Criminal Network Analysis with Interactive Strategies: A Proof of Concept Study using Mobile Call Logs

Peng ZHOU*, Yan LIU†, Mengjia ZHAO‡
School of Software Engineering, Tongji University
Shanghai, China
Email: *1435855@tongji.edu.cn, †yanliu.sse@tongji.edu.cn, ‡1434319@tongji.edu.cn

***Abstract*—The communication data are becoming increasingly important for criminal network analysis nowadays, this data provides a digital trace which can be regarded as a hidden clue to support the crack of criminal cases. Additionally, performing a timely and effective analysis on it can predict criminal intents and take efficient actions to restrain and prevent crimes. The primary work of our research is to suggest an analytical process with interactive strategies as a solution to the problem of characterizing criminal groups constructed from the communication data. It is expected to assist law enforcement agencies in the task of discovering the potential suspects and exploring the underlying structures of criminal network hidden behind the communication data. This process allows for network analysis with commonly used metrics to identify the core members. It permits exploration and visualization of the network in the goal of improving the comprehension of interesting microstructures. Most importantly, it also allows to extract community structures in a appropriate level with the label supervision strategy. Our work concludes illustrating the application of our interactive strategies to a real world criminal investigation with mobile call logs.**

Keywords-criminal network analysis; interactive strategy; network measure; community detection;

I. INTRODUCTION

Communication data are widely used in criminal network analysis to understand direct relationships and identify implicit connections. Many efforts have been devoted to leverage those data in criminal community detection, connection strength evaluation, microstructure discovery, and suspect identification. However, the skill set required for criminal network analysis (CNA) is complex and diverse, such as the application of empirical study and domain knowledge into the data preprocessing, criminal investigation knowledge, intelligence analysis experience, network visualization layout technique in different network scale, and social network analysis knowledge, which leads mismatching of expectation and reality on the power of data analytics.

We proposed an interactive analysis process by observing practical workflows which involving detectives, intelligence officers, data engineers, data scientist, and domain experts, combined with the technique of social network analysis and machine learning. This process is divided into three phases according to important results in each phase, namely,*i) network construction*: the generation of network structure, *ii)metric design*: the core nodes and relations, *iii) structure observation*: structures extraction in different levels. In each phase, the process adopted interactive strategies in order to formulate and assess hypotheses in a rapid, iterative manner—thus supporting exploration Criminal Networks (CN) with the pace of human thought.

DOI reference number: 10.18293/SEKE2016-119

The interactive analysis process is our chosen way to explore the CN with a case study using the mobile call log. In the construction phase, we build the CN by data preprocessing, data cleaning and data extraction, which contains 7840 nodes and 103043 links, and then three layout are introduced to configure and depict the CN. Metric analysis phase illustrated the adoption of some classic metrics in social network analysis seeking to identify the key group members, and the evaluation of these metrics with corresponding result. Finally, in the structure analysis phase, Fast Unfolding is used as our detection algorithm to analyse CN interactively and rapidly. Additionally, we proposed a strategy — label supervision to control the level of the structure extracted.

Our contributions contain: 1) suggesting a generical analytical process for Criminal Network Analysis. It can be divided into three phases. 2) proposing interactive strategies to the analytical process, such as using various visualization layouts to configure the network, reinterpreting different measures from the social network domain to the criminal network domain, and controlling the community structure level with label supervision strategy. 3) conducting a proof of concept study using mobile call logs.

The paper is organized as follows. Section II describes related works about CNA. Section III illustrates how we proposed the interactive analytical process and introduced three visualization layouts for CN. In the following, Section IV, section V and section VI demonstrates the three phases separately with an example of Criminal Network being constructed, analyzed with metrics and explored with detection algorithm. The final section, section IIV, concludes the future work and a conclusion according to our work.

II. RELATED WORK

In this section, an overview of previous works in the domain of criminal network analysis has been firstly provided, and

then comes to the concept of community structure and structure detection method. Finally, we present a comprehension on the two relevant works about its advantages and limits.

