Spatially Embedded Co-offence Prediction Using Supervised Learning

Mohammad A. Tayebi
School of Computing Science
Simon Fraser University
British Columbia, Canada
tayebi@cs.sfu.ca

Martin Ester
School of Computing Science
Simon Fraser University
British Columbia, Canada
ester@cs.sfu.ca

Uwe Glässer
School of Computing Science
Simon Fraser University
British Columbia, Canada
glaesser@cs.sfu.ca

Patricia L. Brantingham
School of Criminology
Simon Fraser University
British Columbia, Canada
pbrantin@sfu.ca

ABSTRACT

Crime reduction and prevention strategies are essential to increase public safety and reduce the crime costs to society. Law enforcement agencies have long realized the importance of analyzing co-offending networks—networks of offenders who have committed crimes together—for this purpose. Although network structure can contribute significantly to co-offence prediction, research in this area is very limited. Here we address this important problem by proposing a framework for co-offence prediction using supervised learning. Considering the available information about offenders, we introduce social, geographic, geo-social and similarity feature sets which are used for classifying potential negative and positive pairs of offenders. Similar to other social networks, co-offending networks also suffer from a highly skewed distribution of positive and negative pairs. To address the class imbalance problem, we identify three types of criminal cooperation opportunities which help to reduce the class imbalance ratio significantly, while keeping half of the co-offences. The proposed framework is evaluated on a large crime dataset for the Province of British Columbia, Canada. Our experimental evaluation of four different feature sets show that the novel geo-social features are the best predictors. Overall, we experimentally show the high effectiveness of the proposed co-offence prediction framework. We believe that our framework will not only allow law enforcement agencies to improve their crime reduction and prevention strategies, but also offers new criminological insights into criminal link formation between offenders.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*

Keywords

Social Network; Link Prediction; Co-offence Prediction

1. INTRODUCTION

Crime is a purposive deviant behavior that is an integrated result of different social, economical and environmental factors [1], and generates substantial costs to society at individual, community, and national levels. In 2007, in the United States, crimes cost $179 billion in government expenditures [2]. In 2008, the total tangible and intangible costs of crime in Canada were an estimated $100 billion [3]. Given such whopping costs, crime reduction and prevention strategies have become a priority in many countries.

Existing crime forecasting methods answer questions pertaining to spatial and temporal patterns of future crime [4, 5]. These methods generally consider properties of crime incidents and behavior of single offenders to recognize criminal patterns. Over the past two decades, law enforcement and intelligence agencies have realized the importance of *co-offending network analysis* for crime investigations [6, 7]. A co-offending network links offenders who have committed crimes together [6]. Understanding co-offending is central to understanding the etiology of crime and the effects of intervention strategies [6] .

In spite of the importance of co-offending network analysis for public safety, the methodology for analyzing large-scale networks is rather premature. In this work we propose a framework for co-offence prediction using supervised learning. Even though supervised learning methods for link prediction have been studied widely [8, 9, 10, 11], to the best of our knowledge, there is no study on supervised learning for co-offence prediction. In [12], an unsupervised method for investigating top-k potential suspects is proposed.

Contrary to other social networks, concealment of activities and identity of actors is a common characteristic of co-offending networks. Still, the network topology is a primary source of information for co-offence prediction. Besides, there are two other major information sources: environmental activity and criminal activity. Offenders who are spatially close are socially close too as this increases the chance of meeting each other and forming new criminal collaborations [13]. Further, common criminal experience (with the same type of offences) also affects co-offending behavior [14].

The proposed framework builds on criminological theories [7, 15, 16, 17, 18] and, considering the available information on offenders, distinguishes three different criminal cooperation opportunities: socially-related, geographically-related and experience-related. We study the co-offence prediction problem in each of these prediction spaces separately, achieving two goals. First, the heavy class imbalance between positive (existing links) and nega-

tive samples (non-existing links) is the main challenge of the link prediction problem [10]. The restriction of the training and test data to the different prediction spaces reduces the class imbalance ratio significantly, while keeping about half of the positive samples (co-offences). Second, the prediction spaces enhance the understanding of co-offence patterns in different criminal cooperation opportunities.

We define the prediction features in four different categories, *social, geographic, geo-social* and *similarity*, and evaluate their prediction strength both individually and as a set. Social features indicate social closeness of offenders based on their position in a co-offending network. Geographic features show spatial proximity of offenders based on their residential locations and the location of offences they have committed. Geo-social features combine social and geographic characteristics of offenders. Finally, similarity features capture homophily-based characteristics of offenders. Evaluating features individually and also as a set shows that the geo-social features we define outperform other features.

Several studies show that supervised link prediction approaches outperform unsupervised methods [9, 10] that use only topological features [8]. In contrast to unsupervised methods, supervised learning methods can overcome the class imbalance problem [10]. Exploiting the geographic information provided by location-based social networks services, some recently proposed link prediction methods consider spatial characteristics of users [19, 20]. In [19], the authors use information about places visited by users, in addition to their social network features, to define prediction spaces which reduce the class imbalance ratio and improve the prediction performance.

Co-offending networks are spatially embedded similar to location-based social networks. However, the environmental effects on the formation of co-offence links and accordingly our approach in defining offenders' spatial closeness are different from those in location-based social networks [19, 20, 21, 22]. The proposed supervised learning framework aims at the public safety and security sector. Although there has been significant research on the reasons for involvement of single offenders in crime [6, 16, 18], there is no comprehensive study on the causes for offender collaboration. We argue that our study on co-offence prediction in different criminal cooperation opportunities opens up a new door to the understanding of co-offence patterns.

For the experimental evaluation, we use a large crime dataset comprising 4.4M records of police reported arrest data, made available for research purposes by the Royal Canadian Mounted Police (RCMP). The data was retrieved from the Police Information Retrieval System (PIRS), a large database keeping information for the regions of the Province of British Columbia which are policed by the RCMP. The co-offending network extracted from this dataset has more than 150k nodes. By applying different classifiers to the defined prediction spaces we can correctly predict up to 90% of all co-offences in the best case scenario.

The co-offence prediction framework proposed in this paper aims at advancing the state-of-the-art in crime data mining by making the following contributions: 1) Defining co-offence prediction spaces to reduce the class imbalance; 2) Introducing novel prediction features for co-offence prediction; and 3) Experimentally evaluating the proposed approach on large real-world crime data. Some of our main findings in this research include: a) Features evaluation have important implications. For instance, preferential attachment is a strong predictor compared to the features extracted from common friends in the network. This implies that the chance of criminal collaboration increases more with the opportunity to commit crimes than with trust or transitivity in the co-offending network. Crime

locations distance is better predictor than home location distance, meaning that being criminally active in the close districts causes new criminal collaboration. b) Geo-social features are better co-offence predictors than geographic and social features. This result implies that we need to focus more on combined patterns in environmental and social features to enhance crime reduction and prevention. c) The experimental results show that the proposed supervised method outperforms the unsupervised methods. d) Although there is variability in the performance of different classifiers, the probability of predicting a co-offence for experience-related offenders is higher than for socially and geographically-related co-offenders.

