SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Spatially Embedded Co-offence Prediction Using Supervised Learning

Mohammad Tayebi[1], Martin Ester[1], Uwe Glässer[1], Patricia Brantingham[2]

[1] Simon Fraser University, School of Computing Science
[2] Simon Fraser University, School of Criminology

Crime

- Crime generates substantial cost.
- Crime reduction and prevention
- Predictive policing
- Data mining approaches for crime prevention
- Forward-thinking, proactive versus reactive
- Micro-level analysis in addition to the macro-level analysis

Predictive Policing

- Predicting offenders
- Predicting victims
- Predicting crime locations
- Predicting criminal collaborations
  - Co-offending network disruption
  - Organized crime detection
  - Co-offence prediction

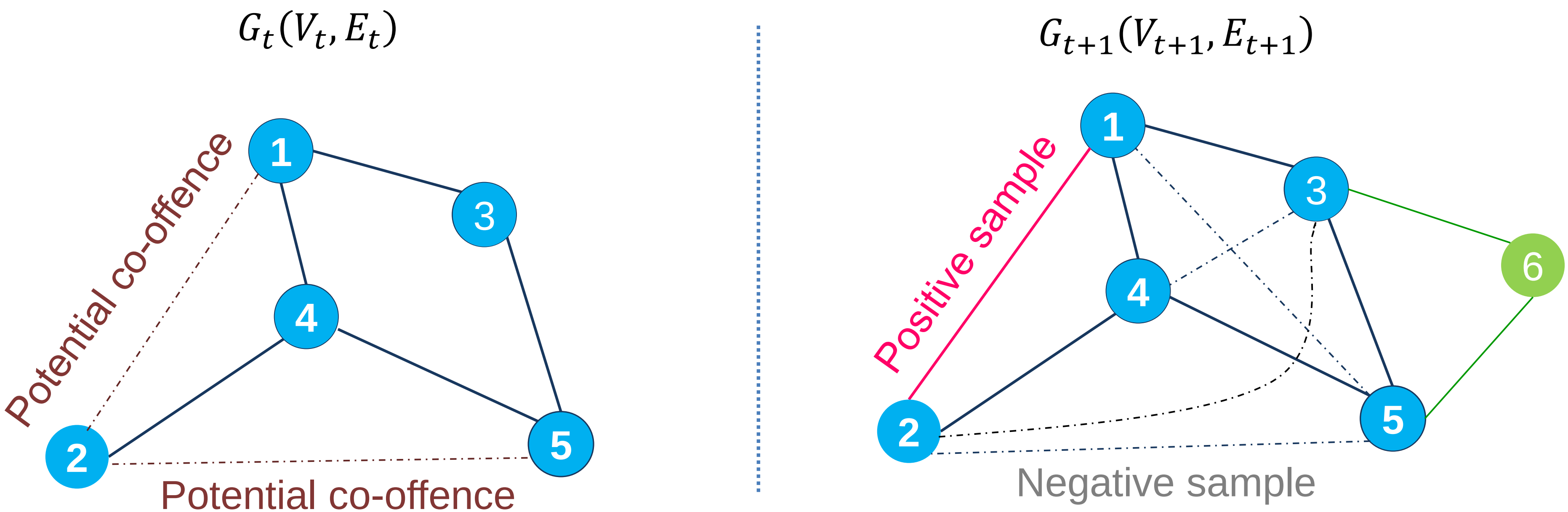
Co-offending Networks

- Network of offenders who have committed crimes together.
  - Node: offender/ Edge: co-offence

> *"Understanding co-offending is central to understanding the etiology of crime and the effects of intervention strategies ". [Reiss, 1988]*

- Most co-offending groups are unstable and the relationships are shortlived [Reiss, 1988].
- There is some stability in co-offending relationships over time [McGloin, 2008].

Co-offence Prediction Problem



Potential co-offence	$(\mathrm{u}, \mathrm{v}) \notin E_t$
Positive sample	$(\mathrm{u}, \mathrm{v}) \notin E_t \wedge (\mathrm{u}, \mathrm{v}) \in E_{t+1}$
Negative sample	$(\mathrm{u}, \mathrm{v}) \notin E_t \wedge (\mathrm{u}, \mathrm{v}) \notin E_{t+1}$

