Spatio-temproal prediction of crimes using network analytic approach

Saroj Kumar Dash*‡, Ilya Safro*¶, Ravisutha Sakrepatna Srinivasamurthy†§
*School of Computing, Clemson University
†Department of Electrical and Computer Engineering, Clemson University
Email: { ‡sdash, ¶isafro, §rsakrep }@clemson.edu

Abstract—It is quite evident that majority of the population lives in urban area today than in any time of the human history. This trend seems to increase in coming years. A study [5] says that nearly 80.7% of total population in USA stays in urban area. By 2030 nearly 60% of the population in the world will live in or move to cities. With the increase in urban population, it is important to keep an eye on criminal activities. By doing so, governments can enforce intelligent policing systems and hence many government agencies and local authorities have made the crime data publicly available. In this paper, we analyze Chicago city crime data fused with other social information sources using network analytic techniques to predict criminal activity for the next year. We observe that as we add more layers of data which represent different aspects of the society, the quality of prediction is improved. Our prediction models not just predict total number of crimes for the whole Chicago city, rather they predict number of crimes for all types of crimes and for different regions in City of Chicago.
Reproducibility: our code is available at https://goo.gl/V1yVLk.

Index Terms—Data Fusion, Network Analysis, Crime Data Analysis, Data Analysis for Government

I. INTRODUCTION

It is quite evident that majority of the population lives in cities [5]. Modern and effective police system plays a key role in city safety that is especially critical in megalopolises with culturally diverse population. In the era of big data, analyzing historical events of criminal activity might give some critical information to implement intelligent policing across the city. Many cities have realized this and crime data have been made publicly available for research. There exist different types and formats of crime datasets [11], [23], [26] which, however, can be summarized as collections of crime records that typically contain time, location and short description type of the crimes without providing too many details due to such obvious reasons as privacy and confidentiality. We discuss some of them in Section II. However, not much research has been done to improve the crime prediction by fusing the crime data with other types of city data which is the goal of our research.

Our contribution

In this paper, we analyze the Chicago crime dataset and show how the quality of crime prediction can be improved if fused with other city datasets [11]. In contrast to previous approaches, we use network analytic techniques to connect different regions of the city and fuse them with different data types to predict next crime pattern in a given region. We fuse crime data with such information as sanitation, schools, police stations, libraries, and 311 services to make train the prediction models. We observe that using different types of data which are not directly related to crime increases the accuracy of models and information from distant regions of the city may give insight on criminal activities of another distant part of the city.

This paper is organized as follows. We discuss the related work in which relevant spatial and temporal predictive modeling is discussed in Section II. We introduce our prediction method including necessary definitions and datasets used in Section III. The computational results that cover the entire Chicago area and all types of crimes are presented in Section IV. Finally, we conclude and discuss potential research directions in Section V.

II. RELATED WORK

The area of crime data mining has shown increasing popularity and demand in recent years. One of the most important reasons is the increasing crime rates. The "Economist" indicates that there was a 22% increase in the number of murders in Chicago in recent times [25]. As a result of massive urbanization, the situation with crime in many cities is similar to that in Chicago. Consequently, there was a rapid increase in research on predictive crime analysis for better policing and to stop gang activities or any other major criminal activities. To predict crime, several classical statistical techniques are being applied, whose results are then used to take correct steps taken by the police to prevent crimes. Statistical methods such as the auto-regression modeling [9], association of crime events [8], and random walks [24] are being used in state of the art analysis of crimes. In this section, we review several relevant crime analysis methods.

In [8], Brown and Hagen discuss an associative approach to find similarity among the crime records in the available data. The authors find rules to associate each crime event data with importance attributes which qualitatively and quantitatively adds weights to the attributes. The total similarity measure between crime records is computed which is the summation of the products of the weights of that attribute and the normal association value between the two crime event attributes. The weights are calculated based on the importance of the attribute

in a particular crime record that is considered for finding the similarity. Once the weights are calculated, the *total similarity measure* (TSM) is computed to find the crimes which are similar to each other in nature. For an interpretability of crime attributes, a similarity measure of attributes was developed. Then using the Bayers Rule method, a relative association between different crime records is found. This method is compared to the manual crime similarity search which is being used in the police department so far. The method was found useful for its improved performance and applicability in the crime trend analysis.

In [24], the goal of Tayebi et al. is to predict the probability of an activity space of a criminal. Similar to our approach, the authors use network analysis to predict the crime activity area at certain location. The paper is based upon hotspot (high intensity crime area) recognition of the criminal and then the spatial information for building an activity space of the criminal. This spatial criminal activity space is used to see the activity of the criminal and then for taking appropriate actions. It is based upon the basic thinking that a criminal does more crime in a vicinity where the criminal is more familiar with and also where the victim is new. Based on the hotspot recognition and pattern, the next crime location is predicted. It uses the random walk model to predict the criminals home location and based upon that, the method identifies the next crime location. This method is also mentioned in all the network analysis approaches discussed in [16].

Spatial hot spot analysis in one of the main streams in crime prediction given a historic data for more than two decades. The idea introduced in [7] can be summarized as clustering training criminal data into hot spots using mixture models and density estimation with the hypothesis that crimes are likely to happen within such clusters (hot spots). However, these methods do not take into account socio-environmental information such community quality, people awareness, police station proximity, and quality of schools. Also, lack of historical information on particular areas make such models useless on them. Other works in this class of methods include [17], [21]. Our work is complementary to such methods because, using the network analytic approach, we find similar areas whose historical information is also used in prediction.

