." *To appear in the J. Comput. Graph. Stat.*, 2009.

[MDS08] Susan M. Mniszewski, Sara Y. Del Valle, Phillip D. Stroud, Jane M. Riese, and Stephen J. Sydoriak. "EpiSims simulation of a multi-component strategy for pandemic influenza." In *SpringSim '08: Proceedings of the 2008 Spring simulation multiconference*, pp. 556–563, San Diego, CA, USA, 2008. Society for Computer Simulation International.

[MEH11] Sebastian Meyer, Johannes Elias, and Michael Höhle. "A Space–Time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence." *Biometrics*, 2011.

[MJ00] Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. John Wiley & Sons Ltd, 2000.

[ML08] David Marsan and Olivier Lengliné. "Extending earthquakes' reach through cascading." *Science*, **319**(5866):1076, 2008.

[ML10] David Marsan and Olivier Lengliné. "A new estimation of the decay of aftershock density with distance to the mainshock." *Journal of Geophysical Research*, **115**(B9):B09302, 2010.

[MMK05] Naoki Masuda, Hiroyoshi Miwa, and Norio Konno. "Geographical threshold graphs with small-world and scale-free properties." *Phys. Rev. E*, **71**(3):036108, March 2005.

[MP93] Joan Moore and Raquel Pinderhughes, editors. *In the Barrios: Latinos and the Underclass Debate*. Russel Sage Foundation, 1993.

[MS12] George O. Mohler and Martin B. Short. "Geographic Profiling from Kinetic Models of Criminal Behavior." *SIAM Journal of Applied Mathematics*, **72**(1):163 – 180, 2012.

[MSB11] George O. Mohler, Martin B. Short, P. Jeffrey Brantingham, Frederic Paik Schoenberg, and George E. Tita. "Self-Exciting Point Process Modeling of Crime." *Journal of the American Statistical Association*, **106**(493):100–108, 2011.

[MW02] Michael W. Macy and Robert Willer. "From factors to actors: computational sociology and agent-based modeling." *Annu. Rev. Sociol.*, **28**:143–166, 2002.

[New01a] Mark E. J. Newman. "Scientific collaboration networks. I. Network construction and fundamental results." *Phys. Rev. E*, **64**(1):016131, June 2001.

[New01b] Mark E. J. Newman. "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." *Phys. Rev. E*, **64**(1):016132, June 2001.

[New01c] Mark E. J. Newman. "The structure of scientific collaboration networks." *Proc. Natl. Acad. Sci.*, **98**(2):404–409, 2001.

[New03] Mark E. J. Newman. "The Structure and Function of Complex Networks." *SIAM Rev.*, **45**(2):167–256, 2003.

[NJS02] Mark E. J. Newman, Duncan J.Watts, and Steven H. Strogatz. "Random graph models of social networks." *Proc. Natl. Acad. Sci.*, **99**(suppl. 1):2566–2572, February 2002.

[NSJ01] Mark E. J. Newman, Steven H. Strogatz, and Duncan J.Watts. "Random graphs with arbitrary degree distributions and their applications." *Phys. Rev. E*, **64**(2):026118, July 2001.

[NSK11] Kevin Nichols, Frederic P. Schoenberg, Jon E. Keeley, Andrew Bray, and David Diez. "The application of prototype point processes for the summary and description of California wildfires." *Journal of Time Series Analysis*, **32**(4):420–429, 2011.

[OA82] Yosihiko Ogata and Hirotugu Akaike. "On Linear Intensity Models for Mixed Doubly Stochastic Poisson and Self- Exciting Point Processes." *Journal of the Royal Statistical Society. Series B (Methodological)*, **44**(1):pp. 102–107, 1982.

[Oga88] Yosihiko Ogata. "Statistical models for earthquake occurrences and residual analysis for point processes." *J. Amer. Statist. Assoc.*, **83**(401):9–27, 1988.

[Oga98] Yosihiko Ogata. "Space-Time Point-Process Models for Earthquake Occurrences." *Annals of the Institute of Statistical Mathematics*, **50**:379–402, 1998. 10.1023/A:1003403601725.

[OL09] Mike O'Leary. "The mathematics of geographic profiling." *J. Invest. Psychol. Offender Profiling*, **6**:253 –265, 2009.

[OL11] Mike O'Leary. "Modeling Criminal Distance Decay." *Cityscape: A Journal of Policy Development and Research*, **13**(3):161–198, 2011.

[Pap09]   Andrew V. Papachristos. "Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide." *Am. J. Sociol.*, **115**(1):74–128, 2009.

[Pit10]   Ashley B. Pitcher. "Adding police to a mathematical model of burglary." *Eur. J. Appl. Math.*, **21**(Special Double Issue 4-5):401–419, 2010.

[Pro09]   Product No. 2009-M0335-001. "National Gang Threat Assessment." *Natl. Gang Intell. Cent.*, January 2009.

[PW11]   Michael D. Porter and Gentry White. "Self-exciting hurdle models for terrorist activity." *The Annals of Applied Statistics*, 2011.

[RFT10]   Steven M. Radil, Colin Flint, and George E. Tita. "Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles." *Ann. Assoc. Am. Geogr.*, **100**(2):307–326, 2010.

[Rip81]   Brian D. Ripley. *Spatial Statistics.* John Wiley & Sons, 1981.

