MASSACHUSETTS INSTITUTE OF TECHNOLOGY

IDS.131 FINAL PROJECT REPORT

GRADUATE CRIMINAL NETWORKS GROUP

Predicting Criminal Behavior with Networks

Authors
HENRY ASPEGREN, SHI KAI CHONG, ADHAM KALILA, SAM MAHLE

December 13, 2017

1 Introduction

Criminal recidivism is a major issue in countries around the world. Efforts to improve rehabilitation are driven by a desire to not only help prisoners and decrease crime rates, but also a need to decrease overcrowded prison populations and expenditures on prisoners. In 2015, for example, the United States spent a combined $ 56.9 billion on correctional facilities and programs [3]. The demand for additional resources poses a major concern for policy makers and correction officers alike. Another study published in 2014 that tracked over 400,000 state prisoners from 30 different states found that over 67 percent were rearrested within 3 years of their release from jail [1]. More informed rehabilitation efforts may help decrease recidivism rates and, by extension, prison population sizes and expenditures. There have been many efforts made to explain recidivism. Researchers have studied how life events such as adverse childhood experiences [6], amount of time spent in prison [5], and many other factors influence recidivism. This research seeks to add to that body of work by using network analysis to develop a deeper understanding of how criminal networks can be used to predict future offenses.

2 Overview

The data used in this research was provided by Carlo Morselli of the University of Montreal. It includes all crimes committed in Quebec, Canada from 2003 to 2010. Critically the arrest data includes a unique identifying number, the case number, which can be used to determine which criminals have been arrested together (this is called a “co-arrest”). The arrest data also provides biographical information about the age and sex of the criminal, the municipality in which the crime was committed, the date of the arrest, as well the crime committed. Using this data, we show that information about co-offenders improves recidivism predictions. We also reveal significant patterns in the co-offender network using community detection techniques. Finally, we use machine learning methods to cluster categories of crime in an attempt to determine how different types of crime are related. The goal of each of these methods is to build greater understanding of the dynamics of a criminal network. If significant results are found, rehabilitators can use this information to prepare more informed intervention and prevention strategies in an effort to reduce recidivism.

3 Predicting Recidivism

3.1 Problem

Criminologists have long speculated about the importance of relationships between criminals. However, it is usually difficult in practice to quantify this impact. In this part of the project, we show that information from the network aspects of crime can be exploited to improve prediction of recidivism. Specifically, we show how features extracted from the co-offending network can be used to improve the accuracy of recidivism predictions.

3.2 Methods

3.2.1 Formal Definition

From the dataset, we selected a subset of users who were first arrested with at least one person (and hence will always have a positive degree in the co-offender network) . Let y be the binary variable that indicates that an offender had been arrested more than once between 2003 and 2010 ($y = 1$ if an offender is arrested more than once and $y = 0$ otherwise). The goal in this prediction task is to predict y given X, a set of feature variables.

3.2.2 Models

A total of 3 models with different sets of feature variables X are considered.

We consider a **Baseline model** where we predict recidivism with only demographic information and arrest data. This model uses the features Age, Gender, Crime Committed in First Arrest and Municipality of First Arrest.

The two other models that incorporated network features are:

Model A : Baseline model + Degree centrality (Degree C.) + Closeness Centrality (Closeness C.) + Eigenvalue Centrality (Eigen C.)+ Local Clustering coefficient (Cluster Coef)

Model B: Baseline model + Average degree centrality of neighbors (avg Degree C.) + Average Closeness Centrality of neighbors (avg Closeness C.) + Average eigenvalue centrality of neighbors (avg Eigen C.) + Average local clustering coefficient of neighbors (avg Cluster Coef1) + Percentage of pairs of neighbors who are connected (avg Cluster Coef2)

The network features are created using the *co-offender network*. Let $G_{x,y} = (V_{x,y}, E_{x,y})$ be an unweighted, undirected graph in which $V_{x,y}$ is the set of all criminals arrested in the time period from year x to year y. We have that vertexes $v, u \in V_{x,y}$ share an undirected edge if the corresponding criminals have been arrested together sometime in the time period between year x and year y. Let the time of an offender's first arrest be z. Network features are extracted on a per-offender basis using $G_{2003,z}$

Using k-fold validation technique (with $k = 10$) , we evaluate the performance of these models in terms of their average validation Area Under the Curve (AUC) score. Critically this means when we evaluate the performance of the model we are only using *past* arrest data to predict *future* arrests. Any future arrest is not represented in co-offender network we used to extract features.