The objective of [9] is similar to ours. The authors take the Chicago crime dataset (which is also part of our input) and creates the time series data to predict number of crimes at a *specific* area on a week by week basis. The experiment is carried out in a particular Eastern region of Chicago. All of the crime data from 2001 to 2012 is used to train the model which is used to predict the crime numbers for year 2013. For this goal, a statistical method named *autoregressive integrated moving average* (ARIMA) was applied. This method differentiates itself from the normal regressive method by subtracting a differencing transformation to the data before applying the regression. The error subtraction in ARIMA makes the prediction more unbiased as the errors are always being subtracted from the data. The prediction has an error rate of 14.7%. For a one-year of prediction of the number of

crimes, it showed an accuracy of 84%. For the second year, the accuracy was dropped to 80%. We also use the autoregression method as one among the three methods to verify the quality of the crime prediction on our fused data.

In [4], Bastomski et al. consider structural embeddedness within co-offending network to predict homicide rate. According to this paper, the social structure of a neighborhood typically impacts the number of crimes that occur in that neighborhood. If two neighborhoods are structurally similar (similar economic conditions, similar quality of education, etc.) then the types of crime and number of crimes will also be similar. *We use and extend this observation in our methods in which we assume that a prediction of crimes can be made using similarities of several attributes of neighborhoods.* In particular, it is important to mention that these neighborhoods need not be adjacent or connected spatially. This also indicates a better chance of crime spread between similar neighborhoods. The network formed when two or more individuals engage in a criminal activity is known as co-offending network. This paper considers residency of co-offenders as nodes and the neighborhoods belonging to these co-offenders as edges. Using this network, the k-core measure is used as an independent parameter and with the help of other measures, the parameter for embeddedness is formed. Then by using ordinary least squares, the homicides rates are predicted. Finally, the paper confirms that inter-neighborhood ties are not explained by geographic proximity alone.

An approach to predict links between pairs of individuals with examples in the Italian mafia was proposed by Bersluconi et al. in [6]. In this paper, the individuals are considered to be the nodes and they are connected based on the number of telephone calls. Three networks are considered, namely, *wiretap records network*, *arrest warrant network* and *judgment networks*. Weight of the links in the wiretap records network is proportional to the number of phone calls between the individuals. Arrest warrant network is obtained by reducing wiretap records. Similarly, judgment network is obtained by reducing arrest warrant network. This reduction is done by removing the connection between individuals if the number of calls is small. The authors argue that the removal of links is characterized by low betweenness in the network. That is, even if a link is removed, the corresponding individuals are connected indirectly in some way. The second hypothesis is that nodes are likely to be connected if they are similar. In this case, individuals who collaborate rarely will have dissimilar attributes such as interests and background. This hypothesis is proved to be correct and hence author uses this concept to predict missing links. If two nodes are have high similarity scores, there is a chance of connection between these criminals. Author also tests Katz index similarity and structural perturbation method which also gives similar results.

In [2], Almanie et al. develop a method to predict criminal hot spots and prominent crime types in a given area and also the frequency of crimes for a given time period.

This paper applies three methods, namely, the Apriori algorithm, Naive Bayes and Decision Tree classifiers on the Denver

and Los Angeles crime datasets. They consider location, time, crime type data to predict the frequency of crime events. Out of all methods applied, the results of Naive Bayes were found of greatest accuracy which is 54%. The paper gives a basis to apply statistical methods to the crime data and also to consider the spatial data for crime prediction.

In [15], Henderson et al. propose a very unique method in which market prediction algorithm [20] is applied to predict Chicago crime rates. The authors hypothesize that if a crime market is setup where people can trade contracts on crime rates, the market with the help of information gathered from people can be used to predict the crime rates. Market prediction employs a natural feedback loop. If someone believes (on the basis of some information) that crime rate may go up, then she will bet highly on the increase in crime rate and vice versa. Hence, it forms a natural feedback loop.

III. Crime prediction method

Our objective is to predict crime levels for different types of crimes and a given year based on previous years criminal activity and other social aspects. We observe that augmenting the crime data with other social aspects such as education, and economic conditions give few insights for predicting the crime and improve the quality of prediction. Also, social aspects of other similar communities give out critical information for predicting the crime pattern for next year. The experiments are conducted on the Chicago crime dataset [11] provided by the Police Department for public use.

A critical component of the proposed method is to fuse various types of social and historical information in a network which helps to find relationships between different communities within the city. The extracted relationships are incorporated in our prediction model. Model performance is demonstrated on predicting crime rates for year 2015 and compared with similar models that are not augmented with additional social information and relationships between communities. We also compare our results with previously proposed autoregressive model [24] which does not take any information other than historical crime data into consideration. In the following subsections, we discuss the datasets used, an approach to the problem, and the prediction model.

A. Datasets

The City of Chicago provides the city data (including the crime dataset) for public use through the Chicago Data Portal at http://data.cityofchicago.org/. The crime dataset spans from a time period of 2001 to present and has record of all criminal activities happened in Chicago city except for last seven days as the data is updated every week. Below is the list of all the data sources that we have used in our analysis.

- Crime dataset includes 6.6 million crime data records from Chicago from 2001 to present. The total size of data is 1.41 GB.
- Chicago library data includes library locations and details with geographical coordinates of 80 libraries in Chicago city.