Section 2 briefly introduces basic concepts and characteristics of co-offending networks as well as the crime dataset used in our work. Section 3 then explains the proposed co-offence prediction approach, and Section 4 describes the prediction feature sets. Next, Section 5 presents our experimental evaluation and results. Criminological implications and related work are discussed in Sections 6 and 7. Section 8 concludes the paper.

2. CO-OFFENDING NETWORKS

This section describes the formal co-offending network model. A crime dataset consists of a collection of crime records, each of which refers to a reported crime incident. A crime record typically identifies the date, time, location and type of a crime. To the extent known, it also identifies offenders, victims, witnesses, bystanders and their addresses, in addition to other crime specifics and circumstantial evidence. To this end, a crime dataset refers to crime incidents over some period of time.

2.1 Basic Concepts and Definitions

A *co-offending network* $G(V, E)$ is a connected graph structure. Each node represents a known offender. Offenders u and v are connected, $u, v \in V$ and $\{u, v\} \in E$, if they are known to have committed one or more offences together, and are not connected otherwise. A co-offending network is derived from a crime dataset referring to reported crime incidents over a time period.

For $G(V, E)$ and $u \in V$, let Γ_u^n be the subset of offenders in V such that their shortest path distance from u is n. $P_u = (p_u^1, p_u^2, \ldots, p_u^K)$ denotes the frequency of crimes committed by u for each of k crime types in the dataset. The subset of offenders who have committed p crime types that coincide with crime types committed by u is referred to by Θ_u^p.

For u in $G(V, E)$, let $H_u = \{h_u^1, h_u^2, \ldots, h_u^k\}$ denote the known home locations of u, and $C_u = \{c_u^1, c_u^2, \ldots, c_u^l\}$ denote the known crime locations for all offences committed by u. Finally, let $\Delta_u = \{\delta_u^1, \delta_u^2, \ldots, \delta_u^k\}$ denote the time periods (intervals) $\delta_u^i = [t_u^{i_1}, t_u^{i_2}]$ that u lived in each of the k home locations given by H_u. $D(h_u^i, h_v^j)$ is the geodesic distance between h_u^i and h_v^j.

2.2 Crime Data

Crime data mining has enormous potential as analytic tool for law enforcement agencies, criminal intelligence agencies, and beyond to facilitate crime investigations by increasing efficiency and reducing mistakes. On the other hand, access to and sharing of crime data is subject to many restrictions and can even be national security concern because of the highly sensitive nature and related personal information. Police arrest data and court-based data [23] are two important types of official crime data. Police arrest data identifies a person arrested or booked for an offense. Court-based data includes information generated by the courts. In this study we use a large real-world police arrest dataset.
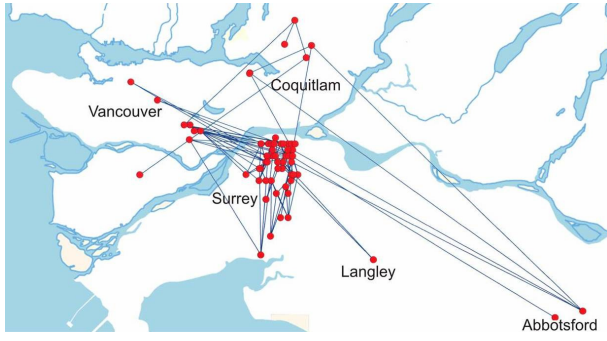
Figure 1: Spatial distance of co-offenders in the BC co-offending network.

As a result of a research memorandum of understanding between ICURS[1] and "E" Division of Royal Canadian Mounted Police (RCMP) and the Ministry of Public Safety and the Solicitor General, five years of real-world crime data was made available for research purposes. This data was retrieved from the RCMP's Police Information Retrieval System (PIRS), a large database system keeping information for the regions of the Province of British Columbia which are policed by the RCMP.

For a time period of five years (2001-2006), the BC police arrest dataset comprises $\approx$ 4.4 million crime records, one for each reported crime incident, and all persons associated with a crime, such as offenders (from complainant to charged), victims, witnesses and bystanders. In total, there are 39 different subject (person) groups. In our experiments, we consider the subjects in four main categories: *charged, chargeable, charge recommended or suspect*. Being in one of these categories means that the police were serious enough about a subjects involvement in a crime as to warrant calling them 'offenders'. From this dataset we extract , using the method described in [24], called BC co-offending network. This network comprises $\approx$ 150,000 nodes and $\approx$ 600,000 edges. The average node degree is four, and about 50% of all the nodes have degree one, meaning these offenders have committed co-offences with only one offender in their criminal life. The largest component links $\approx$ 18% of all the nodes, which is fairly big for this type of social network.

2.3 Offenders' Activity Space

The *Activity Space* of an offender has two main components: *Nodes* and *Paths*. Activity Nodes refer to locations a person frequently visits in the course of daily routine activities. Activity Path represents a common route for a routine trip to a frequently visited location. *Crime Pattern Theory* [16] contends that, rather than venture into new areas to commit crimes, offenders act on criminological opportunities arising in areas they are familiar with, more specifically, areas that are part of their Activity Space.

Inverting research on crime pattern theory [16], geographic profiling [17] seeks to either estimate the location of the residence of an offender or to extrapolate locations where an offender is likely to commit future crime. Centrography is one of the most common search approaches for criminal investigation [17]. In [17], Rossmo shows that some offenders live close to the centroid of their crime locations. The "Circle Hypothesis" by Canter et. al [25] defines a circular area around the location of the first offence in a series of crimes to indicate the general area of an offender's home location.

[1]The Institute for Canadian Urban Research Studies (ICURS) is a university research centre at Simon Fraser University.

In this research our goal is consider offenders' environmental activities for co-offence prediction, while it does not fit with the bases of works in geographic profiling. Considering the definitions in the literature and the general limitations in the crime data we use a simple and safe definition in this regard. For a given offender u with home locations H_u, $A_u^R = \{a_u^1, a_u^2, \ldots, a_u^k\}$ states the activity space of u, where a_u^i is defined as a circle of radius R with h_u^i at its center. We define the *common activity space* of two offenders as the area in which both are active simultaneously, if their activity spaces overlap at any time. For offenders u and v, $A_{u,v}^R = \{a_{u,v}^{1,2}, a_{u,v}^{1,3}, \ldots, a_{u,v}^{p,k}\}$ is their common activity space, that is

$$A_{u,v}^R = \{a_{u,v}^{i,j} \mid a_u^i \cap a_v^j \neq \varnothing \wedge \delta_u^i \cap \delta_v^j \neq \varnothing\}.$$

Intuitively, $a_{u,v}^{i,j}$ represents the intersection of the activity space of u and v during a time they were both active at a_u^i and a_v^j. The offenders who live in $a_{u,v}^{i,j}$ and the crimes in this area are referred to by the set $\chi_{u,v}^{i,j}$ and the set $\phi_{u,v}^{i,j}$, respectively.

