THE MATHEMATICS OF POLICING

TIAN AN WONG

Abstract. This article surveys the mathematics of policing. The first part concerns the mathematics of crime, begins with the PredPol model, then continues on to the many developments following it including co-offending networks and epidemiological criminology, discussing critiques of these approaches where studies have been done. The second part turns to mathematical studies of policing itself, in particular Early Intervention Systems and network models of police misconduct. We propose problems for future research throughout the paper.

Contents

Date: June 28, 2022.
2010 *Mathematics Subject Classification.* 91D30, 91D10, 91C20, 92D30.
Key words and phrases. Predictive policing, co-offending network, co-complaint network, social contagion, early intervention system.

1. Introduction

1.1. **Predictive policing and beyond.** What is predictive pollicing? The RAND Corporation defines predictive policing as "the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions." This paper serves as a mathematical supplement to the recent coverage of predictive policing in Nature [41, 7] and elsewhere [47, 31], and more importantly the boycott by several thousand mathematicians of colaborations with the police appearing in the Notices of the AMS [2], citing for example a 2016 ICERM workshop on predictive policing. But which mathematicians collaborate with police? As as survey of the literature shows, only a handful of primarily statisticians and applied mathematicians do. (Many more number theorists, on the other hand, collaborate with the NSA without much objection.) Significant contributions to the development of predictive policing are made from other fields, such as criminology, sociology, computer science, law, even anthropology and business. Nonetheless, the foundations of predictive policing are indeed mathematical, though modern usages also heavily incorporate machine learning and such methods to optimize models, and coded as algorithms for implementation.

The primary purpose of this article is to present a selective overview of the mathematics underlying various predictive policing models, along with some mathematical approaches to studying police themselves. Being a selective approach to the topic, aimed at mathematicians, it is not intended to be comprehensive and rather mainly concerned with the mathematical aspects, and avoids larger discussions about policing in general and more philosophical questions. We shall at time enter into the weeds so as to open up the black box as much as possible. The interested reader is encouraged to explore surveys and discussions from different points of view, such as statistical physics [12], social science [30], legal [15], and algorithmic injustice [36]. Note that the language use in this paper, particularly that surrounding the language of crime and such is at times taken from the articles in question.

1.2. **Policing the police.** The study of crime leads to the natural question: how can we study police instead? The literature review provided here reveals not only the lack of research on policing, but the lack of involvement of mathematicians in actively studying policing, whether as an antiracist practice or otherwise. In the few studies that do exist, nonetheless, we can see the significant applicability of the mathematical tools used to study crime and criminal networks to policing and police networks, though it is also crucial to study the differences, where such models cannot be easily repurposed. Moreover, the blind spots and biases in models of crime must also be considered when attempting to apply them to police data. In the hopes of inspiring further research, throughout the paper we pose various problems — some more well defined than others — as possible avenues of research.

To be certain, that the focus on this paper is on the mathematics should not obscure the connections to questions about policing at large. In particular, taking as axiomatic that mathematics is not neutral, not least as it relates to questions of politics and policing, an important corollary is that how we think about such matters influences the research questions that are pursued. Similar meta-questions also underlie the recent boom in studies on the mathematics of gerrymandering, or what is now called political geometry [11]. The exciting recent developments in the latter field offer promise for the mathematical problems of independent interest, both applied and otherwise, that can emerge from a deeper study of the mathematics of policing. One underlying motivation for this paper is the sentiment that if mathematicians have collaborated with policing, which has in various ways shown to be biased, in what ways can mathematicians work to say, de-bias policing, provide evaluative benchmarks, or offer quantitative critiques of carceral systems at large? From a policy perspective, the phenomenon of police violence can be and is framed as a public health concern [49]. Indeed, the World Health Organization (WHO) classifies police excessive use of force as a form of violence, which in general is considered a public health issue. Moreover, given the mathematics to criminalization pipeline, one might argue for the moral responsibility of the mathematical community towards the applications of mathematics.

1.3. **Summary.** In Part 1, we survey the mathematics of predictive policing, beginning with the classic PredPol in Section 2 and review its critiques and refinements in Section 3. This model of predictive policing is known as *place-based* or *hotspot* policing, whereby it is geographic locations that are assigned risk values depending on various factors such as arrests. Later studies show that this model can be biased against racial minorities and lower-income communities, and moreover leads to negative feedback loops. The incorporation of environmental factors and sociological data forms, typically incorporated through machine learning, forms an additional layer and is known as Risk Terrain Modelling (RTM). While PredPol (now rebranded as Geolitica) is the most well-known placed-based prediction system, it is far from the only one; others such as HunchLab, ShotSpotter, Oracle, Microsoft, Accenture, IBM, Hitachi, Motorola, and Palantir all have predictive policing products.

In Section 4, we turn to *people-based* policing, whereby it is individuals who are studied rather than geography. The core idea here is the use of a co-offence network, where individuals who share an offence are used to form a network, by which predictions and recommendations are made. A notorious example of this is the Chicago Police Department's Strategic Subjects List (SSL), colloquially known as the "heat list." The final Section 5 of Part 1 touches on the mathematics of gun violence, a particular subset of the mathematics of crime, having much overlap with the earlier sections. We discuss the idea of epidemiological criminology, and the mathematical frameworks that are used.

In Part 2, we turn to the mathematics of police themselves. Here, the literature is much more sparse, and much work remains to be done. It is hoped that this survey will inspire further research into the topic. In Section 6, we discuss the study of police misconduct, which has received the most attention to date. Of particular note is the beginnings of network models that study officer co-complaints in a manner parallel to criminal co-offending networks. In Section 7, we conclude with various topics not covered in depth here, including related statistical work (as opposed to

primarily mathematical work, though the distinctions between the two here are not strict) that is done in this area, and pose questions for further consideration.

Part 1. Predictive Policing

2. Placed-based prediction: The PredPol model

The company PredPol is often singled out in discussions surrounding predictive policing, in part because it might be argued as being the progenitor of the field, though it is by no means the only game in town nor the state of the art. What is the mathematics underlying PredPol? According to Vice reporting, PredPol was the most widely used predictive policing company/software in US as of 2019, having contracts including in Utah, California, and Washington (in particular the University of California, Berkeley). It was born out of a collaboration with the LAPD, the FBI, and UCLA.

In this section, we discuss two models: the epidemic type aftershock (ETAS) model that the PredPol patent is built on [32], and the reaction-diffusion model that ETAS in turn draws upon. While reaction-diffusion equations can be considered to belong to the realm of partial differential equations, with applications to the physical sciences, the ETAS model is the most popular statistical model for earthquake occurrence. PredPol is based on equations that are used to model earthquake occurrences and chemical reactions which produce certain "hotspots," and where adding "hotspot policing" can be viewed as an inhibitor to the process. From a statistical point of view, the introduction of police then to given hotspots are then viewed as "treatments" or interventions which are meant to reduce crime. Such models are also referred to as spatio-temporal models.

2.1. The reaction-diffusion model.

We review the mathematics underlying [43, 44]. Reaction-diffusion models are typically used to describe chemical reactions, in which activators and inhibitors move, mix, and interact. In [43], the model describes houses and burglars, while [44] describes "motivated offenders" and targets or victims as activators and law enforcement as inhibitors. Their reaction-diffusion system involves "mobile criminal offenders" within a square environment with periodic boundary conditions.

2.1.1. *Discrete model.* In the discrete model, houses are placed on a lattice in the plane with constant spacing. Within the plane $s = (x, y)$, the score in question $A_s(t)$ is interpreted either as the attractiveness of a house to a burglar, or the *risk of victimization*, "representing general environmental cues about the feasibility of committing a successful crime and/or specific knowledge offenders possess about target or victim vulnerability in the area." The risk is given by

$$A_s(t) = A_s^0 + B_s(t),$$

where A_s^0 is a fixed value and $B_s(t)$ is a dynamic value that models the idea that if a site has been attacked, it has a higher risk of being re-vicitimized shortly after the first incident. Its first approximation is

$$B_s(t + \delta t) = B_s(t)(1 - \omega\delta t) + \theta E_s(t), \tag{1}$$

where ω sets a time scale over which repeat victimizations are most likely to occur, and θ is a multiplier of $E_s(t)$, the number of burglary events that occurred at site s since time t. The authors then modify this model to account for "near-repeat

victimization, and the broken windows effect", by allowing the quantity $B_s(t)$ to spread spatially to its neighbors. Equation (1) is replaced with

$$B_s(t+\delta t) = \left[B_s(t) + \frac{\eta \ell^2}{z}\Delta B_s(t)\right](1-\omega\delta t) + \theta E_s(t),$$

where Δ is the discrete Laplacian, whereby

$$\Delta B_s(t) = \frac{1}{\ell^2}\left(\sum_{s'\sim s} B_{s'}(t) - zB_s(t)\right).$$

Here z is the number of sites s' neighboring s, and $0 \leq \eta \leq 1$ measures the significance of neighborhood effects. Computer simulations are then run to show that the model produces certain dynamic and stationary 'hotspots.'

The "criminal agent" is modeled as either committing a crime at site s or moving to a neighboring location based on a biased random walk so that site s' is visited with probability

$$q_{s\to s'}(t) = \frac{A_{s'}(t)}{\sum_{s'\sim s} A_{s'}(t)}.$$

The probability of occurrence for each burglar located at site s between times t and $t+\delta t$ given by

$$p_s(t) = 1 - e^{-A_s(t)\delta t}$$

in accordance with a standard Poisson process in which the expected number of events during the time interval of length δt is $A_s(t)\delta t$. In the discrete model, burglars are removed after committing a crime, and regenerated at each lattice site at a constant rate Γ. We write

$$E_s(t) = n_s(t)p_s(t),$$

where $n_s(t)$ is the number of criminals at the site s at time t.

