Trust Estimation of Historical Social Harm Events in Indianapolis Metro Area

Saurabh Pandey, Nahida Chowdhury, Rajeev R. Raje, George Mohler, Jeremy Carter
Indiana University-Purdue University Indianapolis
Indianapolis, Indiana, USA
{*pandey, nschowdh, rraje, gmohler, carterjg*}*@iupui.edu*

***Abstract*—Social harm involves incidents resulting in physical, financial, and emotional hardships such as crime, drug overdoses and abuses, traffic accidents, and suicides. These incidents require various law-enforcement and emergency-responding agencies to coordinate together for mitigating their impact on society. In this paper, we discuss the enhancements made to Community Data Analytic for Social Harm Prevention (CDASH) - a system that we have created for analyzing historical social harm events. CDASH predicts 'hot-spots' and displays them graphically to law-enforcement officials. The enhanced system, called Trusted-CDASH (T-CDASH), superimposes a trust estimation framework on top of CDASH. We discuss the importance and necessity of associating a degree of trust with each social harm incident reported to T-CDASH. We also describe different trust models that can be incorporated for assigning trust while examining their impact on prediction accuracy of future social harm events. To validate the trust models, we run simulations on historical social harm data of Indianapolis metro area, illustrating the behavior of each trust model and exploring their significance.**

***Keywords*-Social harm; Trust management; Hot-spots; Data cross validation.**

I. Introduction

Human interactions lead to diverse social formations establishing lawful processes within the society [1]. Pemberton [2] describes situations in which such social formations can become harmful: non-fulfillment of needs paves the way towards social harm in society. Social harm is a concept that enables criminology to move beyond legal definitions of crime to include immoral, wrongful and injurious acts that are not necessarily illegal [3]. Along with criminal activities, social harm encompasses any harm caused to the society irrespective of it being intentional or not. Thus, Hillyard and Tombs [3], consider social harm more responsive to the causes of human suffering than legally defined crimes.

There is a need to prevent and mitigate such social harm disruptions occurring in the society. Researchers have proposed various ways of alerting societies about social harm incidents. One way of dealing with social harm is through geographic profiling [4] by analyzing regions with connected crimes to identify likely areas of offenders residence. Another way is by creating machine learning modules and software tools for social harm prediction. This paper focuses on Trusted Community Data Analytic for Social Harm Prevention (T-CDASH), a web based system for capturing, analyzing, predicting and thereby mitigating social harm. It is an enhancement of our past work; Community Data Analytic for Social Harm Prevention (CDASH) [5]. T-CDASH assists in bringing together various stakeholders including law-enforcement agencies, health-care organizations, community organizations, and citizens for efficiently mitigating social harm. Such a system not only acts as an information source to these stakeholders but also can help in reducing the impact of social harm events in the society. Thereby, leading a way towards "Frugal Social Smart Cities". T-CDASH utilizes a Hawkes Point Process Service as suggested by Mohler et al. in [6] for generating social harm predictions and communicating risks to various stakeholders in the community.

Trust is an important component in any system, especially in distributed systems; where multiple, possibly unknown, entities interact together to achieve a common goal. In T-CDASH, multiple stakeholders interact with the system providing live social harm inputs. Although incidents reported by the Indianapolis Metropolitan Police Department (IMPD) and Emergency Medical Services (EMS) can be considered highly trustworthy, inputs from others, such as community organizations and citizens, may not be always trusted. Also, inaccuracies may occur while recording data reported to 911 either due to misinterpretation of reported incidents or due to selection of the incorrect incident category. Since these reported incidents are used while predicting future social harm hot-spots, entities with malicious intentions and possible inaccuracies while recording incidents by 911 operators may mislead T-CDASH. To ensure high accuracy of hot-spot predictions and thereby, efficient resource allocations, it is essential to incorporate a trust framework that will associate a degree of trust with every input reported to T-CDASH.

This paper discusses the design of the trust framework and experiments performed with it to evaluate its impact on social harm predictions. The paper also describes historical social harm data that was made available by the IMPD and EMS and the associated pre-processing and correlation of this data. Cross-validation of data using two techniques; Rolling Origin (RO) and Rolling Windows (RW) [7] [8] is performed to examine their impact on the trust framework and predictions generated for the social harm incidents.

