Mining for offender group detection and story of a police operation

Fatih Ozgul [1] Julian Bondy [2] Hakan Aksoy [3]

[1] School of Computing and Technology,
St. Peter’s Campus, University of Sunderland, SR6 0DD, UK
Email: fatih.ozgul@sund.ac.uk

[2] School of Global Studies, Social Science & Planning,
RMIT University, Melbourne, Australia
Email: bondy@rmit.edu.au

[3] Information Processing Unit, Bursa Police Department, Bursa, Turkey
Email:aksoy975@yahoo.com

Abstract

Since discovery of an underlying organisational structure from crime data leads the investigation to terrorist cells or organised crime groups, detecting covert networks are important to crime investigation. As shown in application of Offender Group Detection Model (OGDM), which is developed and tested on a theft network in Bursa, Turkey, use of effective data mining methods can reveal offender groups. OGDM detected seven ruling members of twenty network members. Based on initial findings of OGDM; thirty-four offenders are considered to be in a single offender group where seven of them were ruling members. After *Operation Cash* was launched, the police arrested the seven detected ruling members, and confirmed that the real crime network was consisting of 20 members of which 3 whom had never been previously identified or arrested. The police arrested 17 people, recovered worth U.S. $ 200,000 of stolen goods, and cash worth U.S. $ 180,000.

Keywords: crime data mining, group detection, social network analysis.

1 Introduction

Link analysis and group detection is a newly emerging research area which is at the intersection of link analysis, hypertext and web mining, graph mining (Cook and Holder, 2000) and social network analysis (Scott, 2004). Graph mining and social network analysis in particular attracted attention from a wide audience in police investigation and intelligence (Getoor et al., 2004). As a result of this attention, the police and intelligence agencies realized the knowledge about offender networks and detecting covert networks are important to crime investigation (Senator, 2005). Group detection refers to the discovery of underlying organisational structure that relates selected individuals with each other, in broader context; it refers to the discovery of underlying structure relating instances of any type of entity among themselves (Marcus et al., 2007). Since discovery of an underlying organisational structure from crime data leads the investigation to terrorist cells or organised crime groups, detecting covert networks are important to crime investigation. Detecting an offender group or even a part of group (subgroup) is also important and valuable. A subgroup can be extended with other members with the help of domain experts. An experienced police officer usually knows the friends of well-known offenders, so he can decide which subgroups should be united to constitute the whole group. Another outcome of offender group detection is considered to be pre-emptive strike or crime prevention. For example a drug dealing network prepares all required vehicles and people for transaction where all members are in the process of getting prepared. Such cases can be prevented with offender group detection before it happens. A further advantage of group detection is acting in a group of offenders to commit a crime is regarded as an aggravating factor for a heavier punishment in many country’s laws. For instance, Turkish Crime Code extends six years imprisonment for group leader and one year imprisonment for group members plus the punishment.

Specific software like Analyst Notebook (2007), and Sentient (2007) provide some visual spatio-temporal representations of offender groups in graphs, but they lack automated group detection functionality.

In this paper, we make the following contributions for offender group detection (OGD);

- We identify and discuss converting arrest data to graph format where there is no standardised way of doing this. We suggest the choice of representation for edges and nodes should follow the rules in SNA where mostly one-mode social network representation which is now standard (section 4).

- We explain precisely how to use police arrest data to look for possible offender groups (section 5). Surprisingly this has not been explained precisely before.
- We show how we can apply filters to graph data in order to adhere to countries' criminal law requirements (section 7).
- We show that ruling members, not new recruits, are likely to be detected, but "big brother" of network is unlikely to be detected (section 8).