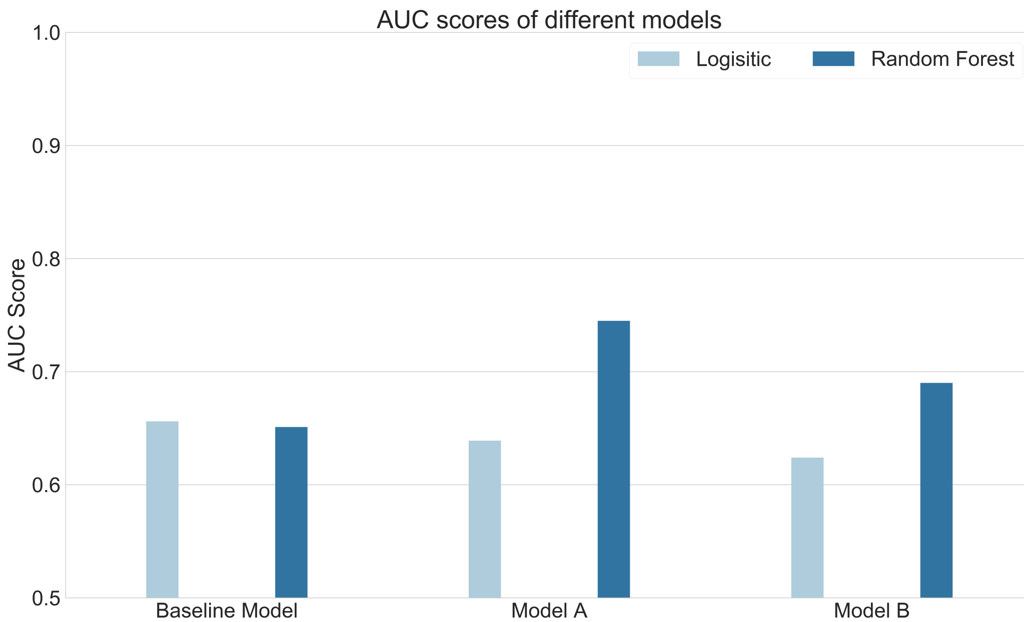
3.3 Results

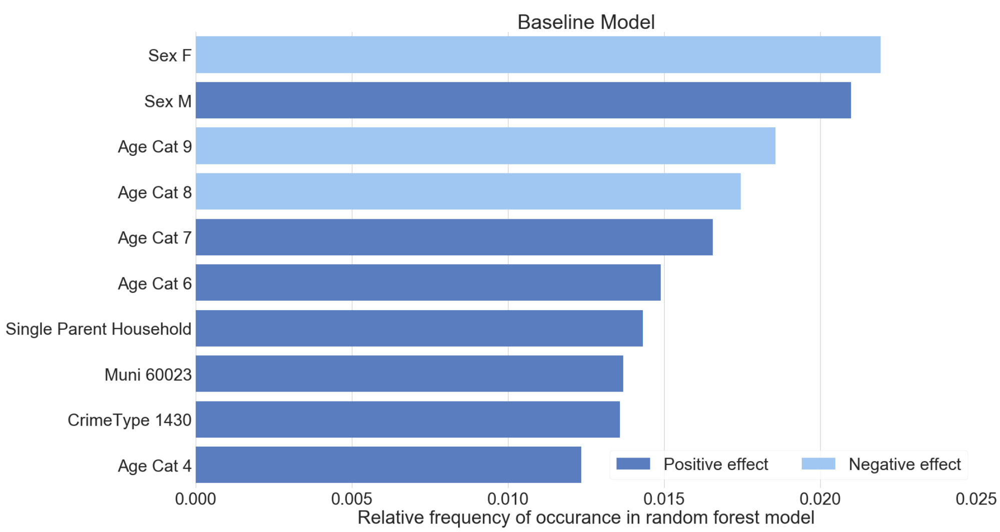
3.3.1 AUC score

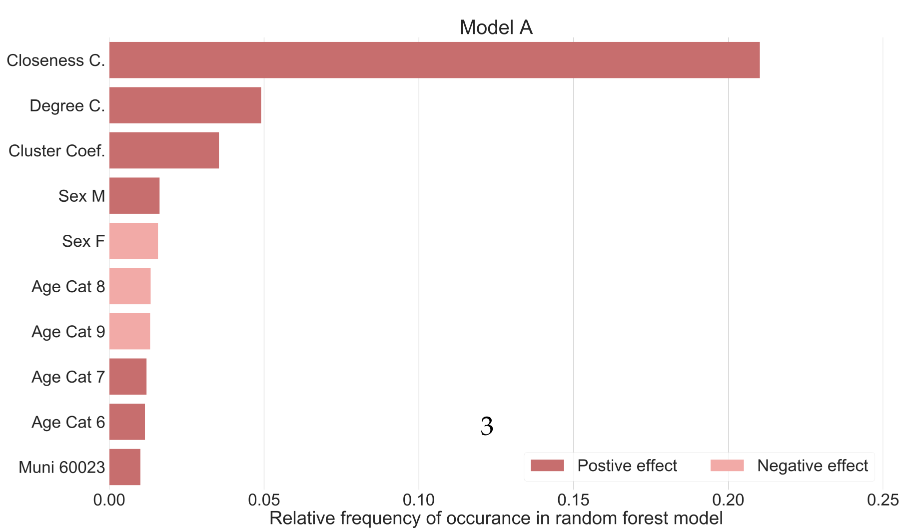
For each set of variables we tried two models - a random forest predictor and a linear model with logistic loss. In this study, the random forest classifier consists of 100 decision trees, each predicting recidivism based on 5 randomly selected variables. The graphs below summarize the performance of these models using the different sets of variables. In general, the random forest models perform better, indicating that there is a non-linear structure in the data. The average area under the ROC (AUC) curve improves from 0.65 to 0.75 when we add network information about the offender directly. On the other hand, if network information about the offender's neighbors is added, the AUC score improves slightly less at 0.7. This shows that the network contains meaningful information that we can utilize to explain recidivism.

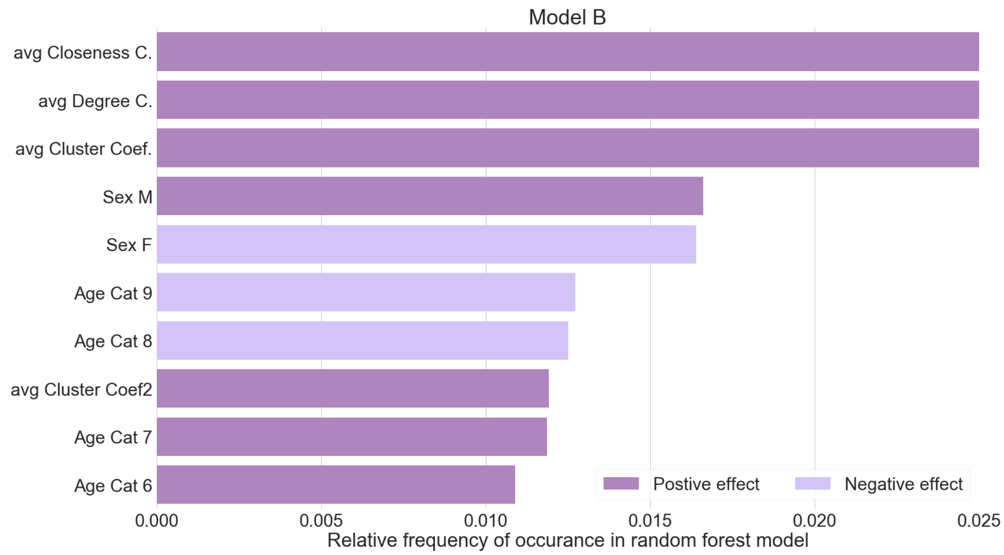
3.3.2 Interpretation of Models

The plots below show the most significant variables in each of the random forest classifiers as measured by the frequency with which the variable appears in the decisions trees. We see that in the baseline model, gender and age (a categorical variable dividing the age into 10 quantiles) are the most significant variables. This makes intuitive sense - men and younger people are more likely to re-offend. For Model A and B, we see that the most significant variables are the network features. In particular, closeness centrality is the most significant, and its frequency is much higher than the demographic variables. It indicates that a high level of closeness centrality would mean a higher chance of recidivism. Intuitively, it could mean that the more embedded an offender is in the criminal network, the more likely he would be influenced by other offenders, increasing his chances of being a repeat offender. Alternatively this could be because the more embedded criminals in the network are less likely to give up a life of crime than someone on the periphery.


AUC scores of different models
1.0
0.9
0.8
0.7
0.6
0.5
AUC Score
Logisitic
Random Forest
Baseline Model
Model A
Model B


Baseline Model
Sex F
Sex M
Age Cat 9
Age Cat 8
Age Cat 7
Age Cat 6
Single Parent Household
Muni 60023
CrimeType 1430
Age Cat 4
Positive effect
Negative effect
0.000
0.005
0.010
0.015
0.020
0.025
Relative frequency of occurance in random forest model


Model A
Closeness C.
Degree C.
Cluster Coef.
Sex M
Sex F
Age Cat 8
Age Cat 9
Age Cat 7
Age Cat 6
Muni 60023
Postive effect
Negative effect
0.00
0.05
0.10
0.15
0.20
0.25
Relative frequency of occurance in random forest model



4 Community Detection

4.1 Problem

Criminals frequently form groups with strong social connections. These groups constitute communities of criminals. For example in organized crime, the community might contain many members working together in a large and sophisticated operation. A community can also capture more subtle relationships and common interests. While community analysis of crime provides an exciting avenue of research it is difficult to reliably identify communities. This is because most relationships between criminals and the communities (if they exist) are latent. However co-offender data provides strong signals: if two criminals are arrested together it is very likely that they have a relationship. Our task is to use this co-offending data to identify communities that are meaningful. While there is no ground truth about what makes a community meaningful, we will define meaningful as providing predictive power about future arrests. We want to detect communities such that (1) future arrests will be more likely between members of the same community and (2) the types of crimes an offender is arrested for in a future is likely to be within the top categories of crime in the community.

