Combined Detection Model for Criminal Network Detection

Fatih Ozgul[1], Zeki Erdem[2], Chris Bowerman[3], and Julian Bondy[4]

[1,3] Faculty of Computing, Engineering &Technology,
University of Sunderland, SR6 0DD, Sunderland, United Kingdom
[2] TUBITAK- UEKAE, Information Technologies Institute,
41470 Gebze, Kocaeli, Turkey
[4] School of Global Studies, Social Science & Planning,
RMIT University, Melbourne, Australia
fatih.ozgul@istanbul.com, chris.bowerman@sunderland.ac.uk,
zeki.erdem@bte.tubitak.gov.tr, bondy@rmit.edu.au

Abstract. Detecting criminal networks from arrest data and offender demographics data made possible with our previous models such as GDM, OGDM, and SoDM and each of them proved successful on different types of criminal networks. To benefit from all features of police arrest data and offender demographics, a new combined model is developed and called as combined detection model (ComDM). ComDM uses crime location, date and modus operandi similarity as well as surname and hometown similarity to detect criminal networks in crime data. ComDM is tested on two datasets and performed better than other models.

Keywords: Criminal networks, crime data mining, clustering, group detection, police arrest data, offender demographics.

Criminal networks have been an interesting domain for computer scientists. Analyzing criminal networks, using various methods to find out relationships between criminals are investigated by social network analysis scientists as well. What is the lacking of research for criminal networks is how to extract possible relationships between criminals and using these relationship links to assume the existence of previously unseen hidden criminal networks. To look for these links, it is best to look for similarities in crime data. Traditionally there are two types of crime data are kept by the police; the first is police arrest records, and the second is offender demographics information.

1 Introduction

Current intelligence security informatics is mainly focused on using social network analysis (SNA) and machine learning techniques for structural and positional analysis [1, 2] of criminal networks [3, 4, 5] where mostly required information is provided

H. Chen et al. (Eds.): PAISI 2010, LNCS 6122, pp. 1–14, 2010.

from non-crime data. More research therefore should be looking for using crime data provided by the police to detect criminal networks. This is because most of criminals have previous crime records and they have similarities for features of their crimes and their demographics information. GDM and OGDM [6, 7] is tested for detecting criminal networks previously and they were successful to extract co-offending knowledge, similarities of crime location, date preferences and modus operandi. Another model Socio cultural Detection Model (SoDM), which is based on criminals' surname and hometown similarity, is tested on two datasets where GDM and OGDM are also tested on. SoDM performed less well than GDM and OGDM, however. Since combining GDM, OGDM, and SoDM into one model can produce better results, a new model (ComDM) is offered to get maximum performance.

In this study, we test Combined Detection Model (ComDM) on two datasets which contain terrorist, drug, mafia, violence, and theft networks from two cities in Turkey; Bursa(including 85 criminal networks) and Diyarbakir (including 40 criminal networks). Detection results are compared against previously known and available criminal networks in datasets; findings of ComDM are measured with precision, recall and f-values. Feedbacks from domain experts in Bursa and Diyarbakir are also added. This paper shows:

- Criminological approach to criminal networks (Section 2).
- A brief literature review for criminal networks (Section 2).
- Formal definition of ComDM (Section 3).
- Results of datasets for ComDM (Section 4)
- Evaluation results for ComDM (Section 5)
- Domain experts' feedback on ComDM (Section5).

2 Criminal Network Detection

Criminal networks can be gangs, drug dealing networks, mafia type violence and theft groups, street pickpockets, hooligan groups, or terrorist cells. As social networks, criminal networks consist of two sorts of elements: actors (such as criminals, suppliers) and relations between these actors. In addition to actors' personal characteristics and skills, criminal network has also special characteristics as an entity; so each network can be holding specific behavioural heuristics, crime committing habits (e.g. modus operandi), and associations with other criminals in order to provide specific tools and skills.

There are criminological approaches for the reason why criminals choose to work together. If we can find those reasons, it is easy to select suitable features in data for using these features in criminal network detection. Some previous research has shown that the most of criminal networks are not dominated by centrally controlled organisations with a clear hierarchy and strict division of tasks where they rather operate in an informal and flexible way [8]. In many ways criminal groups are either similar to each other because of being its members' friend-of-a friend position, referral chain or might be in need of assistance and special expertise from other criminals. The expertise of knowledge is required to operate for their team to function properly. In terms of stability [8] in criminal networks, there are two types of actors (e.g.criminals); *life* *course persistent criminals* begin offending in early ages, engage in an array of offences and persist over the life course. On the other hand *adolescent limited criminals* begin offending later, engage in group based delinquent acts that are indicative of teenage rebellion, but then *adolescent limited criminals* tend to be in general decline of criminality. Canter [9] suggests that two dominant trends are identified for criminal networks; firstly the size of the network, secondly the product of the centrality of leadership. He states that using these criteria, three types of criminal organisation to be specified; *ad-hoc groups, oligarchies,* and *organised criminals*. *Ad-hoc groups* are with relatively little structure, sometimes with just the presence of key central figures. *Oligarchies* are the kind of networks where their communications appear to be controlled by a small group of people. *Organised criminals* are the closest to being an illegal organisation with most of differentiation, indicating a management hierarchy. *Ad-hoc groups* are the smallest sized groups whereas *organised criminals* are the largest sized ones. For instance, he suggests that hooligan groups are less structured whereas drug dealing networks are the most structured were found for all types of criminal activity.