- Library visitors by location: This data gives information about the number of visitors in each month for all 80 libraries in Chicago city.
- Police station data: Gives us an information about the location of 25 police stations for 25 districts in Chicago city.
- Police districts boundaries: The areas for which a police station is responsible for. Used to connect communities, crimes and police stations.
- Dataset of all 311 service (non-emergency municipal services) requests. This includes the all-lights service requests, transit ride requests, pot holes requests, sanitation community requests and vacant house information. The size of this data is approximately 450 MB.
- Chicago public school progress information: This data is available from 2011 to 2017. This dataset contains locations and related data of 188 schools in the Chicago area.
- Chicago school average ACT (American College Testing) scores: The average ACT scores were collected for all the 188 schools.

The crime data is available from year 2001 till now. However, most of the other data can be traced back to 2011 only, so for the training of our models, we used all available data from 2011 to 2014. The test sets were created using the 2015 data.

B. Approach

The objective of this work was not just to predict the overall crime activity in Chicago city. Instead, we are interested in predicting the crime activity for *various regions* of Chicago city. There are several ways to divide Chicago city into geographical regions and one of them, defined by the Social Science Research Committee at the University of Chicago, is called *community division* [22]. Chicago can be divided into 77 communities and the map of the communities in Chicago is readily available from City of Chicago portal (see Figure 1). The communities are officially recognized by the City of Chicago. The 77 communities can be seen in the city of Chicago Portal for Dataset [11]. The community number in the dataset tells about the spatial resolution of Chicago. This spatial resolution could be improvedlater by considering the exact geospatial coordinates of the crime reported. For simplicity we have considered the resolution of the community area in our experiments. We found that among all the datasets that we mentioned in the above section the coordinates or the spatial location of the crime was always available in all the different datasets. Thus we took community or the location of crime reported as a unifying attribute to build the network by connecting all the datasets through there location. Because the reasons of illegal activities depend on social factors, it is clear that two communities might share similarity in crime patterns if they are neighbors to each other since social factors typically do not change drastically between neighboring communities. However, it has been observed that two communities can be similar in crime activity even if
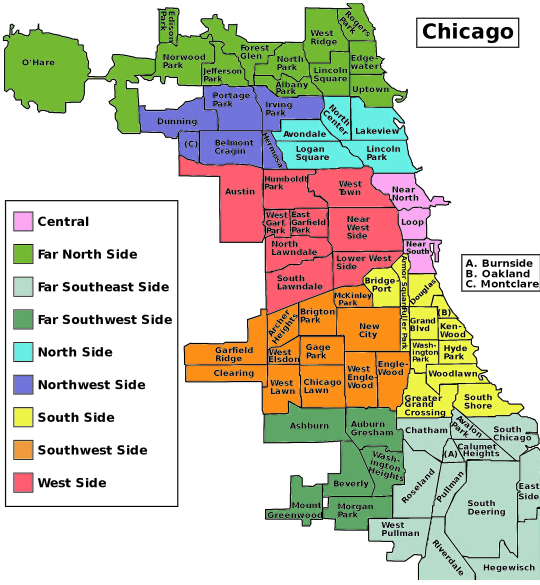
Fig. 1. Chicago is divided in 77 community areas which are well defined and do not overlap. Each community area has one or more neighborhoods in it. Image courtesy www.thechicago77.com
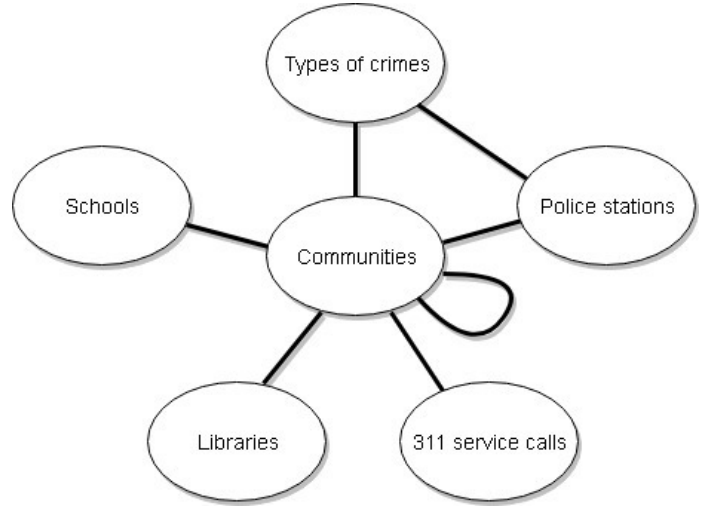

Fig. 2. Schematic representation of all node and edge types.

they are not neighbors. This is because both communities can have similar social structure [4] which may help with a predictability of community attributes (or features). To explore the social structure, we built a network of 77 communities and computed the similarity matrix between the communities which was used to extract two most similar communities for each reference community. All predictions described below are computed using the information from two most similar communities.

Once we get features to predict criminal activity for the given year, we use two models to make the prediction. Those are polynomial regression and support vector machine. We compare our results with auto-regression model as it is just an extrapolation algorithm. The results will be discussed in section IV.