A. Criminal Networks Analysis

In the last thirty years, many efforts have been used in order to analysis the Criminal Networks in a more intelligent way. One of the most important research in the CNA domain is due to Malcolm Sparrow [18], who summarized four features of the Criminal Network, namely, i) limited dimension — CNs are often composed of few thousands entities; ii) incomplete information — criminal networks are unavoidably incomplete and erroneous; iii) undefined border — it is difficult to pick out all the relations of each entity; and, iv) dynamism — new relations always indicate constant evolution of network structure.

Fortunately, this contribution lead to a trend that researchers tried to analyze Criminal Networks with the techniques in the domain of the social network analysis. For instance, Baker and Faulkner [3] studied illegal networks in the field of electric plants and Klerks [11] concentrated on criminal organizations in the Netherlands. Silke [17] and Brannan et al. [5] acknowledged a slow growth in the terrorism network and examined state of the art in criminal network analysis. Arquilla and Ronfeldt [2] summarized previous researches and proposed the concept of Netwar with its application to terrorism.

However, in 2006, a popular work by Valdis Krebs [12] applied network analysis in conjunction with network visualization theory to analyze the 2001-09-11 terrorist attacks. This work represents a starting point of a series of academic papers in which social network analysis methods become applied to a real-world case, differently from previous work where mostly toy models and fictitious networks were used. Krebs paper inspired further research in network analysis for the design of better SNA applications to support intelligence agencies in the fight against terror, and law enforcement agencies in their quest fighting crime.

B. Community Structure and Community Detection

Community structure is one of the most common characteristics in the study of networks [15], which refers to the occurrence of groups of nodes in a network that are more densely connected internally than with the rest of the network [9]. There is one widely adopted concept to investigate the quality of this structures in the network called network modularity, which can be expressed as follows: let consider a network, represented by $G = (V, E)$, which has been partitioned into m communities; its corresponding value Q of network modularity is defined as:

$$Q = \sum_{s=1}^{m} \left[\frac{l_s}{|E|} - (\frac{d_s}{2|E|})^2 \right] \quad (1)$$

assuming l_s the number of links between nodes belonging to the s_{th} community and d_s is the sum of the degrees of the nodes in the s_{th} community. High values of Q indicate high values of l_s for each discovered community, yielding to communities internally densely connected and weakly coupled among each other. The process of detecting community structures in a network called community detection, and it still is regarded as a computationally difficult task. However, several methods for community detection have been proposed and applied with varying extents of success. Such as minimum-cut method [13], hierarchical clustering [1], Girvan-Newman algorithm [14], modularity maximization and clique based methods [6].

Lots of works shown that community detection is proved as a powerful tool to analyze the structure in the criminal networks. Emilio et al. [7] employed Girvan-Newman algorithm and a variant based on modularity optimization called Newmans algorithm to detect and explore the community structures in the CNs reconstructed from phone call logs. Hamed Sarvari et al. [16] performed a large scale analysis with clique based methods to find patterns and substructures of that network based on a publicly leaked set of customer email addresses. However, these studies and early researches somehow neglected the importance of network visualization and the interactions during the analysis process, laying emphasis on aspects related more to statical network characterization. In our research, we stress these twofold by introducing various visualization layouts and adopting crucial interactive points.

III. Interactive Strategies

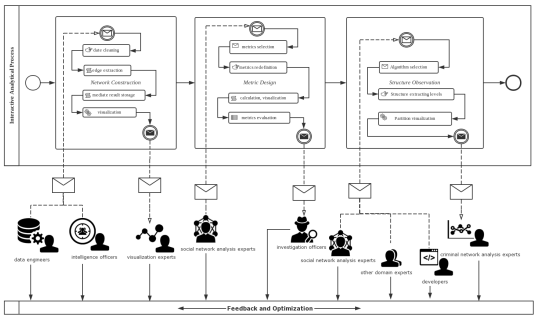


Figure 1. The interactive analytical process

At present, the task of analysing Criminal Networks can not be finished purely relying on the computational analysis or manual analysis. We examined all the analytical process in these works and we found that almost all the processes about CNA are interactive process between humans and computers, but the interactive strategies are weak, limited and even not involved obviously and deeply. In order to gain better understanding of the process with some strong interactive strategies during the CNA, we followed 5 investigations based on a real world criminal cases and also observed a general pattern of work shared by most police officers and intelligence analysts.

