ARTICLE IN PRESS

Physica A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Geographical influences of an emerging network of gang rivalries

Rachel A. Hegemann [a], Laura M. Smith [a], Alethea B.T. Barbaro [a,*], Andrea L. Bertozzi [a], Shannon E. Reid [b], George E. Tita [b]

[a] UCLA Mathematics Department, 520 Portola Plaza, Box 951555, Los Angeles, CA 90095-1555, USA
[b] Department of Criminology, Law and Society, 2340 Social Ecology II, Irvine, CA 92697-7080, USA

ARTICLE INFO

Article history:
Received 3 December 2010
Received in revised form 17 March 2011
Available online xxxx

Keywords:
Agent-based model
Complex system
Dynamic network
Gang rivalry
Biased Lévy walk
Geographical Threshold Graph

ABSTRACT

We propose an agent-based model to simulate the creation of street gang rivalries. The movement dynamics of agents are coupled to an evolving network of gang rivalries, which is determined by previous interactions among agents in the system. Basic gang data, geographic information, and behavioral dynamics suggested by the criminology literature are integrated into the model. The major highways, rivers, and the locations of gangs' centers of activity influence the agents' motion. We use a policing division of the Los Angeles Police Department as a case study to test our model. We apply common metrics from graph theory to analyze our model, comparing networks produced by our simulations and an instance of a Geographical Threshold Graph to the existing network from the criminology literature.

1. Introduction

Street gangs are a growing problem around the world [1–3]. In fact, recent statistics from The National Gang Intelligence Center estimate that there are 1 million active gang members in the United States alone [4]. Violence is intrinsic to street gangs, and rival gangs battle to gain respect and street reputation [5,6]. Criminal activities perpetrated by gang members, including armed robbery, homicide, drug dealing, 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 [1–3].

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 plenty of time [7,8]. It is therefore reasonable to think of each gang as a node embedded in Euclidean space [9,10]. 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 [11]. 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 [9,12]. It is for this reason and others listed in Section 1.4 that we consider Hollenbeck as a case study for our model.

Statistical approaches are often used to analyze gang activities. For example, in their recent paper, the authors of [9] used a network statistical approach called CONCOR to partition the region of Hollenbeck into areas of similar violence activity and

* Corresponding author. Tel.: +1 310 794 6646.
E-mail addresses: dansonr@math.ucla.edu (R.A. Hegemann), lsmith@math.ucla.edu (L.M. Smith), alethea@math.ucla.edu (A.B.T. Barbaro), bertozzi@math.ucla.edu (A.L. Bertozzi), reids@uci.edu (S.E. Reid), gtita@uci.edu (G.E. Tita).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.05.040

geographic proximity. From this, they were able to conclude that a relationship exists between the location of a gang and its rivals. Other authors have utilized various statistical approaches to examine gang-related data [10,7]. The methodology of determining factors that correlate with gang activity has been able to describe certain features of the observed system. This approach has made important contributions in the field, but it cannot make solid causal arguments or test theories [13,14]. Though one could argue that creating an experiment would help validate causal arguments, many of these experiments are not feasible due to monetary and resource constraints and could be unethical and infringe on basic human rights. Even in circumstances where the resources are available, interventions may fail to be implemented as planned. This is evident in the work of [10].

Obtaining complete and valid data sets is a common issue in the field of criminology. Data sets are many times unreliable through inaccuracies, underreporting, and potential bias [15]. This provides a fundamental problem in the conclusions made from faulty data. In [16], a strong case is made to move beyond statistical modeling and instead model social phenomena using a mathematical approach. We use such an approach with the aim to understand the plausible mechanisms for the formation of gang rivalries and to provide social scientists a means by which to test social theories. For comparison, we propose two different mathematical techniques: a pure network and an agent-based approach.

Our goal in this work is to understand how long-term rivalries among gangs develop by using a model including geography, social dynamics, and human mobility patterns. Because we have only one observed rivalry network, this is a very difficult inverse problem. Furthermore, statistical approaches are not able to determine causal effects. As a solution, we propose an agent-based model that is coupled to a dynamically evolving network. This bottom-up approach simulates the mobility of gang members by using the conclusions from current literature in human mobility patterns, see Section 2.1 for more details. These agents interact to form a rivalry network. We compare the resulting simulated network to the gang rivalry network observed in the eastern Los Angeles division of Hollenbeck [10,9].

The outline of our paper is as follows: network and agent-based modeling approaches are described in Sections 1.1 and 1.2, and previous work on crime models is discussed in Section 1.3. Detailed information concerning the particulars of Hollenbeck are found in Section 1.4. In Section 2, we outline the proposed model. In Section 3, we describe two baseline models, one instance of a Geographical Threshold Graph and a network derived from Brownian motion. We contrast these simpler models with our model to demonstrate the need for a more complex approach. In Section 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 5 provides a sensitivity analysis of our model. We conclude and give future directions in Section 6.

1.1. Network models

General network models and the corresponding analysis are useful for describing the behavior of complex systems and have played an increasingly active role [17–19]. 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 [20–22].

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 [23]. The importance of geography is also seen in friendship networks in [24]. This work was descriptive by nature, and was therefore limited in the conclusions that could be drawn. 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 [25]. The study found that an estimated 69% of a person's friends can be described by geography. This paper constructed a simulation and was able to create plausible scenarios and test theories of information spreading on a social network, again highlighting a strength of a mathematical modeling approach.

In the scenario of gang rivalries, the geographic location of gangs play a role in the observed gang rivalry structure [10]. Because of this, one of our approaches is to construct a simple network model that incorporates the proximities between gang set spaces. One method for incorporating geographical information into the random graph construction is by using a Geographical Threshold Graph [26–28]. 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 an 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.

1.2. Agent-based models

Network models can provide a computationally inexpensive means to reproduce an observed network, but these models do not lend themselves to describing phenomena of interest beyond the structure of the network. An alternative modeling approach is to use agent-based models. These models are generally used for complex systems, since they are able to capture details at the level of an individual, or *agent*. In this class of models, agents often move through phase space and interact

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

amongst themselves, producing complex dynamics and patterns from simple behavioral rules. Agent-based models can answer different questions from a network approach since they focus on the way individuals actions can determine the behavior at the system level.

This type of modeling approach is widely used in economics [29–31], epidemiology [32], sociology [33], biology [34], and other situations in which complex systems are encountered. The strength of agent-based modeling is that it allows for responses at the individual level to be directly incorporated into the model. For example, agent-based models have been proposed in the context of searching and swarming [35], and much interesting mathematical analysis has been done on the behavior of these systems [36,37]. In this way, modelers can include information and behavioral dynamics from scientists who study the complex system of interest. Agent-based methods have been proposed to describe social systems and economics since they can provide a means by which to test theories about individual dynamics in cases where the dynamics are not precisely known [38,13].

In our model, 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 a model that uses an agent-based approach to form an emerging network [39,40]. 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 [41,42]. 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.

One of the major strengths of this approach is the flexible framework available for these models. For example, this method can easily incorporate environmental and spatial information inherent to the system. One example of this is shown in [43], where the authors were able to incorporate environmental information in the form of temperature and current, as well as geography of landmasses, to accurately model and predict the migration patterns of a species of pelagic fish. In our agent-based model, we use information about the environment in the form of freeways, rivers, and road density. In a network context, coupling an agent-based model to a network is a novel way to explore 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.

1.3. Previous work on crime modeling

Many mathematical models have been created to study the patterns, mechanisms, and potential interventions associated with crime. Research has been conducted to address various aspects of burglaries such as hot spot formation [44–46], policing strategies [47,48], criminal cooperation [49], and geographical profiling [50]. Some attempts have also been made to model gang behaviors. The study in [51] examines the short-term retaliatory behavior of the rivalries based on between gang violence data from the LAPD. In this work, 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 the model proposed by [52], an agent-based approach was used to simulate the location of violent interactions and gang retaliations in Hollenbeck. Embedded in the model was a rivalry network. Though the model recreated similar features to the violence data, the model made unrealistic assumptions about the mobility of the agents. For instance, the agents' movements were influenced by the location of every rival gang member, an implausible assumption. Further, agents only move toward rival gangs members, ignoring their own set space, or center of activity. However, according to the criminology literature, this retaliatory behavior is only seen on short time scales [51]. Furthermore, gang members tend to avoid the territory of rival gangs and spend large quantities of time at their set spaces [53,54]. Another concern is that the agents ignore geographic features, such as highways and rivers, known to correlate with the rivalry structure [9]. It has been shown in other models simulating human movement that highways restrict movements [55].

We propose a bottom-up approach with an agent-based model that incorporates movement rules from literature on human mobility in order to capture the long-term gang rivalry structure. These rules consider geographical features known to relate to movement dynamics. We also consider current literature on known gang behavior as a basis for directional decisions. More details are seen in Section 2.1. We use Hollenbeck as a case study for our model.

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 Fig. 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 [9,10]. The set spaces for the gangs and the corresponding observed rivalry network are displayed in Fig. 1, as given in [9].

