Learning to rank spatio-temporal event hotspots

George Mohler
Computer and Information Science
Indiana University - Purdue University Indianapolis
gmohler@iupui.edu

Michael D. Porter
Information Systems, Statistics and Management Science
University of Alabama
mporter@cba.ua.edu

Jeremy Carter
Criminal Justice
Indiana University - Purdue University Indianapolis
carterjg@iupui.edu

Gary LaFree
Criminology and Criminal Justice
University of Maryland
glafree@umd.edu

ABSTRACT

Crime, traffic accidents, terrorist attacks, and other space-time random events are unevenly distributed in space and time. In the case of crime, predictive policing algorithms aim to focus limited resources at the highest risk crime hotspots in a city. A crucial step in the implementation of these strategies is the construction of scoring models used to rank spatial hotspots. While these methods are evaluated by area normalized Recall@k (called the Predictive Accuracy Index), models are typically trained via maximum likelihood or rules of thumb that may not prioritize model accuracy in the top k hotspots. Furthermore, current algorithms are defined on fixed grids that fail to capture risk patterns occurring in neighborhoods and on road networks with complex geometries. We introduce CrimeRank, a learning to rank boosting algorithm for determining a crime hotspot map that directly optimizes the percentage of crime captured by the top ranked hotspots. The method employs a floating grid combined with a greedy hotspot selection algorithm for accurately capturing spatial risk in complex geometries. We illustrate the performance using crime and traffic incident data provided by the Indianapolis Metropolitan Police Department, IED attacks in Iraq, and data from the 2017 NIJ Real-time crime forecasting challenge. Our learning to rank strategy was the top performing solution (PAI metric) in the 2017 challenge. We show that CrimeRank achieves even greater gains when the competition rules are relaxed by removing the constraint that grid cells be a regular tessellation.

KEYWORDS

Learning to rank, Gradient boosting, Spatial-temporal point process, Crime forecasting, Scan statistic

ACM Reference Format:
George Mohler, Michael D. Porter, Jeremy Carter, and Gary LaFree. 2018. Learning to rank spatio-temporal event hotspots. In *Proceedings of The 7th International Workshop on Urban Computing (URBCOMP2018).* ACM, London,UK, 9 pages.

URBCOMP2018, August 2018, London

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

1.1 Crime and security event hotspots

Many types of events related to human activity cluster in space and time, forming event "hotspots." Burglary offenders are known to replicate success at nearby, or identical, locations to previous crimes [39] and space-time clusters are observed in patterns of shootings [37] due to retaliation and escalation. Event hotspots also occur in more extreme security settings, for example IED attacks tend to cluster in time [25] due to self-excitation and exogenous effects. In Figure 1, we plot Improvised Explosive Device (IED) attacks in Baghdad from 2004-2009. These events cluster along road networks and at major intersections within the spatial geography of the city.

Hotspot policing is a strategy for deterring crime where police resources are directed to the highest volume crime areas of a city. Experimental studies indicate that elevated policing presence in hotspots comprising a relatively small area of the city can lead to statistically significant crime rate reductions [7]. The standard approach for determining hotspots consists of dividing a city into geographic sub-regions, often grid cells, and scoring hotspots based upon historical crime counts over a specified time window [10]. More recently, point processes have been introduced for ranking crime hotspots [30] and have been shown to lead to further crime rate reductions in field trials over traditional hotspot mapping [31]. Other approaches for ranking crime hotspots include generalized linear models [20, 42, 44], generalized additive models [44], and random forests have been applied to the problem of ranking offenders [3]. Space-time models for event prediction also have been applied to conflict [48] and terrorism [17] datasets.

Since the goal of hotspot and predictive policing is crime rate reduction, the standard metric for assessing a given scoring procedure is the percent of crime captured inside the top ranked hotspots in the absence of police intervention. The predictive accuracy index (PAI) [10, 31, 35],

$$\text{PAI} = \frac{\text{crime in k hotspots}}{\text{total crime}} \cdot \frac{\text{total area}}{\text{area of k hotspots}}, \qquad (1)$$

measures the percent of crime predicted in the top k hotspots normalized so that spatially random predictions have a PAI value of 1. In practice, the value of k is chosen to correspond to policing resources and realistic values may correspond to an area on the order of 1% of a city [31].

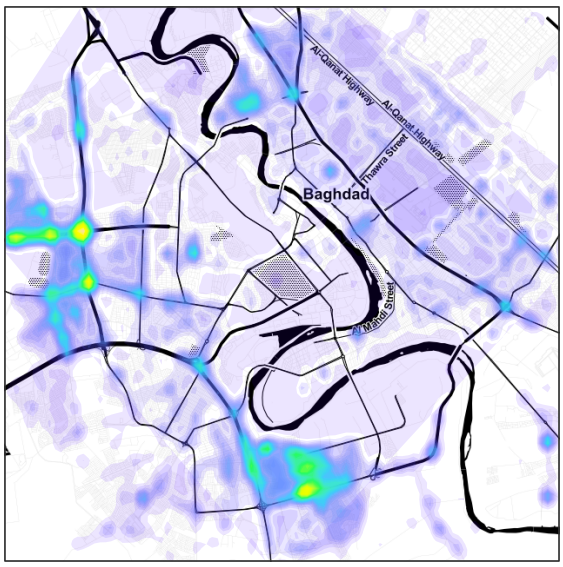

Figure 1: IED attack hotspots in Baghdad from 2004 to 2009.

1.2 Learning to rank

Similar loss functions, such as NDCG@k, Prec@k and Recall@k, are used in information retrieval [27] to measure the effectiveness of scoring algorithms aimed at producing a high percentage of relevant documents in the top k documents returned from a query. The mathematical formulation of the two problems is similar, where the analog of a query is the time unit (window) for which crime hotspot predictions are made, the analog of a document is a single spatial unit (grid cell, neighborhood, block, street corner, etc.) in the city, and the analog of relevance is a binary or integer variable indicating whether or not a crime occurred inside the spatial unit and time window (or how many crimes occurred). We therefore use the notation PAI@k to denote the PAI value when the top k hotspots are flagged for police intervention. Learning to rank algorithms attempt to directly optimize the loss function of interest and have been shown to out-perform regression and likelihood based algorithms that optimize a smooth surrogate loss function [8, 27]. We note that there has been some work on spatial learning to rank in the context of inferring a users location from noisy GPS [38], however to our knowledge no work to date has focused on the learning to rank problem in the context of crime event prediction.