2.4 Geographic and Network Proximity

Neighborhood greatly influences the formation of communities and social networks. Hence, social problems are often studied in connection with neighborhoods. Many researches suggest that crime is as well strongly linked to geographical characteristics. Criminology researches use spatial analysis of crime to understand the distribution of crime incidents and why crime occurs in some places but not others [16, 17]. Figure 1 visualizes home locations of offenders for one of the connected components of the BC co-offending network. The fact that clusters can be seen in this visualization implies that many pairs of co-offenders live in the same city.

The criminology literature intensively studies modeling of offender travel to crime locations. For instance, it has been concluded that most offenders travel short distances to commit crimes [26]. But there is no definite reasoning to explain how the activity space of two offenders influences their potential collaborations. Findings in [27] suggest that co-offenders use their resources to increase the scope of their awareness space.

Generally, it is important to take into account the spatial dimension of co-offending relationships. We have studied this aspect for all observed criminal collaborations in the BC crime dataset. About 39% of the co-offenders live less than 2 km apart, and about 63% of them live less than 10 km apart. Figure 2a shows the probability distribution of home location distance for co-offenders. The probability distribution of home and crime location distance of offenders is shown in Figure 2b. 46% of the crime incidents happen in less than 2 km distance from the home location, and 70% of the crimes happen within a distance of less than 10 km. We conclude that a large percentage of the crime incidents are located close to the offenders' residence. While the maximum values for co-offenders home location distance and offenders' home and crime location distance are about 1000 km, the medians of these distances are only 5 km and 3 km, respectively. This confirms that co-offenders tend to be geographically confined.

In [13], it is concluded that socially clsoe offeners are spatially close too, meaning that offenders who are close in the co-offending network have more overlap in their activity spaces too. Criminological theories and the discussed experimental results motivated us focus more on geographic and geo-social features, besides general social features extracted from network topology for co-offence prediction.
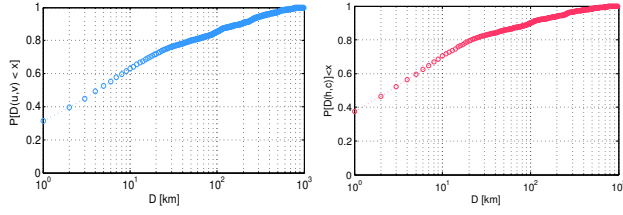
Figure 2: (a) Home location distance of co-offenders; (b) Home and crime locations distance of offenders.

3. CO-OFFENCE PREDICTION

In this section we define the *co-offence prediction problem* and propose a novel analytic framework for solving this problem.

3.1 Problem Definition

A co-offending network $G(V, E)$ refers to all co-offences within time period $[t_0, t_n]$ associated with the underlying crime dataset. For any time $t \in [t_0, t_n]$, one can derive $G_t(V_t, E_t)$ as a substructure of G, which is a co-offending network, by restricting to all crime incidents up to time t. For $G_t(V_t, E_t)$ we now define a *potential co-offence* at time $t+1$ as any pair of offenders (u, v) such that $(u, v) \notin E_t$, meaning that offenders u and v have not committed any crime together prior to $t+1$. A potential co-offence (u, v) is in the *positive class*, if $(u, v) \in E_{t+1}$, and it is in the *negative class*, if $(u, v) \notin E_{t+1}$. The co-offence prediction task is to predict for each potential co-offence in G_t if it belongs to the positive class or the negative class.

Co-offence prediction can be viewed as a link prediction problem for co-offending networks. This aspect is modeled using a binary classification problem that adopts a set of prediction features as described in Section 4. The major challenge in any link prediction problem is the heavily skewed distribution of negative and positive classes. Contrary to unsupervised methods, supervised methods learn class distributions to increase classification performance [10].

3.2 Criminal Cooperation Opportunities

Because of all obvious risks about committing crimes, offenders do not select their collaborators accidentally. Considering offenders' choices and selection opportunities, any co-offence can have social, environmental or experience related causes. To quantitatively model how offenders form a criminal cooperation, for each offender u, we define three opportunity spaces for criminal cooperation defined as follows:

Socially-related. Social interactions influence the behavior of individuals. This observation has been studied widely in areas such as educational choices and labor market outcomes. Arguably, this applies to criminal behavior as well, where social networks compensate for lack of formal institutions in gaining knowledge and criminal skills. Criminal behavior associated with illegal activities is influenced by informal networks and peer interactions. Sutherland, in the theory of *differential association* [18], explains individual criminality with a social-psychological process of learning crime through interaction with social groups. According to [18], criminal behavior is the result of learning an excess of definitions favorable to crime. Offenders use social interactions for sharing information, recruiting young criminals [28] and transferring criminal skills. Although co-offending networks are generally considered short-lived networks, they constitute an important source of criminal cooperation opportunities. In our definition, socially-related cooperation opportunities for an offender u, denoted by S_u,

arise from offenders v such that their distance in the co-offending network is not greater than N, excluding the direct neighbors of u:

$$S_u = \{(u, v) : v \in (\bigcup_{i \leq N} \Gamma_u^i) \setminus \Gamma_u^1\}$$

Geographically-related. Activity space and social environment of offenders greatly influence their criminal decisions and are key factors in forming collaboration opportunities [29, 16]. Activity space plays a crucial role in where crimes are committed and individuals are recruited for the purpose of co-offending. Not only do offenders converge with crime victims, but they also converge with each other in order to find suitable co-offenders. Felson's "offender convergence settings" describe certain locations in which potential offenders meet each other to initiate and maintain co-offending and crime groups [29]. Based on accepted criminological theories we define geographically-related space as an important potential space for co-offending link formation.

Offender u is considered geographically related to any offender v, if they have a common (overlapping) activity space, but they are not connected to each other directly in the co-offending network.

$$G_u = \{(u, v) : v \in (\bigcup_{a_u^i \in A_u^r} \chi_u^i) \setminus \Gamma_u^1\}$$

Experience-related. Network studies on offenders have generally observed evidence of homophily like criminal experience [14]. This is so consistent that group homogeneity is considered one of the key characteristics of co-offending [14]. With similar criminal experience the chance increases that offenders forme new criminal collaborations. This can happen via implicit networks or due to demand for specific criminal expertise. All pairs of offenders who have similar criminal experience but are not directly connected to each other in the co-offending network are categorized into this class:

$$E_u = \{(u, v) : v \in (\bigcup_{i \geq P} \Theta_u^i \setminus \Gamma_u^1\}$$

We use the above criminal cooperation spaces to define the spaces which cover the prediction candidates. Prediction space division has two advantages: First, it helps to reduce the class imbalance ratio. Second, it leads to a clearer understanding of the effects of each of these categories on new co-offence formation. For $G_t(V_t, E_t)$ we generate three spaces for the co-offence prediction task:

SR: This space includes all links emerging between pairs of offenders that are socially-related: $SR^t = \{\forall u \in V_t : \bigcup_u S_u\}$. In our experiments for constructing the SR space, we consider $N = 2$. In other words, a pair of offenders who are exactly 2-hops apart are added to this space.

GR: that contains all links between pairs of offenders who are geographically-related but not socially-related: $GR^t = \{\forall u \in V_t : (\bigcup_u G_u) \setminus SR^t\}$. For constructing the activity space of offenders which defines if two offenders are geographically-related we use $R = 2km$.

