Non-Stationary Model for Crime Rate Inference Using Modern Urban Data

Hongjian Wang, Huaxiu Yao, Daniel Kifer, Corina Graif, Zhenhui Li

Abstract—Crime is one of the most important social problems in the country, affecting public safety, children development, and adult socioeconomic status. Understanding what factors cause higher crime rate is critical for policy makers in their efforts to reduce crime and increase citizens' life quality. We tackle a fundamental problem in our paper: crime rate inference at the neighborhood level. Traditional approaches have used demographics and geographical influences to estimate crime rates in a region. With the fast development of positioning technology and prevalence of mobile devices, a large amount of modern urban data have been collected and such big data can provide new perspectives for understanding crime. In this paper, we use large-scale Point-Of-Interest data and taxi flow data in the city of Chicago, IL in the USA. We observe significantly improved performance in crime rate inference compared to using traditional features. Such an improvement is consistent over multiple years. We also show that these new features are significant in the feature importance analysis. The correlations between crime and various observed features are not constant over the whole city. In order to address this geospatial non-stationary property, we further employ the geographically weighted regression on top of negative binomial model (GWNBR). Experiments have shown that GWNBR outperforms the negative binomial model.

Index Terms—Computer Crime inference, taxi flow, geographically weighted regression, negative binomial model.

1 Introduction

Understanding the factors that predict crime is important because even though crime rates have been generally declining since the early nineties, recent years have started to see lower rates of decline and even some upward fluctuations after 2010 [1]. Moreover, reports of direct and indirect victimization and exposures to crime remain very high [2]. For instance, more than two-fifths of children and youth in a recent national survey reported a physical assault in the previous year [3]. Understanding the neighborhood context of crime is particularly important because victimization and other forms of crime exposures have many severe consequences. Beyond the high medical bills and violent death, consequences include behavioral and mental health problems, aggression, substance abuse, post-traumatic stress disorder, and suicide, lower academic achievement, and engaging in further violence [4].

In this paper, we study the problem of crime rate inference across communities. We select Chicago as the target of study for the following reason. Chicago has more homicides and non-negligent manslaughter rates (15.2) per 100,000 residents than New York (4.0) and Los Angeles (6.5) according to the FBI crime statistics for 2013 and has experienced no decline in the past decade compared to the other two large cities, which have been on a slow declining slope [5].

Traditionally, researchers have used demographic information based on Decennial Census (e.g., population poverty level, socioeconomic disadvantage, racial composition of population) to estimate the crime rate in a community [6]. However, such information only contains the social information of residents in the neighborhoods and misses information on daily population dynamics within and between neighborhoods. In our experiments (Section 6), negative binomial and geographically weighted negative binomial models that only use demographic features and an intercept result in a relative error of as much as 30% for crime rate estimation in Chicago. Existing studies also highlight the importance of geographical influence [7] in estimating crime rates, i.e., the crime in the nearby communities can be propagated to the focal community. But, depending on the geographic scale of analysis, geographical influence does not contribute a great deal in improving the crime inference on top of demographic feature (with at most 0.4% relative improvement in our experiments focused on larger geographic units than census tracts). This is probably because the nearby communities also share similar demographics, which limits the additional benefit of geographical influence.

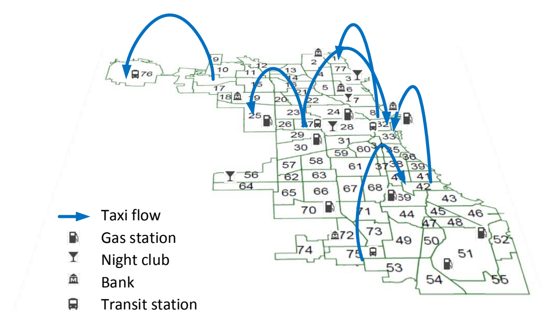


Fig. 1: An illustration of various types of features we use in Chicago. The POI distributions across community areas profile the region functionality. The taxi flows connect non-adjacent regions and act as "hyperlinks" on the space.

- *Hongjian Wang, Huaxiu Yao, and Zhenhui Li are with the College of Information Sciences and Technology, Pennsylvania State University. E-mail: {hxw186, huy144, jessieli}@ist.psu.edu*
- *Daniel Kifer is with the Department of Computer Science and Engineering, Pennsylvania State University. E-mail: dkifer@cse.psu.edu*
- *Corina Graif is with the Department of Sociology and Criminology, Pennsylvania State University. E-mail: corina.graif@psu.edu*

Manuscript received April 19, 2017.

Recently, big data reflecting city dynamics have become widely available [8], e.g., traffic flow, human mobility, social media, and crowd-generated Points-Of-Interest (POI). As shown in Figure 1, such newer types of big data could provide new insights to advance our understanding of traditional socioeconomic urban problems, such as the crime rate inference problem we focus on in this paper. In particular, we propose to study two newer types of urban data: POI and taxi flow.

POI data. POI data provide venue information such as GPS coordinates, category, popularity, and reviews. These POIs mostly belong to categories such as food, shop, transit, education, etc. As one example, the POI data reveal locations of gas stations and convenience stores, which are more likely to be targets for crime because of lack of guardianship, easy to access, and presence of readily attainable valuables [9]. Recent studies have also shown that using such categorical information of POIs are helpful to profile neighborhood functions [10]. Such neighborhood functions could further help us predict crime rate (e.g., communities with less education or entertainment facilities may have a higher rate of crime).

Taxi flow data. A huge amount of taxi flow data reflect important information about how people move across neighborhoods in the city. In previous studies, when using geographical influence [7], scientists assume that a community is affected by other spatially proximate communities. However, even if two communities are distant in geographical space, they could be strongly correlated if many people frequently travel between these two communities [11]. We hypothesize that taxi flows may be considered as “hyperlinks” in the city that connect different areas and we use such data to estimate crime rates. We do not expect taxi flow data to capture the movements of offenders, as the vast majority of taxi trips are probably unrelated to crime. Instead, we view taxi flow as a proxy for broader patterns of population routine activity and mobility, commuting flows, and other forms of social and economic exchanges between two communities over space. Such exchanges may increase the number of potential targets and opportunities for crime [12], [13] or contribute to inter-community diffusion of information about successful local strategies to control or prevent crime (e.g., successful features of neighborhood watch programs).

We apply various regression models to 5 years of crime data in Chicago. The most frequently used model is linear regression; however, because crime count cannot be negative, we also use negative binomial regression. We demonstrate that negative binomial model generally performs better than the linear regression. In addition, adding POI and taxi flow features reduces the relative error by at least 5% in our experiments. This indicates that the new urban data provide additional information about the communities which are not covered by traditional features.

As an extension to the conference version [14], we investigate models that incorporate geographic heterogeneity; that is, we do not expect the same features to have the same relation to crime in every location because crime incidents in different regions may be associated with different social-economic factors. In fact, there are several neighborhoods where negative binomial model gives poor prediction. This tells us that a global model such as the negative binomial model, which assumes a constant correlation between crime and observed features, would not yield accurate estimates. Therefore, we further propose to employ a graphically weighted regression approach to capture the non-stationary property of crime. The intuition behind this strategy is to train many local models instead of one global model to predict the crime. The geographically weighted regression is a useful framework on how to pick samples and weight them for local model training. We thus apply geographically weighted regression in combination with negative binomial model, and the experiments show further improvements over the global model.

In summary, the contributions of this paper are: 1) We study an old but important crime inference problem by utilizing new urban data: POIs and taxi flows. We provide detailed discussions of how to construct features, tests of different combinations of features, and the theoretical interpretations of the result from a social scientist (a co-author in the paper). 2) We find that utilizing these new types of big urban data improves the crime rate inference. 3) We employ a geographically weighted regression framework to capture the non-stationary property of crime. 4) We conduct a systematic comparison between various regression models. The geographically weighted negative binomial model has significantly better performance, and could serve as a new baseline for future crime inference problems.

The rest of this paper is organized as follows. We first review the related work in Section 2. The crime inference problem is formulated in Section 3. We discuss the inference model in Section 4 and feature extraction procedure in Section 5. The Section 6 presents the quantitative evaluation results on real data. Finally, we conclude the paper in Section 7.

2 RELATED WORK

Sensing technologies and large-scale computing infrastructures have produced a variety of big data in urban spaces (e.g. human mobility, POI, and traffic). These heterogeneous data convey rich knowledge about city dynamics and enable us to address many urban challenges. For example, human mobility data could help improve the efficiency of transportation systems such as estimating real-time traffic flow [15], [16], and forecasting travel time for road segments [17], [18], [19] or a trip [20], [21]. The POI and taxi data can be used to infer air quality [22] and city noises [23]. With the similar motivation, we employ such modern urban data for crime rate inference.

In the criminology literature, researchers have studied the relationship between crime and various features (social, demographics, and geographic factors). Examples are historical crime records [24], [25], education [26], ethnicity [27], income level [28], unemployment [29], and spatial proximity [7]. In data mining, newer types of data are used. For example, studies use twitter to predict crime [30], [31], and cellphone data [32], [33] to evaluate crime and social theories at scale. Overall, existing work on crime prediction can be categorized into three paradigms.

Time-centric paradigm. This line of work focuses on the temporal dimension of crime incidents. For example, Mohler et. al. [24] propose to use a self-exciting point process to model the crime and gain insights into the temporal trends in the rate of burglary. In another study, Ratcliffe [34] investigates the temporal constraints on crime, and propose an offender travel and opportunity model. His findings suggest that a proportion of offending is driven by the availability of opportunities presented in the routine lives of offenders.