2.1.2. *Continuous model.* From the discrete model we form the difference quotient,

$$\frac{B_s(t+\delta t) - B(t)}{\delta t}$$

and take the limit as ℓ and δt approach 0 to arrive at the differential equation[1]

$$\frac{\partial B}{\partial t} = \frac{\eta D}{z}\nabla^2 B - \omega B + \epsilon D\rho A. \tag{2}$$

Here we have denoted

$$D = \frac{\ell^2}{\delta t}, \quad \epsilon = \theta\delta t, \quad \rho(s,t) = \frac{n_s(t)}{\ell^2},$$

and ρ is the density of criminal agents

$$\frac{\partial \rho}{\partial t} = \frac{D}{z}\nabla\cdot\left[\nabla\rho - \frac{2\rho}{A}\nabla A\right] - \rho A + \gamma, \tag{3}$$

where offenders exit the system at the rate ρA and are reintroduced at the constant rate $\gamma = \Gamma/\ell^2$. The PDE for ρ is obtained by a difference quotient for $n_s(t)$, using the equation

$$n_s(t+\delta t) = A_s \sum_{s'\sim s} \frac{n_{s'}(t)(1-p_{s'}(t))}{T_{s'}(t)} + \Gamma\delta t,$$

[1]This is essentially the patented algorithm displayed in https://www.predpol.com/technology/

where

$$T_{s'}(t) = \sum_{s''\sim s'} A_{s''}(t),$$

which simply means that any agents that are present at s after one time step must have either arrived from a neighboring site after having not committed a crime there, or have been generated at s at rate Γ. The coupled differential equations (2) and (3) thus describe the continuous model.

In [44] the authors study these coupled PDEs to show that crime risk will form dense, well-spaced hotspots when the diffusion of risk by individual crimes is spatially broad enough. Police suppression is modeled by instantaneously setting the crime rate $\rho A_s(t) = 0$ at the locations of current crime hotspots and maintaining this suppression for a fixed time period. The authors then claim that subcritical crime hotspots may be permanently eradicated with police suppression.

2.2. **Epidemic-type aftershock (ETAS).** This section covers the mathematics underlying [33], which is the basis for the PredPol patent [32]. The authors treat the dynamic occurrence of crime as a continuous time, discrete space epidemic-type aftershock sequence point process. In seismology, point processes are used by considering a "parent earthquake" and subsequent background events or aftershocks. The ETAS model estimates long term and short term hotspots and systematically estimates the relative contribution to risk of each via a Expectation-Maximization (EM) algorithm.

The ETAS model can then be intuitively understood as a branching process: first generation events occur according to a Poisson process with constant rate μ, then events (from all generations) each give birth to N direct offspring events, where N is a Poisson random variable with parameter θ. As events occur, the rate of crime increases locally in space, leading to a contagious sequence of "aftershock" crimes that eventually dies out on its own, or is interrupted by police intervention.

In this model, policing areas are discretized into square boxes. The probabilistic rate of events in box n at time t is defined to be

$$\lambda_n(t) = \mu_n + \sum_{t_n^i < t} \theta\omega e^{-\omega(t-t_n^i)}, \tag{4}$$

where t_n^i are the times of events in box n in the history of the process. The background rate μ is a (nonparametric histogram) estimate of a stationary Poisson process.

The expectation, or E-step, sets

$$p_n^{ij} = \frac{\theta\omega e^{-\omega(t_n^i - t_n^j)}}{\lambda_n(t_n^j)}, \qquad p_n^j = \frac{\mu_n}{\lambda_n(t_n^j)},$$

where $\theta\omega e^{-\omega t}$ is called the triggering kernel that models "near-repeat" or "contagion" effects in crime data.

The maximization, or M-step, sets

$$\omega = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_{i<j} p_n^{ij}(t_n^j - t_n^i)},$$
$$\theta = \frac{\sum_n \sum_{i<j} p^{ij}}{\sum_n \sum_j 1},$$
$$\mu = \frac{\sum_n \sum_j p_n^j}{T},$$

where T is the length of the time window of observation.

2.2.1. *Rederivation après Lum.* Based on Kristian Lum's note [27], we discuss the derivation of the EM-algorithm above inferred from [33], given in detail because it is not explicated by the latter. The general reader is advised to skip this section on first reading, and note that the notation varies slightly from the above, but can be easily seen to correspond. Let

$$X = \{t_i^{(n)}, n = 1, \ldots, k, i = 1, \ldots, N_n\},$$

where $t_i^{(n)}$ is the time of the i-th crime that occurs in location n, and N_n is the total number of crimes that occur in bin n during the period of observation. We also define latent indicator variables $Z_{ij}^{(n)}$, which denote the provenance of the j-th crime. Namely,

$$Z_{0j}^{(n)} = \begin{cases} 1, & \text{if the } j\text{-th crime arose from the baseline process} \\ 0, & \text{otherwise,} \end{cases}$$
$$Z_{ij}^{(n)} = \begin{cases} 1, & \text{if the } i\text{-th crime arose is the parent of the } j\text{-th crime} \\ 0, & \text{otherwise.} \end{cases}$$

The number of crimes arising from the baseline process in location n over a time period of duration T is distributed as Poisson with rate $\mu_n T$. Define

$$\sum_{j=1}^{N_n} Z_{0j}^{(n)} \sim \text{Poisson}(\mu_n T),$$

and the contribution to the full Z-augmented log-likelihood of the events arising from the baseline process is given by

$$l_{\text{base}}(X) = \sum_{n=1}^{k} \left(-\mu_n T + \sum_{j=1}^{N_n} Z_{0j}^{(n)} \log(\mu_n T) - \log\left(\left(\sum_{j=1}^{N_n} Z_{0j}^{(n)} \right)! \right) \right).$$

A crime occurring at time t in bin n, regardless of how it arose, gives rise to c_t child events in bin n, where $c_t \sim \text{Pois}(\theta)$. The time until each of the child events are random draws from an exponential distribution with rate parameter ω, written as

$$p\left(t_j^{(n)} | t_i^{(n)}, Z_{ij}^{(n)} = 1\right) \sim \text{Exp}(\omega),$$

and

$$\sum_{j: t_i^{(n)} < t_j^{(n)}} Z_{ij}^{(n)} \sim \text{Poisson}(\theta),$$

where by definition the lefthand sum is is the number of child events of crime i from location n. The contribution to the log-likelihood of the child events is given by

$$l_{\text{child}}(X) = l_{\text{exp}}(X) + l_{\text{pois}}(X),$$

where

$$l_{\text{exp}}(X) = \sum_{n=1}^{k} \sum_{i=1}^{N_n} \sum_{j:t_i^{(n)} < t_j^{(n)}} Z_{ij}^{(n)} (\log \omega - \omega(t_i^{(n)} - t_j^{(n)})),$$

$$l_{\text{pois}}(X) = \sum_{n=1}^{k} \sum_{i=1}^{N_n} \left(-\theta + \sum_{j:t_i^{(n)} < t_j^{(n)}} Z_{ij}^{(n)} \log(\theta) - \log\left(\left(\sum_{j:t_i^{(n)} < t_j^{(n)}} Z_{ij}^{(n)} \right)! \right) \right).$$

Altogether, the full augmented log-likelihood is given by

$$l(X; \theta, \omega, \mu) = l_{\text{base}}(X) + l_{\text{child}}(X),$$

and under this model, (4) gives the expected rate at time t and location n, given all other previously observed crimes in location n.

The expectation is calculated as

$$p_{ij}^{(n)} = E[Z_{ij}^{(n)} | X, \theta, \omega, \mu_n],$$

with parameter values given at each iteration step. In order to derive the EM algorithm associated with the given log-likelihood, we must compute

$$Q(\theta, \omega, \mu) = E_{Z|X,\theta,\omega,\mu_n} l(X; \theta, \omega, \mu).$$

To do so, we substitute $Z_{ij}^{(n)}$ with $p_{ij}^{(n)}$ in terms that are linear in $Z_{ij}^{(n)}$, and write C_1, C_2 for constants that do not depend on μ_n, ω, or θ, hence

$$E_{Z|X,\theta,\omega,\mu_n} l_{\text{base}}(X) = \sum_{n=1}^{k} \left(-\mu_n T + \sum_{j=1}^{N_n} p_{0j}^{(n)} \log(\mu_n T) \right) - C_1,$$

$$E_{Z|X,\theta,\omega,\mu_n} l_{\text{exp}}(X) = \sum_{n=1}^{k} \sum_{i=1}^{N_n} \sum_{j:t_i^{(n)} < t_j^{(n)}} p_{ij}^{(n)} (\log \omega - \omega(t_i^{(n)} - t_j^{(n)})),$$

$$E_{Z|X,\theta,\omega,\mu_n} l_{\text{pois}}(X) = \sum_{n=1}^{k} \sum_{i=1}^{N_n} \left(-\theta + \sum_{j:t_i^{(n)} < t_j^{(n)}} p_{ij}^{(n)} \log(\theta) \right) - C_2.$$

Finally, in order to update the parameters θ, ω, μ, we optimize Q with respect to each variable by taking partial derivatives and solving. That is,

$$\frac{\partial Q}{\partial \mu_n} = -T + \frac{1}{\mu_n} \sum_{j=1}^{N_n} p_{0j}^{(n)},$$

$$\frac{\partial Q}{\partial \omega} = \sum_{n=1}^{k} \sum_{i=1}^{N_n} \sum_{j: t_i^{(n)} < t_j^{(n)}} p_{ij}^{(n)} \left(\frac{1}{\omega} - (t_i^{(n)} - t_j^{(n)}) \right),$$

$$\frac{\partial Q}{\partial \theta} = \sum_{n=1}^{k} \sum_{i=1}^{N_n} \left(-1 + \frac{1}{\theta} \sum_{j: t_i^{(n)} < t_j^{(n)}} p_{ij}^{(n)} \right).$$

Setting each equation equal to zero and solving then gives the updated parameters as in the M-step above.