In this paper we develop a learning to rank algorithm, CrimeRank, for space-time event hotspot ranking. A general overview of the algorithm is as follows. Features are defined for each potential hotspot in a city at a particular time unit and then used to calculate a risk score that ranks hotspots over the next (future) time unit. Similar to LambdaMart [8], we introduce a pseudo-derivative for PAI@k and then perform gradient ascent boosting to maximize PAI. At each iteration we use decision trees as the weak learner to model the derivative of PAI as a function of the features in each hotspot. At prediction time we compute the score for a collection of potentially over-lapping hotspots and then perform a greedy sort to select the top k non-overlapping hotspots.

1.3 Outline

We apply the CrimeRank method to several space-time event data sets to illustrate the improvement in PAI over existing methodologies. The outline of the paper is as follows: in Section 2 we provide details on the CrimeRank algorithm and in Section 3 we include results for the CrimeRank algorithm on several data sets including crime and traffic incidents in Indianapolis, IED attacks in Baghdad, and data from Portland, Oregon used in the 2017 NIJ Real-time crime forecasting challenge. Our learning to rank strategy under the team name PASDA was the top performing solution (PAI metric) in the 2017 challenge. We show that CrimeRank achieves even greater gains when the competition rules are relaxed and spatial discretizations are not required to be a regular tessellation. We discuss future directions for research in this area in Section 4.

2 CRIMERANK: AN ALGORITHM FOR RANKING CRIME AND OTHER SPACE-TIME EVENT HOTSPOTS

2.1 Feature selection

Given a data set of space time event locations up to the present day, our goal is to flag a set of k spatial areas that have the highest risk for event occurrence in the near future, e.g. the next day, week, month, etc. In this paper we will consider rectangular grids for dividing a city into sub-areas, though our methodology applies to more general polygons and other sub-divisions.

In the case of crime, algorithms typically fall into one of two broad categories for ranking spatial areas, namely nonparametric methods utilizing only event data (kernel hotspot maps and point processes are common methods) or multivariate models that explicitly incorporate additional variables such as demographics [44], income levels [26], distance from crime attractors [20, 26, 44], leading-indicator crimes [11, 18], and auxiliary social sensing data (Twitter, mobile phone locations, Google street view, etc.) [6, 21, 42, 45].

Because the focus of this paper is on the optimization method used to train a hotspot ranking model, rather than feature selection, we restrict our attention to univariate modeling where features are derived from the event data alone. Our methodology would easily extend to other types of features. For univariate feature extraction we compute a 52-week time series consisting of the event count in each grid cell for the 52 weeks leading up to the present. Our learning task is then to rank the grid cells such that the top k cells have the largest number of events in the subsequent week.

A training data set is then created over a historical time period by computing the 52 dimensional feature set for each cell and each week in the historical period, where the label is the number of events in the next week. While each row in the data set is a grid cell-week pair, we note that rows are not independent in the ranking setting and all rows corresponding to the same week must be considered simultaneously to compute PAI. The analog of a week in the information retrieval setting is a query. However, regression based methods will treat all rows as independent during training.

2.2 Optimization of PAI@k

Next we describe our optimization method for maximizing PAI@k, the area normalized fraction of crime in the top k event hotspots.

Let $i \in \{1, 2, \ldots, N\}$ index the N grid cells and $t \in \{1, 2, \ldots, T\}$ index the T time periods in which predictions are being made. Let z_{it} denote the feature vector, s_{it} the score, and y_{it} the label (number of events) for cell i at time t. Note that y_{it} is the number of events in the future time period $t+1$. This gives a total of $N \times T$ observations.

The set of scores induce a ranking on the grid cells for each time period. Let r_{it} be the rank of score s_{it}, with a rank of one being assigned the cell with the *largest* score at time t. Then the top k cells, at time t, are $V_{kt} = \{i : r_{it} \leq k\}$. The resulting PAI is calculated separately for each time period.

We first note that PAI is non-smooth as a function of s_{it}. In particular, consider fixing the scores except for two grid cells in the same week t indexed by i and j and assume $y_{it} > y_{jt}$. Then PAI will be piecewise constant as a function of $s_{it} - s_{jt}$ and will have a jump discontinuity at $s_{it} = s_{jt}$. Therefore PAI has no derivative for performing gradient ascent. However, we follow the approach of [8] and introduce a pseudo-derivative λ_{it},

$$\lambda_{it} = \sum_{j: y_{it} > y_{jt}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{it} - s_{jt}}} - \sum_{j: y_{jt} > y_{it}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{jt} - s_{it}}}, \tag{2}$$

that models the gradient of PAI at cell-week i-t. Here the term $\Delta_{kt}(i,j)$ denotes the change in PAI if the ranking of cells i and j are swapped at time t (leaving all other rankings fixed) and can be written,

$$\Delta_{kt}(i,j) = \begin{cases} c\,(y_{it} - y_{jt})/N_t & r_i \leq k, r_j > k \\ 0 & r_i \leq k, r_j \leq k; r_j > k, r_j > k \\ c\,(y_{jt} - y_{it})/N_t & r_j > k, r_j \leq k \end{cases} \tag{3}$$

where c = (total area)/(area of k grid cells) is the PAI normalizing constant.

The first summation in (2) is over all pairs where grid cell i should be ranked higher than grid cell j and thus is positive in order to increase the score s_{it} and thus increase the PAI. The logistic term evaluated at $s_{it} - s_{jt}$ is introduced to add regularization and in [8] the authors find that it has the effect of adding a margin. The second term is over pairs where i should be ranked lower than j and thus has the effect of lowering the score s_{it} (and therefore increasing PAI).