ER: that includes all pair of offenders that are experience-related but not socially-related: $ER^t = \{\forall u \in V_t : \bigcup_u E_u \setminus SR^t\}$. For creating the ER space, we apply $P = 2$, meaning that (u, v) are considered experience-related offenders if both committed at least two crimes of the same type.

3.3 Reducing Class Imbalance Ratio

For a network $G(V, E)$ the number of links E is often $O(|V|)$, but the number of no-existing links is often $O(|V|^2)$. Accordingly,
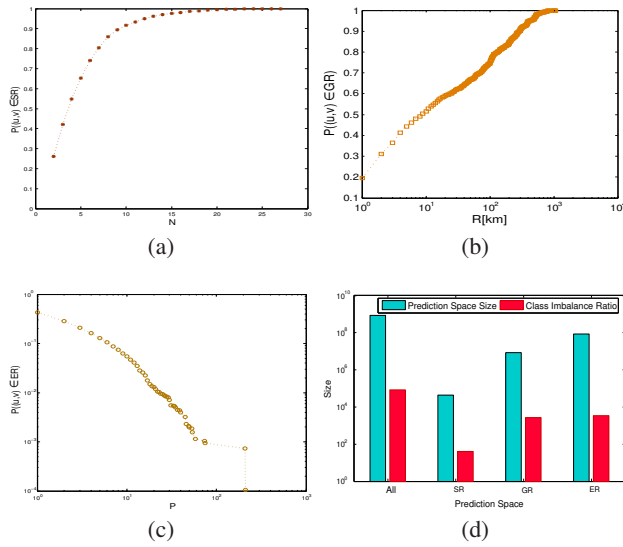
(a)   (b)

(c)   (d)

Figure 3: (a) Cumulative probability distribution of a positive sample in SR space in respect to N; (b) Cumulative probability distribution of a positive sample in GR space in respect to R; (c) Cumulative probability distribution of a positive sample in ER space in respect to P; (d) Prediction spaces size and their class imbalance ratio.

the prior probability of link formation is very small. Since the goal of supervised learning is achieving high precision, class imbalance in the training dataset leads to overfitting to negative samples which reduces the recall of positive samples. Note that in the co-offence prediction task high recall of the positive class (co-offences) is critical, so that overcoming the class imbalance is essential.

In the experimental setting defined in Section 5.1, the number of negative samples is 850M while the number of positive samples is only 11k. Therefore, the *class imbalance ratio*, the ratio of negative samples to positive samples, is about 77K. Using the prediction spaces to reduce the training and test dataset can effectively reduce the class imbalance. In [19] an efficient prediction space division schema for location-based social network is proposed, but it does not apply to co-offence prediction where environment activity of offenders are different than location-based social networks users. Furthermore we consider the homophily effects, in terms of criminal activity similarity, on forming new link which is not considered in the approach proposed in [19]. Although reducing class imbalance is an important objective, the other main concern is keeping as many positive samples as possible.

Restricting the dataset to samples with short graph distance is the most effective solution for the imbalance problem. In the SR space increasing N affects the number of negative samples set superlinearly, because we expect to see more co-offences between offenders close in the network. Figure 3a demonstrates the probability of a positive sample in the SR space for different values of N of the network distance. The probability does not grow linearly as N increases, meaning that similar to other type of social networks [10, 19] majority of positive samples are at a closer network distance. For the maximum value $N = 27$ there are 204M negative samples. But changing this to $N = 2$ decreases the number of negative samples to 47K, while keeping 25% of the positive samples.

Applying different constraints related to the other prediction spaces also helps to reduce the class imbalance ratio. Figure 3c shows the

probability of a positive sample in the GR space for different values R of the radius of the activity space. With $R = 2$ km, $R = 10$ km and $R = 100$ km the GR space covers cumulative 31%, 51% and 75% of the positive samples. Figure 3b shows the probability of a positive sample in the ER space for different values of P. With $P = 2$ of the same crime type, $P = 5$ and $P = 10$ the ER space includes 29%, 13% and 5% of the positive samples. Applying $R = 2$ km and $P = 2$ reduces the $850M$ negative samples in the dataset to 8M and 82M respectively for GR and ER spaces. With these constraints we can keep 31% and 29% of the positive samples in the GR and ER spaces.

As demonstrated in Figure 3d, the original imbalance ratio of 77K for the whole prediction space reduces to 40, 2700 and 3400 respectively for SR, GR and ER spaces. While in each of three spaces roughly we can keep approximately the same percentage of positive samples (26%, 31% and 29% for the SR, GR and ER spaces), we see that the class imbalance reduction is better in the SR space. In total we are able to maintain half of positive samples for training, meaning that the likelihood of predicting a co-offence successfully increases significantly.

4. PREDICTION FEATURES

Prediction features are divided into four categories: social, geographic, geo-social and similarity features. This section describes how these features are extracted. Features are defined formally in Table 1.

4.1 Social Features

The social features set includes all features that are derived using only the topology of the co-offending network and the position of offenders in the network. Preferential is defined as the product of the node degrees of two offenders, meaning that the more connected an offender is, the more likely he forms new criminal links. Common, denotes the number of neighbors two offenders have in common. Overlap divides the number of common neighbors by the overall number of neighbors of two offenders as a normalized indication of common neighbors effect. Adamic [30] assigns higher weight to the common neighbors with smaller node degree.

4.2 Geographic Features

With increasing overlap of the activity space of offenders the chance of forming new criminal collaboration increases. Home locations distance, HDN, is the average distance between the current and past home locations of two offenders. HDT weights home location distance by the time two offenders lived in the corresponding locations. Committing crimes in the same neighborhood may cause a criminal tie between offenders. CDN is the average crime location distance of two offenders.

4.3 Geo-social Features

Geo-social features combine the social and geographical characteristics of offenders. Common activity space of offenders directly affects their behaviors. Being active in an area with many offenders living there can increase the chance that offenders meet each other directly or indirectly, and engage in co-offending. OCT denotes the number of offenders living in a common activity space of two given offenders. In the definition of common activity space time is implicitly restricted to $\delta_u^i \cap \delta_v^j$. OCTT boosts OCT if two offenders are active for a longer period of time. Without any restriction, OCN denoted the number of offenders who live in a common activity space anytime in the time interval $[t_0, t]$.

Characteristics related to the quantity of crimes in the common activity space provide another set of geo-social features. A crime

hotspot located within the common activity space of offenders may increase the chance of collaboration among these offenders. CCT states the number of crimes in the common activity space of two offenders, where, similar to OCT, the time is restricted to the period over which both offenders are active simultaneously. CCTT weights the time interval of two offenders being active. And CCN refers to the total number of crime incidents in the common activity space.

4.4 Similarity Features

The homophily principle states that individuals tend to associate themselves with others in a social network who are similar [31]. This is also a very well-studied topic for co-offending [14]. Age difference, ethnic group similarity and gender similarity, represented by Age, $Ethnic$ and $Gender$, are three features in this category. Finally, $CrimSim$ also expresses similarity of criminal experience based on committed crime types.