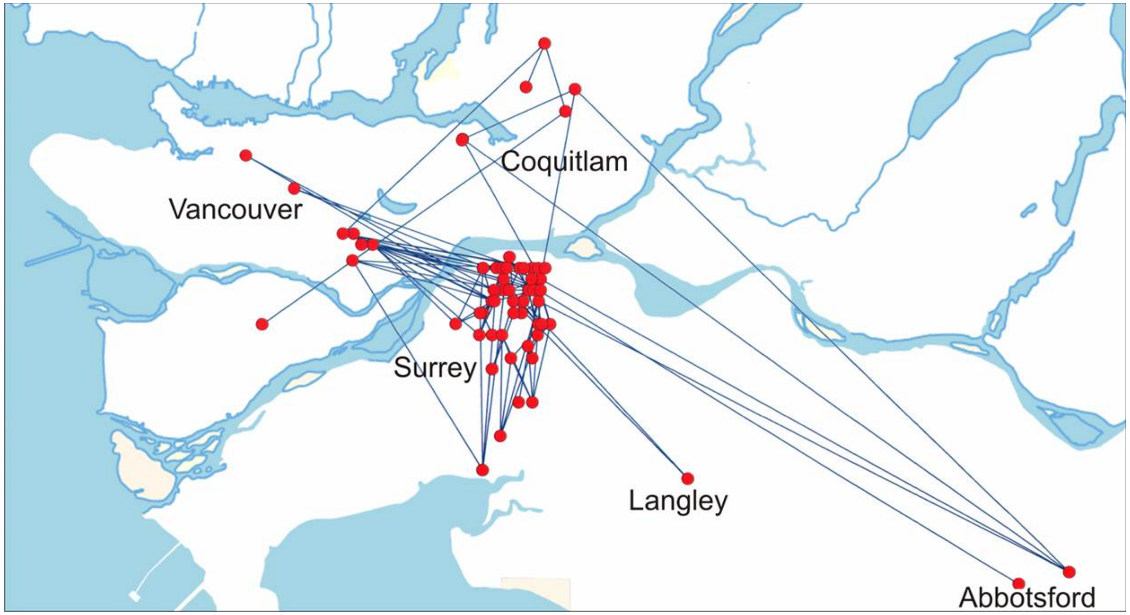
Problem: To predict whether a potential co-offence in G_t belongs to the *positive* class or the *negative* class.

Co-offence Prediction

- Link prediction problem for co-offending networks
- Modeled using a binary classification problem that adopts a set of prediction features
- Main challenge is heavily skewed distribution of negative and positive classes.
  - The prior probability of link formation is very small.
    - Negative class: 890M, Positive class: 11k
  - Classifier overfits to negative samples
  - For co-offence prediction the recall of positive samples is important.

Geographic and Network Proximity

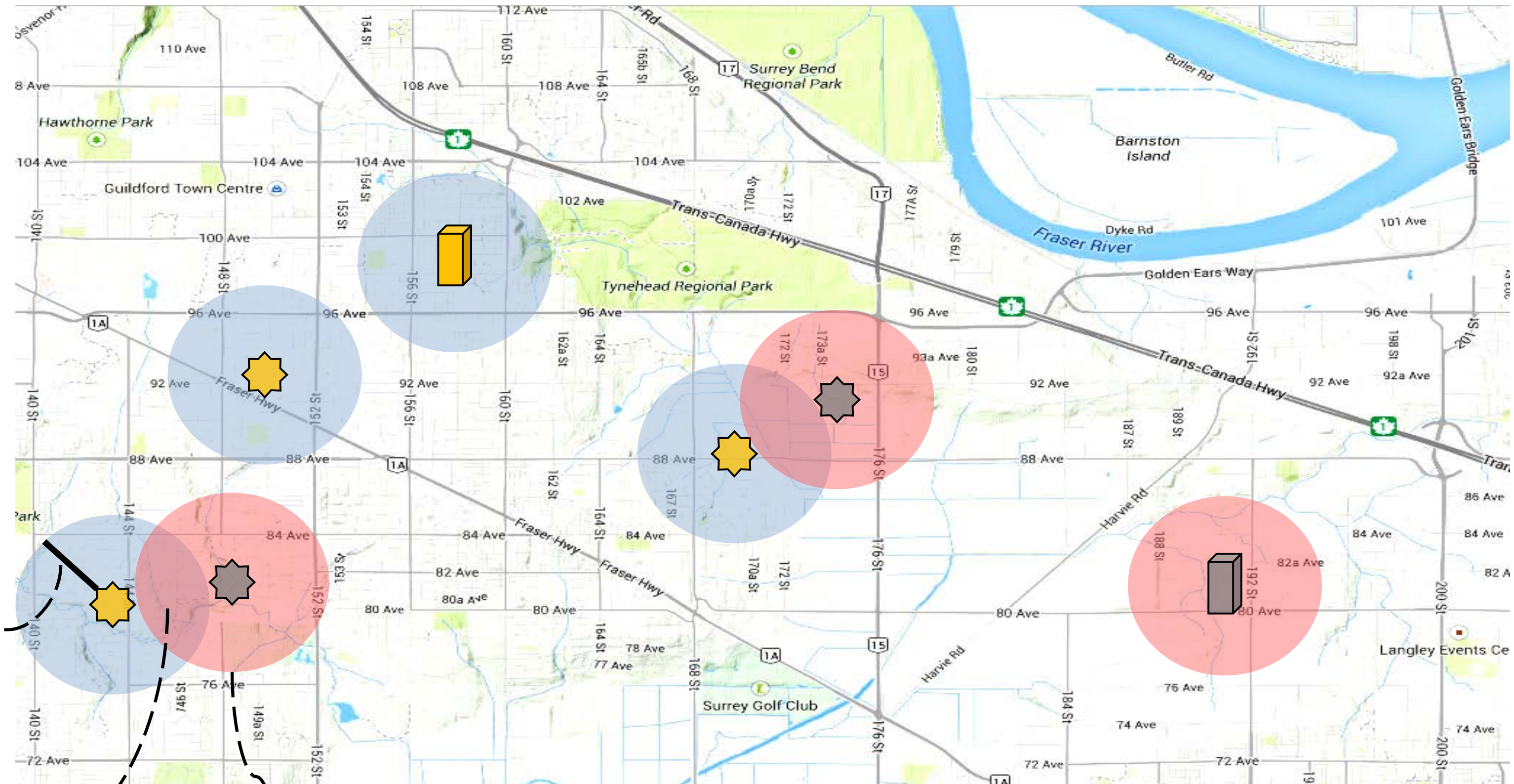
- Crime is strongly linked to geographical characteristics
- British Columbia crime dataset
  - 39% of the co-offenders live in less than 2 km apart.
  - 46% of the crimes happen in less than 2 km distance from the home location.
- Co-offenders tend to be geographically confined.