A. Interactive Analysis Process

After doing induction and summary on the workflow, we concluded that the process of CNA is a human-interaction process, and proposed an interactive analytical process for CNA.

As is shown on the Figure 1. the process can be mainly divided in to three phases according to each phase's promising results, namely, *i) network construction*: the generation of network structure, *ii) metric design*: the core nodes and relations, *iii) structure observation*: the extraction of organization structure. And this process is supposed to assist investigator to analysis Criminal Network in an interactive way.

In the first phase, we clean data interactively depending on the empirical rules suggested by intelligence officers and data engineers, extract the edge table from those data, provide a format of intermediate result storage and finally perform network visualization based on the layouts selected by visualization experts. In the second phase, some metrics are introduced by SNA experts from the domain of SNA into the domain of CNA. And there is also need to reinterpret them before employing in the context of CN, after which we might be able to obtain the ranking results. Additionally, we perform various configurations on network with each entitys corresponding metric value. At the end of this phase, a certain evaluation is defined to suit for our CN to examine the quality of different metrics in consultation with investigation officers. The crucial point related to interaction in this phase is twofold, one is metrics selection which can measure different centrality and influence of each node, another is that we need to configure many kinds of layouts in consideration of the scale of the CN. The following phase is structure analysis of the CN. First, we need to choose a detection algorithm, taking the features of the CNs constructed during the phase one into consideration. For example, different scale might need different methods, different complexity likewise need different method. So we might need the support of SNA experts, other domain experts and programming developers. Then ascertain the extracting structure level to amplify or narrow down graph interests. Finally, a partition visualization results is produced enabling to examine the quality of the structure and even find out the hidden knowledge. Note that the three phase is an integrated whole. The previous stage outputs is regarded as the next phase of the input, so each phase results would interfere with the next phase results. Consequently, we suggest another stage — feedback and optimization to ensure the effective communication between various domain experts and optimization of the whole interactive process.

B. Data Set

For better illustrating the following works, here we present a brief introduction of the dataset. The data source supporting the case study is based on an investigation in the fight against a local burglary criminal groups, and it is by no means completed considering the limited collection techniques and criminals' deliberate disguise.

A brief description of the data set is shown in Table I. With a preliminarily discovery on the dataset, we find that the dataset contains 73 detailed phone call tickets, among them containing 25 known suspect and 48 persons involved with this investigation case. All phone call tickets are all in the period from 2014.6 to 2015.6 summing up all the phone call logs from all the call tickets, we got totally 1016833 logs.

TABLE I. Date set summary

Detailed ticket	Known suspect profile	Time period	Total call logs
73	25	2014.1 2015.6	1016833

IV. Network Construction

In order to assure the quality of the network construction, we proposed a generical interactive workflow to support this phase task. This workflow shows how the phone call logs are preprocessed, transformed, extracted and constructed into a network.

A. Data Preprocessing and Network Construction

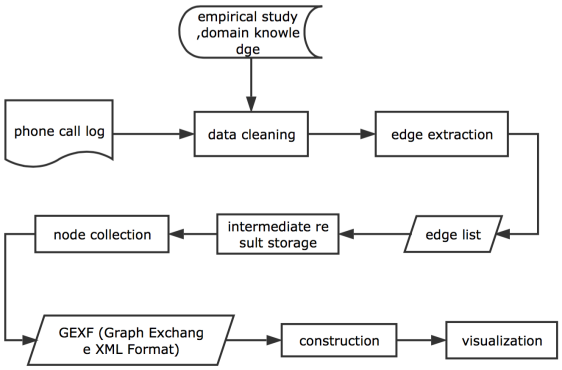


Figure 2. The data preprocessing workflow

Performing a good preprocessing on the dataset can get rid of the noise as much as possible, and it permits the network structure to be constructed easily without cost of quality. It allows the network metrics to depict a member more exactly and make visualization closer to the nature of the criminal communication network. Additionally, it likewise can reduce the following analytical errors and the workload of the structure analysis phase. Owing to the great importance of data preprocessing, we think highly of this stage.

