DEVELOPMENT OF ENHANCED BAYESIAN MODEL FOR DETECTION OF COVERT MEMBERS IN CRIMINAL NETWORKS USING TELECOMMUNICATION METADATA

BY

ISMAIL Abideen Adekunle
PhD/SEET/2015/830

DEPARTMENT OF TELECOMMUNICATION ENGINEERING
FEDERAL UNIVERSITY OF TECHNOLOGY MINNA

AUGUST, 2021

ABSTRACT

Crime has become a global challenge in recent times. The phenomenon has become a difficult task that military war-fare approach alone can address effectively without intelligence. Criminal intelligence involves gathering data on criminal activities and participants for preparing deplorable strategies and interventions. Social Network Analysis (SNA) offers supportive tools for analysing Organised Criminal Groups (OCGs) and identifying important nodes with conspicuous relationship as its priority. SNA-based techniques arrived at key players in criminal network with nodes that have high SNA metric values. Apart from datasets challenge, SNA is a weak scheme for key players in OCGs because conspicuous links raise susceptibility of vibrant participants while silent key actors are concealed. Also, status of key actors in OCGs are unrelated with SNA metrics. Scatter-graph of vulnerability and strategic positions was devised to mitigate unrelatedness of SNA metrics for detection of key players in Criminal Social Network (CSN). The scheme identifies actors that have both high vulnerability and high strategic position values at the same time. This is synonymous to Influence Maximization (IM) – set of nodes that have high influence. Silent key players or legitimate actors in adversary network still remain unresolved. Missing node concept works towards set of nodes not known initially as part of a social criminal group. It has high affinity for well-connected nodes than marginal nodes. Node discovery scheme unravels latent structure behind key players within CSN. The scheme pinched on multiple sources of data about a criminal group yet legitimate actor are not captured. Inference approach offers probability-based prediction for detecting covert nodes yet only well-connected nodes with conspicuous relationships are still identifiable. The development of Enhanced Bayesian Model aimed at predicting key players like financial aiders and ammunition suppliers with evasive attitudes. It was conceived towards inherent problem of erratic behaviour and structural equivalence abating key-players from theoretical graph-based. Bayesian model and Recursive Bayesian Filter (RBF) algorithm were combined to have Enhanced Bayesian Network Model (EnBNM) with RBF to lower error rate and improve prediction. EnBNM scheme re-ranks participant's attribute by assigning inference to nodes base on conditional probability of Bayesian model. EnBNM's algorithm was validated using ground truth and SNA-Q model adopted for classifying Criminal Profile Status (CPS). EnBNM was tested using dataset of participants in November 17 Greece revolutionary group - (N'17) and data of participants in September 11 Al-Qaeda terrorist group - (9/11). For N'17 dataset, EBNM detected all alleged and convicted leaders. Additional two actors were detected who had the same CPS with convicted leaders. EnBNM also detected marginal actors; participants with high tendency to evasion. Out of four (4) detected fugitives, two of them belong to the first-generation leadership (G) faction. For 9/11: nine (9) out of nineteen (19) central participants detected by EBNM have the same CPS with convicted leaders. It means that seven (7) more actors are detected as additional key players by EnBNM that previous models did not detect. Six of these actors detected are conspirators. A financial aider to the group was detected among fugitives. The results corroborate that terrorist organisations are self-organised with decentralised key players as a measure to minimize effect of security perturbation. The simulation results showed that the court judgement of the N'17 group was 40% in error as additional two actors were detected by EBNM apart from the three convicted leaders by court. It shows that support of intelligence is highly needed for effective disruption of OCG and terrorism. The EnBNM algorithm also detected over 80% of legitimate actors - less vulnerable participants in the 9/11 terrorist group and has 59.09% accuracy score in detection of conspirators.

TABLE OF CONTENTS

CHAPTER TWO

CHAPTER THREE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BNM - Bayesian Network Model

CCTV - Close Circuit Television

CDR - Call Detail Records

CMR - Call Management Records

CNA - Criminal Network Analysis

CPS - Criminal Profile Status

DOR - Diagnostic Odd Ratio

FDR - False Discovery Rate

FNR - False Negative Rate

FOS - Flat Organisational structure

FPR - False Positive Rate

HOS - Hierarchical Organisational Structure

IM - Influence Maximization

KPP - Key player problem

KPP-POS - Key player problem positive

KPP-NEG - Key player problem negative

LR- - Negative Likelihood ratio

LR+ - Positive likelihood ratio

MAP - Maximum-a-posteriori

MNO - Mobile Network Operator

NPV - Negative Predictive Value

OCG - Organised crime group

P_d - Probability of detection

P_{fa} - Probability of false alarm

Poi – Point of interest

PPV - Positive Predictive Value

Q1 - Prominent

Q2 - The most prominent

Q3 - Inconspicuous

Q4 - Less prominent

RBF - Recursive Bayes Filter

RG - Random Graph

ROC - Receiver Operating Characteristics

SNA - Social Network Analysis

SNR - Signal to Noise Ratio

SPC - Specificity

TN - Terrorist network

TNR - True Negative Rate

TPR - True Positive Rate

LIST OF SYMBOLS

C_D Degree centrality

C_B Betweenness centrality

C_C Closeness centrality

C_E Eigenvector centrality

G (V, E) Network Graph

H(X) Shannon entropy

H_{co} Connectivity entropy

H_{ce} Centrality entropy

$p(x_i)$ probability mass distribution

P(A|B) – Bayes' theorem or conditional

probability V_{cp} conspirator

V_{nl} network leader

V_{rm} regular nodes

V_{sp} sleeper partner

V_T terrorist nodes

Gamma

λ eigenvalue

Definition of terms

Affiliate	is an important participant that is not a regular member; affiliate could be seen as an external member to a social group
Covert node	referred to hidden node; important node
Criminal network	is a structural representation of social relationship in a criminal organisation
Critical conspirator	a key player with greater influence in criminal network
Evasive node	referred to a node with low susceptibility
Fugitive	an important node that it is difficult to detect by simple tools; it requires extra effort
Influential node	a node that can influence other nodes, it is similar to disease or rumour spreaders, market promoters-influencers
Key player	Is an important node; internally influence decision or activities of a social group
Link	represents an edge that connects two nodes; it depicts those two connected nodes share the same attribute or related. It can be used to denote calls, messages, conversation or a meeting
Maximum-a-posteriori	is a peak value of inference curve in posterior probability distribution
Network leader	a key player that controls network resources; it depicts a well-known criminal.

Node	represents a participant, phone or device in a social network
Participant	is a node in social network; an individual involved in a social group
Recursive Bayes Filter	is iteration loop for calculating posterior probability distribution
Regular member	an active participant in a social group; it could depict an overt node.
Sleeper partner	is an affiliate key player with low participation in criminal network
SNA metric	is a tool for measuring worth of a node: influence capacity, importance, sometimes it is used for classifying nodes into overt and covert nodes
Social network	is a graph showing connections/links between nodes; is the smallest unit of a complex network

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background to the Study

Different parts of the globe have one security challenge or the other in the form of persistent conflicts. Organised crime has become a global phenomenon, represented in a confluence of conflicts from Africa, to the Middle East and the Americas, with distinct linkage response to international terrorism (Interpol, 2018). Consequences of these crimes are on the increase considering the rising number of victims and Internally Displaced People (IDP) (United Nations, 2014; Barnes, 2017). For instance, over 3000 died in the USA September 11, 2001 attacks. Approximately 27,000 lives have been lost to the Boko-Haram insurgency in Nigeria, while money earmarked for containing insecurity are too exorbitant (Tayebi, 2015; Ashby, 2016).

Perpetrators of various heinous crimes are described as dark networks or criminal organisations (Morselli, 2009; Manning, 2010; Brunetto *et al.*, 2016). Dark networks are known to be the interconnection of individuals or Organised Crime Groups (OCGs) (Malm & Bichler, 2011). The interconnections among members of criminal groups had been identified as factors responsible for resilience of criminal organisations (Behzadan, 2016; Salvatore *et al.*, 2016). It has also been observed that conventional military warfare approaches are becoming ineffective for combating criminal organisations without incorporation and support of intelligence about criminal organisations (Malm & Bichler, 2011; Minor, 2012; Gunnell *et al.*, 2016).

Social Network Analysis (SNA) is found supportive to criminal intelligence investigations (Keller, 2015; Jones *et al.*, 2018). It was initially designed as a model for describing various relationships, ties, and transactions among members of organisations

(Le, 2012; Molinero *et al.*, 2018). It has become a method for rendering solutions for complex related systems (Basu, 2014). Some academia advocate for incorporation of SNA into criminal investigation(Sparrow, 1991; Kreb, 2002).This had been intensified in some recent works (Brunetto *et al.*, 2016; Bright *et al.*, 2017; Grassi *et al.*, 2019).

However, some works have identified the inadequacy of SNA. Karthika and Bose (2011) considered SNA as inappropriate data mining techniques for criminal networks. It was stated that SNA only discovers patterns from the known structures and not from the hidden structure like terrorist networks. Kitsak *et al.* (2010) observed that the best spreaders or influencers are not necessary being the most highly connected or central node. Husslage *et al.* (2012) opined that the Flat Organisational Structure (FOS) and leaderless principle imbibed by criminal organisations constituted factors that conceal high-profile criminals and affiliates. This suggests that functional covert networks do not have a high distinction between the centrality of the individuals, that is, they are leaderless.

Borgatti (2006) provided definitions of the Key Player Problem (KPP). It is a crystal clear concept that dislodges SNA metrics capacities in identifying all key players within a social network (Ortiz-arroyo, 2010).

Ozgul revealed different topologies on terrorist organisations based on their formations from literature (Ozgul, 2016). The work stresses that variation in terrorist topologies plays an essential role in the positions of most members who play key roles. Terrorist groups are resilience because critical players who are vital to the organisations' existence and recuperations are missed or evade detection. Eiselt and Bhadury (2015) identified dynamic positions of key players as one carried out through manipulating conversation. The manipulation pave way for important members to look like unimportant actors while

less important members look like prominent ones. In short, manipulation has potential of distracting attention of detectives from real vital players.

Positions of participants in social networks are dynamic. This has also attracted the attention of academia and criminal investigators by devising techniques for mining high-profile criminals from decentralized and dynamic networks. Dynamic relationships exist among participants in OCGs. Study and adopting common topological analysis are becoming inadequate strategies for disrupting criminal network because relationships among syndicates are absurd. Yao explores dynamic network to identify hidden relationships (Yao *et al.*, 2016). Basaras *et al.* (2017) developed a technique that supports a dynamic network. The development of dynamic centrality metrics provides alternative to address dynamicity in a complex network (Yao *et al.*, 2016; Huang & Yu, 2017).

Dynamism in social network is more connected to rumour spreading, computer virus attack and human infection disease than criminal network organisations (Basaras, 2013). It is about identifying influential spreader and spreading capacity of actors; that is actors capable of infent other actors or transmit infectious diseases. This is an attribute that high-profile and affiliate key players to terrorist groups might not be bound to (Liu *et al.*, 2016; Zhang *et al.*, 2016; Wang *et al.*, 2017).

Data is becoming prominent for developing preventive strategies against incessant crimes orchestrated by OCG (Ashby, 2016). Preventive strategies are meant to find either long term or short-term measures to forestall future reoccurrence (Maeno & Ohsawa, 2007a). Such mechanism lowers confrontations between security agencies and foot soldier terrorists. It offers law enforcement agencies ample time to study members in the data and to identify hidden members - especially those that are more pertinent to the existence of the organisation.

Analysing crime data is one of strategies for obtaining criminal intelligence to support conventional warfare approaches. Crime data contain various information about criminal activities including covert member those that overt members work for (Hulst, 2009). There is more potential to identify high-profile criminals within an OCG through analysing their crime data than in the warfare confrontation which they hardly participated (Ismail *et al.*, 2017). Crime data have fundamental challenges related to sources and reliability (Berlusconi, 2013).

Data sources and reliability dominated the notion on data defectiveness. These have potential of influencing network structures of participants. The principal suspect is that defective data conceal influential participants and affiliates (Butt *et al.*, 2014; Berlusconi *et al.*, 2016). Objectively, defectiveness is generic and inevitable on any data. For crime data, it connotes omission of inconspicuous relationships or missing links (Kossinets, 2006;Parisi *et al.*, 2018). It also denotes exclusion of some participants called missing nodes (Maeno, 2007, Eyal *et al.*, 2011; Sina *et al.*, 2013). The two insinuations - omission of relationships and high-profile participants undermine the efficiency of security intervention and intelligence (Duijn *et al.*, 2014).

This research aimed at using telecommunication metadata of a terrorist or militant group as a reliable, formidable and robust data for tackling data defectiveness. Gunnell (2016) presents different classes of grading police intelligence data. The ubiquitous use of mobile phones was identified as a reliable source for crime data (Varese, 2013; Basu, 2014). Telecommunication metadata as underlying information, contain blueprint on activities of mobile phone users. It is regarded as the best for predicting behaviours of mobile phone users as well as to replicate individuals with respects to social relations be it in groups, online or offline transactions(Campana & Varese, 2012; Butt *et al.*, 2014; Ferrara *et al.*, 2014).

Besides, huge data are produced daily from the use of mobile phones which should be used by security operatives in fighting terrorism and other crimes. Unfortunately, majority of researches caried out on terrorism make use of open-source data. A number of research done on the 9/11 attacks also made use of open source data (Kreb, 2002; Levi, 2007; Eilstrup-Sangiovanni & Jones, 2008; Course & Hill, 2014). Data about phone users participating in various criminal activities can be extracted from telecommunication gadget(Memon *et al.*, 2011; Ferrara *et al.*, 2014; Onwuka; *et al.*, 2016).

Telecommunication metadata is a collection of recorded information about phone-user such as location, altitude, time of call and duration of calls. These contain useful tips to identify phone users participating in illegal activities and to prepare adequate interventions. Thus, detection of high-profile criminals not well known to security agencies could be tracked from telecommunication metadata, and it can aid security efforts towards combating criminal organisation resilience.

Access to intelligence data alone is not sufficient to tackle the problem of critical players; as knowledge of theoretical graph and conception analysis are also inadequate to solve it. Hulst (2009) opined that SNA is a promising tool needed by law enforcement agencies. Specific methodological problems associated with criminal intelligence data and lack of experience with SNA applications hampered researcher's ability to improve knowledge of organised crime and terrorism. Key players in terrorism have contravening attributes to that of influential actors in the open organisations – none criminal organisations (Ismail *et al.*, 2017).

Lampe (2009) proposed two features for identifying prominent members of adversary networks. The features are human capital and social capital attributes. The human capital attribute is to supplement social capital attribute of criminal actors. The social capital

attributes are obtainable through SNA metrics (Bright *et al.*, 2015). Methods for identifying human capital attributes are rare in literature, but it had been cajoled through SNA illustrated in (Gunnell *et al.*, 2016; Malm *et al.*, 2016; Bichler *et al.*, 2017).

Data defectiveness is a hindrance to detection of covert nodes. Telecommunication metadata has no information directly related to human capital attribute that that is, personal attributes of mobile phone users. This is part of missing information in telecommunication metadata. Covert members as used in this research denote affiliates, high-profile collaborators or co-offenders inside a dataset of terrorist groups. It is highly challenging to identify these set of participants by SNA tool because they always lie low. These are set of participators who engage with OCGs through inconspicuous relationships. Hardly noticeable by criminal investigators as key players. This is also affecting significance covert members from detective techniques. Real-life social status of the affiliate criminals submerges their status making covert members becoming unnoticed as key players.

1.2 Statement of the Research Problem

The problem of insecurity confronting the world is emanating as persistent-conflict. Its intermittent nature and sporadic occurrence indict the current security operative approach as ineffective in bringing it under total control or complete eradication (Manning, 2010; Ferrara *et al.*, 2014). This is because the conventional approach is after overt members of a criminal group who execute organisation's agenda. But conventional approach is not after hidden members whose roles are not exposed to public and security agents (Bright, 2015). The activities of hidden members, otherwise known as covert members, are pertinent to the existence and recuperation of the criminal group whenever overt members are eliminated. Covert members enjoy secrecy on their identity because relationships with

overt members are not easy to establish due to covert ways of communications (Butt *et al.*, 2014). Also, social networks of participants, hide hierarchies of the members due to flat organisational structure (Chatterjee, 2005; Clauset *et al.*, 2008) and leaderless principle they imbibe (Husslage *et al.*, 2012). The problem now is how the covert members (covert nodes in a network) of a criminal group can be identified by inference from social networks of overt members; this is a missing node problem.

1.3 Aim and Objectives of the Study

This research work aims to develop enhanced Bayesian network-based algorithm for detecting covert members of a criminal group using mobile telecommunication metadata. The objectives for achieving this aim are set as follows:

(i) To develop a Bayesian model for covert node detection.

(ii) To develop an algorithm for covert node detection based on the model developed in (i) above

(iii) To evaluate the performance of the developed algorithm using network attributes of criminal's mobile phone call metadata,

(iv) To use SNA - Quadrant model for validation of nodes detected in (iii) above, and

(v) To compare the algorithm with the existing covert node detection algorithms.

1.4 Significance of the Study

This work will be beneficiary to individuals, groups and nations. The work is to complement security agents' intervention and provide adequate intelligence to fight crimes or persistent conflicts. The proposed approach is to identify covert members in a criminal organisation; those making criminal groups become resilient. Identifying those covert members can lead to successful disruption of OCGs.

This approach will facilitate and promote socio-economic stability of society, as it will expose key actors behind persistent conflicts. With this approach, less resources and security personnel will be effective for disrupting OCGs. This will definitely cut down the security votes.

1.5 Scope and Limitation of Study

Criminal organisations are regarded as complex systems. Relationships among participants in communication networks are also complex. With the concept of complexity, the following scopes, assumptions and limitations are put into consideration in this thesis. This work considered OCGs in general but streamline the target to terrorist groups.

(i) Telecommunication data of terrorist groups are intended for evaluation.

(ii) Participants are presumably bounded in telecommunication metadata;

(iii) The selected attributes are those defined in literature, that also related to features of salient criminals especially key players in OCGs.

(iv) Development of enhanced Bayesian model was made to infer the salient actors.

(v) The implementation and testing of the models are limited to the advanced laboratory-scale testbed.

1.6 Thesis Organisation

This work is divided into five chapters; chapter one presents an overview of criminal organisations as organised crime groups orchestrating persistent conflict. Factors preventing security agency from annihilating OCGs were mentioned. Also, the contents of chapter one includes the problem statement, aim and objectives, significance as well as scope and limitation.

Chapter Two starts with overview. It opens discussion on organisational structure and typologies of social networks. Impacts of the dark organisations are briefly enumerated to substantiate the need for this research. A section of the chapter discusses various techniques, models and algorithms for detection of key players. Sources of data, dynamism concepts related to criminal organisation structure; strategic position, vulnerability, network disruption and resilience were reviewed. A detailed review of related works on key players attributes, as well as techniques for identification and prediction of covert members from covert social networks or dark networks were presented. Chapter Two was concluded with a summary of the research gaps this work addressed.

Chapter three focuses on development of proposed methodology to address part of research gaps identified in the reviewed works. It shows steps towards actualising Enhanced Bayesian Network Model (EnBNM) for prediction of covert nodes. It also presents modality for identifying profiles of relevant participants in OCGs using SNA - Quadrant (SNA-Q). The SNA-Q serves a validation tool. Roles and contribution of participants in a criminal network are projected on four criminal profiles: Q1 to Q4. All metrics for evaluating the performance were obtained from social networks of criminal groups.

Chapter Four presents results of EnBNM algorithm and its performance on two criinal datasets of N'17 revolutionary group and 9/11 dataset of a faction of Al-Qaeda terrorist group. Results of SNA-Q algorithm validate EnBNM detection. Comparative analysis of EnBNM was done with existing entropy variation algorithm

Finally, chapter Five concludes this thesis, with review on aim and objectives achieved, the contribution of the study to knowledge and recommendation for future works.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Chapter Overview

This chapter discusses state of the art on the use of advanced technology emerging from the science of complex networks for the provision of security and fighting OCGs through intelligence gathering. It encompasses a review of complex networks, covert networks' theory and algorithms for the detection of covert nodes. Covert networks refer to the blending of criminal groups within an extensive complex network (Maksim & Carley, 2003; Kramer, 2007; Memon *et al.*, 2011; Husslage *et al.*, 2012; Smith *et al.*, 2013; Basu, 2014; Butt *et al.*, 2014; Smith *et al.*, 2014). The chapter discusses covert network taxonomy and associated challenges for crime fighting and counter-terrorism. Related works on various techniques and algorithms for detecting covert nodes were discussed. A discussion on data sources for analyses of covert networks and the importance of Telecommunications metadata for uncovering hidden criminals is presented. This chapter is concluded with research gaps and the research proposition.

2.2 The Global Face of Crime

Methodology for fighting crime has become a multidisciplinary subject. The nature and extent of today's global crime calls for more intelligence gathering and better intelligence-gathering tools. Various disciplines such as Physics, Mathematics, Biology, Chemistry, Chemical Engineering, Telecommunication Engineering, Computer Engineering and Computer Science are getting involved in study and design of different types of intelligence gathering tools to assist law enforcement agents in crime-fighting. This is because, locally and globally crime rate has been on the increase with reference to Tables 2.1 and 2.2.