2 Group Detection

Group detection task is defined and different methods applied in data mining, in social network analysis, and in graph theory. For example, Getoor and Diehl (2005) state group detection aims clustering of object nodes in a graph into groups that share common characteristics. But to some extent, subgraph discovery does the same job for finding interesting or common patterns in a graph. On the other hand social network analysis tries to detect cohesive subgroups among which there are relatively strong, direct, intense, frequent, or positive ties (Wasserman and Faust, 1994). Graph matching (Cook and Holder, 2007) methods are also recommended for group detection tasks. There are also many specific group detection models. Adibi et al. (2004, 2005) propose KOJAK group finder which firstly positioning possible groups, expanding these groups using knowledge-based reasoning techniques and then adding more candidates relying on observed interactions that shows possible associations. Kubica et al. (2002, 2003) first proposes a generative model for multi-type link generation, called collaborative graph model (cGraph) and introduce a scalable group discovery algorithm called k-groups, which is similar to k-means algorithm.

3 OGD

When we focus on offender group detection, the most remarkable works are CrimeNet Explorer, which is developed by Xu et al. (2005), and Terrorist Modus Operandi Detection System (TMODS), which is developed by 21st Century Technologies (Moy, 2005).

3.1 CrimeNet Explorer

Xu et al. (2005) defined a framework for automated network analysis and visualization. Using COPLINK connect and COPLINK detect (Chen et al., 2002) structure to obtain link data from text, CrimeNet Explorer used an Reciprocal Nearest Neighbour (RNN) based clustering algorithm to find out links between offenders, as well as discovery of previously unknown groups. CrimeNet Explorer framework includes four stages: network creation, network partition, structural analysis, and network visualization. CrimeNet Explorer uses concept space approach for network creation, RNN-based hierarchical clustering algorithm for group detection; social network analysis based structural analysis, and multi dimensional scaling for network visualisation. CrimeNet Explorer is the first model to solve offender group discovery problem and its success comes from the powerful functionality of overall COPLINK structure. On the other hand, since CrimeNet Explorer was evaluated by university students for its visualization, structural analysis capabilities, and its group detection functionality, the operationally actionable outputs of CrimeNet Explorer has not been proved on real-time police investigations.

3.2 Terrorist Modus Operandi Detection System (TMODS)

TMODS, which is developed by 21st Century Technologies (Marcus et al., 2007), automates the tasks of searching for and analysing 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. TMODS is mature and powerful distributed java software that has been under development since October 2001 (Marcus et al., 2007). But it needs a pattern graph and an analyst to run the system. Like a supervised learning algorithm, TMODS tries to tailor the results according to pre-defined threatening activity. Another possible drawback is graphs used in TMODS are multi-mode and can be disadvantageous for further analysis. Multi-mode graph means that nodes in multi-mode graphs are more than two types of entities. A person, a building, an event, a vehicle are all represented as nodes; when for instance we want to detect key players in multi-mode graph, a building can be detected as key player, not a person. This can be a cause of confusion. To overcome this confusion the definition of a one-mode (friendship) social network should be used rather than representing all entities as nodes.

4 Offender Group Representation

Wasserman and Faust (1994) pp.35 states that the modes of a network as the number of sets of entities on which structural variables are measured. One-mode (friendship) networks, the predominate type of network, study just a single set of actors while two-mode (affiliation) networks focus on two sets of actors, or one set of actors and one set of events. One could ever consider (three and higher) mode networks but rarely have social network methods has been designed for such complicated data structures. According to these definitions it is better to represent actors (offenders) as nodes and rest of the relations as edges in one-mode (friendship) social networks. This can produce many link types such as "co-defendant link", "spatial link", "same weapon link", and "same modus operandi link". Thereby many graph theoretical and SNA solutions can be used on one-mode (friendship) networks effectively such as friendship identification, finding key actors.

5 Police Arrest Data

We recommend looking for common characteristics of offenders in police arrest data. Do they commit the same crime somewhere sometime together, and then any of these offenders has also committed another crime with another offender? This information can be obtained from a relational database table, text-based arrest report, or CCTV footage.

In *Operation Cash* we obtained this information from Bursa Police Arrest Data where the table included the fields for: P_ID (person id), C_ID (crime reference number), BRANCH (police branch that deals with), CRT_ID (Crime type it belongs to), CR (Name of the offence), MOT_ID (Modus Operandi it belongs to), MO (name of the modus operandi), D (date stamp), DIS (district), NG (neighbourhood), and NG_ID (neighbourhood number).