Place-centric paradigm. Most existing works adopt a place-centric paradigm, where the research question is to predict the location of crime incidents. The predicted crime location is sometimes referred as *hotspot*, conceptualized at various geographical sizes. For example, Toole et. al. [35] use criminal offense records to identify spatio-temporal patterns at multiple scales. They employ various quantitative tools from mathematics and physics and identify significant correlation in both space and time in the crime behavioral data. Short et. al. [36] study the dynamics of crime hotspots and identify stable hotspots, where criminals are modeled as random walkers. Bogomolov et. al. [33] use human behavioral data derived from mobile network and demographic sources, together with open crime data to predict crime hotspots. They compare various classifiers and find random forests have the best prediction performance. Wang et. al. [30] use automatic semantic analysis to understand natural language in Tweets, from which the crime incidents are reported. Some other work [37], [38] employ kernel density estimation (KDE) to identify and analyze crime hotspots. Those studies form another form of crime prediction, which relies on the retrospective crime data to identify areas of high concentrations of crime. Nakaya et. al. [39] extend the crime cluster analysis with a temporal dimension. They employ the space-time variants of KDE to simultaneously visualize geographical extent and duration of crime clusters.

Population-centric paradigm. In the last paradigm, research focuses on the criminal profiling at individual and community levels. At the individual level, Wang et. al. [25] aim to automatically identify crimes committed by the same individual from a historical crime database. The proposed system, called *Series Finder*, is designed to find and classify the modus operandi (M.O.) of criminals. At the community level, Buczak et. al. [40] use fuzzy association rule mining to identify crime patterns. The rules they found are consistent across all regions. They identify association rules from population demographics in communities. In another paper, Traunmueller et. al. [32] use computational methods to validate various social theories at a large scale. They used mobile phone data in London, from which they mine the people dynamics as features to correlate with crime.

The problem we tackle is to estimate crime rate in a community, which is different from the first two categories of work, mainly because our innovation lies in using newer type of data to enhance the commonly used traditional counterparts. More specifically, we use POI to enhance the demographics information and use taxi flow as hyperlinks to enhance the geographical proximity correlation. Although in our problem we do not consider the temporal dimension of crime in depth, it could be a promising supplement to better profile crime. Our problem is not location prediction of any particular crime incident. Therefore the methods proposed in place-centric methods are not applicable in our problem. However, the features we propose may be incorporated in those crime prediction models.

Our approach falls into the third paradigm because we try to predict the crime rate for Chicago community areas. In our study, the community areas are well-defined and stable geographical regions. The newly proposed POI features and taxi flows provide new perspectives in advancing our understanding of crime rates across community areas.

It is worthy noting that our problem is different from the spatial interpolation (kriging) problem in geostatistics field [7]. Kriging method in general is used for spatial interpolation, where the goal is to estimate the value of a target variable at a certain location given the observations of the same variables on nearby locations [41]. The original kriging technique only involves one variable, and aims to interpolate missing observations of the target variable on a continuous plane [42]. Later on, various extensions on kriging are developed [43]. Among those extensions, regression-kriging [44] and co-kriging [45], [46] have been widely used in many applications, such as estimating soil nitrogen [47] and real-time precipitation prediction [48]. Both regression-kriging and co-kriging incorporate auxiliary variables to estimate the target variable. The co-kriging assumes that auxiliary variables strongly correlate with target variable and there are abundant observations of auxiliary variables. The regression kriging learns a regression function between auxiliary variables and target variables, and then applies kriging method to estimate the regression residuals. In our problem, the co-kriging method is not appropriate, because some of the auxiliary variables are not strongly correlate with crime rate by itself. The regression kriging method, on the other hand, could be used as a baseline for comparison. However, it is worth mentioning that the kriging method aims to interpolate missing values to minimize overall variance, which is in contrast with our goal of optimal prediction.

3 Overview

The crime data collected in Chicago has detailed information about time, location (i.e., latitude and longitude), and types of crime. In our problem, the term *crime count* refers to number of crime incidents in a region (i.e., community area) in a year. The *community area* is used as our geographical unit of study, since it is well-defined, historically recognized and stable over time [49]. In total, there are 77 community areas in Chicago. *Crime rate* is the crime count normalized by the population in a region. We use vector $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$ to denote the crime rates in regions. The crime rate inference problem is to estimate the crime rate in one region using the crime rate of other regions in the same year by considering the features of regions and correlations among regions.

The crime data of Chicago are obtained from the City of Chicago data portal [50]. Chicago is one of few cities with detailed crime data that are made public online. The crime dataset contains the incident date, location (street name and GPS coordinates), and primary type from year 2001 to 2015. In total there are $5{,}856{,}414$ recorded crime incidents over 15 years, or on average $390{,}417$ crimes incidents per year. We visualize the crime rate in Figure 2, from which we can see that the downtown area has the highest crime rate.

In this paper we study the crime rate inference problem. More specifically, we estimate the crime rate of some regions given the information of all the other regions. Without loss of generality, we assume there is one community area t with crime rate y_t missing, and we use the crime rate of all the other regions $\{y_i\}\backslash y_t$ to infer this missing value. Our problem is mathematically formalized as follows

$$\hat{y_t} = f(\{y_i\}\backslash y_t, X), \tag{1}$$

where X refers to observed extra information of all those community areas.

We consider two types of features X for inference:

- **Nodal features:** Nodal features describe the characteristics of the focal region. Such features include demographic information and Point-of-Interest (POI) distribution. Demographics are frequently used in literature, but POI is a newer type of big

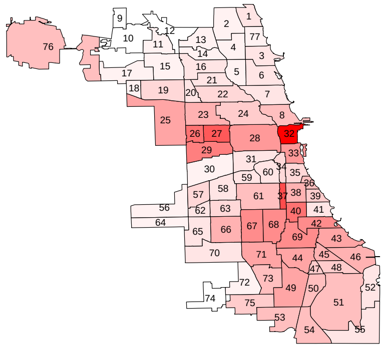


Fig. 2: Crime rate of Chicago by community areas. The community area #32 is Chicago downtown, which has the highest crime rate.

data, which we find significantly improve the crime inference accuracy.

- **Edge features:** (1) Geographical influence. Geographical influence considers the crime rate of the nearby locations. This feature has been extensively used in literature as well. To estimate the focal region, the crime rate of nearby regions are weighted according to spatial distances. (2) Hyperlink by taxi flow. Locations are connected through the frequent trips made by humans, which can be considered as the hyperlinks in space. This type of feature has not been previously studied in the criminology literature. We propose to use taxi trips to construct the social flow. Our hypothesis is that two regions that are more strongly connected through social flow will influence each other's crime rate.

In the following sections, we first discuss the inference models based on these two types of features in Section 4 and then discuss how to construct these features using the real-world data in Section 5.

4 Inference Model

4.1 Linear Regression

The most straightforward prediction model is linear regression. This model assumes that the error term for y_i follows a Gaussian distribution $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Equation (2) gives the linear regression formulation of our problem:

$$\mathbf{y} = \alpha^T X^N + \beta^f W^f \mathbf{y} + \beta^g W^g \mathbf{y} + \epsilon. \tag{2}$$

$X^N \in R^{d_N \times n}$ is the nodal feature matrix where column i is the nodal feature vector of region i, d_N is the dimension of nodal features, and n is the number of regions. Both demographic features and POI distribution features are included in X^N as nodal features. $W^f \in \mathbb{R}^{n \times n}$ is the flow matrix of taxi flow, and $W^g \in \mathbb{R}^{n \times n}$ is the spatial matrix representing the geographical adjacency. In addition, $\alpha \in \mathbb{R}^{d_N}$ and $\beta^f, \beta^g \in \mathbb{R}$ are the coefficients for corresponding features. Note that $\epsilon \in \mathbb{R}^n$ is the only stochastic variable on the right-hand side; all other terms are fixed observation values. Therefore, we incorporate all the fixed observations into one term $X \in \mathbb{R}^{(d_N+2) \times n}$, and we get the standard regression problem:

$$\mathbf{y} = \mathbf{w}^T X + \epsilon,$$

where $\mathbf{w} = [\alpha^T, \beta^f, \beta^g]^T$ is the concatenation of all coefficients.

4.2 Negative Binomial Regression

In our problem, we aim to infer the crime rate, which is guaranteed to be a non-negative integer. However, linear regression does not ensure this property. *Poisson regression* is another form of regression, more appropriate for non-negative data than linear regression [51], [52]. With shortened notation $\mathbf{x}_i$, which represents all features in a region, the Poisson regression model has the exponential function as link function

$$E(y_i) = e^{\mathbf{w}^T \mathbf{x}_i}. \tag{3}$$

In the following, we omit the index i wherever it is clear to refer to the variable of a single region. The link function comes from the assumption that y follows the Poisson distribution with mean λ. Additionally, the mean λ is determined by observed independent variables $\mathbf{x}$, i.e. $\lambda = e^{\mathbf{w}^T \mathbf{x}}$. Adding all together, the joint probability of y is

$$P(y|\mathbf{w}) = \frac{e^{-e^{\mathbf{w}^T \mathbf{x}}} (e^{\mathbf{w}^T \mathbf{x}})^y}{y!}. \tag{4}$$

However, Poisson regression enforces the mean and variance of dependent variable y to be equal. This restriction leads to the "over-dispersion" issue for some real problems, that is the presence of larger variability in data set than the statistical model expected. To address this, we use the Poisson-Gamma mixture model, which is also known as *negative binomial regression*. Negative binomial regression is frequently used in crime research [53].

Given that the crime rate y follows Poisson distribution with mean λ, in order to allow for larger variance, λ itself is a random variable having a Gamma distribution with shape k and scale $\theta = \frac{p}{1-p}$. The probability density function of y becomes

$$\begin{aligned} P(y|k,p) &= \int_0^\infty P_{Poisson}(y|\lambda) \cdot P_{Gamma}(\lambda|k,p) d\lambda \\ &= \int_0^\infty \frac{\lambda^y}{y!} e^{-\lambda} \cdot \lambda^{k-1} \frac{e^{-\lambda(1-p)/p}}{(\frac{p}{1-p})^k \Gamma(k)} d\lambda \\ &= \frac{\Gamma(k+y)}{y!\Gamma(k)} p^y (1-p)^k. \end{aligned} \tag{5}$$

This is exactly the probability density function of negative binomial distribution.