Table 2.1: Snapshot of Nigeria's position in Global Crime Index

Rank	Country	Crime index	Safety Index
1	Venezuela	84.25	15.75
2	Papua New Guinea	80.24	19.76
3	South Africa	77.07	22.93
4	Afghanistan	76.37	23.63
5	Honduras	74.78	25.22
6	Trinidad and Tobago	70.95	29.05
7	El Salvador	68.82	31.18
8	Guyana	68.74	31.26
9	Syria	68.09	31.91
10	Brazil	67.85	32.15
11	Jamaica	67.53	32.47
12	Angola	66.63	33.37
13	Peru	66.61	33.39
14	Namibia	65.89	34.11
15	Bangladesh	64.22	35.78
16	**Nigeria**	**63.86**	**36.14**
17	Puerto Rico	63.35	36.65
18	Argentina	63.31	36.69
19	Bahamas	62.25	37.75

(Source: https://www. numbeo.com/crime/rankings_by_country.jsp; date 26/05/2021)

Table 2.2: Snapshot of Nigeria's position in Africa Crime Index

Rank	Country	Crime index	Safety Index
1	South Africa	77.07	22.93
2	Angola	66.63	33.37
3	Namibia	65.89	34.11
4	**Nigeria**	**63.86**	**36.14**
5	Libya	62.00	38.00
6	Kenya	61.40	38.60
7	Zimbabwe	58.88	41.12
8	Tanzania	56.64	43.36
9	Uganda	56.07	43.93
10	Somalia	56.04	43.96
11	Botswana	52.84	47.16
12	Algeria	51.88	48.12
13	Ethiopia	50.03	49.97
14	Morocco	49.10	50.90
15	Ghana	48.52	51.48
16	Mauritius	47.89	52.11

(Source: https:// www.numbeo. com/crime/rankings_by_country.jsp date 26/05/2021)

Table 2.1 shows that Nigeria occupies the sixteenth position on the globe and it occupies the fourth in Africa from Table 2.2 with crime index of 63.86 percent and safety index of 36.14 percent. The positions of Nigeria in the two tables are likely to be true when considering activities of different OCGs - from the Militants, Street-gang, Fulani-herdsmen, Cattle rustlers, Oil-bunkering, Boko-Haram and recently sporadic operation of Kidnappers. There is high tendency for further lowering of safety index for Nigerians. The summation of crime index and safety index is equal to hundred percent.

Terrorism is the most difficult OCG to track due to diverse operational activities criminals involved in (Eilstrup-Sangiovanni & Jones, 2008). Terrorist groups are fond of concealing their criminal activities from security operatives and the public that is, spreading fear (Behzadan *et al.*, 2017). Threats manifest from massive destructive attacks. These range from killings, maiming, raping and sometimes looting. Terrorist groups are not the only criminal group that spread threats. Other OCGs like Drug Trafficking Organisation (DTO) also do. Although their own attacks could be selective to their opponents in business. Petta opined that terrorism and OCGs are the same (Petta, 2018).

Terrorist groups are driven by different ideologies (Barnes, 2017; Bichler *et al.*, 2017). Some terrorist groups are known to be profit or money oriented when others are tribalistic-based or ideological-based. Terrorists like launch massive destructive attacks on society and openly admit such actions in order to get money, plan and execute other their criminal activities. Towards the end of cold war, some of OCGs sponsored to the war started experiencing more lesser support from the state that sponsored them. This forced majority of OCGs to devise new sources of funds to support activities. It was speculated that this turn in events had led to increase in criminal activities of various kinds to generate money by OCGs (Levi, 2007; United Nations, 2014).

Often, activities of OCGs are illicit and violent against the State and society where they thrive. Sometimes OCGs amass wealth and power, despite terrorists do not like being identified. Occasionally, some criminal groups launch open attacks (referred to as gang wars) against their opponents in retaliation or as a show of superiority. Vice-a-vice, OCGs act like a terrorist organisation. All these imply that organised crime groups and terrorist groups have areas of similarity in structure and activities. This craves for common type of intelligence gathering and methodology to be applied to both. Hence, both can be referred to interchangeably as terrorist groups or OCG (Petta, 2018).

Criminal activities were initially restricted within local and geographical reaches. However, globalization and the advent of information infrastructure had enabled criminal organisations to grow into a national and transnational organisation (Brunetto *et al.*, 2016; Abazia, 2017). The internet and advances in telecommunication infrastructure made it possible for OCGs to spread their tentacles beyond local boundaries, making it possible to receive supports and alliance across the globe making those behind national boundaries become less tractable and apprehended.

Currently, a lot of multidisciplinary researches are trying to explore telecommunication infrastructures for disrupting OCGs, the same facilities that make criminals become invisible and to engage in net-war with security operative (Malm & Bichler, 2011; Welser *et al.*, 2011; Madeira & Joshi, 2013). This is done by collecting their digital footprints called telecommunication metadata. This pieces of data contain information about participants and activities involved in online of the communication networks (Ferrara *et al.*, 2014). Proper analysis of metadata produces facts related to people involved in criminal activities and roles individual played. These are intelligence data needed by law enforcement for effective destabilization of criminal groups.

2.3 Complex Networks

Networks have recently become a paradigmatic way of representing complex systems. It works with interactions or relationship that exist between a system's constituent parts. The pattern of relationship is itself an intricate and is evolving together with the system's dynamics. A network is an underlying structure and it exists in natural and human-made systems. Resulting networks from natural and man-made system are characterized by randomness and structure. Dynamic systems also confine to a complex network (Holme, 2003). A network or sub-network always has a set of nodes that are connected. The complexity is not on the size but intricate connections and topological properties systems shared.

A complex network is an aggregation of sub-networks or community of networks (Duch & Arenas, 2005; Costa *et al.*, 2006; Fortunato, 2010; Gliwa *et al.*, 2012; Klemm *et al.*, 2012). Many complex systems of interest, such as the internet, social and biological relations are depicted with network structure. The network structure sometimes illustrates a system's function. This is also applied to abstract systems or models. Threat networks are illustrated with arbitrary connections of nodes (Smith *et al.*, 2012; Carter *et al.*, 2014). Human relations also have these described features illustrated in Figure 2.1.

Figure 2.1 shows complex networks formation and evolvement of subnetworks from human relationships or interaction. A family is a complex network. It occurs from union of a man and a woman which are elements from different families, through marriage another family evolves that is another community of network - union or relationship. Such representation also exists in schoolmates, hobby, friends and scientific community networks. Each name given in Figure 2.1 represents a community of network or a subnetwork that evolved from one of the relationships in entire human races.

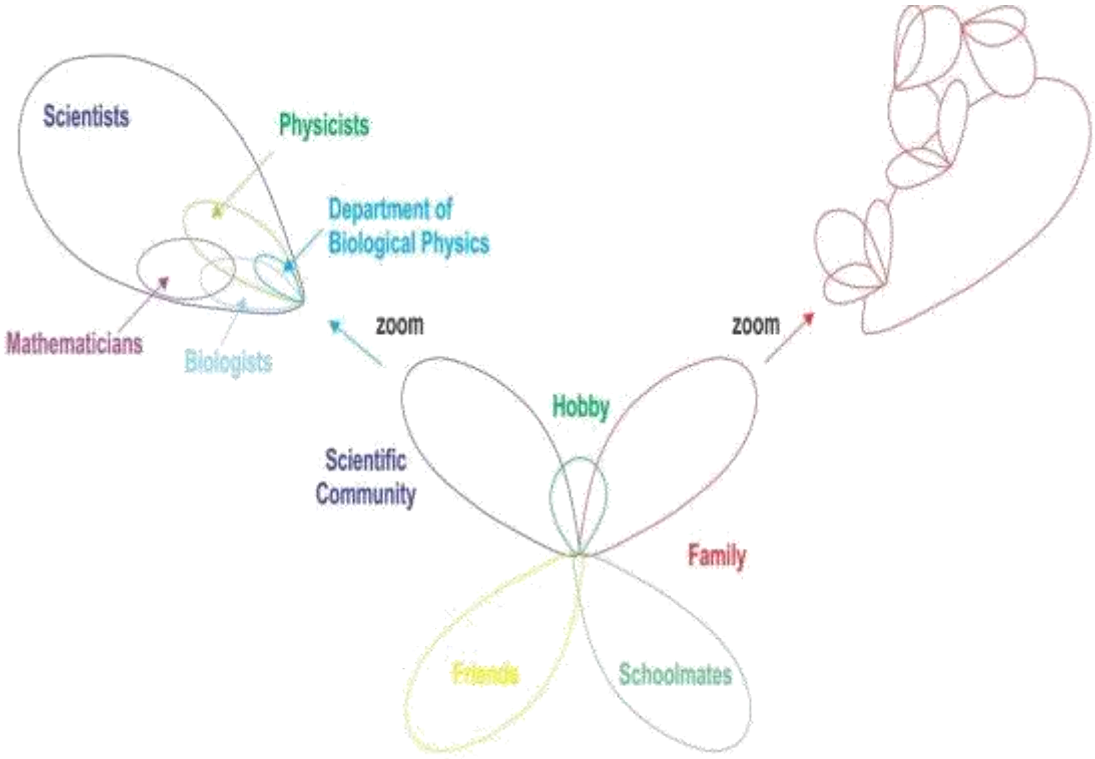

Figure 2.1: Complex Network Evolution
(Source: Palla *et al.*, 2005)

Different community of networks have potential of evolving as a result of members' interaction. The interactions of scientific community members lead to various Engineering professions, Biological Physics and others related professions. Criminal network also evolves from interaction of members from different communities of networks. It shows that OCGs are isolated communities (Maksim & Carley, 2003; Smith *et al.*, 2013). OCGs have its elements from different networks. This is a reason OCGs are regarded as covert networks (Hussain, 2009; Butt *et al.*, 2014).

Complex systems are often described with models. Models reduce intricacy associated with links and offers simple description about a system. However, it is considered as insufficient concept due to few components used for description of an entire system. This was still adopted for criminal networks. The models make use of only conspicuous relationships and participants. From model description, significant parts of information are omitted or missing (Fortunato, 2010). There is tendency for larger missing

components unincorporated to influence decision drawn negatively over the model that have few parts description (Bliss & Schmidt, 2013).

Today, complex networks are everywhere - from national power grids and airline networks to social contact disease networks, neuronal networks and protein-protein interactions (Ghasemi *et al.*, 2014; Zhao *et al.*, 2015; Yao *et al.*, 2016; Parisi *et al.*, 2018). To understand behaviour of complex systems, it is imperative to first chart the structure of network. In neuroscience, the study had gradually seen a transition from reductionist studies of individual neurons and structures to a holistic approach that is, charting the overall interaction of the system. Figure 2.2 presents few examples of complex networks in diverse settings.

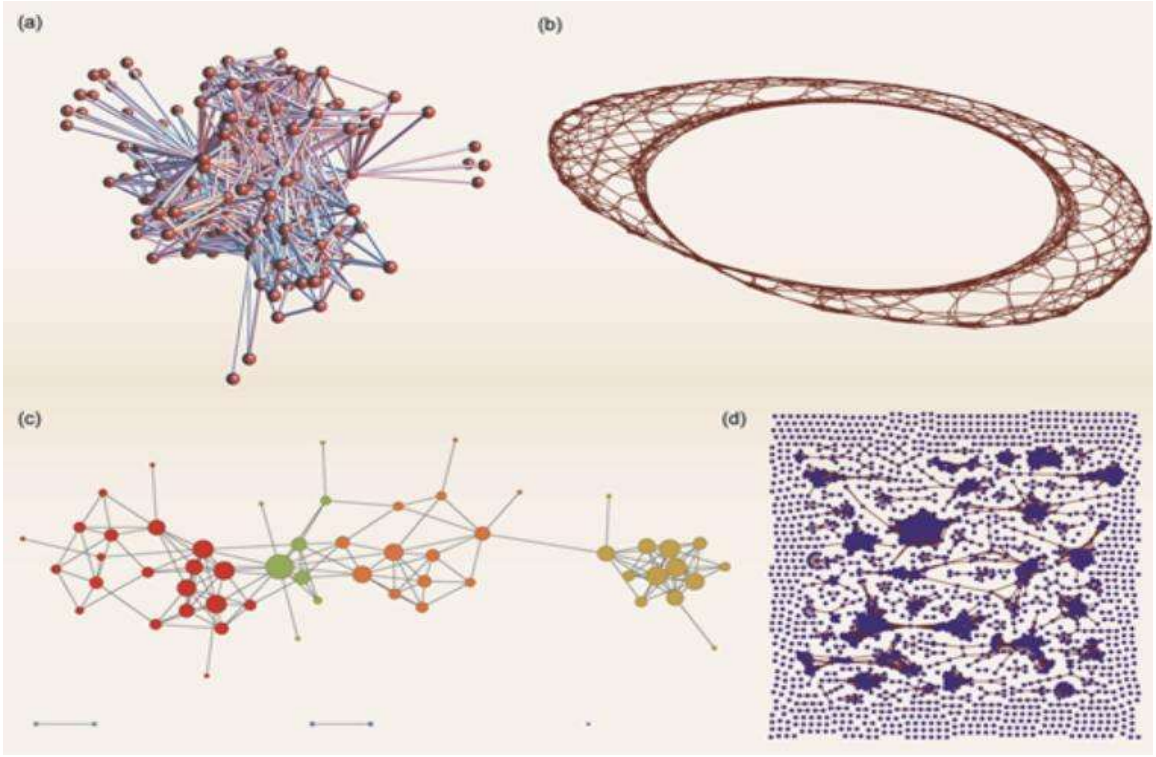

Figure 2.2: Four examples of Complex networks: (a) represents a wiring diagram of the nematode worm brain, (b) a complex network constructed from a chaotic Rössler circuit, (c) a partial representation of a person's Facebook friend network, coloured according to clustering and (d) a (fragmented) network of potential infection pathways for avian influenza (Source: Fortunato, 2010)

2.3.1 Classification of complex networks

Complex networks are governed by the rules and conditions of evolvement (Belinda, 2010). Complex networks are classified using social network perspective. Its classifications are based on types of edge, nodes and different ways nodes are connected. A full, partial or egocentric are part classification based on number of nodes, while unimodal, bimodal and affiliation networks are classes based on involved entities (Hensen, 2011b). Multiplex network is a classification based on different ways people that is, nodes are connected. Multiplex gives a multiple relational networks (Zignani *et al.*, 2015; Sharma & Singh, 2016).

A full or complete network contains all entities of interest. Full networks are multiplex in structure – entities are connected in different ways. Multiplex structures are difficult to analysis without rendering them into simple structure like unimodal type. Entities in unimodal structure are of the same feature type that is, user-to-user or document-to-document type. This is the simplest mode to represent a complex network.

Partial network can be created from full network by taking few entities of multimodal networks. Both unimodal and partial network lower intricacy and aid comprehensive analysis than bimodal and multimodal networks (Guillaume *et al.*, 2006; Kossinets, 2006). Bimodal network can be transformed into two unimodal networks: a user-to-user and an affiliation-to-affiliation network. The transformation prevents omission and suppression of relations in unimodal. But a unimodal network ensures that all entities have the same rules and conditions. The unimodal structure permits to view components of complex networks as only nodes and links. The two components are another source of complex network's classifications.

An edge connects two nodes or vertices. There are three types of edges: directional, bi-directional or non-directional. A directional edge exists in communication networks illustrated in Figure 2.3 (a) and (b).

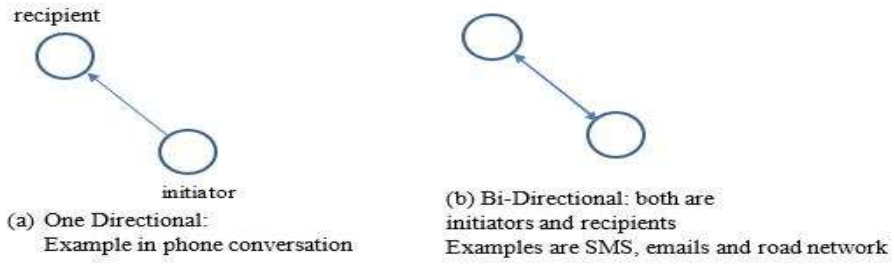

Figure 2.3: Directional Relationship between an initiator and a recipient

Figure 2.3(a) depicts type of edge in communication conversation. An initiator of conversation that is, a caller is differentiated from a receiver using a directional edge - one arrow-head. Figure 2.3(b) illustrates an edge that depicts conversation between two phone users; between a caller and it's called; a transaction or a relationship between two persons. Sometimes, bi-directional edges represent relationships that combine initiating link and response link. This also goes with ties, transactions, Short Message Services (SMS) and email transferred between two people. This is also applicable to non-directional edges. Bi-directional edge can be replaced non-directional.

Edges are in different sizes. Base on edge size, complex networks are also classified as a weighted digraph, unweighted digraph, weighted graph, and unweighted graph. Any network can be transformed from one of these formats to another (Ying *et al.*, 2017; Zejun *et al.*, 2017). Unweighted edges are used for relationships that are assigned the same weight. This edge is common in a unimodal network. Random networks are described with unweighted link between its dyad. When equal weight is assigned to edges in network, it makes links within to be uniform. Unweighted edges sometimes avoid preferential treatment on relationship; every edge is treated equal and assigned the same

weight. But the weighted network considers frequency of interactions in phone conversation or any other factors in weighing relationship between a dyad.

2.3.2 Types of complex networks

Complex networks are categorized into three types: random, small-world, and scale-free networks(Xu & Chen, 2008; Qiao *et al.*, 2017). The trio belong to unimodal class of networks. Figure 2.4 shows an overview of complex networks taxonomy using three metrics: heterogeneity, randomness and modularity. Position of terrorist network is behind the frame which implies that properties are not compatible (Ilachinski, 2005). Terrorist and OCGs are hierarchical network while modularity randomness and heterogeneity are less concerned with hierarchies (Xu and Chen, 2008).

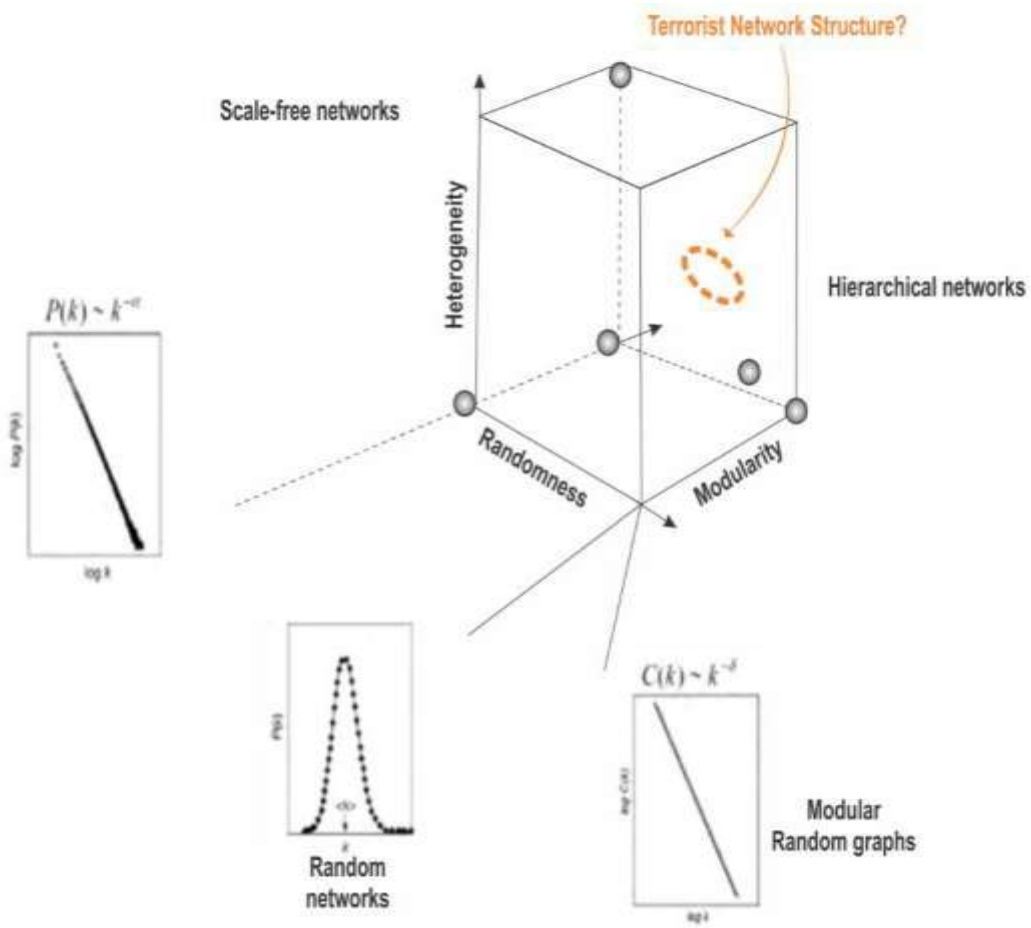

Figure 2.4: Overview of Complex Networks Taxonomy
(Source: Ilachinski, 2005)

(i) Random Networks

A random graph is a network whose connections are based on probability and number of nodes. The theory of Random Graph (RG) was developed by Erdos and Renyi (Ilachinski, 2005). Erdos and Renyi considered a sample space that is feasible for actualizing ($\frac{N}{2}$) order N labeled graphs $(,)$ with equiprobable size . RGs follow the binomial model that is defined by equation (2.1). The addition of edges requires equation (2.2). A random graph roughly has N number of links with probability, (Piraveenan, 2010).

$$( ) = (1 -)^{(} \quad (2.1)$$

$$( ) \propto \quad (2.2)$$

where P (G) is probability of graph G; P(N) is link probability scales and is a tunable parameter. The is an independent probability $0 \leq \geq 1$.

When analysing a network, one approach is to view a network as a single fixed entity. But sometimes, edges are viewed as random variables. Irrespective of perspective, a given network has more components than an individual object within the graph. A random network can be viewed as a sample from probability distribution and should be studied as a whole rather than in pieces when to gain an insight about the network (Clauset & Woodard, 2013). This perspective permits analysts to see probability distribution of random networks as a whole ensemble of many different networks, whereas probability distribution determines probability of any particular network. Rather than looking at the particular properties of a single network, it is better to study the properties of the whole ensemble of networks.

One advantage of looking at an ensemble of networks defined by a probability distribution is to minimize or lower the influence of different statistical properties. One can imagine taking many samples from the probability distribution and then look for properties that are common to most of the samples. Sometimes, it could be to take an ensemble average of a quantity, that is, taking an average of the probability distribution (Ghasemi *et al.*, 2014).