6 Offender Group Detection Model (OGDM)

OGDM is mainly developed for detecting gangs and theft networks. As exhibited in Figure 1. the source of link information is gathered from police arrest records where a link table; consisting of From (From Offender), To (To offender), and W (how many times this offender pair caught together by the police) is produced with an inner join SQL query.

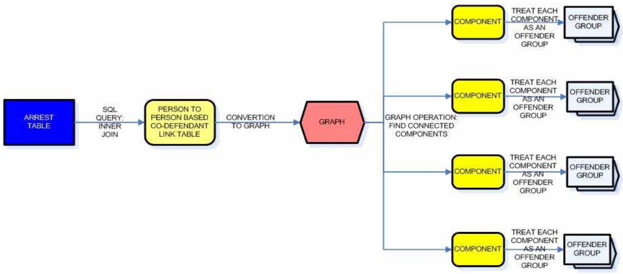
ARREST TABLE
SQL QUERY: INNER JOIN
PERSON TO PERSON BASED CO-DEFENDANT LINK TABLE
CONVERTION TO GRAPH
GRAPH
GRAPH OPERATION: FIND CONNECTED COMPONENTS
COMPONENT
TREAT EACH COMPONENT AS AN OFFENDER GROUP
OFFENDER GROUP
COMPONENT
TREAT EACH COMPONENT AS AN OFFENDER GROUP
OFFENDER GROUP
COMPONENT
TREAT EACH COMPONENT AS AN OFFENDER GROUP
OFFENDER GROUP
COMPONENT
TREAT EACH COMPONENT AS AN OFFENDER GROUP
OFFENDER GROUP

Figure 1

Inner join query result, which we call co-defendant link table, then converted to graph where nodes represent offenders, edges represent crimes committed together using offender group representation exhibited in section 4. Number of times caught together is counted to be used for edge weight (W). At this point a subgraph detection operation is needed; various social network analysis algorithms such as k-clique, k-core (Wasserman et al., 1994) can be used for this purpose. We used strongly connected components (SCC) algorithm in *Operation Cash* because it is scalable and gives concrete results. SCC algorithm is defined as (Cormen et al., 2001); a directed graph is called strongly connected if for every pair of vertices U and V in a graph there is a path from U to V and a path from V to U. The strongly connected components of a directed graph are its maximal strongly connected subgraphs.

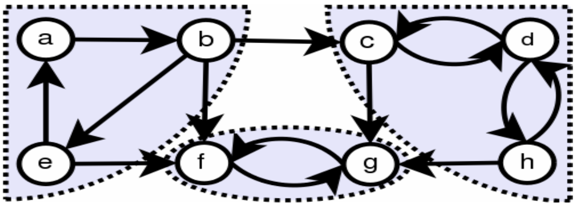

Figure 2. This figure shows graph with its strongly connected components are marked

In a graph generated from an arrest table where there are at least couple of hundred thousand of crimes (edges) and thousands of offender (nodes) makes scalability and performance issue very important. At last, every component represents a unique offender group because one offender can only belong to one group thereby concrete a result of group membership is obtained.

7 Filtering for Legal Requirements

Turkish Crime Code requires that an criminal organisation (offender group) must consist at least of three members, and two members in an offender group must have been convicted together for committing the same crime at least two times (Turkish Crime Code, Article Number:261). According to this definition, where edge weight is W and number of members is N;

$$W_{group} >= 2, N_{group} >= 3$$

is the threshold to constitute a criminal organisation. This requirement can be different in different countries but it is essential to create a filter for a legally accepted criminal organisation.