In negative binomial regression, the link function is

$$E(y) = e^{\mathbf{w}^T \mathbf{x} + \epsilon}. \tag{6}$$

The error term e^ϵ is the mixture prior from the Gamma distribution, and we assume its mean is 1, i.e. $E(e^\epsilon) = 1$. This setting ensures that $E(y) = e^{\mathbf{w}^T \mathbf{x}} \cdot e^\epsilon = e^{\mathbf{w}^T \mathbf{x}}$. Meanwhile, given the probability density function of negative binomial distribution in Equation (5), the mean of negative binomial distribution is $\frac{pk}{1-p}$. Combining the theoretical mean with the link function, we have $p = \frac{e^{\mathbf{w}^T \mathbf{x}}}{e^{\mathbf{w}^T \mathbf{x}} + k}$. Therefore, the probability mass function of y becomes

$$P(y|\mathbf{w}, k) = \frac{\Gamma(k+y)}{y!\Gamma(k)} \left(\frac{e^{\mathbf{w}^T \mathbf{x}}}{e^{\mathbf{w}^T \mathbf{x}} + k} \right)^y \left(\frac{k}{e^{\mathbf{w}^T \mathbf{x}} + k} \right)^k. \tag{7}$$

The log-likelihood function of negative binomial model is given in Equation (8), where w and θ can be estimated by maximizing likelihood.

$$\mathcal{L}(\mathbf{w}, k; \mathbf{y}, X) = \sum_{i=1}^{n} \left\{ y_i \ln \left(\frac{e^{\mathbf{w}^T \mathbf{x}_i}}{e^{\mathbf{w}^T \mathbf{x}_i} + k} \right) + k \ln \left(\frac{k}{e^{\mathbf{w}^T \mathbf{x}_i} + k} \right) + \ln \Gamma(y_i + k) - \ln \Gamma(y_i + 1) - \ln \Gamma(k) \right\}. \quad (8)$$

4.3 Non-Stationary Model

The two regression models described above assume the statistical correlations between crime rate and observed features are constant over space, because they learn one set of parameters for all community areas. In the real world, it is very likely that some statistical correlations between crime rate and observed features are not stationary over space. In this section we propose to apply a non-stationary model, called geographically weighted regression (GWR) [54], to capture the different crime correlations at different places.

Formally, a global spatial regression model such as the aforementioned two models has the following form

$$y = f(\mathbf{x}, \mathbf{w}), \quad (9)$$

where $\mathbf{w}$ is the parameter of the regression function f. Given a set of data points $\{y_i, \mathbf{x}_i\}_{i=1}^n$ sampled at locations $l_1, \cdots, l_n$, the maximum likelihood estimation of parameter $\mathbf{w}$ is given by

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \sum_{i=1}^{n} \mathcal{L}(y_i, f(\mathbf{x}_i, \mathbf{w})). \quad (10)$$

This global model is stationary, because the weights used for predictions are the same at all locations, when we fit the model to find the optimal parameter.

Instead, the GWR learns a *local* regression function f with parameter $\mathbf{w}_i$ at each location of interest l_i:

$$y = f(\mathbf{x}, \mathbf{w}_i), \quad \forall l_i \in \{l_1, l_2, \cdots, l_n\}, \quad (11)$$

where l_i is usually a geospatial coordinate in the two dimensional space. In order to train a lot of local models, we need a larger number of samples at each location l_i, which are usually not available. To address this issue, GWR uses the spatially nearby samples and gives each sample a weight according to the distance between sample point and target location l_i. The objective for the local model at location l_i is

$$\mathbf{w}_i^* = \arg\min_{\mathbf{w}_i} \sum_{j=1}^{n} \gamma_{ij} \mathcal{L}\left(y_j, f(\mathbf{x}_j, \mathbf{w}_i)\right), \quad (12)$$

where γ_{ij} is the spatial kernel to weight the neighboring data point at location l_j for regression model at location l_i.

Choice of spatial kernel γ. There are several spatial kernels we can choose from. The most straightforward solution is to exclude samples that are further away from target location. Namely,

$$\gamma_{ij} = \begin{cases} 1 & \text{if } d_{ij} < \tau \\ 0 & \text{otherwise,} \end{cases} \quad (13)$$

where d_{ij} is the distance between l_i and l_j, and τ is a distance threshold. Clearly, such a solution suffers from the discontinuity.

A better solution is to specify the weight γ as a continuous function of distance d, which is

$$\gamma_{ij} = \exp\left(-\frac{d_{ij}^2}{2h^2} \right), \quad (14)$$

where h is referred to as the bandwidth of the Gaussian kernel. Intuitively, when the samples are dense near the target location l_t, the h can be set smaller, so that we give lower weights to those samples far away. On the other hand, if the samples are sparse, h should be set larger, so that we consider those further away samples as well to train our model. When h is set to infinity, the GWR becomes a global model, since all samples have equal weight 1.

One issue with the Gaussian kernel in Equation (14) is that when the samples are dense, it over-smooths local models by considering too many samples at each location. A popular alternative kernel utilizes the bi-square function,

$$\gamma_{ij} = \begin{cases} \left(1 - \frac{d_{ij}^2}{\tau^2}\right)^2 & \text{if } d_{ij} < \tau \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

The bi-square kernel functions provides continuous weight for samples up to distance τ. In our problem, since we do not have too many samples available, therefore we use the Gaussian kernel, and in experiment we will show the bandwidth tuning to get the best results.

Applying GWR on existing methods. The GWR is more like a framework rather than a method, which can be applied to many existing regression methods. The classic GWR is applied to linear regression model resulting in the following objective for location l_i

$$\mathbf{w}_i^* = \arg\min_{w_i} \sum_{j=1}^{n} \gamma_{ij} (y_j - \mathbf{w}_i^T \mathbf{x}_j)^2. \quad (16)$$

Similarly, the GWR framework can be employed with the negative binomial regression model, and we call this *geographically weighted negative binomial regression* (GWNBR). Here, the objective for model at location l_i is to optimize the weighted log-likelihood function:

$$\mathcal{L}(\mathbf{w}_i, k_i; \mathbf{y}, X) = \sum_{j=1}^{n} \gamma_{ij} \left\{ y_j \ln \left(\frac{e^{\mathbf{w}_i^T \mathbf{x}_j}}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} \right) + k_i \ln \left(\frac{k_i}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} \right) + \ln \Gamma(y_j + k_i) - \ln \Gamma(y_j + 1) - \ln \Gamma(k_i) \right\}. \quad (17)$$

4.4 Optimization

The objective in Equation (17) can be solved using a block coordinate gradient descent method, by alternatively solving $\mathbf{w}_i$ and k_i. Details for solving each step are given below.

Fix k_i, solve $\mathbf{w}_i$:

When k_i is fixed, the objective function can be simplified as follows:

$$\min_{\mathbf{w}_i} - \sum_{j=1}^{n} \gamma_{ij} \left\{ y_j \ln \left(\frac{e^{\mathbf{w}_i^T \mathbf{x}_j}}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} \right) + k_i \ln \left(\frac{k_i}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} \right) \right\}. \quad (18)$$

The gradient is:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_i} = \sum_{i=1}^{n} \gamma_{ij} \{ y_i \frac{\mathbf{x}_j e^{\mathbf{w}_i^T \mathbf{x}_j}}{e^{\mathbf{w}_i^T \mathbf{x}_j}} - \frac{(y_i + k_i)\mathbf{x}_j e^{\mathbf{w}_i^T \mathbf{x}_j}}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} \}. \tag{19}$$

Then,

$$\mathbf{w}_i^t = \mathbf{w}_i^{t-1} + \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{w}_i^{t-1}}. \tag{20}$$

Fix $\mathbf{w}_i$, solve k_i:
When $\mathbf{w}_i$ is fixed, the objective function becomes:

$$\begin{aligned} \min_{k_i} = & \\ & - \sum_{j=1}^{n} \gamma_{ij} \left\{ y_j \ln \left(\frac{e^{\mathbf{w}_i^T \mathbf{x}_j}}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} \right) + k_i \ln \left(\frac{k_i}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} \right) \right. \\ & \left. + \ln \Gamma(y_j + k_i) - \ln \Gamma(k_i) \right\}. \end{aligned} \tag{21}$$

The gradient is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial k_i} = & \sum_{i=1}^{n} \gamma_{ij} \{ \frac{y_i}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} + \ln(\frac{k_i}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i}) \\ & - \frac{k_i}{e^{\mathbf{w}_i^T \mathbf{x}_j} + k_i} + 1 + \psi(y_j + k_i) - \psi(k_i) + y_i \}, \end{aligned} \tag{22}$$

where $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is digamma function. Then,

$$k_i^t = k_i^{t-1} + \alpha \frac{\partial \mathcal{L}}{\partial k_i^{t-1}}. \tag{23}$$

5 Feature Extraction

In this section, we will discuss the details of features used in our method. In the literature, the most commonly used features are demographics and geographical influence. We also extract two types of new features from Point-Of-Interest data and taxi flow data. Below we describe the datasets used to construct features and the characteristics of these features.

5.1 Nodal Feature: Demographics

Socioeconomic and demographic features of neighborhoods have been widely used to predict crime [33], [55], [56], [57]. Previous studies have shown that crime rate correlates with certain demographics. For example, scholars [6], [58] suggest that population diversity leads to less crime in certain neighborhoods. In our study, we include demographic information from the US Census Bureau's Decennial Census [59]. Using 2010 census information would overlap with the time in which crime is measured. Instead, we use year 2000 demographic data because we are interested in predictors that precede temporally the period in which crime rates are evaluated. The demographics include the following features:

total population, population density, poverty index, disadvantage index, residential stability, ethnic diversity, race distribution.