[RSH08]   Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, and Song Chong. "On the Lévy-walk nature of human mobility: Do humans walk like monkeys?" In *IEEE INFOCOM 2008 - IEEE Conference on Computer Communications*, pp. 924–932. IEEE, April 2008.

[SBB12]   Laura M. Smith, Andrea L. Bertozzi, P. Jeff Brantingham, George E. Tita, and Matthew Valasik. "Adaptation of an ecological territorial model to street gang spatial patterns in Los Angeles." *Discrete and Continuous Dynamical Systems*, **32**(9):3223–3244, September 2012.

[Sch03]   Frank Schweitzer. *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences.* Springer, 2003.

[SDP08]   Martin B. Short, Maria R. D'Orsogna, V.B. Pasour, George E. Tita, P. Jeffrey Brantingham, Andrea L. Bertozzi, and Lincoln B. Chayes. "A statistical model of criminal behavior." *Mathematical Models and Methods in Applied Sciences*, **18**:1249–1267, 2008.

[Seb02]   Fabrizio Sebastiani. "Machine learning in automated text categorization." *ACM Comput. Surv.*, **34**(1):1–47, 2002.

[Sil86]   Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman & Hall/CRC, April 1986.

[SKW10]   Laura Smith, Matthew Keegan, Todd Wittman, George Mohler, and Andrea Bertozzi. "Improving Density Estimation by Incorporating Spatial Information." *EURASIP Adv. Signal Process: spec. issue Adv. Image Process. Def. Secur. Appl.*, 2010.

[SSB11] Alexey Stomakhin, Martin B. Short, and Andrea L. Bertozzi. "Reconstruction of missing data in social networks based on temporal patterns of interactions." *Inverse Problems*, **27**(11):115013, 2011.

[SSH09] Hermann M. Singer, Irina Singer, and Hans J. Herrmann. "Agent-based model for friendship in social networks." *Phys. Rev. E*, **80**(2):026113, August 2009.

[ST02] Frank Schweitzer and Benno Tilch. "Self-assembling of networks in an agent-based model." *Phys. Rev. E*, **66**(2):026113, August 2002.

[SU09] Didier 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**(6):61110, 2009.

[SW87] Robert J. Sampson and John D. Wooldredge. "Linking the micro- and macro-level dimensions of lifestyle-routine activity and opportunity models of predatory victimization." *Journal of Quantitative Criminology*, **3**:371–393, 1987. 10.1007/BF01066837.

[TCE05] George E. Tita, Jacqueline Cohen, and John Engberg. "An Ecological Study of the Location of Gang "Set Space"." *Soc. Probl.*, **52**(2):272–299, 2005.

[Tes06] Leigh Tesfatsion. "Chapter 16 Agent-Based Computational Economics: A Constructive Approach to Economic Theory." volume 2 of *Handbook of Computational Economics*, pp. 831–880. Elsevier, 2006.

[TG07] Zoltán Toroczkai and Hasan Guclu. "Proximity networks and epidemics." *Physica A: Stat. Mech. Appl.*, **378**(1):68 – 75, 2007.

[Thr27] Frederic M Thrasher. *The Gang: A Study of 1313 Gangs in Chicago*. University of Chicago Press, 1927.

[TLL10] Jie-Jun Tseng, Chih-Hao Lin, Chih-Ting Lin, Sun-Chong Wang, and Sai-Ping Li. "Statistical properties of agent-based models in markets with continuous double auction mechanism." *Physica A: Stat. Mech. Appl.*, **389**(8):1699–1707, 2010.

[TR10] George Tita and Steven Radil. "Making Space for Theory: The Challenges of Theorizing Space and Place for Spatial Analysis in Criminology." *J. Quant. Criminol.*, pp. 1–13, 2010. 10.1007/s10940-010-9115-5.

[TR11] George E. Tita and Steven M. Radil. "Spatializing the social networks of gangs to explore patterns of violence." *Journal of Quantitative Criminology*, pp. 1–25, 2011.

[TRR03] George Tita, K. Jack Riley, Greg Ridgeway, Clifford A. Grammich, Allan Abrahamse, and Peter W. Greenwood. "Reducing Gun Violence: Results from an Intervention in East Los Angeles." *Natl. Inst. Justice, RAND*, 2003.

[VS08]   Alejandro Veen and Frederic P. Schoenberg. "Estimation of SpaceTime Branching Process Models in Seismology Using an EM Type Algorithm." *Journal of the American Statistical Association*, **103**(482):614–624, 2008.

[Wel96]   Barry Wellman. "Are personal communities local? A Dumptarian reconsideration." *Soc. Netw.*, **18**(4):347 – 354, 1996.

[WF09]   Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, 2009.

[Wil06]   Allen Wilhite. "Chapter 20 Economic Activity on Fixed Networks." volume 2 of *Handbook of Computational Economics*, pp. 1013–1045. Elsevier, 2006.

[YL99]   Yiming Yang and Xin Liu. "A re-examination of text categorization methods." In *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*, pp. 42–49, New York, NY, USA, 1999. ACM.

[ZOV02]   Jiancang Zhuang, Yosihiko Ogata, and David Vere-Jones. "Stochastic declustering of space-time earthquake occurrences." *Journal of the American Statistical Association*, **97**(458):369–380, 2002.