The rest of the paper is organized as follows: Section II describes the architecture of the T-CDASH system. Section III discusses the social harm data used in the analysis along

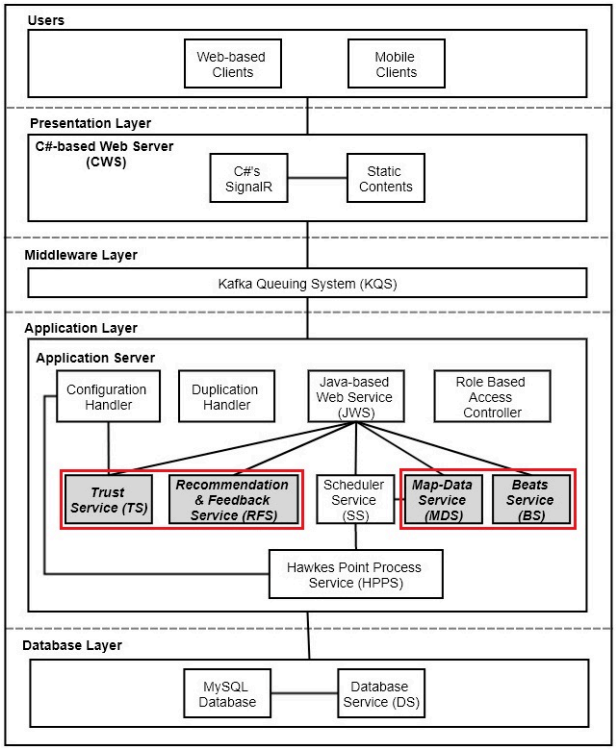

Figure 1: T-CDASH System Architecture.

with pre-processing and correlation operations performed on them. Section IV details the trust management framework of T-CDASH. Section V presents results from several experiments indicating the performance of different trust models developed as part of the framework. The paper concludes by providing insights gathered and possible directions for further research.

II. T-CDASH System Architecture

Figure 1 depicts the system architecture of T-CDASH. It is based on the principle of Service Oriented Architecture (SOA). Four layers of T-CDASH include:

Presentation Layer: The Presentation Layer helps in enhancing the User Experience by presenting social harm information in a user friendly manner.

Middleware Layer: The Middleware Layer, consisting of a Kafka Queuing System (KQS) [9], acts a communication link between the Presentation and Application layers. It helps in enhancing the fault tolerance capability of T-CDASH.

Application Layer: The Application Layer helps in processing social harm information. A key service that helps in generating social harm predictions is the Hawkes Point Process Service (HPPS). The HPPS is a self-exciting point process that allows modeling of risks and forecasting trends in social harm [6].

Database Layer: The Database Layer helps in storing and retrieving social harm information. It also stores feedbacks obtained from the police officers.

Most of the services and components in each layer are borrowed from CDASH and described in our earlier work [5]. Following are the additional services, highlighted in Figure 1, in the Application Layer of T-CDASH.

Trust Service (TS). This service implements the trust framework and helps in estimating the trust of social harm events reported to T-CDASH. Various trust models are created and experimented with as explained in Sections IV and V. The TS protects T-CDASH from misleading inputs and data recording inaccuracies during the prediction process.

Recommendation and Feedback Service (RFS). The RFS helps in achieving two key functionalities associated with T-CDASH. Firstly, it provides useful recommendations of possible actions to police officers while patrolling in hot-spot locations. Secondly, it also helps in capturing feedback from police officers regarding the actions taken by them while patrolling the hot-spots.

Beats Service (BS). The patrolling area under the jurisdiction of the IMPD is divided into several geographical sections called beats. Currently, the IMPD has divided the area of the Marion County into 78 beats. The geographic information relating to the boundary of each beat is provided by the IMPD. Each police officer is assigned to a beat and BS helps in fetching boundary information of the beat.