Certain properties of Hollenbeck make it accessible to modeling the gang rivalry networks outlined in [10,9]. First, it is a closed system in that the gang activity within Hollenbeck is generally isolated from gang activity outside of Hollenbeck.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

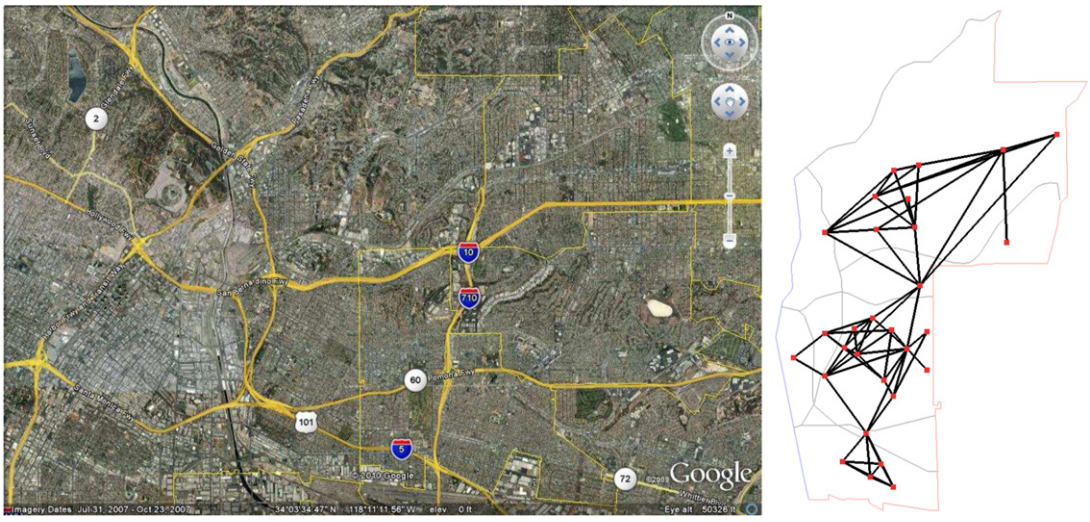

Fig. 1. Google Earth™ Image of the Hollenbeck area (left). Map of the Hollenbeck area with the location of the gang set spaces, or centers of a gang's activity, and the corresponding rivalry network approximated by [9], where a node of the network represents a set space, and an edge represents a rivalry between two gangs (right). Major roads, highways, the Los Angeles river, and division lines are also seen in both images.

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.

2. Our model

Our objective is to model the long-term gang rivalry structure and gang member mobility by incorporating simple behavioral rules and geographical factors, such as road density, highways, and locations of gangs' set spaces (centers of activity). The movement of each agent is not intended to model each detail involved in an individual's mobility, but rather capture the statistical behavior of people's movements observed from the literature. 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.

2.1. 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. Several studies give compelling evidence that when people move unconstrained environment, the jump lengths between movements is distributed like a power law [56,57]. Further, in the presence of obstacles such as roads and buildings, the jump lengths more accurately follow a bounded power law distribution [58].

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 [57]. Gonzalez et al. confirmed in their data that humans do tend to frequent a small number of locations often [58]. 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 [7]. Literature on gang activity suggests that, in general, gang members tend to stay away from their rival gangs' set spaces [54]. 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 [53]. 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

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

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. Simulations have shown that they provide an obstacle to mobility [55]. It has been posited that these features play a role in gang rivalry networks [9]. Therefore in our model, we view major roads, highways, and rivers as semi-permeable boundaries that affect the agents' movements.

2.2. Model details

2.2.1. 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, or centers of activity. 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. Embedded in the city is a rivalry network among the gangs in the city. This network is encoded in the rivalry matrix R. Each element (i, j) of R corresponds to the number of interactions between gang i and gang j. When two agents are within interaction range, we consider them to have interacted and the corresponding element of the rivalry matrix, R, is updated. Refer to Section 2.2.3 for details on R.

2.2.2. Environment

The environment of interest is on the scale of a small city. Agents and gang set spaces, or centers of gang activity, have a coordinate location in free space. The rest of the physical features of the city are approximated by an NxM grid. Each point in free space is identified with the nearest grid element. The size and number of grid elements are constant throughout the simulation and are limited by the available data and the memory of the computer.

Two specific features encoded by this grid are the road density and semi-permeable boundaries represented by a region map. Each grid element of the road density 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. We pair this region grid with a transition matrix that stores the associated probability of an agent to cross from one region to another. These probabilities are determined at the start and remain constant throughout the simulation. This is implemented to discourage agents from crossing freeway boundaries.

2.2.3. 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

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

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. Although the final rivalry network is symmetric, the influence of the rivalry matrix R_{ij} on the agents' movements during the course of the simulations are not symmetric. The data available to us was in the form of a symmetric network, and so we chose the thresholding rule to be comparable with the data. Other thresholding rules that result in asymmetric networks could be implemented. For example, if the thresholding rule for gang i solely depended on the quantity $\rho_i(j) > T$, then the resulting network could be asymmetric.

2.3. Process overview and scheduling

At each iteration an agent is chosen from the set of all agents with equal probability. The selected agent performs one step of a random walk by choosing a jump length and direction from probability distributions for its new prospective location. Depending on the distributions for the jump length and direction, different random walks will occur. The literature on human mobility suggests humans move according to a truncated Lévy walk, motivating our model selection. Further, it is unreasonable to assume that an person's direction of movement is solely determined by set space locations. Therefore, we use a statistical distribution to simulate an agent's directional choice. See Section 2.1 for more motivational details.

The jump length, x, is chosen from the Bounded Pareto probability density function,

$$P(x; k, x_m, x_M) = \frac{k x_m^k x^{-k-1}}{1 - \left(\frac{x_m}{x_M}\right)^k} \quad k > 0,\ x_M \geq x \geq x_m > 0. \tag{2}$$

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

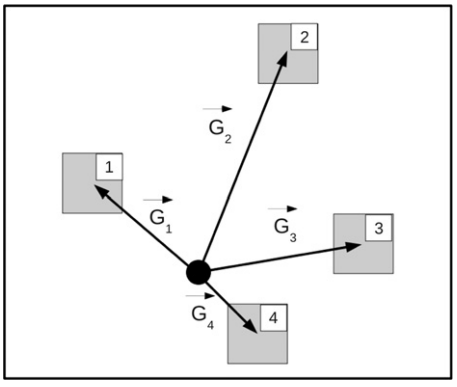

Fig. 2. Cartoon example of the direction vectors incorporated in the direction of bias formula, Eq. (4). The agent in this example is located at the dot. Here G_1, G_2, G_3, and G_4 show the vectors pointing toward the set spaces of gangs 1 through 4, respectively. Depending on the choices of H_i and D_{ij}, different movement dynamics are possible.

The minimum and maximum jump lengths, x_m and x_M respectively, provide the bounds for the jump length x. The scaling parameter, k, determines how quickly the probability density function decays from the minimum jump length to the maximum jump length. For all agents the minimum jump length, x_m, and scale, k, are fixed parameters. 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, with x_M ranging from a to A. These values remain constant throughout the simulation. Here, A is called the largest maximum jump length and a is the smallest maximum jump length. Then x_M is calculated via

$$x_M = (1 - \delta) \cdot A + a, \tag{3}$$

where the road density at the agent's location, δ, is between 0 and 1.

The second quantity needed to move the agent to the next location is the new direction, θ. This is obtained by constructing a deterministic direction of bias. This bias incorporates the agent's location with respect to its home set space and the location of its rival gangs' set spaces. Agents have a higher probability of moving in this direction. However, with lower probability, they have the ability to move any direction. To account for this mobility pattern, the von Mises distribution is used to simulate the direction the agent will move, θ.

More specifically, given an agent in gang i, the bias direction, μ_i, is defined as

$$\langle y, z\rangle = H_i(\|\overrightarrow{G_i}\|_2)\frac{\overrightarrow{G_i}}{\|\overrightarrow{G_i}\|_2} + \sum_{j\neq i} D_{ij}(\|\overrightarrow{G_j}\|_2)\frac{\overrightarrow{G_j}}{\|\overrightarrow{G_j}\|_2} \tag{4}$$

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

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 toward the agent's home set space, and when $l \neq i$, it points toward a different gang's set space. This concept is shown in the cartoon example in Fig. 2.

In Eq. (4), H_i gives the rules for weighting toward 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:

$$H_i(\|\overrightarrow{G_i}\|_2) = h_i\,\|\overrightarrow{G_i}\|_2, \tag{5}$$

$$D_{ij}(\|\overrightarrow{G_j}\|_2) = w_{ij}(R)\,\frac{1}{\|\overrightarrow{G_j}\|_2}. \tag{6}$$

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.

After determining the direction of bias from Eq. (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) [59–61]. For $\theta \in [-\pi, \pi]$, the von Mises distribution is given by

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

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

Table 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.