1) Building Network: We define a multi-layered network $\mathcal{N}$ with an underlying edge-weighted undirected graph $G = (V, E)$ with $|V| = n$ and $|E| = m$, in which we fuse different information sources. The set of nodes is defined as a union of several layers

$$V = \mathcal{C} \cup \mathcal{T} \cup \mathcal{S} \cup \mathcal{P} \cup \mathcal{L} \cup \mathcal{R},$$

where, in our specific experiments with Chicago city,

- $\mathcal{C}$ is the set of 77 communities.
- $\mathcal{T}$ is the set of crime primary types defined by the Police Department (represented by Illinois Uniform Crime Reporting (IUCR) Codes [11]), $|\mathcal{T}| = 34$. Examples of the primary types include theft, narcotics and robbery.
- $\mathcal{S}$ is the set of 188 schools.
- $\mathcal{P}$ is the set of 25 police stations.
- $\mathcal{L}$ is the set of 80 libraries in Chicago.
- $\mathcal{R}$ is the set of 8 types of 311 service requests, namely, sanity, pot holes, lights one, light all out, lights alley,

trees, and vacant.

The set of edges E (see schematic backbone of the network in Figure 2) includes different types of connections between nodes, namely,

- A community $c \in \mathcal{C}$ is connected to a crime type $t \in \mathcal{T}$ if that t was registered in c and number of such crimes determines the edge weight. Communities are also connected with each other if they share a border.
- A school $s \in \mathcal{S}$ is connected to community $c \in \mathcal{C}$ if s is located in c. Edge weights between schools and communities are determined by the schools average ACT scores. The higher the ACT scores, the higher the standard of education.
- Police stations can belong to more than one community. A police station $p \in \mathcal{P}$ is connected to all communities in $\mathcal{C}$ that it belongs to according to [11] and also to the neighboring communities in case they are not connected directly.
- The police station node $p \in \mathcal{P}$ is connected to crime type $t \in \mathcal{T}$ based on where the crime has happened. The pt edge weight equals the number of times crime type t has been reported to police station p.
- Each library $l \in \mathcal{L}$ is connected to its community $c \in \mathcal{C}$ in which it is located. The edge weight is the number of visitors visiting that library. Large edge weights indicate educated community and, hence, may impact levels and types of crime.
- Every service type request to 311, $r \in \mathcal{R}$, is connected to its community $c \in \mathcal{C}$ in which request was registered. The edge weight is the number of requests of type r in c. A big number of requests indicates that community is active in fixing their issues.

The weighting function on edges is denoted by $w : E \to \mathbb{R}_{>0}$. Prior adding all layers of edges to G, they are normalized with respect to themselves using the min-max normalization. This normalization brings all values to $[0, 1]$ interval and
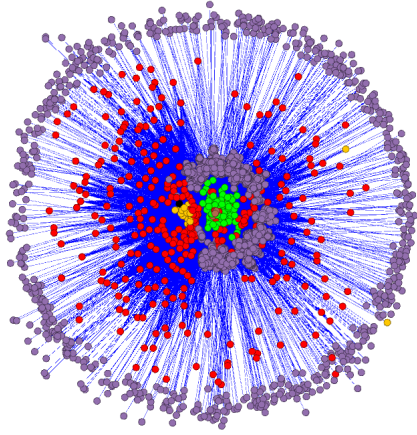
Fig. 3. Visualization of the complete network for year 2011. Shows complete network with crime type nodes and the police network connected to each other.Nodes for this network are community(green), crime types(red), Schools(violet),Police stations(orange) and sanitations(brown).

is important to avoid numerical problems that affect the prediction algorithms. An example of a resulting network is visualized in Figure 3.

All events described in our datasets contain information about the time. For our analysis we build one network for each month for the period of 2011-2015 (total 60 networks, $\mathcal{N}_i$, $1 \leq i \leq 60$). The dynamic nature of criminal activities is addressed with the analysis of all $\mathcal{N}_i$ from which corresponding similarity matrices M_i (see Section III-B2) are built.

2) Similarity: To get the top two similar communities for each community in $\mathcal{C}$ and to generate similarity matrices $\mathcal{N}_i$ per month, we used random walk similarity [24]. Random walk similarity is a metric frequently used in network analysis in which the Markov chain model is defined on transition probability for a random walk to move from one node to another. Defining a transition matrix $P \in \mathbb{R}^{n \times n}$, where $p_{ij} = \frac{w_{ij}}{\sum_k w_{ik}}$ is the probability of transition from i to j, we can compute the similarity between two nodes in V using the pseudo-inverse of the Laplacian of G [13].

Using the above method to compute the random walk similarity between two nodes we obtained the similarity between all pairs of communities and formed a similarity matrix. The features from top two similar communities are also used as features for the prediction model explained in the next section. It is important to mention that our experimental observation confirms that neighbor communities may not always be similar and similar communities may not always be neighbors.

C. Features

Besides the crime data, for each community, we define additional features for the prediction models. These features include the number of police stations in that community, number of 311 service calls registered, number of schools and libraries that are connected to that community. We also consider this information for top two similar communities.

Below we describe the set of features for each data point $p_{i,c}$ that corresponds observations in month i, $1 \leq i \leq 60$, in community $c \in \mathcal{C}$ whose top two similar communities are found to be $a, b \in \mathcal{C}$. The vector $p_{i,c}$ is a concatenation of the following vectors and values:

- a vector whose entries are $cr_{i,c,t}$, the numbers of crimes of types $t \in \mathcal{T}$ observed in c in month i, where $|\mathcal{T}| = 34$ according to Illinois Uniform Crime Reporting (IUCR) Codes [11];
- same vector with entries $cr_{i,a,t}$ for community $a \in \mathcal{C}$;
- same vector with entries $cr_{i,b,t}$ for community $b \in \mathcal{C}$;
- $(p_{i,c}, p_{i,a}, p_{i,b})$ - numbers of police stations connected to a, b and c;
- $(l_{i,c}, l_{i,a}, l_{i,b})$ - numbers of visitors to libraries connected to a, b and c in month i;
- $(s_{i,c}, s_{i,a}, s_{i,b})$ - numbers of schools connected to a, b and c in month i;
- $(sc_{i,c}, sc_{i,a}, sc_{i,b})$ - numbers of 311 service calls registered in a, b and c in month i.