In computer science literature, there are few social computing models developed for detection of criminal networks. The nearest similar research area is online social networks. For instance, Goldbeck [10] worked on online social networks, where users maintain lists of friends and express their preferences for items like movies, music, or books, are very popular. They say that for online social network systems to be effective, it is important to understand *the relationship between social and personal preferences*. Similar to that, Crandall [11] and his friends say that a fundamental open question in the analysis of social networks is to understand the interplay between *similarity and social ties*. He says people are similar to their neighbors in a social network for two distinct reasons: first, they grow to resemble their current friends due to social influence; and second, they tend to form new links to others who are already like them. This phenomenon is called as *selection* by sociologists. People tend to have attributes similar to those of their friends. There are two underlying reasons for this. First, the process of social influence leads people to adopt behaviors exhibited by those they interact with; this effect is at work in many settings where new ideas diffuse by word-of-mouth or imitation through a network of people. Second, distinct reason is that people tend to form relationships with others who are already similar to them.

One of the pioneering projects about criminal networks was COPLINK Connect, Detect and CrimeNet Explorer [12] in Arizona which did entity extraction text mining from police narrative reports, then created links between entities from same documents and created possible criminal networks, plotted these networks using multi dimensional scaling techniques. Xu et al. [13, 14] defined a framework for automated network analysis and visualization. Using COPLINK Connect and COPLINK Detect [12] structure to obtain link data from text, CrimeNet Explorer used a Reciprocal Nearest Neighbour (RNN) based clustering algorithm to find out links between offenders, as well as discovery of previously unknown groups. CrimeNet Explorer used concept space approach for network creation, RNN-based hierarchical clustering algorithm for group detection; social network analysis based structural analysis and Multi Dimensional Scaling (MDS) for network visualization.

FLINTS project [15] used soft behavioural and hard forensic (fingerprints, DNA) to give analysts the ability to build a graphical image of relations between crimes and criminals. FinCEN project [16] also aimed to reveal money laundering networks by comparing financial transactions. Oatley et al. [15] did some link analysis work on burglary cases in the OVER project. Skillicorn [17] did similar work on detection of the clusters within clusters to filter the surplus of information on possible terrorist networks and present the police a viable subset of suspects to work on. Another remarkable work is done by Adderly and Mushgrove [18, 19] applied clustering techniques and Self Organising Maps to model the behaviour of sex offenders.

TMODS, which is developed by 21st Century Technologies [20, 21, 22, 23, 24, 25] automates the tasks of searching for and analyzing instances of particular threatening activity patterns. With TMODS, the analyst can define an attributed relational graph to represent the pattern of threatening activity he or she is looking for. TMODS then automates the search for that threat pattern through an input graph representing the large volume of observed data. TMODS pinpoints the subset of data that match the threat pattern defined by the analyst thereby transforming a manual search into an efficient automated graph matching tool. User defined threatening activity or pattern graph can be produced with possible terrorist network ontology and this can be matched against observed activity graph. At the end, human analyst views matches that are highlighted against the input graph.

3 ComDM

Combined Group Detection Model (ComDM) is developed in order to benefit from maximum use of similarities in criminal behaviour (e.g. choice of crime location, time and modus operandi) between criminals and use of demographic similarities such as family bonds, relative bonds, and coming from the same hometown circumstances. ComDM is a combined model of OGDM [7] and SoDM. Police arrest records and offender demographic data are the input source of this model. Namely; crime location, crime date, crime modus operandi from arrest data, offender surnames and hometown information from offender demographics data are used as input in ComDM. Requirements of applying ComDM are two; the first is availability of crime location, date and modus operandi information in police arrest data and offender surname and place of birth or hometown information in offender demographics data. The second requirement is availability of these data in relational table format.

Six steps are taken in ComDM as presented in figure 1. In general, spatial, temporal, modus operandi, surname, and hometown links are created and then a clustering approach is applied for criminal network detection. The first step is linking similarly behaving criminals (operating in the same location, on the same dates, using similar modus operandi) who also come from the same family and hometown with SQL inner join queries. Using three fields of arrest table are crime features which are spatial, temporal and modus operandi fields for inner join queries; spatial, temporal, and modus operandi links are created. Similar to this operation, using two fields of demographics table are criminal features, which are surname and hometown fields; surname and hometown links are created.