The data preprocessing stage was shown on the Figure 2. The data cleaning comes firstly to our work, the empirical study and domain knowledge that we used as follows:

a) eliminate the call logs containing bank notifications, communication providers and other public service providers. For instance, the number 10086 is a communications service providers in China, and 95595 is the call number of Agricultural Bank of China. *b)* remove redundant logs produced by peer to peer calling. When one entity calls another entity, both of their phone detailed tickets will generate almost the same log with a certain time difference and different call types that one is calling and another is called. *c)* wipe off the call logs whose call duration is zero. This type of logs indicate that is not a successful connection and should not be considered in the following works.

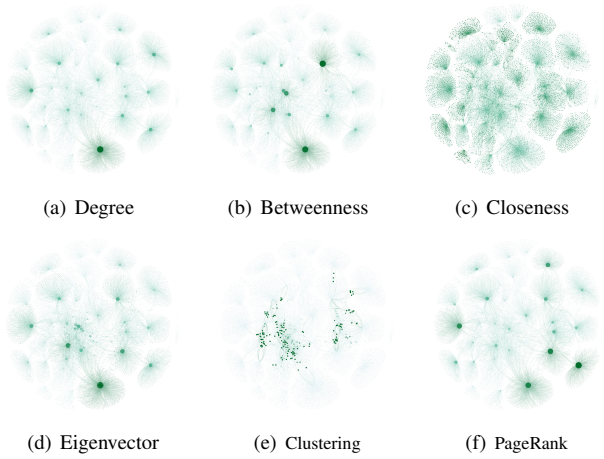
After data cleaning, the workflow extracts the data with few noise to build the edge list of the network. In this step we need

to change call type according to the corresponding call type in order to identify the source nodes and the target nodes. When finishing edge extracting, our work chooses a intermediate result storage. Then we write a python script to collect the nodes to build the node set and importing the nodes set and edge list by the python-igraph package and constructing our CNs, finally we export GEXF file in order to visualize by a software called Gephi. At length we construct a burglary criminal network shown on the figure 3 (a), which has totally 7840 nodes and 1016833 links.

V. Metric Design

In order to exactly capture the influence of core members and its roles playing in the criminal organizations, we introduce a series of measures from the domain of Social Network Analysis and interpret them in the context of our certain CN. After calculating these metrics, we render the network view with different color intensity and size based upon the metric value of each entity and spotlight certain nodes.

A. Network Measures Reinterpretation in CNs



(a) Degree (b) Betweenness (c) Closeness

(d) Eigenvector (e) Clustering (f) PageRank

Figure 4. Network metrics

Degree centrality measures the activity and influence by calculating the number of direct relations a node has. It is convinced that a call number with a high degree could be regarded as a hub which act as an import information channel in the communication network and a node with high activity. It is defined as:

$$Dgr_i = \frac{d_i}{S-1} = \frac{\sum_{i \in M} m_{ij}}{S-1} \tag{2}$$

where d_i is the directly links to other call numbers of a call number i, and m_{ij} is the ij_{th} element of the adjacency matrix M and S is the sum of the whole call numbers. $S-1$ is the normalization factor.

Betweenness centrality is another indicator to measure nodes centrality which equals the number of the shortest paths from all nodes to all other nodes which pass through that node. In the context of CN, a call number with high betweenness can be regarded as broker of messages, indicating the great importance of this call number in the information transfer. Most importantly, this kind of call numbers usually connects two or more densely call number clusters, removal of them might lead to the destabilization, abruption and even destroy of the CNs. The betweenness centrality of a call number (i) is redefined as:

$$Btw_i = \frac{\sum_{j<k \in N} \frac{p_{jk}(i)}{p_{jk}}}{(S-1)(S-2)} \tag{3}$$

where p_{jk} is the total number of the shortest paths from call number j to call number k and the $p_{jk}(i)$ is the number of those paths that pass through call number i, the $(S-1)(S-2)$ is the normalization factor.