3. Critiques and refinements of PredPol

3.1. **Statistical bias using a synthetic population.** The first notable quantitative study of the PredPol system is the study of Lum and Isaac [28], which simulates a synthetic population in Oakland, CA based upon census data, apply a model based on data from the 2011 National Survey on Drug Use and Health (NSDUH) in order to predict an individual's probability of drug use within the past month based on their demographic characteristics. The resulting data set acts as a replacement for the "ground truth" of drug crime use data, giving estimates of illicit drug use from a non-criminal justice, population-based data source. Compared with police records, the authors find that drug crimes known to police are not a representative sample of all drug crimes. This disparity is visualized in Figure 1.

Applying their reconstruction of the PredPol algorithm as outlined above, the authors conclude that rather than correcting for the apparent biases in the police data, the model reinforces these biases, suggesting that predictive policing of drug crimes results in increasingly disproportionate policing of historically over-policed communities. In particular, this is despite PredPol's claim that ""only three data points in making predictions: past type of crime, place of crime and time of crime. It uses no personal information about individuals or groups of individuals, eliminating any personal liberties and profiling concerns." See also [18] for a critique of other assumptions in the PredPol model.

Problem 3.1. Which explicit and implicit assumptions in predictive policing models are problematic and why?

3.2. **Runaway feedback loops via a generalized Pólya urn model.** Ensign et al. [14] model develop an urn model of predictive policing showing that runaway feedback loop occurs in predictive policing. This is a general problem of what is called traditional batch-mode machine learning, and the theory of urns is a common frame-work in reinforcement learning.

In the generalized Pólya urn model, an urn contains balls of two colors, say red and black. At each time step, a ball is drawn, and based on its color a number of balls are replaced. If red, we add a red and b black balls; and if black we add c red

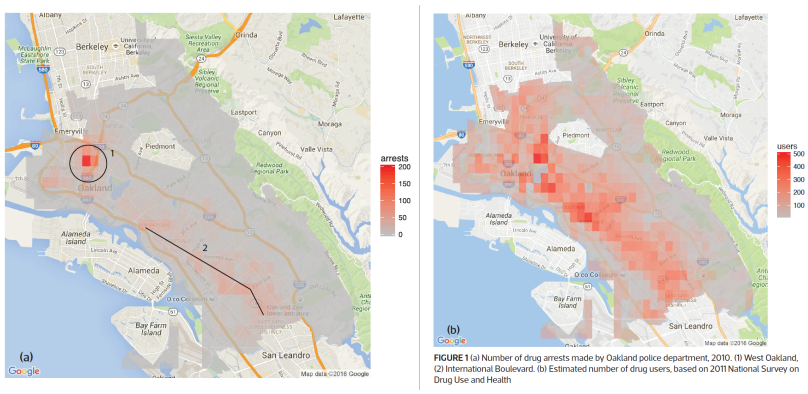

FIGURE 1. Comparison of PredPol predictions versus NSDUH predictions [27, Figure 1].

and d black balls. This is represented by the replacement matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where the standard case is when $a = d = 1$ and $c = b = 0$. As a toy model for predictive policing, A and B are two policing blocks, and the goal is to distribute police officers according to the proportion of crime in each area. Let d_A be the rate at which police in A discover crimes, r_A the rate at which crimes are reported in A, and w_d, w_r the respective weights such that $w_d + w_r = 1$ and $w_d d_A + w_r r_A$ represents the total rate of incident data from A.

The following assumptions are made:

(0) (Effective policing) A region with x percent of total crime in the precinct should receive x percent of policing.
(1) (Predictive model) The officer tosses a coin based on current statistics to decide where to go next. This means that the model uses some form of statistical information to make predictions on crime.
(1) (Context) The only information retained about a crime is a count.
(2) (Truth in crime data) If an officer goes to a location A with an underlying ground truth crime rate of λ_A, the officer discovers crime at a rate of $d_A = \lambda_A$. Reported incidents are also reported at a rate that matches the underlying ground truth crime rate, i.e., $r_A = \lambda_A$.
(3) (Discovery only) Incident data is only collected by an officer's presence in a neighborhood, i.e., $w_d = 1$ and $w_r = 0$.
(4) (Non/uniformity of crime rate) A visit to area A has probability λ_A of encountering a crime, and a visit to area B has probability λ_B of encountering a crime. If $\lambda_A = \lambda_B$, we say the crime rate is uniform.

3.2.1. *Uniform crime.* First assume that crime rate is uniform, so $\lambda = \lambda_A = \lambda_B$. In this case, sending police to area A or B is given by drawing a red or black ball. Then flipping a coin with probability λ that if heads, in which case we simulate one step of the standard Pólya urn, and if tails, we simply the replace the ball that was drawn. In this case, the probability of drawing a red ball has a limiting distribution equal to the Beta distribution that only depends on the initial number of red and

black balls (c.f. [39]). Recall that the Beta distribution is the distribution on $[0,1]$ given by probability distribution function

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}x^{\alpha-1}(1-x)^{\beta-1},$$

where $\Gamma(x)$ is the usual gamma function, and here $\alpha = n_A$, the number of red balls and $\beta = n_B$, the number of black balls initially in the urn. In particular, this means that the long-term probability of visiting an area is a random draw based on this initial data, and does not learn that the crime rates are the same.

3.2.2. *Nonuniform crime.* Now we drop the uniformity assumption. The urn is now modelled by the stochastic addition matrix

$$\begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix},$$

where X_A is a Bernoulli variable taking the value 1 with probability λ_A and 0 with probability $1-\lambda_A$, and similarly X_B. Let $n_A^{(t)}, n_B^{(t)}$ be the number of red and black balls respectively at time t. The probability of adding any ball to the urn is given by

$$P(\text{adding a ball}) = \frac{n_A^{(t)}\lambda_A + n_B^{(t)}\lambda_B}{n_A^{(t)} + n_B^{(t)}},$$

while the probability of adding a red ball conditioned on adding any ball, is given by

$$\frac{P(\text{adding a red ball})}{P(\text{adding a ball})} = \frac{n_A^{(t)}\lambda_A}{n_A^{(t)}\lambda_A + n_B^{(t)}\lambda_B}, \tag{5}$$

and similarly for a black ball. In particular, this is the same as the deterministic Pólya urn in which an i-colored ball is sampled, replaced and then add in λ_i more balls of the same color. Thus the stochastic matrix reduces to

$$\begin{pmatrix} \lambda_A & 0 \\ 0 & \lambda_B \end{pmatrix}.$$

Using a result of Renlund, the runaway feedback loop of this toy model can then be shown.

Proposition 3.2 ([14, Lemma 4])**.** *the asymptotic probability of sampling a red ball is* 1 *if* $\lambda_A > \lambda_B$ *and* 0 *if* $\lambda_A < \lambda_B$*.*

This tells us that as long as A has a ground truth crime rate that is even a little higher than that of B, the update process will lead to police being eventually completely sent to A.

3.2.3. *Incorporating feedback.* In order to learn the crime rate, the Pólya urn should contain balls in proportion to the relative probability of crime occurrence. The following update rule guarantees that the urn proportion will converge to the ratio of replacement (i.e. crime) rates.

Consider the probabilities λ_A and λ_B now conditioned on a ball of the respective color having been sampled. This makes the probability of adding a red ball equal to

$$\frac{n_A^{(t)}\lambda_A}{n_A^{(t)} + n_B^{(t)}}$$

rather than λ, and the expected fraction of red balls being added to the urn after one step of the process equal to (5) instead of $\lambda_A/(\lambda_A+\lambda_B)$.

The following change is now introduced: instead of always adding the new balls, we first sample another ball from the urn, and only add the new balls if the colors are different. This makes the probability of adding a red or black ball

$$\frac{n_A^{(t)}\lambda_A}{n_A^{(t)}+n_B^{(t)}}\frac{n_B^{(t)}}{n_A^{(t)}+n_B^{(t)}} \quad \text{or} \quad \frac{n_B^{(t)}\lambda_B}{n_A^{(t)}+n_B^{(t)}}\frac{n_A^{(t)}}{n_A^{(t)}+n_B^{(t)}}$$

respectively, where we see that the probabilities are proportional to λ_A, λ_B up to the common factor $n_A^{(t)}n_B^{(t)}/(n_A^{(t)}+n_B^{(t)})^2$. This is an example of rejection sampling, where sampled values are dropped according to some probability scheme to affect the statistic collected. Another related scheme is importance sampling, where balls are weighted inversely proportional to the rate at which police are sent.

3.2.4. *Reported indidents.* It is also possible to remove Assumption 3, so that w_d and w_r can take on different values. That is, we allow both discovered and reported incidents to be used as input to the urn model, as is more usually the case in predictive policing systems. In this case, the total weight of incidents from A would be $w_d d_A + w_r r_A$ if it were visited and $w_r r_A$ otherwise. This yields the following urn replacement matrix

$$\begin{pmatrix} w_d d_A + w_r r_A & w_r r_B \\ w_r r_A & w_d d_B + w_r r_B \end{pmatrix}.$$

It is then possible to study the limiting fraction of balls in the urn and account for feedback again, which we refer to [14, §3.4–3.5] for details.