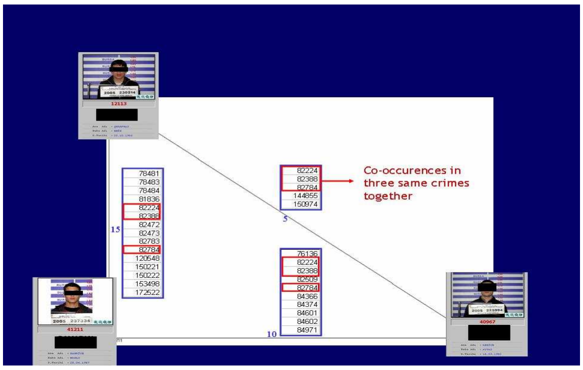

Figure 3. This triad of thieves committed various crimes together. The person in the top left has committed 15 crimes together (W=15) with the person in the left bottom and 5 crimes together (W=5) with the person in the right bottom. The person on the left bottom has also committed 10 crimes together (W=10) with the person on the right bottom. Overall, these three persons have committed 3 crimes together as a group which is shown in crime reference numbers 82224, 82388, and 80784 highlighted in red boxes.

8 Operation Cash

Offender group detection action is started with preparation of Bursa Police arrest data. Initial data preprocessing and data cleaning are done in cooperation with Bursa Police Department on more than 300000 crimes and 6000 offenders. Starting from 1994 to 2007, arrest data included all offenders with a unique person-id number. This uniqueness allowed us to track all offenders' activities. We had opportunity to find out an offender's history over time with all his/her crimes had committed. We produced first the link table, and then converted it to a massive graph; at the end all components in the graph are obtained with SCC. Accepting that even two offenders caught by the police is enough to be a component, total number of components were 33004 (199728 crimes; with an average of 6.05 crimes per component). When W_{group} threshold is put to 2, number of components is dropped to 4488 (15482 crimes; with an average of 3.45 crimes per group). When N_{group} threshold is put to 3, number of offender groups, which is adherent to Turkish Criminal Law definition, is dropped to 1416. Reminding the fact that these groups included many offenders committed various types of crimes from theft to violence, from gangs to terrorists; we only focused on active theft groups who committed crimes in the last five years. As a result, 63 theft groups are detected and these findings were introduced to the police experts for further examination. According to police experts, our findings were very valuable but not enough. There was a consensus to search group members, gather enough evidence for arrest and prepare the case for a sentence. Besides, in parallel, the effectiveness of our method was also a question for the police so just one random theft group out of 63 is focused, a judge verdict is obtained for electronic surveillance and telephone conversations of all members of selected group are eavesdropped for ten weeks. Our findings for this theft group are exhibited in figure 4 as offenders by person-id numbers, and with degrees of members in brackets. Degree is a metric in social network analysis which is count of incoming and outgoing links for an actor (Wasserman et al., 1994). High degree value for an actor suggests that actor is likely to be a key player in the network.

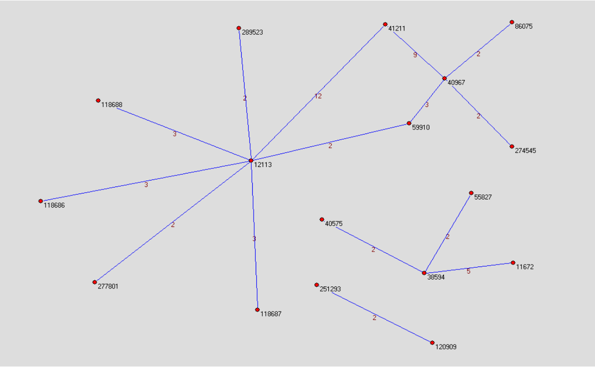

Figure 4. As filtering is applied in order to meet Turkish Criminal Code requirement, the theft group, consisting of 17 offenders with degrees in brackets:12113**(54)**, 41211**(42)**,40967**(32)**, 38594**(18)**, 11672**(10)**,59910**(10)**, 118686**(6)**,118687**(6)**,118688**(6)**,40575**(4),**55827**(4)**,86075**(4)**, 120909**(4)**,251293**(4)**,274545**(4)**,277801**(4)**, 289523**(4)**

After this electronic surveillance, verification of who is who in the network and gathering enough convincing evidence, *Operation Cash* is launched. The police arrested 17 people, recovered worth US $ 200000 stolen jewelleries, PCs, laptops, mobile phones, and some cash worth US $ 180000.