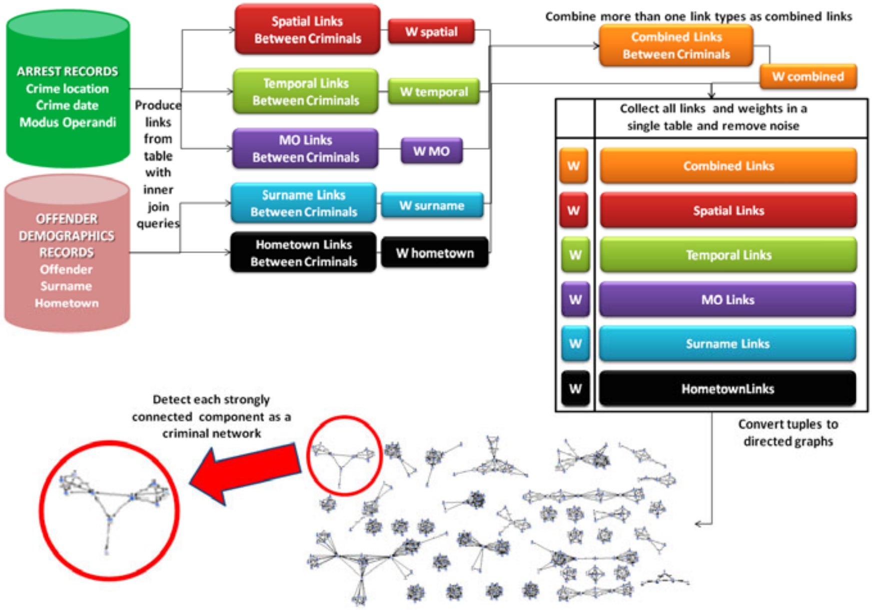

ARREST RECORDS
Crime location
Crime date
Modus Operandi
OFFENDER DEMOGRAPHICS RECORDS
Offender
Surname
Hometown
Produce links from table with inner join queries
Spatial Links Between Criminals
W spatial
Temporal Links Between Criminals
W temporal
MO Links Between Criminals
W MO
Surname Links Between Criminals
W surname
Hometown Links Between Criminals
W hometown
Combine more than one link types as combined links
Combined Links Between Criminals
W combined
Collect all links and weights in a single table and remove noise
W
Combined Links
W
Spatial Links
W
Temporal Links
W
MO Links
W
Surname Links
W
HometownLinks
Convert tuples to directed graphs
Detect each strongly connected component as a criminal network

Fig. 1. Combined Detection Model (ComDM)

For each two offenders (e.g. Offender A, and Offender B) two offenders are linked with spatial link (A, B), temporal link (A, B), MO link (A, B), combined link (A, B), surname link (A, B), hometown link (A, B) as well as spatial link (B,A), temporal link (B,A), MO link (B,A), combined link (B,A), surname link (B,A), hometown link (B,A).

$$\begin{gathered}
spatial\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\
spatial\ Link\ (B,A) = Offender\ B \rightarrow Offender\ A \\
temporal\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\
temporal\ Link\ (B,A) = Offender\ B \rightarrow Offender\ A \\
MO\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\
MO\ Link\ (B,A) = Offender\ B \rightarrow Offender\ A \\
Combined\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\
Combined\ Link\ (B,A) = Offender\ B \rightarrow Offender\ A \\
Surname\ Link\ (A,B) = Offender\ A \rightarrow\ Offender\ B \\
Surname\ Link\ (B,A) = Offender\ B \rightarrow Offender\ A \\
Hometown\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\
Hometown\ Link\ (B,A) = Offender\ B \rightarrow Offender\ A \\
where\ Surname\ Link(A,B) = Surname\ Link\ (B,A) \\
and\ Hometown\ Link(A,B) = Hometown\ Link\ (B,A)
\end{gathered}$$

The second step is giving spatial, temporal, modus operandi, surname, and hometown link weights according to distribution of links for Offender A. Offender A has five link weights compared against all similar offenders as,

$$W_{spatial\ Link(A,B)} = \frac{\sum number\ of\ spatial\ links\ between\ (A,B)}{\sum number\ of\ spatial\ links\ for\ A} \quad (1)$$

$$W_{spatial\ Link(B,A)} = \frac{\sum number\ of\ spatial\ links\ between\ (A,B)}{\sum number\ of\ spatial\ links\ for\ B} \quad (2)$$

$$W_{temporal\ Link(A,B)} = \frac{\sum number\ of\ temporal\ links\ between\ (A,B)}{\sum number\ of\ temporal\ links\ for\ A} \quad (3)$$

$$W_{temporal\ Link(B,A)} = \frac{\sum number\ of\ temporal\ links\ between\ (A,B)}{\sum number\ of\ temporal\ links\ for\ B} \quad (4)$$

$$W_{MO\ Link(A,B)} = \frac{\sum number\ of\ modus\ operandi\ links\ between\ (A,B)}{\sum number\ of\ modus\ operandi\ links\ for\ A} \quad (5)$$

$$W_{MO\ Link(B,A)} = \frac{\sum number\ of\ modus\ operandi\ links\ between\ (A,B)}{\sum number\ of\ modus\ operandi\ links\ for\ B} \quad (6)$$

$$W_{Surname\ Link(A,B)} = \frac{1}{\sum number\ of\ Surname\ links\ for\ A} \quad (7)$$

$$W_{Surname\ Link(B,A)} = \frac{1}{\sum number\ of\ Surname\ links\ for\ B} \quad (8)$$

$$W_{Hometown\ Link(A,B)} = \frac{1}{\sum number\ of\ Hometown\ links\ for\ A} \quad (9)$$

$$W_{Hometown\ Link(B,A)} = \frac{1}{\sum number\ of\ Hometown\ links\ for\ B} \quad (10)$$

The third step is identifying combined links between two offenders where there are more than one type of link exist (e.g. there are links between two offenders as spatial link, temporal link, modus operandi link, surname link, and hometown link). In case of more than two link types between Offender A and Offender B;

$$\begin{gathered} spatial\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\ temporal\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\ MO\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \\ Surname\ Link\ (A,B) = Offender\ A \rightarrow\ Offender\ B \\ Hometown\ Link\ (A,B) = Offender\ A \rightarrow Offender\ B \end{gathered}$$