Map-Data Service (MDS). All the social harm hot-spots, along with beats and recommendations, are displayed on an interactive Google map. Hot-spots are updated periodically, currently every 8 hours to match police shifts, and these changes must be reflected on the map. Such updates are communicated to the users through the MDS.

III. Social Harm Data and Processing

Social Harm Data. T-CDASH uses social harm data obtained from two sources: Computer Aided Dispatch (CAD) and Uniform Crime Reporting (UCR). The data is for the year 2012-2013 and was provided by the IMPD for the Indianapolis metropolitan area.

- *CAD*: The CAD data includes social harm events reported to the IMPD through 911 calls for service [10]. The incidents reported in CAD are initial assumptions about a social harm situation. However, the actual incident and its authenticity may not be known until investigated by the police. Also, it is assumed that the description provided for the incident correctly resembles the actual incident but that may not be the case. Thus, the CAD records may not be entirely trustworthy.
- *UCR*: Each State in the United States can have its own schema for maintaining social harm records. For analysis and maintenance, it is necessary to maintain records with a common schema. For this, the Federal Bureau of Investigation (FBI) collects, publishes, and archives

social harm records in a UCR repository [11]. Since these entries are recorded post police investigations, they can be considered highly trustworthy.

Data Pre-processing. The records from each of the above data sources have their own schema and thus, it is necessary to transform them in a schema that can be used with the HPPS. Currently, T-CDASH supports hot-spot predictions for 18 different incident types [6]. It is thus necessary to map the CAD and UCR records to these 18 categories.

With CAD, the description field in the schema represents the type of social harm. Therefore, the description field is used to map the CAD records into corresponding T-CDASH records. The mapping is achieved through pattern matching; matching a description pattern with a particular T-CDASH incident code. Similarly, the UCR data is transformed. The UCR data is streamlined and hence, a direct mapping between UCR and T-CDASH incident codes without using any pattern matching technique is possible.

Data Correlation. In T-CDASH, a level of trust is associated with social harm incidents while maintaining the anonymity of the reporter. The trust is computed through an opinion model as suggested by Jøsang in [12]. With opinion model, the trust is based on three components: *belief* (*b*), *disbelief* (*d*), and *uncertainty* (*u*). These components, in turn, depend on the positive and negative evidences available for an incident [13]. To gather these evidences, T-CDASH considers three aspects associated with social harm.

- *Location*: Geo-coordinates (latitude and longitude) of the location where the reported incident occurred.
- *Day*: Date (day and month) on which the reported incident occurred.
- *Incident Type*: The category of the reported incident.

Based on these aspects, the live social harm incidents are correlated with historical social harm incidents. The historical incidents that got correlated, act as evidences for the live incident. The idea behind correlating social harm events is the assumption that if a large number of incidents, similar to the reported incident (with respect to the above aspects of the event), occurred historically, it is likely that the reported incident can be considered trustworthy.

For computing total evidences, two aspects, location and/or day, are considered. With location, a circular range of 110m (or three decimal places accuracy with respect to latitude and longitude [14]) around the reported incident is taken into consideration. This range is chosen to allow a small neighborhood area to be considered while gathering evidences. All the historical social harm incidents within this range are assumed to be contributing to the total evidences. Similarly, historical social harm incidents that occurred within a range of days (4 to 7 days) before or after the day of the reported incident, in the same month from previous years, also contributed towards the total evidences. It is important to note that these range values are parameters to T-CDASH and can be tuned for different situations.

For positive evidences, the type of incidents is considered. All the incidents present as total evidences, and having the same incident type as that of the live incident, are considered as positive evidences for the live incident. This correlation of live social harm incident with historical social harm records helps in associating trust with the live incident.

IV. Trust Management Framework in T-CDASH

Different stakeholders including IMPD, community organizations and citizens, interact with T-CDASH. As stated earlier, to ensure that the predictions generated by T-CDASH are trustworthy, there needs to be a trust framework in place. The trust framework establishes trust on social harm incidents and permits only trustworthy incidents to be considered while generating predictions. Within the trust framework of T-CDASH, five different trust models are created, compared and experimented with.