Parameters	Acceptable values	Hollenbeck values	Tested range	Description
Agent				
x_m	$0 < x_m < a$	0.1		Minimum jump length
k	$0 < k$	1.1	[1, 1.9]	Bounded Pareto scaling parameter
κ	$0 \leq \kappa$	3.5	[1.5, 5]	von Mises scaling parameter
h_i	$-\infty < h_i < \infty$	1		Home weighting
$w_{ij}(R)$	$-\infty < w_{ij}(R) < \infty$	$-\rho_i(j)$		Rival gang weighting
Environment				
N_i	$N_i \in \mathbb{Z}^+$	$14 \leq N_i \leq 598$		Number of gang members in gang i
S_i	$S_i \in \mathbb{R}^2$	See Fig. 1		Location of gang i set space
A	$a < A$	200	[100, 400]	Largest maximum jump length
a	$x_m < a < A$	100	[100, 200]	Smallest maximum jump length
B	$0 \leq B \leq 1$	0.2	[0, 0.5]	Permeability of boundaries
Network				
T	$0 \leq T$	0.04	[0, 0.06]	Threshold for existence of an edge

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, θ, and jump length, x, a prospective location is calculated. The new location is then checked to see if the result would move the agent into a different region. Within the same region movement is unrestricted. 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 location of interactions is also recorded and could be of interest to other applications, see discussion in Section 6 and Fig. 14.

The model is run until limiting behavior is observed in all of the metrics. In the absence of an observed network, it could be helpful to run simultaneous simulations with different random seeds and calculate the variance of each metric over the course of the simulation run. When the variances of each metric levels off, terminate the simulation. In the case of Hollenbeck, the final network is taken after 20,000,000 iterations and then thresholded to ignore infrequent interactions. For more details on long-term behavior of the model, see Section 4.2.

2.4. 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 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.

2.5. Hollenbeck parameters

The grid of environment features of Hollenbeck was approximated from the Google Earth™ image in Fig. 1. Hollenbeck is about 39.4 km^2 [9,10]. 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 Fig. 3. The boundaries of the Hollenbeck region were approximated using points from the geographic features visible from Google Earth™. These boundaries were used to construct the region grid. To approximate the road density, the Weighted H^1 Maximum Penalized Likelihood Estimation method was used [62]. Although the area of Hollenbeck does not have large invalid regions of agent movement, alternative cities could have regions where human mobility is not expected to occur, such as in lakes, mountains, and oceans. 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 [63–66]. 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 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.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

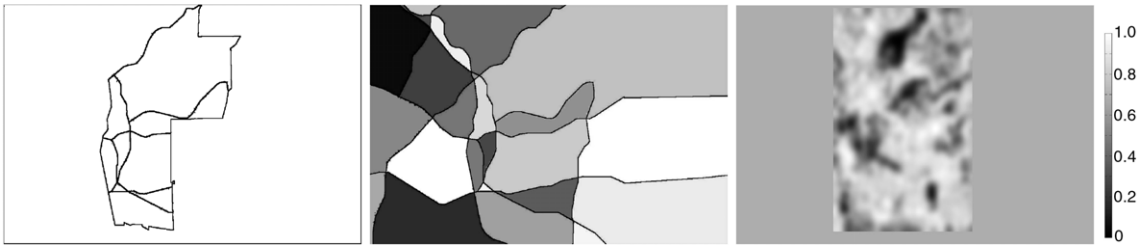

Fig. 3. The image on the left shows the location of Hollenbeck in the $N \times M$ environment grid. The semi-permeable boundaries encoded in the model are displayed in the center image. The shades in gray of this image are used to distinguish among regions. On the right, we used a Weighted H^1 Maximum Penalized Likelihood Estimation method with a road map as the initial data to approximate the road density of Hollenbeck [62]. 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.

3. Baseline comparison models

3.1. Geographical Threshold Graphs (GTGs)

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 [26–28]. 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.

Geographical Threshold Graphs are not deterministic in general. However, we are using this framework to construct one network in order obtain a reasonable comparison to our proposed model. To this end, we provide deterministic weights, η_i, taken from data to be the size of each gang. 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 and choose $\beta = 2$. The threshold was chosen to give the same number of rivalries as the observed rivalry network.

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. Experiments without encoding the semi-permeable boundaries were conducted with unsatisfactory results. Therefore, the semi-permeable boundaries of the model are incorporated. In this model, 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.

3.3. Baseline network graphs

Fig. 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.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

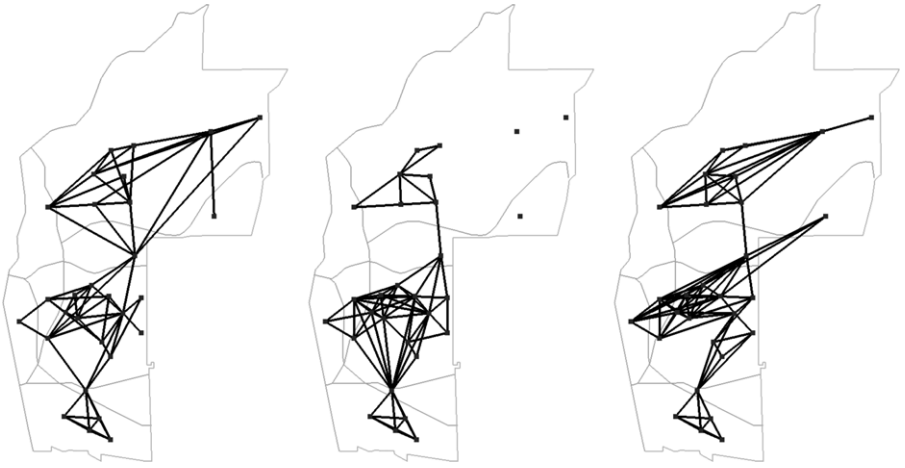

Fig. 4. A visual comparison of the observed rivalry network (left), GTG (center), and BMN (right). Here, a node of the network represents a set space, and an edge represents a rivalry between two gangs.

4. Results

The proposed model produces strong results when compared to the baseline models. For analysis of the model, this section is divided into two parts, the internal properties of the model and the comparison among models. Due to the stochastic nature of the movement rules, the final network is not deterministic. Despite fluctuations among simulation runs, within a single run the model exhibits long-term stable behavior in the metrics used for analysis. The stochasticity and long-term behavior are the internal properties we examine in detail in Sections 4.1 and 4.2, respectively. The stochastic nature of the model allows for a more realistic scenario, in the case where the observed network is just one instance of a random process. Further, the existence of stable long-term behavior in our model is important to replicating the observed system, since research has demonstrated that the rivalry networks among gangs tend to be stable over time [11,9,8]. A more detailed discussion of this can be seen at the end of Section 4.2.

After an examination of the internal properties of the model, we subject the models to a number of metrics in order to capture the features of the observed network as well as the accuracy of model networks. Keeping in mind the potential for variants of an observed network, the measures we chose to evaluate the performance of our model are fairly robust to small perturbations to the observed network. The metrics used to assess the shape of the network and the accuracy are defined in Section 4.3, and their corresponding results are displayed in Section 4.4.

For analysis and comparison, we took one simulation run as a showcase of the model. This network was obtained by searching the parameter space within the ranges specified in the fourth column of Table 1, allowing for dependences between parameters. The 34,128 simulated networks were then sorted according to accuracy, defined in Eq. (7). Because each of the gangs in Hollenbeck are active, the graph with the highest accuracy with all non-zero degree nodes was chosen. The parameter values for the optimal run are found in the third column of Table 1. We will to refer to this as Simulated Biased Lévy walk Network (SBLN). Fig. 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.

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. Fig. 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

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

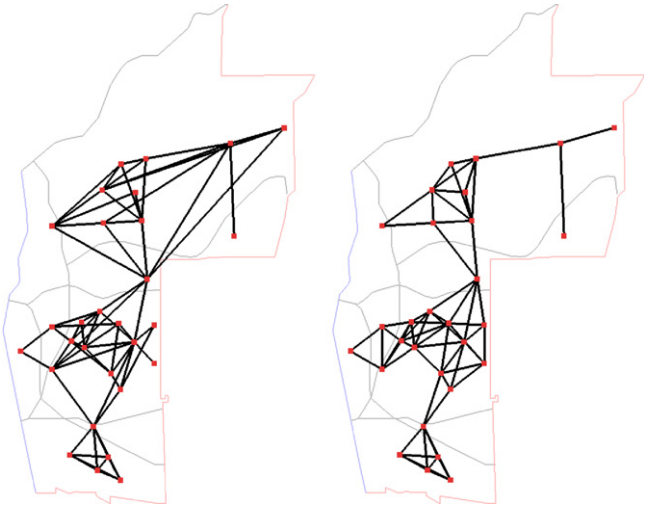

Fig. 5. Comparison of the observed rivalry network (left) and the SBLN (right). Here, a node of the network represents a set space, and an edge represents a rivalry between two gangs. The SBLN has a shape and structure similar to the observed network, but does not capture many of the longer edges.