Besides the mathematical convenience, looking for properties over a whole network ensemble makes sense in a real-world application. Despite brains are wired differently, people perform the same activity through similar motor tasks. For analysts, there could be more need to study common properties facilitating motor task action than seeking variation on how each brain was wired. An ensemble network has random network framework. Different networks can be sampled with the same probability distribution. This probability distribution is the property over the entire network.

Research conducted on complex networks by physicists, computer and social scientists revealed that complex networks are not completely random graph but they also exhibit small-world and scale-free properties (Ilachinski, 2005; Behzadan, 2016).

(ii) ***Scale-Free Networks***

A scale-free network is a graph whose degree distribution asymptotically (Huang & Yu, 2017). That is, the fraction of ( ) of nodes in the network having connections to other nodes goes for large values of k given as equation (2.3)

$$( ) = ^{-} \tag{2.3}$$

where c is normalisation constant and γ is a parameter whose value in the range $2< \gamma >3$.

A network is called scale-free if the characteristics of the network are independent of the size of the network, that is, the number of nodes. It means, when the network grows, the underlying structure remains the same. That is; the ratio of the number of very connected nodes to the number of nodes in the rest of the network remains constant as the network changes in size. Compared with the Erdos-Renyi (RG) where most nodes typically have several links near a small or average value. It is observed that the evolvement of RG leads to the relative number of very connected decreases.

A scale-free network is extremely inhomogeneous. Majority of nodes in scale-free graphs have only one or two links when a few nodes have a large number of links. Another fact is that both random graph and scale-free exhibit small-world property but only the scale-free network has the short path length that passes through one of the highly-connected hubs (Xu & Chen, 2008). This gives information about the system behaviour and tolerance to outside intrusion or a targeted attack. Scale-free network is found in many application areas related to science and engineering. These include:

- the topology of web pages - where the nodes are web pages, and the links are hyper-links,
- the collaborative network of actors - where the nodes are actors, and the links are co-stars in the same movie,
- the power grid system - where the nodes are generators, transformers, and substations and the links are power transmission lines,
- the peer-reviewed scientific literature - where the nodes are publications, and the links are citations (Ilachinski, 2005).

Figure 2.5 is an example of a random graph and a scale-free graph. A scale-free graph has few shaded nodes indicating highly connected nodes. But, nodes in the random graph had number of links in proportional to (2.3)

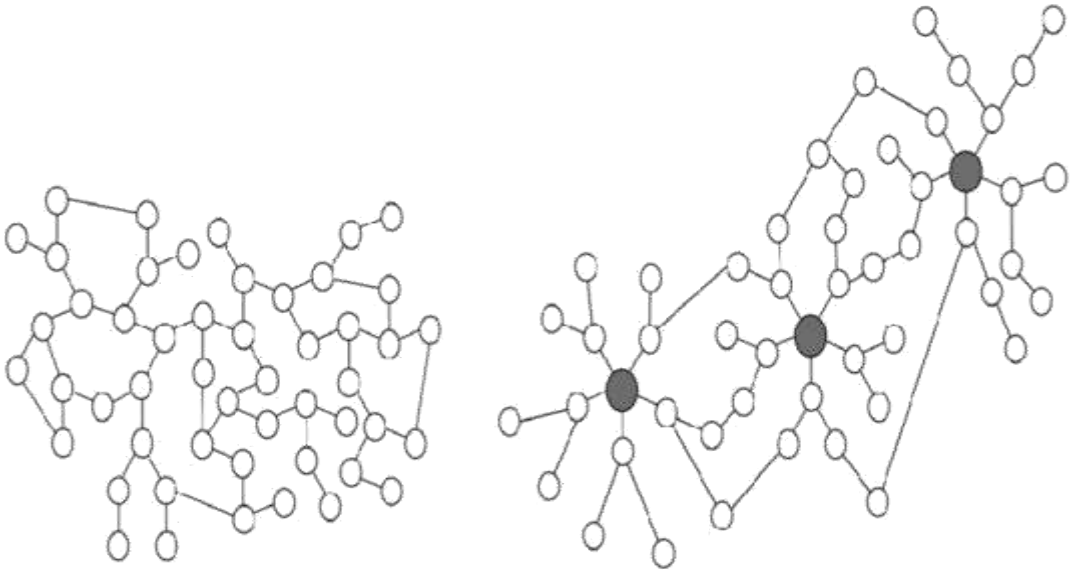

Figure 2.5: An example of Random network vs the Scale-free network
(Source: Alvarez *et al*., 2015)

(iii) ***Small-world Networks***

A small-world network is a type of mathematical graph in which most nodes are neighbours to one another - that is neighbours of any given node could be neighbours of each other. Majority of the nodes can be reached from every other node by a small number of hops or steps as shown in Figure 2.6.

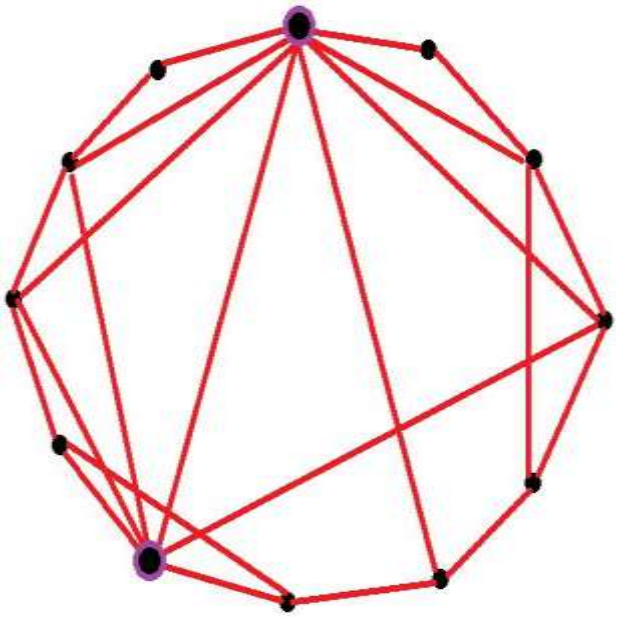

Figure 2.6: An example of the small-world network system model

The small-world network is constructed by randomly rewiring the edges of a ring lattice with nodes. It represents a graph with a large proportion of nodes with a few links. Only small percentage of nodes having a large proportion of links. Examples of small-world networks include collaborative social-networks of film actors, the US electric power grid and the neural network of a nematode (Ilachinski, 2005). The small-world model has been actively applied to the communications networks research due to resulting network topology with features such as smaller average transmission delay and more robust network connectivity (Xu & Chen, 2008).

2.3.3 Application of complex networks

Huge data are emanating from various human activities including criminal data. Supercomputer and high processing speed devices have aided collection and processing of big data produces from complex relationships (Kasture, 2012). Over the last decade, numerous applications based on new understandings of complex networks have been reported. Human-made systems like communication networks exhibit complex network properties (Chatterjee, 2005; Eiselt & Bhadury, 2015; Zignani *et al.*, 2015). Communication networks confirm existence of relationships among participants or organisation entities. A complex network graph is defined mathematically as $= (,)$, where V is set of vertices (nodes) and E as set of edges (links). The following are just few complex systems.

(i) ***Computer Networks***

Computer networks refer to a set of computer terminals interconnected together. The interconnection permits other terminal computers to have access to resources on other systems, as this prevents duplication of resources. Computer networking ensures that there is communication between one system and others via media that is either wireless

or cable (Maeno & Ohsawa, 2007b). It also shows that there is a structural path to the flow of information. Some rules or protocols can guide communication. The structure in the computer network is known as topology (Xu & Chen, 2008). There are different topologies. Each topology has its own merits and demerits. Topology of computer networks is a structure that provide information about important terminals. If the topology is unknown, it may be difficult to identify important computer that serves as a hub. Effect of disconnecting or removing a computer from a network can vary from one topology to another.

*(ii) **Criminal Networks***

Criminal networks are organised groups of people perpetrating nefarious activities. Crimes cannot be single-handedly committed without alliance with some expert co-offenders (Ouellet *et al.*, 2013; Tayebi, 2015). The co-offenders use their experiences, personalities and positions to facilitate crimes (Behzadan *et al.*, 2017; Grassi *et al.*, 2019). Co-offenders in OCGs are like terminals in computer networks and WSN. Unlike computer terminals, participants in OCGs are of different personalities.

Some OCGs have individuals that have personal interest in crime. This includes members who are benefitting from crimes directly. OCGs thrive on collaborations and effective communication (Behzadan *et al.*, 2017; Grassi *et al.*, 2019). Criminal relationships are built on trust and the trust leads to collaboration. Relationships between criminals are multi-dimensional and hardly noticed (Lin & Chalupsky, 2003).

Criminals conceals their relationship. This prevent public and security agents from having knowledge about people who are involved (Everton, 2009; Manning, 2010; Memon *et al.*, 2011). Figure 2.7 is a network graph of Global Salafi Jihadists (GSJ). It shows different sub-networks with the GSJ. The sub-networks were coloured according

geographical locations of participants. The links between nodes represent communication media or ties.

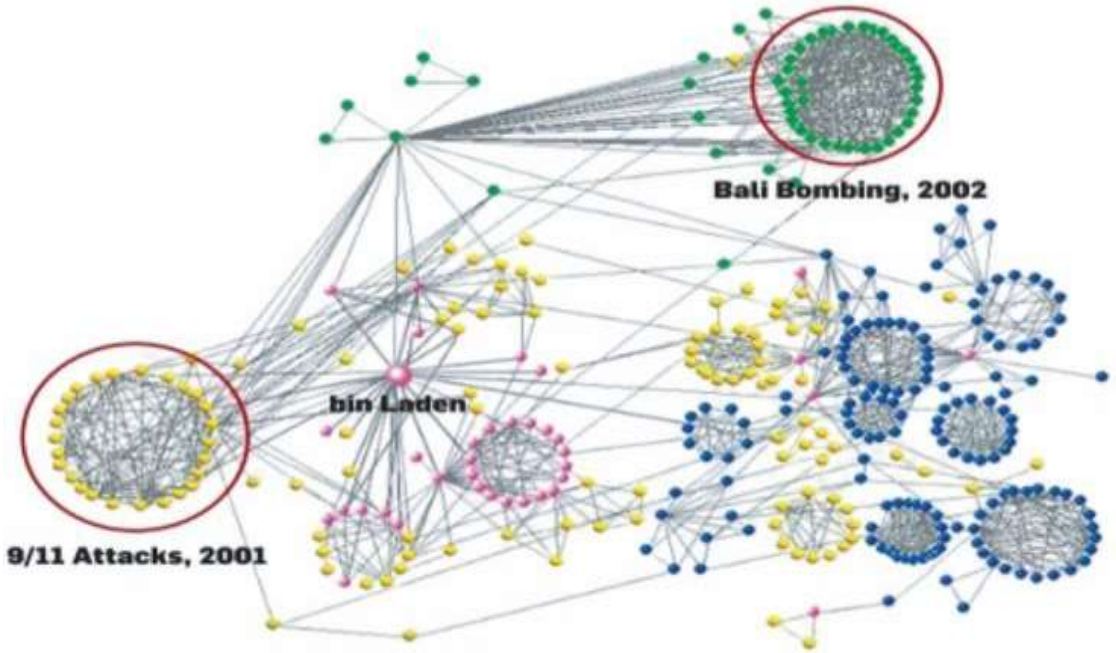

Figure 2.7: Global Salafi Jihadi Criminal Network
(Source: Xu & Chen, 2008)

(iii) ***Social Networks***

Social networks concern relationships, interactions and communication exercised by human or devices. The relationship between two persons can be that of birth, family, relatives, school attended, religion gatherings and geographical location (Calderoni, 2012; Saxena *et al.*, 2018). Interactions do involve two persons who use a medium of communication to share ideas (Zignani *et al.*, 2015). Interaction also encompasses using language, signals or signs by two persons or group of people. Social network is small unit of a complex network. This, sometimes established around people somebody is known with (Jones *et al.*, 2018).

Social network shows how people are related people. The relationship is not restricted to blood relations alone. Relationships across all parts of the globe have made world look like a global village. There are different types of social network media that are in use - 26

such as twitter, IMO, Telegram, and WhatsApp (Madeira & Joshi, 2013;). These media also produce their own complex networks through the people using them. Social media cannot distinguish personal attributes of their subscribers. Different organisation groups use social network media to discuss concerned issues.

*(iv) **Sensor Networks***

Sensor networks refer to a group of space dispersed dedicated devices for sensing and recording of environmental related activities. Sensor networks are organized for collecting data and disperse it to a central location (Sun *et al.*, 2016). The communication between sensors is wireless. The name was coined from the medium of communication between the sensors - wireless communication sensors metamorphosed to wireless sensor network (WSN). A typical WSN is shown in Figure 2.8. The WSN measure could be for measuring environmental conditions like temperature, sound, humidity, and pollution level. It can be deployed as surveillance devices for monitoring and collecting crime data. Wireless sensor networks can be implemented using different topological structures.

Figure 2.8: Wireless Sensor network
(Source: Wireless Sensor Network Archtecture and its Application, 2020)

*(v) **Transportation Networks***

The globalization of transportation, communication and finance have benefitted both legal business and professional criminals. Roads also, have formed part of complex networks connecting cities. These cities are vertices. Transportation networks form

in human relationships as people move from one city to another. Identification of important vertices could be connected with a particular interest (Holme, 2003; Smith *et al.*, 2014; Liu *et al.*, 2016). Figure 2.9 is a vehicular traffic network. The data were constructed from the imaging sensor (Bliss & Schmidt, 2013)

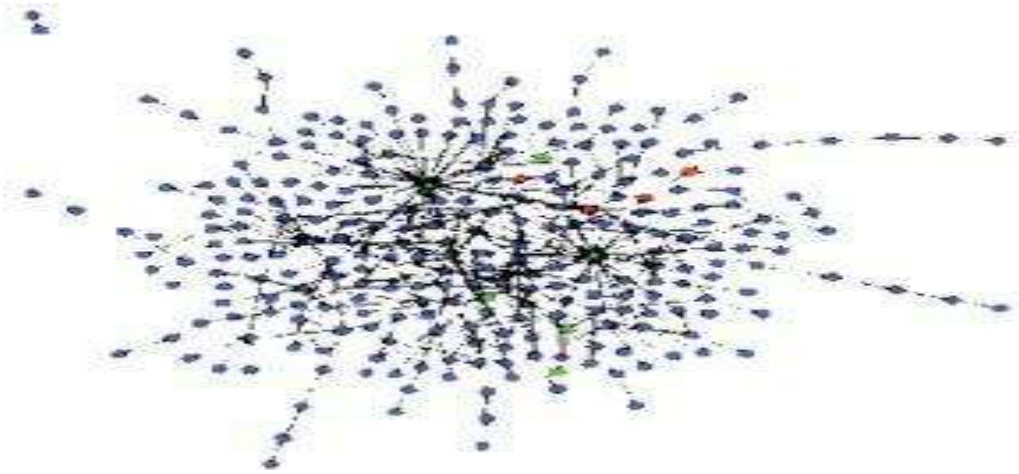

Figure 2.9: Vehicle Traffic Network
(Source: Bliss & Schmidt, 2013)

2.3.4 Relationship between criminal networks and communication networks

OCGs are frequently described with the term network which shows that OCGs have some features in common with complex networks (Le, 2012; Basu, 2014; Behzadan, 2016; Gunnell *et al.*, 2016; Leuprecht *et al.*, 2016; Burcher & Whelan, 2017; Robinson & Scogings, 2018). Criminal networks and communication networks have cordial structures apart from their participants having freedom of joining or quitting networks. Characterization of communications networks are adequate for describing criminal networks as well (Burcher & Whelan, 2017). Common elements in both networks are nodes, links and messages (Le, 2012). Nodes are participants in networks which could be persons, computers or mobile phones (Basu, 2014). Links are connections between persons through which messages flow (Le, 2012).

Analysis of mobile phone calls data has offered insight about human interactions and behaviour of mobile phone users. It had also been used for studying the spread of diseases and human mobility patterns(Ren *et al.*, 2016; Wang *et al.*, 2016; Huang & Yu, 2017; Pei *et al.*, 2017; Jalayer *et al.*, 2018; Namtirtha *et al.*, 2018). It was found useful in counter terrorism and study criminal networks(Ren *et al.*, 2014; Blondel *et al.*, 2015). Data of mobile phone users is needed to construct network that is, relationship among mobile phone users. From network structure, cordial relationships are studied through strong and weak links or relationships (Catanese *et al.*, 2013; Ferrara *et al.*, 2014; Agreste *et al.*, 2016).

Mobile Network Operator (MNO) provides data about mobile phone users. It can also be collected through handheld devices (Onwuka *et al.*, 2016). Data about mobile phone users are used in phone calls network shown in Figure 2.10. Three possible layers are presented in Figure 2.10. The local connectivity presents mobile-phone users that are connected through wireless or Bluetooth. Social contacts illustrate phones users and people contacted through their phones as social contacts. The calling network illustrates a network when mobile phone users make calls. The Both communications networks and criminal networks belong to the family of complex networks that can be studied and analysed using any type of network tools.

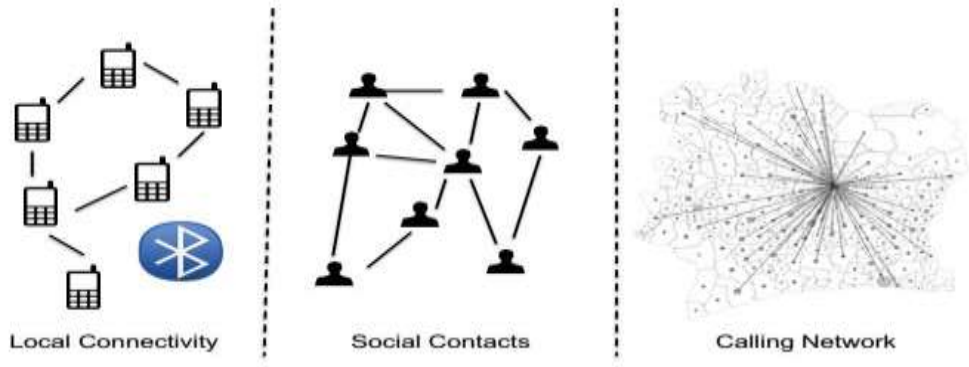

Figure 2.10: Structures within Mobile Phone Datasets
(Source: Ferrara *et al.*, 2014)

2.4 Tools for Complex Network Analysis

Materials for Complex Network Analysis (CNA) are grouped into software tools, datasets and metrics. Software tools encompasses application packages deploying in data collection and construction of social networks graphs (Hensen, 2011b; Borgatti *et al.*, 2012; Thangaraj & Amutha, 2018; Reserved *et al.*, 2019). Datasets contain underlying information about mobile phone users or participants in social groups. Dataset are actual material used for construction of social networks (Yang *et al.*, 2014; Blondel *et al.*, 2015; Rostami & Mondani, 2015; Zignani *et al.*, 2015; Agreste *et al.*, 2016; Robinson & Scogings, 2018). Software is used for extracting dataset and transforming of dataset into network graphs. Metrics are tools for measuring properties (Ilachinski, 2005; Xu & Chen, 2008; Smith *et al.*, 2009; Rodrigues & Milic-Frayling, 2011). Some of these tools are briefly discussed here.

2.4.1 Software tools

Software tools are application packages designed for collating raw data. Some of these tools collate data about events sequentially from the source. Some application packages used at pre-processing stage of data collection and computation of statistic measures are parts of software tools. Spreadsheets packages like Microsoft Excel - MS-xlx is one of such tools. MS excel can be used to view content of dataset, convert a dataset to network graph or for carrying out statistical estimation. Microsoft excel can do these tasks only when datasets are in comma separated value (csv.) format. It is important to note that not all spreadsheet applications can be used for constructing network graph or for visualizing relationships in datasets.

NodeXL is an extendible toolkit for network overview. Its discovery and exploration were implemented as an add-in to the Microsoft Excel version 2007 spreadsheet software. The

tool adds ‘network graph’, as a chart type to nearly ubiquitous Excel spreadsheet. These provide needed supportive features for CNA. A number of them are online and offline for pre-processing of data, constructing network graph and extracting network attributes of nodes in the dataset (Hensen, 2011a; Rotman & Golbeck, 2011; Borgatti *et al.*, 2012; Catanese *et al.*, 2013; Park, 2018). These include UCINET, Python library and MATLAB. UCINET has a free subscription version when there is online subscription with payment.

2.4.2 Datasets and sources of datasets

Crime data are also part of dataset. This constitutes essential material for design of effective strategies against OCGs (Memon *et al.*, 2011; Roberts & Everton, 2011; Berlusconi *et al.*, 2016). Datasets contain pieces of information about organisations activities and participating members in it. Telecommunication metadata as a dataset usually offers pieces of information about participants and relationships. The relationship in telecommunication metadata mostly concerns calls. This is the most reliable source of datasets about OCGs because all calls are recorded when criminals suspicious individual are less concerned of it (Rostami & Mondani, 2015). Inaccessibility to reliable sources of data limit criminal intelligence about OCGs; adversely affect decision-making and indirectly aids resilience (Maksim & Carley, 2003; Bright *et al.*, 2015; Salvatore *et al.*, 2016).

Significances of datasets for fighting OCGs had been investigated on (Belinda, 2010; Berlusconi, 2013). Berlusconi (2013) reported multiple sources as reliable datasets for investigating and analysing OCGs than a single source. Multiple sources detected different key players that a single source was not identified. Another inadequacy of a single source is reporting of a particular relationship that can lead to missing other

relationships like ties, transactions and meeting. One observed challenge of multiple sources dataset of terrorist groups is that they are scarce and rare in deployment. Intermittent nature of terrorist operations prevent access to multiple relations and multi-source terrorist datasets. Bright *et al.*, (2015) and Butt *et al*. (2014) carried out different researches confirmed significance of using multiple sources of datasets over one source.

There is high tendency of missing comprehensive organisational structure through missing ties or relationships due to non-availability of multiple sources (Clauset *et al.*, 2008; Maeno, 2009). Irrespective of information omitted in the dataset, it is called missing information (Maeno, 2009; Eyal *et al.*, 2011; Sina *et al.*, 2013; Berlusconi *et al.*, 2016).