Ground-truth Model. This model considers all the inputs to be completely trustworthy and passes them to the HPPS for generating predictions. No processing or filtering is performed on any input. However, since everything is trusted by the model, it does not filter out any misleading inputs. Thus, the hot-spots generated by using this trust model may not be acceptable or correct.

Optimistic Model. In this model, a high percentage (80% to 95%) of user inputs are considered to be trustworthy. The inputs that are to be trusted are chosen randomly and passed to the HPPS for generating predictions. Remaining inputs (5% to 20%) are ignored. Since most of the inputs are accepted, this model too may allow many misleading inputs to contribute towards hot-spots generation. Hence, hot-spots generated by this model too may not be acceptable or correct.

Pessimistic Model. This model is the opposite of the Optimistic model. In this model, a high percentage (80% to 95%) of user inputs are ignored. Only a small percentage (5% to 20%) of inputs (chosen randomly) are considered trustworthy and passed on to the HPPS for generating predictions. Since, this model ignores most of the inputs, it is safe to assume that it filters out all the misleading inputs to T-CDASH. However, it may also ignore many genuine inputs thereby negatively impacting the prediction accuracy.

Average Model. In this model, half of the inputs are considered trustworthy while the remaining half are simply ignored. The choice of selecting or ignoring the input for generating predictions is random. Since, 50% of the inputs are considered, it may perform better by considering genuine inputs while ignoring misleading inputs. However, since inputs are randomly chosen, the accuracy of predictions would still be questionable.

Random Model. In this model, a set of randomly chosen inputs are considered trustworthy and used while generating

predictions. This model may perform best in a scenario when historical social harm data is not available to train the HPPS. Similar to the Average model, inputs are randomly chosen, based on a randomly generated number, and hence the accuracy of predictions would be questionable.

Opinion-based Model. This model is based on the opinion model of trust as suggested by Jøsang in [12]. As stated earlier, Jøsang's opinion model is based on *b*, *d*, and *u* which in turn depend on the positive and negative evidences as shown below.

$$b = \frac{positive\_evidence}{total\_evidence + n} \quad (1)$$

$$d = \frac{negative\_evidence}{total\_evidence + n} \quad (2)$$

$$u = \frac{n}{total\_evidence + n} \quad (3)$$

Here, *n* is the number of possible outcomes. In our work, n=2, as the incident is either trusted or it is ignored.

Any reported incident is viewed as not being either true or false but rather on the basis of subjective belief (b), disbelief (d) and uncertainty (u). Positive evidences support the incident and contribute towards higher belief while negative evidences oppose the incident and contribute towards higher disbelief. The *b*, *d*, and *u* values are generated using two methods. One method (named Random) randomly assigns values to *b*, *d*, and *u*. Thus, similar to the Random model, the accuracy of predictions generated using this random method would be indeterminate. The other method (named Heuristic) utilizes the correlation between live and historical incidents as detailed in the Data Correlation subsection of this paper for computing *b*, *d*, and *u* values. Since this Heuristic method is based on actual event attributes and their correlations with the historical incidents, it is expected to result in the generation of most accurate predictions.

V. EXPERIMENTS AND ANALYSES

Various trust models described in Section IV are implemented and experimented with, to evaluate their accuracy. In these experiments, real-time CAD and UCR data are used.

Before comparing the trust models, it is important to train the HPPS. Since the UCR data is highly trustworthy, the HPPS is trained on the 2012 UCR data. Also, real-time data is required to test the trust models. Since the CAD data is a real-time reporting of social harm incidents, CAD records of 2013 are considered for evaluating the trust models.

A baseline model having accurate predictions is required to compare the performance of trust models. Accurate predictions are generated using completely trustworthy data. This paper considers the UCR data to be completely trustworthy. It is necessary to consider all the UCR records for generating accurate hot-spots. Thus, the Ground-truth model is chosen to be the baseline model with the UCR data. As stated earlier, the CAD records are reported in real-time and prone to errors. Thus, they mimic the live incidents that will be fed to T-CDASH. With this in consideration, CAD records are fed to other models and hot-spots generated by them are compared with the hot-spots generated by the Ground-truth model. Multiple iterations are performed while comparing the models. Each iteration consists of data belonging to a particular month of 2013.