4.2 Methods

Let $G_{x,y} = (V_{x,y}, E_{x,y})$ be the unweighted, undirected co-offender network in the time period from year x to year y. Let the time of an offender's first arrest be z. To simulate realistic conditions of data availability, we construct individual networks for each offender using data collected up to z (i.e for each user we analyze the graph $G_{2003,z}$. The steps of the analysis are shown in the flowchart below. An implication of our methodology is that we will only be analyzing users who have been arrested with others (solo offenders are excluded from our study).

In general, the problem of community detection is to assign a label $c_i \in C$ to each vertex $v \in V_{2003,z}$ based only on $G_{2003,z}$ to maximize the within community edges and minimize the edges across communities. In this study, we used two approaches to determine communities in $G_{2003,z}$.

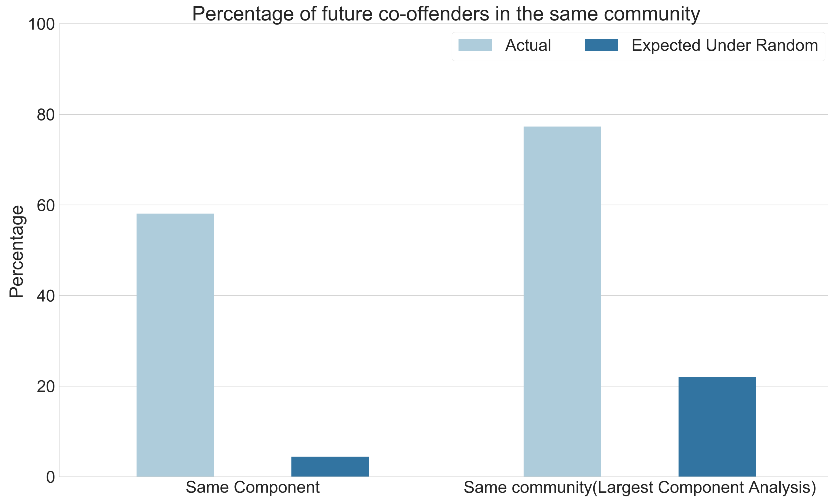
The first approach is conceptually simple and intuitive. We chose the communities to be the strongly connected components of G. This is a very natural choice because if two people are not arrested together, but are both arrested separately with the same person, it is likely that they share some relationship. This is based on the insight that human social networks tend to have relatively high clustering coefficients.

The second approach resolves a significant limitation of the first approach. In the first approach all members of a connected component are in the same community, even if that connected component is very large. For example in $G_{2003,2007}$ the largest connected component has 1459 offenders. However this is unlikely to be a single community because most people can only maintain a much smaller number of relationships. To be able to detect communities within a connected component we needed a new approach.

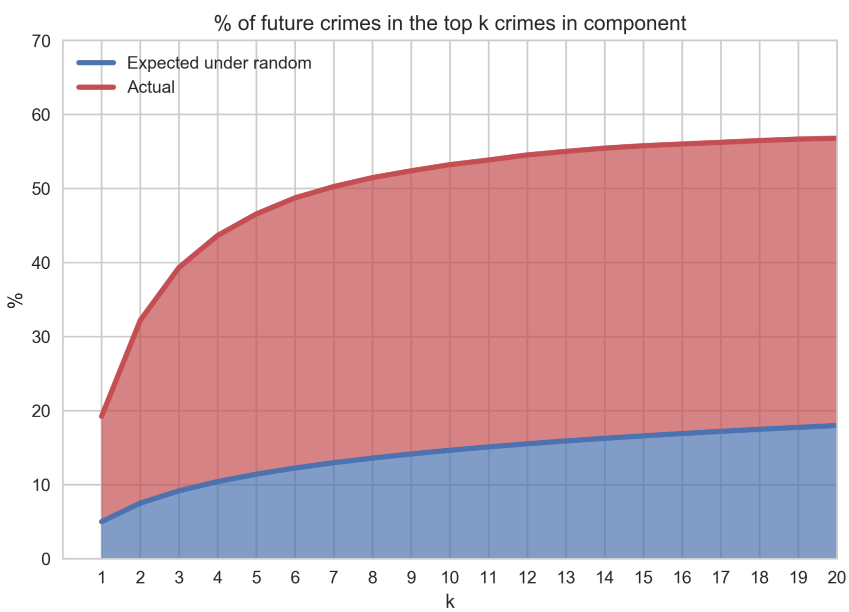
To resolve this, we narrow our analysis down to the subset of users in the largest component, and implement the stochastic block model [2]. In this model each vertex is assigned to a community, and a matrix of edge probabilities P. Edges are then added randomly with independent probability P_{ij} where i and j are the communities of the vertices. The Erdos-Renyi model is just a special case of the stochastic block model where $P_{ij} = p$ for all i and j. In our case we are not given the parameters (e.g. the communities) of the model. Instead we estimate these parameters by trying to maximize the log-likelihood of observing the co-offender network. This is an NP-Hard problem so we must settle for an approximation. Of course more parameters will give a strictly higher log-likelihood so we add a penalty term to punish more complex models. We run this entire inference procedure several times and select the model that maximizes this information criteria. Finally we use the community assignments in the estimated model as our communities.

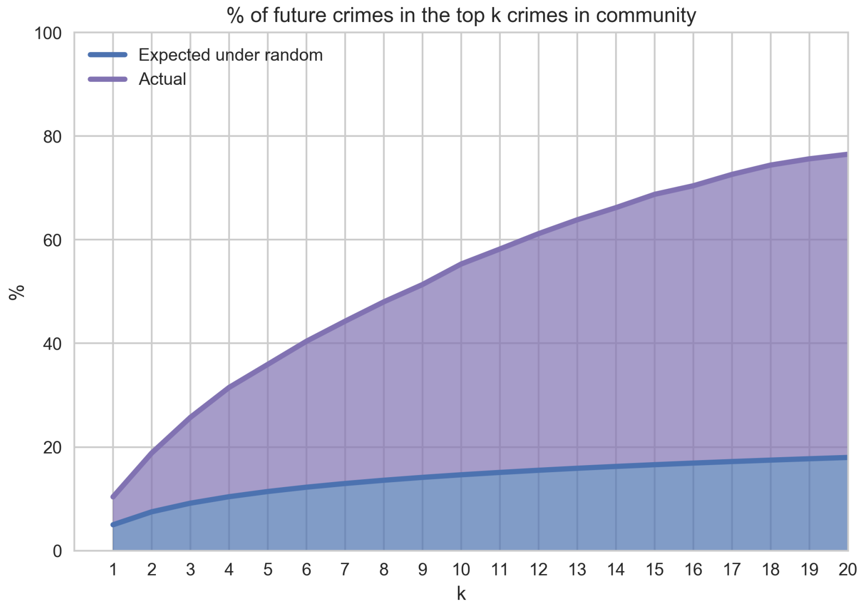
4.3 Results

Averaged across all offenders, 58% of future co-offenders belong to the same component of the offender. By contrast, we would expect this number to be 4.4% if the co-offenders are randomly distributed within the components (with probability proportional to the size of the components). If we focus our analysis to offenders that appear in the largest component, on average 77.3% of the people who are arrested with an offender after his first arrest belong to the same community of the offender. By contrast, we would expect this number to be 21.9% if the co-offenders are randomly distributed within the communities. This is shown in the Figure 5 below.