The resulting vectors $p_{i,c}$ were corrected using min-max normalization and used as an input for prediction models.

D. Crime prediction

To predict the number of crimes in a given community for a given year and month, we apply and compare three types of regression models, namely, polynomial regression, support vector regression, and auto-regressive model. Each technique is briefly explained below.

1) Polynomial Regression: Initially, we tried to use linear regression but because the obtained results were not satisfactory, we shifted to polynomial regression of degree two [18]. Experiments with degree greater than two indicate that this method suffers from over-fitting the data. A polynomial regression of degree two is a process of fitting a quadratic polynomial to the data. In our problem, we do not anticipate high-oscillatory shape of the data because under very rare circumstances the number of crimes in the megalopolis can quickly change in any direction. However, it is typically also not a straight line (like in linear regression) because the number of crimes exhibit periodic behavior. If $x_o, x_1, ..., x_n$ represent our feature set and y represents the number of crimes of a certain type then the relationship between y and feature set can be expressed as

$$y = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_2 + a_4 x_2^2 + ... + a_{n-1} x_n + a_n x_n^2 \quad (1)$$

where $a_0, a_1, ..., a_n$ are the coefficients of features x_i which are also called weights. By using either pseudo-inverse or gradient descent, weights can be determined.

2) Support Vector Regression: The support vector regression (SVR) technique [12] uses same principles as the famous classification algorithm support vector machine (SVM) [27]. In the core of SVM there is a convex quadratic programming problem (QP) in which support vectors that identify a separating hyperplane from the rest of the training data are discovered. In the support vector regression, the goal is to predict the data point value rather than its discrete class.

In our implementation, we used LibSVM [10] to predict the number of crimes of a certain type for a given community and RBF function was chosen as the kernel. For our size of the data, the complexity of the kernel-based SVR was not prohibitive, however, we observed that in some cases, a tuning of regularization and kernel parameters with model selection methods is beneficial. For this purpose and for a larger data there exist several scalable techniques such as [19], [29].

3) Auto-regressive model: In [9], the auto-regressive model was suggested for the analysis of crime data. Similar to other linear models, the variable of interest in auto-regressive model is predicted using a linear combination of its previous values. However, this model also adds a stochastic term on which the variables depend which is attractive for crime prediction because it makes an attempt to model data uncertainty that we have in crime prediction which in reality depends on many social factors. Because we just had 5 years of data, we chose the lag in this model to 2. That is, previous two year's crime values becomes the input for the prediction model.

IV. COMPUTATIONAL RESULTS

Crime type	2011	2012	2013	2014	2015
Arson	504	469	364	397	453
Assault	20411	19898	17971	16900	17041
Battery	60458	59134	54003	49447	48910
Burglary	26619	22844	17894	14570	13183
Concealed carry license violation				15	34
Crime sexual assault	1471	1409	1272	1325	1365
Criminal damage	37332	35854	30853	27798	28672
Criminal trespass	8659	8215	8135	7539	6401
Deceptive practice	12569	13515	13581	15466	15676
Gambling	736	724	596	393	310
Homicide	437	505	422	426	499
Human trafficking			2	2	13
Interference with public officer	1048	1228	1281	1398	1308
Intimidation	171	156	134	116	122
Kidnapping	266	236	242	220	190
Liquor law violation	619	573	465	397	292
Motor vehicle theft	19387	16492	12582	9912	10070
Narcotics	38605	35488	34127	28995	23837
Non-criminal		6	7	27	35
Non-criminal (subject specified)		2		1	
Obscenity	40	26	24	38	46
Offense involving children	2329	2197	2331	2358	2265
Other narcotic violation	5	6	5	10	5
Other offense	20189	17479	17988	16972	17541
Prostitution	2424	2204	1652	1626	1322
Public indecency	13	17	10	10	14
Public peace violation	3095	3007	3135	2903	2422
Robbery	13982	13485	11820	9800	9638
Sex offense	1071	1051	1019	958	972
Stalking	181	207	153	140	154
Theft	75146	75458	71524	61548	57319
Weapons violation	3880	3907	3246	3114	3362

TABLE I

TOTAL NUMBER OF CRIMES OF ALL TYPES REGISTERED IN THE CITY OF CHICAGO POLICE DEPARTMENT IN TRAINING (2011-2014) AND TEST (2015) SETS.

The goal of computational experiments was to demonstrate

a high-quality prediction of the number of crimes for each crime type in each community within the city of Chicago. This significantly extends experiments presented in [9] and introduces new research directions. Because predicting the crimes is a quite complicated task because of the dependence of data on a variety of social, economic and political factors, our second goal was to demonstrate that network constructed and used to fuse various data sources does improve the quality of prediction. To the best of our knowledge, no work demonstrates a prediction of all crime types for each community on this and other comparable datasets, so, unfortunately no fair comparison with other methods can be presented. Our code and results are available at https://goo.gl/V1yVLk.