Rostami and Mondani (2015) attempted to verify discrepancies in using multiple sources of datasets and its consequences on intelligence policing. Different individuals emerged as prominent members under different sources of data for the same Swedish street gang. It was concluded that complexity characterized multiple datasets deployed. This provided another evidence supporting latent structure phenomenon raised on key players' evasiveness (Butt *et al.*, 2014; Bright *et al.*, 2015).

Accuracy and reliability of law enforcement data sources had also been researched. It was aimed at verifying reliability of wiretapping with respect to other sources of datasets that law enforcement used as criminal intelligence. Reliability was confirmed on wiretapping of phone conversation of drug cartel members. It was opined that those activities of DTO are supportive. Participants with high profiles status were easily noticed from conversation (Malm & Bichler, 2011; Calderoni, 2012; Varese, 2013). This approach was rare in terrorist research.

Surveillance, court verdict records, telecommunication metadata are parts of police sources of criminal datasets (Catanese *et al.*, 2013; Ferrara *et al.*, 2014). But open

are commonly used in terrorist research (Xu & Chen, 2008; Le, 2012). Another set of sources are information from victims and witnesses; communities and members of publics; Close Circuit Television CCTV or automated number plate recognition; media and internet; commercial statutory and non-statutory agencies (Gunnell *et al.*, 2016). The next paragraphs throw light on communication dataset.

Communications metadata is succinctly captured as "data about data"(Ferrara *et al.*, 2014). It provides underlying information about online and offline activities of mobile phones users - such as geographical location, altitude, time of calling, duration of call, receiver, churn calls and rejected calls (Ferrara *et al.*, 2014; Eiselt & Bhadury, 2015; Ismail *et al.*, 2019). Technically, personality of phone users is inscribed in communication metadata that can aid security operatives locate positions and places frequently visited by users. It is regarded as data for information transportation (Eilstrup-Sangiovanni & Jones, 2008). Various forms of electronic communications generate metadata (Rodrigues & Milic-Frayling, 2011; Alvarez *et al.*, 2015). Telecommunications metadata comprises two classes of data – Call Detail Record (CDR) and Call Management Records (CMR) (Catanese *et al.*, 2013).

CDR records information about each call processed in calling network. In contrast, CMR contains information about the Quality of Service (QoS) and diagnostic information about the call. Specifically, information like amount of data sent, received, jitter, latency and losses. Different networks have varying fields in respective CDR. CDR data are usually cleaned by removing personal information of phone users. This is called anonymous. Anonymous CDR data safeguards violation of user privacy. Personalities and identities of people in a dataset are shielded from researchers. Abridged version of CDR content is shown in Table 2.3, and a typical raw CDR from a network database is provided in Appendix C.

Table 2.3: Abridged CDR Contents

Call	Base station	Calling number	Called number	Start time	Duration	Cost	On net call
1	nyc-1234	8881112234	9992223345	01/01/2014 14:35:23	38	3.8	FALSE
2	paris-2512	9992223345	8881112234	01/03/2014 18:35:23	100	10.00	FALSE
3	Chicago-3412	8882345678	8883345722	01/03/2014 18:40:30	50	0.00	TRUE

(Source: Ferrara *et al*., 2014)

Metadata has formed the basis for postulating users' social behaviour in social network outfits (Hulst, 2009). CDR was used in studying mobility patterns of mobile phone users (Ren *et al.*, 2014; Blondel *et al.*, 2015). Reasonable insights were gathered from the mobility pattern. Location, time and group of people involved in a particular event of interest were obtainable. When CDR holds relevant information on participants in OCGs, a suitable model is also needed to identify who are key stakeholders in the group.

2.4.3 Metrics for complex network analysis

Complex networks are subjects of evaluation. The assessments are done to verify empirically some of properties ascribed to types or classes of networks. Evaluation on complex network is classified into three: topological analysis, links analysis and nodes analysis. Each assessment requires specific metrics. Tools for each class of assessment are briefly discussed in this section to provide reliable relationships between some of these tools and this research work: detection of covert members in OCGs.

(i) Topological Analysis Metrics

Topological analysis metrics are tools for evaluating cliques, clusters, sub-networks or group of nodes. Table 2.4 presents descriptions about the metrics. Some of properties that classify a network into one group or others were taken from the topological metrics most especially random graph, small-world and scale-free networks.

Table 2.4: Topological Analysis Metrics

Metrics	Description
Clustering Coefficient	Measures the average "cliquishness" of a node within the graph or subgraph; estimates the degree to which a graph is modular that is, is organized hierarchically
Component	The most substantial connected subset of nodes and links; all nodes within the component graph are connected, either directly or indirectly, and none of the nodes has any connections to parts of the graph outside the component.
Connectivity	Measures the extent to which agents are linked to one another either direct or indirect routes; typically defined using the maximum or average path distance
Density	The ratio of the actual number of links in a network to the total possible number
Inclusiveness	Measures the total number of nodes in a network minus and sometimes the ratio to the number of isolated or minimally connected nodes
Path Length	Measures the typical separation between nodes
Size	Number of agents and links in a graph
Spanning Tree	A subgraph of a network that is, a tree that contains all nodes; of particular interest for weighted graphs is the minimum spanning tree. Note that the spanning tree minimizes the weights along all links in the network
Symmetry	The ratio of the number of symmetric to asymmetric links or the total number of links in a network
Transitivity	Measures the ratio of the number of transitive triples over the total number of possible transitive triples number of parts of length two; x, y, z is transitive if whenever x is linked to y and y is linked to z, z is also linked to x

(Sources: Ilachinski, 2005; Xu & Chen, 2008)

*(ii) **Link Analysis Metrics***

Link analysis metrics evaluate properties of edges connecting nodes. It can also be deployed for evaluating properties of links in dynamic networks like road network and telecommunication links (Holme, 2003). Likewise, it also provides reliability test for dynamic links such as transactions and relationships between criminals. Link analysis had been identified with criminal investigation and in provision of criminal intelligence to law enforcement agencies (Rostami and Mondani, 2015; Berlusconi *et al.*, 2016). Table 2.5 presents summary of metrics for evaluating links.

Table 2.5: Links Analysis Metrics

Metrics	Description
Capacity	Measures the load capacity of a link to carry information (or general "resource") of a given type
Duration	Measures the duration of a link; (permanent or transient) the decay of strength in time, general stability over time
Frequency	Measures how often a link is active
Multiplexity	The extent to which a link between two nodes represents multiple kinds of relationships
Strength	Measures the intensity of the relationship between nodes; communicative, emotional, degree of sharing, reciprocity
Symmetry	Measures the extent to which a link (or the information that it is a conduit for) is bidirectional
Type	Direct/Indirect, directed/undirected, weighted/unweighted
Visibility/Vulnerability	Measures the degree to which a link is vulnerable to eavesdropping, jamming, physical disruption or destruction

(Sources: Ilachinski, 2005)

(iii) ***Node Analysis/Centrality Metrics***

Centrality metrics are tools for analysing nodes' importance in networks. These tools are used to rank nodes and a node that has the highest scores in the assessment is usually picked as the most important of all nodes. Assessment of nodes analysis is based on centrality concept – that is, a node at central of a network structure. The concept ascribes importance to a node at the centre. The far way nodes to centre have low centrality scores. This phenomenon was formed on open organisations but adopted as well for criminal groups.

The concept can identify few central nodes in networks most especially when all nodes have different number of links connected to them. A node that has highest number of links or positioned at central of a subnetwork is identifiable. Table 2.6 presents some of the centrality metrics. But basic ones that connected with this work are briefly discussed.

Table 2.6: Centrality Metrics

Metrics	Description
Betweenness	Measures the extent to which a node mediates, or plays the role of "information broker" between two nodes or clusters of nodes
Brokerage	Measures a node's "brokerage" strength; that is, the degree to which a node manages the information flow between two or more groups that otherwise would not be linked
Centrality	Measures the degree at which a node plays a significant (or "central") role in a network
Closeness	Measures the extent to which a given node is "close to" other nodes in the network; typically defined by averaging over all possible paths to other nodes
Degree	Number of links to other nodes
Diversity	Number of links to a different node (where "different" means either that they are not linked to one another or otherwise represent agents that have different internal states
Eccentricity	Measures maximal distance between a given node and any other node in the graph
Effective network size	The measure is used in analysis of "structural holes" in-network; based on the supposition that links among a node's neighbours attenuate the sufficient size of that node's local network
In-degree	Number of directional links that point towards a given node
Isolation	Measures the degree to which a node is isolated, relative to others in the group to which it belongs
Out-degree	Number of directional links that point away from a given node
Prestige	Measures how influential a given node on the receiving end of information flow is; it is defined only for directions graph

(Source: Ilachinski, 2005)

(a) Degree centrality

This is the most straightforward approach for measuring nodes' importance (Freeman, 1978; Bright, 2015; Wang *et al.*, 2017; Jalayer *et al.*, 2018). It counts on the number of straight edges or links incident on the individual actors (Kitsak *et al.*, 2010). Often, a node with a high degree centrality score is referred to as a hub (Minor, 2012). In rumour spreading and epidemic models, a high nodal degree individual is called either influential spreader, influencers or influential node (Ren *et al.*, 2016; Sun *et al.*, 2016; Pei *et al.*, 2017). But in a criminal network, high degree actors are assumed to be network leaders – indicating actors that give orders, instruction to subordinates (Calderoni, 2012; Bright *et al.*, 2017). Degree centrality is defined as (2.4) (Keller, 2015)

$$(\quad)=\sum (,) \tag{2.4}$$

where ( ) is degree centrality of node ( ), n is total number of neighbouring nodes and (,) is connectivity rule between node ( ) and its neighbour node ( ).

(,) = 1 if and only if and are connected and 0 otherwise. Some leaders in drug cartels were not identified with highest nodal degree.(Calderoni, 2012; Bright *et al.*, 2015)

(b) Betweenness centrality

Betweenness is another socio-metric tool for node's evaluation. It tends towards position than direct number of links connected to nodes. This metric informs about number of times a node lies between adjacent nodes (Freeman, 1978; Belinda, 2010; Lee *et al.*, 2012). It seeks the path for flow of information within a network structure (Grassi *et al.*, 2019). A high betweenness centrality node act like a bridge connecting isolated lands that is, clusters of nodes or two subnetworks. Figure 2.11 illustrates high betweenness nodes in a network structure.

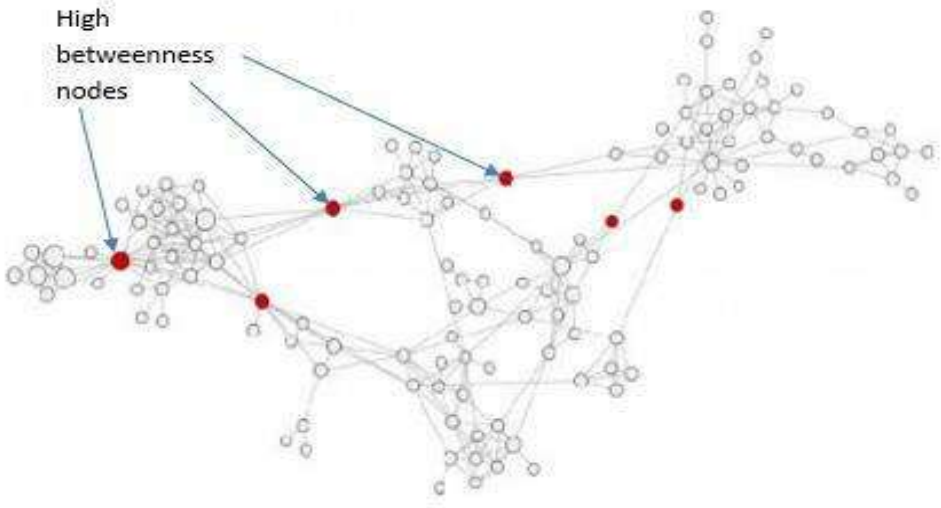

Figure 2.11: High Betweenness nodes in a Social Network
(Sources: Ahsan *et al.*, 2015)

Criminologists and network analysts are fond attacking high betweenness nodes in criminal networks. The nodes are target in order to disintegrate the network into smaller subnetworks when they are removed. This has become adopted strategy for disrupting criminal networks(Bright, 2015). Technically, removal of high betweenness nodes increase distance apart of disjointed nodes. It was opined that removal of high betweenness nodes does not cause total disruption of criminal network but it leads to disintegration of smaller groups with much difficult to curtail henceforth (Duijn *et al.*, 2014; Leuprecht *et al.*, 2016; Salvatore *et al.*, 2016; Bright *et al.*, 2017). Betweenness centrality is defined as (2.5) (Keller, 2015).

$$( )= \Sigma\Sigma \frac{( )}{} \tag{2.5}$$

where $( )$ is the betweenness centrality of a node , is the number of geodesic or shortest paths between two nodes and ; and $( )$ is the number of such paths containing and is the number of nodes in the network. It is also expressed as (2.6)

$$(.) = \frac{2}{(-1)(-2)} \Sigma_{\neq,} (,) \frac{( )}{( )} \tag{2.6}$$

where $( )$ $\overline{( )}$ denotes the probability that falls on a random selected geodesic connecting and .

A high betweenness node is also recognised as a broker. Brokers connect structural holes – isolated clusters (Reingen & Zinkhan, 1994; Hu and Mei, 2017). Drug market niche flourishes on hinge of brokers (Malm & Bichler, 2011; Berlusconi *et al.*, 2016). Brokers are hardly noticed when serving as connectors between two layers of a network (Everton, 2009; Bright *et al.*, 2017). Nodes connecting between two layers are safe from

criminologist's attention (Grassi *et al.*, 2019). The shield waxed when all links of a broker are unknown or inconspicuous. This gives a broker an advantage to becomes a non-centric and uninfluential node at sight (Matous & Wang, 2019). Maksim and Carley (2003) opined that a falling criminal group revamps the group by using unnoticed ties referred to as a 'sleeper' link.

Bright (2015) explored betweenness tool in a multi-relation network of a drug cartel on nodes connecting layers of a criminal network. Brokers that have potent to evade detection from one-mode social network analysis were detected. This tool works well with a network which multi-relational network and not a mode relational network that terrorist networks are characterised. The account in which elusory of key players in terrorist group is attributed (Du *et al.*, 2014; Grassi *et al.*, 2019). Multi-relation based network exposes other relations probably suppressed or omitted in a single-mode (Bright *et al.*, 2015; Wang & Zhao, 2015).

(c) Closeness centrality

Closeness centrality measures the proximity of a node to all other nodes in the entire network. It is technically defined as inverse sum of distances an actor has to all other actors on the closest possible paths. This is expressed in equation (2.7).

$$(\quad)^{-1} = \sum_{=1} \quad (\ ,) \qquad (2.7)$$

where (,) denotes number of shortest paths connecting and . Closeness centrality can also be expressed as equation (2.8) with (, ;) denoting the geodesic distance between and that pass-through g. Figure 2.12 illustrates a graph with some nodes that have high closeness values.

$$(,) = \frac{-1}{\Sigma_{*}(.:)} \tag{2.8}$$

A vertex that is closer to centre of a cluster gets higher closeness centrality score. This phenomenon is very prominent with a unimodal network. Nodes are referred to as "shallow" when get low closeness centrality scores as indication of their shortest geodesic distances to other nodes. It has not been applied or implemented in a multi-layer network.

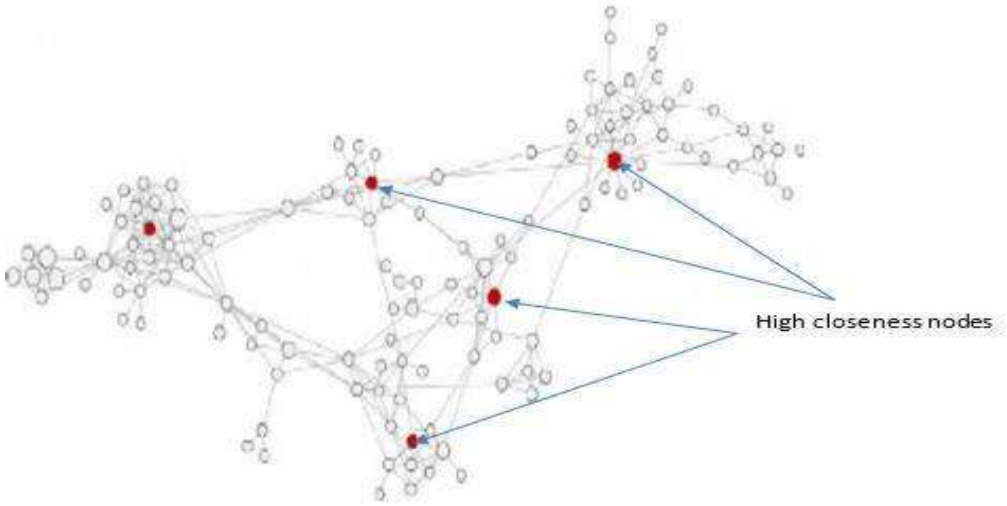

Figure 2.12: High Closeness nodes in a Social Network
(Source: Ahsan *et al.*, 2015)

(d) Eigenvector centrality

Eigenvector centrality was conceived on Hierarchical Organisational Structure (HOS) which is concerned profiling of members in open organisations to identify important personnel (Bonacich & Lloyd, 2001). This was deployed on unstructured datasets like internet, social networks, communication networks that form Flat Organisational Structure (FOS) (Carley *et al.*, 1998; Gregory, 2007; Clauset *et al.*, 2008). The concept replicates pyramid to justify hierarchies of members in traditional organisations.

Bonacich introduced eigenvector centrality, which is generalisation of degree centrality to identify influential nodes(Ahsan *et al.*, 2015). It is a phenomenon base on assumption that one's "importance" is not solely a function of number of people one

knows, but a function of how many people that one knows are themselves also relevant(Ilachinski, 2005). Each actor with importance raises significance of centrally connected actor. This corroborates profile of topmost node in pyramid or HOS. Eigenvector centrality of node is defined as ( ) in equation (2.9).

$$( ) \propto \sum \in ( ) ( ) \tag{2.9}$$

where ( ) denotes set of 's neighbours. It is also formalised as th component of the principal eigenvector of 's adjacency matrix, = [ ] defined in (2.10).

$$( ) = ^{-1} \quad \Sigma \tag{2.10}$$

where ( ) denotes eigenvector centrality of node , is a constant and is the largest eigenvalue. Equation (2.11) shows that eigenvector is proportional to the row sums of a matrix, M, equal to the sum of powers of A, weighted by corresponding powers of the inverse of eigenvalue.

$$= + ^{-1}{}^{2} + ^{-2}{}^{3} + \cdots = \sum ^{1-} \tag{2.11}$$

It is defined Chronologically, the next in scores follows the highest one. In principle, a node that attains the highest eigenvector score when its neighbours are also high in eigenvector values. This concept worth determines a leader in open organisation but not in OCGs(Hulst, 2009; Bright *et al.*, 2015).

2.5 Review of Related Works

Covert networks are social networks that often consists of harmful users. Memon *et al* (2011) defined covert networks as knowledge about the structure and organisation of terrorist groups. Due to the covertness of terrorist activities, the absence of individuals

and relationships often occurs in the construction of interpersonal relationship network of terrorists using social network theory, thus affecting the effectiveness of analysis.

Covert nodes literarily refer to unknown or hidden nodes in social networks (Paul, 2012; Smith *et al.*, 2013; Ren *et al.*, 2016). It could refer to critical members of OCGs. Covert nodes are indispensable actors from their roles. They are also non-vulnerable even under sophisticated detective techniques(Minor, 2012). Despite covert networks share features with conventional networks, its nodes are difficult to map due to masking of their activities like transactions and relationships. Terrorist groups are loosely organised networks with decentralised command structure (Berzinj *et al.*, 2012).

Criminal organizations work in small groups. Despite OCGs imbibe decentralised structure, they communicate, coordinate and conduct their campaigns in a network-like manner (Belinda, 2010; Dawoud *et al.*, 2010). There is a need to automate the collection of data and analysis of the terrorist network to find hidden relationships and participants. While overt nodes in a criminal network are known criminals - those responsible for physical operations and executing the group agenda. Majority of OCG's foot soldiers are expendable participants in OCG because their importance is limited. They are less involved in critical activities like planning and providing logistics. Security agents often apprehend overt members. But covert members who provide logistics for training of foot soldiers are hardly apprehended (Minor, 2012).

2.5.1 Related works on detection methods

Network analysis is a knowledge discovery process for identification of entities and properties from social networks. Most discovery processes are classified into supervised and unsupervised learning techniques (Hasan *et al.*, 2006). Both involved in classification of entities and properties. Unsupervised approach was described with low

compared to supervised approach (Lin & Chalupsky, 2003). Supervised learning approach requires trained dataset, sometime with attributes while unsupervised learning approach does not (Hasan *et al.*, 2006).

Parameters depict attributes used in discovery process (Lin & Chalupsky, 2003). Centrality constitutes parts of these attributes. Centrality attributes are graph-based parameters. These are not actual attribute but being deployed in lieu of genuine ones (Lampe, 2009). There is challenge in depicting nodes attributes accurately using graph (Maksim & Carley, 2003). An unsupervised framework employed directional semantic approach for description of nodes and links as it cater for any node has more than one relationship with others (Lin & Chalupsky, 2003).

The behaviours of nodes are analysed based on the semantic profile generated (Kriegler, 2014). The semantic profile deals with collection of condensed paths generated through variable relaxation approach(Hasan *et al.,* 2006). Condensed paths are path types with unique format. Identification of key players are chosen by considering outlier. Often, preference is given to highly communicated node as the most influential node (Butt *et al.*, 2014). This considered one side of outliers as a measure that minimize false detection. The remaining part of this section presents methods and algortihms for detecting covert nodes from literature.