The poverty index measures the proportion of community area residents with income below the poverty level. The disadvantage index is a composite scale based on prior work [60], a function of poverty, unemployment rate, proportions of families with public assistance income, and proportion of female headed households. The residential stability measures home ownership (the proportion of owner occupied housing units over all occupied housing units) and the proportion of residents 5 years old and older who resided in the same house 5 years earlier [49], [61]. Racial and ethnic diversity is an index of heterogeneity [6] based on the combination of Hispanic vs. non-Hispanic and the racial categories. Those who checked Hispanics are included in that category independent of what race they noted. Those who checked non-Hispanic are separated by race (White, Black, Asians, Pacific Islanders, and Others). In 2000, respondents could pick more than one racial categories. Our categories reflect the corresponding race noted as the only one. If respondents checked more than one race, they were included in the category of "Others".

Figure 3 visualizes the crime rate and demographics features in Chicago by community areas. Comparing with Figure 2, it is clear that the crime rate and poverty index and disadvantage index are consistent, the ethnic diversity shows an inverse correlation, and the total population has little correlation with crime.

Table 1 shows the Pearson correlation coefficient between various demographics features and the crime rate at community area level. The corresponding p-value is also calculated and shown in the table to indicate the significance of the correlation coefficient. There are in total 77 community areas in Chicago. We can see that the poverty index and disadvantage index positively and strongly correlate with crime, while the ethnic diversity negatively correlates with crime. Other features such as total population, population density, and residential stability have weaker correlations. One counter-intuitive observation is that the total population has a weak and negative correlation with crime. The reason is that we use crime rate in each community area, which is already normalized by the population, and therefore the total population and population density have less impact.

TABLE 1: Pearson correlation between demographic features and crime rate (* indicates significant correlations with p-value less than 5%).

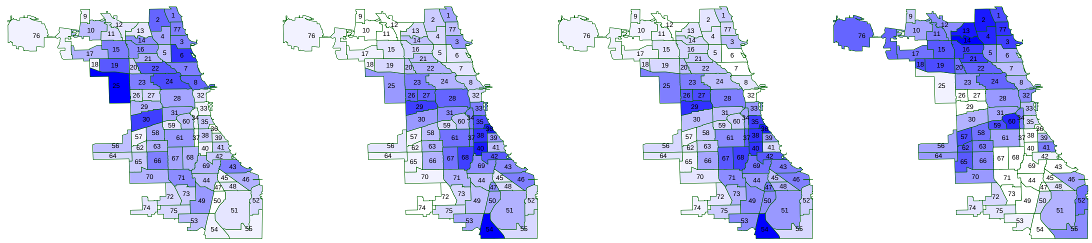
Feature	Correlation	p-value
Total Population	-0.1269	0.2716
Population Density	-0.1972	0.0855
Poverty Index	**0.5573***	1.403e-07
Disadvantage Index	**0.5959***	1.082e-08
Residential Stability	-0.0453	0.6965
Ethnic Diversity	**-0.5545***	1.678e-07
Percentage of Black	**0.6696***	2.779e-11
Percentage of Hispanic	**-0.3820***	0.0006

5.2 Nodal Feature: Point-of-Interest (POI)

While demographics are traditional census data, POI is a type of modern data that provide fine-grained information about locations. We collect POI from FourSquare [62]. POI data from FourSquare provide the venue information including venue name, category, number of check-ins, and number of unique visitors. We mainly use the major category information because categories can characterize the neighborhood functions. There are 10 major categories defined by FourSquare:

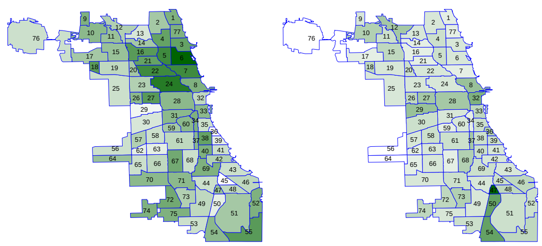
food, residence, travel, arts & entertainment, outdoors & recreation, college & education, nightlife, professional, shops, and event.

In total, we have crawled 112,000 POIs from FourSquare for Chicago. Most of these POIs are in the downtown area of Chicago. For the purpose of visualization, we normalize the POIs count per category by the total POI count in a neighborhood and plot two



(a) Total population (b) Poverty index (c) Disadvantage index (d) Ethnic diversity

Fig. 3: (a)-(d) Demographics in Chicago by community areas. Darker colors indicate higher values. Features are normalized into $[0, 1]$.



(a) Nightlife (b) Professional

Fig. 4: POI ratio per neighborhood. The saturation of color is proportional to the ratio value. The "professional" category distribution is more consistent with the crime distribution, and therefore it is the most correlated with crime. Meanwhile, the "nightlife" category is negatively correlated with Chicago crime. The POI ratios are independently normalized for different POI categories.

selected categories, i.e. nightlife and professional, in Figure 4. The darker colored neighborhoods in Figure 4 are the ones with a higher proportion of residence POIs.

TABLE 2: Pearson correlation between POI category and crime rate (* indicates significant correlations with p-value less than 5%).

POI category	Correlation	p-value
Food	-0.1543	0.1803
Residence	-0.0610	0.5984
Travel	-0.0017	0.9883
Arts & Entertainment	-0.0049	0.9661
Outdoors & Recreation	0.0668	0.5637
College & Education	-0.0078	0.9473
Nightlife	-0.1553	0.1775
Professional	**0.3221***	0.0043
Shops	-0.1676	0.1450
Event	0.2196	0.0549

In Table 2 we show the Pearson correlation between POI category and crime rate. The category "professional" is most significantly correlated with the crime rate. Under the professional POI category, there are some venues with a large population concentration, such as transportation center, convention center, community center, and co-working space. In those venues, the population volume is high and residential stability is low, therefore the professional POI counts positively correlates with crime rate. One counter-intuitive observation is that "nightlife" category is not positively correlated with crime (-0.1553). This can be seen in Figure 4(a). The majority of nightlife venues in Chicago are located in the northern area, while most crime incidents occur in the downtown area.

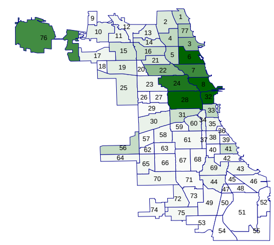


Fig. 5: Absolute POI count distribution. In our crawled POI dataset, most community areas have less than 100 venues. Meanwhile, the downtown area there are over $10,000$ venues for one community area, e.g. #8, #32.

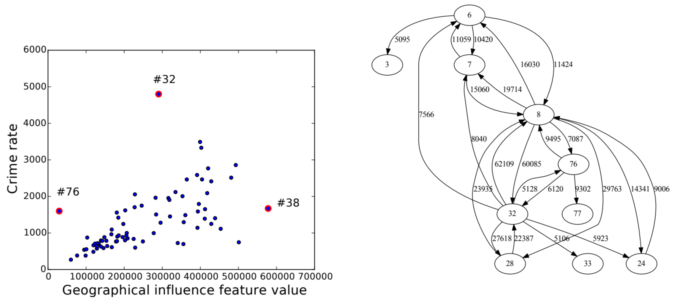
There are different ways to use the POI data. The straightforward definition of POI distribution is calculated by normalizing the POI count in each category by the total POI counts. However, the POIs in Chicago are not evenly distributed. As shown in Figure 5, most POIs are in the downtown area and some areas only have a few POIs. If normalized by the total number of POIs in a neighborhood, two neighborhoods may show similar distributions but they are quite different. For example, a downtown neighborhood and a distant neighborhood may both have a high ratio of the food category but the downtown neighborhood has many more POIs in total and is more dynamic in population constitution. Therefore, using the raw count instead of normalized distribution is more effective. This is also demonstrated in estimation accuracy as shown in Table 3, where the POI count feature has a performance gain of 10% over POI percentage feature.

TABLE 3: Using POI count instead of POI percentage improve the estimation accuracy. Estimation for crime in 2014 with all other features.

Scheme	MAE	MRE
Count	272.51	0.268
Percentage	302.61	0.298

5.3 Edge: Geographical Influence

Together with the US census demographics data, we also collected the boundary shape files of Chicago, which are used to calculate the geographical influence feature. Previous studies have also



(a) (b) (c)

Fig. 6: **(a)** The correlation between geographical influence feature $\mathbf{f}^g$ and crime rate y. We marked out three outliers and their corresponding community area ID. **(b)** Major taxi flows between neighborhoods from January to March in 2014. The number on the edge is the traffic volume. We set a threshold of 5000 and only plot high volume flows. The label on a node is the corresponding community area ID. We can see that there are several hub community areas, such as #6, #8, #32, which are all in the downtown areas. **(c)** Correlation between taxi flow feature $\mathbf{f}^f$ and crime rate y. We marked out two outliers and their corresponding community area ID.

shown that the crime rate at one location is highly correlated with nearby locations [63], [64]. Such geographical influence is also frequently used in the literature [65], [66]. It is calculated as:

$$\mathbf{f}^g = W^g \cdot \mathbf{y}, \tag{24}$$

where W^g is the spatial weight matrix. If region i and j are not spatially adjacent, $w^g_{ij} = 0$; otherwise, $w^g_{ij} \propto distance(i,j)^{-1}$. Here $distance$ refers to the distance function to calculate the distance between two regions. Without loss of generality, in this paper we use the euclidean distance between the centroids of two regions as $distance(i,j)$.

In Figure 6a, we plot crime rate with respect to geographical influence calculated in Equation (24). We observe an obvious positive correlation, which means if nearby neighborhoods have a high crime rate, the focal neighborhood is more likely to have a high crime rate. We also observe a few outliers in Figure 6a. These neighborhoods have very different crime rates in their nearby neighborhoods compared to their own. For example, in Figure 2, community area #38 locates in an area where the neighbors have high crime rates but its crime rate is relatively low; in contrast, neighborhood #32 has a high crime rate even though its neighbors have relatively low crime. The community area #76, home of the O'Hare International Airport, is far from most of other community areas, however its own crime rate is relatively high.