Common Activity Space

Crime location

Home location



R: predefined Radius

Activity Space: $A_u^R = \{a_u^1, a_u^2, \ldots, a_u^k\}$

Common Activity Space: $A_{u,v}^R = \{a_{u,v}^{i,j} | \ a_u^i \cap a_v^j \neq \varnothing \wedge \delta_u^i \cap \delta_v^j \neq \varnothing\}$

Criminal Cooperation Opportunities

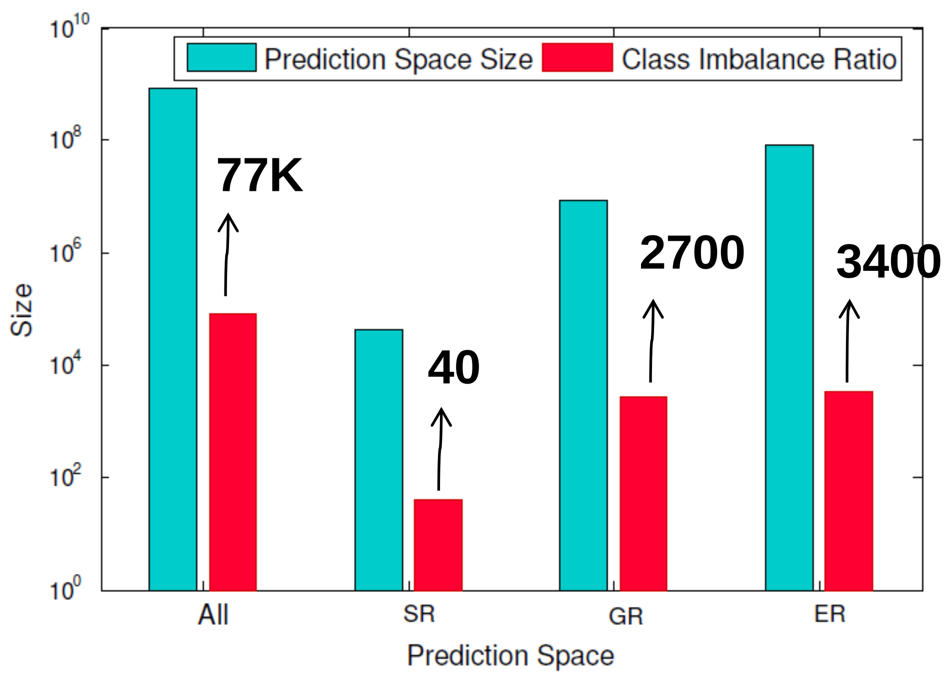
- Offenders do not select their collaborators accidentally.
  - **Socially**-related: distance in the co-offending network (smaller than **N**)
  - **Geographically**-related: common activity space (with radius **R**)
  - **Experience**-related: similar criminal experience (**P** similar crime types)

- Prediction spaces
  - SR Space: socially-related, and N = **2**.
  - GR Space: geographically-related but not socially-related, and R = **2km**.
  - ER Space: Experience-related but not socially-related, and P = **2**.