The comparison is presented in Figures 4, and 5 using box plots of all three techniques and all crime types for 12 months in 2015. The training set was chosen to be 2011-2014. In Table I, we present the number of registered crimes in both sets. The quality of prediction is evaluated using Root Mean Square Error (RMSE) over all communities for a certain crime type $t \in \mathcal{T}$ (i.e., the more the result or mean value is closer to zero the more accurate is the prediction), namely,

$$\mathsf{RMSE}(t) = \sqrt{\frac{\sum_{m=1}^{12}\sum_{c\in\mathcal{C}}(cr_{i,c,t} - \tilde{cr}_{i,c,t})^2}{\text{total \# of crimes of type } t}},$$

where $\tilde{cr}_{i,c,t}$ is the predicted number of crimes of type t in month i, at community c. For each crime type, the corresponding RMSE box was computed over 77 (communities) $\times$ 12 (months in test year 2015) = 924 values.

In order to demonstrate how other social information can affect the prediction we have created two types of networks which help us to identify two most similar communities, namely, *only-crime* and *full* networks. The only-crime network's set of nodes contains only three layers: communities, crimes, and police stations, $V = \mathcal{C}\cup\mathcal{T}\cup\mathcal{P}$, i.e., the information that can be extracted from the police department dataset. The edge set generation and the rest of network construction remained the same. In the full network, we fuse information from all datasets as we describe in Section III.

In Figure 4, we show 11 most frequent types of crimes that the police identifies in megalopolises. The usefulness of full network and data fusion in the prediction is immediately observed when the polynomial regression (green boxes) is applied. All box indicators, namely, median, maximum, minimum, first and third quartiles indicate an improvement of the prediction quality. The support vector regression (red boxes) always demonstrate the highest quality prediction that in many cases significantly improves the auto-regression proposed in [9].

In Figure 5, we present a similar comparison for all 34 types of crimes and observe similar results with SVR being a clear winner out of three regression methods. An important conclusion we make based on our results is as follows. The SVR method (similar to regular support vector machines) maximizes the separating margin using support vectors discovered in training data. Usually, the number of support vectors is
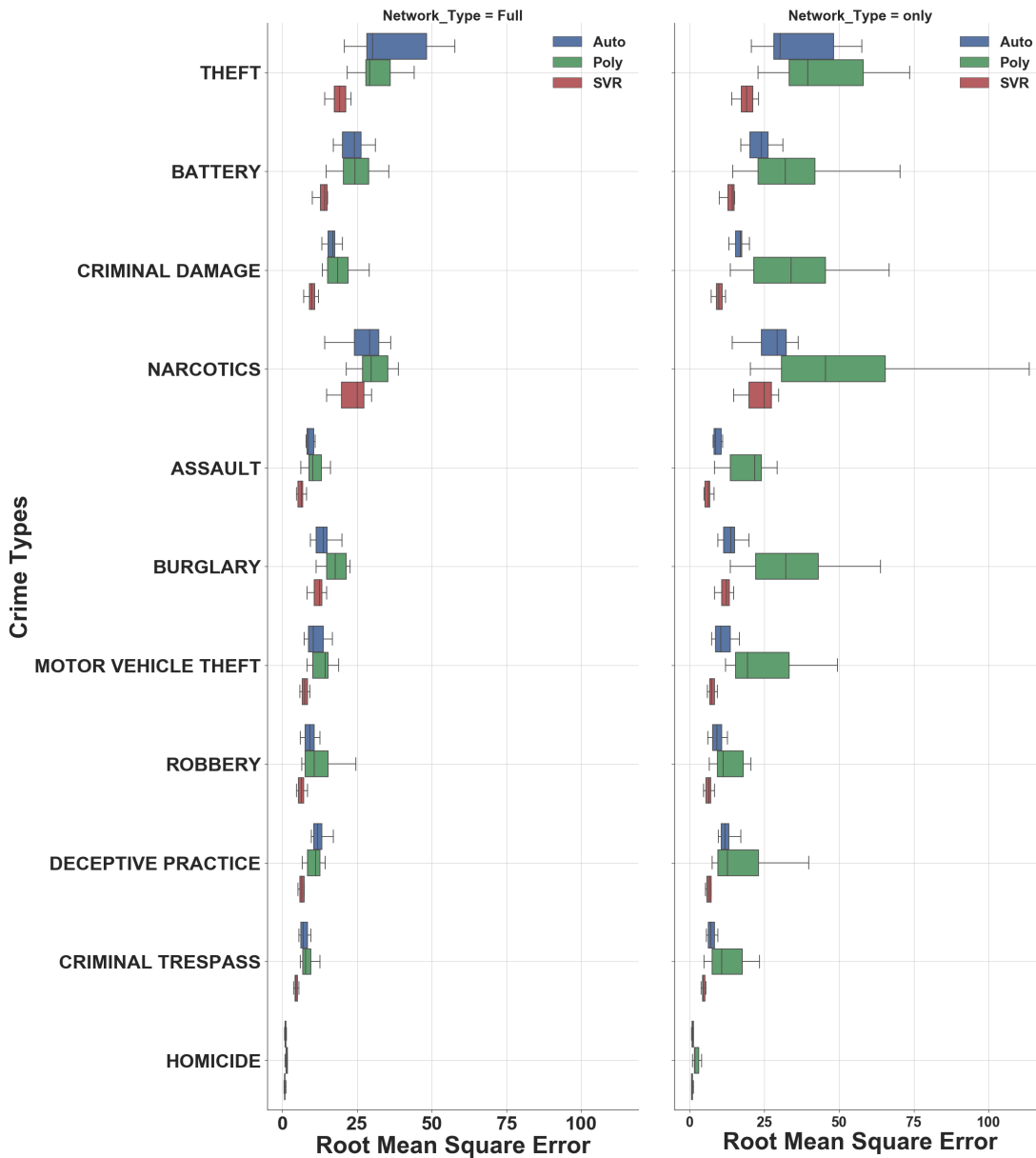
Fig. 4. Comparison of *full* and *only crime* network based predictions of all three methods. Boxes colored blue, green, and red represent auto-, polynomial, and support vector regression methods, respectively. Only top 11 (i.e., most frequent) crime types are presented in this comparison. Each box represents the results of RMSE prediction over all 12 months of the year in all communities in the City of Chicago. The left and right parts of the figure correspond to *full* and *only crime* network based predictions.