5. EXPERIMENTS AND RESULTS

In this section we present the experimental evaluation, starting with a description of the experimental design.

5.1 Experimental Design

The spacial nature of co-offending networks makes co-offence prediction a difficult task. Offenders not only do hide their illegal activities and contacts, but their accomplice relationships are usually short-lived [6].

For our experiments, we divide the dataset into two disjoint sets of incidents, one for the first 50 months and the second one for last 10 months. Excluding noisy data, such as traffic related offenses, the number of incidents for the aforementioned time periods are 1.8M and 800K respectively. Considering only offences with more than one offender reduces these numbers to 67K and 17K. We extract the co-offending networks corresponding to each of these time periods. Next, criminal cooperation opportunities are identified and each pair of potential co-offences are assigned to one of three prediction spaces: SR, IR or ER. Eventually, the prediction features of each pairs of potential co-offenders are extracted.

We use the open-source machine-learning library Weka [32] for different classifier building. The performance of classifiers and single feature prediction are assessed based on the Receiver Operating Characteristics (ROC), which shows the trade-off between the true positive rate over the false positive rate [33]. The area under the ROC curve, called AUC, is an appropriate measure for comparing two ROC curves. AUC measures the probability that a classifier ranks a randomly chosen positive sample higher than a negative sample.

Various definitions of offender activity space are known in the literature. We use a simple definition: a circle of radius R with the home location of an offender in the center point (based on the available information on offenders in the dataset). Common activity space of two offenders is defined as the intersection of their activity space and is used for defining geo-social features. Given that there is no consensus in the criminology literature as to the appropriate value of R, we tried different values of R, ranging from 1 km to 10 km. Interestingly, for all features, a value of $R = 2$ km maximizes the prediction performance, which we therefore chose in our experiments.

5.2 Single Feature Significance

We compare the predictive power of individual features described in Section 4 in each of the prediction spaces. This is important to diagnose which features in each prediction space play a more im-

Social Features	
Preferential	$\lvert \Gamma_u^1 \rvert \times \lvert \Gamma_v^1 \rvert$
Common	$\lvert \Gamma_u^1 \rvert \cap \lvert \Gamma_v^1 \rvert$
Overlap	$\frac{\lvert \Gamma_u^1 \rvert \cap \Gamma_v^1}{\lvert \Gamma_u 1 \rvert \cup \lvert \Gamma_v^1 \rvert}$
Adamic	$\sum\limits_{z \in \Gamma_u^1 \cap \Gamma_v^1} \frac{1}{log(\Gamma_z^1)}$
Geographic Features	
HDN	$\frac{\sum\limits_{i=1}^{i=m} \sum\limits_{j=1}^{j=n} e^{\frac{-D(h_u^i, h_v^j)}{\lambda}}}{\lvert H_u \rvert \times \lvert H_v \rvert}$
HDT	$\frac{\sum\limits_{i=1}^{i=m} \sum\limits_{j=1}^{j=n} e^{\frac{-D(h_u^i, h_v^j)}{\lambda}} \times \lvert(\delta_u^i \cap \delta_v^j)\rvert}{\lvert H_u \rvert \times \lvert H_v \rvert}$
CDN	$\frac{\sum\limits_{i=1}^{i=m} \sum_{j=1}^{j=n} e^{\frac{-D(c_u^i, c_v^j)}{\lambda}}}{\lvert C(u) \rvert \times \lvert C(v) \rvert}$
Geo-social Features	
OCT	$\sum\limits_{i=1}^{i=p} \sum\limits_{j=1}^{i=k} \lvert \chi_{u,v}^{i,j} \rvert$
OCTT	$\sum\limits_{i=1}^{i=p} \sum\limits_{j=1}^{i=k} \lvert \chi_{u,v}^{i,j} \rvert \times \lvert \delta_u^i \cap \delta_v^j \rvert$
OCN	$\sum\limits_{i=1}^{i=p} \sum\limits_{j=1}^{i=k} \lvert \chi_{u,v}^{i,j} \rvert : [t_0, t]$
CCT	$\sum\limits_{i=1}^{i=p} \sum\limits_{j=1}^{i=k} \lvert \phi_{u,v}^{i,j} \rvert$
CCTT	$\sum\limits_{i=1}^{i=p} \sum\limits_{j=1}^{i=k} \lvert \phi_{u,v}^{i,j} \rvert \times \lvert \delta_u^i \cap \delta_v^j \rvert$
CCN	$\sum\limits_{i=1}^{i=p} \sum\limits_{j=1}^{i=k} \lvert \phi_{u,v}^{i,j} \rvert : [t_0, t]$
Similarity Features	
Age	$\lvert Age(u) - Age(v) \rvert$
Gender	$\begin{cases} 1, & \text{if Gender}(u) = \text{Gender}(v) \\ 0, & \text{if Gender}(u) \neq \text{Gender}(v) \end{cases}$
Ethnic	$\begin{cases} 1, & \text{if Ethnic}(u) = \text{Ethnic}(v) \\ 0, & \text{if Ethnic}(u) \neq \text{Ethnic}(v) \end{cases}$
CrimSim	$\frac{\sum\limits_{i=1}^{K} P_u^i P_v^i}{\sqrt{\sum_i^K (P_u^i)^2} \times \sqrt{\sum_{i=1}^K (P_v^{i=1})^2}}$

Table 1: Prediction features.

portant role in co-offence prediction. For this purpose, we compute the feature values for all negative and all positive potential co-offenders. Then, using a range of decision thresholds and computing the false/true positives ratios, we generate the ROC curves for each single feature as presented in Figure 4. Note that the social features can be extracted only for SR space.

As shown in Figure 4a, in the SR space, the $Preferential$ feature is the best predictor with AUC value 0.82 and is superior to other social features. The performance of $Adamic$ and $Jaccard$ are worse than a random predictor. This shows that being a frequent offender and having broader criminal relationships, rather than common relationships, increases the chance of engaging in new criminal cooperations. Following $Preferential$ are geo-social and geographic features CDN, OCN, CCN, and HDN, with AUC values between 0.74 and 0.79. The performance of similarity features is lower than the mentioned set.