5.4 Edge: Hyperlinks by Taxi Flow

In our Chicago taxi dataset, there are $1{,}038{,}476$ taxi trips in total from January to March in 2014[1]. For each trip, the following information are available: pickup/dropoff time, pickup/dropoff location, operation time, and total amount paid. We requested the taxi trip records from Chicago under the Illinois Freedom of Information Act. Figure 6b shows a visualization of the major flows at community level.

One of our hypotheses is that the social interaction among two community areas propagates crime from one region to another. The Chicago taxi data captures the social interactions among various community areas. To calculate this, we first map all taxi trips to community areas to get the taxi flow $w^f_{ij}, \forall i,j \in \{1,2,\cdots n\}$. Then the taxi flow lag is constructed by the product of taxi flow and the crime rate of neighboring regions as follows

$$\mathbf{f}^f = W^f \cdot \mathbf{y}. \tag{25}$$

The taxi flow W^f is a matrix with entry w^f_{ij} denoting the number of taxi trips from j to i. Note that $\forall i$, $w^f_{ii} = 0$ in matrix W^f, because we have to exclude the crime in the focal area from its own predictor. The semantic of this taxi flow feature is how much crime in the focal area is contributed by its neighboring areas through social interaction.

The correlation between taxi flow and crime rate is shown in Figure 6c. From the scatter plot, we can see that overall the crime rate is positively correlated with the taxi flow. There are two outliers clearly shown in Figure 6c. The community area #32 is downtown, which has the highest crime rate and is hard to predict by taxi flow. Another anomalous community area (#47) has relatively low crime rate by itself. However, this area has a lot of inflows from high-crime communities.

6 Experiments

6.1 Settings

We use four types of data as features, including demographics, POI, geographical influence, and taxi flow, to predict the total crime rates. The details of feature data are described in Section 5, and a description of crime data is available in Section 3. We conduct the crime prediction on five consecutive years, 2010 – 2014. There are over 30 categories of crime, and many categories have sparse values over regions. Therefore, we only study the effect of crime categories in Section 6.7, and in the rest experiments, we predict the total crime rate.

The following four methods, e.g. regression kriging (RK), linear regression (LR), negative binomial regression (NB), and geographically weighted negative binomial regression (GWNBR), are evaluated. The regression kriging method employs a regression model to incorporate all features as auxiliary variables, and then applies kriging to estimate the regression residual. We add kriging as one of the baseline, mainly because kriging is one of the most widely used method for geospatial interpolation [41].

1. This taxi dataset is different from the taxi dataset used in our KDD paper. In KDD we used taxi trips from October to December in 2013. We changed to a newer dataset, because there are many taxi trips in 2013 having wrong destination information (they report the same coordinates outside of USA).

We evaluate the estimation accuracy under various feature combinations. The bandwidth of Gaussian kernel h used for GWNBR is tuned separately under different settings. Refer to Section 6.4 for more details on parameter tuning.

We adopt leave-one-out evaluation to estimate the crime rate of one geographic region given all the information of all the other regions. When we construct the spatial/social lag variable for the training data, the effect of testing region is completely removed. For example, if region y_t is the testing region, the remaining $\{y_i|\forall i \text{ s.t. } i \neq t\}$ become the training set. For any y_i in the training set, its geographical influence feature and taxi flow feature are constructed only from $\{y_i|\forall i \text{ s.t. } i \neq t\}$.

In the evaluation, we estimate the crime rate for testing community areas. The accuracy of estimation is evaluated by mean absolute error (MAE) and mean relative error (MRE).

$$MAE = \frac{\sum_i |y_i - \hat{y}_i|}{n} \tag{26}$$

$$MRE = \frac{\sum_i |y_i - \hat{y}_i|}{\sum_i y_i} \tag{27}$$

6.2 Performance on Different Feature Combinations

The leave-one-out evaluation results are shown in Table 4. Ideally, we should show all the possible combinations of feature groups, which will result in 15 combinations. Due to the space limit, we show 4 combinations in Table 4. Setting 1 (**D+G**) represents the traditional features used in criminology literature. Setting 2 (**D+G+P**) and setting 3 (**D+G+T**) are used to examine the individual effects of new features on POI and Taxi flows. Setting 4 (**D+G+P+T**) studies the combined effects of all features.

6.2.1 POI Feature

Adding POI features improves the accuracy of the NB model (see setting 2 vs. setting 1 and setting 4 vs. setting 3 in Table 4). The POI distribution reflects the functionality of a region. The most correlated POI major category is "professional", under which there are a lot of venues like transportation center and conventional center. These are locations with more dynamic movements of people. Such location information is not reflected in any of other features. POI thus provides unique information and it shows that using big data can benefit us in advancing the study of traditional crime inference problems.

Another issue worth discussing is whether POI is a surrogate of population features from demographics. That is, a region with more POIs is a region with higher population. However, as we see from Table 4, the NB model using POI feature in addition to the demographic and geographic features always outperforms the model without the POI feature. This is because population from demographics reflects the number of residents in that region, but POI reflects dynamics of population (e.g., people go to venues for food, entertainment, or travel). Therefore, the dynamic population in POI further complements the residential population in demographics.

6.2.2 Taxi Flow

To test our hypothesis that crimes do not only correlate with nearby regions but also correlate through hyperlinks on the space (i.e., the taxi flow), we examine if considering taxi flow improves the inference accuracy. Comparing setting 3 with setting 1 in Table 4, we find that the improvement by taxi flow is not obvious for NB. However, comparing setting 4 with setting 2, we observe a much more significant accuracy boost. The reason could be that the taxi flow further complements the POI data. When POI information is missing from the predictor, the city dynamics captured by taxi flow are weakened as well.

Meanwhile, we note that incorporating the new urban data (POIs and taxi flow) significantly improves the performance of the NB model. Specifically, comparing setting 4 with setting 1, the improvements in MRE are 6.6%, 5.1%, 4.3%, 5.7%, and 5.7% for the five years, respectively.

TABLE 4: Performance evaluation on total crime rate prediction. Various feature combinations are shown in each column. The linear regression model, negative binomial model, and geographically weighted negative binomial model are compared by year group.

			Settings			
Column ID			1	2	3	4
Features[1]		**D**emo	✓	✓	✓	✓
		Geo	✓	✓	✓	✓
		POI		✓		✓
		Taxi			✓	✓
Year	Model[2]	Error				
2010	RK	MAE	463.68	461.67	452.39	421.59
		MRE	0.346	0.344	0.337	0.314
	LR	MAE	394.78	432.45	413.80	404.00
		MRE	0.295	0.323	0.309	0.301
	NB	MAE	402.82	343.55	404.49	315.28
		MRE	0.301	0.256	0.302	0.235
	GWNBR	MAE	343.44	332.38	327.20	**265.93**
		MRE	0.256	0.248	0.244	**0.198**
2011	RK	MAE	460.83	422.01	424.05	430.77
		MRE	0.358	0.328	0.330	0.335
	LR	MAE	380.08	422.94	402.63	402.81
		MRE	0.296	0.329	0.313	0.313
	NB	MAE	391.26	340.49	396.59	325.34
		MRE	0.304	0.265	0.308	0.253
	GWNBR	MAE	330.04	325.56	322.77	**275.61**
		MRE	0.257	0.253	0.251	**0.214**
2012	RK	MAE	455.94	418.04	438.13	412.55
		MRE	0.368	0.338	0.354	0.333
	LR	MAE	375.62	423.88	402.69	404.06
		MRE	0.304	0.343	0.325	0.327
	NB	MAE	395.82	349.09	399.99	342.41
		MRE	0.320	0.282	0.323	0.277
	GWNBR	MAE	332.70	342.95	334.83	**330.51**
		MRE	0.269	0.277	0.270	**0.267**
2013	RK	MAE	408.90	468.39	457.20	413.77
		MRE	0.360	0.412	0.402	0.363
	LR	MAE	369.24	433.48	387.58	400.75
		MRE	0.325	0.381	0.341	0.352
	NB	MAE	390.09	348.16	389.79	324.94
		MRE	0.343	0.306	0.343	0.286
	GWNBR	MAE	319.75	334.58	316.70	**316.34**
		MRE	0.281	0.294	0.279	**0.278**
2014	RK	MAE	404.76	365.14	453.07	380.83
		MRE	0.398	0.359	0.446	0.374
	LR	MAE	329.93	386.9	355.83	358.47
		MRE	0.325	0.381	0.35	0.353
	NB	MAE	347.39	303.59	348.71	289.52
		MRE	0.342	0.299	0.343	0.285
	GWNBR	MAE	282.35	280.58	274.16	**272.51**
		MRE	0.277	0.276	0.270	**0.268**

[1] Demo – demographic features, Geo – geographical influence, POI – POI features, Taxi – taxi flow feature.
[2] RK – Regression Kriging, LR – Linear Regression, NB – Negative Binomial Regression, GWNBR – Geographically Weighted Negative Binomial Regression.

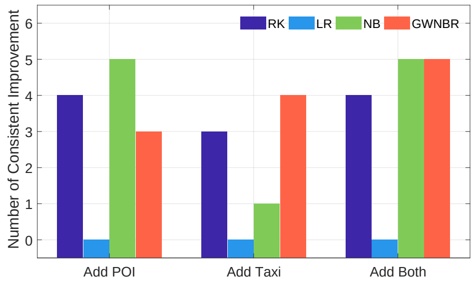


Fig. 7: The feature consistency results. The x-axis shows the effect of adding different features. The y-axis shows number of years that shows improvement.

6.2.3 Feature Consistency

In this section, we study the consistency of features as shown in Figure 7. We get the effect of adding POI feature by comparing setting 2 and setting 1 in Table 4. Similarly, setting 3 vs setting 1 shows the effect of adding taxi, and setting 4 vs setting 1 shows the effect of adding both. For each method we show number of years in y-axis, in which the added feature improves the model performance.