Closeness centrality is the reciprocal of the sum of the shortest paths between that node and all other nodes in the network. In the analysis of our CN, a high closeness indicates this call number can reach majority of the other nodes easier and faster. The closeness centrality of a call number (i) is redefined by the expression:

$$Clsn_i = (H_i)^{-1} = \frac{S-1}{\sum_{j \in N} d_(i,j)} \tag{4}$$

where $d_(i,j)$ is the distance between node i with node j, H_i is the normalized distance.

Eigenvector centrality defines another centrality with the consideration of the importance of their neighbours. In the framework of our CN, a call number with high eigenvector centrality means that this call number can reach a group of other call numbers easily and quickly. The eigenvector centrality of a node (i) is redefined as:

$$Eign_i = \frac{1}{\alpha} \sum_{j \in L(i)} m_j = \frac{1}{\alpha} \sum_{j \in N} a_{ij} x_j \tag{5}$$

where $L(i)$ is the direct link set of call number i, α is a constant, and a_{ij} is the ij_{th} element of the adjacency matrix M.

Clustering coefficient is a measure of the degree to the aggregation of nodes in a graph. In the context of our CN, a call number with high clustering coefficient means a high likelihood of that the direct links of the given call number can reach each other. It is redefined as:

$$Clst_i = \frac{|e_{jk}|}{l_i(l_i-1)} \tag{6}$$

where e_{jk} is the link existing in the neighbours of call number i and l_i is the number of direct links of the call number i.

PageRank is a variant of the eigenvector centrality measure in some degree. Our work employ this measure to the importance of a certain call number globally. It is redefined as:

$$Page(i) = (1-f) + f * \sum_{j \in H(i)} \frac{P_i(j)}{L_j} \tag{7}$$

TABLE II. Different network metrics of the known suspect

Dgr	Btw	Clsn	Eigvt	Clst	PgRk
9699*	1002*	7071	5636*	5117	7777*
0691*	9699*	1083	8662*	3229	9699*
1509*	5636*	8890	0691*	4903	8662*
7777*	2406*	5703	1002*	5064	0691*
8662*	8662*	4848	7777*	7037	7789*
8199*	4751*	7515	5676*	9868	1002*
1002*	8199*	8986	8199*	1115	8199*
2223*	6526*	9549	9699*	7777*	9250*
1359*	7777*	5362	2406*	8318	3605*
5636*	0691*	4097	2223*	5782	5636*
9250*	1509*	9891	0399*	9699*	2223*
7789*	2223*	8582	4076*	6409	1509*
3605*	7970*	3401	9250*	0691*	8391*
8391*	8391*	0206	7789*	5435	1359*
2406*	8318	1223	4494	8662*	2406*
5676*	1359*	2601	8391*	5347	6526*
7970*	5676*	6179	1509*	7071	7970*
7574*	4494	0393	1359*	8199*	5676*
6526*	7789*	3012	3605*	1083	7574*
3157*	3605*	0722	6859	8890	7603*
4076*	7588	5894	6526*	7789*	4076*
7692	9250*	3040	8033	1002*	0399*
2647*	3157*	5049	4751*	9250*	3157*
0399*	4076*	2602	2397	3605*	2647*
4751*	7574*	4892	1050	5636*	7692

where $H(i)$ are the set of call numbers directly linking to the call number i , L_j is the number of outgoing links in j and f is the damping factor.

B. Metrics Visualization and Ranking Interpretation

Our visualization results with the combination view of Fruchterman Reingold layout and Fisheye layout was shown on the Figure 4, and the color intensity and the size of the nodes both are rendered by their corresponding metric value. To check the usability of different metrics, we defined the hit rate of the known burglary suspect in the top 25 and top 100 rank as:

$$h = \frac{n}{K} * 100\% \tag{8}$$

where n is the number of known suspect in the top K ranking of the corresponding metric. The hit rate results were shown on the Table III.