We note that the computational cost of λ_{it} over all i is quadratic, however in practice the performance is approximately linear. First, only grid cells in the same time period need to be considered when computing $\{\lambda_{it}\}_{i=1}^{N}$. Second, for many event data sets and reasonably small grid cells only a small percentage of cells will contain non-zero counts. Because (2) only involves pairs in which $y_i \neq y_j$ the cost is $O(M_0 M_1)$ where M_1 is the number of non-zero labels for a given t and M_0 is the number of zero label cells.

Given the model for the derivative λ of PAI@k, we then use decision tree based gradient boosting to optimize the loss function. We call our method CrimeRank and provide pseudo-code in Algorithm 1. Starting with an initial guess for scores s_{it}, we then perform boosting iterations where 1. the pseudo-derivative λ_{it} is computed using the current score guess, 2. a regression tree is fit to the derivative λ_{it} as a function of the features z_{it}, and 3. the score s_{it} is updated by a gradient ascent step. In practice we find that using *stochastic* gradient ascent [16] performs better where a random subset of λ_i are used to estimate the regression tree Γ at each iteration. In Figure 2 we plot an example of boosting iterations for robbery incidents in Indianapolis. Empirically we find that the pseudo-derivative is effective in maximizing the PAI (proportional to the fraction of crime predicted) on training data. We provide more results in Section 3.

Algorithm 1 CrimeRank

Input: features z_{it}, labels y_{it} for $i = 1, ..., N$ and $t = 1, \ldots, T$. Number of trees M, learning rate η. Initial guess score $s_{it} = 0$.
for $l = 1, ..., M$ **do**
  for $t = 1, \ldots, T$ **do**
    for $i = 1, ..., N$ **do**

$$\lambda_{it} = \sum_{j: y_{it} > y_{jt}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{it} - s_{jt}}} - \sum_{j: y_{jt} > y_{it}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{jt} - s_{it}}}$$

    end for
  end for

  $\Gamma_l \leftarrow \text{RegTree}(\lambda, z)$
  where $\lambda = [\lambda_{11}, \ldots, \lambda_{N1}, \lambda_{12}, \ldots, \lambda_{NT}]$,
  $z = [z_{11}, \ldots, z_{N1}, z_{12}, \ldots, z_{NT}]$

  for $i = 1, ..., N$ and $t = 1, \ldots, T$ **do**
    $s_{it} = s_{it} + \eta\, \Gamma_l(z_{it})$
  end for
end for

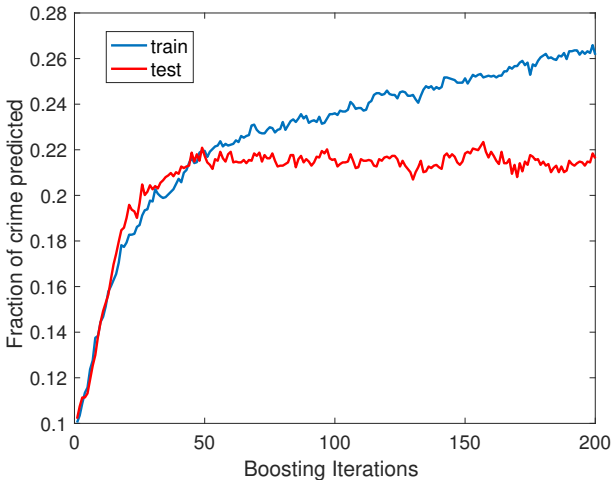

Figure 2: CrimeRank boosting iterations using stochastic gradient ascent for robbery hotspot ranking in Indianapolis over 2013-2015 (split for training and testing).

2.3 Offgrid space-time ranking

The second component of CrimeRank is an "offgrid" approach that we introduce for dealing with complex geometries that are associated with event patterns along road networks and other urban structures. In Figure 3 we provide an illustration of the problem

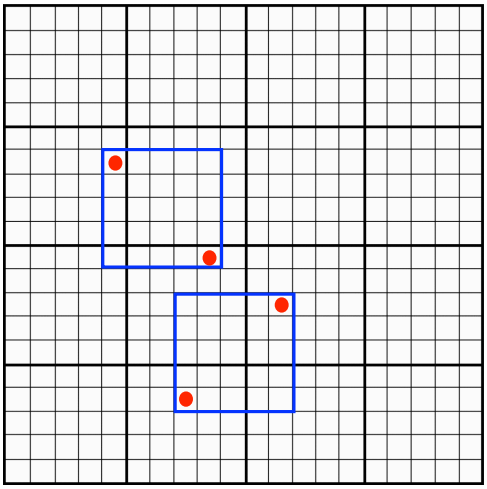

Figure 3: 4x4 grid scenario. Maximum PAI@2 on the fixed grid is 4 (1/2 of crime captured divided by 2/16 of area flagged). Maximum PAI@2 on a floating grid is 8.

that arises with fixed grids used in spatial hotspot ranking. Here four events are plotted over a regular grid (thick black lines) and we let $k = 2$. Then four grid cells each have one event, the others have zero, so that the maximum possible PAI@2 is four (two crimes out of four predicted area normalized by two cells out of sixteen). However, cells chosen without respect to a regular grid can achieve a PAI@2 of eight even with the same size and shape.

We introduce a simple heuristic for moving to an offgrid approach while taking advantage of the CrimeRank algorithm introduced in Section 2.2. In particular, we train CrimeRank on a fixed regular grid obtaining the fitted CrimeRank model (i.e., the collection of regression trees). The CrimeRank model is then used to estimate the risk score for a larger collection of grid cells and a greedy sort algorithm is used to find the set of k non-overlapping cells with the largest scores.

The CrimeRank model is fit one time, on a given grid, and then used to estimate the score at additional grid cells. The additional collection of grid cells can be generated, e.g., by translating and rotating the original grid used for model fitting. Because the model features must be calculated for the new grid cells, it is important to use the same size cells. In Sections 3.1 and 3.2 we use $g \times g$ over-lapping grids identical to the original fixed grid except that they are offset by a multiple of $\Delta x/g$ from the fixed grid where Δx is the length of the side of a grid cell. Figure 3 illustrates the setting of $g = 5$; the thick lines shows the original 16 grid used for training the model and the collection of 200 additional grid cells are the square regions obtained by centering on each small square. In practice we find that $g = 10$ works well in balancing accuracy and storage/computational costs. In Section 3.3, we also incorporated rotated grid cells to expand the number of potential hotspots.