Obtained evidences and interrogations showed that ruling members were detected using OGDM. It has been proved that the real network was consisting of 21 members and 3 of them (AB, MRK, and SE) have never been arrested by the police so their names were not available in the database. We managed to get only 4 ruling members (12113, 38594, 41211, and 277801). Four leaders were basically the chief of gun-jewellers thieves (12113), the skilled expert thief specialized in electronic goods (277801), chief of electronic goods thieves (38594), chief of car and gadget supplier for the network (41211). Interestingly, "big brother" of the network (220868) has only two records in police database. His leader position is identified after interrogations and cross examination of members' statements.

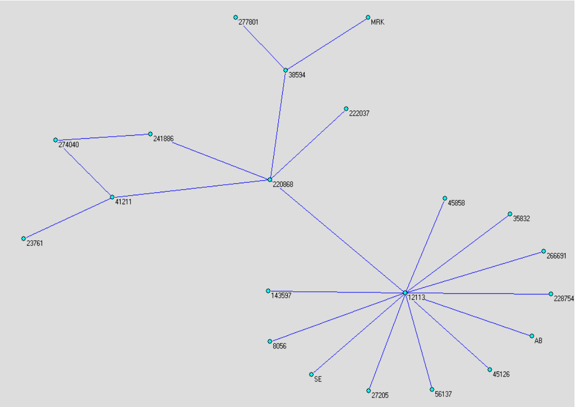

Figure 5. Theft network after verification of evidences. 12113**(24)**,38594**(6)**,41211**(6)**,241886**(4)**,274040**(4)**, 8056**(2)**, 23761**(2)**, 27205**(2)**, 35832**(2)**,45126**(2)**,45858**(2)**,56137**(2)**, 143597**(2)**,220868**(2)**, 222037**(2)**, 228754**(2)**, 266691**(2)**, 277801**(2)**, AB**(2)**, MRK**(2)**, SE**(2)**

Operation Cash has attracted wide attention and positive feedback in local and national newspapers (Zaman, Olay, PolisHaber, 2006). The police commissioner of Bursa city stated that *Operation Cash* was the most successful operation among all operations by Bursa Police in 2006.

9 Conclusion

It has been shown that co-defendant information in police arrest data is beneficial for the police to detect ruling members of offender groups. It has been also shown that detecting an underlying criminal network is possible with link mining and group detection techniques.

OGDM has been successful for partly detection of offender groups. But it is clear that domain expertise is still needed for complete detection of groups. This shows the necessity of semi-supervised models for OGD.

The result achieved depends on the details of the OGDM come from offender group representation success (see section 5). By representing actors as nodes and rest of the relations as edges in one-mode (friendship) social networks can produce many link types such as "co-defendant link", "spatial link", "same weapon link", "same modus operandi link". This helped many graph theoretical and SNA solutions can be used in *Operation Cash*.

Additional criminological conclusions reached after discussions with domain experts are;

- Group members likely to come from the same family (e.g. small-aged pickpocketing group).
- Group members likely to cooperate and come together for required skills to commit crimes.(e.g. theft from offices group, theft from residences group, fraud group, violence group).
- Group members are high likely coming from the same age group and peer group.
- Group members' origins are high likely coming from the same home cities and towns.
- Group members are likely to live in the same areas.
- Group members are likely to operate in the same areas.
- Group members are likely to work in the same industries(e.g. Scrap Dealer Auto theft Group).

References

Adderley, R. (2004), The use of data mining techniques in operational crime fighting *in* 2nd Symposium on Intelligence and Security Informatics, ISI‐ 2004. Tucson, AZ, USA. 3073, pp. 418–425.

Adibi, J., P. Pantel, et al. (2005), 'Report Link Discovery: Issues, Approaches and Applications', (KDD-2005 Workshop - LinkKDD-2005), *SIGKDD Explorations* **7**(2), pp. 123-125.