To begin with, we can found that the degree, betweenness, eigenvector and PageRank have a similar and excellent layout view. After checking the hit rate of these metrics, it indicated that 96% of suspect belongs to the top 25 of the degree, 88% to betweenness, 84% to eigenvector, 96% to PageRank. Therefore in the top 100 ranking, they consequently got 100%. These four metrics were proved a valid measure for burglary group CN. But the remaining two metrics were not good enough and even no one suspect was ranked in the top 25 of closeness metric. Although the hit rate of clustering coefficient only got 60% in the top 25. After we performed a manual check on the top 100, it works well. But for closeness centrality, it still got 0% in the top 100. The most likely factor resulting in this phenomenon is the way of data collection.

In summary, the metrics selection should take the way of the data collecting into consideration ensuring that the effective

TABLE III. The hit rate of top-K ranking

Rank	Dgr	Btw	Clsn	Eigvt	Clst	PgRk
Top 25	96%	88%	0%	84%	60%	96%
Top 100	100%	100%	0%	100%	100%	100%

metrics for measuring the CNs would be adopted. In the context of burglary group phone CN, these four indicators which are degree, betweenness, eigenvector, and PageRank, are good enough to analysis the network in an effective way. The clustering coefficient might work well if we broaden the ranking scope.

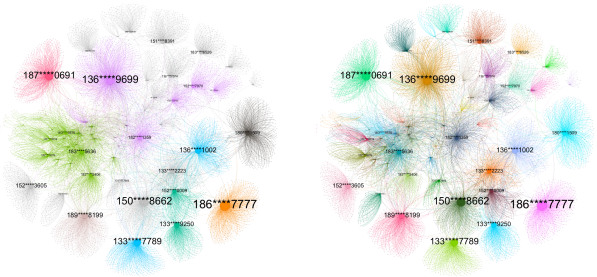
VI. Structure Observation

After the metric design phase, this section is focusing on the structure analysis of the CN and is expected to select a suitable algorithm to finish the task of community extraction rapidly and interactively.

A. Fast Unfolding with Label Supervision Strategy

In the consideration of the scale of our burglary CN, which contains 7840 nodes and 103043 links, and the demands for interactive point mentioned in the section III, Fast Unfolding [4] has been adopted to detect the communities structure in burglary CN. Besides, we provide a label supervision strategy to put priori knowledge and evidence into the structure extraction. Thus, the level of the community structure can be well controlled. Finally, we visualize our community results with combination of FR layout and fisheye layout. Besides the extremely high speed, another reason why we choose Fast Unfolding algorithm is that this heuristic method provides a parameter of resolution to control the scale of the communities detected. In most situations the default value of the resolution is 1, if the value is lower, then the size of the communities detected would become smaller, which also means that it will produce more communities. If we set a higher value than the default, it means bigger communities structure and less communities. Therefore, in this way, our process can gradually increase or decrease the value of resolution in order to extract subgroups with appropriate scales in the burglary CN.

However, it is unknown when we should stop reduction or increasements of the resolution to gain the appropriate structure level. we proposed a general method based on label supervision to solve this question which can be suitable for all the detection algorithm. According to detection algorithm initial stage, it can be divided into forward and reverse strategy. If the detection is start from one community to appropriate mounts of community, it means that the process of the detection is one to more, and the label supervision should inspect the partition of the nodes labeled in the same level, we called this type forward strategy; but if the algorithm is from more to less, the label supervision should inspect the mergence of the nodes in the same level. this type called reverse strategy. The following steps describe the process of the FU with the reverse label supervision strategies: 1) assign each node to a different community, for each node i, consider the neighbours j of i, put node i to its neighbour j when reaching the maximum



(a) r=1 m =0.813 nc = 22 (b) r=0.18 m =0.778 nc =41

Figure 5. The left figure shows a community detection with resolution = 1 and the number of communities is 22.The right part of the figure is the situation with resolution of 0.18

positive gain of modularity 2) check the nodes labeled in the same level whether be assign into the same structure. 3) take the communities found during the step 1 as a nodes and build a new network, then back to step 1.

In the context of the CNA, when we know several entities in a same structure level, then we label this node and examine them after each phase of the Fast Unfolding working on the network till that they are arranged into different communities. and this detection can be seen as an effective analysis.