(i) Review of Works on SNA-based Detection

Social Network Analysis (SNA) is a resourceful tool to security agents as it supports criminal intelligence (Sparrow, 1991; Klerks, 2001; Basu, 2014; Kriegler, 2014; Berlusconi *et al.*, 2016; Burcher & Whelan, 2017). Most of techniques presented here depend on SNA metrics and they are classified into single centrality metric and mult-centrality metrics approaches. Multi-centrality techniques involved plotting of two

centrality metrics on x and y axes. The combination aimed at using strategic positions and vulnerability to identify important nodes (Calderoni, 2010; Morselli, 2010; Bright, Greenhill and Ritter, 2015; Bright *et al.*, 2017). SNA-based techniques are reviewed as following:

Kreb (2002) deployed four centrality metrics to identify important terrorist cells that is, participants in September 11 2001 attacks. The four metrics identified different actors – assumed to be key players. Kreb used open-source data usually categorised as unreliable.

Borgatti(2006) was an inventor of Key Player Problem (KPP). The concept was developed to identify set of key players from graph. It was realized that all key players are not bound to central of network. The KPP concept was observed on two domains: KPP – Positive (KPP-Pos) and KPP – Negative (KPP Neg.). KPP-Pos deals with connectivity such as social network of viral market, fast transmission, rumour spreading and epidermic disease, while KPP-Neg concerns network disruption. KPP shows that key player's position differs – that is, positions of key players are not fixed nor tied to centrality concept. Network disruption is unachieved when only set of nodes that have highest centrality scores are removed. KPP emphasized on node's positions and not on personality.

Lampe (2009) proposed incorporation of human-capital attribute that is, personality to augment and complement detection of important nodes through social-capital attributes. This intended to offer concrete evidence to support social-capital detection rather than using speculation. Human-capital attributes are inaccessibility and difficulty to deploying.

Morselli (2010) proposed and developed concept of vulnerability and strategic positions for identification of vulnerable and less-vulnerable actors. Degree centrality denotes

vulnerable attribute while betweenness denotes strategic position. Four quadrants were obtained after plotting vulnerability and strategic positions(Morselli, 2010). Two upper quadrants are point of interest (poi) to Morselli. Nodes in those two quadrants were counted as vulnerable while nodes in the other two were regarded as less-vulnerable. The problem of hidden relationships was still unresolved. Legitimate actors also remain undetected (Hulst, 2009).

Karthika and Bose (2011) deployed multi-centrality approach on heterogenous nodes in the 9/11 terrorist network. Nodes with highest centrality scores was identified as covert nodes. These include actors and a place frequently visited by terrorists(Karthika & Bose 2011). Preference is given to nodes with conspicuous relationships that is, links.

Berzinj *et al.,* (2012) targeted financial manager in a terrorist network. Berzinj et al 's model revolved around terrorist operational activities; where finance manager takes central part of terrorist structure illustrated with Figure 2.13. Five different new centrality metrics were developed but unexperimented.

Figure 2.13: Decentralized Structure of a Terrorist Organization
(Source: Berzinj *et al*; 2012)

Campana and Varese(2012) used single centrality metric for ranking phone conversation of mafia involving in drug peddling or DTO. It had a tremendous result. The highest

hierarchy was assigned to actor that has the longest period of conversation as well as call-rate while low vulnerability was ascribed to an actor with low phone conversation. The approach played down on low conversation of legitimate actors.

Husslage *et al.*,(2012) exprimented flat organisational structure (FOS) with flat structural-like figures to show that key players in terrorist groups cannot be identified with ordinary metrics. The work craves for sophisticated appraoch to identify covert members. It was inferred from correlation results that terrorism groups are leaderless group and also imbibe FOS to salvage the group.

Catanese *et al.*, (2013) explored mobile phone traffic calls of actors in a criminal network. Many statistical analysis tools were incorpoated in LogAnalysis. The suite generated hierarchies of phone users in metadata. It also displays visual graphs of actors. LogAnalysis can handle large datasets effectively. Hierarchies and frequency of calls are determinant of key players. Conspicuous links and call time-frame are setbacks. LogAnalysis lack capacity to detect a key actor with low call-rate.

Ferrara *et al.*, (2014) offered a theoretic framework on key players problem. Phone call records is used in reconstructing network structure of participants in criminal activities. Basic centrality metrics were incorporated in expert system. The application can compute and costruct hierarchies like LogAnalysis. Both expert system and LogAnalysis cannot detect legitimate actors that is, a key player with relatively low phone calls cannot.

Butt *et al.*,(2014) approach was directed towards key players that can evade detection. Only degree centrality metric was deployed in the approach. The metric identified at least one actor from each layer of multi-relations network. The method identified an actor that had highest bank transaction. But that actor was among nodes that had low degree in the calling network. No further analysis was given on personalities of detected nodes.

Bright *et al.*, (2015) employed centrality metrics to rank resources of participants in DTO. The target was to find resources that make high-profile participants become vulnerable. Eight resources were ranked and classified as tangible and intangible. The weight scores of resources were plotted against degree and betweenness separately using concepts of graph in (Morselli, 2010). Although, some legitimate actors featured in detection of prominent. The ranking scheme was biased and prevalent actors in detection are actors that have high centrality scores in both measures.

Bright *et al.*(2015) presents another scheme for detecting set of key players who use to lie between layers of a criminal network. Multiple links are indispensable in OCGs. This denotes multi-relational network. Each link or network layer facilitates transfer of resource like money, drug and information. Each was examined using (Morselli, 2010)'s scheme to identified actors occupying strategic positions by connecting layers of a criminal network. Some of targets actors are potential legitimate actors if single link network was analysed.

Ozgul (2016) analysed terrorist network structure – topology. The analysis was taken from historical background of terrorist organisations. Positions of key players was analysed with centrality metrics. Analysis was less concerned about terrorist datasets, dynamic behaviours and probability of structural changes. This analysis was appropriate because few key players are still playing to the gallery of centrality that is, they submissive to centrality concept. Yet, there is high chance of failure if caution is not applied.

Gunnell *et al.* (2016) also analysed street gang using Morselli's scheme (Morselli, 2010). The scheme was deployed on datasets of gangsters and victims. Four quadrants in the scheme were designated as characteristic of people within each. The upper two quadrants

were tagged ‘Gateway’ and ‘Central participants’, while the lower two quadrants were designated as ‘Peripheral’ and ‘highly visible’. Gunnell *et al.*, point of interest are gateway and central participants. The datasets of victims were not separated from gangsters and the personality of emerged gangs were not given.

Grassi *et al.*,(2019) deployed eight different betweenness centrality algorithms for detection of network leaders. The scheme was deployed on participants’ attendance in the mafia meeting. The detection was validated by correlation between results of algorithms that is, nodes commonly detected. Some close-associates of leaders were ranked significantly high few schemes. Unfortunately, these actors fell out of Grassi’s scope.

Ismail *et al.* (2019) deployed SNA-Quadrant (SNA-Q) for detection of smart criminals in OCGs. The scheme is illustrated with Figure 2.14 with its four quadrants and features.

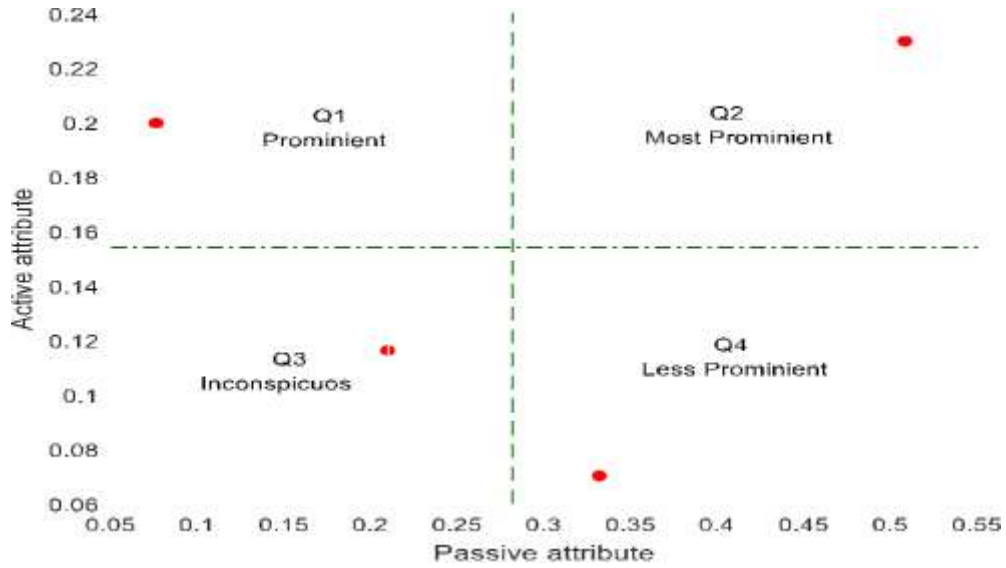

Figure 2.14: SNA-Quadrant model for Classification of Importance

The four quadrants in Figure 2.14 denote derived attributes used for defining relevance of participants in OCGs. Q1 designates actors who are ‘high’ in indicting activities but ‘low’ in covert roles. Q2 represents actors that are ‘high’ in both overt and covert roles. Q3 designates actors who are low in both overt and covert roles that is, low

in indicting activities and that of criminal group survival. An example of these actors is an insider to a crime victim group.

Finally, Q4 represents actors who are low in overt roles but high in covert roles that is, having low participation in indicting activities but highly involved in activities relevant to group survival. This description illustrates an important participant that is less-vulnerable. Such actor is therefore referred to a 'smart actor'. The point of interest (poi) is fourth quadrant - that has high active attribute and low passive attribute on y and x axes respectively. The poi adequately describes characteristics of legitimate actors in OCGs (Hulst, 2009) as illustrated in Figure 2.14. The summary of the reviewed works is given in Appendix A – Meta-analysis of SNA-based Algorithms.

(ii) Review of Works on Non-SNA-Based Detection

This section list set of works detecting covert nodes that did not compute SNA – they are regarded as non-SNA-based detection. Missing nodes and Node Discovery are prevalent here.

Zhao *et al.*, (2006) proposed the use of Relational Markov Networks (RMN) to describe the entities and the relations among them in an affiliation network. The work used Profile in Terror (PIT) database to study the entity and relationship labelling. The technique had no means of detecting covert members but PIT is vital data for analysing members of defunct terrorist groups.

Maeno and Ohsawa (2007a) proposed and developed heuristic method for analysing covert social network behind terrorism. The technique employed set theory concept to identify latent structure that is, key player's position. Records of terrorist's attendances were worked on. An actor that appeared in more than one record of attendance was

identified as occupier of latent structure. The scheme used a scarce dataset for terrorism and only actor with conspicuous links is detectable.

Maeno(2007) developed Statistical Inference Method (SIM) for identification of latent structure. The algorithm tends towards identifying elusory actors due to using medium of communication not well-known to influence a criminal group. Inability to identify the medium used by affiliate criminal who stir and influence organisation's affairs was referred to as a latent structure. The structure makes affiliate becomes less vulnerable to detection.

SIM computes maximal likelihood of all actors and identify outlier as latent structure. The interactive process of the node discovery algorithm is shown in Figure 2.15. The approach was highly classic but mathematically unrealistic for OCGs. Ranking and clustering procedure along with expert investigator knowledge enhanced understanding in calculating suspicious inter-cluster relationships. But the SIM is found to be too sophisticated. It makes the system unrealistic.

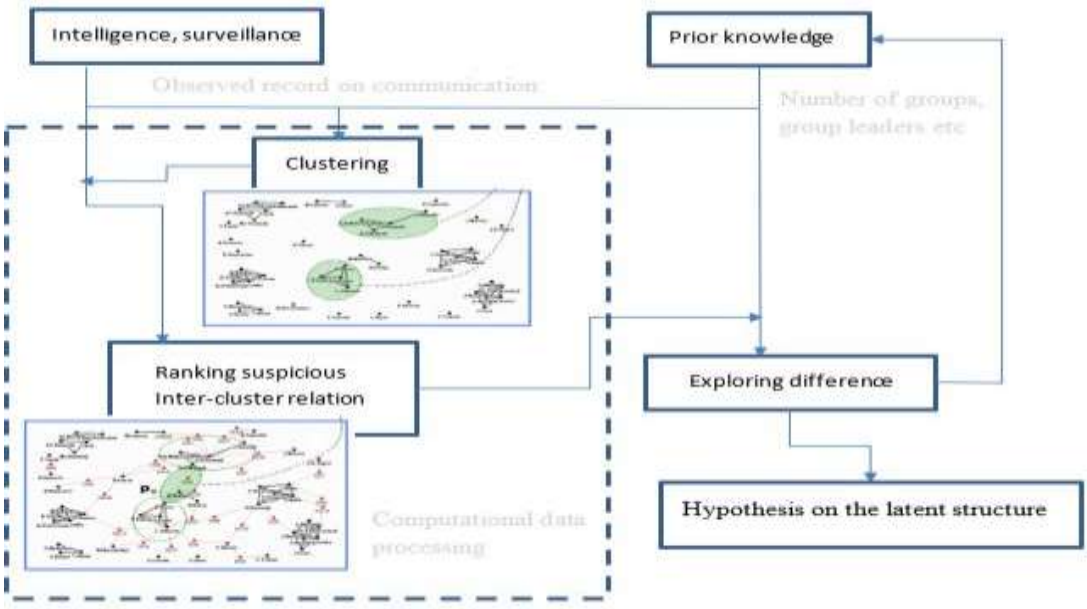

Figure 2.15: Interactive Process for Detection of Latent Structure
(Source: Maeno, 2009)

Eyal *et al.,* (2011) addressed the issue of covert nodes from missing information that is, identify nodes that were not known before. Clustering was deployed on nodes that have similar attributes. The problem was structured on a factor that can help in identifying similar nodes or set of nodes to be clustered. The method runs from nodes clustering to computation of affinity matrix that is, level of similarity between missing nodes. Marginal nodes are of less advantage.

Figure 2.16 illustrates the overview of missing node problem. Nodes in the cloud denote participators in online service or apprehended criminals. Nodes outside the cloud represent target friends need to be identified by security operatives, unknown criminals or co-offenders who are yet to be apprehended.

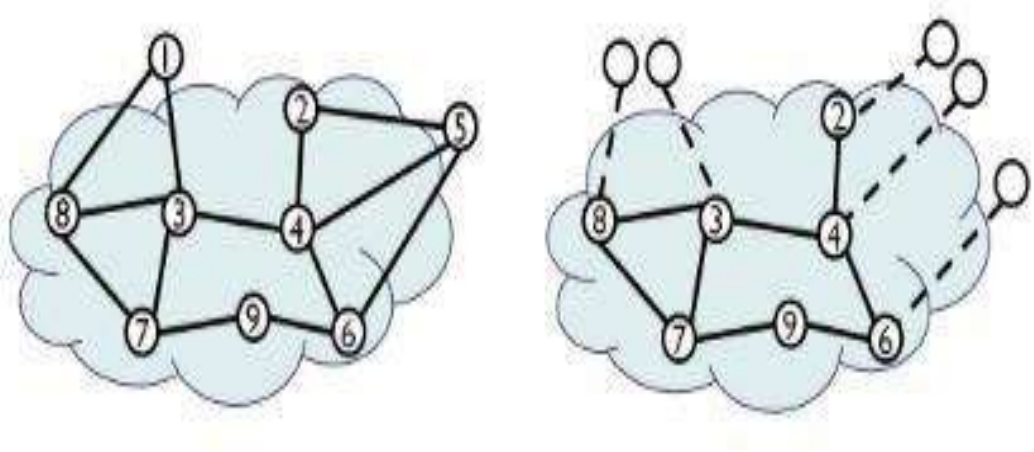

Figure 2.16: Formulation of the Full network and that of Missing nodes
(Source: Sina *et al.*, 2013)

Sina *et al.,* (2013) addressed the challenge from specific node attributes by implementing Structure and Attribute Missing Node Identification using Attribute's similarity- SAMI-A and Structure and Attribute Missing Node Identification using Social-Attribute Network -SAMI-N. The missing information is a deplorable concept when attributes of concerned nodes are easily accessible. The algorithm is suitable and adequate for online marketing especially when identifying more friends of known customers to be lured into online services by online services agencies.

The missing nodes scheme does not address node's importance. Insider nodes – that is, nodes in the cloud - were used to identify outsider nodes or additional unknown nodes. The method takes care of data defectiveness associated with criminal organisations. However, it is not a robust scheme for high-profile terrorist who hardly share attribute with overt terrorists. There is high tendency of false alarm for outsiders that accidentally share attributes with insiders. Table 2.7 presents summary of meta-analysis on reviewed work on None SNA-based techniques.

Table 2.7: Meta-Analysis of Non- SNA-based Algorithms

Authors	Title	Method	Strength	Weakness
Zhao *et al.*, (2006)	Entity and Relationship Labeling in Affiliation Networks	Using Relational Markov Networks (RMN) to investigate relational classification	Profile In Terror (PIT) was identified as a dataset that has relational structure to study affiliation. It is a useful for statistical relational networks	It did not identify covert nodes
Maeno and Ohsawa(2007a)	Analyzing the covert social network foundation behind terrorism disaster	Searching for latent structure of covert nodes; set theory was deployed as approach	Detect latent structure for influential nodes; Detect most influential nodes, multiple records of attendance are required	Only few actors were detected; Marginal nodes were undefined; legitimate actors are evasive;
Maeno (2009)	Node discovery in a networked organization	Searching for influential nodes, using heuristic method with data crystallization and Statistical inference method (SIM)	identify influential actors; multiple data records of crimes were explored; interactive mode and expert knowledge were involved	Mathematical complexity demerit; concern highly recurring nodes; marginal actors are undefined
Eyal *et al.*,(2011)	Identifying missing node information in social networks	Deployed missing link problem approach to detect covert nodes; clustering algorithm; nodes similarity for missing node identification	Attribute's similarity favoured high nodal actors; Accessibility to data: structural attribute in lieu of personal attribute; high affinity for central node	High prone to false alarm; marginal nodes have low similarities in a cluster; evasiveness of legitimate actors
Sina *et al.*,(2013)	Solving the missing node problem using the structure and attribute information	SAMI-A and SAMI-N were developed and applied to identify covert nodes in any structure	It is highly deplorable to social marketing and advertisement; Only popular actors are used to identify unknown actors but not unpopular actors to detect influential node	high-profile actors in OCG hardly share attribute with overt criminals; Legitimate actors are undefined in the scheme; difficulty in access personal attributes of actors

(iii) ***Review of Works on Inference-Based Detection Techniques***

The techniques articulated here employed inferences. Inference techniques are probabilistic-based methods. Probabilistic models aim at abstracting the underlying structure from the observed network. For predicting missing links, inference uses learning of patterns in a given network G = (V, E). After that, the learned patterns are used for building a target function through optimization of observed or given parameters that established the model composed of a group of parameters Θ (Lü & Zhou, 2011; Sharma
& Singh, 2016). The inference has been applied to missing information - link prediction (Rhodes & Keefe, 2007), and it was also applied to node discovery (Maeno & Ohsawa, 2007a; Maeno, 2009).

Hasan *et al.*, (2006) focused on link prediction from the supervised learning perspective. The work addressed link prediction problems using classification techniques. It used the information gain, gain ratio and average rank as performance metrics. The work offered list of features required for link prediction in co-authorships domain and terrorist domain. Unfortunately, there was no trained datasets for evaluating terrorist dataset due to the incompleteness and fuzzy boundaries.

Rhode and Keefe's developed a Bayesian-based algorithm for prediction of links (Rhodes
& Keefe, 2007). The prediction was carried out on two classes of problems: (i) predicting deliberately omitted links - those that were removed and (ii) predicting future links – links currently not existing. It was designed to pre-empt the growth of covert organisations. Equation (2.12) defines the likelihood of newly predicted or omitted links. As likelihood predict unknown links, it shows how the covert network will grow. The likelihood is calculated using probabilities of positive (|) and negative (|).

$$(\ldots) = \prod \frac{()}{} \quad (2.12)$$

where (1 ...)is the likelihood of a predicted link when positive probability (|) is larger than negative probability (|)

It was claimed that prediction is uniformly distributed throughout the network. The prediction was contrary to rich get more precious phenomenon in the small-scale network. The Bayesian-based link prediction is not a useful tool for detecting omitted links between a network leader and high-profile nodes.

Hussain and Ortiz-arroyo (2008) used Bayes theorem to infer actors' importance. The posterior probability was used to calculate node's entropy in the network. The algorithm followed Shannon's entropy and that entropy variations. Entropy of the network was computed first, then entropy of individual nodes was computed by subtracting entropy computed when a node was removed from entropy of the entire network. A node identify with the highest degree of uncertainty or lowest entropy value is picked as important node. The algorithm measured uncertainty of a node that was not in the network.

From Bayes' theorem of (2.13)

$$(|) = \frac{(|) ( )}{()} \quad (2.13)$$

Denominator ( ) of equation (2.13) was expanded to give (2.14)

$$(|) = \frac{(|) ( )}{() () + () ()} \quad (2.14)$$

where ( ) is prior probability or marginal probability of node regardless of any information which is computed by considering the total number of nodes present in the network, ( ) is the probability that the node is not a key player given by (1 − ( )).

(|)is the conditional probability of given , meaning that the node is a key player, which is computed based on the number of links incident on that particular node, so if there are nodes in the network then to be a central node of the network, it has to be linked with other (− 1) nodes.

(|) is the conditional probability given N meaning that node is not a key player, which is obtained by computing (1 − (|)). This method provides a simple and straight forward way for computing nodes' entropy - using Shannon's entropy as uncertainty.

Serin *et al.*, (2009) developed sensitivity analysis technique for analysing social networks of cities. The technique depends on Shannon entropy definition of random variable X. Entropy of nodes in the network is computed using three centrality entropies defined degree entropy, betweenness entropy and closeness entropy (2.15) to (2.17).

Degree entropy:

$$( ) = \frac{( )}{\Sigma_{} ( )} \tag{2.15}$$

where ( ) is probability mass function, E is the number of edges in a graph, probability mass function ( ) of node is computed using ( )- number of links connected to over sum of links in entire graph.