Then these links are reduced to one single link as;

$$Combined\ Link\ (A,B) = Combined\ Link\ (B,A) = Offender\ A \rightarrow Offender\ B \quad (11)$$

and link weight for combined link is calculated as geometric mean of spatial, temporal, modus operandi, surname and hometown link weights (or just any two of them) as follows;

$$W_{Combined\ Link(A,B)} = W_{Combined\ Link(B,A)} = \sqrt{\begin{array}{c} W_{spatial\ Link\ (A,B)}2 \\ + W_{temporal\ Link\ (A,B)}2 \\ + W_{MO\ Link\ (A,B)}2 \\ + W_{Surname\ Link\ (A,B)}2 \\ + W_{Hometown\ Link\ (A,B)}2 \end{array}} \tag{12}$$

The fourth step is removing weaker links which hold link weighting below a threshold level. Threshold level is 0.1 and links below this threshold is removed as noise. This is represented as,

$$\begin{gathered} W_{spatial\ Link(A,B)}\ and\ \ W_{spatial\ Link(B,A)} \geq 0.1, \\ W_{temporal\ Link(A,B)}\ and\ \ W_{temporal\ Link(B,A)} \geq 0.1\ , \\ W_{Mo\ Link(A,B)}\ and\ \ W_{Mo\ Link(B,A)} \geq 0.1, \\ W_{Surname\ Link(A,B)}\ and\ \ W_{Surname\ Link(B,A)} \geq 0.1, \\ W_{Hometown\ Link(A,B)}\ and\ \ W_{Hometown\ Link(B,A)} \geq 0.1, \\ W_{Combined\ Link(A,B)}\ and\ \ W_{Combined\ Link(B,A)} \geq 0.1 \end{gathered}$$

The fifth step is gathering all combined links and other remaining links which are holding more than 0.1 link weights (e.g. threshold value) in graph format. Resulting graph contains criminal networks which have at least three offenders linked to each other with different type of link weights above the selected threshold (e.g. 0.1). Such as;

$$\begin{aligned} Graph_{criminal\ network} = A, B, C\ and\ spatial\ (A, B), spatial(B, A), surname(A, C), \\ surname\ (C, A), combined\ (B, C), combined(C, B) \end{aligned} \tag{13}$$

Finally the sixth step is detecting individual criminal networks using strongly connected components (SCC) algorithm [26]. We finally get a criminal network in graph format. When a detected graph contains two (or more) criminal networks and they are equal number of criminals in two subgroups, then feature selection method [27] is applied when deciding which criminal network is prevailing. It is based on comparison of means and variances of total link weights. Using selection scores exhibited in the following equations for subgroups, the higher scoring subgroup is decided as detected major criminal network within this graph.

$$Test\ Score\ (1st\ \ subgroup) = \sqrt{\frac{var\ (W_{links\ of\ 1st\ subgroup})}{number\ of\ subgraps\ for\ 1st\ subgroup}} \tag{14}$$

$$Test\ Score\ (2nd\ \ subgroup) = \sqrt{\frac{var\ (W_{links\ of\ 2nd\ subgroup})}{number\ of\ subgraps\ for\ 2nd\ subgroup}} \tag{15}$$

$$\text{let } mean(W_{links\ of\ 1st\ subgroup}) > mean(W_{links\ of\ 2nd\ subgroup})$$

$$Selection\ Score\ (1st\ subgroup) = \frac{|mean\ (W_{links\ of\ 1st\ subgroup}) - mean\ (W_{links\ of\ 2nd\ subgroup})|}{Test\ Score\ (1st\ subgroup)} \quad (16)$$

$$Selection\ Score\ (2nd\ subgroup) = \frac{|mean\ (W_{links\ of\ 1st\ subgroup}) - mean\ (W_{links\ of\ 2nd\ subgroup})|}{Test\ Score\ (2nd\ subgroup)} \quad (17)$$

Confusion matrix for ComDM is constructed as similar to metrics offered by Kaza et al.[28];

Table 1. Confusion matrix for ComDM

	Offenders considered as Criminal Network Members	**Offenders considered as not Criminal Network Members**
Offenders considered refer to the same Criminal Network	**TP**- True Positive	**FP** - False Positive
Offenders considered refer to other Criminal Networks, not to the same Criminal Network	**FN** - False Negative	**TN** - True Negative

In a detected strongly connected component, when detected criminals are all within detected graph, it is accepted as true positive (TP). In case of detected criminals are mostly in the same graph but there are also other unrelated individual criminals, unrelated individual criminals are accepted as false negatives (FN), rightly detected criminals are accepted are still true positive (TP). In case of detected criminals are mostly in strongly connected component but they are belong to more than one criminal networks, such as two or three, criminals which are in the second (or third) unexpected criminal network are accepted as false positive (FP). Precision, recall and f-measure are given in equations below. Precision means that retrieved group is relevant. Recall means that relevant group is retrieved. F-measure is harmonic mean of precision and recall values.

$$precision = \frac{TP}{TP + FP} \quad (18)$$

$$recall = \frac{TP}{TP + FN} \quad (19)$$

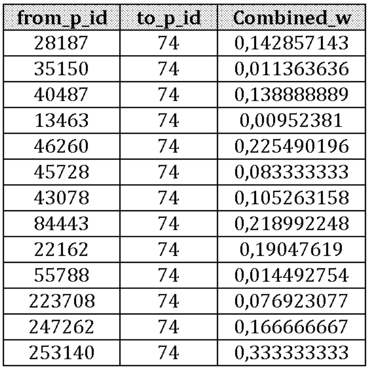
$$F - measure = 2.\frac{precision\ .recall}{precision + recall} \quad (20)$$

4 ComDM Testbeds

To test whether ComDM performs well for detecting criminal networks, two datasets are used as testbeds. They are Bursa Criminal Networks (BCN) and Diyarbakir Drug Networks (DDN) and each has different characteristics. BCN includes 85 criminal network of various types; mostly theft and violence networks, including 8 terrorist networks. DDN includes 40 drug dealing and -mafia type- organised crime networks. Both of the datasets are extracted from massive police databases and they also include some unrelated crimes and criminals. BCN and DDN both are gathered from two sources of crime data; police arrest records, and offender demographics records.