Crucially, the study [14] retains the Assumption 2, namely, that that reported and discovered incident rates track the true crime rates, an assumption that is empirically false in general.

Problem 3.3. Extend the analysis of [14] without Assumption 2.

3.3. **A fairness penalty for PredPol.** In response to criticism of PredPol, Mohler introduced a fairness penalty for the original model. As a variation on [32], following [34], the parameters of λ_n in (4) can be estimated by maximizing the log-likelihood function

$$L(a,\omega,\theta) = \sum_{i=1}^{N} \log(\lambda_{n_i}(t_i)) - \sum_{n\in G}\int_0^T \lambda_n(t)dt \tag{6}$$

where n_i is the grid cell event i and the integral is taken over the window of observation $[0,T]$. The Hawkes process intensity process λ_n is used in practice by ranking all grid cells n at a given time t and then directing police patrols to the top k grid cells, called hotspots. We will denote the set of grid cells comprising the top k hotspots at time t as K_t. We will refer to λ_n as neutral if it is estimated by maximizing (6).

Let $m = 1,\dots,M$ be an indexing set for the number of subgroups (e.g., racial groups) surveyed, and let z_n^m be the population count of racial group m in grid cell n. Assuming that each of the grid cells in K_t receives the same amount of patrol, then the amount of patrol a particular racial group receives, per individual of that

group per day,

$$p_m = \frac{\sum_{t=1}^{T} \sum_{n \in K_t} z_n^m}{T \sum_{n \in G} z_n^m}.$$

The time interval over which hotspots are defined can be taken to be days, so that t is a discrete sum over days. We then define a measure of fairness by comparing the patrol statistics p_m between pairs of groups

$$F(a, \omega, \theta) = \sum_{m > m'} (p_m - p_{m'})^2,$$

so that when $F = 0$, each group m receives the same amount of patrol per individual.

We then add F to the log-likelihood (6) as a fairness penalty and maximize with respect to a, ω, θ,

$$L(a, \omega, \theta) - \chi F,$$

where $\chi \in \mathbb{R}$ is a penalty parameter that controls the balance between accuracy and fairness in the point process model. Note that this latter loss function is non-differentiable due to the sort and threshold required for defining the top k hotspots K_t for each day t.

Problem 3.4. What other fairness penalties (or other corrections) might be used to mitigate bias?

3.4. **Other methods.** To what extent is the spatio-temporal modeling of crime reasonable? Several studies address the appropriateness of place-based interventions and other factors lending to the risk of certain crimes occuring.

3.4.1. *Spatial clustering.* To detect homicide clusters in Newark NJ, Zeoli et al. scanned an ellipse window across the centroids of all census tracts ($n = 90$), calculating the number of observed and expected homicides inside the window at each location [51]. The expected number of cases was calculated as

$$E[h] = p \times \frac{H}{P}$$

where h is the observed number of homicides, p the population in the census tract, and H and P the total numbers of homicides and population, respectively. The relative risk of homicide for each census tract month by dividing the number of observed homicides by the expected number of homicides; the alternative hypothesis was that the risk of homicide inside the scanning window was higher than that outside the scanning window. Under a Poisson assumption, the likelihood function for a specific window was proportional to

$$\left(\frac{h}{E[h]}\right)^e \left(\frac{H-h}{H-E[h]}\right)^{c-e} I,$$

where H is the total number of homicides, h is the observed number of homicides within the window, and $E[h]$ is the expected number of homicides within the window under the null hypothesis of there being no difference. Also, $I = 1$ when the window has more cases than expected under the null hypothesis and 0 otherwise. As a result, the authors found significant clusters for nonintimate familial, escalating dispute, revenge, and drug-motivated homicides, but not for intimate partner homicide and robbery homicide.

3.4.2. *Risk Terrain Modelling.* RTM is a geospatial crime analysis tool that is designed to examine environmental risk factors associated with crime and to identify the areas where their spatial influence is linked with vulnerability to criminal behaviour. The basic RTM process involves incorporating environmental features such as educational institutions, bars, and public transportation stops into the assessment places' crime vulnerability

4. People-based prediction: Network approaches

4.1. **The co-offending network model.** We describe the co-offending network model of crime, following [46] which this section is based on. A broader non-technical survey is also given in [4]. Here, a crime dataset $\mathscr{C}$ is modelled based on a collection of police records that document crime events in a geographic area of interest reported over a period of time. Each record refers to a single crime event; more than one record may refer to the same event.

We may use the notion of a hypergraph $H(N, E)$ consisting of a set of nodes or vertices N and a set of hyperedges E, which is a collection of subsets of N. The set N consists of three disjoint sets, corresponding to actors A, incidents I, and resources R. A hyperedge $e \in E$ is a non-empty subset of N such that the following conditions hold:

$$|e \cap I| = 1, \quad |e \cap A| \geq 1, \quad |e \cap R| \geq 1.$$

Intuitively, a hyperedge associates a set of actors and a set of resources with a single incident. Given $H(N, E)$, we can obtain bipartite graphs by first defining $H'(N', E')$ where $N' = N$ and

$$E' = \{(a, i, r) : \exists\ e \in E \text{ s.t. } \{a, i, r\} \subset e, a \in A, i \in I, r \in R\},$$

then decomposing H' into 3 bipartite graphs AI, IR, and AR, modeling the relations between actors and incidents and so on.

Definition 4.1. Starting from the graph AI, we define a *co-offending network* as a graph $G(V, E)$ where V is the subset of known offenders in A and E indicates known co-offences, meaning that $u, v \in V$ are connected by an edge in G if there is some $i \in I$ such that $\{u, i\}$ and $\{v, i\}$ are edges in AI. We can weight the edges $e \in E$ with the number of known co-offences shared by the same co-offenders, denoted strength(e).

Given k offenders and m offences, define a $k \times m$ matrix M such that each entry $m_{ij} = 1$ if offender i is involved in event j, and 0 otherwise. We can then express the co-offending network as a $k \times k$ matrix

$$N = MM^T, \tag{7}$$

and

$$n_{ij} = \sum_{l=1}^{k} n_{il} n_{lj}.$$

The diagonal of this matrix shows for each offender the number of related crime events.

Remark 4.2 (Scale-free networks). The degree distribution

$$P(k) = \frac{1}{|V|} |\{v \in V : \deg(v) = k\}|$$

of a graph gives the probability that a randomly selected node (i.e., vertex) has degree k. Studies show that many real-world networks from diverse fields ranging from sociology to biology to communication follow a power-law distribution

$$P(k) \sim k^{-\lambda},$$

where λ is called the exponent of the distribution. This distribution implies that there are many nodes of low degree, and few nodes of high degree. Such networks (graphs) are called scale-free networks. The example of a co-offending network studied in [46, §3.1.3] is one such.

The following are some properties of the co-offending network that are studied, which are drawn from the general study of networks.

4.1.1. *Co-offending Strength Distribution.* The co-offending strength of two offenders i and j is defined to the number of offences committed together. We can then define a subgraph $G(V, E, \alpha)$ whose edges consist only of links between $i.j \in V$ whose co-offending strength exceeds a chosen threshold α.

4.1.2. *Connecting paths.* If two offenders are connected in a co-offending network, we can ask what is the shortest path between them. Given $i, j \in V$, define l_{ij} to be the length of shortest path connecting i and j if it exists, and ∞ otherwise. The average distance of network G is defined as the average path distance of connected pairs of nodes,

$$\text{avgd}(G) \frac{1}{|\{\{i,j\} : l_{ij} < \infty\}|} \sum_{\{i,j\}: l_{ij} < \infty} l_{ij}.$$

It can be interpreted as the speed of spreading a message in a network.

Also define the diameter of a network G as

$$\text{diam}(G) = \max\{l_{ij} : l_{ij} < \infty\},$$

giving the longest shortest path between any pair of nodes, which describes the compactness and connectivity of the network. For removing the effect of outliers another measure called effective diameter is sometimes used, given as the minimum number of steps in which at least 90% of all connected pairs of nodes can reach each other [37].

4.1.3. *Clustering Coefficient.* The local clustering coefficient calculates the probability of neighbors of a node v to be neighbors to each other, given by

$$C_v = \frac{a_v}{|\Gamma_v|(|\Gamma_v| - 1)},$$

where a_v is the number of edges between neighbors of v and $|\Gamma_v| = \deg(v)$ is the number of neighbors of v [3]. The quantity $|\Gamma_v|(|\Gamma_v| - 1)$ is the maximum number of edges that can exist between neighbors of v. The clustering coefficient of the network is then computed by simply averaging C_v over all v,

$$C(G) = \frac{1}{|V|} \sum_{v \in V} C_v.$$

The clustering coefficient of a network shows to what extent friends of a person are also friends with each other.

4.1.4. *Connected components.* A connected component is a subset of G where any two nodes in it are connected by a path. Studying characteristics of connected components is an initial step in analysis of epidemic spreading through a social network. Various algorithms can then be applied to detect communities in the network, such as the Girvan–Newman algorithm which progressively removes edges connecting highly clustered communities.

4.1.5. *Centrality measures.* These measures attempt to identify nodes with the greatest structural importance in a network, and can be divided into three groups based on how they are calculated: node degree, shortest path, and node ranking.