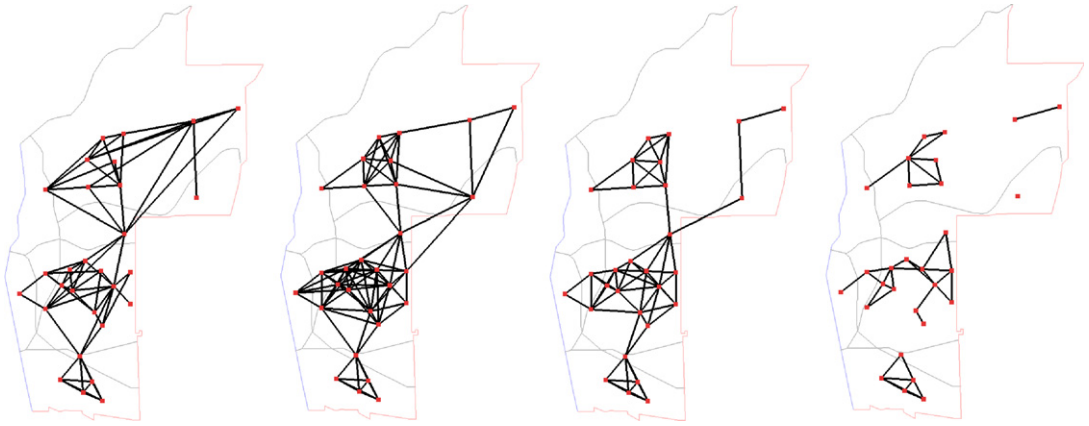

Fig. 6. Percent edge agreement for the ensemble of runs for the SBLN parameter values. These four images give a comparison of, from left to right, the observed rivalry network, the Ensemble SBLN 1% edge agreement, the Ensemble SBLN 50% edge agreement, and the Ensemble SBLN 100% edge agreement. Here, a node of the network represents a set space, and an edge represents a rivalry between two gangs.

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.

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. Fig. 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 Fig. 7). The 100% edge agreement network in Fig. 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 Fig. 7). These are the edges not appearing in the 1% edge agreement network in Fig. 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

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

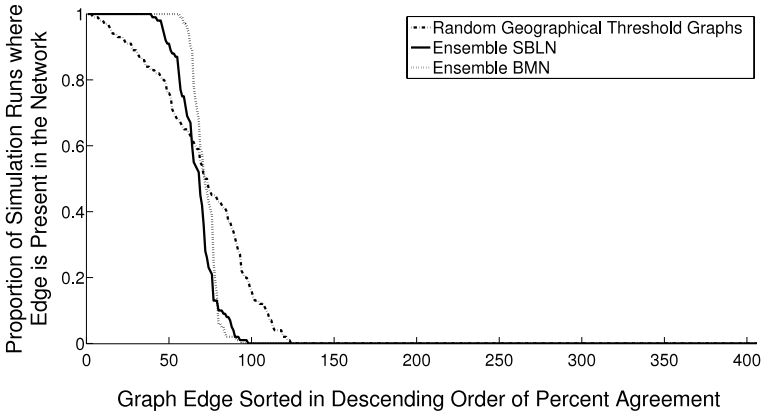

Fig. 7. Plot of the edge persistence for the Ensemble SBLN (solid), Ensemble BMN (thin-dash), and an ensemble of random Geographical Threshold Graphs (thick-dash). The randomly weighted Geographical Threshold Graphs were constructed with random weights and have a median of 69 edges present. The edges 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.

models, there is agreement among the edges of the Ensemble BMN and Ensemble SBLN, maintaining some structure within the simulated networks.

Further analysis was conducted on the stochastic nature of the proposed SBLN model by looking at the metrics calculated on the network produced with varying levels of percent edge agreement. Plots of the accuracy, density, and nodal degree variance for the collection of randomly weighted Geographical Threshold Graphs, the Ensemble SBLN, and the Ensemble BMN are displayed in Fig. 8. A definition of these metrics can be found in Section 4.3. As the percent edge agreement increases, the accuracy metric generally increases for the randomly weighted Geographical Threshold Graphs. However, the Ensemble SBLN and Ensemble BMN do not show much variation, though it is notable that the Ensemble SBLN consistently has higher accuracy across percent edge agreements. For the density metric, as the percent edge agreement increases, the density decreases as expected. The randomly weighted Geographical Threshold Graphs have the greatest variation in the density metric and nodal degree variance metric. This implies that there is a greater degree of stochasticity for this model. Alternatively, more structure is seen in the Ensemble SBLN and Ensemble BMN. The two ensembles show small variations in the density and nodal degree variance metrics for percent edge agreements between 15 and 75. Note that the Ensemble BMN has little change in nodal degree variance across all percent edge agreements. This indicates a lower degree of stochasticity in the model, which agrees with the steep decline in Fig. 7.

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. Despite this, the results indicate there is limiting behavior of the observed metrics as the simulation progresses. Fig. 9 displays the density and accuracy over the progression of the simulations for the Ensemble SBLN. In general, the accuracy metric we used is a measure of how close the simulated network is to the observed network. The density of a graph is proportional to the average number of rivals of all gangs; for further definitions of these metrics, refer to Section 4.3. Each run is observed every 1000 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 seem to stabilize. For the average values of the density and accuracy of the last iteration, refer to Tables 2 and 3. Further, we tracked the variance of the metrics over the course of the simulation; plots of the variance for density and accuracy are shown in Fig. 10. We observe that the variance of the network metrics across simulations levels out, indicating an appropriate time to terminate the simulation.

The Ensemble SBLN exhibits 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 [11,9,8], and that the activity spaces of gangs are anchored to specific places [7,67]. 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 [10]. 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

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

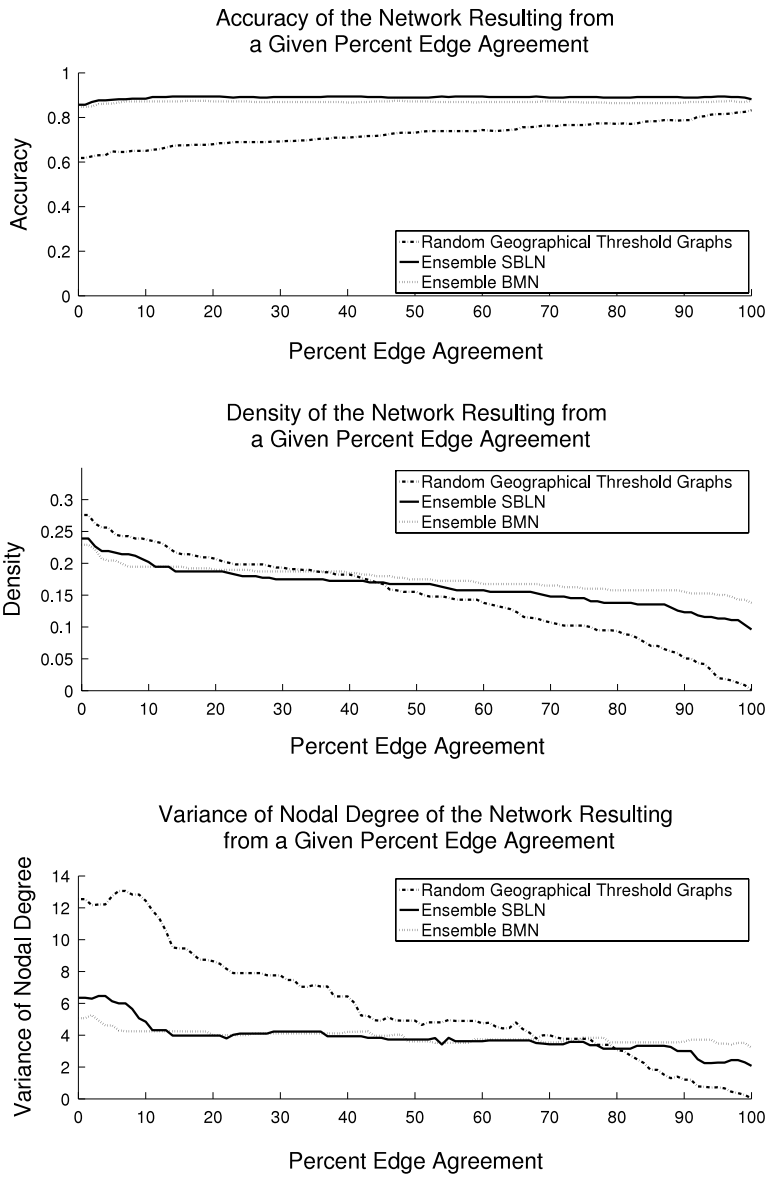

Fig. 8. Plots of the measured accuracy (top), density (middle), and variance (bottom) of the networks constructed from a given percent edge agreement. The metric values corresponding to the Ensemble SBLN (solid), Ensemble BMN (thin-dash), and an ensemble of random Geographical Threshold Graphs (thick-dash) are plotted together. The randomly weighted Geographical Threshold Graphs were constructed with random weights and have a median of 69 edges present.

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.