Once all of the grid cells are scored, we utilize a greedy sort algorithm (Algorithm 2) to identify the top k non-overlapping hotspots. First we select the cell with the highest score over all grids. Second we select the cell with the next highest score such that it does not overlap with the first cell. We continue on in this fashion, where the jth cell is selected with the highest score such that it does not overlap with cells $1, \ldots, j-1$.

Algorithm 2 Greedy sort for hotspot selection

Input: number of hotspots k, grid cell scores s_j, grid cell polygons U_j, hotspot cell set $V = \emptyset$
  while $\text{Area}(V) < k \cdot \text{Area}(U_1)$
    $j \leftarrow \arg\max\{s_j : U_j \cap V = \emptyset\}$
    $V \leftarrow V \cup U_j$
  end while

We note that there is a connection between the offgrid methodology we have proposed here and spatial scan statistics used to detect anomalies (for example disease outbreaks) in spatial-temporal event data [2, 22, 32]. The goal of the scan statistic approaches is to detect emerging spatio-temporal clusters that have anomalous event rates by *scanning* over many possible spatial regions and time periods. Because the focus is often on disease outbreaks, the clusters identified with scan statistic methods are usually constrained to be connected, or nearly connected, spatial regions. While our goal is different, namely identifying the regions with the largest expected event rate in the future rather than identifying the regions that have the most unusual event rates in the recent past, the scan statistic methods developed to search for irregularly shaped clusters [13, 14, 33, 40, 41] could be used to generalize the rectangular regions we considered and speed the search process. We will return to this idea in the discussion in Section 4.

3 RESULTS

3.1 Indianapolis crime hotspot ranking

In our first example we test the CrimeRank methodology using crime and vehicle crash incident data from the city of Indianapolis, Indiana. Crime incidents for years 2012-2015, specifically robbery and residential burglary, were provided electronically by the Indianapolis Metropolitan Police Department (IMPD). Vehicle crash data for years 2012-2013 were provided electronically from the Indiana State Police using the Automated Reporting Information Exchange System (ARIES). One of two characteristics must occur for collisions to be included in ARIES; if the incident resulted in personal injury or death, or property damage to an apparent extent greater than one thousand dollars. Both crime and crash data included date and time stamp as well as state-plane coordinates from a composite address locator that were converted to WGS84 coordinates. Robbery [19, 37, 47], residential burglary [5, 34, 36], and vehicle crashes [9, 12, 23] have demonstrated spatiotemporal patterns in criminological research that are likely to inform strategic police operations to mitigate risk and deter offending. Thus, these three incident types are the focus of the present demonstration.

In the data set there are 35,225 burglary incidents, 13,135 robbery incidents, and 42,328 traffic accidents and we model and evaluate each event type separately. We consider weekly time periods and,

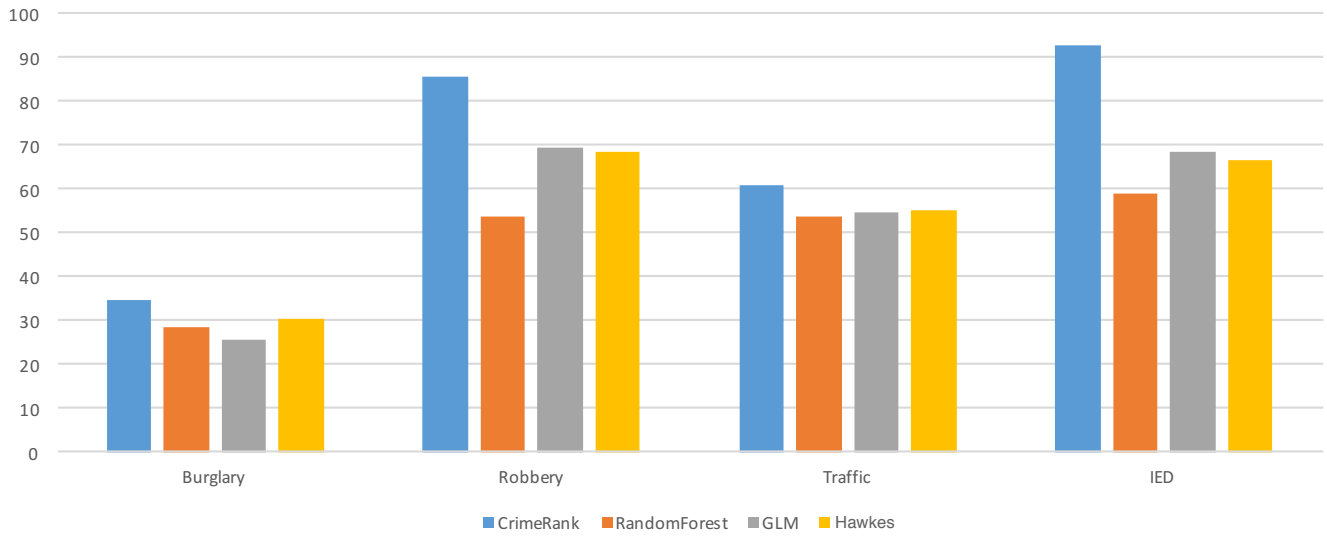

Figure 4: PAI results of CrimeRank vs. three methods of comparison for ranking hotspots of burglary, robbery, traffic incidents, and IED attacks.

following [31], use grid cells of size $150m \times 150m$. We compare CrimeRank to several existing methods including random forest, generalized linear model (GLM), and a Hawkes point process [30, 31]. CrimeRank, random forest, and GLM use the same features (weekly event counts in the grid over the last 52 weeks). For the Hawkes model, we follow [31] where the conditional intensity $f_i(t)$ of events in grid cell i at time t is modeled by,

$$f_i(t) = \mu_i + \sum_{t_i^j < t} \theta\omega e^{-\omega(t-t_i^j)}, \tag{4}$$

where t_i^j are the times of events in cell i in the history of the process. The Hawkes model has two components, one modeling place-based environmental conditions that are constant in time and the other modeling dynamic changes in risk. The parameters μ_i, θ and ω are estimated using an Expectation-Maximization algorithm [31]. Hawkes processes are used to model a variety of social phenomena where events increase the likelihood of future events. In addition to crime, recent social applications include Twitter resharing [49], IPTV viewing behavior [46], and human mobility [43].