(1) Degree centrality measure is simply measured by the degree of v,
$$C_{\text{deg}}(v) = \frac{\deg(v)}{|V| - 1}.$$
(2) Closeness centrality is the inverse of the average distance of any node to v,
$$C_{\text{close}}(v) = \frac{|V| - 1}{\sum_{u \in V} d(u, v)},$$
where $d(u, v)$ is the usual distance between u and v.
(3) Betweenness centrality is the number of shortest paths between pairs of nodes that pass through v,
$$C_{\text{bet}}(v) = \sum_{\substack{u,w \in V \\ u \neq w \neq v}} \frac{\sigma_{uw}(v)}{\sigma_{uw}},$$
where $\sigma_{uw}(v)$ is the number of shortest paths between u and w passing through v, and σ_{uw} is the total number of shortest paths between u and w.
(4) Eigenvector centrality is the the principal eigenvector of the adjacency matrix A_G of G, computed as follows. The relative centrality of node v is given by
$$x_v = \frac{1}{\lambda} \sum_{u \in V} a_{vu} x_t,$$
which can be arranged as an eigenvector equation
$$Ax = \lambda x,$$
where λ is an eigenvalue and x an eigenvector. The v-th component of the related eigenvector then gives the relative centrality score of the vertex v in the network.
(5) The PageRank centrality gives a variant of this measure, computed iteratively as
$$C_{\text{PageRank}}(v) = \frac{1 - d}{|V|} - d \sum_{u \in \Gamma_v} \frac{C_{\text{PageRank(u)}}}{|\Gamma_v|},$$
where d is a fixed probability of continuing the process of moving on the network and not jumping to a random node, and Γ_v is the set of neighbors of v.

The appropriate choice of centrality measure varies according to context. Depending on choice, key players can be identified in the network, and the effects of their removal can be studied.

4.1.6. *Network evolution.* . The evolution of the network G_t can be studied over a period of time, and the invariants above can thus be viewed as variable with respect to a parameter t. For example, centrality measures can be aggregated over time by weighting the centrality of an node v over a discrete time period $t = 1, \ldots, T$ as a function of t, such as a linear weighting

$$C_{x,\text{linear}}(v) = \sum_{t=0}^{T} t \cdot C_x(v_t),$$

where $C_c(v(t))$ computes the centrality measure C_x for v_t in the graph G_t.

4.2. **Community detection and Organized Crime.** Algorithms for community detection in (static) graphs usually look for a 'good' partition of the nodes, where 'good' depends on the purposes at hand. Community detection methods can be divided into three types:

(1) *Node-based.* An example of a node-based community is a clique, which is a maximal complete subgraph in which every two vertices are connected by an edge. (It is NP-complete to identify the maximum clique in a graph.) Other examples include k-plexes, which are subgraphs of size n where each vertex has at least $n - k$ neighbors.
(2) *Group-based.* An example of a group-based community is a density criteria for edges, where if the density of the edges in a group of nodes is larger than a threshold then that group of nodes is defined as a community. This allows for some low-connectivity for some nodes.
(3) *Network-based methods.* This amounts to partitioning the graph into disjoint groups according to some criterion such as connectedness. Here, quality measures are defined on partitions, such as modularity

$$Q = \sum_i (e_{ii} - a_i)^2,$$

where e_{ij} is the fraction of edges that connect nodes in community i to nodes in community j and $a_i = \sum_j a_{ij}$. This is a commonly used measure, though it is NP-complete, thus most solutions are based on approximations.

An approach to offender group detection is proposed in [46, §4.3], which we briefly describe and refer to the reference for details and results. Given a co-offender network G_t, one obtains co-offending networks at time $t_1, \ldots, t_n$ and seeks to detect organized crime groups based on certain thresholds α and β for activity (frequency of offences) and criminality (seriousness of offences) respectively.

First, we define an offender group C_i as consisting of adjacent k-cliques, sharing at least $k-1$ nodes with each other. (In this model, an offender group should have at least three members, so $k \geq 3$.) Let R_t be the set of edges (co-offences) in C_i at time t. Define the activity of C_i at time t_1 with respect to time t_2 to be

$$A_{t_1,t_2}(C_i) = \frac{|R_{t_1}|}{|R_{t_2}|}.$$

Also define the criminality of C_i at time t to be

$$S_t(C_i) = \frac{1}{n} \sum_{k=1}^{n} s(i_k),$$

where $i_1, \ldots, i_n$ enumerates the offences committed by members of C_i at time t and $s(i_k)$ is a measure of the seriousness of i_k. The seriousness can be assigned, for example, using legal classifications. Then C_i can be considered to be active if $A_{t_1,t_2}(C_i) > \alpha$ and serious if $S_{t_2}(C_i) > \beta$ for chosen $\alpha, \beta > 0$ and time frames t_1, t_2. Thus C_i can be considered a possible organized crime group if it is both serious and active.

4.3. **Criminal Profiling: Random walks on graphs.** Given a partial set of crime suspects and a known co-offending network, the problem of crime suspect recommendation aims to recommend the top-N potential suspects. This is related to problems of recommendation in social networks, and can be formulated as follows: Given a crime dataset $\mathscr{C}$, a co-offending network $G(V, E)$ defined on $\mathscr{C}$, and a new crime event e with a set A of offenders already charged, recommend the top-N suspects not included in A that are co-offenders in e with high probability.

In [46, §5.3], the authors introduce CrimeWalker, an extension of TrustWalker [25]. A set of random walks are performed for a single offender u to compute the top-N suspects recommended to be co-offending with u. First, define the probability of stopping at v at the k-th step of a random walk starting from u,

$$P(v, u, k) = \frac{\mathrm{sim}(u, v)}{1 + e^{-k/2}},$$

where $\mathrm{sim}(u, v)$ is a similarity measure between u and v. An example of such a measure is

$$\mathrm{sim}(u, v) = 1 - \left(w_d \frac{D_{uv}}{\max_{u,v}(D_{uv})} + w_a \frac{A_{uv}}{\max_{u,v}(A_{uv})} + w_g G_{uv} \right),$$

where D_{uv} (resp. A_{uv}) denotes the distance between the home locations (rep. ages) of u and v, $G_{uv} = 0$ if u, v share the same gender and 1 otherwise, and w_d, w_a, w_g are weights to be computed and assigned according to importance, for example using a χ^2 method. Secondly, we define the probability of walking from v to w in the k-th step of the random walk starting from u,

$$P(v \to w | u, k) = \frac{n_{vw}}{\sum_{y \in \Gamma_v} n_{vy}},$$

where $n_{vw} = \mathrm{strength}(vw)$.

For each $a \in A$, after performing a set of random walks, top-N suspects with the highest frequency of being returned as the result of a random walk are considered as the top-N recommended co-offenders R_a for the input offender u. To each $r_a \in R_a$, we have a probability $P(r_a|a)$ of reaching r_a in a random walk starting from a. One way to obtain a single list out of the set $\{R_a : a \in A\}$ is taking the aggregate probability

$$P(v|A) = \frac{1}{|A|} \sum_{a \in A} p(v|a),$$

and another approach is to assign a higher probability to suspects that are recommended for more,

$$P(v|A) = \prod_{a \in A} p(v|a).$$

Remark 4.3. It is also possible to use random walks to model crime location prediction, which we refer to [46, §7] for details.

4.4. **Co-offence prediction.** Co-offence prediction is also a natural question to ask, given a co-offence network $G_t = G_t(V_t, E_t)$. We define a potential co-offence at time $t+1$ to be any pair $u, v \in V$ such that $uv \notin E_t$. We say a potential co-offence is in the positive class if $uv \in E_{t+1}$ and in negative class otherwise. The task of co-offence prediction is then to predict for each potential co-offence in G_t, whether it belongs to the positive or negative class. Then [46, §6.3] proposes a supervised learning framework for learning potential positive co-offenders. We also mention [23] which compares ensemble versus regression methods in predictive policing.

As an example, Green et al. [19] study the social contagion of gunshot violence using a co-offending network, where the violence is modelled again as a (Bayesian) Hawkes point process (4). Here, each node occupies its own coordinate of the Hawkes Process and each gunshot victimization is an event of the process occurring on the coordinate that corresponds to the victim (repeated victimizations of the same individual correspond to multiple events on the same node, and are treated the same as single victimizations). The exciting function is given by the product

$$g_\alpha(u, v)\omega e^{-\omega(t-t_i)}$$

where for $u, v \in V$,

$$g_\alpha(u, v) = \begin{cases} \alpha d(u, v)^{-2}, & d(u, v) \leq 3 \\ 0 & d(u, v) > 3 \end{cases}$$

models the assumption that contagious events are localized and that the transmission probability increases closer to the source. They also account for seasonal variations by defining the expected number of total victimizations occurring on day t as

$$M(t) = A\left(1 + \rho \sin\left(\frac{2\pi t}{365.24}\right) + \varphi\right),$$

where the parameters A, ρ, ϕ are learned using non-linear least squares estimates with the Gauss-Newton algorithm. It is then related to the node-level exogenous intensity μ_t by

$$M(t) = |V| \int_{t-1}^{t} \mu_s ds,$$

and assuming μ_t to be approximately constant over the course of one day, one has simply $M(t) = |V|\mu_t$. Finally, the authors then use this model to predict for future events.

We conclude this section with a general problem, which we will find partially answers to in the next section.

Problem 4.4. Apply the co-offending network model to police offences.