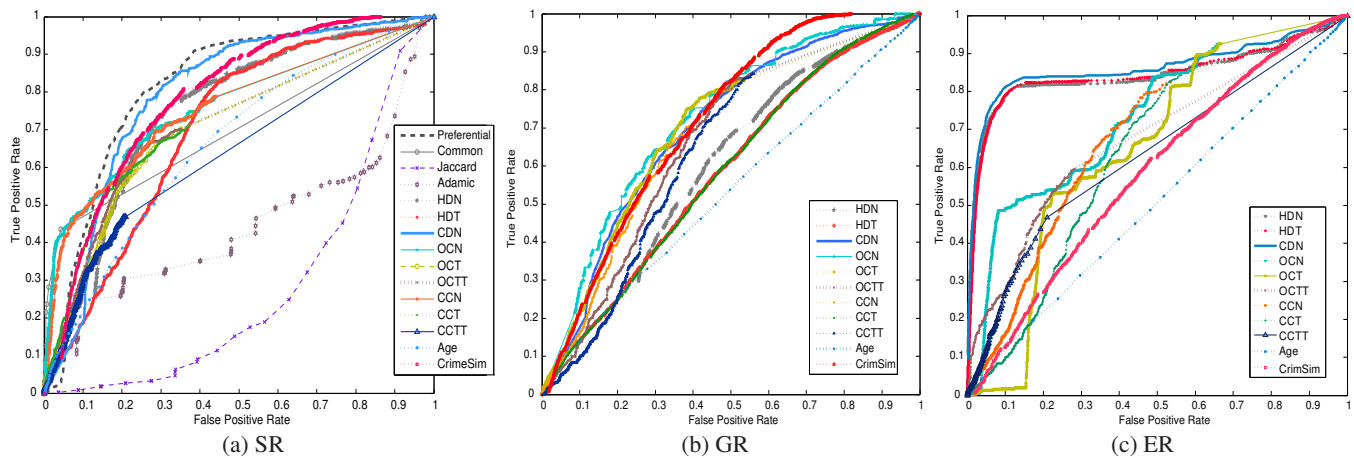| (a) SR | (b) GR | (c) ER |

Figure 4: Single feature significance in three different predictions spaces.

In the GR space, performance of most of the features is weaker than their performance in the SR space. This is not unexpected for geographic or geo-social features, where in the negative potential class we gather all pairs that are geographically close and the predictor cannot be as successful as in the SR space. Nevertheless, the AUC values of these features fall within the range 0.60 to 0.70. Compared to the SR space, in the ER space geographic and geo-social features work better. Their AUC values are between 0.72 and 0.86.

In the similarity features generally Gender works as good as a random predictor in all three spaces. Age and Ethnic have similar performance in the ER and GR spaces. But in SR, AUC values of these two features are 0.66 and 0.60. This shows that in the SR space one can see stronger signs of the homophily principle in forming new relationships. CrimSim has the best performance in the GR space with AUC value of 0.71.

Counterintuitively, time-based features perform weaker than the original version of the corresponding feature. For instance, in the SR space, the performance of HDT compared to HDN, OCT and OCTT compared to OCN, and, finally, CCT and CCTT compared to CCN, all drop down to five percent. The same trend applies to the other prediction spaces. Parameter regularization of time-based features for promoting their prediction strength is subject to future research. An interesting pattern in all spaces is that CDN outperforms HDN, which shows that being criminally active in areas that are in close proximity is a stronger indicator for forming new criminal collaboration than living in close proximity to each other.

Each single feature can be considered an unsupervised link predictor. Unsupervised link prediction methods assign scores to potential links based on node attributes or network structural proximity measures [8]. Then, the link prediction problem is defined as a binary classification problem by choosing some probability threshold, and predicting that potential links with a probability above the threshold will be formed, whereas those below the threshold will not be formed. We use the best AUC value of the single features to compare with the supervised predictors in the next section.

As discussed in [8], the performance of unsupervised link predictors varies in different social networks, meaning that unsupervised link predictors are domain-specific. For instance, among the structural network features, preferential attachment works best for co-offending networks in our study, while this single feature predictor has the worst performance for the network of cellular phone users in [10].

5.3 Prediction Evaluation

The previous section analyzes the performance of individual features. Now, how good works a supervised method using a group of extracted features? Naturally, one would expect to see performance improvements. For this purpose, we use four different classification methods: Naïve Bayes, J48 (equivalent to C4.5), random forests (10 trees, each constructed while considering 4 random features), and bagging (10 bags).

Similar to the work in [9, 19], we run 10-fold cross validation over 10 different randomly sampled training sets for each of the three prediction spaces, SR, GR and ER. We consider the average values of AUC, and precision and recall over positive samples. The prediction results are listed in Table 2. One can see variability in the different classifier performances, and also in the results for the different prediction spaces. All classifiers for all spaces outperform single features. Generally, prediction works best in the ER space.

Two ensemble methods, bagging and random forest classifiers, work better than the other classifiers and Naïve Bayes is the weakest one in all spaces. In ensemble learning for decreasing the variance error the results of a number of classifiers are combined to make a prediction. An ensemble method improves the prediction performance if the overlap of misclassification of the single classifiers is small. On average we see 5.1% AUC improvement in random forest classifiers comparing to single decision tree classifier which shows the variance reduction influence.

Since all features are available in the SR space and the bagging classifier works better than other methods, we continue the following experiments with the SR space and bagging classifier. To study prediction strength of each of the feature sets, social, geographic, geo-social and similarity, each time we keep one of the feature sets and evaluate it using the bagging classifier with the same setting. As shown in Table 3, the geo-social features outperform the other three sets, and the geographic feature set has the worst performance. Comparing the prediction performance using all features to different subsets shows integrating all features from social, environmental and personal can contribute to co-offence prediction.

Most existing unsupervised link prediction methods assign scores to the potential links and rank them. The top-k links are categorized as new links, and the remaining as missing links. The evalua-

Algorithm	Space	Precision	Recall	AUC
	SR	0.888	0.807	0.907
J48	GR	0.869	0.834	0.901
	ER	0.935	0.81	0.898
	SR	0.836	0.514	0.825
Naïve Bayes	GR	0.825	0.441	0.817
	ER	0.945	0.706	0.895
	SR	0.898	0.843	0.944
Random Forest	GR	0.864	0.883	0.944
	ER	0.941	0.944	0.982
	SR	0.908	0.84	0.951
Bagging	GR	0.863	0.854	0.952
	ER	0.946	0.942	0.986

Table 2: Recall, precision and AUC for different classification algorithms on the three different prediction spaces, SR, GR and IR, using 10-fold cross validation over 10 different random training sets.

Features Set	Precision	Recall	AUC
Social	0.903	0.792	0.919
Geographic	0.721	0.786	0.811
Geo-social	0.863	0.853	0.942
Similarity	0.849	0.851	0.928
All Features	0.908	0.84	0.951

Table 3: Prediction strength of different feature sets for SR space using the bagging classifier with 10-fold cross validation over 10 different random training sets.

tion of unsupervised methods is same as single features evaluation in Section 5.2. Preferential, Jaccard and Adamic are common unsupervised link prediction methods as addressed in [8]. While Preferential has the highest the AUC value among individual features we see more then 13% improvement in the bagging classifier result, which means supervised classifiers outperform the unsupervised predictors significantly.

For a better insight into class imbalance issues, we vary the size of negative samples for the SR space using same set of positive samples, where the positive samples proportion ranges from 2% to 100%. As illustrated in Figure 5a, the AUC value is relatively stable. Unsurprisingly, with negative sample size and accordingly imbalanced ratio growth, precision increases and recall decreases. As expected, this is because of negative class overfitting, meaning that the trained classifier is more likely to predict non-existing links.

6. CRIMINOLOGICAL IMPLICATIONS

Crime forecasting models are divided into short-term and long-term categories in terms of predicted time periods. Short-term models are beneficial for tactical decision making, whereas long-term models are helpful for planning and policy development. Co-offence prediction is an important aspect for short-term models. Studying new link formation patterns can help designing effective long-term crime reduction and prevention strategies.