4.3. Metrics used for analysis

We analyze our model according to several common metrics for accuracy and shape. Since we are examining a rivalry network, there are certain popular measures that are not applicable here. For example, the clustering coefficient is often used in social and friendship networks, describing the proportion of a node's neighbors who are also neighbors [20,68,69]. In the analysis of gangs this metric is not relevant because the rival of a rival gang is not necessarily a rival. In addition, there may be errors in an observed rivalry network. With these in mind, we carefully choose measures that are applicable to gang rivalry networks and robust to small perturbations in the observed data. For example, the degree distribution is commonly used in social network analysis [70,71,68,21,19,20]. In the context of gangs, the degree distribution of the network describes the number of rivalries of each gang. Knowing the distribution of rivalries at the city level may be helpful to policy makers for determining how to allocate resources. The measures we used are motivated and defined in Sections 4.3.1 and 4.3.2.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

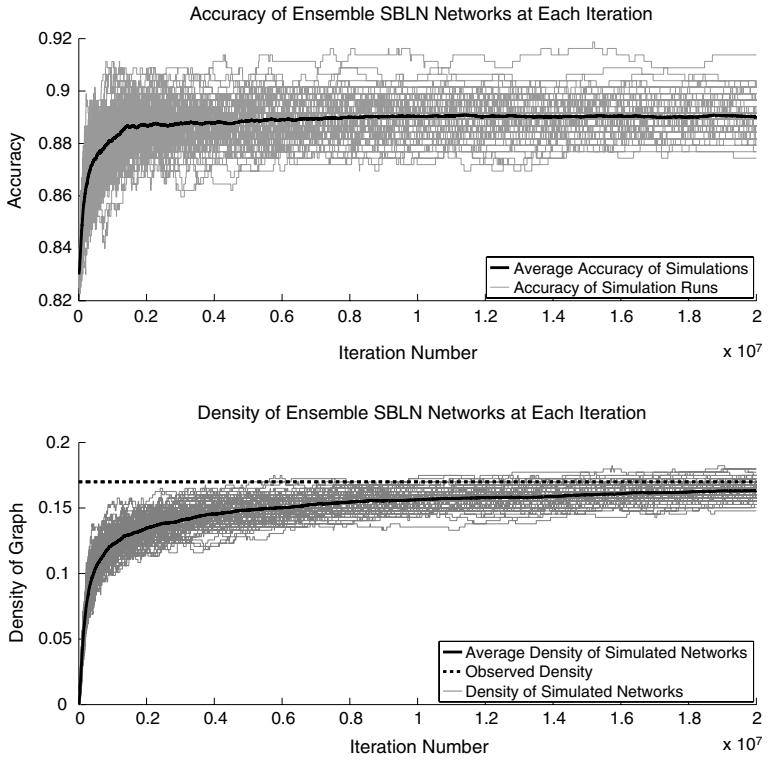

Fig. 9. Plots of the accuracy (top) and the density (bottom) of the SBLN over the $2 \cdot 10^7$ iterations. Each of the 100 Ensemble SBLN runs are plotted by thin lines. The average over all the runs at each sampled iteration is shown with the solid, thick line. The density of the observed network is 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.

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

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{7}$$

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 [72]. The *F1 score* provides another measure to analyze the accuracy of the predicted network, [73,74], and is defined as

$$\mathrm{F1} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \tag{8}$$

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)* [75,72]. This measurement varies between -1 and 1, where a value of 1 is a perfect prediction. The MCC is defined as follows:

$$\mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}}. \tag{9}$$

The ACC, F1 score, and MCC provide a summary of the TP, TN, FP, and FN. These measures are fairly robust to changes in the observed data. For example if one edge were added or removed in the observed network, at most the ACC would change by $\frac{2}{N(N-1)}$, where N is the number of nodes.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

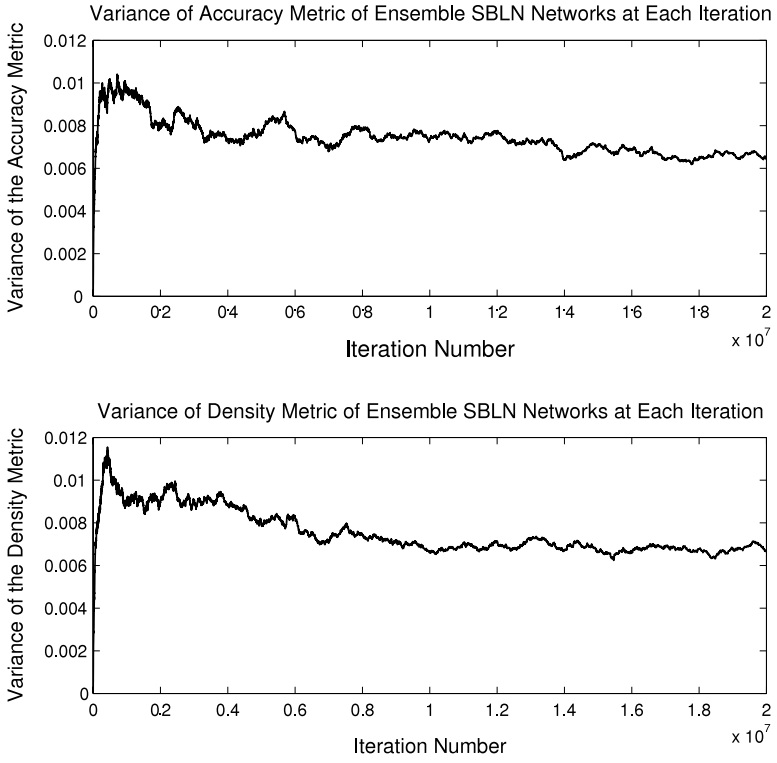

Fig. 10. Plots of the variance of the accuracy (top) and the density (bottom) measures of the Ensemble SBLN over the $2 \cdot 10^7$ iterations. The variances of the 100 Ensemble SBLN runs were calculated at 1000 iteration increments.

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.

4.3.2. Shape metrics

We would like to verify that the simulated network has a similar shape to that of the observed network. To do this, we calculate the graph *density*, standard *variance* of nodal degree and Freeman's *centrality measure* of the graph [69,76]. Given N nodes, the density of a graph is defined by

$$\sum_{i=1}^{N} \frac{\text{degree}(i)}{N(N-1)}. \tag{10}$$

In the context of gangs, the degree of a gang is equivalent to the number of rivals of the gang. The density of the rivalry network is the average number of rivalries scaled by a normalization, $\frac{1}{N-1}$. Networks with the same number of edges and nodes have the same density measure. Further, this metric is fairly robust to perturbations of the graph. For instance, if one edge were added or removed, the density metric would only change by $\frac{2}{N(N-1)}$.

Two other metrics, the variance of nodal degree and centrality measure, give an indication of the spread of degrees among the gangs. The variance of nodal degree for a graph is

$$\sum_{i=1}^{N} \frac{(\text{degree}(i) - \text{aveDegree})^2}{N}. \tag{11}$$

The centrality measure of the graph is defined to be

$$\sum_{i=1}^{N} \frac{\text{maxDegree} - \text{degree}(i)}{(N-1)(N-2)}. \tag{12}$$

These measures provide summary statistics with which to describe the general shape of the graph. This becomes useful when comparing the output to multiple models. For instance, if the observed network has a few gangs with high degrees and the rest of the gangs with low degrees, then this will be reflected high values in the spread statistics. A good model should capture this feature.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

Table 2
Accuracy measures for the SBLN, Ensemble SBLN, GTG, BMN, and Ensemble BMN. The σ denotes the standard deviation of the ensemble metric values.

		TP	TN	FP	FN	ACC	F1 score	MCC
SBLN		50	320	17	19	0.9113	0.7353	0.6822
Ensemble	Average	45.50	316.1	20.90	23.50	0.8906	0.6722	0.6069
SBLN	$\pm\sigma$	$\pm$1.269	$\pm$2.424	$\pm$2.424	$\pm$1.269	$\pm$0.0077	$\pm$0.020	$\pm$0.025
GTG		48	316	21	21	0.8966	0.6957	0.6333
BMN		47	313	24	22	0.8867	0.6714	0.6031
Ensemble	Average	43.61	309.2	27.76	25.39	0.8691	0.6213	0.5424
BMN	$\pm\sigma$	$\pm$1.380	$\pm$1.390	$\pm$1.390	$\pm$1.380	$\pm$0.0051	$\pm$0.016	$\pm$0.019

Table 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.

		Density	Variance of nodal degree	Centrality
Observed		0.16995	4.32105	0.20106
SBLN		0.16503	3.54578	0.16799
Ensemble	Average	0.16355	3.66423	0.15040
SBLN	$\pm\sigma$	$\pm$0.005593	$\pm$0.48395	$\pm$0.01883
GTG		0.16995	9.97622	0.27778
BMN		0.17488	3.88585	0.15741
Ensemble	Average	0.17579	3.93926	0.16065
BMN	$\pm\sigma$	$\pm$0.004546	$\pm$0.41351	$\pm$0.02635

The *degree distribution* has been widely used to understand the overall network structure [68,21,19,20]. We compare the nodal degree cumulative distribution function (CDF) of our simulations with the observed network.

4.4. Evaluating models using graph metrics

4.4.1. Accuracy metric results

Table 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.

4.4.2. Shape metric results

Table 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.

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 Fig. 11. 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.