5. The mathematics of gun violence

In 1993, Kellerman et al. [26] found that guns in the home were associated with an increased risk of homicide in the home. The study was funded by the CDC National Center for Injury Prevention and Control (NCIPC), and lobbying by the NRA produced the 1996 Dickey Amendment, disallowing the funds from being used "advocate or promote gun control," and equivalent language was used to expand this to the NIH in 2012, effectively eliminating external funding for research on gun violence. In 2018, the bill was reinterpreted to allow for research but not to "specifically advocate for gun control," and the fiscal year 2020 federal budget

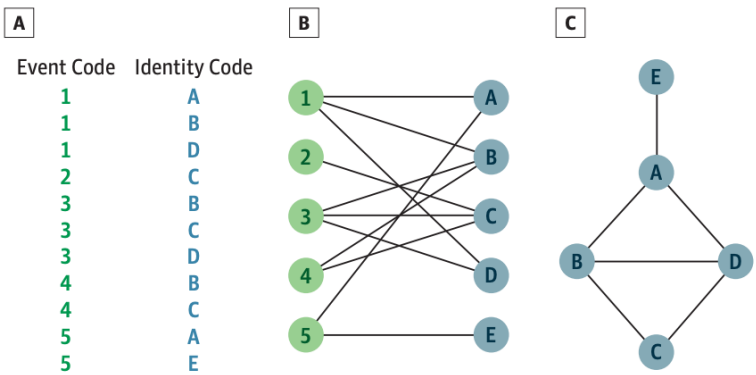

FIGURE 2. An example of constructing a co-offending network [19, Figure 1].

provided the first funding since 1996. Despite this, the mathematics of gun violence is has been researched in certain respects, much of it statistical, and we refer to [40] for an excellent survey of the state of the art and directions for future research.

5.1. **Epidemiological criminology.** There is much to be said about the connections and differences between criminology and epidemiology. Indeed, Akers and Lanier [1] propose *epidemiological criminology* as a framework to integrate the two disciplines, and urge that"war" metaphors must be replaced with "health" metaphors at the highest levels of policy. Slutkin has also argued in many places, for viewing violence as a contagious disease [45].

5.1.1. *The contagious nature of imprisonment.* We first describe an agent-based model of incarceration based on the susceptible-infected-suspectible (SIS) model by Lum et al. [29]. This requires three main components: a contact network through which individuals stochastically transmit the disease, transmission probabilities that dictate the rate at which agents transmit to each other and a period of infectivity. Here, incarcerated people are considered to be 'infectious,' and transmit the 'disease' of incarceration to others; individuals released from prison cease to be 'infectious' and return to a 'susceptible' state

The probability that infected agent i transmits to susceptible agent j by $p(i \to j)$. The probability that i transmits the disease to agent j over the whole course of the infectious period s is given by

$$p_s(i \to j) = 1 - (1 - p(i \to j))^s.$$

Similar to [28], a synthetic multe-generation population is simulated based upon US Census, the CDC, and the Social Security Administration, for the distributions for the sex, lifespan, and the number of children of each agent respectively. Transmission probabilities $p(i \to j)$ are derived directly from the survey of prison inmates giving the probability that an inmate's mother, father, sister, brother or adult child are also incarcerated by inmate gender. Sentencing distributions are generated using Bureau of Justice Statistics data.

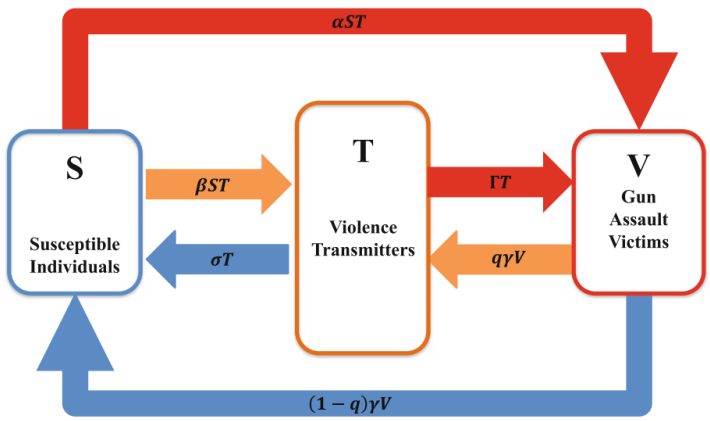

FIGURE 3. Flow diagram for the STV model as in [48, Figure 1].

Assuming random mixing and homogeneity of transmission rate, the also consider an ODE approach, where the number of people infected when an outbreak reaches a steady state I, is determined by the expected number of transmissions per infected person, which in turn is given by the product ps of transmission rate and the duration of infectivity. Ignoring births and deaths, one has

$$I = \begin{cases} I = 0, & s < \frac{1}{p}, \\ I = 1 - 1/ps & s \geq \frac{1}{p}. \end{cases}$$

One can define a critical duration of infectivity $s_c = \frac{1}{p}$ such that for $s < s_c$, the outbreak dies out, and for $s > s_c$, it achieves a non-zero steady-state prevalence. In sum, the authors show that the "dramatic disparities" in incarceration rates of Black and White Americans can be generated by the 'transmission' of incarceration from an incarcerated person to his or her family and close friends combined with modest differences in sentencing.

Problem 5.1. How does the contagion model of incarceration interact with policing and surveillance methods?

5.1.2. *Impact of Violence Interruption on the Diffusion of Violence.* Wiley et al. [48] introduce a Susceptible–Transmitter–Victimized (STV) epidemic model to explore the impact of violence interruption on contagious violence. Let S be the susceptible population, vulnerable to adopting a culture of violence and vulnerable to violence victimization, T be the transmitter population, currently engaged in a culture of violence, and V the victmized population.

Individuals in each class can move from one to another following Figure 3. For example, those in T can infect those in S in one of two ways: influence others to adopt a violent lifestyle with a transmission efficiency of β, in which case the susceptible moves from S to T; or (2) violence transmitters can assault a susceptible at a transmission rate of α, after which the susceptible moves from S to V. Those in T can move to S through violence interruption efforts at a rate of σ. Assumptions

are made s consistent with the classic SIR compartmental model and its variations, such as homogeneous mixing (not necessarily appropriate) and a constant population $S + T + V = K$. Based on model simulations, the authors argue that their results suggest that the most effective method for reducing gun violence are interventions that target all individuals at risk of becoming violence transmitters irrespective of whether or not individuals have immediate access to a firearm.

Problem 5.2. Can violence interruption be applied to police violence? If so, how?

Part 2. Predicting Police

6. Police misconduct

A broad question now presents itself: how can we use the tools described above, or develop new tools in order to study the police? For example, data journalists at The New Inquiry made a simple model to predict financial crime in New York City [9], following the idea of predicting white-collar crime. In this second part, we review efforts to study police themselves through various means and for various purposes.

6.1. **Early Intervention Systems (EIS).** Specific programs have been widely adopted, broadly known as Early Intervention Systems designed to reduce officer misconduct. More recent examples of EIS systems such as in Charlotte-Mecklenburg Police Department [6, 21] and also in Australia [10] use machine learning enabled recommendations systems, taking various factors into account such as officer characteristics, situational factors, and neighborhood factors. See also [20] for a recent survey of EIS systems.

6.2. **Co-complaint networks.** In recent years, several studies on the network structure of police misconduct have been carried out. For example, in a series of works by Papachristos and collaborators [35, 50, 52], complaint records of the Chicago Police Department (CPD) used to first estimate the effect of individual level attributes such as gender, tenure, race, and ethnicity on the frequency of complaints received, and second, construct a co-complaint network that is then analyzed. There are two types of misconduct: civilian-facing complaints, such as excessive use of force and illegal searches, and department-facing complaints such as drug abuse, lost weapons, and other operations and personnel violations, modelled as the left and right graphs respectively in Figure 4. In [50], the authors employ a Bayesian exponential random graph model (BERGM), estimating

$$P(y|\theta) = \frac{1}{z(\theta)} \exp\left(\theta_1 + s_1(y) + \theta_2 s_2(y) + \cdots + \theta_k s_k(y)\right),$$

where θ_k are parameter estimates for the sufficient statistics $s_k(y)$ (including covariate-specific statistics for race and ethnicity, gender, tenure, and weeks at risk of receiving a complaint) and $z(\theta)$ is a likelihood normalizing constant. A positive estimate for θ_k implies that the frequency of s_k in the observed data is greater than would be expected by chance, conditional on the other parameters in the model.

6.2.1. *Network position of police who shoot.* A finer analysis is undertaken in [52], where the position of police who shoot civilians within the co-complaint network are shown to act as "brokers" in the network, here used in the sense of lying on shortest paths between other nodes in the network, measured by the betweenness centrality measure described in Section 4.1.5(3). Note that since betweenness is only defined within path-connected graphs, a nodes betweenness centrality is defined relative to is connected component. It is then shown that betweenness centrality in either network is a significant predictor of faster time to first shooting, and that this holds even after controlling for individual level attributes.

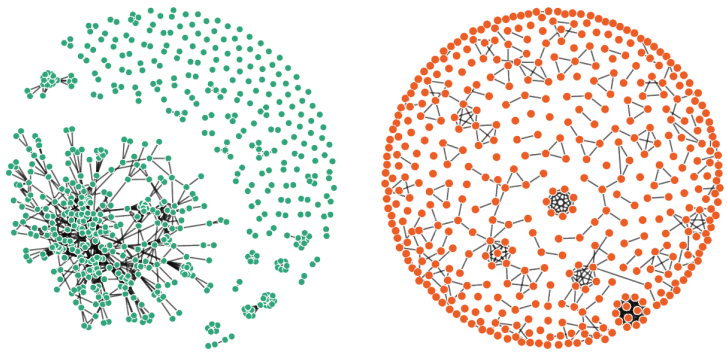

FIGURE 4. The network structure of CPD misconduct [50, Figure 3].