We observe that adding both features shows the best performance for both NB and GWNBR, meanwhile it is not necessarily the best for other models. For example, LR shows that setting 1 is better than setting 4 over different years. RK shows that setting 1 is better than setting 4 in year 2013 but not other years. We argue that it is crucial to use the right model to captures extra information and achieves consistent results.

6.3 Performance on Different Methods

In this section, we compare the prediction error of different methods, and we have the following observations.

6.3.1 Regression Kriging vs. Other Regressions

We observe that kriging method usually performs worse than other regression methods. The reason is that the kriging method is designed for interpolation and the objective is to minimize the estimation variance. Kriging method usually overestimates a local minimum and underestimates a local maximum due to the fact the kriging uses average to interpolate. For the crime rate prediction problem, other regression methods directly optimize the prediction error, and therefore outperforms the kriging method.

6.3.2 Negative Binomial Regression vs. Linear Regression

In Table 4, we can see that under most settings, the negative binomial regression significantly outperforms the linear regression (with only a few exceptions when using demographic features and geographic features). When using all the features, NB is significantly better than LR with at least 5% improvement in MRE. There are two reasons why the negative binomial regression is more appropriate for crime rate estimation than linear regression. First, negative binomial regression guarantees the prediction variable is non-negative. Second, it is difficult to get very precise estimates of crime rate, and the negative binomial regression allows a large variance in the estimated crime rate.

6.3.3 GWNBR vs. NB, and the Effectiveness of Non-Stationary Model

Next, we compare the negative binomial regression with the geographically weighed negative binomial regression. As shown in Table 4, the GWNBR model consistently outperforms the NB model in all experiment settings, which validate our hypothesis that the correlation among crime rate and other features are non-stationary.

In addition, comparing setting 2 (**D+G+P**) or setting 3 (**D+G+T**) with setting 1 (**D+G**) in Table 4, we observe that the performance improvement for GWNBR by POI feature or taxi flow separately is not obvious. However, when all features (**D+G+P+T**) are used, GWNBR consistently gives lower estimation error than using the traditional features only (**D+G**). In the best case (years 2010 and 2011), GWNBR reduces the MAE by over 15%. This again suggests that the POI feature and taxi flow complement each other, and that incorporating both features yields the best inference accuracy.

In view of the superior performance of GWNBR, in all the following experiments we only refer to the performance of GWNBR.

6.4 Parameter Sensitivity

In the GWNBR model, the bandwidth parameter h in Equation (14) controls the influence of a nearby training sample. There are several approaches to determine the bandwidth h [67]. Here, we adopt the cross validation approach to estimate a best h from data. More specifically, we fit a model on the training data and report the prediction error on the testing data. The best h should lead to the lowest prediction error.

In this experiment, we use data from year 2010 to study the effect the bandwidth h on the performance of GWNBR. Using a data-driven approach, we adopt the mean absolute error as a measure of fit. In Figure 8, we plot the MAE against bandwidth under leave-one-out setting. It is clear that the optimal bandwidth is 5.8, which gives us the lowest MAE. Furthermore, we observe that when h becomes larger, the performance of GWNBR model approaches to the NB model (with a MAE of 310). This observation is consistent with the model, because an infinitely large bandwidth gives all samples the same weight, which makes the GWNBR essentially an NB model.

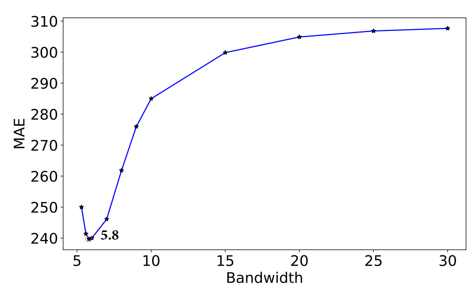


Fig. 8: Bandwidth sensitivity analysis for geographically weighted negative binomial regression.

6.5 Feature Importance

In this section, we study the importance of features through significance tests.

From previous results, we see that combining POI features and taxi flow will help improve the estimation accuracy. Now we try to measure the significance of this accuracy boost by permutation tests. If a feature correlates with crime, when we randomly permute the values of this feature among neighborhoods, we will

TABLE 5: Estimated p-value for each feature. The p-value is defined as the possibility that a smaller error measure is observed under the null hypothesis. Permutation test is conducted on 2014 data with all features (D+G+P+T) used.

Settings: D+G+P+T	LR		NB		GWNBR	
	MAE	MRE	MAE	MRE	MAE	MRE
	358.47	0.353	289.52	0.285	272.51	0.268
Feature	p-value					
D	0.000	0.000	0.000	0.000	0.000	0.000
G	0.001	0.001	0.005	0.005	0.004	0.004
P	0.022	0.021	0.005	0.005	0.006	0.007
T	0.000	0.000	0.080	0.079	0.049	0.052

expect a higher error in crime rate estimation. So in each round of permutation, we can get an error in estimation. We compare the error with the original feature to the error distribution obtained from permutations. We conduct 1,000 rounds of permutations to approximately estimate the error distribution. The position of the original error in this distribution indicates the significance of this feature. For example, if the original error is smaller than 99% of the errors from the permutations, the p-value is 0.01.

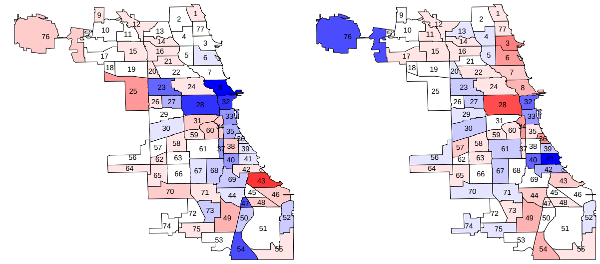
In Table 5, the p-values of different features are reported. The demographics feature is the most significant feature with estimated p-value being 0.00. In all the $1,000$ random permutations of demographic feature, we never observe an error lower than the original error. The proposed POI distribution and taxi flow are significant as well, with p-values of 0.6% and 4.9% for the GWNBR. We notice that the taxi flow has a p-value around 5% instead of close to 0. One reason is that the taxi flow overlaps with geographic feature. Thus, permuting taxi flow may not have a critical influence on the estimation error in certain cases.

6.6 Improvements in Different Regions

The POI distributions and taxi flow patterns are different from region to region. It is interesting to find out whether they have a consistently positive impact in crime rate estimation. For POI feature, we calculate the difference in estimation error (MAE) between two setting 3 (D+G+T) and setting 4 (D+G+P+T). Similarly, the MAE difference between setting 2 and setting 4 is calculated for the taxi flow feature. The results are shown in Figure 9. A positive difference (blue area) indicates that adding the new feature helps reduce the estimation error, while a negative difference (red area) indicates that the new feature adds more noise to the data.

It is interesting to observe that in the downtown area, i.e. community areas #8, #32, and #28, POI significantly improves the estimation accuracy. The reason is two fold. 1) The demographics information from census is mostly about the residing population in the focal area. However, in the downtown area there are a lot of floating population groups conducting various social activities, and this is not reflected by the census demographics. The POI information, on the other hand, reflects the functionality of a region, hence complements the demographic information. 2) In the downtown area, there are more POIs than other places, which provide more complete information about the community profile.

As for the taxi flow feature, it helps the most in those suburb area, because the taxi flow reflects the social interaction in those areas. In the downtown area #28 and #8, the taxi flow feature incurs a relatively large estimation error. The reason is that the taxi flow distribution in Chicago is extremely skewed. Roughly 61% of the Chicago taxi trips have a destination in the downtown area, which may result in the model over-propagating crime estimates from all of Chicago into the downtown area.



(a) POI (b) Taxi flow

Fig. 9: Performance improvement per region in year 2014. **Left:** The MAE difference of setting 3 and 4. **Right** The MAE difference of setting 2 and 4. The color blue means the MAE is reduced by adding the corresponding feature, while color red means the MAE is increased. The color saturation indicates the value of difference.

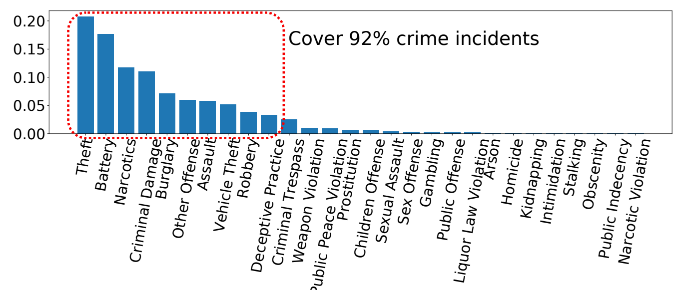


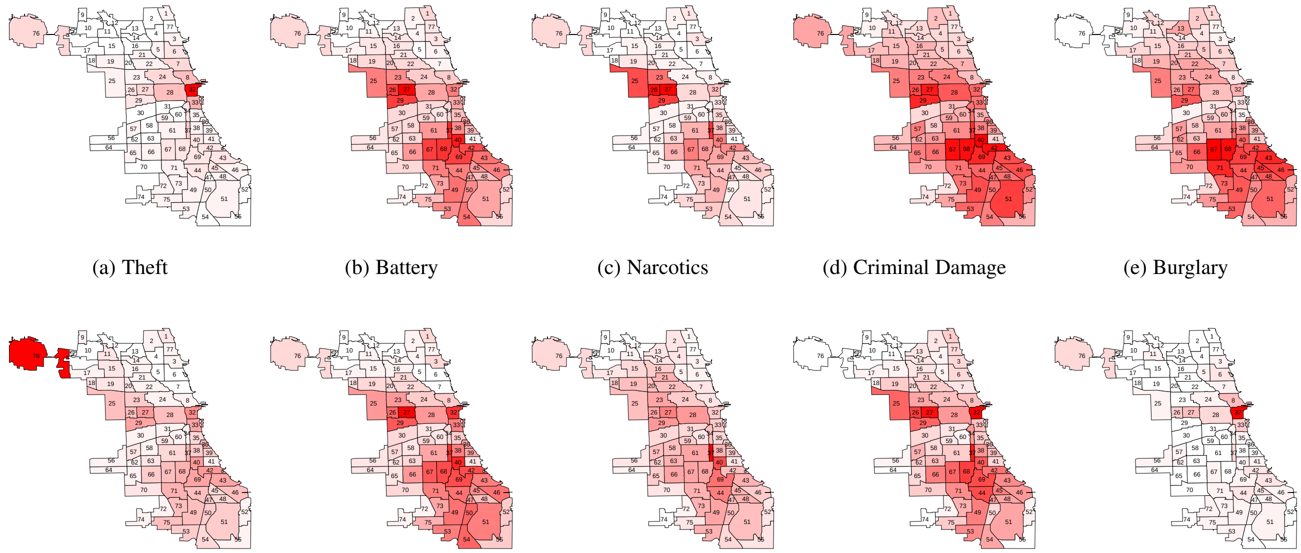
Fig. 10: Crime rate distribution of different categories in year 2014. Top-10 categories cover 92% of total crime incidents.