We can also look at the predictive power of the communities in terms of how much they tell us about the types of crimes associated with an offender's future arrests. The graph below shows the percentage of future crimes conducted by an offender that falls within the top k types of crimes in the offender's first arrest community. We see that 40% of crimes fall within the top 3 types of crime in the offender's community. Narrowing the analysis to communities detected within connected components , we see that 40% of crimes fall within the top 6 types of crime in the offender's community. Nevertheless, these values are much higher than that under a random model where the future crimes committed are random (and proportional to the frequency of the occurrence of the crime in the dataset).





These results show that networks constructed at the time of an offender's first arrest exhibit strong predictive power about future arrests, both in terms of the people who a person is likely to be rearrested with and the type of crimes they are likely to commit. This provides evidence of crime as a social phenomenon, as opposed to the view that crimes conducted are the result of momentary lapse in personal judgment.

5 Clustering Crimes

5.1 Problem

We have shown that arrest data captures meaningful information about crime at an individual level by predicting recidivism and future co-arrests. Could it also tell us something meaningful about the relationship between different types of crime? Criminologists typically divide the nearly 400 or so different crimes into three categories: crimes against a person, crimes against property and crimes against the market. We wondered whether this was the best categorization and if it was supported the by arrest data. In this section we formalize crime categorization as an unsupervised learning problem in which crimes must be assigned to clusters using only arrest data. Note that there is no single definition of what makes crimes similar and so there is no of "ground truth" about how crimes should be grouped together. Our goal instead is to try one definition of similarity and interpret the results. Crucially we will examine whether our approach learns similar categories to the classic categorization or discovers different relationships entirely.

5.2 Methods

In our first approach we used the primary charge for each arrest. We defined the similarity between two crimes to be the number of offenders that have been arrested for both crimes at least once. Note that this requires the criminal to have been arrested at least twice. Intuitively this definition captures the notion that two crimes are similar if they are committed by the same person. We formulated the problem in terms of an undirected graph G = (V,E) in which V is the set of all crimes and two crimes share an edge if they have been committed by the same person. We removed all isolated vertices (this left a total of 169 crimes). We used two versions of G: an unweighted version and a weighted version. In the weighted version the edge weight was the log of the number of offenders who have been arrested for both crimes. We then used the stochastic block model to perform community detection of crimes [see description of this in section

4.2]. We used this model in two different ways. First we fixed the 'number of communities parameter' at three and ran an optimization procedure to maximize the log-likelihood of observing the network. We then used the resulting 3 communities as our clusters. In the second approach we relaxed our restriction on the number of clusters and maximized an information criteria that punished more parameters. We again used the resulting communities as our clusters.

As a separate method for clustering similar crimes based on the co-offender network, we used a low dimensional embedding of a normalized co-occurrence matrix. The co-occurrence matrix was built over all crimes in the co-offending dataset. A unit increase in the co-occurrence between two crimes corresponds to either an arrest event that included both crime types or a case in which the same criminal was arrested for both crime types in separate instances. This affinity matrix was normalized so that crimes that appear more frequently do not get a disproportionate representation in the embedding. The normalization method chosen was association strength, which is a type of probabilistic similarity measure that is recommended for co-occurrence matrix normalization in the literature [4]). The resulting matrix was reduced to 2 dimensions using both principal component analysis and spectral embedding. We then performed hierarchical clustering as well as k-means clustering on the low-dimensional representation. To be consistent with the first clustering method described above, we also removed isolated vertices. Unfortunately the clustering resulted in lumping most of the crimes into one group which is of little practical value. For this reason, we disregard this method of clustering and continue with the results of the stochastic block model described above.

We analyzed our clusterings in two ways. First, for each cluster we computed the percentage of crimes in the standard 3 categories with the following labels:

1. crimes against a person
2. crimes against property
3. crimes in the market

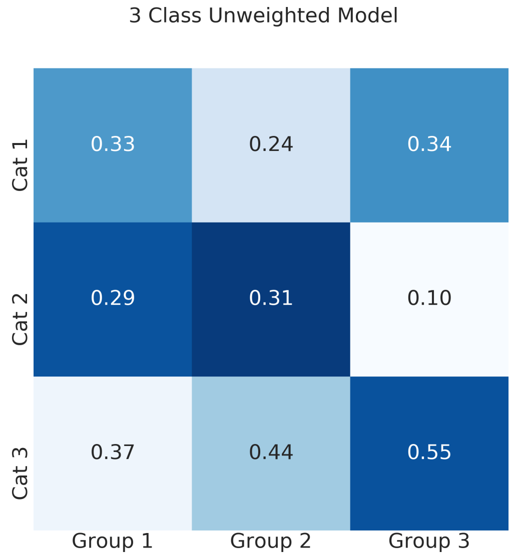
We then compared these to overall distribution of the standard three categories all the crimes before clustering. If our clustering does not capture any aspect of the standard categorization then we would expect the distribution of the categories in each cluster to be similar to the overall distribution. We plot this analysis as a heatmap where the values correspond to the share of each crime and the color corresponds to the deviation from the overall distribution.

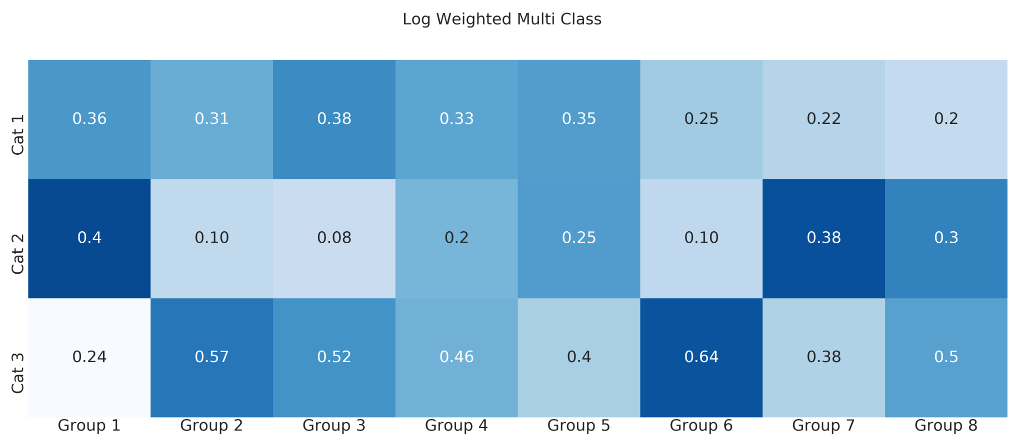
Secondly, we calculate the following metric. For every criminal we look at the primary crime of his first arrest. Then we look at the the primary crimes for up to 3 of his subsequent arrests. We compute the ratio of subsequent arrests that are in the same cluster as the first arrest out of all subsequent arrests. We take the average ratio across all criminals. The results of the rating are normalized for the number of clusters by dividing by the expected percentage of those crimes being in the correct cluster at random. Using this metric, we were able to rate the performance of 6 possible models used for clustering. The results are summarized in the table below. The Normalized ratings, which are a multiple of the random clustering, show that the highest scoring model under this metric is the Log Weighted Multi Class with 8 clusters.