We use the time period 1/1/2013 to 6/31/2014 for training and evaluate the methods on each week during the time period 7/1/2014 to 12/31/2015 (for traffic accidents we use 1/1/2013 to 6/31/2013 for training and 7/1/2013 to 12/31/2013 for testing). For CrimeRank we use a max leaf size of 500 for the regression trees and subsample 1/4 of the training data when constructing each tree. We use $k = 200$ grid cells for evaluation, comprising approximate 0.4% of the city, on the same order of magnitude as realistic predictive policing deployments [31].

In Figure 4 we plot the PAI results for the four methods applied to crime incident data in Indianapolis. For all three crime incident types CrimeRank outperforms the other methodologies. In the case of burglary, CrimeRank captures 21%, 36% and 14% more crime in the top 200 hotspots compared to the random forest, GLM, and Hawkes models respectively. The largest improvement is in the case of robbery, where CrimeRank improves by over 20% crime captured compared to the next best method. An explanation for these results is that in the case of robbery, crime is highly clustered on street networks and CrimeRank is able to adapt to the geometry of the network (see Figure 5). Burglary is more spatially disaggregated and thus the PAI values are lower and there is less room for improvement, though Crimerank still improves over the next best method (Hawkes) by 5 PAI points (35 vs 30).

3.2 Improvised Explosive Device (IED) attacks in Baghdad, Iraq

In our second example we test the CrimeRank methodology using IED incident data from central Baghdad, including date, latitude and longitude of attacks, during the Iraq War from 2004 to 2009. In the data set there are 16,495 IED attacks. The attack data are based on Significant Activity (SIGACT) reports by Coalition forces in Iraq. Unclassified data from the MNU-I SIGACTS III database were provided to the Empirical Studies of Conflict (ESOC) project [4]. The data set includes a wide range of activity but our analysis here is limited to IEDs. The SIGACT data have two weaknesses that are relevant here. First, they capture violence against civilians and between nonstate actors only when U.S. forces are present and so likely undercount sectarian violence [24][15]. Given that our emphasis is on IEDs, missing sectarian violence should not bias our results. Second, these data almost certainly suffer from measurement error in that units vary in their thresholds for reporting specific events as significant activity. Fortunately, there is no evidence that such error is nonrandom with regard to the IED locations.

We again make weekly predictions and use grid cells of size $150m \times 150m$. For CrimeRank we use a max leaf size of 500 for the regression trees and subsample 1/4 of the training data when constructing each tree. We compare CrimeRank to the same three methods as in Section 3.1 including a random forest and generalized linear model (GLM) using identical 52 week time series features, along with the Hawkes point process. We use the time period 1/1/2006 to 6/31/2007 for training and we evaluate the methods over the time period 7/1/2007 to 12/31/2008. We again use $k = 200$ grid cells for evaluation, comprising approximately 0.4% of the central area of Baghdad (chosen for the study to be a similar size to Indianapolis).

Figure 5: CrimeRank determined IED hotspots for a week in 2008 in an area of central Baghdad. Hotspots align to road network and certain intersections to maximize PAI.

In Figure 4 we plot the PAI results for the four methods applied to the IED incident data. Similar to robbery, CrimeRank outperforms the other methodologies by 57%, 35% and 39% (random forest, GLM, and Hawkes model respectively). In Figure 5 we provide an example of the CrimeRank hotspot distribution on a given week in the testing period for a section of central Baghdad. We note that grid cells are able to align to intersections and diagonal roads in a manner such that the corners of the grid cell are aligned with the street, thus maximizing PAI (for example the left most cluster of four cells illustrate this effect).

In Figure 6 we plot the average number of IED incidents captured in the top k grid cells (as a function of k). One interesting effect to note is that the highest grid cells of CrimeRank contain less incidents compared to the other three methods. This is likely due to the fact that PAI is not changed by a re-ordering of the top grid cells ranking, but instead is sensitive to cells either being inside or outside of the top k. After the top 10 cells, CrimeRank cells contain significantly more incidents than the other three methods, explaining the overall improvement in PAI.

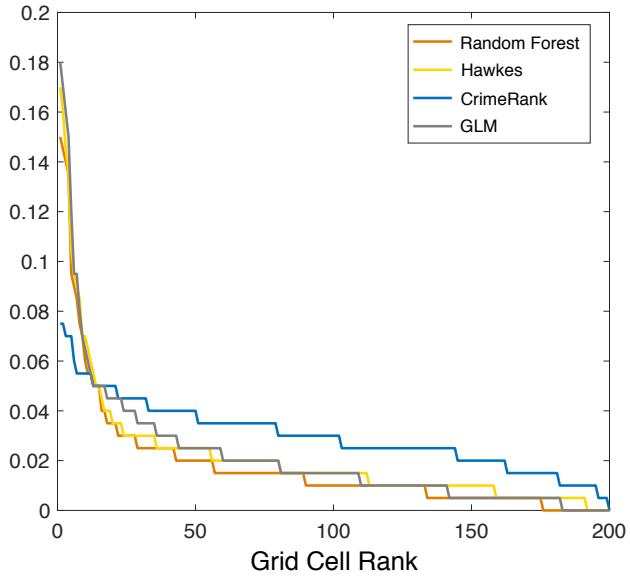

Figure 6: Average number of incidents captured in the top k grid cells.