4.1 ComDM Testbed: Bursa Dataset

In this task, all types of links for Bursa Criminal Networks, which are obtained in the previous tasks, are collected. Those are namely spatial, temporal, modus operandi links, and surname and hometown links. All these links are merged altogether, and if there is more than one link between two criminals are available, and then geometric mean of those links weighting values are calculated and given as combined link weight between those two criminals. According to this definition, links and link weights are calculated and they are partly presented in the figure 2. In general, with ComDM, the number of links is increased and there are many more relations between criminals in ComDM.



from_p_id	to_p_id	Combined_w
28187	74	0,142857143
35150	74	0,011363636
40487	74	0,138888889
13463	74	0,00952381
46260	74	0,225490196
45728	74	0,083333333
43078	74	0,105263158
84443	74	0,218992248
22162	74	0,19047619
55788	74	0,014492754
223708	74	0,076923077
247262	74	0,166666667
253140	74	0,333333333

Fig. 2. Combined links and link weights in Bursa Dataset

When these links are collected for building graphs, some members are observed (see Figure 3) to have “brokerage” roles. Detection findings for ComDM are better than GDM, OGDM, and SoDM for Bursa dataset. As presented in figure 4, just one criminal network (e.g. BCN#85) is not detected. Most of the networks are detected with high precision. Another success is about high accuracy of recall score. The only problem is about merging many networks into a single big network, thus detections are typical subgraph detections and some network members are highly connected to some “outer nodes”, which causes them to be “brokers”.

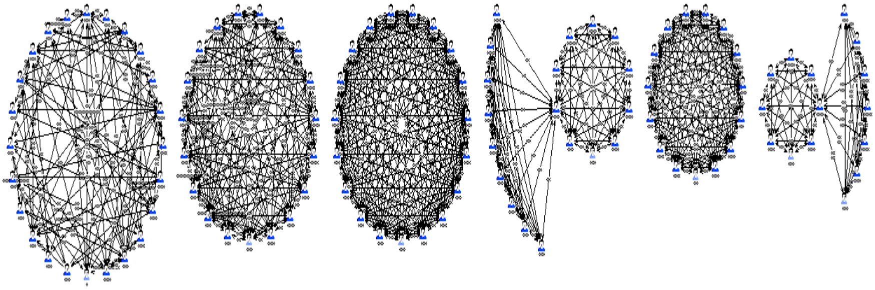
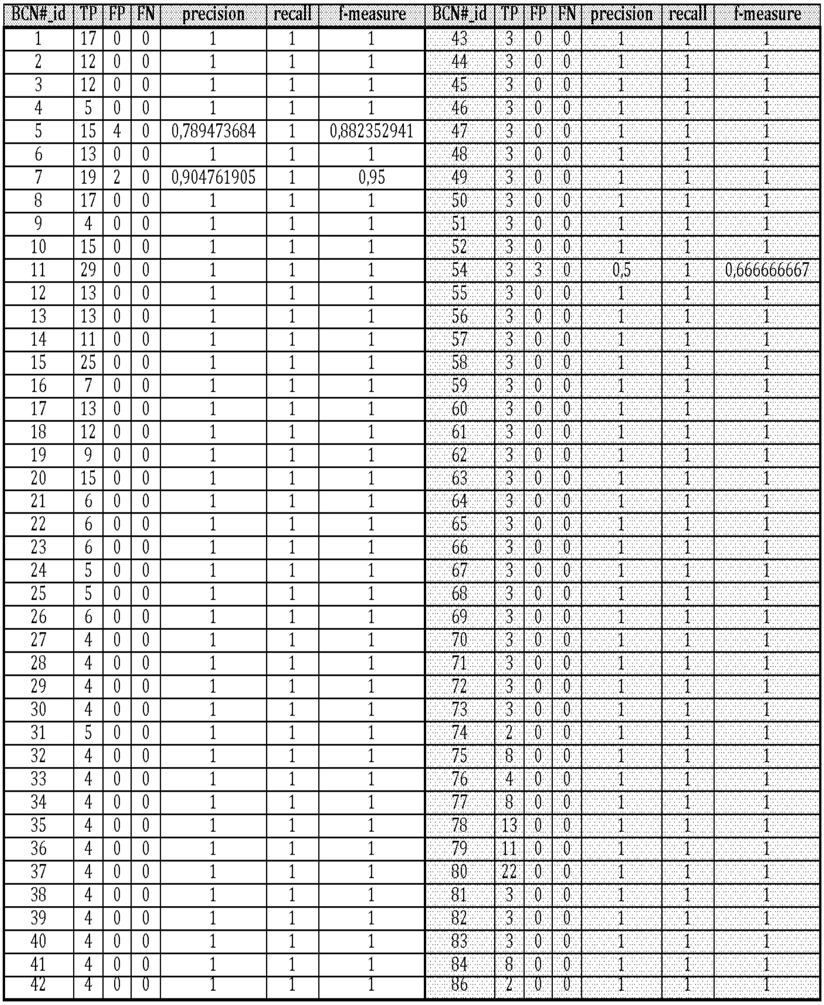