5.3 Results

Table 1: Metric Results for Analyzing Clustering Accuracy

Model Details	Number of Clusters	Average Rating [%]	Normalized
Unweighted 3 Class	3	91.86	2.76
Unweighted Multi Class	3	91.86	2.76
Weigted 3 Class	3	75.83	2.27
Weighted Multi Class	14	26.74	3.74
Log Weighted 3 Class	3	89.96	2.70
Log Weighted Multi Class	8	64.18	5.44





For each of the heatmaps, the groups are the clusterings we produced and the categories correspond to crimes against a person (Cat 1), property (Cat 2), and market (Cat 3).

The complete clusterings are given in the appendix. However we select two groups from the multi-class approach to interpret.

Table 2: Group 1 Cluster Details

	share of Category 1:	0.36
	share of Category 2:	0.40
	share of Category 3:	0.24
	share of drug crimes:	0.04
	Crime	Crime code
0	MEURTRE PREMIER DEGRE	1110
5	TENTATIVE DE MEURTRE	1210
8	AGRESSION SEXUELLE ARMEE	1320
18	VOIES DE FAIT GRAVES NIV3	1410
24	AUTRES VOIES DE FAIT	1480
33	AUTRE CRIME AVEC VIOLENCE	1630
41	PROXENETISME	3120
48	USAGE ARME FEU FAUSSE ARM	3360
53	USAGE DANGEREUX ARME A FE	3385
54	DOC.ADM RELATIVE ARME A F	3390
57	UTIL DE MONN CONTREF.	3420
90	VOL QUALIFIE DE VEHICULE	16104
92	AUTRES VOLS QUALIFIES	16109
95	INTIMID.GEN.ART.423 1.A-G	16701
98	INCENDIE DE VEH. AUTO.	21102
100	INCENDIE AUTRES BIENS	21109
104	AUTRES INTRO AVEC EFFRAC.	21209
105	VOL + 5000$ DANS SUR VEH.	21301
114	VOL TRACTEUR,REMORQUE	21354
116	VOL MOTONEIGE	21356
118	VOL AUTRE VEHICULE	21359
140	MEFAIT DOMMAGE + 5000$	21701
142	MEFAITS + 5000$ SUR VEH.	21703
145	MEFAIT (GRAFFITI) 5000$ -	21706
161	CRACK TRAFIC	42302

Table 3: Group 4 Cluster Details

	share of Category 1:	0.33
	share of Category 2:	0.20
	share of Category 3:	0.47
	share of drug crimes:	0.33
	Crime	Crime code
9	AGRESSION SEXUELLE	1330.0
43	AUTRE ACTE PROSTITUTION	3130.0
51	POSSESSION D”ARMES	3375.0
62	AUTRE SUBSTANCE POSSESS.	4130.0
67	COCAINE TRAFIC	4220.0
69	CANNABIS TRAFIC	4240.0
83	VOIE DE FAIT POLICIER	14601.0
86	SEQUESTRATION	15102.0
87	VOL QUALIFIE SUR PERSONNE	16101.0
93	EXTORSION - PERSONNE	16201.0
102	INTRO EF. CAMP CHA. ROUL.	21202.0
117	VOL VEHICULE TOUT TERRAIN	21357.0
123	VOL 5000$ ET - DE BICYCL.	21406.0
157	CRACK POSSESSION	41301.0
168	CANNABIS EN TERRE	44402.0

5.3.1 Interpretation

We see that our clustering of 3 groups was surprisingly consistent with the standard 3 categories. Our second cluster was mostly property crimes, our third cluster was mostly market crimes and our first cluster was a mix of personal and property crimes. This is interesting because our method appears to have learned the standard 3 categories from the arrest data. However we note that all of our clusters are still mixtures of the standard categories. In particular clusters 1 and 3 contain a mixture of two categories. This could be because many people tend to commit crimes from multiple categories. Under our definition this would result in crimes from different categories being grouped together.

When we let the information criteria select the number of clusters, it returned a total of 8 clusters. This is interesting because it suggests that 3 categories might be an oversimplification. Looking at the individual crimes in each cluster, the results are surprisingly interpretable. We have selected clusters 1 and 4 to demonstrate this. Notice how cluster 1 is primarily violent crimes such as murder, assault and the use of dangerous weapons. These are very serious crimes that carry long sentences. Group 4, on the other hand is predominantly drug crime and more minor violent crimes consistent with drug dealing. This is of course anecdotal and not conclusive, but it provides intuitive validation of the approach.

6 Summary

This study clearly reveals that the social component of crimes can be used to explain future crime events. The results of the regression indicate that the network information is significant in predicting recidivism. The variable with the most significant impact is closeness centrality, indicating that a person’s relationships with other convicted criminals provides the most information about the likelihood that he/she will commit future crime. These results, however, do not provide information about which crimes a criminal will commit in the future, or with whom they will commit future crimes. The community detection method makes progress towards answering these questions. By identifying meaningful groups within the criminal network, we confirm the intuitive fact that people will likely commit future crimes with others in their community, and often repeat the same or similar kinds of crime. This is a significant finding for rehabilitators, who can more selectively monitor with whom they allow their released criminals to interact. While this study does not prove causality, it may be the case that preventing such interactions will lead to decreased recidivism. We also show that the most common crimes within a community are good predictors of the types of crimes people in that community will commit in the future. Finally, our clustering method shows that there are different ways of categorizing crimes other than the current method of manual assignment. While there are many potential applications of the clusters we found, perhaps the most important conclusion is the implication is that correction officers and councilors should reconsider the way in which they think about crimes and crime categories.

7 Future Work

This research revealed that network information is significant in predicting recidivism, and lays the foundation for predicting with whom a criminal is likely to be rearrested. Further research should incorporate crime types using the clustering models introduced in this research to make more accurate predictions about what crime a person is likely to commit in the future. Additionally, further research should focus on methods of identifying the "key players" in each community that we identified. Doing so may reveal those criminals which correction officers should focus on arresting in order to prevent them from negatively influencing the rest of the community

References

[1] Trends in u.s. corrections. the sentencing project. Accessed: December 10, 2017.

[2] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[3] Matthew R Durose, Alexia D Cooper, and Howard N Snyder. *Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010*. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics Washington, DC, 2014.

[4] Nees Jan van Eck and Ludo Waltman. How to normalize cooccurrence data? an analysis of some well-known similarity measures. *Journal of the Association for Information Science and Technology*, 60(8):1635–1651, 2009.

[5] Daniel P Mears, Joshua C Cochran, William D Bales, and Avinash S Bhati. Recidivism and time served in prison. *J. Crim. L. & Criminology*, 106:83, 2016.

[6] Kevin T Wolff and Michael T Baglivio. Adverse childhood experiences, negative emotionality, and pathways to juvenile recidivism. *Crime & Delinquency*, page 0011128715627469, 2016.