3.3 2017 NIJ Crime Forecasting Challenge

The 2017 NIJ Crime Forecasting challenge tasked participants with forecasting the spatial locations containing the highest volume of crime-related calls for service in Portland, OR. Specifically, the contestants were given event data comprising projected geographic coordinates, date, and category (burglary, street crime, theft of auto, other) for the period of March 1, 2012 through February 28, 2017. Separate forecasts were made for 4 event types: burglary (Burg), street crime (Street), theft of auto (MVT), and all calls for service (ACFS) and 5 forecast horizons: 1 week (March 1-7), 2 weeks (March 1-14), 1 month (March 1-31), 2 months (March 1-April 30), and 3 months (March 1-May 31). The submitted forecast was specified to be a set of regular grid cells that covered all of the study region with some of the cells flagged as a "hotspot". The grid cells were required to be a regular tessellation of the Portland, OR administrative region in which all grid cells must have the same size, shape, and orientation. Rectangles, triangles, and hexagons were the permitted grid shapes. Furthermore, the grid cells were required to have an area between 62,500 ft^2 and 360,000 ft^2 with the smallest dimension being at least 125 ft. The cells flagged as hotspots were required to have aggregate area between 0.25 mi^2 - 0.75 mi^2, but there was no requirement that the hotspot cells be connected.

For the competition, we developed a Rotational Grid PAI maximization strategy (RGPM) [29] under the team name PASDA that was designed for jointly learning an optimal grid and scoring function for the purpose of maximizing PAI in crime forecasts under the rules of the NIJ competition. We used a regular grid of equally sized rectangles with the minimum allowable area (62,500 ft^2). The grid was parametrized with three parameters: cell height h, a grid translation parameter γ and a rotation angle θ. The overall procedure is captured in Algorithm 3, where the model $\mathcal{M}$ mapping features to the target variable was either a point process based

Table 1: Aggregate number of 1st, 2nd and 3rd place PAI finishes across divisions along with total number of overall 3rd and higher finishes (A) and number of 3rd and higher finishes within division (B).

Name	1st	2nd	3rd	A	B
PASDA	4	5	4	13	20
TAMERZONE	4	5	2	11	15
GRIER	1	4	0	5	8
JeremyHeffner	2	0	3	5	9
ANDY_NIJ	1	2	1	4	9
KUBQR1	0	1	3	4	7
pennaiken	2	0	2	4	10
Codilime	3	0	0	3	7

GLM or a random forest (depending on crime category). A simplex method was used to maximize PAI with respect to the rotational grid parameters.

In Table 1 we include overall competition results illustrating the accuracy of our RGPM approach. In the table we list the number of overall (across the three divisions) 1st, 2nd and 3rd place PAI finishes for teams having placed at least once. We note that the RGPM tied for the most 1st and 2nd place finishes and had the most 3rd place finishes across the crime type categories and forecasting windows. We also include in Table 1 the total number of finishes (3rd place and higher) within our division (large business) and overall, in both cases the RGPM method had the most finishes.

Algorithm 3 Rotational grid PAI maximization

1: **Function PAI**(h,θ,γ,$\vec{x}_i$,t_i,ω,A_{min})
a. Set up grid with cell height h, cell area A_{min}, grid angle θ, and offset γ.
b. Calculate event based features on grid using crime locations $\vec{x}_i$ and times t_i.
c. Fit a supervised model $\mathcal{M}$, using tuning parameters ω, on event features defined on the training set.
d. Predict $\mathcal{M}$ on test data features and output PAI.
Return PAI
2: **Function OptimizeGrid**($\vec{x}_i$,t_i,ω,A_{min})
Run simplex method to maximize PAI($h, \theta, \gamma, \vec{x}_i, t_i, \omega, A_{min}$) over h, θ, and γ.
Return h, θ, and γ.

Next we compare CrimeRank to the top performing methods of the NIJ competition. CrimeRank uses features similar to the RGPM method, namely event counts in the week, month, 90 days, 1 year, and 5 years leading up the the forecasting window date. For training we use the time period 3/1/2013 to 5/31/2016 and then we evaluate the CrimeRank method using the competition validation data set. To reduce the memory requirements of using the offgrid search, we generate the additional grid cells by creating rectangles centered at a sub-sample of the event locations in the training period (10000 events). We consider (250ft $\times$ 250ft) squares and (125ft $\times$ 500ft) rectangles with four orientations (0, $\pi/4$, $\pi/2$ and $3\pi/4$). We use a max leaf size of 100 for street crime and 50 for all calls for service for the regression trees and subsample 1/4 of the training data when constructing each tree. Examples of the hotspot cells are shown in the top right of Figure 7. The code to reproduce our CrimeRank results is available at Github [1].

We restrict our attention to the categories street crime and all calls for service over the 3 month forecasting window. In Figure 7 we plot CrimeRank PAI values (NIJ validation data set) vs. boosting iterations in comparison to the top performing solutions in the competition. In the case of street crime, CrimeRank achieves a PAI of 90 compared to the 1st place solution PASDA (PAI 87) and the 2nd place solution TAMERZONE (PAI 84). For all calls for service, CrimeRank achieves a PAI of 64 compared to the 1st place solution CODILIME (PAI 60.5). In Figure 7 we also plot examples of CrimeRank hotspots and note that rectangles at diagonal angles are heavily favored in certain areas of Portland where major streets run diagonally, an artifact that was not possible within the rules of the NIJ competition (but meets the spirit of the rules in terms of cell shape, size, and non-overlapping requirements). Given the high societal cost of crime [28], we believe a PAI improvement of 4 (over competition winning methods) is a significant result.

4 DISCUSSION

We developed a spatial-temporal learning to rank algorithm, CrimeRank, for identifying high risk "hotspots" in human activity data. The method directly optimizes the PAI@k loss function from criminology using gradient boosting. Although the loss function is non-smooth, a pseudo derivative is used in the boosting algorithm that empirically maximizes PAI. CrimeRank also deals with the geometry of hotspots in urban environments using a novel greedy sorting algorithm at the time predictions are made. We show that CrimeRank improves the % of events captured in hotspots by up to 35% compared to commonly used methods for crime and IED event data.

In this work we restricted our attention to searching for rectangularly shaped hotspots. While we do develop the offgrid approach that considers shifting, rotating, and scaling the rectangles, hotspots with more general shapes may better capture location specific geometries and lead to higher PAI scores. Future research that draws on the scan statistics literature may lead to such further improvements.

The methods introduced here will compliment recent work on the incorporation of social sensing data into crime predictions [6, 21, 42, 45]. For example, real-time human movement data collected via smart phones or fixed city sensors has been shown to improve crime hotspot prediction accuracy. Implementing real-time, offgrid learning to rank and spatial scan methods at scale presents several computational and algorithmic challenges that have yet to be solved.