Adibi, J. & Chalupsky, H. (2004), The KOJAK Group Finder: Connecting the dots via integrated knowledge based and statistical reasoning, *in* IAAI.

Analyst Notebook (2007), 'i2 Analyst Notebook', i2 Ltd, <http://www.i2.co.uk/> Viewed at 31 July 2007.

Chen, H., Chung, W., et al. (2004), Crime data mining: a general framework and some examples, *in* Computer **37**(4), pp. 50-56.

Chen, H., J. Schroeder, et al. (2002), 'COPLINK Connect: information and knowledge management for law Enforcement', *Decision Support Systems* **34**, pp. 271-285.

Cook, D.J. & Holder, L.B. (2000), 'Graph-Based data mining', *IEEE Intelligent Systems* **15**(2), pp. 32-41

Cook, D.J. & Holder, L.B. (2007), *Graph Mining*, Wiley-Interscience, John Wiley Sons, Hoboken, New Jersey.

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (2001), *Introduction to Algorithms*. Second Edition. MIT Press and McGraw-Hill

Getoor, L. & Diehl, C.P. (2005), 'Link Mining: A Survey', *SIGKDD Explorations* **7**(2), pp. 3–12

Getoor, L. et al. (2004), 'Link Mining: a new data mining challenge', *SIGKDD Explorations* **5**(1), pp. 84-89.

Guest, S. D., Moody, J., Kelly, L., Rulison, K.L., (2007), 'Density or Distinction? The Roles of Data Structure and Group Detection Methods in Describing Adolescent Peer Groups', *Journal of Social Structure*, **8**(1), Viewed at 28 July 2007,< http://www.cmu.edu/joss/content/articles/volindex.html >.

Kubica, J., Moore, A., et al. (2003), cGraph: A fast graph-based method for link analysis and queries, *in* IJCAI 2003 Text Mining and Link Analysis Workshop.

Kubica, J., Moore, A., et al. (2002), Stochastic Link and Group Detection, *in* 18th National Conference on Artificial Intelligence, AAAI Press/ MIT Press

Marcus, S.M., Moy, M. & Coffman, T. (2007), Social Network Analysis, *in* Diane J.Cook and Lawrence B. Holder, 'Mining Graph Data', John Wiley & Sons.

Moy, M. (2005), 'Using TMODS to run the best friends group detection algorithm', 21st Century Technologies Internal Publication.

Olay (2006), 'Technological tracking to criminal groups', Bursa Olay Local Newspaper, 19th of December 2006 , Viewed at 31 July 2007, <http://www2.olay.com.tr/blocks/haberoku.php?id=5990&cins=Spot%20Bursa>.

PolisHaber (2006), 'Operation 'Cash' By Police', Turkish Police News Portal, Viewed at 31 July 2007, <http://www.polis.web.tr/article_view.php?aid=3666>.

Scott, J. (2004), *Social Network Analysis: A Handbook*, SAGE Publications, London, UK.

Senator, T.E. (2005), 'Link Mining Applications: Progress and Challenges', *SIGKDD Explorations*, **7**(2), pp. 76–83.

Sentient (2007), 'Sentient Data Detective', Sentient Information Systems, Viewed at 31 July 2007, <http://www.www.sentient.nl/>.

Taipale, K. A. (2003), 'Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data', Columbia Science and Technology Law Review 5.

Wasserman, S. & Faust, K. (1994), *Social Network Analysis Methods and Applications.* Structural Analysis in the Social Sciences, Cambridge University Press.

Xu, J. J. & Chen, H. (2005), 'CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery', *ACM Transactions on Information Systems* **23**(2), pp. 201-226.

Zaman (2006), 'Police tracked down 63 crime groups with new technology help', Zaman National Newspaper, 9th of January 2007, Viewed at 31 July 2007, <http://www.zaman.com.tr/webapp-tr/haber.do?haberno=437444>.