8 Appendix

Unweighted 3 Class Crime Cluster

Table 4: Group 1 Cluster Details

	share of Category 1:	0.33
	share of Category 2:	0.29
	share of Category 3:	0.38
	share of drug crimes:	0.13
	Crime	Crime code
5	TENTATIVE DE MEURTRE	1210.0
8	AGRESSION SEXUELLE ARMEE	1320.0
10	AUTRE INFRACTION SEXUELLE	1340.0
18	VOIES DE FAIT GRAVES NIV3	1410.0
21	INFLIGER ILLEGAL. LESIONS	1440.0
23	NEGLIGENCE CRIM. LESIONS	1470.0
24	AUTRES VOIES DE FAIT	1480.0
32	APPELS TEL. INDEC.HARASS.	1626.0
33	AUTRE CRIME AVEC VIOLENCE	1630.0
41	PROXENETISME	3120.0
43	AUTRE ACTE PROSTITUTION	3130.0
48	USAGE ARME FEU FAUSSE ARM	3360.0
53	USAGE DANGEREUX ARME A FE	3385.0
55	ENTREPOSAGE NON SECUR ARM	3395.0
57	UTIL DE MONN CONTREF.	3420.0
64	METHAMPHETAM. POSSESSION	4150.0
80	ECSTASY PRODUCTION	4460.0
85	ENLEVEMENT	15101.0
89	VOL QUALIFIE INST. FINANC	16103.0
90	VOL QUALIFIE DE VEHICULE	16104.0
91	VOL QUALIFIE SAC A MAIN	16106.0
92	AUTRES VOLS QUALIFIES	16109.0
95	INTIMID.GEN.ART.423 1.A-G	16701.0
96	INTIMID. PERS. SYST. JUST	16702.0
98	INCENDIE DE VEH. AUTO.	21102.0
100	INCENDIE AUTRES BIENS	21109.0
104	AUTRES INTRO AVEC EFFRAC.	21209.0
105	VOL + 5000$ DANS SUR VEH.	21301.0
113	VOL MOTOCYCLETTE	21353.0
114	VOL TRACTEUR,REMORQUE	21354.0
115	VOL VR CONSTRUCTION,FERME	21355.0
116	VOL MOTONEIGE	21356.0
118	VOL AUTRE VEHICULE	21359.0
120	VOL 5000$ - DE SAC A MAIN	21403.0
121	VOL 5000$ - A LA TIRE	21404.0
128	OBTENTION FRAU.GITE-NOUR.	21604.0
129	OBTENTION FRAU. TRANSPORT	21605.0
130	FRAUDE CHANGEM. D”ETIQ.	21606.0
140	MEFAIT DOMMAGE + 5000$	21701.0
142	MEFAITS + 5000$ SUR VEH.	21703.0
145	MEFAIT (GRAFFITI) 5000$ -	21706.0
150	DISTRIBUT. PORNO JUVENILE	34552.0
151	POSSESSION PORNO JUVENILE	34553.0
155	AUT INF CONT. FRAUD. COMM	37909.0
161	CRACK TRAFIC	42302.0
162	CRACK POS TRAFIC	42303.0
164	METHAMPHETAM. POS TRAFIC	42501.0
167	CANNABIS HYDROPONIQUE	44401.0

Table 5: Group 2 Cluster Details

	share of Category 1:	0.34
	share of Category 2:	0.11
	share of Category 3:	0.55
	share of drug crimes:	0.22
	Crime	Crime code
0	MEURTRE PREMIER DEGRE	1110.0
1	MEURTRE DEUXIEME DEGRE	1120.0
2	HOMICIDE INVOL. COUPABLE	1130.0
3	NEGLIGENCE CRIMIN. MORT	1150.0
4	AUTRES INFRACTIONS MORT	1160.0
6	CONSPIRATION MEURTRE	1220.0
7	AGRESSION SEXUELLE GRAVE	1310.0
11	CONTACT SEXUEL	1345.0
12	INCITATION CONTACT SEXUEL	1350.0
13	INCESTE	1360.0
14	CORRUPTION D”ENFANT	1365.0
15	LEURRE AVEC UN ORDINATEUR	1370.0
16	RELATION SEXUELLE ANALE	1375.0
17	VOYEURISME	1385.0
22	DECH. ARME AVEC INTENTION	1450.0
25	PRISE D”OTAGE	1520.0
26	TRAITE DE PERSONNE	1525.0
27	ENLEVEMENT MOINS 14 ANS	1530.0
28	ENLEVEMENT MOINS 16 ANS	1540.0
29	ENLEV. ORDONNANCE GARDE	1550.0
30	ENLEV. SANS ORDON. GARDE	1560.0
35	MORT/LESION INCEND/MEFAIT	1650.0
36	MORT/LESIONS - EXPLOSIFS	1660.0
37	INTRO EF. VOL ARMES A FEU	2121.0
39	MAISON DE DEBAUCHE	3110.0
40	PROST-18,VIVRE PROD PROST	3115.0
42	PROST-18,PROXENETISME	3125.0
44	MAISON DE PARIS	3210.0
45	MAISON DE JEUX	3220.0
46	EXPLOSIFS	3310.0
47	ARME A FEU:VENTE ACQUISI	3340.0
49	TRAFIC D”ARMES	3365.0
52	IMPORT EXPORT NON AUTO AR	3380.0
54	DOC.ADM RELATIVE ARME A F	3390.0
58	PROD.DISTR PORNO JUVENILE	3455.0
59	RECYCL PROD CRIMIN(C.CR)	3890.0
60	HEROINE POSSESSION	4110.0
65	ECSTASY POSSESSION	4160.0
66	HEROINE TRAFIC	4210.0
70	METHAMPHETAMINES TRAFIC	4250.0
71	ECSTASY TRAFIC	4260.0
72	HEROINE IMPORT-EXPORT	4310.0
73	COCAINE IMPORT-EXPORT	4320.0
74	AUTRE SUBSTANCE IMP-EXP	4330.0
75	CANNABIS IMPORT-EXPORT	4340.0
76	METHAMPHETAMINES IMP EXP	4350.0
77	ECSTASY IMPORT EXPORT	4360.0
78	AUTRES DROGUES PRODUCTION	4430.0
79	METHAMPHETAM. PRODUCTION	4450.0
81	USAGE ARME A FEU CRIME	14551.0
82	BRAQUER UNE ARME A FEU	14552.0
94	EXTORSION - AUTRE	16209.0

99	INCENDIE V.R EN MOUVEMENT	21103.0
106	VOL + 5000$ DE SAC A MAIN	21303.0
107	VOL + 5000$ A LA TIRE	21304.0
108	VOL + 5000$ A L"ETALAGE	21305.0
109	VOL + 5000$ DE BICYCLETTE	21306.0
132	FAUSSE RECLAM. GOUVERN.	21608.0
133	AUTRE FRAUDE	21609.0
134	FRAUDE PAR TELEMARKETING	21610.0
135	FRAUDE VAL. MOB. FINANCE	21611.0
136	FAUSSE RECLAM. ASSURANCE	21612.0
137	FRAUDE PAR ORDINATEUR	21613.0
138	CLONAGE CARTES DE SERVICE	21614.0
144	MEFAIT (GRAFFITI) + 5000$	21705.0
146	BIENS CULTE RELIG. -5000$	21708.0
147	LOTERIE ILLEGALE	32302.0
148	AUTRE JEU ET PARIS	32303.0
149	PRODUCTION PORNO JUVENILE	34551.0
152	CORRUPTION FONCTIONNAIRES	37902.0
153	MALVERSATION	37903.0
154	INTIMID. PERS. JUSTICE	37906.0
156	PARTIC. ACT. ORG. CRIMIN.	38403.0
158	HEROINE POS TRAFIC	42101.0
165	ECSTASY POS. TRAFIC	42601.0
166	CRACK IMPORT-EXPORT	43301.0