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,

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

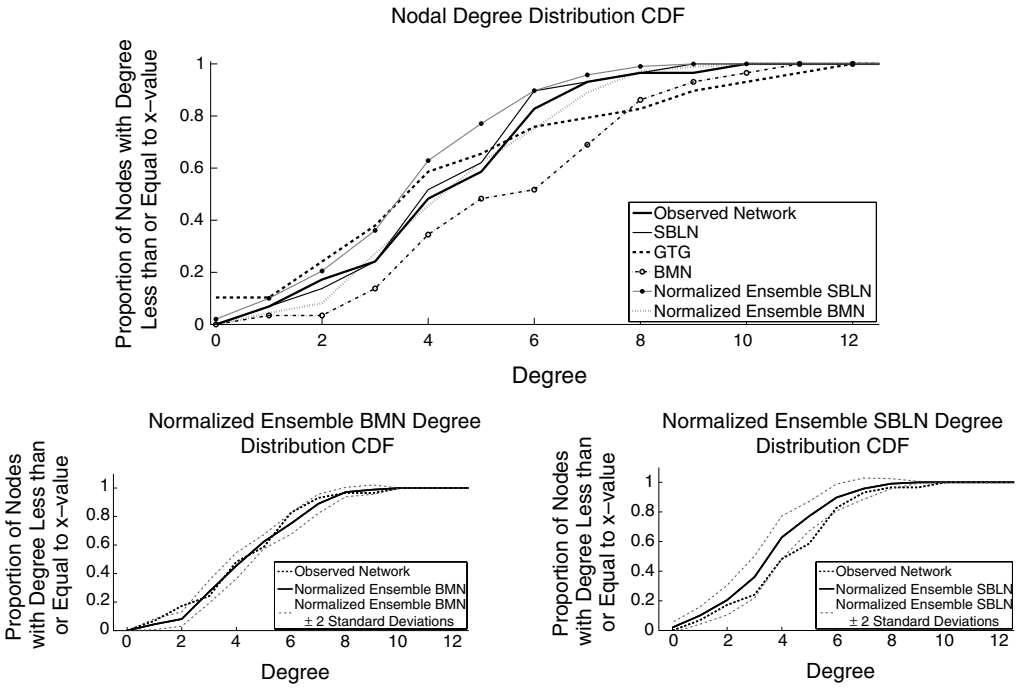

Fig. 11. The top figure plots together the cumulative distribution functions of the degree distribution for the observed network (thick-solid), GTG (thick-dashed), BMN (dot–dash), normalized Ensemble BMN (thin-dash), SBLN (thin-solid), and normalized Ensemble SBLN (dot-solid). A normalized ensemble CDF 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).

Table 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 1.

Bounded Pareto scaling parameter	$k \in [0.77, 1.43]$
von Mises scaling parameter	$\kappa \in [2.45, 4.55]$
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	$T \in [0.028, 0.052]$

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.

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 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 4.

For each simulation run, we compute the accuracy, Matthews Correlation Coefficient, F1 score, centrality measure, variance of nodal degree, and density 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 more dramatic results are plotted in Fig. 12. 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.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

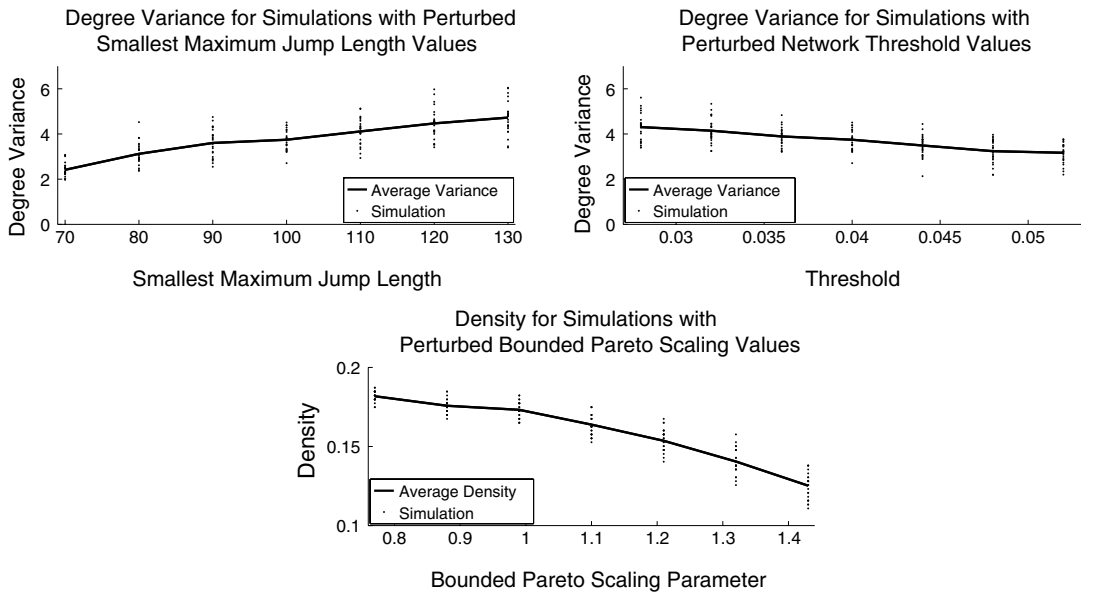

Fig. 12. Plots of the nodal degree variance versus the smallest maximum jump length (top left), and the network threshold (top right). We also display the density versus the Bounded Pareto scaling parameter (bottom), where the vertical axis has been rescaled for visualization. The solid curve indicates 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.

Table 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 Fig. 12 are highlighted in bold font. Fig. 13 displays this information in a color map.

	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

As seen in Fig. 12, 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 Fig. 12. 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 Fig. 12, 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 cannot compare the trend of the 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 5 and visualized in Fig. 13. In Table 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 Fig. 13. The dark, and light, intensities of the color map represent large positive, and negative, values of the best fit line slope.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

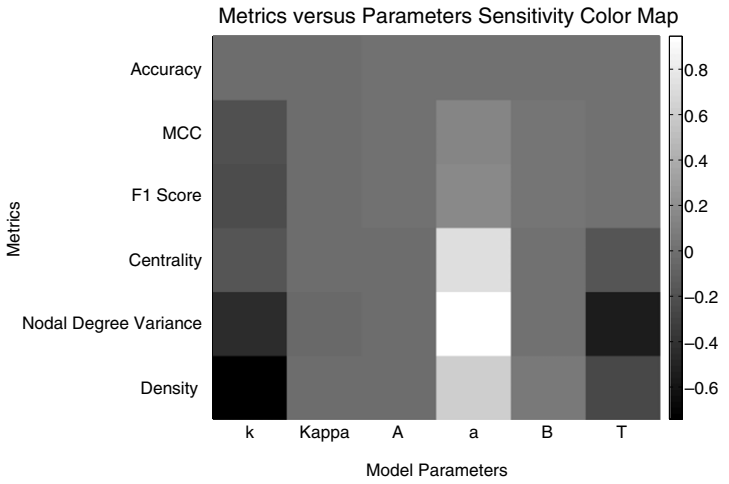

Fig. 13. Slopes of the best fit line to the rescaled data for each of the parameter and metric combinations depicted in a color map. The parameters varied include the Bounded Pareto scaling parameter, k, the von Mises scaling parameter, Kappa, the largest maximum jump length, A, the smallest maximum 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 5.

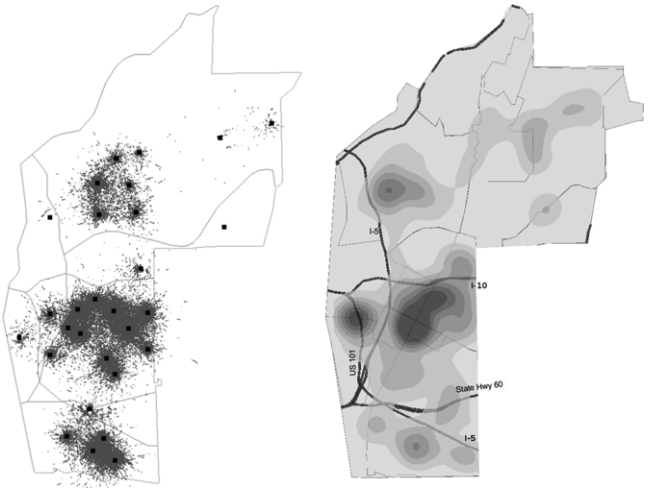

Fig. 14. Locations of all the interactions between agents for one of the Ensemble SBLN runs (left). Density map of gang-related violent crimes in Hollenbeck between 1998 and 2000 (right).

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 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 metrics. By increasing the network threshold, the number of connections decreases. This in turn decreases the density, nodal degree variance, and centrality. 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 140–260, it becomes very

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

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 nonexistence 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.

6. Discussion

Using biased truncated Lévy walks with semi-permeable boundaries, we have designed an agent-based model for gang members 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 [10,9]. 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 and shape.