5 ACKNOWLEDGEMENTS

This work was supported in part by NSF grants S&CC-1737585, SES-1343123, ATD-1737996 and CCF-1659488. G.M. is a co-founder of PredPol, a company offering predictive policing services to law enforcement agencies and serves on the board of directors.

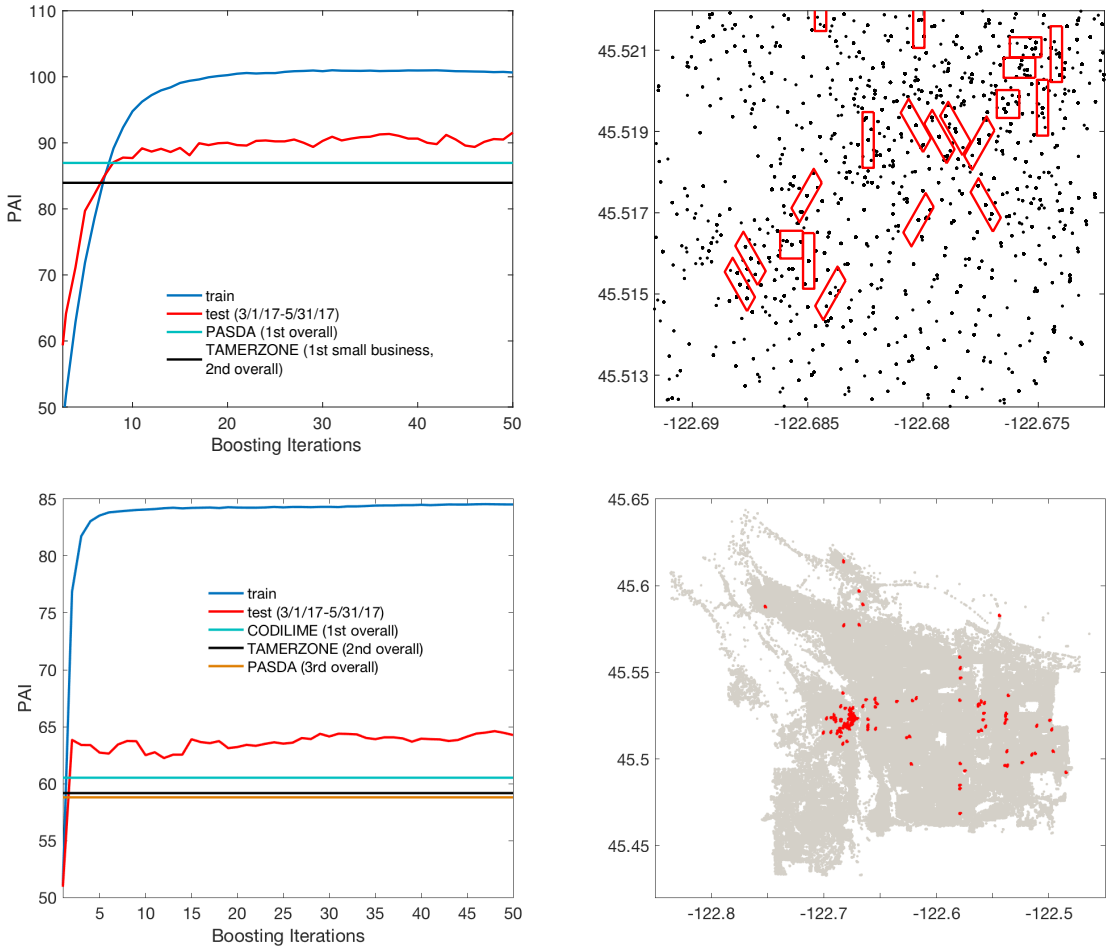

Figure 7: Top left: PAI results for CrimeRank applied to Portland street crime vs. NIJ top performing solutions. Top right: Example street crime hotspots selected via CrimeRank. Lower left: PAI results for CrimeRank applied to Portland all calls for service (ACFS) vs. NIJ top performing solutions. Lower right: Locations (red) of CrimeRank ACFS hotspots for the 3-month NIJ forecasting windows (locations of all incidents in gray).

REFERENCES

[1] 2018. CrimeRank. https://github.com/gomohler/crimerank. (2018).

[2] Renato Assunção and Thais Correa. 2009. Surveillance to detect emerging space–time clusters. *Computational Statistics & Data Analysis* 53, 8 (2009), 2817–2830.

[3] Richard Berk, Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz, and Lindsay Ahlman. 2009. Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172, 1 (2009), 191–211.

[4] Eli Berman, Jacob N Shapiro, and Joseph H Felter. 2011. Can hearts and minds be bought? The economics of counterinsurgency in Iraq. *Journal of Political Economy* 119, 4 (2011), 766–819.

[5] Wim Bernasco. 2008. Them again? Same-offender involvement in repeat and near repeat burglaries. *European Journal of Criminology* 5, 4 (2008), 411–431.

[6] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. 2014. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction*. ACM, 427–434.

[7] A.A. Braga. 2001. The effects of hot spots policing on crime. *The ANNALS of the American Academy of Political and Social Science* 578, 1 (2001), 104–125.

[8] Christopher JC Burges. 2010. From ranknet to lambdarank to lambdamart: An overview. *Learning* 11, 23-581 (2010), 81.

[9] Jeremy G Carter and Eric L Piza. 2017. Spatiotemporal Convergence of Crime and Vehicle Crash Hotspots: Additional Consideration for Policing Places. *Crime & Delinquency* (2017), 0011128717714793.

[10] S. Chainey, L. Tompson, and S. Uhlig. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal* 21, 1 (2008), 4–28.

[11] Jacqueline Cohen, Wilpen L Gorr, and Andreas M Olligschlaeger. 2007. Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geographical Analysis* 39, 1 (2007), 105–127.

[12] Grant Drawve, Michelle Belongie, and Hannah Steinman. 2017. The Role of Crime Analyst and Researcher Partnerships: A Training Exercise in Green Bay, Wisconsin. *Policing: A Journal of Policy and Practice* (2017).