Table 6: Group 3 Cluster Details

	share of Category 1:	0.24
	share of Category 2:	0.31
	share of Category 3:	0.44
	share of drug crimes:	0.24
	Crime	Crime code
9	AGRESSION SEXUELLE	1330.0
19	AGR.ARMEE OU LESIONS NIV2	1420.0
20	VOIES DE FAIT NIV.1	1430.0
31	HARCELEMENT CRIMINEL	1625.0
34	PROFERER DES MENACES	1640.0
38	RECEL	2150.0
50	POS ARMES EN CONTRAV ORDO	3370.0
51	POSSESSION D"ARMES	3375.0
56	INFRACTION CAUTIONNEMENT	3410.0
61	COCAINE POSSESSION	4120.0
62	AUTRE SUBSTANCE POSSESS.	4130.0
63	CANNABIS POSSESSION	4140.0
67	COCAINE TRAFIC	4220.0
68	AUTRE SUBSTANCE TRAFIC	4230.0
69	CANNABIS TRAFIC	4240.0
83	VOIE DE FAIT POLICIER	14601.0
84	VOIE DE FAIT AGENT PAIX	14602.0
86	SEQUESTRATION	15102.0
87	VOL QUALIFIE SUR PERSONNE	16101.0
88	VOL QUALIFIE DANS COMMERC	16102.0
93	EXTORSION - PERSONNE	16201.0
97	INCENDIE BIEN IMMOBILIER	21101.0
101	INTRO PAR EFFR. DANS RES.	21201.0
102	INTRO EF. CAMP CHA. ROUL.	21202.0
103	INTRO EF. ETA. COM. PUBL.	21203.0
110	AUTRES VOLS + 5000$	21309.0
111	VOL AUTOMOBILE	21351.0
112	VOL CAMION,AUTOBUS	21352.0
117	VOL VEHICULE TOUT TERRAIN	21357.0
119	VOL 5000$ - DANS SUR VEH.	21401.0
122	VOL 5000$ - A L"ETALAGE	21405.0
123	VOL 5000$ ET - DE BICYCL.	21406.0
124	AUTRES VOLS 5000$ ET -	21409.0
125	FRAUDE PAR CARTE SERVICE	21601.0
126	FRAUDE PAR CHEQUE	21602.0
127	FRAUDE PAR GUICHET AUTO..	21603.0
131	SUPPOSITION DE PERSONNE	21607.0
139	AUTRES FRAUDES	21699.0
141	MEFAIT DOMMAGE 5000$ OU -	21702.0
143	MEFAITS 5000$ - SUR VEH.	21704.0
157	CRACK POSSESSION	41301.0
159	COCAINE POS TRAFIC	42201.0
160	AUTRE SUBSTANCE POS TRAF.	42301.0
163	CANNABIS POS TRAFIC	42401.0
168	CANNABIS EN TERRE	44402.0

Log Weighted Multi Class Crime Cluster

Table 7: Group 1 Cluster Details

	share of Category 1:	0.36
	share of Category 2:	0.40
	share of Category 3:	0.24
	share of drug crimes:	0.04
	Crime	Crime code
0	MEURTRE PREMIER DEGRE	1110.0
5	TENTATIVE DE MEURTRE	1210.0
8	AGRESSION SEXUELLE ARMEE	1320.0
18	VOIES DE FAIT GRAVES NIV3	1410.0
24	AUTRES VOIES DE FAIT	1480.0
33	AUTRE CRIME AVEC VIOLENCE	1630.0
41	PROXENETISME	3120.0
48	USAGE ARME FEU FAUSSE ARM	3360.0
53	USAGE DANGEREUX ARME A FE	3385.0
54	DOC.ADM RELATIVE ARME A F	3390.0
57	UTIL DE MONN CONTREF.	3420.0
90	VOL QUALIFIE DE VEHICULE	16104.0
92	AUTRES VOLS QUALIFIES	16109.0
95	INTIMID.GEN.ART.423 1.A-G	16701.0
98	INCENDIE DE VEH. AUTO.	21102.0
100	INCENDIE AUTRES BIENS	21109.0
104	AUTRES INTRO AVEC EFFRAC.	21209.0
105	VOL + 5000$ DANS SUR VEH.	21301.0
114	VOL TRACTEUR,REMORQUE	21354.0
116	VOL MOTONEIGE	21356.0
118	VOL AUTRE VEHICULE	21359.0
140	MEFAIT DOMMAGE + 5000$	21701.0
142	MEFAITS + 5000$ SUR VEH.	21703.0
145	MEFAIT (GRAFFITI) 5000$ -	21706.0
161	CRACK TRAFIC	42302.0

Table 8: Group 2 Cluster Details

	share of Category 1:	0.31
	share of Category 2:	0.11
	share of Category 3:	0.58
	share of drug crimes:	0.21
	Crime	Crime code
1	MEURTRE DEUXIEME DEGRE	1120.0
23	NEGLIGENCE CRIM. LESIONS	1470.0
29	ENLEV. ORDONNANCE GARDE	1550.0
35	MORT/LESION INCEND/MEFAIT	1650.0
40	PROST-18,VIVRE PROD PROST	3115.0
42	PROST-18,PROXENETISME	3125.0
49	TRAFIC D"ARMES	3365.0
59	RECYCL PROD CRIMIN(C.CR)	3890.0
66	HEROINE TRAFIC	4210.0
70	METHAMPHETAMINES TRAFIC	4250.0
74	AUTRE SUBSTANCE IMP-EXP	4330.0
81	USAGE ARME A FEU CRIME	14551.0
94	EXTORSION - AUTRE	16209.0
99	INCENDIE V.R EN MOUVEMENT	21103.0
130	FRAUDE CHANGEM. D"ETIQ.	21606.0
137	FRAUDE PAR ORDINATEUR	21613.0
138	CLONAGE CARTES DE SERVICE	21614.0
144	MEFAIT (GRAFFITI) + 5000$	21705.0
156	PARTIC. ACT. ORG. CRIMIN.	38403.0

Table 9: Group 3 Cluster Details

	share of Category 1:	0.38
	share of Category 2:	0.09
	share of Category 3:	0.53
	share of drug crimes:	0.29
	Crime	Crime code
2	HOMICIDE INVOL. COUPABLE	1130.0
3	NEGLIGENCE CRIMIN. MORT	1150.0
4	AUTRES INFRACTIONS MORT	1160.0
6	CONSPIRATION MEURTRE	1220.0
7	AGRESSION SEXUELLE GRAVE	1310.0
11	CONTACT SEXUEL	1345.0
12	INCITATION CONTACT SEXUEL	1350.0
13	INCESTE	1360.0
17	VOYEURISME	1385.0
22	DECH. ARME AVEC INTENTION	1450.0
25	PRISE D''OTAGE	1520.0
30	ENLEV. SANS ORDON. GARDE	1560.0
46	EXPLOSIFS	3310.0
60	HEROINE POSSESSION	4110.0
65	ECSTASY POSSESSION	4160.0
71	ECSTASY TRAFIC	4260.0
73	COCAINE IMPORT-EXPORT	4320.0
75	CANNABIS IMPORT-EXPORT	4340.0
78	AUTRES DROGUES PRODUCTION	4430.0
80	ECSTASY PRODUCTION	4460.0
82	BRAQUER UNE ARME A FEU	14552.0
106	VOL + 5000$ DE SAC A MAIN	21303.0
108	VOL + 5000$ A L''ETALAGE	21305.0
109	VOL + 5000$ DE BICYCLETTE	21306.0
133	AUTRE FRAUDE	21609.0
134	FRAUDE PAR TELEMARKETING	21610.0
136	FAUSSE RECLAM. ASSURANCE	21612.0
149	PRODUCTION PORNO JUVENILE	34551.0
150	DISTRIBUT. PORNO JUVENILE	34552.0
154	INTIMID. PERS. JUSTICE	37906.0
155	AUT INF CONT. FRAUD. COMM	37909.0
158	HEROINE POS TRAFIC	42101.0
164	METHAMPHETAM. POS TRAFIC	42501.0
165	ECSTASY POS. TRAFIC	42601.0