In choosing the metrics for the analysis of the models, we took into account that the observed network may not represent the actual set of rivalries in a given region. If the true rivalry network were only a small perturbation from the observed network, then ideally, we would like the metrics to only vary slightly. For this reason, we evaluated these models with the accuracy and shape metrics presented. The accuracy metrics provide a means to determine whether the simulations are recreating the same rivalries that are observed. The variance of nodal degree and centrality measures indicate the spread in degrees for each gang. These measures could be used to help policy makers know how to distribute resources. For example, if one gang has several rivals and the rest of the gangs only have one, then the resources should be mainly directed toward the one gang that is highly centralized within the network. Alternatively, if all gangs have the same number of rivals, then resources should be spread more equally through the area. The density metric gives an indication as to the connectivity of the network. In terms of gang rivalries, a higher density indicates that a larger number of rivalries are present in the system. Creating a model that produces a network with similar shape metrics is desirable, since the connectivity and degree distributions could be useful for determining intervention strategies and methods for implementation.

We implemented simpler, baseline models so that we could contrast them with the SBLN model. The baseline models did not perform as well as the SBLN, but they did provide some insight into the modeling of gang rivalry networks. The GTG is a simple model designed to compare with the agent-based models. This 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 3.1. The GTG model indicates that geography and size of gangs matter for the rivalry structure. 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. It is not obvious how to extend the modeling framework of a GTG to include policing strategies, the location of violence, retaliatory behavior, and effects of injunctions, unlike agent-based models.

The BMN is a simplified version of the SBLN model. Although the BMN accuracy results were not as strong as the GTG and SBLN results, this method was able to reproduce a similar shape as the observed rivalry network. This model incorporated geographical features, but ignored directional decisions of the agents. The presence of semi-permeable boundaries in this model reduced the number of connections between regions, giving the simulated network a more similar shape to the observed. This suggests that the incorporation of boundaries plays an important role in the replication of the observed network. This is corroborated in [9]. In the absence of the directional decisions of the agents, we see a fundamental difference between the SBLN and the BMN. Without directional decisions the BMN has too many high degree nodes. Another major problem with the BMN model is that the stopping criterion 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 criterion. Unlike the BMN, our proposed SBLN model exhibits long-term stabilization of the accuracy and density metrics. This is a direct consequence of the directional decisions of the agents. The need for incorporating directional decisions into a mobility model is consistent with the literature [58].

The SBLN is the best model in replicating the observed network. 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. This model is admittedly more complex than the other models, however, it provides better results in terms of accuracy and shape of the networks. The stochasticity in the model is beneficial not only to estimating the expected resulting network, but also potential networks that could arise from the seemingly random movements of individuals. One important feature is that this model produces stable long-term behavior, meaning there is a reasonable point at which to stop the number of iterations. This has practical consequences when the true rivalry network is not known. In the absence of an observed network with which to calibrate the parameters of the model, various parameters could be initialized using data when available. There has been some work done on approximating the probability density function for jump lengths in a given region. Using similar methodology as [77], one could approximate the parameters associated with this region. In addition, an analysis of traffic flow under highways and within regions could be used to approximate the boundary cross probability. This would greatly reduce the number of variables of the model and would be an interesting avenue for future study.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040

Our modeling framework allows for a variety of data output, depending on the questions of interest. For example, our 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. Using this violence data would provide an additional data set to use for testing the external validity of our model. The issues of internal and external credibility are discussed in [78]. Fig. 14 shows the locations of the interactions among agents for one of the Ensemble SBLN simulation runs and the density estimation 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.

This highly flexible model provides a framework with which to test sociological theories related to gang activities. One question of great interest to social scientists is the role of territories in the motivation for gang violence [10]. One could encode territories into the model by having each agent place a marker, associated with its gang, on the locations where it has been. By leaving these markings, one could see territories begin to form. Once these territories were established, the behavioral rules could be changed to avoid or attack the territories of gangs, instead of one point in space. From this, social scientists would be able to play out various scenarios and test hypotheses. Another interesting phenomenon observed is the presence of alliances or truce between gangs, as has been observed in Chicago [79] and Los Angeles [10,80]. 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. A known truce between two gangs could be incorporated into the model by flagging that pair of gangs as allies with negative values to the corresponding element of the gang rivalry matrix R. Then, all future interactions between gangs with an alliance would not influence the movement decisions for agents of these gangs. Our model could also be used to examine this question.

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 [10,12]. There are several advantages of approaching this serious problem using a mathematical model. One major strength over a network or statistical approach is that once the model has been sufficiently calibrated, it provides a powerful tool for social scientists to test theories and hypotheses. The paper [16] highlights that correlation does not imply causality, and though statistical analysis provides important contributions to social science as a field, it falls short in terms of identifying underlying mechanisms and testing causation hypotheses. Though one could argue that creating an experiment would help validate or reject causal arguments, such experiments are expensive and may bring up ethical concerns. Finally, if the simulation can accurately model the social phenomena of interest, then we gain insight into how intervention strategies could alter the existing gang rivalry system. The costs of implementing these changes in the simulation are 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.

Acknowledgments

We would like to thank Allon Percus for his useful feedback and Peterson Trethewey for his generous programming advice. We would also like to thank the LAPD, in particular Sean Malinowski, Gabriel Barboza, Arnold Suzukamo, Javier Macias, and the Hollenbeck Gang Division for their time and the enlightening conversations. In our work we used Google Earth™ images. This work was sponsored by NSF grant DMS-0907931, NSF grant DMS-0968309, AFOSR MURI grant FA9550-10-1-0569, ARO grant 58344-MA, ONR grant N000141010221, and ARO MURI grant 50363-MA-MUR.

References

[1] H.C. Covey, Street Gangs Throughout the World, Charles C Thomas, Publisher, 2010.
[2] M.W. Klein, F.M. Weerman, T.P. Thornberry, Street gang violence in Europe, Eur. J. Criminol. 3 (4) (2006) 413–437.
[3] M.W. Klein, H.-J. Kerner, C.L. Maxson, E.G.M. Weitekamp, Euro Gang Paradox: Street Gangs and Youth Groups in the US and Europe, Kluwer Academic Publisher, 2001.
[4] Product No. 2009-M0335-001, National gang threat assessment, Natl. Gang Intell. Cent. (2009).
[5] F.M. Thrasher, The Gang: A Study of 1313 Gangs in Chicago, University of Chicago Press, 1927.
[6] S.H. Decker, B. Van Winkle, Life in the Gang: Family, Friends, and Violence, Cambridge University Press, 1996.
[7] G. Tita, J. Cohen, J. Engberg, An ecological study of the location of gang set space, Soc. Probl. 52 (2) (2005) 272–299.
[8] A.V. Papachristos, Murder by structure: dominance relations and the social structure of gang homicide, Am. J. Sociol. 115 (1) (2009) 74–128.
[9] S. Radil, C. Flint, G. 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) (2010) 307–326.
[10] G. Tita, K. Riley, G. Ridgeway, C. Grammich, A. Abrahamse, P. Greenwood, Reducing gun violence: Results from an intervention in East Los Angeles, Natl. Inst. Justice, RAND (2003).
[11] G. Tita, S. Radil, Making space for theory: the challenges of theorizing space and place for spatial analysis in criminology, J. Quant. Criminol. (2010) 1–13. doi:10.1007/s10940-010-9115-5.
[12] Information Technology Division Management Report Unit. Statistical digest. Los Angeles Police Department, 2008.
[13] M.W. Macy, R. Willer, From factors to actors: computational sociology and agent-based modeling, Annu. Rev. Sociol. 28 (2002) 143–166.
[14] P. Torrens, Geography and computational social science, GeoJournal 75 (2010) 133–148. doi:10.1007/s10708-010-9361-y.
[15] J. Eck, L. Liu, Contrasting simulated and empirical experiments in crime prevention, J. Exp. Criminol. (2008).
[16] M.B. Gordon, A random walk in the literature on criminality: a partial and critical view on some statistical analyses and modelling approaches, European J. Appl. Math. 21 (Special Double Issue 4–5) (2010) 283–306.
[17] M.E.J. Newman, Scientific collaboration networks. I. Network construction and fundamental results, Phys. Rev. E 64 (1) (2001) 016131.