Fig. 3. Example detected criminal networks using ComDM in Bursa Dataset

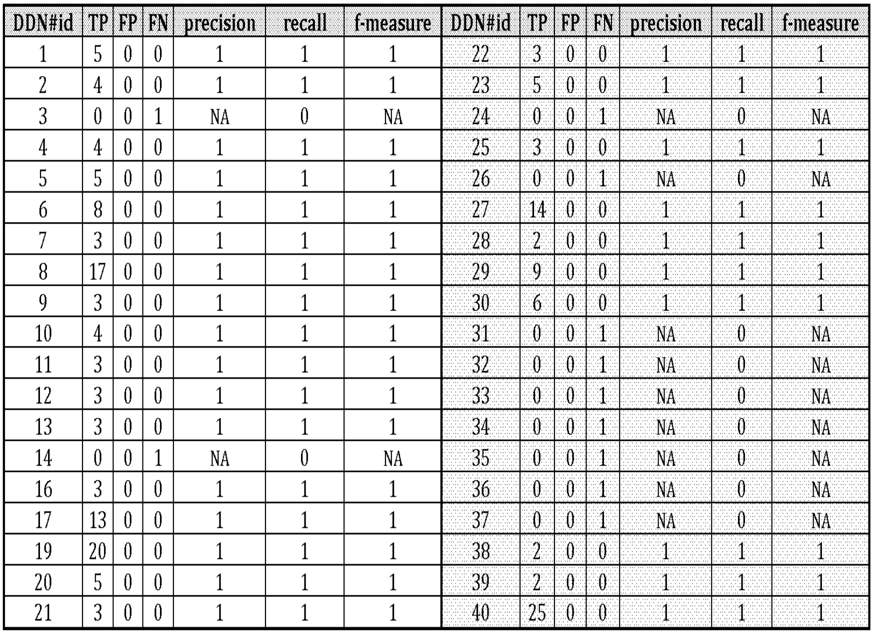


BCN#_id	TP	FP	FN	precision	recall	f-measure	BCN#_id	TP	FP	FN	precision	recall	f-measure
1	17	0	0	1	1	1	43	3	0	0	1	1	1
2	12	0	0	1	1	1	44	3	0	0	1	1	1
3	12	0	0	1	1	1	45	3	0	0	1	1	1
4	5	0	0	1	1	1	46	3	0	0	1	1	1
5	15	4	0	0,789473684	1	0,882352941	47	3	0	0	1	1	1
6	13	0	0	1	1	1	48	3	0	0	1	1	1
7	19	2	0	0,904761905	1	0,95	49	3	0	0	1	1	1
8	17	0	0	1	1	1	50	3	0	0	1	1	1
9	4	0	0	1	1	1	51	3	0	0	1	1	1
10	15	0	0	1	1	1	52	3	0	0	1	1	1
11	29	0	0	1	1	1	54	3	3	0	0,5	1	0,666666667
12	13	0	0	1	1	1	55	3	0	0	1	1	1
13	13	0	0	1	1	1	56	3	0	0	1	1	1
14	11	0	0	1	1	1	57	3	0	0	1	1	1
15	25	0	0	1	1	1	58	3	0	0	1	1	1
16	7	0	0	1	1	1	59	3	0	0	1	1	1
17	13	0	0	1	1	1	60	3	0	0	1	1	1
18	12	0	0	1	1	1	61	3	0	0	1	1	1
19	9	0	0	1	1	1	62	3	0	0	1	1	1
20	15	0	0	1	1	1	63	3	0	0	1	1	1
21	6	0	0	1	1	1	64	3	0	0	1	1	1
22	6	0	0	1	1	1	65	3	0	0	1	1	1
23	6	0	0	1	1	1	66	3	0	0	1	1	1
24	5	0	0	1	1	1	67	3	0	0	1	1	1
25	5	0	0	1	1	1	68	3	0	0	1	1	1
26	6	0	0	1	1	1	69	3	0	0	1	1	1
27	4	0	0	1	1	1	70	3	0	0	1	1	1
28	4	0	0	1	1	1	71	3	0	0	1	1	1
29	4	0	0	1	1	1	72	3	0	0	1	1	1
30	4	0	0	1	1	1	73	3	0	0	1	1	1
31	5	0	0	1	1	1	74	2	0	0	1	1	1
32	4	0	0	1	1	1	75	8	0	0	1	1	1
33	4	0	0	1	1	1	76	4	0	0	1	1	1
34	4	0	0	1	1	1	77	8	0	0	1	1	1
35	4	0	0	1	1	1	78	13	0	0	1	1	1
36	4	0	0	1	1	1	79	11	0	0	1	1	1
37	4	0	0	1	1	1	80	22	0	0	1	1	1
38	4	0	0	1	1	1	81	3	0	0	1	1	1
39	4	0	0	1	1	1	82	3	0	0	1	1	1
40	4	0	0	1	1	1	83	3	0	0	1	1	1
41	4	0	0	1	1	1	84	8	0	0	1	1	1
42	4	0	0	1	1	1	86	2	0	0	1	1	1