6.2.2. *Repurposing recidivism models.* In study by different authors, [24] proposed repurposing the PredPol to model police use of force.[2] Here, a co-offending network is constructed as in §4, here referred to as an officer co-complaint network. The excessive use-of-force is again modelled as a Hawkes process as in (4). Besides applying a machine learning framework to forecast the use-of-force complaints, the authors also propose an interpretable model taken from recidivism risk that produces a risk score card. Here, one first generates a large candidate set of binary features X_i, then finds a sparse set of integer coefficients θ that solve the optimization problem of the form

$$\min_{\theta \in \mathbb{Z}^n} \sum_{i=1}^{N} 1_{\{y_i X_i^T \theta \leq 0\}} + c_0 ||\theta||_0 + c_1 ||\theta||_1,$$

where the 0-norm encourages sparsity and the 1-norm encourages the coefficients to be small. The problem can be cast as a mixed-integers linear programming problem and solved efficiently.

6.2.3. *Misconduct as social contagion.* Co-complaint networks only provide a skewed snapshot of a police department. A broader context for the misconduct network can reveal more about how misconduct might spread throughout the broader policing network.

Problem 6.1. Use more comprehensive data to contextualize a co-complaint network.

Two approaches to latter problem have been pursued. In a UK context [38], peer groups were defined by linking officers and staff assigned to the same line manager, and estimate the variable

$$y_{it} = \begin{cases} 1 & \text{if police } i \text{ has a misconduct event at time } t \\ 0 & \text{otherwise} \end{cases},$$

and the independent variable of interest is the proportion of peers of i in $t-1$ receiving misconduct reports at $t-1$,

$$\text{Peer}(y_{i(t-1)}) = \frac{\sum_{j \neq i} y_{j(t-1)}}{\sum_{j \neq i} 1}.$$

[2]Note that the one of the co-authors, George Mohler is also the architect of PredPol itself.

One then estimates the linear probability model

$$P(y_{it} = 1) = \alpha \text{Peer}(y_{i(t-1)}) + \beta W,$$

where W is a vector of control variables that include demographic characteristics (such as gender, length of service, employee's business group, employee type and employee performance) and additional controls for annual and seasonal effects, and β a vector of coefficients for these variables. Applied to London's Metropolitan Police Service (2011–2014), the authors find that a 10% increase in prior peer misconduct increases an officer's later misconduct by 8%.

On the other hand, [42] examines the social contagion property of police misconduct by combining misconduct complaints with 911 calls for service to the Dallas Police Department (2013–2014). One first constructs an officer response to calls for service with an $n \times z$ matrix where M where n is the number of police in the sample and z the number of unique call-generating incidents occurring on a given day t. Then similar to (7), we have a co-response matrix

$$W = MM^T$$

encoding the number of times any two officers respond to the same 911 incident on day t. One then defines a spatial lag for exposure to deviance

$$\sum_{j:j \neq i}^{n} w_{ij}^t y_{jt}$$

where the entry w_{ij}^t in W is the number of joint responses by police i and j across day t, and y_t is a binary vector indicating whether or not each of the n officers engaged in any misconduct that was ultimately sanctioned on day t. A decay factor is also introduced to incorporate forgetting, rather than a sharp cutoff window of observation.

Now let $N_i(t)$ be an increasing, right-continuous counting process, counting the number days of misconduct of police i up to day t. It is modelled as a Poisson process, so we are interested in its intensity, that is, the instantaneous probability of sanctioned misconduct. Let

$$Y_i(t) = \begin{cases} 1 & \text{if } i \text{ is at risk of misconduct at } t \\ 0 & \text{otherwise} \end{cases}$$

and let Z_i be the unobserved frailty, i.e., the random/varying effect, shared across all risk intervals for i. Then the intensity of N_i for the k-th strata, i.e., the k-th day of misconduct) at time t is given as

$$\lambda_{ik}(t|Z_i) = Y_i(t) Z_i \exp(\beta^T x_i(t)) \lambda_{0k}(t),$$

where $x_i(t)$ is a vector of p covariates for i at t, β the corresponding vector of p unknown parameters (the regression coefficients), and λ_{0k} is the unspecified, nonnegative baseline intensity for the k-th strata at time t when all covariates are 0. These parameters are conditional log intensity ratios which summarize the positive or negative association between some covariate of interest x_{ip} and the intensity of the k-th event for a one-unit increase in x_{ip}, where this association is multiplicative. But in contrast to [38], the authors conclude from their findings that there is little evidence for the contagion effect in peer networks built from 911 calls to service; we refer to the discussion in [42] for a comparison of these and other results.

7. Further discussion

7.1. **Statistical work.** Besides the work of Lum and Isaac on PredPol [27], other statistical studies have been carried out in various related settings, of which we mention a few: [22, 13] analyses police stops, searches, and arrests in Oakland, [17] analyses data from NYPD's stop-and-frisk program (which, importantly, was used by the ACLU in fighting it), [5] studies the effect of Black Lives Matter protests on police use-of-force complaints, and [16] applies a network meta-regression to fatal police violence data to to estimate that the National Vital Statistics System (NVSS) data underreports about a half of police killings. We also note that [8] studies the effect of removing certain high-risk offenders ("bad apples"), though their methodology is not entirely clear to us. There are many fruitful avenues to pursue here and elaborate upon, but in this paper we have chosen to focus on the mathematical aspects in contrast to the statistical aspects of policing. We hope that someone more well-positioned than us will be able to provide a complementary survey of the statistical approaches.

7.2. **Final questions.** We conclude with some general questions for further consideration, besides the more specific questions raised throughout the paper.

(1) What are all the assumptions used in these models, explicitly or implicitly? What do these models reveal and what do they obscure?
(2) Where can bias occur? How might it be taken it to account? We can think of a mathematics to criminalization pipeline, where the intermediary is technological implementation of an underlying mathematical model. The question for mathematicians then primarily regards where mathematics
(3) Many models are created based on geospatial data. How might one account for the affects of redlining, economic segregation, and other such factors? How does broken windows theory and its critiques apply, if at all, to the study of police?
(4) Ensemble methods have proven power in the mathematical approaches to gerrymandering. Might they also find some application in this area?
(5) Could this kind of modeling be applied in an abolitionist framework? A crucial paradigm in this regard that it is about *presence* rather than absence. Using the language of public health such as in [40], this relates in part to interventions or interruptions of criminalisation and violence. What if the predictions were used to allocate resources for community support rather than policing? How might models take make predictions based on various forms of intervention or interruption?

Perhaps most importantly, the larger question is: what *kinds* of questions should we be asking about policing? As discussed in the introduction, the ways in which we think about the function and purpose of policing informs the ways in which we consider it and the questions that we ask of it. What sort of world are we imagining?

We can take the recent development of the mathematics of development as a model for how one might develop a research program with regards to the mathematics of carceral systems at large. The problem of gerrymandering might be phrased as follows: given a partition p of a space X into N districts $X_1, \ldots, X_N$, each with an election with 2 candidates and M voters, show that the partition p is 'fair' in the sense that it produces proportional representation. Of course, the

difficulty arises in defining what 'fair' means, and how it is measured. Methods such as political geometry, ensemble methods, topological data analysis, and the efficiency gap have been applied in this context to study this problem.

What would constitute a 'core' meta-research question in the context of policing? One possible answer is the following imprecise question: *When is policing necessary?*. Under this banner might be questions such as

(1) What incidents motivate people to call the police? For example, one might report an incident like a break-in, a vehicle crash, a physical assault, but also noise complaints and suspicious characters. What is important here is the fact that police are not always called when an individual is directly at risk of harm, but very often to alert the authorities of incidents of various natures. Most notably, this includes when police may be called to certain nonviolent events (e.g., mental health crises involving unarmed persons) that escalate and become police killings. How might a different kind of first responder (say, an unarmed one) change the outcome?
(2) Where are police concentrated and why? What are the ecological factors that correlate to the presence of policing, such as race, income, density, and resources? Is it possible to determine the causal relationships between levels of policing and crime? For example, can causal methods shed light on whether increased crime leads to increased policing, and if this increased policing has the effect of reducing crime?
(3) The fundamental question of what *is* crime is also a deep question in radical criminology. For the mathematician, the classification question here is perhaps less important than the intervention: what other methods are there, such as violence interruption, that reduce calls to 911? Can we interrupt police violence? What about transformative justice?
(4) A practical side of the previous question is the matter of data: typical crime statistics tend to exclude incidents that are perpetrated by police, correctional officers, military personnel, and so on (sometimes called the exclusionary rule for blue data). How does the data analysis change when we consider blue data? How does it compare with non-blue data?
(5) What can be done with studying police funding data? Does resource allocation according to city translate to "clearing cases"? How is "effective spending" of public funds determined in this case? The question of de/funding the police is of course a sensitive question at the moment, but a more basic question is simple: how do we determine if police funding is being well-spent? What are the benchmarks for this?
(6) How do we interface police data with incarceration rates? Recall the work of Lum et al. [29] that models incarceration risk as contagion, but it doesn't deal with its relation to policing. What else can we learn from incarceration data? Is there a data story that can be told about the legacy of slavery to mass incarceration, as articulated by the 13th amendment?

In total, these questions attempt to bear out the problem of what is the function of policing, what is the definition of crime, and what is the proper response to such events. Whereas much literature on policing assumes that the only response to crime is police intervention, broader alternatives have been raised such as mental health practitioners, social workers, and other nonviolent responders who may be more adequately prepared to respond to various incidents such as mental health

crises, domestic disputes, and so on. As a function of social interactions, this may perhaps be modeled as a transitional state from non-crisis to crisis, and what are the crisis conditions that necessitate police intervention. That is, we have a step function

$$f : S_t \to \{0, 1\},$$

where S_t is a state depending on time (and other implicit variables), and the output 1 indicates a call to 911. The problem then would be to determine under what conditions the transition occurs.