incomparably smaller than the training set size, and in our case, we observe that the number of support vectors is very small for all crime types and most communities. Because the number of support vectors is small and the prediction quality is high, their analysis can be *fundamentally interpreted* by the domain experts and decision makers which can be used is applicable and novel crime preventing policies.

V. FUTURE WORK AND CONCLUSION

We presented a prediction method for spatio-temporal information on crimes based on publicly available city data. We fuse the crime records with information about schools,

libraries, police stations, and 311 service calls using network analytic approach to identify observations that improve the quality of prediction. After all the features of data are identified, regression-based prediction can be applied. We experimented with polynomial, auto-regressive and support vector regression methods and found that the quality of support vector regression significantly outperforms other methods. During our experiments, we have analyzed the Chicago crime events of all available 34 types and predicted them for all 12 months of 2015 for all 77 communities that divide the City of Chicago.

Although the data sources we proposed to work with are updated on-line (or at least there is always an option for
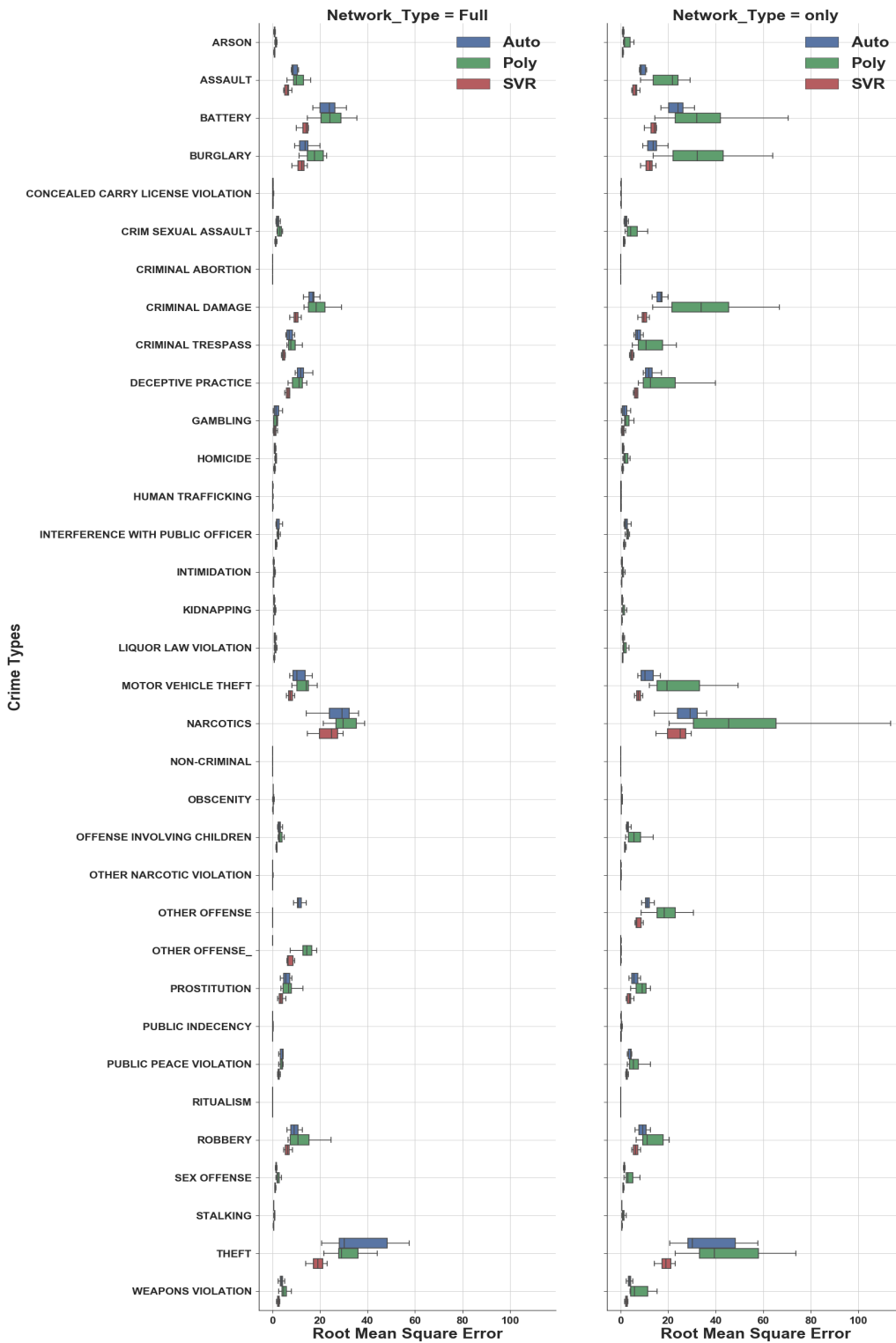
Fig. 5. Comparison of *full* and *only crime* network based predictions of all three methods. Boxes colored blue, green, and red represent auto-, polynomial, and support vector regression methods, respectively. All crime types are presented in this comparison. Each box represents the results of RMSE prediction over all 12 months of the year in all communities in the City of Chicago. The left and right parts of the figure correspond to *full* and *only crime* network based predictions.