Betweenness entropy:

$$( ) = \Sigma \neq \neq \frac{( )}{} \tag{2.16}$$

where ( ) is betweenness probability mass function, between nodes and is the geodesic distance

Closeness entropy is defined as (2.17):

$$( ) = \frac{-1}{\Sigma \quad (.)} \qquad (2.17)$$

where ( ) is closeness probability mass function denoting closeness entropy and is

length of geodesic from/to a given node .

Sensitivity of all nodes were computed using set of defined entropies. Nodes that are important in the network were inferred from sensitivity computation. Each entropy was presented in a visualization form. Combined approach was also used to present different effects each node under different entropy. For combine approach, all entropies were multiplied. Visualisation was used to show each entropy and significance of nodes.

Entropy is average information a node has within a network. With entropy, a node that has the highest impact definitely will have low entropy computation. This indicates that entropy of entire network fall when the node was not in the network. Entropy algorithm requires that entropy of a network is first computed with all involved nodes. Then, subsequent entropy computation is for individual node in the network. The computed entropy of a node is then subtracted from the entropy of network. The metrics include degree entropy, betweenness entropy, and closeness entropy.

Ortiz-arroyo Daniel simplified Borgatti's definitions of Key Player Problem (KPP) positive and negative (Ortiz-arroyo, 2010). Two new metrics were developed by Ortiz-arroyo to replace the metric used in KPP definition(Borgatti, 2006). The metrics were used in Shannon's entropy expressed as equation (2.18).

$$( ) = -\sum \quad ( ) \times \log_2 ( ) \tag{2.18}$$

where ( )defines Shannon entropy of a discrete random variable has, ( ) which is probability mass function of state for a system with different states.

( ) and ( ) are two probability distribution used in (2.18). Each replaces manual KPP metric. ( ) is connectivity of a node defined by ( ) = $^{\deg(}{}_2{}^{)}$ where; ( ) is the number of incident edges to node and N is the total number of edges in the graph;

defined by where

h ( ) is the number of shortest path from node to all other nodes in the graph and h (1, 2, ...) is the total number of shortest paths M that exists across all the nodes in the graph.

This led to two simplified tools for implementing KPP and get an optimal set of key players for disrupting criminal networks or to identify node that is high in transmitting information. Equation (2.19) defines and equation (2.20) defines

$$( ) = -\sum \quad ( ) \times \log_2 ( ) \tag{2.19}$$

$$( ) = -\sum \quad ( ) \times \log_2 ( ) \tag{2.20}$$

(2.19) and (2.20) are still under structural equivalence problem emphasized in. The set of nodes detected distinct from other nodes structurally. It is observed that disruption of networks cannot be achieved by removing only central actors. That is, some peripheral actors are also significant to network disruption. The summary of inference-based Algorithms and Techniques is presented in Table 2.8.

Table 2.8: Meta-Analysis of Inference-based Algorithms and Techniques

Authors	Title	Method	Strength	Weakness
Hasan *et al.*,(2006)	Link prediction using supervised learning	predicting link through a supervised learning approach; classify attributes for prediction of links	identify possible links among actors; nodes that have the same proximity can be identified	it meant for links and not nodes discovery;
Rhodes and Keefe(2007)	Social network topology: a Bayesian approach	Using Bayesian to compute likelihood of missing and future links in covert social networks	It accurately predicted more deliberately removed links;	Link prediction lack capacity to identify important link; it fails in identifying key players
Hussain and Ortiz-arroyo (2008)	Locating key actors in social networks using Bayes' Posterior Probability Framework	Entropy of nodes was computed using Bayes probability theory; Uncertainty of nodes was used to infer important actors in the network	Node that has large number of links was identified. It was significant with conspicuous links	Key players with low links are undefined;
Serin *et al.*,(2009)	Entropy-based sensitivity analysis and visualization of social networks	Shannon entropy model was used for computing sensitivity analysis of nodes; deployed three centrality measure entropies	Central nodes and node with high direct links are prevalent in different entropies;	Visualization was poor; the connection between sensitivity and colour visualization was unknown;
Ortiz-arroyo,(2010)	Discovering sets of key players in social networks	Entropy centrality and Entropy connectivity	Effective on KPP Pos. and KPP Neg.	Unverified personality; marginal nodes are undefined;

2.6 Research Gaps

A number of techniques for detecting covert nodes had been reviewed. The state of art on covert nodes detection shows that there are still open areas. Covert members in OCGs require techniques that can detect key players from marginal or peripheral. Another reason for new technique is that most techniques cannot handle inconspicuous relationship that OCGs are fond of. Inconspicuous relationships pave way for key player's evasiveness. Lastly, some key players evade detection as a result of forming 'structural equivalent' with overt members. Key players are difficult to distinguish from overt members when both have the same structural property. Figure 2.17 summarizes the research gaps. Key players in OCGs are not central elements like other application area. 59

Therefore, there is need for a technique or an algorithm that can differentiate key players from set of overt members – identify key player hiding with marginal nodes.

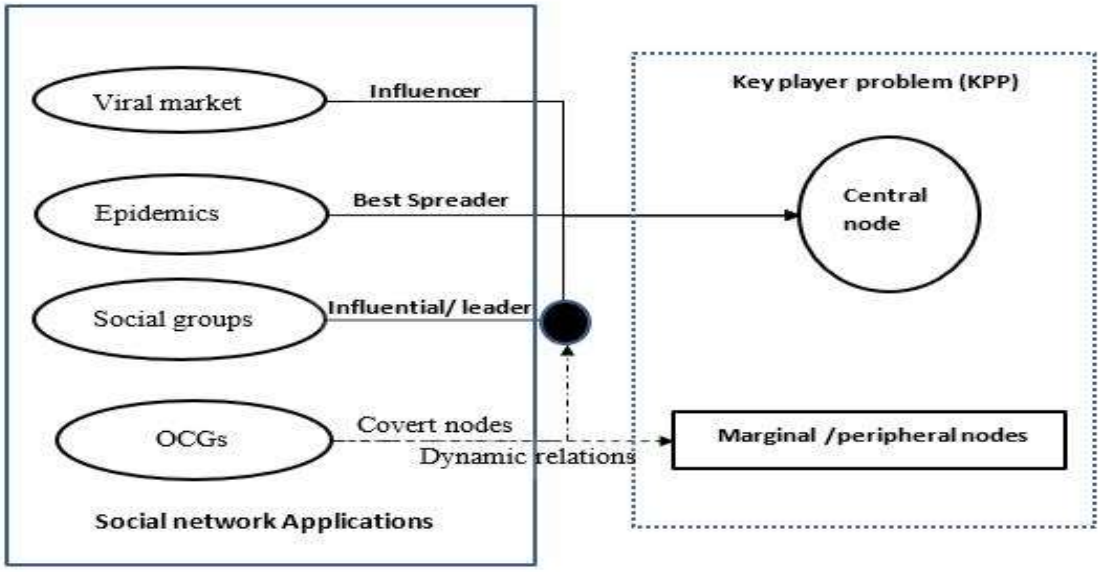

Figure 2.17: Application of Central nodes and Key Players

2.7 Uniqueness of the Study

It manifested from research gaps that OCG's key players run away from central positions in order to evade detection. Key players to OCGs are not only regular members. It extends to outsiders who provide financial aid or supply ammunition or an informant to criminal group. Roles of outsiders demand no central position. Operating and collaborating from outside make them less-vulnerable to detection.

Figure 2.18 illustrates relationship that make an outsider key player less vulnerable to detection. All nodes in the cloud represent regular members or foot soldiers of OCGs. Node 5 is an outsider hardly known to other regular members except node 4. All nodes in the cloud are vulnerable but node 5 is not. Detecting and removing node 5 can disrupt OCGs. Participant with this feature is regarded as legitimate actor.

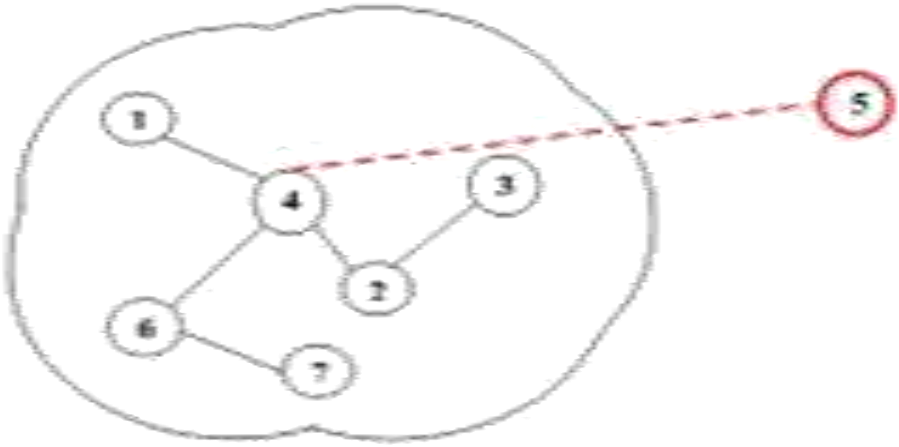

Figure 2.18: An affiliate member's relationship with a terrorist network

This type of actor was not class of nodes detected in reviewed works. Conspicuous relationships fail on low key players. Bayesian model had better prediction even with any degree of uncertainty data. This capacity is yet to be explored on actors with low susceptibility. Due to inaccessibility to multi-relational data of terrorist organisations, different datasets with varying degrees of uncertainty need to be experimented.

2.8 Chapter Summary

This section presents different models for detecting covert nodes. All reviewed works and their techniques were grouped into SNA-based, non-SNA-based and inference based. All algorithms were able to detect a node or set of nodes as covert nodes or key players. It was observed that all identified nodes are central nodes and low degree centrality were not at advantage of being detected. It was only in Quadrant approach that low-degree actors emerged as key players.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Preamble

This chapter presents procedures for actualizing the research aim and objectives in this thesis. The chapter is divided into two sections: materials and methodology. The material discussed hardware and software used in actualizing the research aim. The methodology section contains all sub sections for actualizing the set of objectives. There are three major subsections. The preliminary sections include: acquisition of data, construction of network graphs, and extraction of network attributes. The two main sections are Bayesian network model (BNM) and SNA-Q model. The BNM has three sub-processes: development of Bayesian model for covert nodes detection, development of Enhanced Bayesian Network Model (EnBNM) algorithm and experimental evaluation of EnBNM algorithm, while SNA-Q model also has three sub-processes: application of SNA-Q model, SNA-Q algorithm and experimental evaluation of SNA-Q algorithm.

3.2 Materials

All experiments are implemented using MATLAB R2015a and Python library conducted on 1.90GHz Intel® machine with 6 GB RAM. The operating system is Window 10 edition. Two real terrorist datasets: N'17 and 9/11 network groups are used for experimenting the developed model. The datasets are described as follow:

1) N'17 dataset(*UCINET Software -17 November Greece Bombing*, 2017) is a relatively small dataset of a criminal group. The November 17 has twenty-two entities. And each entity represents a participant. The twenty-two participants involved were apprehended and they were also prosecuted.

2) 9/11 dataset (*UCINET Software - 9/11 Hijackers*, 2017) is also a relatively big dataset in terrorist domain. The dataset for September 11 terrorist has 60 entities and one hundred and ninety-nine (199) edges. Nineteen (19) out of sixty participants were attackers and the remaining forty-one (41) participants were conspirators in the dataset.

3.3 Methodology

Method for detection of covert nodes is presented here. The target covert nodes are those in OCGs. These covert nodes are special and different from those in other application areas. These are human being participating in illegal activities. They either benefiting directly or indirectly from crimes. It is important to participants in OCGs to provide measure to escape security operative traps. Regular members are easily becoming vulnerable but affiliate members are not.

Dataset is only available mean to catch-up with affiliates participants. Unfortunately, profiles of participating nodes are not indicated, this still give way for high-profile members to evade detection. In order to lower evasion of key players, prediction was proposed. The prediction is to give all nodes the same opportunity to be predicted. In order to ensure high-profile member is the one predicted, status of picked nodes is checked to validate the detection. The comprehensive procedure of method for detecting covert members in OCGs is presented in the flow diagram Figure 3.1.

The first three processing blocks: acquisition of dataset about OCGs, construction of network graph and extraction of network attributes of nodes are stages for obtaining data for experimenting algorithm of EnBNM and SNA-Quadrant (SNA-Q) model. The SNA-Q model was a validating tool deployed.

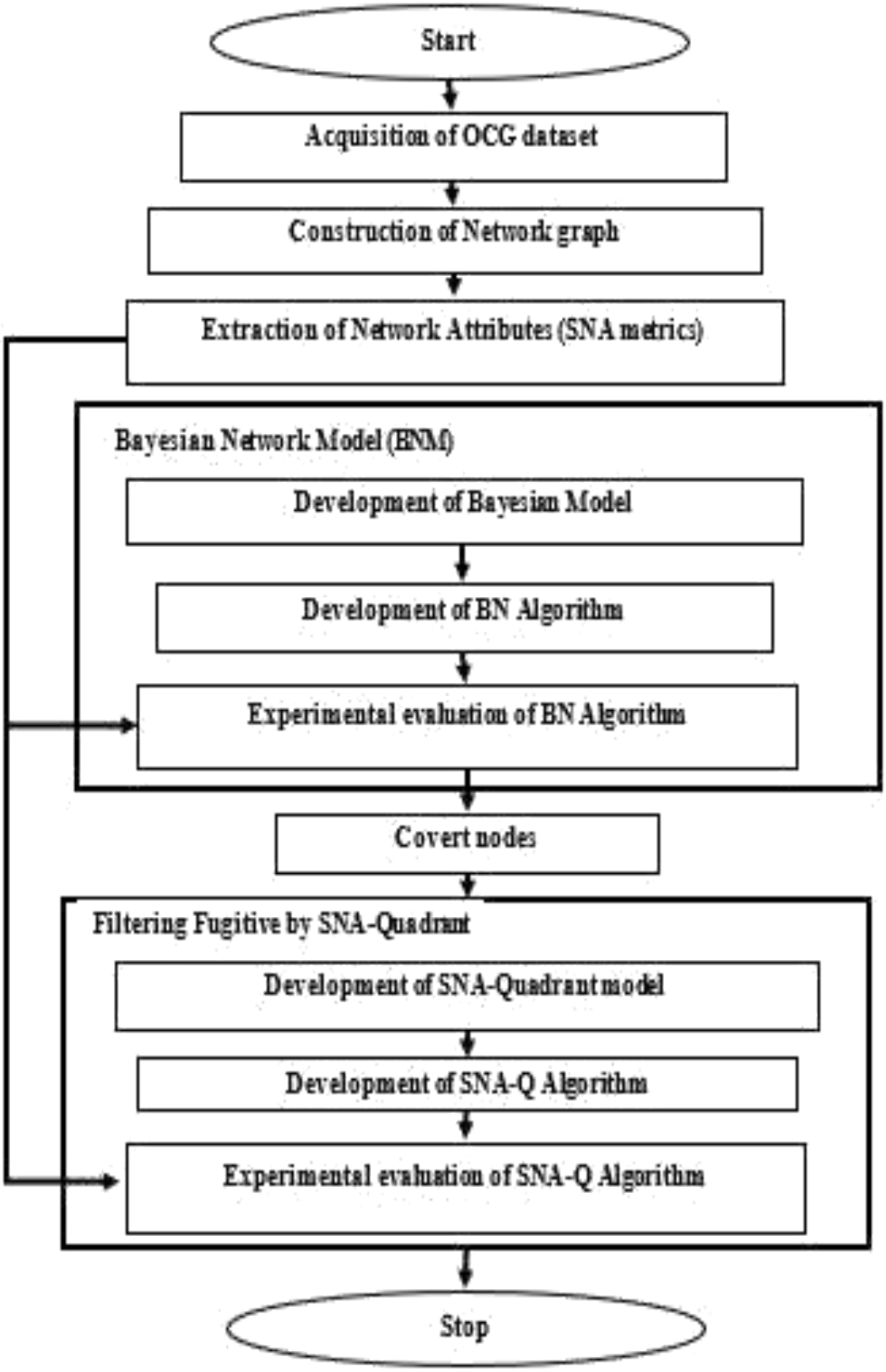

Figure 3.1: Flow Diagram of Methodology

3.3.1 Acquisition of OCG's dataset

Acquired datasets are forms information about participants in OCGs. It is very sensitive information and it is out of the scope of this work. The researcher makes use of datasets available on UCINET site (www.ucinet.com). The site has a number of datasets on defunct covert networks. Two of these datasets were downloaded and use them to test performance of EnBNM. The dataset is available in comma separated value(csv.) format.

3.3.2 Construction of network graphs

After acquiring criminal datasets, the datasets were plot into network graphs using Python library. Participating individual were depicted as a point. The relationships of participants are links connecting two points together. From the network graph, relationships among the participants are easy to navigate. This network graph is a unimodal network graph. Detail and type of links connecting two nodes are not specified. The relationship between the points is expressed as equation (3.1). This defined a terrorist graph where; is a network graph of phone communication. is a subset of large graph G.

$$=(,) \quad (3.1)$$

where denotes set of suspected and apprehended terrorists and denotes set of links between terrorist nodes. Equation (3.2) gives categories of participants in terrorist activities.

$$= \{ 1, 2 \quad \dots \quad \} \quad (3.2)$$

$\in$ consists of : regular terrorist, :

Such that any network leader, : critical conspirators and : sleeper partners. The network graphs shown in Figures 3.2 and 3.3 represent visual relationships among participants in each terrorist dataset as described in (section 3.2). Figure 3.2 is a graph of relationship among participants in the N'17 criminal

network and Figure 3.3 is a graph of relationship among participants in the 9/11 criminal network.

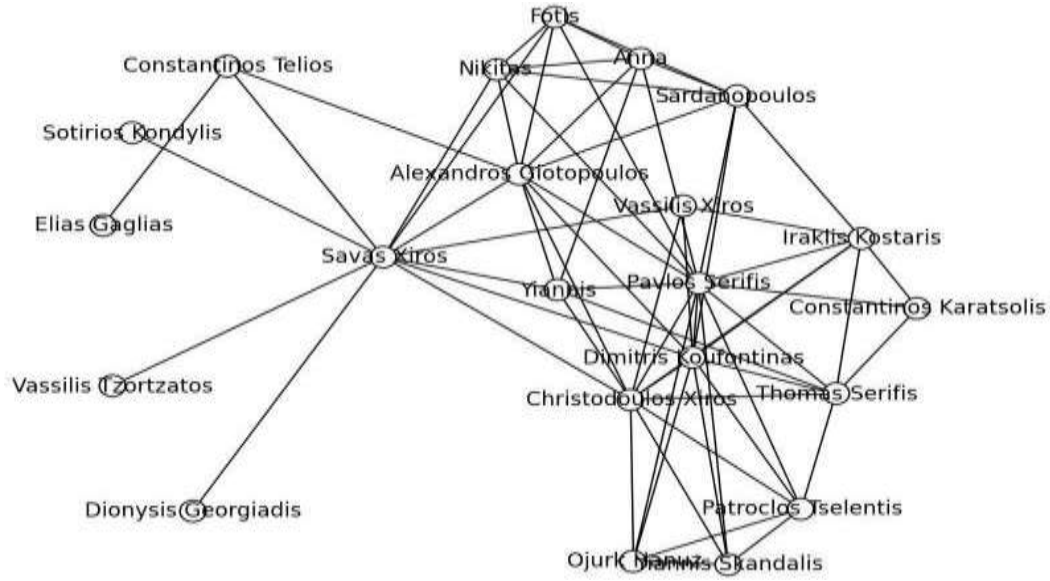

Figure 3.2: Network Graph of actors in the N'17 Greece Revolutionary Group

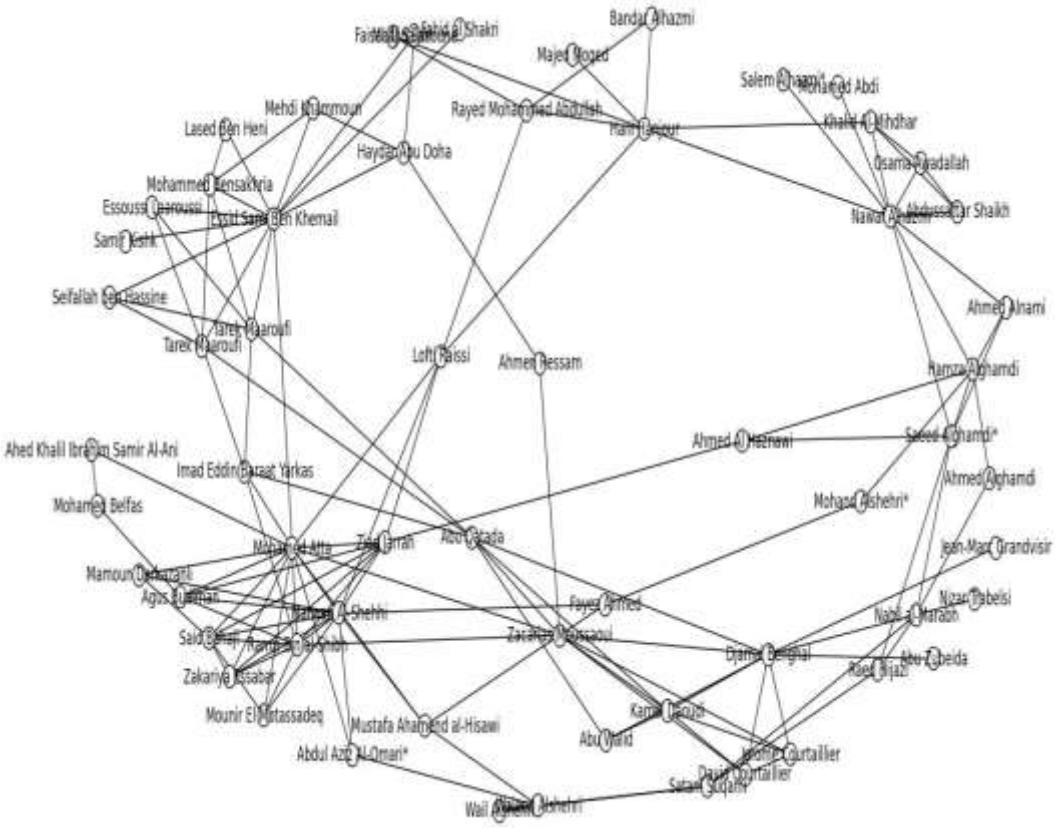

Figure 3.3: Network Graph of actors in the Al Qaeda 9/11 attacks

3.3.3 Extraction of network attributes

Attribute of participants is the actual data needed for evaluating model. Network attributes were adopted in lieu of real participant attribute. Centrality metrics of nodes were extracted using some of Python library commands. There are different tools in python library for extracting network attributes. Each attribute depends on name according to centrality concept metrics otherwise known as SNA.