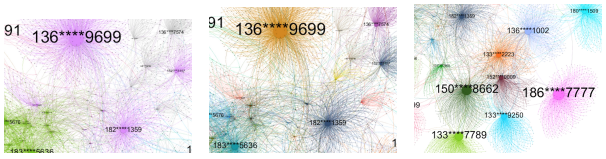
B. Structure Partition Results

The Figure 5 (a) has shown the result after performing FU on the burglary criminal network with the default resolution value, where the modularity value Q is 0.813 and the number of the communities is 22. The network was partitioned into different substructures and each communities, including nodes and edges, were rendered by different colors and the size of the node varies with the community scale. In order to gain a appropriately structural levels, then we marked the known suspects call number 136****9699 and 182*****1359 shown on the Figure 6 (a) and put this knowledge into each iteration of the Fast Unfolding algorithm with a step of 0.1 reduction of the resolution. After 83 iterations, we examined these call numbers were assigned into different communities shown on the Figure 6 (b), where the resolution is 0.18, modularity is 0.778 also a high value and the number of communities is 41. This partition result is shown on the Figure 5 (b).

An another goal of this phase is to find out potential suspects in burglary gang. In the Figure 6 (c), we can find the entity with phone number of 152****0099 which was rendered by brown kept large amounts of relation with top ranking criminal entities, which are 150****8662, 186****7777, 133****2223, 133****9250. Consequently this entity maybe a another member of the burglary group, we recommend such entities to officer for support the following investigation.

VII. Conclusion

In this paper, we introduce an interactive analytical process to explore the CNA with a case study using mobile phone logs. This process generally works well with the CN in our case study. In detail, the core members of the burglary group usually



(a) Label supervision process-1 (b) Label supervision process-2 (c) The protential suspect

Figure 6. The sample of label supervision method

get a high ranking in network metrics, but not all measures are effective for our analysis, such as closeness centrality and clustering coefficient, and three visualization layouts helps a lot during the whole process. Most importantly, our framework can extract community structures in a appropriate level with the application of label supervision into the Fast Unfolding algorithm.

References

[1] Alejandro J Alvarez, Carlos E Sanz-Rodríguez, and Juan Luis Cabrera. Weighting dissimilarities to detect communities in networks. *Phil. Trans. R. Soc. A*, 373(2056):20150108, 2015.

[2] John Arquilla and David Ronfeldt. *Networks and netwars: The future of terror, crime, and militancy*. Rand Corporation, 2001.

[3] Wayne E Baker and Robert R Faulkner. The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American sociological review*, pages 837–860, 1993.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[5] David W Brannan, Philip F Esler, and NT Anders Strindberg. Talking to" terrorists": Towards an independent analytical framework for the study of violent substate activism. *Studies in Conflict and Terrorism*, 24(1):3–24, 2001.

[6] Tim S Evans. Clique graphs and overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(12):P12037, 2010.

[7] Emilio Ferrara, Pasquale De Meo, Salvatore Catanese, and Giacomo Fiumara. Detecting criminal organizations in mobile phone networks. *Expert Systems with Applications*, 41(13):5733–5750, 2014.

[8] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[9] Mohammad Hamdaqa, Ladan Tahvildari, Neil LaChapelle, and Brian Campbell. Cultural scene detection using reverse louvain optimization. *Science of Computer Programming*, 95:44–72, 2014.

[10] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6):e98679, 2014.

[11] Peter Klerks. The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? recent developments in the netherlands. *Connections*, 24(3):53–65, 2001.

[12] Valdis E Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.

[13] Mark EJ Newman. Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):321–330, 2004.

[14] Mark EJ Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133, 2004.

[15] Mason A Porter, Jukka-Pekka Onnela, and Peter J Mucha. Communities in networks. *Notices of the AMS*, 56(9):1082–1097, 2009.

[16] Hamed Sarvari, Ehab Abozinadah, Alex Mbaziira, and Damon McCoy. Constructing and analyzing criminal networks. In *Security and Privacy Workshops (SPW), 2014 IEEE*, pages 84–91. IEEE, 2014.

[17] Andrew Silke. The devil you know: Continuing problems with research on terrorism. *Terrorism and Political Violence*, 13(4):1–14, 2001.

[18] Malcolm K Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social networks*, 13(3):251–274, 1991.