the federal authorities to receive this information on-line), predicting crimes may benefit from features added from social networks as it has been shown in [1], [28], [14]. One of the promising research direction is fusing social network data that can provide socio-behavior "signals" for crime prediction. While the hypothesis that publicly available data in social networks may include predictive variables for crime and security events [1], [3] is known, we argue that adding community-based information will improve the prediction quality.

REFERENCES

1 Somayyeh Aghababaei and Masoud Makrehchi. Mining social media content for crime prediction. In *Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on*, pages 526–531. IEEE, 2016.

2 Tahani Almanie, Rsha Mirza, and Elizabeth Lor. Crime prediction based on crime types and using spatial and temporal criminal hotspots. *CoRR*, abs/1508.02050, 2015.

3 Neela Avudaiappan, Alexander Herzog, Sneha Kadam, Yuheng Du, Jason Thatcher, and Ilya Safro. Detecting and summarizing emergent events in microblogs and social media streams by dynamic centralities. In *2017 IEEE International Conference on Big Data*, pages 1627–1634. IEEE, 2017.

4 Sara Bastomski, Noli Brazil, and Andrew V. Papachristos. Neighborhood co-offending networks, structural embeddedness, and violent crime in chicago. *Social Networks*, 51:23 – 39, 2017. Crime and Networks.

5 Nate Berg. US urban population is up... but what does' urban'really mean. *The Atlantic*, 2012.

6 Giulia Berlusconi, Francesco Calderoni, Nicola Parolini, Marco Verani, and Carlo Piccardi. Link prediction in criminal networks: A tool for criminal intelligence analysis. *PLOS ONE*, 11(4):1–21, 04 2016.

7 Carolyn Rebecca Block. Stac hot-spot areas: A statistical tool for law enforcement decisions. In *Crime analysis through computer mapping. Washington, DC: Police Executive Research Forum*, pages 15–32. Citeseer, 1995.

8 Donald E. Brown and Stephen Hagen. Data association methods with applications to law enforcement. *Decis. Support Syst.*, 34(4):369–378, March 2003.

9 Eugenio Cesario, Charles Catlett, and Domenico Talia. Forecasting crimes using autoregressive models. pages 795–802, 08 2016.

10 Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. acm transactions on intelligent systems and technology, 2: 27: 1–27: 27, 2011. *Software available at http://www. csie. ntu. edu. tw/cjlin/libsvm*, 2011.

11 City of Chicago. Data portal for City of Chicago . https://data.cityofchicago.org, 2018.

12 Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161, 1997.

13 F. Fouss, A. Pirotte, J. Renders, and M. Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):355–369, March 2007.

14 Matthew S Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014.

15 M Todd Henderson, Justin Wolfers, and Eric Zitzewitz. Predicting crime. *Ariz. L. Rev.*, 52:15, 2010.

16 Malcolm K. Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects, 09 1991.

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

18 Oleg Roderick and Ilya Safro. Learning of Highly-Filtered Data Manifold Using Spectral Methods. In Christian Blum and Roberto Battiti, editors, *Learning and Intelligent Optimization*, pages 154–168, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

19 Ehsan Sadrfaridpuor, Talayeh Razzaghi, and Ilya Safro. Engineering fast multilevel support vector machines. *submitted, preprint arXiv:1707.07657*, 2018.

20 Emile ServanSchreiber, Justin Wolfers, David M. Pennock, and Brian Galebach. Prediction markets: Does money matter? *Electronic Markets*, 14(3):243–251, 2004.

21 Michael A Smith and Donald E Brown. Discrete choice analysis of spatial attack sites. *Information Systems and e-Business Management*, 5(3):255–274, 2007.

22 Social Sciences Research Centre. 77 community area defined. 2012.

23 Sunlight Foundation. US City Open Data Census . http://us-city.census.okfn.org/dataset/crime-stats, 2018.

24 Mohammad A Tayebi, Martin Ester, Uwe Glässer, and Patricia L Brantingham. Crimetracer: Activity space based crime location prediction. In *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 472–480. IEEE Press, 2014.

25 The Economist. Chicago accounted for 22% of a nationwide increase in murders last year. https://www.economist.com/democracy-in-america/2017/09/26/chicago-accounted-for-22-of-a-nationwide-increase-in-murders-last-year, 2017.

26 U.S. General Services Administration, Technology Transformation Service. Government Datasets published at data.gov. https://catalog.data.gov/dataset?tags=crime, 2018.

27 V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

28 Xiaofeng Wang, Donald E Brown, and Matthew S Gerber. Spatiotemporal modeling of criminal incidents using geographic, demographic, and twitter-derived information. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*, pages 36–41. IEEE, 2012.

29 Yang You, James Demmel, Kenneth Czechowski, Le Song, and Richard Vuduc. Ca-svm: Communication-avoiding support vector machines on distributed systems. In *Parallel and Distributed Processing Symposium (IPDPS), 2015 IEEE International*, pages 847–859. IEEE, 2015.