6.7 Experiments on Different Crime Categories

In this section, we evaluate the accuracy of crime rate estimation for different crime categories. The crime data consists of 28 crime categories. The percentage of each category is shown in Figure 10, where we can see that theft is the top crime category taking 22% of all crimes. Top-10 categories cover 92% of total crime, so we only focus on these categories in this paper.

In Table 6, we show the performance of GWNBR on the top-10 crime categories under various feature settings. There are some interesting results that are different from the total crime. The main reason is that different crime categories usually have very different spatial distribution. Figure 11 shows the distributions for top-10 categories.

First, comparing setting 2 to setting 1, POI features make the results worse on the following categories: narcotics, criminal damage, burglary, and motor vehicle theft. From Figure 11 we can see the distributions of these categories are different from the total crime distribution. The most notable difference is that downtown area is a center for overall crime, but not for these categories. Recall that downtown areas have the highest POIs count from Figure 5. Therefore, the POI features actually correlate less with those categories. In Table 7, we quantitatively measure the correlation between crime rate and POI features. Since the POI features are a vector and the crime rate is a scalar, we use a pairwise setting to calculate the correlation. More specifically, we report the Spearman correlation of pairwise difference on crime rates and the pairwise cosine similarity on POI features. The Spearman correlation is typically negative, because if two communities have



(a) Theft (b) Battery (c) Narcotics (d) Criminal Damage (e) Burglary

(f) Other Offense (g) Assault (h) Motor Vehicle Theft (i) Robbery (j) Deceptive Practice

Fig. 11: (a)-(j) Crime rates of overall and top-10 crime categories in Chicago by community areas in year 2014. Darker colors indicate higher values.

TABLE 6: Performance of GWNBR on top-10 crime categories in year 2014.

		Settings			
Column ID		1	2	3	4
Features	**Demo**	✓	✓	✓	✓
	Geo	✓	✓	✓	✓
	POI		✓		✓
	Taxi			✓	✓
Category	Error				
Theft	MAE	97.64	**88.01**	97.88	92.83
	MRE	0.347	**0.313**	0.348	0.330
Battery	MAE	59.65	57.74	59.95	**55.30**
	MRE	0.251	0.243	0.252	**0.232**
Narcotics	MAE	61.58	65.05	**59.54**	62.38
	MRE	0.416	0.439	**0.402**	0.421
Criminal Damage	MAE	29.37	32.65	**29.11**	32.40
	MRE	0.200	0.222	**0.198**	0.221
Burglary	MAE	21.90	23.49	**21.79**	23.13
	MRE	0.237	0.254	**0.235**	0.250
Other Offense	MAE	19.93	**16.68**	20.20	17.03
	MRE	0.241	**0.202**	0.244	0.206
Assault	MAE	19.48	18.37	18.93	**16.36**
	MRE	0.240	0.226	0.233	**0.202**
Motor Vehicle Theft	MAE	19.79	21.42	**19.60**	21.61
	MRE	0.289	0.313	**0.286**	0.316
Robbery	MAE	**18.94**	25.97	19.18	23.27
	MRE	**0.376**	0.516	0.381	0.452
Deceptive Practice	MAE	19.82	**15.15**	20.12	15.19
	MRE	0.428	**0.327**	0.435	0.328

similar POI distributions (large cosine similarity), then their crime rate difference should be small. Overall, we observe that crime rates in narcotics, criminal damage, burglary, and motor vehicle theft have small or close to 0 correlations with POI distributions.

Second, under the theft, deceptive practice, and other offense crime categories, **D+G+P** (setting 2) achieves the best performance. From Table 7 we can see that the correlations between POI and crime rate of theft, deceptive practice, and other offense are actually much stronger than other crime categories. The total crime has a correlation of -0.109, and all these three categories have much lower correlation values around -0.2. This demonstrates that the POI feature is a dominant feature in crime prediction for these three categories. As a result, the **D+G+P** feature setting has the best performance for these three categories.

Third, under narcotics, criminal damage, burglary, and motor vehicle theft crime categories, **D+G+T** (setting 3) gives the best performance. These crime categories usually have a high biased distribution toward suburb area according to Figure 11. Recall from Figure 9 that taxi flow helps the most in suburb area, because the social interactions in those areas have a significant influence on crime.

Fourth, the robbery is an anomaly category, where neither POI features nor taxi flow features improve the inference accuracy. While this anomaly is hard to explain due to the limitation of our data, we should note that in most cases the POI features and taxi flow features are indeed helpful for crime rate inference.

7 Conclusion

In the social science literature, demographic factors and geographic neighbors are known to exhibit strong correlations with crime. In this paper we address the problem of crime rate inference using new features from urban data. More specifically, we propose to use POI features to complement the demographic features, and to use taxi flow as hyperlinks to supplement the geographical influence. The intuition behind the POI features is that the POI distributions of community areas profile the region functionality. The intuition behind the hyperlinks is that the taxi flow models the social interaction among nonadjacent regions, which potentially propagate offenders, victims, or resources and information used in crime control. We adopt a negative binomial regression model over the linear regression model, because the count-based regression model addresses issue of non-negative outcomes and deals with over-dispersion. We further propose to use a geographically weighted regression model to handle the non-stationary across space. Both POI and taxi flow features from a publicly accessible dataset in Chicago are evaluated to be helpful. In the best scenario,

TABLE 7: Spearman correlation between POI and crime rate. We calculate the correlation in a pairwise setting, because the POI is a vector while the crime rate is a scalar. More specifically, for each pair of regions, we calculate the cosine similarity between their POI features and the difference between their crime rates. Then we report the Spearman correlation of the pairwise difference in crime and the similarity of POI feature. The correlation value is typically negative, indicating that when two communities have similar POI distributions (large cosine similarity), their crime rate difference should be small. For POI performance, "-" indicates neutral performance, and "↑" indicates improved performance.

Category	Correlation	POI Performance
THEFT	-0.216	↑↑
BATTERY	-0.103	↑
NARCOTICS	-0.019	-
CRIMINAL DAMAGE	-0.072	-
BURGLARY	-0.027	-
OTHER OFFENSE	-0.226	↑↑
ASSAULT	-0.145	↑
MOTOR VEHICLE THEFT	-0.050	-
ROBBERY	-0.077	-
DECEPTIVE PRACTICE	-0.216	↑↑
total	-0.109	↑

the POI distribution and taxi flow reduce the prediction error (MAE) by over 15%.

8 Acknowledgments

The work was supported in part by NSF award #1544455, #1054389, and funding from NICHD R24-HD044943. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing any funding agencies.

References

[1] BJS, ""rates of violent victimizations, 1993-2015" and "rates of property victimizations, 1993-2015"," Generated using the NCVS Victimization Analysis Tool at http://www.bjs.gov, 2016, bureau of Justice Statistics, 17-Nov-2016.

[2] J. Truman and L. Langton, "Criminal victimization, 2014," Bureau of Justice Statistics (BJS), U.S. Department of Justice, Tech. Rep. NCJ 248973, 2015.

[3] D. Finkelhor, H. A. Turner, A. Shattuck, and S. L. Hamby, "Violence, crime, and abuse exposure in a national sample of children and youth: an update," *JAMA pediatrics*, vol. 167, no. 7, pp. 614–621, 2013.

[4] C. Graif, "Toward a geographically extended perspective of neighborhood effects on children's victimization," *American Society of Criminology Annual Meeting*, 2015.

[5] C. Tribune, "A tale of 3 cities: La and nyc outpace chicago in curbing violence," 2015. [Online]. Available: http://www.chicagotribune.com/news/ct-violence-chicago-new-york-los-angeles-met-20150918-story.html

[6] C. Graif and R. J. Sampson, "Spatial heterogeneity in the effects of immigration and diversity on neighborhood homicide rates," *Homicide Studies*, 2009.

[7] L. Anselin, "Under the hood: issues in the specification and interpretation of spatial regression models," *Agricultural economics*, vol. 27, no. 3, pp. 247–267, 2002.

[8] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM TIST*, vol. 5, no. 3, p. 38, 2014.

[9] J. E. Eck and D. L. Weisburd, "Crime places in crime theory," *Crime and place: Crime prevention studies*, vol. 4, 2015.

[10] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *ACM SIGKDD*. ACM, 2012, pp. 186–194.

[11] C. Graif, A. S. Gladfelter, and S. A. Matthews, "Urban poverty and neighborhood effects on crime: Incorporating spatial and network perspectives," *Sociology Compass*, vol. 8, no. 9, pp. 1140–1155, 2014.

[12] L. E. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," *American sociological review*, pp. 588–608, 1979.

[13] P. Brantingham and P. Brantingham, "Criminality of place," *European journal on criminal policy and research*, vol. 3, no. 3, pp. 5–26, 1995.

[14] H. Wang, D. Kifer, C. Graif, and Z. Li, "Crime rate inference with big data," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 635–644.