Since social harm data is a time series data, two techniques, Rolling Origin (RO) and Rolling Windows (RW) [7] [8], of time series data cross-validation are applied on the social harm records to analyze their impact on the prediction accuracy of trust models.

Table I: Performance of Optimistic, Pessimistic, Random and Average Models

Model	System	Inputs Allowed (%)	Hot-spots Matched (%)
Optimistic	RW	80	37.46
Optimistic	RW	90	36.98
Optimistic	RW	95	36.33
Optimistic	RO	80	35.60
Optimistic	RO	90	34.93
Optimistic	RO	95	34.54
Pessimistic	RW	5	49.66
Pessimistic	RW	10	47.94
Pessimistic	RW	20	46.46
Pessimistic	RO	5	49.02
Pessimistic	RO	10	48.53
Pessimistic	RO	20	45.14
Average	RW	50	42.93
Average	RO	50	39.28
Random	RW	Random	42.85
Random	RO	Random	41.62

A. Experiments with Trust Framework

In our experiments, the Ground-truth model acts as a baseline model and the accuracy of all the other models is defined in terms of hot-spots matching percentage. The hot-spots matching percentage is the percentage of hot-spots, generated by a model, that match (have the same location and incident type) with the hot-spots generated by the Ground-truth model.

Optimistic Model. With Optimistic model, three different percentages, 80, 90 and 95, of inputs were considered trustworthy. On average, the matching percentage was 35.97.

Pessimistic Model. With Pessimistic model, three different percentages, 5, 10 and 20, of inputs were considered trustworthy. On average, the matching percentage was found to be 48.29.

Average Model. With Average model, it is expected that the hot-spots matching percentage will be approximately the average of the matching percentages of the Optimistic and Pessimistic models. On average, the matching percentage was found to be 41.10, which is as expected.

Random Model. The Random model is non-deterministic as it randomly considers a set of inputs to be trustworthy. On average, the matching percentage was found to be 42.23.

These experimental results for the above models are summarized in Table I.

Opinion-based Model. In Opinion-based model, two methods (Random and Heuristic) are used to assign values to b, d and u. In Random method, if the randomly generated belief value for an incident is above a chosen threshold belief value, the incident is considered for generating hot-spots. Similarly, if the randomly generated disbelief value is above a chosen threshold disbelief value, the incident is ignored. In all other scenarios, the trust on the incident is uncertain and it is either considered or ignored randomly. In Heuristic method, data correlation, as described in Data Correlation subsection of section III, is considered for assigning values to b, d and u. Table II depicts the percentage of hot-spots matched between the hot-spots computed by the two methods of Opinion-based model and the Ground-truth model while considering different threshold percentages of belief and disbelief. On an average, the matching percentage of Random method was 40.63 and Heuristic method was 47.59.

B. Observations

Best Model. From Tables I and II, it can be seen that the Pessimistic model (allowing 5% inputs for processing) performs best when compared to all the other models. The Pessimistic model is followed by the Opinion-based model with the Heuristic method, the Average model, and lastly the Optimistic model in that order of matching percentages. Since the performance of the Random model and the Random method of Opinion-based model are indeterminate, it is not appropriate to compare them directly with other models.

One reason for such hot-spot matching behavior is due to the fact that many incidents reported to CAD are not reported in the UCR in the same way. This is because the incident may have never occurred or after investigation, it was found that some incident other than the actual one was reported. For example, an incident of Simple Assault is reported in CAD. However, during the investigation, it was found that it was a case of Homicide. Another reason is that many incidents are investigated directly by the IMPD without ever being reported in CAD. Thus, CAD and UCR records differ considerably. This justifies the fact that models considering smaller percentages of CAD data for generating hot-spots present higher hot-spot matching accuracy.