Table 10: Group 4 Cluster Details

	share of Category 1:	0.33
	share of Category 2:	0.20
	share of Category 3:	0.47
	share of drug crimes:	0.33
	Crime	Crime code
9	AGRESSION SEXUELLE	1330.0
43	AUTRE ACTE PROSTITUTION	3130.0
51	POSSESSION D"ARMES	3375.0
62	AUTRE SUBSTANCE POSSESS.	4130.0
67	COCAINE TRAFIC	4220.0
69	CANNABIS TRAFIC	4240.0
83	VOIE DE FAIT POLICIER	14601.0
86	SEQUESTRATION	15102.0
87	VOL QUALIFIE SUR PERSONNE	16101.0
93	EXTORSION - PERSONNE	16201.0
102	INTRO EF. CAMP CHA. ROUL.	21202.0
117	VOL VEHICULE TOUT TERRAIN	21357.0
123	VOL 5000$ ET - DE BICYCL.	21406.0
157	CRACK POSSESSION	41301.0
168	CANNABIS EN TERRE	44402.0

Table 11: Group 5 Cluster Details

	share of Category 1:	0.35
	share of Category 2:	0.25
	share of Category 3:	0.40
	share of drug crimes:	0.20
	Crime	Crime code
10	AUTRE INFRACTION SEXUELLE	1340.0
21	INFLIGER ILLEGAL. LESIONS	1440.0
32	APPELS TEL. INDEC.HARASS.	1626.0
50	POS ARMES EN CONTRAV ORDO	3370.0
55	ENTREPOSAGE NON SECUR ARM	3395.0
64	METHAMPHETAM. POSSESSION	4150.0
68	AUTRE SUBSTANCE TRAFIC	4230.0
85	ENLEVEMENT	15101.0
89	VOL QUALIFIE INST. FINANC	16103.0
91	VOL QUALIFIE SAC A MAIN	16106.0
96	INTIMID. PERS. SYST. JUST	16702.0
97	INCENDIE BIEN IMMOBILIER	21101.0
113	VOL MOTOCYCLETTE	21353.0
115	VOL VR CONSTRUCTION,FERME	21355.0
120	VOL 5000$ - DE SAC A MAIN	21403.0
121	VOL 5000$ - A LA TIRE	21404.0
128	OBTENTION FRAU.GITE-NOUR.	21604.0
129	OBTENTION FRAU. TRANSPORT	21605.0
162	CRACK POS TRAFIC	42303.0
167	CANNABIS HYDROPONIQUE	44401.0

Table 12: Group 6 Cluster Details

	share of Category 1:	0.25
	share of Category 2:	0.11
	share of Category 3:	0.64
	share of drug crimes:	0.18
	Crime	Crime code
14	CORRUPTION D”ENFANT	1365.0
15	LEURRE AVEC UN ORDINATEUR	1370.0
16	RELATION SEXUELLE ANALE	1375.0
26	TRAITE DE PERSONNE	1525.0
27	ENLEVEMENT MOINS 14 ANS	1530.0
28	ENLEVEMENT MOINS 16 ANS	1540.0
36	MORT/LESIONS - EXPLOSIFS	1660.0
37	INTRO EF. VOL ARMES A FEU	2121.0
39	MAISON DE DEBAUCHE	3110.0
44	MAISON DE PARIS	3210.0
45	MAISON DE JEUX	3220.0
47	ARME A FEU:VENTE ACQUISI	3340.0
52	IMPORT EXPORT NON AUTO AR	3380.0
58	PROD.DISTR PORNO JUVENILE	3455.0
72	HEROINE IMPORT-EXPORT	4310.0
76	METHAMPHETAMINES IMP EXP	4350.0
77	ECSTASY IMPORT EXPORT	4360.0
79	METHAMPHETAM. PRODUCTION	4450.0
107	VOL + 5000$ A LA TIRE	21304.0
132	FAUSSE RECLAM. GOUVERN.	21608.0
135	FRAUDE VAL. MOB. FINANCE	21611.0
146	BIENS CULTE RELIG. -5000$	21708.0
147	LOTERIE ILLEGALE	32302.0
148	AUTRE JEU ET PARIS	32303.0
151	POSSESSION PORNO JUVENILE	34553.0
152	CORRUPTION FONCTIONNAIRES	37902.0
153	MALVERSATION	37903.0
166	CRACK IMPORT-EXPORT	43301.0

Table 13: Group 7 Cluster Details

	share of Category 1:	0.22
	share of Category 2:	0.39
	share of Category 3:	0.39
	share of drug crimes:	0.06
	Crime	Crime code
19	AGR.ARMEE OU LESIONS NIV2	1420.0
20	VOIES DE FAIT NIV.1	1430.0
31	HARCELEMENT CRIMINEL	1625.0
34	PROFERER DES MENACES	1640.0
38	RECEL	2150.0
56	INFRACTION CAUTIONNEMENT	3410.0
63	CANNABIS POSSESSION	4140.0
101	INTRO PAR EFFR. DANS RES.	21201.0
111	VOL AUTOMOBILE	21351.0
119	VOL 5000$ - DANS SUR VEH.	21401.0
122	VOL 5000$ - A L"ETALAGE	21405.0
124	AUTRES VOLS 5000$ ET -	21409.0
125	FRAUDE PAR CARTE SERVICE	21601.0
126	FRAUDE PAR CHEQUE	21602.0
131	SUPPOSITION DE PERSONNE	21607.0
139	AUTRES FRAUDES	21699.0
141	MEFAIT DOMMAGE 5000$ OU -	21702.0
143	MEFAITS 5000$ - SUR VEH.	21704.0

Table 14: Group 8 Cluster Details

	share of Category 1:	0.20
	share of Category 2:	0.30
	share of Category 3:	0.50
	share of drug crimes:	0.40
	Crime	Crime code
61	COCAINE POSSESSION	4120.0
84	VOIE DE FAIT AGENT PAIX	14602.0
88	VOL QUALIFIE DANS COMMERC	16102.0
103	INTRO EF. ETA. COM. PUBL.	21203.0
110	AUTRES VOLS + 5000$	21309.0
112	VOL CAMION,AUTOBUS	21352.0
127	FRAUDE PAR GUICHET AUTO..	21603.0
159	COCAINE POS TRAFIC	42201.0
160	AUTRE SUBSTANCE POS TRAF.	42301.0
163	CANNABIS POS TRAFIC	42401.0