[15] R. Herring, A. Hofleitner, P. Abbeel, and A. Bayen, "Estimating arterial traffic conditions using sparse probe data," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, 2010, pp. 929–936.

[16] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in rome," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 141–151, 2011.

[17] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert systems with applications*, vol. 36, no. 3, pp. 6164–6173, 2009.

[18] X. Fei, C.-C. Lu, and K. Liu, "A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1306–1318, 2011.

[19] A. Simroth and H. Zahle, "Travel time prediction using floating car data applied to logistics planning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 243–253, 2011.

[20] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2009, pp. 85–98.

[21] H. Wang, Y.-H. Kuo, D. Kifer, and Z. Li, "A simple baseline for travel time estimation using large-scale trip data," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2016, p. 61.

[22] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1436–1444.

[23] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang, "Diagnosing new york city's noises with ubiquitous data," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 715–725.

[24] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, 2012.

[25] T. Wang, C. Rudin, D. Wagner, and R. Sevieri, "Learning to detect patterns of crime," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2013.

[26] I. Ehrlich, "On the relation between education and crime," in *Education, income, and human behavior*. NBER, 1975, pp. 313–338.

[27] J. Braithwaite, *Crime, shame and reintegration*. Cambridge University Press, 1989.

[28] E. B. Patterson, "Poverty, income inequality, and community crime rates," *Criminology*, vol. 29, no. 4, pp. 755–776, 1991.

[29] R. B. Freeman, "The economics of crime," *Handbook of labor economics*, vol. 3, pp. 3529–3571, 1999.

[30] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, 2012, pp. 231–238.

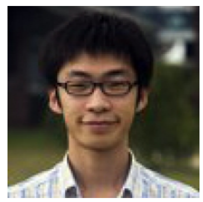
[31] M. S. Gerber, "Predicting crime using twitter and kernel density estimation," *Decision Support Systems*, vol. 61, 2014.

[32] M. Traunmueller, G. Quattrone, and L. Capra, "Mining mobile phone data to investigate urban crime theories at scale," in *Social Informatics*. Springer, 2014, pp. 396–411.

[33] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pentland, "Once upon a crime: towards crime prediction from demographics and mobile data," in *Proceedings of the 16th international conference on multimodal interaction*. ACM, 2014, pp. 427–434.

[34] J. H. Ratcliffe, "A temporal constraint theory to explain opportunity-based spatial offending patterns," *Journal of Research in Crime and Delinquency*, vol. 43, no. 3, pp. 261–291, 2006.

[35] J. L. Toole, N. Eagle, and J. B. Plotkin, "Spatiotemporal correlations in criminal offense records," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 4, p. 38, 2011.
[36] M. B. Short, M. R. D'ORSOGNA, V. B. Pasour, G. E. Tita, P. J. Brantingham, A. L. Bertozzi, and L. B. Chayes, "A statistical model of criminal behavior," *Mathematical Models and Methods in Applied Sciences*, vol. 18, no. supp01, pp. 1249–1267, 2008.
[37] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Security Journal*, vol. 21, no. 1, pp. 4–28, 2008.
[38] J. Eck, S. Chainey, J. Cameron, and R. Wilson, "Mapping crime: Understanding hotspots," National Institute of Justice, Tech. Rep. NCJ 209393, 2005.
[39] T. Nakaya and K. Yano, "Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics," *Transactions in GIS*, vol. 14, no. 3, pp. 223–239, 2010.
[40] A. L. Buczak and C. M. Gifford, "Fuzzy association rule mining for community crime pattern discovery," in *ACM SIGKDD Workshop on Intelligence and Security Informatics*. ACM, 2010, p. 2.
[41] M. A. Oliver and R. Webster, "Kriging: a method of interpolation for geographical information systems," *International Journal of Geographical Information System*, vol. 4, no. 3, pp. 313–332, 1990.
[42] N. Cressie, "The origins of kriging," *Mathematical geology*, vol. 22, no. 3, pp. 239–252, 1990.
[43] M. Knotters, D. Brus, and J. O. Voshaar, "A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations," *Geoderma*, vol. 67, no. 3-4, pp. 227–246, 1995.
[44] G. Hudson and H. Wackernagel, "Mapping temperature using kriging with external drift: theory and an example from scotland," *International journal of Climatology*, vol. 14, no. 1, pp. 77–91, 1994.
[45] D. E. Myers, "Estimation of linear combinations and co-kriging," *Mathematical Geology*, vol. 15, no. 5, pp. 633–637, 1983.
[46] S. Koziel, A. Bekasiewicz, I. Couckuyt, and T. Dhaene, "Efficient multi-objective simulation-driven antenna design using co-kriging," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 11, pp. 5900–5905, 2014.
[47] K. Wang, C. Zhang, and W. Li, "Predictive mapping of soil total nitrogen at a regional scale: a comparison between geographically weighted regression and cokriging," *Applied Geography*, vol. 42, pp. 73–85, 2013.
[48] I. Sideris, M. Gabella, R. Erdin, and U. Germann, "Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of switzerland," *Quarterly Journal of the Royal Meteorological Society*, vol. 140, no. 680, pp. 1097–1111, 2014.
[49] R. J. Sampson and C. Graif, "Neighborhood social capital as differential social organization resident and leadership dimensions," *American Behavioral Scientist*, vol. 52, no. 11, pp. 1579–1605, 2009.
[50] Research & Development Division of the Chicago Police Department, "City of chicago data portal," https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2, 2015.
[51] W. Gardner, E. P. Mulvey, and E. C. Shaw, "Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models." *Psychological bulletin*, vol. 118, no. 3, p. 392, 1995.
[52] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.
[53] D. W. Osgood, "Poisson-based regression analysis of aggregate crime rates," *Journal of quantitative criminology*, vol. 16, no. 1, pp. 21–43, 2000.
[54] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.
[55] C.-C. Hsieh and M. D. Pugh, "Poverty, income inequality, and violent crime: a meta-analysis of recent aggregate data studies," *Criminal Justice Review*, vol. 18, no. 2, pp. 182–202, 1993.
[56] M. K. Wolfe and J. Mennis, "Does vegetation encourage or suppress urban crime? evidence from philadelphia, pa," *Landscape and Urban Planning*, vol. 108, no. 2, pp. 112–122, 2012.
[57] O. Sahbaz and B. Hillier, "The story of the crime: functional, temporal and spatial tendencies in street robbery," in *Proc of 6th International Space Syntax Symposium, Istanbul*, 2007, pp. 4–14.
[58] J. Jacobs, *The death and life of great American cities*. Vintage, 1961.
[59] United States Census Bureau, "Decennial census," http://www.census.gov, 2010.
[60] R. J. Sampson, S. W. Raudenbush, and F. Earls, "Neighborhoods and violent crime: A multilevel study of collective efficacy," *Science*, vol. 277, no. 5328, pp. 918–924, 1997.
[61] R. J. Bursik and H. G. Grasmick, *Neighborhoods and Crime: The Dimensions of Effective Community Control*. Richard Rosenfeld University of Missouri-St. Louis: Lexington Books, 1993.
[62] Foursquare, "Foursquare venues service," https://developer.foursquare.com/overview/venues.html, 2015.
[63] D. M. Gorman, P. W. Speer, P. J. Gruenewald, and E. W. Labouvie, "Spatial dynamics of alcohol availability, neighborhood structure and violent crime." *Journal of studies on alcohol*, vol. 62, no. 5, pp. 628–636, 2001.
[64] J. D. Burnell, "Crime and racial composition in contiguous communities as negative externalities: prejudiced household's evaluation of crime rate and segregation nearby reduces housing values and tax revenues," *American Journal of Economics and Sociology*, vol. 47, no. 2, pp. 177–193, 1988.
[65] L. Anselin, J. Cohen, D. Cook, W. Gorr, and G. Tita, "Spatial analyses of crime," *Criminal justice*, vol. 4, no. 2, pp. 213–262, 2000.
[66] J. D. Morenoff and R. J. Sampson, "Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990," *Social forces*, vol. 76, no. 1, pp. 31–64, 1997.
[67] F. Wu, Z. Li, W.-C. Lee, H. Wang, and Z. Huang, "Semantic annotation of mobility data using social media," in *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015, pp. 1253–1263.



Hongjian Wang is a Ph.D. candidate in the College of Information Sciences and Technology at the Pennsylvania State University. His research interest is data mining, especially mining spatiotemporal data. Mostly he works with trajectory data, social network check-in data, and many other urban data. He received his Bachelor degree in computer science from Shanghai Jiao Tong University in 2013.

Huaxiu Yao is a Ph.D. student in the College of Information Sciences and Technology at the Pennsylvania State University. He received the Bachelor of Engineering degree in electronic information engineering from University of Electronic Science and Technology of China. He was a visiting student at the Pennsylvania State University from September 2016 to March 2017. His research interests are in statistical machine learning and data mining.

Daniel Kifer is an associate professor in the Department of Computer Science and Engineering at the Pennsylvania State University. He received his Ph.D. in computer science from Cornell University in 2006. His research interest include machine learning applications in various areas, such as statistical privacy, security, databases, streaming data processing, text modeling, deep learning, and computational social science.

Corina Graif is an assistant professor in the Department of Sociology and Criminology and a Research Associate at the Population Research Institute at the Pennsylvania State University. She received the PhD in Sociology from Harvard University in 2011 and was a Robert Wood Johnson Health and Society Scholar at the University of Michigan, Ann Arbor during 2011-2013. She studies spatial stratification, mobility, and neighborhood effects on risk and crime.

Zhenhui Li is an associate professor in the College of Information Sciences and Technology at the Pennsylvania State University. Her research interest lies in data mining and database systems. She has been working on designing effective and scalable methods for mining various kinds of complex patterns from spatial and temporal data. She received her Ph.D. degree in computer science from University of Illinois at Urbana-Champaign in 2012. She is a member of IEEE.