- Effects of prediction space division
  - Clearer understanding of the effect of criminal cooperation opportunities
  - Reducing class imbalance ratio

Reducing Class Imbalance Ratio

- Reducing class imbalance ratio and keeping as many positive samples as possible.
- With (N=2, R=2km, P=2) the class imbalance ratio decreases significantly for SR, GR and ER spaces.
- We keep about 30% of positive samples in each prediction space, and 50% in total.



SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Prediction Features

Social features: Derived using only the topology of co-offending networks and the position of offenders in the network.

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	$\sum_{z \in \Gamma_u^1 \cap \Gamma_v^1} \frac{1}{log(\Gamma_z^1)}$

Similarity features: Derived from offenders' demographic attributes.

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	$\frac{\sum_{i=1}^{K} P_u^i P_v^i}{\sqrt{\sum_i^K (P_u^i)^2} \times \sqrt{\sum_{i=1}^K (P_v^{i=1})^2}}$

Prediction Features

Geographic features: With increasing the overlap of the activity space of offenders the chance of forming new criminal collaboration increases.

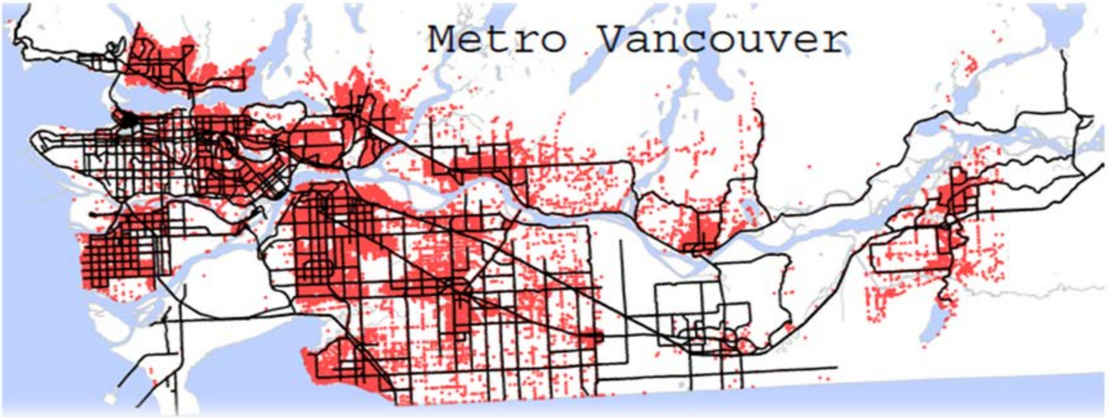
HDN	$\dfrac{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n} e^{\frac{-D(h_u^i,h_v^j)}{\lambda}}}{\lvert H_u\rvert \times \lvert H_v\rvert}$
HDT	$\dfrac{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n} e^{\frac{-D(h_u^i,h_v^j)}{\lambda}} \times \lvert(\delta_u^i \cap \delta_v^j)\rvert}{\lvert H_u\rvert \times \lvert H_v\rvert}$
CDN	$\dfrac{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n} e^{\frac{-D(c_u^i,c_v^j)}{\lambda}}}{\lvert C(u)\rvert \times \lvert C(v)\rvert}$

Geo-Social features: Combines the social and geographic characteristics of offenders.

OCT	$\sum_{i=1}^{i=p}\sum_{j=1}^{i=k} \lvert\chi_{u,v}^{i,j}\rvert$
OCTT	$\sum_{i=1}^{i=p}\sum_{j=1}^{i=k} \lvert\chi_{u,v}^{i,j}\rvert \times \lvert\delta_u^i \cap \delta_v^j\rvert$
OCN	$\sum_{i=1}^{i=p}\sum_{j=1}^{i=k} \lvert\chi_{u,v}^{i,j}\rvert : [t_0, t]$
CCT	$\sum_{i=1}^{i=p}\sum_{j=1}^{i=k} \lvert\phi_{u,v}^{i,j}\rvert$
CCTT	$\sum_{i=1}^{i=p}\sum_{j=1}^{i=k} \lvert\phi_{u,v}^{i,j}\rvert \times \lvert\delta_u^i \cap \delta_v^j\rvert$
CCN	$\sum_{i=1}^{i=p}\sum_{j=1}^{i=k} \lvert\phi_{u,v}^{i,j}\rvert : [t_0, t]$