Fig. 4. Results, evaluation of ComDM clustering in Bursa Dataset

4.2 ComDM Testbed :Diyarbakir Dataset

In this task, all types of links for Diyarbakir Drug Networks' are collected, merged, and links that are more than one pertaining to the same two nodes are treated by combined links and combined weighting score is used. Spatial, temporal, modus operandi, surname, hometown and combined links are merged altogether. Resulting links and link weights are calculated. Just like in Bursa dataset, the number of links is increased and there are many more relations between criminals than single link types any more. After completion of graph building, disconnected components are identified and treated as individual criminal networks. Results and evaluation metrics for the results are presented in figure 5. Contrary to Bursa dataset, there are twelve networks which cannot be totally detected. DDN#18 is entirely undetected at all. Other eleven criminal networks (DDN#3, DDN#14, DDN#24, DDN#26, DDN#31, DDN#32, DDN#33, DDN#34, DDN#35, DDN#36, DDN#37) are wrongly detected, and just one member is offered which is not member.



DDN#id	TP	FP	FN	precision	recall	f-measure	DDN#id	TP	FP	FN	precision	recall	f-measure
1	5	0	0	1	1	1	22	3	0	0	1	1	1
2	4	0	0	1	1	1	23	5	0	0	1	1	1
3	0	0	1	NA	0	NA	24	0	0	1	NA	0	NA
4	4	0	0	1	1	1	25	3	0	0	1	1	1
5	5	0	0	1	1	1	26	0	0	1	NA	0	NA
6	8	0	0	1	1	1	27	14	0	0	1	1	1
7	3	0	0	1	1	1	28	2	0	0	1	1	1
8	17	0	0	1	1	1	29	9	0	0	1	1	1
9	3	0	0	1	1	1	30	6	0	0	1	1	1
10	4	0	0	1	1	1	31	0	0	1	NA	0	NA
11	3	0	0	1	1	1	32	0	0	1	NA	0	NA
12	3	0	0	1	1	1	33	0	0	1	NA	0	NA
13	3	0	0	1	1	1	34	0	0	1	NA	0	NA
14	0	0	1	NA	0	NA	35	0	0	1	NA	0	NA
16	3	0	0	1	1	1	36	0	0	1	NA	0	NA
17	13	0	0	1	1	1	37	0	0	1	NA	0	NA
19	20	0	0	1	1	1	38	2	0	0	1	1	1
20	5	0	0	1	1	1	39	2	0	0	1	1	1
21	3	0	0	1	1	1	40	25	0	0	1	1	1

Fig. 5. Results, evaluation of ComDM clustering in Diyarbakir Dataset

General detection view is similar to those of co-offending and demographics clustering view (figure 6). The only advantage for ComDM detection is its high accuracy for precision and recall values. In general, ComDM results are better than SoDM results. But GDM and OGDM detection results are far better than ComDM detection results for Diyarbakir drug networks.

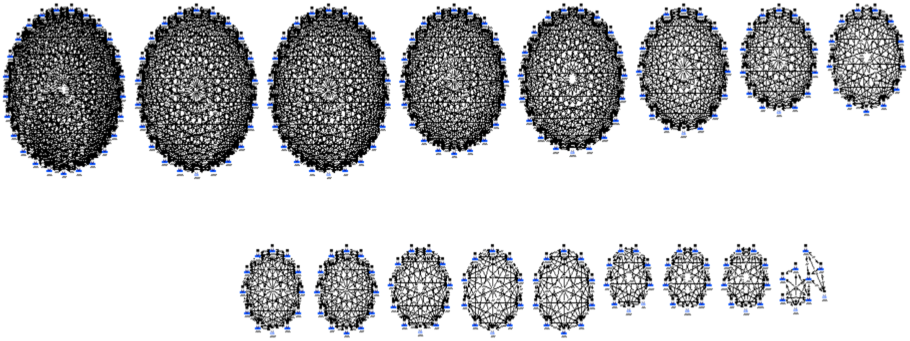


Fig. 6. Example detected criminal networks using ComDM in Diyarbakir Dataset

5 Evaluations and Domain Expert Feedback

Precision, recall and f-test values for Bursa Criminal Networks are 0.99, 0.99, 0.99, for Diyarbakir Drug Networks they are 0.71, 0.71, 0.71. In average, precision, recall and f-test scores for ComDM are 0.85, 0.85, 0.85. ComDM is accepted as successful according to these results. ComDM performs better on BCN to compare against DDN.

Diyarbakir domain experts viewed ComDM as the best detection model compared to others (e.g. GDM, OGDM, and SoDM). Diyarbakir domain experts always recommended using location, modus operandi, surname, and hometown features as potentially the most lucrative features when detecting criminal networks. Contrary to Diyarbakir domain experts, Bursa domain experts offered co-offending, location, and modus operandi as the most valuable features of crime for detecting criminal networks. Bursa domain experts found ComDM as the second most successful model after OGDM. But they said using too much features such as in ComDM might cause bulky results, which is undesirable.

6 Conclusion

We used crime features and offender demographics for detection of criminal networks. ComDM, which uses crime location, date, modus operandi, surname and hometown of criminals to look for similarities between criminals, is designed for better performance on many features of crime and offender demographics. These similarities are processed to obtain criminal networks. Experimental tests on two datasets showed that ComDM performs well on terrorist, theft, violence networks, where it performs less well on drug and mafia networks.