Based on the discussion of offenders' activity space issue in Section 2.3 for all six geo-social features we tried different values of radius R from 1 km to 10 km. As depicted in Figure 5b for all of the features and $R > 2$ km the prediction performance decreases by increasing R. This finding is consistent with a well-know criminology fact [17] which states crimes are likely to occur closer to an offender's anchor points and follow a distance-decay function.

With the settings defined in Section 5.1, $E_{t+1} \cong 120K$, while the portion of positive samples is about 9%. And in the reality the prediction is possible only in this portion where we had information about both of co-offender previously. The propose framework was able to predict 45% of all positive samples.

Figure 6a shows the cumulative degree distribution $P(k)$ in the co-offending network G_{50} for three sets of offenders: V_{50}, $\{u \mid \exists v \in V_t : (u,v) \in SR \cap E_{t+1}\}$ and $\{u \mid \exists (u,v) \in SR \wedge (u,v) \notin E_{t+1}\}$. One can see that for the same node degree k, compared to the two other sets, the set of offenders who contribute to forming co-

offences generally has a greater value of $P(k)$. Section 5.2 shows that Preferential for SR space is a strong predictor. This implies the relevance of co-offending networks for making new criminal cooperation.

Assume that $\tau_{u,v}$ refers to the first time that a positive co-offence (u,v) occurs in time interval $[51, 60]$, and τ is the time of the most recent offence of u or v in $[1, 50]$. Figure 6b plots the distribution of $\tau_{u,v} - \tau$ for all positive potential co-offences. We see a power law distribution, meaning that there are many offences with a small value of $\tau_{u,v} - \tau$ and few with a larger value of $\tau_{u,v} - \tau$. This results shows that offenders who were recently active have a higher potential to form new criminal cooperation.

We see some variance in the prediction strength of different prediction features, with weak and strong predictors in each of the spaces. For all spaces we see high prediction results, which shows that there are strong patterns in co-offending. This supports our idea of dividing criminal cooperation opportunities, this way, finding stronger co-offence patterns that improve predictions. However, co-offending is a product of the intricate interplay between
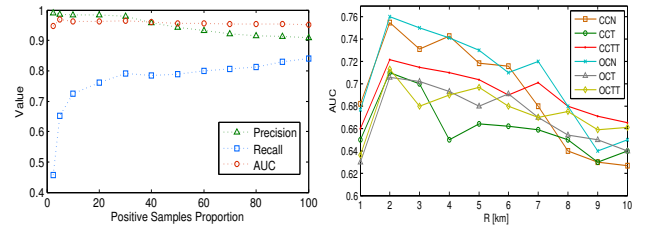

Figure 5: (a) Performance changes of the bagging classifier for different imbalance ratios. The x-axis states the ratio of the positive class size over the negative class size. (b) Single features siginicance using different activity space radius
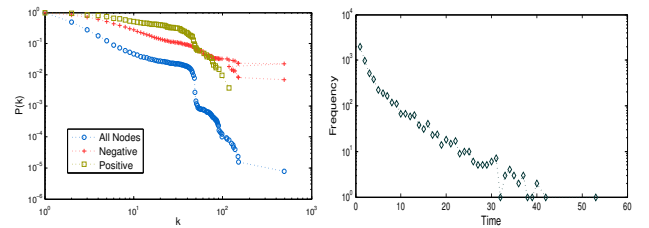

Figure 6: (a) Cumulative degree distribution of offenders (b) Frequency distribution of time difference between two offenders most recent crime and their first co-offence.

many factors, and drawing any firm conclusions requires more substantial and in-depth research beyond the scope of this paper.

Co-offending research currently is non-spatial, except in conflicting edges of gang neighborhoods [34], and usually not associated with crime signatures of crime attributes. Future research should improve the predictive power of link prediction approaches by adding more information about an individual's non-co-offending patterns and co-offending patterns with more guidance for the introduction of preventative programs to reduce gang offences. This research opens the door for advanced approaches to the co-offending networks analysis by identifying those situations where the probability of future co-offending increases. This has potential value in predictive policing by indicating the attributes of individuals who are more likely to engage in future crimes jointly.

7. RELATED WORK

This section discusses published works on co-offending network analysis, crime prediction and link prediction.

7.1 Co-offending Networks

Crime prevention and reduction is the major concern of law enforcement agencies in order to protect the lives and properties of citizens. Just like the abundance of available data on virtually all parts of society can cause an information overload for many, the rapidly growing volume of crime data makes it increasingly challenging for law enforcement and criminal intelligence agencies to analyze all of the data and extract relevant information and knowledge. Data mining constitutes the scientific foundation for managing the large volume of crime data so as to discover new insights for developing effective crime reduction and prevention strategies.

In criminology, several empirical studies that use social network analysis methods to analyze co-offending networks focus on the stability of associations in such networks. Morselli [7] offers a thoughtful insight into 'criminal organizational systems' form a criminal network perspective and applies social network analysis to a number of case studies of criminal groups and organizations. Reiss [6] concludes that the majority of co-offending groups are unstable, and their relationships are short-lived. This finding is corroborated by McGloin et al. [15], who show that there is some stability in co-offending relationships over time for frequent offenders, but delinquents do in general not tend to reuse co-offenders. However, it should be pointed out that the above findings are based on small datasets: 205 individuals in [6], and 5,600 individuals in [15], and may therefore not be representative.

Certain characteristics of co-offending networks, such as being short-lived and having hidden links, make it difficult to adopt and use social network analysis algorithms. Lack of real-world crime data for research purposes and the necessity of cross-disciplinary knowledge are further limiting factors that may explain why social network analysis studies rarely focus on criminal network analysis problems that are important for public safety, such as organized crime groups detection, criminal network destabilization and co-offence prediction.

7.2 Crime Prediction

Two main criminology theories claim that involvement in crime is the result of: 1) an individual's crime propensity, and 2) criminogenic features of the environment to which an individual is exposed. While propensity towards crime has long been studied, in the last few decades criminogenic features of the environment received specific attention. Apart from individuals, spatial aspects increasingly gain momentum, and environmental criminology [16] plays an essential role in crime reduction and prevention tactics.

New research areas emerge, like crime mapping [35], geographic profiling [17] and crime forecasting [4, 5], that support growing and imperative applications of this research field for law enforcement and criminal intelligence agencies.

Crime prediction methods in the literature completely ignore the role of co-offending in committing crime. Rather their goal is to spatially and temporally model observed crimes to predict the time and location of future crimes. For instance in [5], the authors use a point-pattern-based transition density model for crime space-event prediction considering criminal preferences from previous crimes. Given partial information about a crime incident, Crimewalker [12] is an unsupervised method for top-k suspect recommendation, which applies a random walk method on the co-offending network. As concluded in the link prediction literature, supervised methods always outperform the unsupervised competitors. To the best of our knowledge, the proposed approach is the first supervised learning framework for co-offence prediction.