Table II: Opinion-based Model

Method	System	Location?	Day?	b Threshold (%)	d Threshold (%)	Hot-spots Matched (%)
Random	RW	No	No	50	50	42.03
Random	RO	No	No	50	50	39.24
Heuristic	RW	Yes	Yes	50	50	49.47
Heuristic	RW	Yes	No	50	50	49.59
Heuristic	RW	No	Yes	50	50	48.18
Heuristic	RW	Yes	Yes	70	50	47.90
Heuristic	RW	Yes	Yes	50	70	46.53
Heuristic	RW	Yes	Yes	80	80	46.06
Heuristic	RW	Yes	Yes	10	10	46.73
Heuristic	RW	Yes	Yes	30	30	49.82
Heuristic	RO	Yes	Yes	50	50	48.33
Heuristic	RO	Yes	No	50	50	48.81
Heuristic	RO	No	Yes	50	50	47.42
Heuristic	RO	Yes	Yes	70	50	46.98
Heuristic	RO	Yes	Yes	50	70	45.64
Heuristic	RO	Yes	Yes	80	80	45.17
Heuristic	RO	Yes	Yes	10	10	45.87
Heuristic	RO	Yes	Yes	30	30	49.03

These experiments highlight that more the number of inputs ignored, higher is the hot-spot match percentage. Accordingly, both the Pessimistic model and the Opinion-based model with the Heuristic method have the highest match percentages. However, it may not be always advisable to ignore a large percentage of inputs. Consider a scenario where a critical live incident is reported. Since both models ignore most of the inputs, even multiple reports by different users reporting a critical incident may get ignored. This may negatively impact the predictions generated by the system.

It is also important to note that both the Pessimistic model and the Opinion-based model with the Heuristic method have approximately equal hot-spot matching percentages. Since, the Opinion-based model with the Heuristic method takes a more informed decision while considering or ignoring inputs for generating predictions rather than deciding randomly (e.g., the Pessimistic model), it is considered better when compared to the Pessimistic model.

Seasonal Performance of Models. All the experiments are performed on the monthly data from 2013 and then averaged out over the entire year. The hot-spots generated for each month are analyzed and compared. A critical observation is that the percentage match remained close to the average value and did not display any drastic deviations in any month of the year. Thus, a key insight with these experiments is that the performances of various models are agnostic from seasonal changes that may occur in social harms occurring in the society.

Effect of Data Cross Validation. Two cross validation techniques for the time series data: RO and RW are used in our experiments. The difference between the techniques is that the RO method considers all the records while generating predictions while the RW method eliminates the oldest records. The result of the experiments performed with both techniques are depicted in Tables I and II. Tashman in [7] indicated that pruning of old records may be unnecessary if the prediction service considers data in a weighted manner, mitigating the influence of any data from distant past. The HPPS service generating hot-spots in T-CDASH considers data in a weighted manner. The experiments indicate that the matching percentages remain almost the same no matter which cross validation technique is used. This confirms to the observations presented by Tashman in [7].

VI. RELATED WORK

A lot of research has been carried outexperimental in analyzing and predicting social harm. Bogomolov et al. [15] predicted crimes using mobile and demographics data. Crime hot-spots were predicted using the Random Forest algorithm with 70% accuracy in the metropolitan city of London. Yu et al. [16] created a Cluster-Confidence-Rate-Boosting (CCRBoost) algorithm for generating spatio-temporal crime patterns by analyzing historical crime records. CCRBoost predicted residential burglary with 80% accuracy in a north-eastern US city. T-CDASH, however, utilizes an approach proposed by Mohler et al. [6] that focuses on using modulated Hawkes Process Model for predicting social harm.

This paper focuses on the trust aspect of social harm events. Significant literature is available on establishing trust in distributed systems. Furtado et al. [17] describe the reputation-based trust management methodology in Wiki-Crimes system. WikiCrimes, an application for reporting live crimes, uses a reputation model [18] for generating reputation scores for the registered users. The reputation score increases with each genuine crime reported and it is used by the application for associating trust with the live reported events. However, in WikiCrimes, users are required to register with their name and email address. Jøsang [12] introduced an opinion model for estimating the trust of events based on b, d, and u. Ceolin et al. [13] created a trust algorithm for computing b, d, and u as introduced by Jøsang in [12]. The algorithm was applied in the maritime domain for estimating trust of messages to track ships. To maintain user anonymity, T-CDASH utilized Jøsang's opinion model for estimating trust of social harm events.