References

1. Brandes, U.: A Faster Algorithm for betweenness centrality. Journal of Mathematical Sociology 25(2), 163–177 (2001)
2. Coffman, T.R., Marcus, S.E.: Pattern Classification in Social Network Analysis: A case study. In: 2004 IEEE Aerospace Conference, March 6-13 (2004)

3. Hunter, A.: Leninist Cell Data Analysis. 21st Century Technologies Inc., Austin (2002)
4. Smith, M.N., King, P.J.H.: Incrementally Visualising Criminal Networks. In: Sixth International Conference on Information Visualisation (IV'02), IEEE, Los Alamitos (2002)
5. Wang, G.A., Xu, J.J., Chen, H.: Using Social Contextual Information to Match Criminal Identities. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences, vol. 04, pp. 81.2. IEEE Computer Society, Los Alamitos (2006)
6. Ozgul, F., Bondy, J., Aksoy, H.: Mining for offender group detection and story of a police operation. In: Sixth Australasian Data Mining Conference (AusDM 2007), Australian Computer Society Conferences in Research and Practice in Information Technology (CRPIT), Gold Coast, Australia (2007)
7. Ozgul, F., Erdem, Z., Aksoy, H.: Comparing Two Models for Terrorist Group Detection: GDM or OGDM? In: Yang, C.C., Chen, H., Chau, M., Chang, K., Lang, S.-D., Chen, P.S., Hsieh, R., Zeng, D., Wang, F.-Y., Carley, K.M., Mao, W., Zhan, J. (eds.) ISI Workshops 2008. LNCS, vol. 5075, pp. 149–160. Springer, Heidelberg (2008)
8. Mcglorin, J.M., Sullivan, C.J., Piquero, A.R., Bacon, S.: Investigating the stability of co-offenders among a sample of youthful offenders. Criminology 46(1), 155–187 (2008)
9. Canter, D.: A partial order scalogram analysis of criminal network structures. Behaviormetrika 31(2), 131–152 (2004)
10. Golbeck, J.: Trust and nuanced profile similarity in online social networks. ACM Trans. Web 3, 4, Article 12, 33 pages(2009)
11. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback Effects between Similarity and Social Influence in Online Communities. In: KDD'08, Las Vegas, Nevada, USA, August 24-27. ACM, New York (2008)
12. Chau, M., Xu, J., Chen, H.: Extracting meaningful entities from police narrative reports. In: National Conference on Digital Government Research (2001)
13. Xu, J., Chen, H.C.: Fighting Organised Crimes: using shortest-path algorithms to identify associations in criminal networks. Decision Support Systems 38(3), 473–487 (2003)
14. Xu, J., Chen, H.C.: CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery. ACM Transactions on Information Systems 23(2), 201–226 (2005)
15. Oatley, G.C., Zeleznikov, J., Ewart, B.W.: Matching and predicting crimes. In: AI 2004- The 24th SGAI International Conference on Knowledge Based Systems and Applications of Artificial Intelligence (2004)
16. Goldberg, H.G., Wong, R.W.H.: Restructuring transactional data for link analysis in FinCEN AI System. In: AAAI Fall Symposium (1998)
17. Skillicorn, D.B.: Clusters within clusters: SVD and counterterrorism. In: Workshop on Data Mining for Counterterrorism and Security (2003)
18. Adderley, R., Badii, A., Wu, C.: The automatic identification and prioritization of criminal networks from police crime data. In: Ortiz-Arroyo, D., Larsen, H.L., Zeng, D.D., Hicks, D., Wagner, G. (eds.) EuroIsI 2008. LNCS, vol. 5376, pp. 5–14. Springer, Heidelberg (2008)
19. Adderley, R., Mushgrove, P.B.: Data mining case study: Modeling the behavior of offenders who commit sexual assaults. In: ACM SIGKDD 2001 International Conference on Knowledge Discovery and Data Mining, New York, pp. 215–220 (2001)
20. Coffman, T., Greenblatt, S., Marcus, S.: Graph-based technologies for intelligence analysis. Communication of ACM 47(3), 45–47 (2004)
21. Marcus, S., Coffman, T.: Terrorist Modus Operandi Discovery System 1.0: Functionality, Examples, and Value. 21st Century Technologies, Austin (2002)
22. Marcus, S.E., Moy, M., Coffman, T.: Social Network Analysis. In: Cook, D.J., Holder, L.B. (eds.) Mining Graph Data. John Wiley & Sons, Inc., Hoboken (2007)

23. Moy, M.: Using TMODS to run best friends group detection algorithm. 21st Century Technologies, Austin (2005)
24. Coffman, T.R., Marcus, S.E.: Dynamic Classification of Suspicious Groups using social network analysis and HMMs. In: 2004 IEEE Aerospace Conference, March 6-13 (2004)
25. Greenblatt, S., Coffman, T., Marcus, S.: Emerging Information Technologies and enabling policies for counter terrorism. In: Behaivoural Network Analysis for Terrorist Detection. Wiley-IEEE Press, Hoboken (2005)
26. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 2nd edn. (2001)
27. Kantardzic, M.: Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons, New York (2003)
28. Kaza, S., Hu, D., Atabakhsh, H., Chen, H.: Predicting criminal relationships using multivariate survival analysis. In: Proceedings of the 8th annual international conference on Digital government research: bridging disciplines & domains, pp. 290–291. Digital Government Society of North America, Philadelphia (2007)