[13] Luiz Duczmal, André L F Cançado, and Ricardo H C Takahashi. 2008. Delineation of irregularly shaped disease clusters through multiobjective optimization. *Journal of Computational and Graphical Statistics* 17, 1 (2008), 243–262.

[14] Luiz Duczmal, Martin Kulldorff, and Lan Huang. 2006. Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational and Graphical Statistics* 15, 2 (2006), 428–442.

[15] Hannah Fischer. 2008. Iraqi Civilian Casualties Estimates. LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE.

[16] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38, 4 (2002), 367–378.

[17] Peng Gao, Diansheng Guo, Ke Liao, Jennifer J Webb, and Susan L Cutter. 2013. Early detection of terrorism outbreaks using prospective space–time scan statistics. *The Professional Geographer* 65, 4 (2013), 676–691.

[18] Wilpen L Gorr. 2009. Forecast accuracy measures for exception reporting using receiver operating characteristic curves. *International Journal of Forecasting* 25, 1 (2009), 48–61.

[19] Cory P Haberman and Jerry H Ratcliffe. 2012. The predictive policing challenges of near repeat armed street robberies. *Policing: A Journal of Policy and Practice* 6, 2 (2012), 151–166.

[20] L.W. Kennedy, J.M. Caplan, and E. Piza. 2011. Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology* 27, 3 (2011), 339–362.

[21] Aditya Khosla, Byoungkwon An An, Joseph J Lim, and Antonio Torralba. 2014. Looking beyond the visible scene. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3710–3717.

[22] Martin Kulldorff. 2001. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164, 1 (2001), 61–72.

[23] Pei-Fen Kuo, Dominique Lord, and Troy Duane Walden. 2013. Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data. *Journal of Transport Geography* 30 (2013), 138–148.

[24] Barry Leonard. 2009. *Measuring stability and security in Iraq*. DIANE Publishing.

[25] E. Lewis and G. Mohler. 2011. A nonparametric EM algorithm for multiscale Hawkes processes. *preprint* (2011).

[26] Hua Liu and Donald E Brown. 2003. Criminal incident prediction using a point-pattern-based density model. *International journal of forecasting* 19, 4 (2003), 603–622.

[27] Tie-Yan Liu and others. 2009. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval* 3, 3 (2009), 225–331.

[28] Kathryn E McCollister, Michael T French, and Hai Fang. 2010. The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug & Alcohol Dependence* 108, 1 (2010), 98–109.

[29] George Mohler and Michael D Porter. 2017. Rotational grid, PAI-maximizing crime forecasts. *NIJ Report* (2017).

[30] GO Mohler, MB Short, P.J. Brantingham, FP Schoenberg, and GE Tita. 2011. Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* 106, 493 (2011), 100–108.

[31] George O Mohler, Martin B Short, Sean Malinowski, Mark Johnson, George E Tita, Andrea L Bertozzi, and P Jeffrey Brantingham. 2015. Randomized controlled field trials of predictive policing. *J. Amer. Statist. Assoc.* 110, 512 (2015), 1399–1411.

[32] Daniel B Neill. 2009. Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting* 25, 3 (2009), 498–517.

[33] Daniel B Neill. 2012. Fast subset scan for spatial pattern detection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74, 2 (2012), 337–360.

[34] Matt R Nobles, Jeffrey T Ward, and Rob Tillyer. 2016. The impact of neighborhood context on spatiotemporal patterns of burglary. *Journal of Research in Crime and Delinquency* 53, 5 (2016), 711–740.

[35] National Insititue of Justice. 2017. NIJ Real-time crime forecasting challenge. (2017). https://nij.gov/funding/Pages/fy16-crime-forecasting-challenge.aspx

[36] Eric L Piza and Jeremy G Carter. 2017. Predicting Initiator and Near Repeat Events in Spatiotemporal Crime Patterns: An Analysis of Residential Burglary and Motor Vehicle Theft. *Justice Quarterly* (2017), 1–29.

[37] Jerry H Ratcliffe and George F Rengert. 2008. Near-repeat patterns in Philadelphia shootings. *Security Journal* 21, 1-2 (2008), 58–76.

[38] Blake Shaw, Jon Shea, Siddhartha Sinha, and Andrew Hogue. 2013. Learning to rank for spatiotemporal search. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 717–726.

[39] MB Short, MR D'Orsogna, PJ Brantingham, and GE Tita. 2009. Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology* 25, 3 (2009), 325–339.

[40] Skyler Speakman, Sriram Somanchi, Edward McFowland III, and Daniel B Neill. 2016. Penalized fast subset scanning. *Journal of Computational and Graphical Statistics* 25, 2 (2016), 382–404.

[41] Toshiro Tango and Kunihiko Takahashi. 2005. A flexibly shaped spatial scan statistic for detecting clusters. *International journal of health geographics* 4, 1 (2005), 11.

[42] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 635–644.

[43] Pengfei Wang, Yanjie Fu, Guannan Liu, Wenqing Hu, and Charu Aggarwal. 2017. Human Mobility Synchronization and Trip Purpose Detection with Mixture of Hawkes Processes. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 495–503.

[44] X. Wang and D.E. Brown. 2012. The spatio-temporal modeling for criminal incidents. *Security Informatics* 1, 1 (2012), 1–17.

[45] Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. 2012. Automatic crime prediction using events extracted from twitter posts. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 231–238.

[46] Hongteng Xu and Hongyuan Zha. 2017. A Dirichlet Mixture Model of Hawkes Processes for Event Sequence Clustering. In *Advances in Neural Information Processing Systems*. 1354–1363.

[47] Tasha J Youstin, Matt R Nobles, Jeffrey T Ward, and Carrie L Cook. 2011. Assessing the generalizability of the near repeat phenomenon. *Criminal Justice and Behavior* 38, 10 (2011), 1042–1063.

[48] Andrew Zammit-Mangion, Michael Dewar, Visakan Kadirkamanathan, and Guido Sanguinetti. 2012. Point process modelling of the Afghan War Diary. *Proceedings of the National Academy of Sciences* 109, 31 (2012), 12414–12419.

[49] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. 2015. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1513–1522.