Values ascribed to nodes are probability mass function of participating members under the four-centrality metrics extracted. The values were used to quantify relational attitude of participants or actors in a social group – called network attributes. These were extracted from the network graphs - Figure 3.2 and Figure 3.3. The four network attributes are given as following: degree centrality, betweenness centrality, closeness centrality and eigenvector centrality.

Most research works stopped at extraction of network attributes that is, SNA metrics, as final stage for detection of covert nodes. But this was taken further by re-ranking those SNA metrics before setting rule for identifying covert nodes and their relevancies. Network attributes serve as inputs into Bayesian network model for computation of inference before drawing covert nodes. Table 3.1 presents network attributes obtained with respect to network graph of Figure 3.2 and network attributes of Figure 3.3 is presented in Appendix B.

Table 3.1: Network Attributes of actors in the N'17 Greece Revolutionary group

Actor Name	Actor ID	Degree Centrality ( )	Closeness Centrality ( )	Betweenness Centrality ( )	Eigenvector Centrality ( )
Alexandros Giotopoulos	1	0.4762	0.6563	0.1202	0.3032
Anna	2	0.2857	0.5122	0.0040	0.1937
Christodoulos Xiros	3	0.5240	0.6563	0.1043	0.3352
Constantinos Karatsolis	4	0.1429	0.4286	0.0000	0.1070
Constantinos Telios	5	0.1429	0.4667	0.0952	0.0691
Dimitris Koufontinas	6	0.5238	0.6563	0.1101	0.3374
Dionysis Georgiadis	7	0.0476	0.4038	0.0000	0.0299
Elias Gaglias	8	0.0476	0.3231	0.0000	0.0087
Fotis	9	0.2857	0.5676	0.0147	0.1963
Iraklis Kostaris	10	0.3333	0.5122	0.0242	0.2243
Nikitas	11	0.2857	0.5676	0.0147	0.1963
Ojurk Hanuz	12	0.2381	0.4884	0.0000	0.1842
Patroclos Tselentis	13	0.2857	0.5000	0.0024	0.2099
Pavlos Serifis	14	0.6667	0.6563	0.1749	0.3963
Sardanopoulos	15	0.3333	0.5385	0.0145	0.2326
Savas Xiros	16	0.5238	0.6563	0.3483	0.2371
Sotirios Kondylis	17	0.0476	0.4038	0.0000	0.0299
Thomas Serifis	18	0.3333	0.5122	0.0180	0.2296
Vassilis Tzortzatos	19	0.0476	0.4038	0.0000	0.0299
Vassilis Xiros	20	0.1905	0.5250	0.0093	0.1428
Yiannis	21	0.2857	0.5676	0.0261	0.2134
Yiannis Skandalis	22	0.2381	0.4884	0.0000	0.1842

3.3.4 Development of Bayesian network model

The Bayesian model computation is based on framework expressed as equations (3.3) and (3.4)

$$(\mid) = \frac{(\mid) ( )}{ ( )} \tag{3.3}$$

$$(\mid) \propto (\mid) ( ) \tag{3.4}$$

Equations (3.3) and (3.4) are frameworks for computing conditional and marginal probabilities of stochastic events A and B. where ( ) the marginal probability or prior probability of event A and (|) is conditional probability of A given B; it is also called posterior probability. (|) is conditional probability B given A, ( ) is prior probability usually considered as normalizing constant. (|) is likelihood of A given fixed B, here (|) is equal to (|) however, at times likelihood L can be multiplied by a factor so that it is proportional to, but not equal to probability P.

Equation (3.4) becomes a framework for computation of posterior probability when ( ), marginal probability is used as a normalized constant. Posterior probability can be computed using only standardized likelihood or normalized likelihood (|) and prior probability P(A). Therefore, posterior probability distribution (|) is directly proportional to (|), normalized likelihood and prior probability ( ).

The use of Bayesian model requires data and target event. The target event is covert node designated as . From simple rule of probability, probability of event is over sum of possible outcome defined as ( ) = , Equation (3.5) expresses

posterior probability of given data, where (,) is joint probability distribution and ( ) is marginal probability

$$(|) = \frac{(,)}{( )} \tag{3.5}$$

By applying chain rule of probability theory to numerator of (3.5) gives (3.6).

$$(|) = \frac{(|) ( )}{( )} \tag{3.6}$$

where (|) is conditional probability given , (|) is likelihood or conditional probability given , ( ) is marginal probability of covert node and ( ) is normalized constant.

Posterior probability (|) is an inference needed for identifying covert nodes. Equation (3.6) shows that calculation of (|) depends on likelihood (|), prior probability ( ) and normalized constant ( ). But only prior probability is independent of parameter. This implies that is an important parameter in the Bayesian model. P(data) in equation (3.6) is used to find other parameters for computation of posterior probability.

Distribution Types and Involved Parameters

The denominator in (3.6) is expressed as normalized constant (3.7);

$$( ) = \int (,) = \int (|) ( ) \tag{3.7}$$

Note: is considered as a vector that is, contains values $_1$, $_2$ ... , joint probability distribution of vector values is given as (,), applying integral on joint probability distribution normalize the data distribution; Equation (3.7) is integrating likelihood, (|) and prior probability, ( ) with respect to , likelihood (|) was replaced with binomial distribution: $( )(1 -)^{-}$ and prior probability ( ) was replaced with beta distribution $\frac{ (+)}{ ( ) ( )} {}^{-1}(1 -) {}^{-1}$ in (3.7)

gives equation (3.8);

$$( ) = \int_0^1 ( ) (1 -) - \frac{ (+)}{ ( ) ( )} {}^{-1}(1 -) {}^{-1} \tag{3.8}$$

By factorizing the independent terms from (3.8) gives (3.9)

$$( ) = \frac{(+1) (+)}{(- +1) (+1) ( ) ( )} \int_0^1 + -1(1 -) + - -1 \quad (3.9)$$

From (3.8), beta distribution has -1 power; and $(1 -)$ has -1power; from equation (3.9) now has $+ - 1$ power, and $(1 -)$ has $+ - - 1$power after factorized common terms. To replicate -1 and -1 power respectively in (3.9) then equations (3.10) and (3.11) defined parameters $'$ and $'$ that combined parameters of binomial and beta distributions.

$$' = + \quad (3.10)$$

$$' = + - \quad (3.11)$$

By substituting (3.10) and (3.11) into (3.9) gives (3.12),

$$( ) = \frac{(+1) (+)}{(- +1) (+1) ( ) ( )} \int_0^1 '-1(1 -) '-1 \quad (3.12)$$

Integrate (3.16) with respect to gives (3.13)

$$( ) = \frac{(+1) (+)}{(- +1) (+1) ( ) ( )} \cdot \frac{(') (')}{(' + ')} \int_0^1 \frac{(' + ')}{(') (')} '-1(1 -) '-1 \quad (3.14)$$

For special case in beta distribution when $=$ $= 1$ the part of equation (3.13) given as

(3.14)

$$\int_0^1 \frac{(' + ')}{(') (')} '-1(1 -) '-1 = 1 \quad (3.14)$$

Therefore (3.13) becomes (3.15)

$$( ) = \frac{(+1)}{(+2)} = \frac{!}{(+1)!} = \frac{!}{(+1) !} = \frac{1}{+1} \quad (3.15)$$

This is a prior predictive about . Equations (3.8) to (3.15) was used to find parameters and for computation of posterior distribution. Equation (3.15) is predictive probability about covert nodes in a population N. Centrality metrics provided in Table 3.1 are node attributes. The Bayesian model for covert nodes detection is expressed in (3.16)

$$(\mid) = \frac{(\mid) ( )}{( )} \tag{3.16}$$

Where is centrality of the network graphs in Figure 3.1 and 3.2. include degree centrality ( ), betweenness centrality ( ), closeness centrality ( ), and eigenvector centrality ( ). The centrality metrics were defined as cases. Each was deployed as input to the Bayesian model, just to identify more covert nodes.

Case A is when degree centrality ( ) replaces in (3.16) gives (3.17)

$$(\mid) = \frac{(\mid) ( )}{( )} \tag{3.17}$$

Case B is when betweenness centrality ( ) replaces in (3.16) gives (3.18)

$$(\mid) = \frac{(\mid) ( )}{( )} \tag{3.18}$$

Case C is when closeness centrality ( ) replaces in (3.16) gives (3.19)

$$(\mid) = \frac{(\mid) ( )}{( )} \tag{3.19}$$

Case D is when eigenvector centrality ( ) replaces in (3.16) gives (3.20)

$$(\mid) = \frac{(\mid) ( )}{( )} \tag{3.20}$$

Equations (3.17) to (3.20) have the same priori ( ). Each probability mass function is based on circumspect and { , , , } s set of network attributes - for predicting

propensity of actors' vulnerability or evasiveness. serves as needed provision for estimating posterior probability which is inference.

3.3.5 Development of BN algorithm

Algorithm here show the procedures for detecting covert members in OCGs. All participants in OCG or in presented dataset are depicted as points. All points are presented in the same way that is they are equal except in the number of links incident on them. The developed BN is then invoked to predict covert nodes from dataset. Since all points are represented equally, prior probability is provided as one of parameters needed. The likelihood is also provided. The likelihood represents network attribute as a normalized likelihood. Normalised likelihood and prior probability are merged with parameters 'a' and 'b' and go to Recursive Bayesian Filter (RBF) for computation of inference. The inference is then checked for satisfactory test that is, if a marginal node is detected, the detection process terminates by exist and new network attribute will be loaded. If not, parameters a and b will be re-initialized. Figure 3.4 illustrates the flow diagram of algorithm.

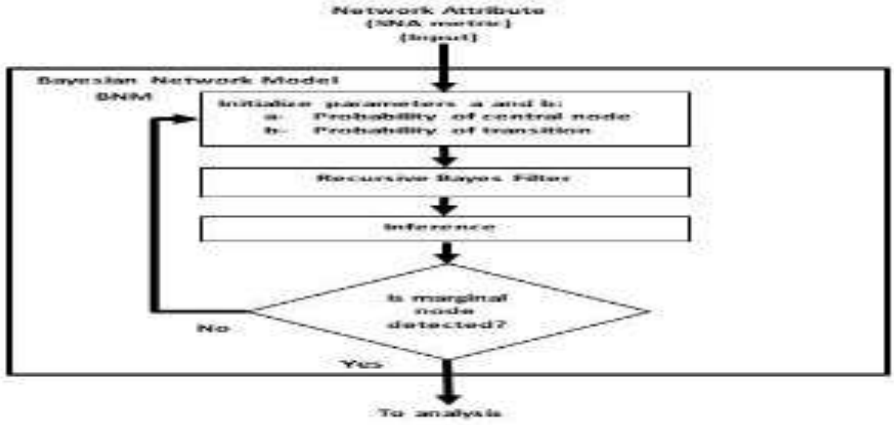

Figure 3.4: Bayesian Network Model Computation Framework

Recursive Bayesian Filter (RBF) minimizes error by refeeding error back. RBF algorithm is given as 3.1. The process within RBF is strictly swapping posterior with prior. The detail of the RBF algorithm is given as Algorithm 3.1. The decision block returns the flow process to the initializing block where parameters ‘a’ and ‘b’ can be adjusted or reinitialized. This permits to observe performance when using different set of values for parameter ‘a’ and ‘b’. Parameters and is [0 1]. The ‘Yes’ takes the flow outside after satisfying with set of nodes produced as covert nodes. The detail of the Bayesian Network Algorithm is described as Algorithm 3.2 and its source codes is in Appendix D.

Algorithm 3.1 RBF ((|))

%Recursive Bayesian Filter (RBF) takes (|) as parameter and returns posterior probability

Input: (|) % centrality metric distribution

( ) ∈ (0,1)

Output: (|) % posterior probability

for = 1 *h*((|))

(|) = ( ). (|))

(|). ((| (|)

( ) =

return (|)

Algorithm 3.2 Bayesian Network Algorithm

%Bayesian-networks-based algorithm for covert nodes Detection using SNA %metrics as inputs

1 Input:% degree centrality

(a) % betweenness centrality

(b) % closeness centrality,

(c) % eigenvector centrality

2 %Initialize, number of network attributes

3 ={ , , , }

4 %Initialize N, number of nodes

5 s= . h()

6 *for each*∈

7 (|) = RBF ( )

8 Outlier-inference (Plot ((|),))

3.3.6 Performance evaluation of BNM

Confusion matrix was deployed for evaluating proposed BNM's performance. Its tools include: sensitivity, probability of false alarm, miss rate, and specificity as presented in Figure 3.5.

Definition of terms - The model performance is observed on tools provided in confusion matrix-Figure 3.5. The chart has two conditions: True condition and Predicted condition. The True condition denotes physical, potential or observable attribute of affiliate criminals like conspirators, sleeper partners and overt members. Condition positive denotes fugitive – conspirators and sleeper partners, while condition negative denotes non-fugitive - attackers. Predicted condition denotes statistical values used for condition of detection. Predicted condition positive is range of values for detected nodes while predicted condition negative is range of values for rejected nodes. A node is taken as detected when it has high maximum-a-posterior (MAP) value. And a node is rejected

when it has low MAP. The predicted condition positive is used for nodes that have relatively high MAP while predicted condition negative is used for nodes that have relatively low MAP. The predicted condition negative represents rejected.

		True Condition			
	Total population	Condition positive	Condition negative	$Prevalence = \sum \frac{condition\ (+)}{\sum total\ population}$	$Accuracy\ (ACC) = \sum \frac{True(+) + True(-)}{\sum Total\ population}$
Predicted condition	Predicted condition positive	True positive, Power	False positive Type I error	$Positive\ predictive\ value\ (PPV), Precision = \sum \frac{True\ (+)}{\sum Predictied\ condition\ (+)}$	$False\ discovery\ rate\ (FOD) = \sum \frac{False(+)}{\sum Predicited\ condition\ (+)}$
	Predicted condition negative	False negative Type II error	True negative	$False\ omission\ rate\ (FOR) = \sum \frac{False -}{\sum Predicted\ condition\ (-),}$	$Negative\ predictied\ value\ (NPV) = \sum \frac{True(-)}{\sum Predicted\ condition(-)}$
		$True\ positive\ rate\ (TPR), Probability\ of\ detection\ (PoD), Sensitivity, Recall = \sum \frac{True(+)}{\sum condition\ (+)}$	$False\ positive\ rate\ (FPR)), Fallout, Probability\ of\ false\ alarm = \sum \frac{False(+)}{\sum condition(-)}$	$Positive\ likelihood\ ratio\ (LR+) = \frac{TRP}{FPR}$	$Diagonistic\ odd\ ratio\ (DOR) = \frac{LR+}{LR-}$
		$False\ negative\ rate\ (FNR), MIss\ rate = \sum \frac{false\ (-)}{\sum condition\ (+)}$	$True\ negative\ rate\ (TNR), Specificity\ (SPC) = \sum \frac{True\ (-)}{\sum condition\ (-)}$	$Negative\ likelihood\ ratio\ (LR-) = \frac{FNR}{TNR}$	$F1\ score = \frac{2}{\left(\frac{1}{Recall}\right) + \left(\frac{1}{Precision}\right)}$

Figure 3.5: The Confusion Matrix

TP denotes number of correctly detected conspirators that is, conspirators that have high MAP. FP is number of wrongly detected conspirator that is, attackers that have high MAP. FN is number of wrongly rejected conspirators that is, conspirators that have low MAP. And finally, TN is number of attackers correctly rejected that is, attackers that have low MAP. The four terms: TP, FP, TN and FN are used in computation of equation (3.21) to (3.34). Each provides different information on performance evaluation. They are presented as following:

Sensitivity, recall, Probability of detection or True Positive Rate (TPR) is expressed in (3.21)

$$= \frac{\Sigma}{\Sigma + \Sigma} \tag{3.21}$$

Miss rate or False Negative Rate (FNR) is expressed as (3.22)

$$= \frac{\Sigma}{\Sigma + \Sigma} \tag{3.22}$$

Probability of false alarm, fall-out or False Positive Rate (FPR) is expressed in (3.23)

$$= \frac{\Sigma}{\Sigma + \Sigma} \tag{3.23}$$

Specificity - SPC or True Negative Rate (TNR) is expressed as (3.24)

$$\Sigma \quad = \Sigma + \Sigma \tag{3.24}$$

Prevalence is defined in (3.25) $= \Sigma \quad \frac{}{\Sigma + \Sigma}$

Precision or Positive Predictive Value (PPV) is defined in (3.26) $= \Sigma + \Sigma \quad \frac{\Sigma}{}$

(3.25)

False Omission Rate (FOR) is expressed in (3.27)

$$\Sigma \quad = \Sigma + \Sigma$$

Positive likelihood Ratio (LR+) is expressed in (3.28)

$$+= \quad \frac{}{} \tag{3.26}$$

Negative likelihood Ratio is expressed in (3.29)

$$-= \quad \frac{}{}$$

(3.27)

(3.28)

(3.29)

Accuracy (ACC) is defined in (3.30)

$$\Sigma + \Sigma \qquad (3.30)$$

False Discovery Rate (FDR) is expressed as (3.31)

$$\Sigma \qquad (3.31)$$

Negative Predictive Value (NPV) is expressed in (3.32)

$$\Sigma \qquad (3.32)$$

Diagnostic Odd Ratio (DOR) is expressed in (3.33)

$$= \qquad (3.33)$$

And finally, $_1$ is expressed in (3.34)

$$_1 = \qquad (3.34)$$

Tools given in eqn. (3.21) to (3.34) were used for evaluating direct information on the BNM's performance with respective to two terrorist networks analysed. In addition, graphs of SNR are plotted using detection probability (TPR). Receiver Operation Characteristic curves (ROC) is also used to access common features or information that cannot be directly accessed by plotting detection probability with probability of false alarm. It allows to compare the impact of different inputs on the BNM and to find roc that has better performance.

3.3.7 Application of SNA-Quadrant model

This section presents SNA-Quadrant (SNA-Q) model for validation of EnBNM's detection. The validation was carried out on detected nodes by comparing set of nodes detected by EnBNM with the SNA-Q model. The SNA-Quadrant (SNA-Q) model was part of literature review section (2.5.1). The Quadrant approach for classification of profiles in OCGs (Ismail *et al.*, 2019). It was adopted for validation because scatterplot of attributes - Figure 2.13 provide a framework that support relevance of all participators in OCGs. Roles and contributions were taken as attributes. Each attribute was rated and used to determine how relevancy an actor is (Ismail *et al.*, 2019).

3.3.8 SNA-Q algorithm

The SNA-Q algorithm started with defining a pair of attributes for plotting scatter graph of active attribute against passive attribute(Ismail *et al.*, 2019). The graph has quadrants Q1 to Q4 describing relevancy of node that fall within it. Q2 and Q4 are points of interest (poi). Actors in Q2 are regarded as vulnerable key players while actors in Q4 are regarded as less-vulnerable key players or smart criminals. The two quadrants: Q2 and Q4 are pois for validating set of nodes detected by EnBNM. Both Q2 and Q4 are also validated with ground truth data of participants. The validation permits researchers to drawing correlation between two models and to identify potential fugitive. The detail of SNA-Q algorithm is given as algorithm 3.3 and its source code is given in appendix E.

Algorithm 3.3 SNA-Q Algorithm

%SNA-Q is an algorithm for covert nodes Detection using SNA metrics as inputs and return points of interests as outputs

Input: % degree centrality,

% betweenness centrality

% closeness centrality,

% eigenvector centrality

Output: Quadrants , pointsInQ4 % list of values

% assign network attribute to active input

% assign network attribute to passive input

```
quadrant =find Quandrant (X_mean, Y_mean)
poi=Identify Q Number of Interests (X_mean, Y_mean )
pointsInQ4.append(poi)
Quadrants. append(quadrant)
return pointsInQ4, Quadrants
```

3.4 Chapter Summary

This chapter presented procedures for actualizing the aim and objectives in this work. The flow chart had three sections: data section, BN model section and SNA-Q model section. Bayesian model was developed for detection of covert nodes and metrics needed for the detection of covert nodes were defined. These metrics were extracted from criminal network graphs and they were presented. The last segment of the methodology presented SNA-Q model for classification of participants' relevancy to criminal groups. This is to be used for verifying set of nodes detected by the developed BNM and to know participants who are potential fugitive.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

This chapter presents experimental results on developed methodology for detection of covert nodes. The developed algorithms were tested with datasets of criminal groups. Results were presented according to the two terrorist datasets that is, the N'17 revolutionary group and the 9/11 terrorist group datasets. The results were presented in the following order: experimental results of N'17 revolutionary group in section 4.1, experimental results of 9/11 terrorist group in section 4.2 and comparative analysis of algorithm's performance was presented in section 4.3.

Section 4.1 and section 4.2 have three different results each. The first segment presents result of evaluation of the BNM algorithm, the second segment presents result of evaluation of the SNA-Q algorithm and the third segment presents verification of inferred nodes from both the BNM using the SNA-Q classification of profiles.

The BNM's segment has four cases of results. Cases defined type of network attribute used as input. Case 1 represents eqn. (3.21) that is, when degree centrality was used as an input; case 2 represents eqn. (3.22) that is, when betweenness centrality was used as input; case 3 represents eqn. (3.23) that is, when closeness centrality was used as input and case 4 represents eqn. (3.24) that is, when eigenvector centrality was used as input. Each input was used to predict participants' evasion.