British Columbia Crime Data

- Police arrest data
  - Covers all of British Columbia policed under contract with the RCMP
  - Between mid-2001 and mid-2006
  - 4.4 million events, 1000 crime types
- Extracted co-offending network
  - 150,000 nodes
  - Average degree of four
  - 50% of the nodes have degree one.
  - The largest connected component links about 18% of the nodes.

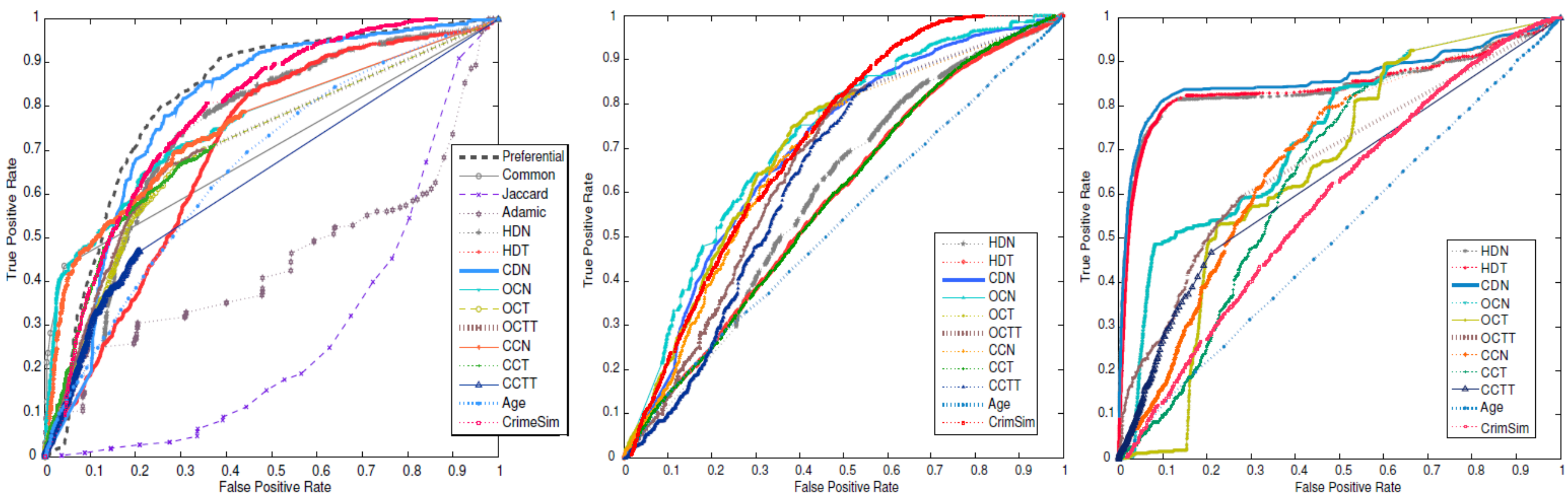


Experimental Design

- Train (first 50 months) and test set (last 10 months)
  - 1.8 M and 800K records
  - 67K and 17K offences with more than one involved offender
- Classification methods: Naïve Bayes, C4.5, random forests, and bagging
- Running 10-fold cross validation over 10 different randomly sampled training sets
- Evaluation measures
  - ROC and AUC

Single Feature Significance

- In SR, Preferential attachment is the best among social feature
- In GR, performance of geographic and geo-social features are weaker.
- Similarity features works better in SR compared to GR and ER.
- Crime locations distance works better than home location distance.



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

Prediction Evaluation

- 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
- Naïve Bayes is the weakest one in all spaces.

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

Prediction Strength

- Geo-social feature set outperforms the other three sets.
- Geographic feature set has the worst performance.
- Prediction performance is the best when we integrate all three feature 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

Conclusion

- Our defined prediction spaces reduces the class imbalance ratio significantly.
- Our novel geo-social feature set outperforms the other feature sets.
- The proposed framework can correctly predict roughly 90% of the co-offences in the prediction spaces.
- Data mining provides valuable insights and novel methods for short-term and long-term crime reduction and prevention strategies.

- In theory, there is no difference between theory and practice. But, in practice, there is.
  - Jan L. A. van de Snepscheut

Thanks !