VII. CONCLUSION AND FUTURE WORK

This paper presents T-CDAH along with a trust framework and associated data mappings. The experiments indicate that live incidents reported to T-CDASH cannot be blindly trusted as they can mislead the system. The experimental results also highlight that considering or ignoring incidents based on certain heuristics can help in making better predictions. Another key outcome is that the accuracy of various models does not depend on the seasonal changes. CDASH system is currently being used by the IMPD for field trials to analyze its impact on reducing social harm. With T-CDASH, we aim at establishing trust between various stakeholders while achieving optimal resource allocation. This will help in reducing social harm costs in society. Thus, it will lead towards "Frugal and Smarter Cities".

Future efforts will incorporate additional trust models that consider other aspects associated with social harm such as the number of times an incident is reported and the incident severity while estimating an incident's trustworthiness. Other model comparison metrics such as Earth Movers Distance [19] will also be incorporated for measuring the hot-spot matching accuracy of the models. Training the HPPS with recent UCR records and testing the trust models with incidents reported in real-time while analyzing them is another future direction. Additionally, the trust framework can be applied in other domains such as telecommunications and social media, to assess its usability.

VIII. ACKNOWLEDGEMENTS

This project is supported in part by NSF grants CNS-1737585, SES-1343123, and DMS-1737996. G.M. is a co-founder and serves on the board of PredPol, a predictive policing company.

REFERENCES

[1] K. Lasslett, "Crime or social harm? a dialectical perspective," *Crime, Law and Social Change*, vol. 54, no. 1, pp. 1–19, 2010.
[2] S. Pemberton, "Social harm future (s): exploring the potential of the social harm approach," *Crime, Law and Social Change*, vol. 48, no. 1-2, pp. 27–41, 2007.
[3] "Criminology beyond crime," *The Open University*, 2016.
[4] D. K. Rossmo, *Geographic profiling*. CRC press, 1999.
[5] S. Pandey, N. Chowdhury, M. Patil, R. R. Raje, C. Shreyas, G. Mohler, and J. Carter, "Cdash: Community data analytics for social harm prevention," pp. 1–8, 2018.
[6] G. Mohler, J. Carter, and R. Raje, "Improving social harm indices with a modulated hawkes process," *International Journal of Forecasting*, vol. 34, no. 3, pp. 431–439, 2018.
[7] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," *International journal of forecasting*, vol. 16, no. 4, pp. 437–450, 2000.
[8] "Time Series Cross-Validation," https://cran.r-project.org/web/packages/greybox/vignettes/ro.html.
[9] "Apache Kafka," https://kafka.apache.org.
[10] "Standard Functional Specifications for Law Enforcement Computer Aided Dispatch (CAD) Systems," https://it.ojp.gov/documents/LEITSC_Law_Enforcement_CAD_Systems.pdf.
[11] "Uniform Crime Reporting," https://www.fbi.gov/services/cjis/ucr.
[12] A. Jøsang, "Artificial reasoning with subjective logic," vol. 48, p. 34, 1997.
[13] D. Ceolin, P. T. Groth, and W. R. Van Hage, "Calculating the trust of event descriptions using provenance." 2010.
[14] "Latitude and Longitude Precision," https://gizmodo.com/how-precise-is-one-degree-of-longitude-or-latitude-1631241162.
[15] 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.
[16] C.-H. Yu, W. Ding, P. Chen, and M. Morabito, "Crime forecasting using spatio-temporal pattern with ensemble learning," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2014, pp. 174–185.
[17] V. Furtado, L. Ayres, M. De Oliveira, E. Vasconcelos, C. Caminha, J. DOrleans, and M. Belchior, "Collective intelligence in law enforcement–the wikicrimes system," *Information Sciences*, vol. 180, no. 1, pp. 4–17, 2010.
[18] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision support systems*, vol. 43, no. 2, pp. 618–644, 2007.
[19] "Earth Movers Distance," http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/RUBNER/emd.htm.