[18] M.E.J. Newman, Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, Phys. Rev. E 64 (1) (2001) 016132.
[19] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[20] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (1) (2002) 47–97.
[21] M.E.J. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E 64 (2) (2001) 026118.
[22] M.E.J. Newman, D.J. Watts, S.H. Strogatz, Random graph models of social networks, Proc. Natl. Acad. Sci. 99 (Suppl. 1) (2002) 2566–2572.
[23] M.T. Gastner, M.E. Newman, The spatial structure of networks, Eur. Phys. J. B 49 (2006) 247–252. doi:10.1140/epjb/e2006-00046-8.
[24] B. Wellman, Are personal communities local? A Dumptarian reconsideration, Soc. Netw. 18 (4) (1996) 347–354.
[25] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, Geographic routing in social networks, Proc. Natl. Acad. Sci. 102 (33) (2005) 11623–11628.
[26] N. Masuda, H. Miwa, N. Konno, Geographical threshold graphs with small-world and scale-free properties, Phys. Rev. E 71 (3) (2005) 036108.
[27] M. Bradonjić, A. Hagberg, A.G. Percus, Giant component and connectivity in geographical threshold graphs, in: Algorithms and Models for the Web-Graph, WAW 2007, San Diego, CA USA, 2007, pp. 209–216.
[28] M. Bradonjić, A. Hagberg, A.G. Percus, The structure of geographical threshold graphs, Internet Math. 5 (2009) 113–140.
[29] J.-J. Tseng, C.-H. Lin, C.-T. Lin, S.-C. Wang, S.-P. Li, Statistical properties of agent-based models in markets with continuous double auction mechanism, Physica A: Stat. Mech. Appl. 389 (8) (2010) 1699–1707.
[30] L. Tesfatsion, Chapter 16 Agent-Based Computational Economics: A Constructive Approach to Economic Theory, in: Handbook of Computational Economics, vol. 2, Elsevier, 2006, pp. 831–880.
[31] A. Wilhite, Chapter 20 Economic Activity on Fixed Networks, in: Handbook of Computational Economics, vol. 2, Elsevier, 2006, pp. 1013–1045.
[32] S. Eubank, H. Guclu, V.S.A. Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, N. Wang, Modelling disease outbreaks in realistic urban social networks, Nature 429 (2004) 180–184.
[33] H.M. Singer, I. Singer, H.J. Herrmann, Agent-based model for friendship in social networks, Phys. Rev. E 80 (2) (2009) 026113.
[34] C.K. Hemelrijk, K. Hanspeter, Density distribution and size sorting in fish schools: an individual-based model, Behav. Ecol. 16 (1) (2004) 178–187.
[35] W. Liu, M.B. Short, Y.E. Taima, A.L. Bertozzi, Multiscale collaborative searching through swarming, in: Proc. 7th Int. Conf. on Inform. Control, Autom., Robot., ICINCO, June 2010.
[36] M.R. D'Orsogna, Y.-L. Chuang, A.L. Bertozzi, L. Chayes, Self-propelled particles with soft-core interactions: patterns, stability, and collapse, Phys. Rev. Lett. 96 (2006) 104302.
[37] J.A. Canizo, J.A. Carrillo, J. Rosado, A well-posedness theory in measures for some kinetic models of collective motion, 2010.
[38] J.D. Farmer, D. Foley, The economy needs agent-based modelling, Nature 460 (2009) 685–686.
[39] F. Schweitzer, Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences, Springer, 2003.
[40] F. Schweitzer, B. Tilch, Self-assembling of networks in an agent-based model, Phys. Rev. E 66 (2) (2002) 026113.
[41] Z. Toroczkai, H. Guclu, Proximity networks and epidemics, Physica A: Stat. Mech. Appl. 378 (1) (2007) 68–75.
[42] S.M. Mniszewski, S.Y. Del Valle, P.D. Stroud, J.M. Riese, S.J. Sydoriak, EpiSims simulation of a multi-component strategy for pandemic influenza, in: SpringSim'08: Proceedings of the 2008 Spring Simulation Multiconference, Society for Computer Simulation International, San Diego, CA, USA, 2008, pp. 556–563.
[43] A. Barbaro, B. Einarsson, B. Birnir, S. Sigurdsson, H. Valdimarsson, O.K. Palsson, S. Sveinbjornsson, T. Sigurdsson, Modelling and simulations of the migration of pelagic fish, ICES J. Mar. Sci. 66 (5) (2009) 826–838.
[44] M.B. Short, P.J. Brantingham, A.L. Bertozzi, G.E. Tita, Dissipation and displacement of hotspots in reaction–diffusion models of crime, Proc. Natl. Acad. Sci. 107 (9) (2010) 3961–3965.
[45] M.B. Short, M.R. D'Orsogna, V.B. Pasour, G.E. Tita, P.J. Brantingham, A.L. Bertozzi, L.B. Chayes, A statistical model of criminal behavior, Math. Models Methods Appl. Sci. 18 (Suppl.) (2008) 1249–1267.
[46] P.J. Brantingham, G. Tita, Offender mobility and crime pattern formation from first principles, in: L. Liu, J. Eck (Eds.), Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems, Idea Press, 2008, pp. 193–208.
[47] P.A. Jones, P.J. Brantingham, L.R. Chayes, Statistical models of criminal behavior: the effects of law enforcement actions, 2010.
[48] A.B. Pitcher, Adding police to a mathematical model of burglary, European J. Appl. Math. 21 (Special Double Issue 4–5) (2010) 401–419.
[49] M.B. Short, P.J. Brantingham, M.R. D'Orsogna, Cooperation and punishment in an adversarial game: how defectors pave the way to a peaceful society, Phys. Rev. E 82 (6) (2010) 066114.
[50] M. O'Leary, The mathematics of geographic profiling, J. Invest. Psychol. Offender Profil. 6 (2009) 253–265.
[51] G. Mohler, Gang rivalry dynamics via coupled point process networks. Unpublished results.
[52] M. Egesdal, C. Fathauer, K. Louie, J. Neuman, Statistical modeling of gang violence in Los Angeles. SIAM Undergrad. Res. (2010) (Online).
[53] E. Anderson, Code of the street: decency, violence, and the moral life of the inner city, 2000.
[54] D. Ley, R. Cybriwsky, Urban graffiti as territorial markers, Ann. Assoc. Am. Geogr. 64 (4) (1974) 491–505.
[55] E.R. Groff, Situating simulation to model human spatio-temporal interactions: an example using crime events, Trans. GIS 11 (4) (2007) 507–530.
[56] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (2006) 462–465.
[57] I. Rhee, M. Shin, S. Hong, K. Lee, S. Chong, On the Lévy-walk nature of human mobility: do humans walk like monkeys? in: IEEE Infocom 2008—IEEE Conference on Computer Communications, IEEE, 2008, pp. 924–932.
[58] M. González, C.A. Hildalgo, A.-L. Barabási, Understanding individual human mobility patterns, Nat. Lett. 453 (2008) 779–782.
[59] K.V. Mardia, P.E. Jupp, Directional Statistics, John Wiley & Sons Ltd., 2000.
[60] S.R. Jammalamadaka, A. SenGupta, Topics in Circular Statistics, World Scientific Publishing Co. Pvt. Ltd., 2001.
[61] D.J. Best, N.I. Fisher, Efficient simulation of the von Mises distribution, J. R. Stat. Soc. Ser. C (Appl. Stat.) 28 (2) (1979) 152–157.
[62] L. Smith, M. Keegan, T. Wittman, G. Mohler, A. Bertozzi, Improving density estimation by incorporating spatial information, in: EURASIP Adv. Signal Process: spec. issue Adv. Image Process. Def. Secur. Appl., 2010.
[63] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC, 1986.
[64] P.P.B. Eggermont, V.N. LaRiccia, Maximum Penalized Likelihood Estimation, Springer, 2001.
[65] G.O. Mohler, A.L. Bertozzi, T.A. Goldstein, S.J. Osher, Fast TV regularization for 2D maximum penalized likelihood estimation, J. Comput. Graph. Stat. 20 (2) (2011) 479–491.
[66] I.J. Goodd, R.A. Gaskins, Nonparametric roughness penalties for probability densities, Biometrika 58 (2) (1971) 255–277.
[67] J. Moore, R. Pinderhughes (Eds.), In the Barrios: Latinos and the Underclass Debate, Russel Sage Foundation, 1993.
[68] M.E.J. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. 98 (2) (2001) 404–409.
[69] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, 2009.
[70] L.H. Wong, P. Pattison, G. Robins, A spatial model for social networks, Physica A: Stat. Mech. Appl. 360 (1) (2006) 99–120.
[71] M. Tomassini, E. Pestelacci, L. Luthi, Mutual trust and cooperation in the evolutionary Hawks–Doves game, Biosyst. 99 (1) (2010) 50–59.
[72] P. Baldi, S. Brunak, Y. Chauvin, C.A. Andersen, H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview, Bioinform. Rev. 16 (5) (2000) 412–424.
[73] F. Sebastiani, Machine learning in automated text categorization, ACM Comput. Surv. 34 (1) (2002) 1–47.
[74] Y. Yang, X. 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, ACM, New York, NY, USA, 1999, pp. 42–49.
[75] B.W. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochem. Biophys. Acta 405 (2) (1975) 442–451.
[76] L.C. Freeman, Centrality in social networks conceptual clarification, Soc. Netw. 1 (3) (1978–1979) 215–239.
[77] B. Jiang, J. Yin, S. Zhao, Characterizing the human mobility pattern in a large street network, Phys. Rev. E (2009).
[78] R. Berk, How you can tell if the simulations in computational criminology are any good, J. Exp. Criminol. 4 (2008) 289–308. doi:10.1007/s11292-008-9053-5.
[79] C.R. Block, R. Block, Street gang crime in Chicago, Technical report, December 1993.
[80] J. Moore, D. Vigil, R. Garcia, Residence and territoriality in chicano gangs, Soc. Probl. 31 (2) (1983) 182–194.

Please cite this article in press as: R.A. Hegemann, et al., Geographical influences of an emerging network of gang rivalries, Physica A (2011), doi:10.1016/j.physa.2011.05.040