Acknowledgments. The author was partially supported by NSF grant DMS-2212924.

References

[1] Akers, T.A. and Lanier, M.M., 2009. "Epidemiological criminology": Coming full circle. American journal of public health, 99(3), pp.397-402.

[2] Aougab, T., Ardila, F., Athreya, J., Goins, E., Hoffman, C., Kent, A., Khadjavi, L., O'Neil, C., Patel, P. and Wehrheim, K., 2020. Boycott collaboration with police. The Notices of the American Mathematical Society, 67(9), p.1293.

[3] A.L. Barabasi, R. Albert, Emergence of scaling in random networks. Science 286(5439), 509–512 (1999)

[4] Burcher, Morgan. Social Network Analysis and Law Enforcement: Applications for Intelligence Analysis. Springer Nature, 2020.

[5] Campbell, Travis. "Black Lives Matter's Effect on Police Lethal Use-of-Force." Available at SSRN (2021).

[6] Carton, S., Helsby, J., Joseph, K., Mahmud, A., Park, Y., Walsh, J., Cody, C., Patterson, C.E., Haynes, L. and Ghani, R., 2016, August. Identifying police officers at risk of adverse events. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 67-76).

[7] Castelvecchi, Davide. "Mathematicians urge colleagues to boycott police work in wake of killings." Nature (2020).

[8] Chalfin, A. and Kaplan, J., 2021. How many complaints against police officers can be abated by incapacitating a few "bad apples?". Criminology & Public Policy, 20(2), pp.351-370.

[9] Clifton, B., S. Lavigne, and F. Tseng. "Predicting Financial Crime: Augmenting the Predictive Policing Arsenal, in The New Inquiry." (2017).

[10] Cubitt, Timothy IC, Ken R. Wooden, and Karl A. Roberts. "A machine learning analysis of serious misconduct among Australian police." Crime Science 9.1 (2020): 1-13.

[11] Duchin, M. and Walch, O., 2021. Political Geometry. Rethinking Redistricting in the US with Math, Law, and Everything In Between. Springer.

[12] M.R. D'Orsogna and M. Perc, Statistical physics of crime: A review. Physics of life reviews, (2015) 12, pp.1-21.

[13] Eberhardt, Jennifer Lynn, ed. Strategies for change: Research initiatives and recommendations to improve police-community relations in Oakland, Calif. Stanford SPARQ, 2016.

[14] D. Ensign, S.A. Friedler, S. Nevlle, C. Scheidegger, and S. Venkatasubramanian, 2018, February. Runaway Feedback Loops in Predictive Policing. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Vol. 81).

[15] Ferguson, A.G., 2016. Policing predictive policing. Wash. UL Rev., 94, p.1109.

[16] GBD 2019 Police Violence US Subnational Collaborators, 2021. Fatal police violence by race and state in the USA, 1980–2019: a network meta-regression. The Lancet, 398(10307), pp.1239-1255.

[17] Goel, Sharad, et al. "Combatting police discrimination in the age of big data." New Criminal Law Review 20.2 (2017): 181-232.

[18] Goodson, H. and Hoyer-Leitzel, A., 2021. Examining the Modeling Framework of Crime Hotspot Models in Predictive Policing. Journal of Humanistic Mathematics, 11(2), pp.465-476.

[19] Green, B., Horel, T. and Papachristos, A.V., 2017. Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014. JAMA, 177(3), pp.326-333.

[20] Gullion, C.L. and King, W.R., 2020. Early intervention systems for police: A state-of-the-art review. Policing: An International Journal.

[21] Helsby, J., Carton, S., Joseph, K., Mahmud, A., Park, Y., Navarrete, A., Ackermann, K., Walsh, J., Haynes, L., Cody, C. and Patterson, M.E., 2018. Early intervention systems: Predicting adverse interactions between police and the public. Criminal justice policy review, 29(2), pp.190-209.

[22] Hetey, Rebecca, et al. Data for change: A statistical analysis of police stops, searches, handcuffings, and arrests in oakland. Calif., 2013-2014. Technical report, Stanford University, SPARQ: Social Psychological Answers to Real-World Questions, 2016.

[23] Kathuria, Ronit, and Vinish Kathuria. "Predictive Policing—Are Ensemble Methods More Accurate Than Regression Methods?." Artificial Intelligence and Technologies. Springer, Singapore, 2022. 145-152.

[24] Khorshidi, Samira, Jeremy G. Carter, and George Mohler. "Repurposing recidivism models for forecasting police officer use of force." 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020.

[25] M. Jamali, M. Ester, TrustWalker: a random walk model for combining trust-based and itembased recommendation, in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09), pp. 135–142, 2009

[26] Kellermann, A.L., Rivara, F.P., Rushforth, N.B., Banton, J.G., Reay, D.T., Francisco, J.T., Locci, A.B., Prodzinski, J., Hackman, B.B. and Somes, G., 1993. Gun ownership as a risk factor for homicide in the home. New England Journal of Medicine, 329(15), pp.1084-1091.

[27] K. Lum, Derivation of the Expectation-Maximization Algorithm for the ETAS Model https://github.com/arun-ramamurthy/pred-pol/blob/master/doc/Rederivation%20of%20Mohler%20et%20al.pdf

[28] K. Lum, and W. Isaac. To predict and serve?. Significance, (2015) 13(5), pp.14-19.

[29] Lum, K., Swarup, S., Eubank, S. and Hawdon, J., 2014. The contagious nature of imprisonment: an agent-based model to explain racial disparities in incarceration rates. Journal of the Royal Society interface, 11(98), p.20140409.

[30] Mugari, I. and Obioha, E.E., 2021. Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing. Social Sciences, 10(6), p.234.

[31] Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them, The Markup https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them

[32] G. O. Mohler (2015). U.S. Patent No. 8,949,164. Washington, DC: U.S. Patent and Trademark Office.

[33] G. O. Mohler et al., Randomized controlled field trials of predictive policing, J. Amer. Statist. Assoc. **110** (2015), no. 512, 1399–1411. MR3449035

[34] Mohler, George, et al. "A penalized likelihood method for balancing accuracy and fairness in predictive policing." 2018 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, 2018.

[35] Ouellet, M., Hashimi, S., Gravel, J. and Papachristos, A.V., 2019. Network exposure and excessive use of force: Investigating the social transmission of police misconduct. Criminology & Public Policy, 18(3), pp.675-704.

[36] C. O'neil. 2016. Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.

[37] C.R. Palmer, P.B. Gibbons, C. Faloutsos, Anf: a fast and scalable tool for data mining in massive graphs, in Proceedings of the 8th ACM SIGKDD International Conference on Knowledge discovery and Data Mining (KDD'02) (2002), pp. 81–90

[38] Quispe-Torreblanca, E.G. and Stewart, N., 2019. Causal peer effects in police misconduct. Nature human behaviour, 3(8), pp.797-807.

[39] Renlund, H., 2010. Generalized Pólya urns via stochastic approximation. arXiv preprint arXiv:1002.3716.

[40] Scott, S. 2020. A Report from the NIMBioS/DySoC Investigative Workshop on the Mathematics of Gun Violence. http://www.nimbios.org/ifiles/WS_gunviolence_fullreport.pdf

[41] Shapiro, A. Reform predictive policing. Nature 541, 458–460 (2017). https://doi.org/10.1038/541458a

[42] Simpson, C.R. and Kirk, D.S., 2022. Is Police Misconduct Contagious? Non-trivial Null Findings from Dallas, Texas. Journal of Quantitative Criminology, pp.1-39.

[43] M. B. Short et al., A statistical model of criminal behavior, Math. Models Methods Appl. Sci. **18** (2008), suppl., 1249–1267. MR2438215

[44] M. B. Short, A. L. Bertozzi and P. J. Brantingham, Nonlinear patterns in urban crime: hotspots, bifurcations, and suppression, SIAM J. Appl. Dyn. Syst. **9** (2010), no. 2, 462–483. MR2670006

[45] G. Slutkin, Violence is a contagious disease, in Contagion of Violence: Workshop Summary (2012)

[46] Tayebi, M.A. and Glässer, U., 2016. Social network analysis in predictive policing. In Social Network Analysis in Predictive Policing (pp. 7-14). Springer, Cham.

[47] www.vice.com/en_us/article/xwbag4/academics-confirm-major-predictive-policing-algorithm-is-fundamentally-f and www.vice.com/en_us/article/d3m7jq/dozens-of-cities-have-secretly-experimented-with-predictive-policing

[48] Wiley, S.A., Levy, M.Z. and Branas, C.C., 2016. The impact of violence interruption on the diffusion of violence: a mathematical modeling approach. In Advances in the Mathematical Sciences (pp. 225-249). Springer, Cham.

[49] Obasogie, O.K. and Newman, Z., 2017. Police violence, use of force policies, and public health. American Journal of Law & Medicine, 43(2-3), pp.279-295.

[50] Wood, G., Roithmayr, D. and Papachristos, A.V., 2019. The network structure of police misconduct. Socius, 5, p.2378023119879798.

[51] Zeoli, A.M., Grady, S., Pizarro, J.M. and Melde, C., 2015. Modeling the movement of homicide by type to inform public health prevention efforts. American journal of public health, 105(10), pp.2035-2041.

[52] Zhao, L. and Papachristos, A.V., 2020. Network position and police who shoot. The ANNALS of the American Academy of Political and Social Science, 687(1), pp.89-112.

Email address: tiananw@umich.edu