7.3 Link Prediction

Link prediction is important aspect of social network analysis to better understand the network structure. Link prediction methods can be used to extract missing information, identify hidden interactions, evaluate network evolution mechanisms, and beyond.

Most unsupervised link prediction methods [8] rely solely on the network topology and assign scores to potential links based on structural proximity measures such as node neighborhoods or path information. On the other hand, any classification method can be used for link prediction [9, 10, 11]. In a detailed study [10], Lichtenwalter et. al. examine key factors in the link prediction problem, and propose a framework for supervised link prediction.

More recently, location-based social networking services made it possible to study and predict spatial behaviors of social network actors. Wang et. al. [20], using trajectory and communication patterns of users, concluded that combining mobility and network features enhances the link prediction results. Scellato et. al. [19], based on their study of the link prediction problem in online location-based social networks, conclude that using information about places people visited boosts the link prediction performance.

Although the challenges and concerns of our study is similar to other link prediction studies [9, 10, 19, 20], there are fundamental distinctions as well. Crime data and co-offending network characteristics are inherently different from affiliation networks, such as co-authorship networks or mobile phone networks. Although environmental activity plays an important role in creating new links, these activities are also very different from location-based social networks. Therefore the feature or prediction spaces defined based on criminological theories are different from the ones known in the literature. After all, the co-offence prediction framework proposed in this research and supported by experimental evaluation is an innovative application of social network analysis that offers a new perspective for crime reduction and prevention strategies.

8. CONCLUSIONS

This paper proposes a supervised learning framework for co-offence prediction. We define the co-offence prediction problem as a link prediction problem in co-offending networks. For our study, we use co-offending networks with more than 150K known criminal offenders. These networks were extracted from a large police-reported crime dataset containing about 4.4M offences over a period of five years in British Columbia, Canada. Considering criminological theories about social, environmental and homophily roots of offending in a comprehensive way, we assign each pair of offenders to socially-related, geographically-related or experience-

related criminal cooperation opportunities. Using these sets, we create three prediction spaces. While we are able to retain half of the co-offences, the prediction space division allows us to significantly reduce class imbalance, which is a major challenge in link prediction. We do not only use homophily, social and spatial characteristics of pairs of offenders to define prediction features, but we also define novel geo-social features combining social and spatial characteristics. Evaluating single feature significance, we conclude that geo-social features generally are more effective than others, while we also observe considerable prediction strength in other features. Employing all features, our classifiers achieve a recall of 84%, 88%, and 94% respectively, for three different prediction spaces, which means that we can correctly predict roughly 90% of the co-offences. We believe that our findings provide valuable insights and novel methods for short-term and long-term crime reduction and prevention strategies.

Acknowledgements

9. REFERENCES

[1] R. Boba. *Crime analysis and crime mapping*. SAGE Publications, 2005.

[2] U. S. Department of Justice. Crime in the United States: uniform crime reports. *Federal Bureau of Investigation; Washington, DC.*, 2008.

[3] J. Zhang. Costs of crime in canada, 2008. *Research and Statistics Division, Department of Justice Canada.*, 2011.

[4] W. Gorr and R. Harries. Introduction to crime forecasting. *International Journal of Forecasting*, 19(4):551–555, 2003.

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

[6] A. J. Reiss Jr. Co-offending and criminal careers. *Crime and justice*, pages 117–170, 1988.

[7] C. Morselli. *Inside criminal networks*, volume 73. Springer, 2008.

[8] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.*, 58(7):1019–1031, May 2007.

[9] M. A. Hasan, V. Chaoji, S. Salem, and M. Zaki. Link prediction using supervised learning. In *Proc. SDM'06*, 2006.

[10] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New perspectives and methods in link prediction. In *Proc. KDD '10*, 2010.

[11] C. Wang, V. Satuluri, and S. Parthasarathy. Local probabilistic models for link prediction. In *Proc. ICDM '07*, 2007.

[12] M. A. Tayebi, M. Jamali, M. Ester, U. Glässer, and R. Frank. Crimewalker: a recommendation model for suspect investigation. In *Proc. RecSys'11*, 2011.

[13] M. A. Tayebi, R. Frank, and U. Glässer. Understanding the link between social and spatial distance in the crime world. In *Proc. SIGSPATIAL GIS'12*, 2012.

[14] F. M. Weerman. Co-offending as social exchange: Explaining characteristics of co-offending. *British Journal of Criminology*, 43(2):398–416, 2003.

[15] J. McGloin, C. J. Sullivan, A. R. Piquero, and S. Bacon. Investigating the stability of co-offending and co-offenders among a sample of youthful offenders. *Criminology*, 46(1):155–188, 2008.

[16] P. J. Brantingham and P. L. Brantingham. *Environmental criminology*. Sage Publications, 1981.

[17] D.K. Rossmo. *Geographic Profiling*. CRC Press, 2000.

[18] E.H. Sutherland. *Principles of criminology*. Philadelphia: Lippincott, 1947.

[19] S. Scellato, A. Noulas, and C. Mascolo. Exploiting place features in link prediction on location-based social networks. In *Proc. KDD'11*, 2011.

[20] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A. Barabasi. Human mobility, social ties, and link prediction. In *Proc. KDD'11*, 2011.

[21] C. Zhang, L. Shou, K. Chen, G. Chen, and Y. Bei. Evaluating geo-social influence in location-based social networks. In *Proc. CIKM'12*, 2012.

[22] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proc. KDD '11*, 2011.

[23] A. Reid, M. A. Tayebi, and R. Frank. Exploring the structural characteristics of social networks in a large criminal court database. In *Proc. ISI'13*, 2013.

[24] P. L. Brantingham, M. Ester, R. Frank, U. Glässer, and M. A. Tayebi. Co-offending network mining. *Counterterrorism and Open Source Intelligence*, pages 73–102, 2011.

[25] D. V. Canter and A. Gregory. Identifying the residential location of rapists. *Journal of the Forensic Science Society*, 34(3):169–175, 1994.

[26] M. Felson. *Crime and Nature*. SAGE Publications, 2006.

[27] B. Snook. Individual differences in distance travelled by serial burglars. *Journal of Investigative Psychology and Offender Profiling*, 1(1):53–66, 2004.

[28] A. Calvó-Armengol and Y. Zenou. Social networks and crime decisions: The role of social structure in facilitating delinquent behavior. *International Economic Review*, 45(3):939–958, 2004.

[29] M. Felson. The process of co-offending. In *M. Smith and D. Cornish, eds, Theory and Practice in Situational Crime Prevention*. Monsey, NJ: Criminal Justice Press, 2003.

[30] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.

[31] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[32] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, pages 10–18, 2009.

[33] F. J. Provost, T. Fawcett, and R. Kohavi. The case against accuracy estimation for comparing induction algorithms. In *Proc. ICML'98*, 1998.

[34] P. J. Brantingham, G. Tita, M. B. Short, and S. Reid. The ecology of gang territorial boundaries. *Criminology,*, 50, 2012.

[35] K. Harries. Mapping crime principle and practice. *Washington D.C, U.S. Department of Justice Office of Justice Programs*, 1999.