The BNM predicts level evasiveness of a given node. The value assigned to an actor depends on score that an actor has from a particular network attribute. The peak of distribution curve is maximum-a-posteriori (MAP). Each actor has a MAP value. The MAP value indicates if a node has chances of evading detection or considered as vulnerable. A node that has low MAP is considered as vulnerable while a node that has

high MAP is considered to be evasive. From the highest and lowest MAP, only outliers are considered that is, inferred / detected nodes.

Likewise, the results of SNA-Q algorithm were defined on four cases A to D. Each defined a pair of network attributes used as inputs to the SNA-Q algorithm. Case A represents when degree centrality and betweenness centrality were used as inputs; case B represents when degree centrality and closeness centrality were used as inputs; case C represents when eigenvector centrality and betweenness centrality served as inputs. Finally, case D represents when eigenvector centrality and closeness centrality were used as inputs to the SNA-Q.

4.1 Experimental Results of Algorithms Using N'17 Criminal Dataset

The results of testing the BNM and the SNA-Q algorithms using the N'17 datasets are presented. Section 4.2.1 presents result of evaluating the BNM algorithm, section 4.2.2 presents result for classification of nodes using the SNA-Q algorithm and section 4.2.3 presents comparative analysis for verification of nodes inferred by the BNM from the N'17 network.

4.1.1 BN Model evaluation results using network attributes of the N'17 criminal group

Results presented in this section were obtained from evaluating the BN model with network attributes of N'17 criminal group. The network attributes are referred to as case 1 to 4. Each case has different MAP values for participating nodes which indicates their level of evasiveness. The results are presented from case 1 to 4 along with performance evaluation on corresponding case.

Case 1: Degree Centrality Input in BNM

Figure 4.1 presents the graph of MAP distribution of N'17 participants when degree centrality was used as input to the BNM - eqn. (3.21). From the graph, actor ID 14 has the least MAP value. This implies that actor Id 14 was the most vulnerable node predicted by BNM using degree centrality as input. Theoretically, a node that is located at the center of a network structure or having relatively high number of links incident on it, usually turns out to be a vulnerable node. Actor IDs 3, 6, and 16 emerged as structurally equivalent actors on MAP value of 380.6 above actor ID 14. These actors are close to center of the network structure of Figure 3.2, as they were inferred as the next vulnerable actors after actor ID 14.

Actor IDs 2, 4, 5, 7, 8, 9, 11, 13, 17, 19, and 21 emerged as outliers with MAP value of 402.6. They are structurally equivalent actors at the highest MAP of 402.6. Though they differ in number of links connected to them, they were inferred as the most evasive actors that is, they are less vulnerable to security operatives' detection.

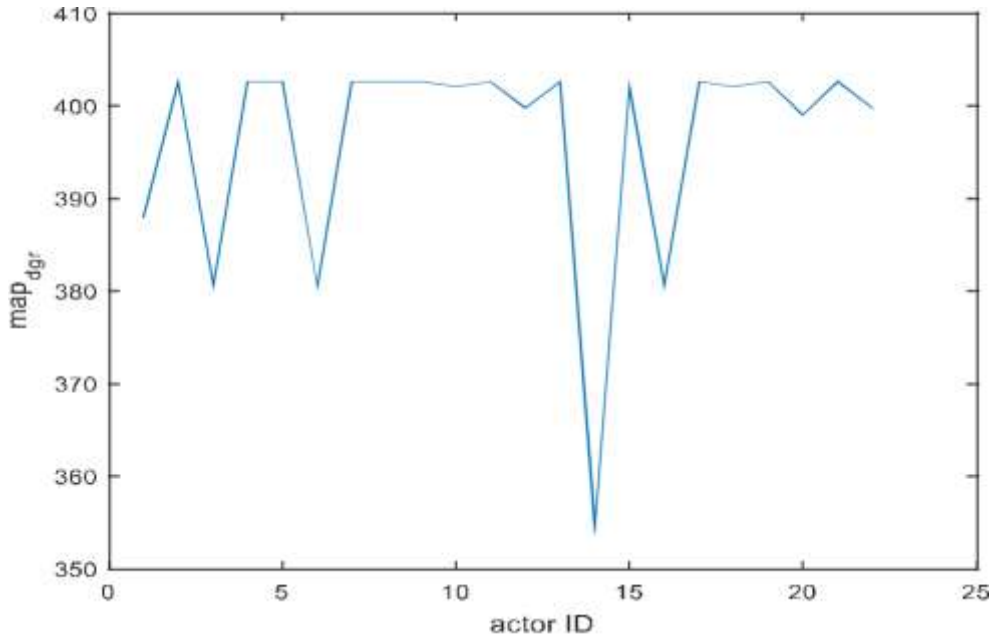

Figure 4.1: MAP Distribution of the N'17 Network with Case 1

Figure 4.2 presents assessment on BNM's detection with degree centrality as input. The chart has 5 TP, 13 FP, 1 FN and 3 TN. It shows that BNM has 0.833 TPR that is, probability of detection; 0.1667 FNR that is, miss rate; 0.8125 FPR that is, Probability of False Alarm (Pfa); 0.1875 TNR that is, specificity; 0.2778 precision and 0.3636 accuracy.

		True Condition (N17 network)			
	Total population (22)	Condition positive (Conspirators)	Condition negative (Attackers)	*Prevalence* = 0.2727	*ACC* = 0.3636
Predicted condition (degree)	Predicted condition positive (High MAP)	TP 5	FP 13	*Precision* = 0.2778	*FOD* = 0.7222
	Predicted condition negative (Low MAP)	FN 1	TN 3	*FOR* = 0.25	*NPV* = 0.75
		TPR = 0.8333	*FPR* = 0.8125	*LR+*= 1.0256	*DOR* =
		FNR = 0.1667	*TNR* = 0.1875	*LR−*= 0.8891	*F1 score* = 0.4167

Figure 4.2: Performance evaluation of BNM's Detection with Case 1

Case 2: Betweenness Centrality Input in BNM

Figure 4.3 presents the MAP distribution of participants in N'17 when betweenness centrality was used as input to the BNM - eqn. (3.22). The plot shows that actor IDs 4, 7, 8, 12, 17, 19 and 20 emerged as structurally equivalent IDs on the least MAP of zero. They were inferred as central participants that is, vulnerable actors. Theoretically, only nodes that have the highest betweenness metrics are usually taken as key players and as vulnerable actors. But here, the BNM was able to identify some actors with low betweenness metrics as vulnerable actors. But their key player status can be determined by their profile.

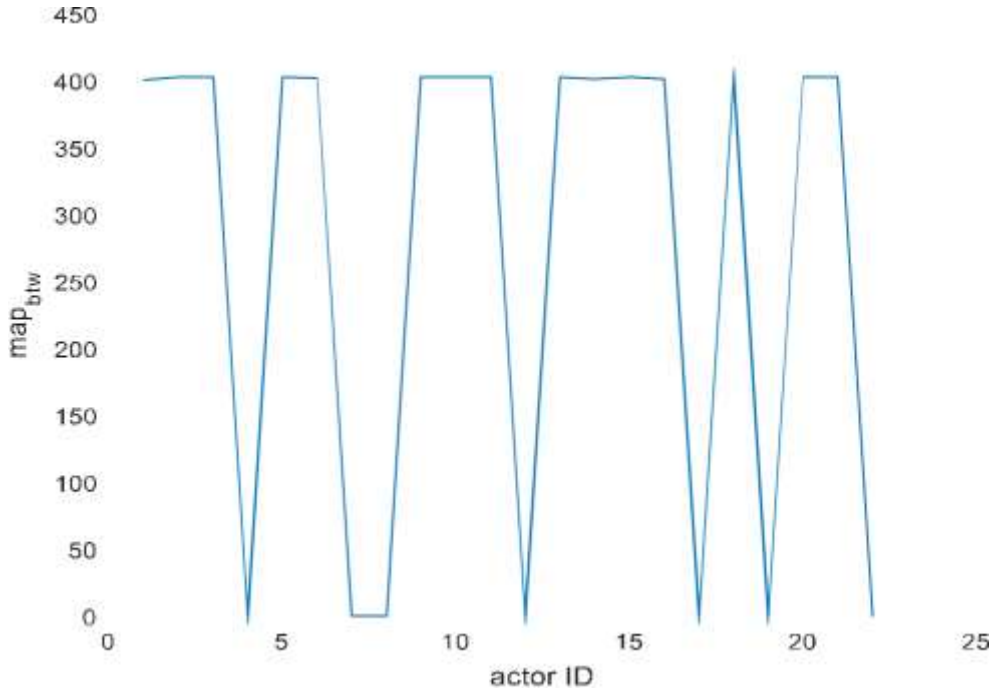

Figure 4.3: MAP Distribution of the N'17 Network with Case 2

Actor IDs 1, 2, 3, 5, 6, 9, 10,11, 13, 14, 15, 16, 18, 20, and 21emerged as structurally equivalent actors on the highest MAP of 402.6. These are less vulnerable actors inferred. BNM takes care of hidden data which is inherent in criminal domain through probability. Therefore, IDs with zero MAP might have engaged in indicting acts that make up for that inference. The BNM was able to infer less-vulnerable actors as central participants which deterministic approach could not detect.

Figure 4.4. presents assessment on BNM’s detection using betweenness centrality as input. The evaluation chart has 6 TP, 9 FP, 0 FN and 7 TN. These values correspond to 1 TPR that is, detection probability; 0 FNR that is, miss rate, 0.5625 FPR that is, probability of false alarm (Pfa), 0.4375 TNR that is, specificity, 0.4 precision and 0.5909 accuracy.

		True Condition (N17 network)			
	Total population (22)	Condition positive (Conspirators)	Condition negative (Attackers)	*Prevalence* = 0.2727	*ACC* = 0.5909
Predicted condition (btw)	Predicted condition positive (High MAP)	TP 6	FP 9	*Precision* = 0.4	*FOD* = 0.6
	Predicted condition negative (Low MAP)	FN 0	TN 7	*FOR* = 0	*NPV* = 1
		TPR = 1	*FPR* = 0.5625	*LR+*= 1.7778	*DOR* =
		FNR = 0	*TNR* = 0.4375	*LR−*= 0	*F1 score* = 0.5714

Figure 4.4: Performance Evaluation of BNM's Detection with Case 2

Case 3: Closeness Centrality Input in BNM

Figure 4.5 presents the MAP distribution of N17 network using closeness centrality as input to the BNM – eqn. (3.23). Actor IDs 1, 3, 6, 14 and 16 emerged as structurally equivalent actors with the lowest MAP of 356.6. They are inferred as central participant as they emerged with the least MAP. The set of IDs identified with low MAP here are in line with theory that is, they were inferred as vulnerable actors where actors with high closeness were being taken as vulnerable actors.

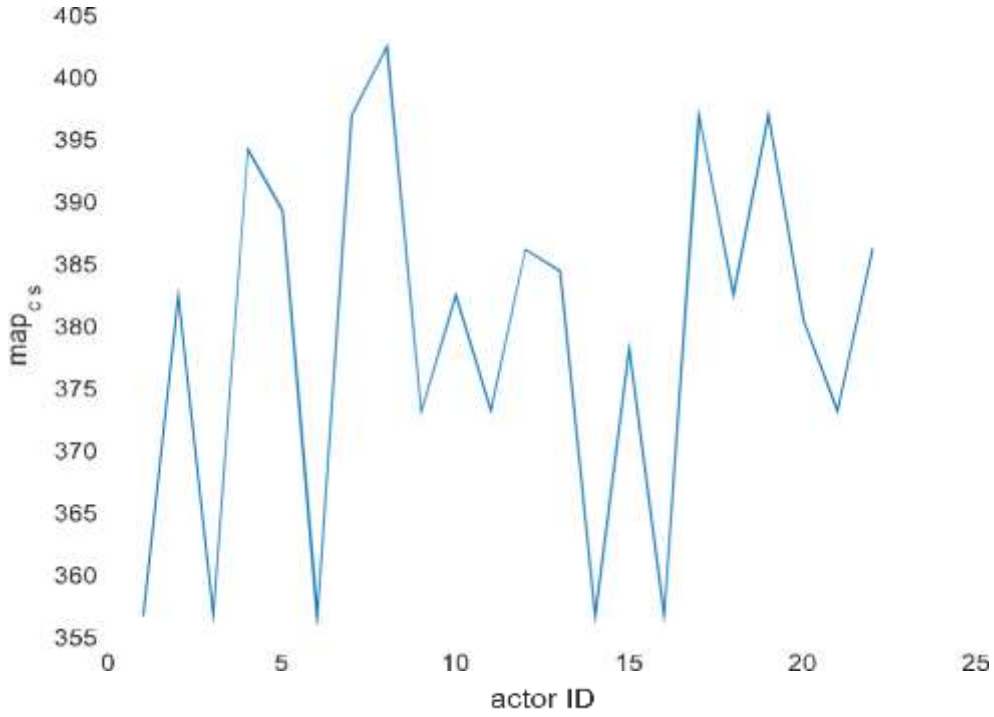

Figure 4.5: MAP Distribution of the N17 Network with Case 3

Actor ID 8 has the highest MAP. This indicates that, it is the most evasive actor. It does not become structurally equivalent with any of actors that it shared structural equivalence property with in earlier detection with other network metrics. Actor IDs 7, 17 and 19 emerged with the same MAP below actor ID 8. The same MAP denotes that they are structurally equivalent. Four actor IDs: 7, 8, 17 and 19 were identified in Figure 4.1 as structurally equivalent actors on the same MAP value. Here actor ID 8 became distinguished from other actors. This is the least suspected actor in the group from its position to the central of the network.

Figure 4. 6 presents assessment of BNM's detection when closeness centrality was used as input. The Figure has 4 TP, 13 FP, 2 FN, and 3 TN. This shows that BNM has 0.6667 TPR that is, detection probability, 0.3333 FNR that is, miss rate; 0.8125 FPR that is, Probability of False Aalarm (Pfa), 0.1875 TNR for specificity; 0.2353 precision and 0.3182 accuracy.

		True Condition (N17 network)			
	Total population (22)	Condition positive (Conspirators)	Condition negative (Attackers)	*Prevalence* = 0.2727	*ACC* = 0.3182
Predicted condition (cls)	Predicted condition positive (High MAP)	TP 4	FP 13	*Precision* = 0.2353	*FOD* = 0.7647
	Predicted condition negative (Low MAP)	FN 2	TN 3	*FOR* = 0.4	*NPV* = 0.6
		TPR = 0.6667	*FPR* = 0.8125	*LR+*= 0.8206	*DOR* = 0.4615
		FNR = 0-3333	*TNR* = 0.1875	*LR−*= 1.7778	*F1 score* = 0.3478

Figure 4.6: Performance Evaluation of BNM's Detection with Case 3

Case 4: Eigenvector Centrality Input in BNM

Figure 4.7 presents the MAP distribution of N'17 network using eigenvector centrality as input to the BNM – eqn. (3.24). Actor ID 13 has the least MAP value of 395.9. It is not structurally equivalent with any actors. This implies that it was inferred as the most vulnerable actor. Actor ID. 21 has the second lowest MAP value. Actor IDs 13 and 21 were inferred as vulnerable actors through prediction by BNM. None of these two actors acceded to the centre of network presented in Figure 3.2.

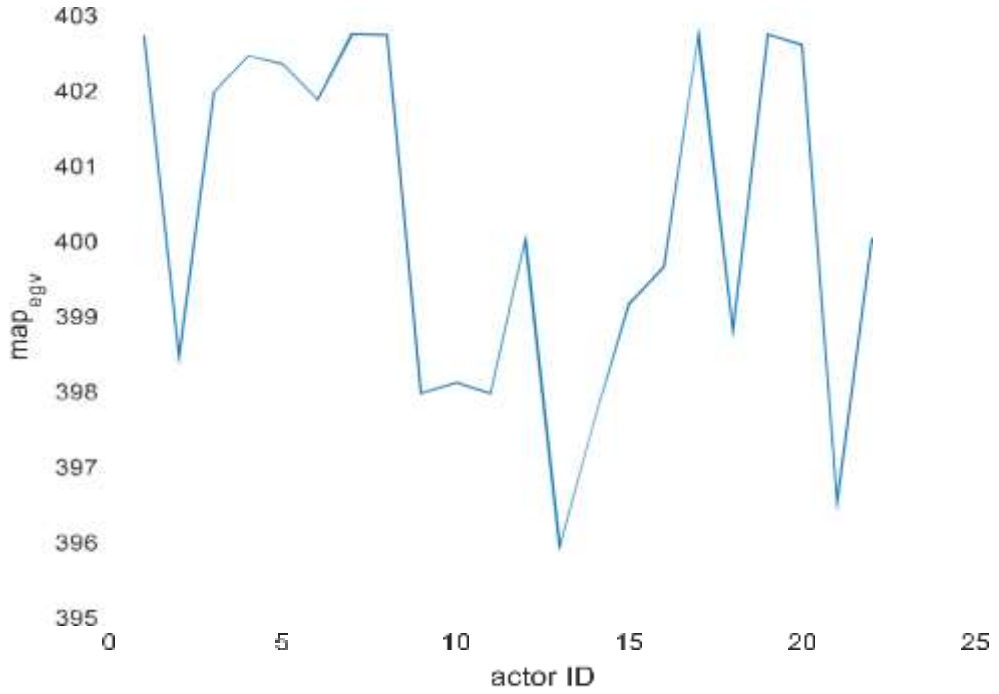

Figure 4.7: MAP Distribution of the N'17 Network with Case 4

In theory, when an actor has high eigenvector metric it is taken as a key player and also as a vulnerable actor. But the BNM was able to infer less-susceptible actors that are neither at the centre of a network structure nor have the highest links connected to them. Detection of less-vulnerable follows the phenomenon that centrality may be unrelated to key players status in criminal networks (Bright *et al.*, 2015). From the upper part of the Figure 4.7, five (5) actor IDs become structurally equivalent on the highest MAP of 402.7. The actors are 1, 7, 8, 17 and 19. Out of these five actors, only actor ID 1 has the highest number of links in Figure 3.2 and also has the highest eigenvector value in Table 3.1.

The BNM shows that only actors with low eigenvector centrality metrics are less-vulnerable to detection. Inclusion of actor ID 1 among nodes that has high MAP shows that. This particular actor was alleged and convicted as chief ideologue leader of the N’17 group. But he denied of being the denied leader or a participating member. He claimed that the allegation of his participation was cooked by the security agent (Kassimeris, 2007). It is a manifestation that not all actors that have the high influence impact in a

criminal group could be vulnerable to detection. The BN model identified actor IDs 13 and 21 as the most vulnerable to detection by using eigenvector centrality for prediction.

Figure 4.8 presents assessment of BNM's detection when eigenvector centrality was used as input. The following were obtained from the Figure: 3 TP, 14 FP, 3 FN, 2 TN. It implies that BNM with eigenvector centrality has 0.5 TPR that is, probability of detection; 0.5 FNR that is, miss rate; 0.875 FPR that is, probability of false alarm (Pfa), 0.125 TNR that is, specificity, 0.1764 precision and 0.2272 accuracy.

		True Condition (N17 network)			
	Total population (22)	Condition positive (Conspirators)	Condition negative (Attackers)	$Prevalence =$ 0.2727	$ACC =$ 0.2272
Predicted condition (eigen)	Predicted condition positive (High MAP)	TP 3	FP 14	$Precision =$ 0-1764	$FOD =$ 0.8235
	Predicted condition negative (Low MAP)	FN 3	TN 2	$FOR =$ 0.6	$NPV =$ 0.4
		$TPR =$ 0.5	$FPR =$ 0.875	$LR+=$ 0.5714	$DOR =$ 0.1428
		$FNR =$ 0.5	$TNR =$ 0.125	$LR-=$ 4	$F1\ score =$ 0.2608

Figure 4.8: Performance evaluation of BNM's Detection with Case 4

Summary of Inferred Nodes from the N'17 Network

Table 4.1 presents the list of actor IDs considered for status verification. They are categorised into central and evasive participants. The central participants consist of actor IDs with low MAPs while the evasive participants are those with the highest MAPs. The former exists in the lowest part of the graphs while the latter occupies the upper part. Some actors emerged as structurally equivalent actors in a layer, while some emerged as non-structurally equivalent actor. In Figure 4.1 actor ID 14 has the least MAP while in

Figure 4.5 actor ID 8 has the highest MAP. These two are not structurally equivalent with any actor. Finally, Figure 4.7 has inferred actor IDs 13 and 21. The two did not form structurally equivalent with other actors. Actor ID 13 was inferred with the lowest MAP therefore it is prominent that can be easily detected by security.

Table 4.1: Nodes Detected by Inference from the N'17 Network using BNM Algorithm

Figure number	Central participants	Evasive participants
Case 1 - Figure 4.1	14, 16, 6, 3	2, 4, 5, 7, 8, 9, 11, 13, 17, 19, 21
Case 2 - Figure 4.3	4,7,8,12,17,19, 22	1, 2, 3, 6, 9, 10, 11, 13, 14, 15, 16, 18, 20
Case 3 - Figure 4.5	1, 3, 6, 14, 16	8, 7, 17, 19
Case 4 - Figure 4.7	13, 21	1, 7, 8, 17, 19

Figure 4.9 presents a Venn diagram of actor IDs inferred as evasive participants in Table 4.3. The Venn diagram compressed re-occurring evasive actors. No actors were duplicated and the place where they were identified was adequately captured. For instance, actor IDs 7, 8, 17 and 19 were inferred in cases 1, 3 and 4. They are the only IDs having highest number of re-occurrences. They emerged in three instances of cases out of four. They are described as legitimate actors as they can hardly be detected by security operatives because each has a link connecting to it. Fortunately, BNM predicted these four actors as vulnerable in Figure 4.3. Identifying the least susceptible actors is a plus to the BNM algorithm.

Actor IDs 2, 9, 11, and 13 occurred in cases 1 and 2 only. They are structurally equivalent on number of links connected to each. Actor ID 1 occurred in case 2 and case 4 only. Actor IDs 4, 5, and 21 occurred in case 1 only and actor IDs 3, 6, 10, 14, 15, 16, 18 and 20 occurred in case 2 only.

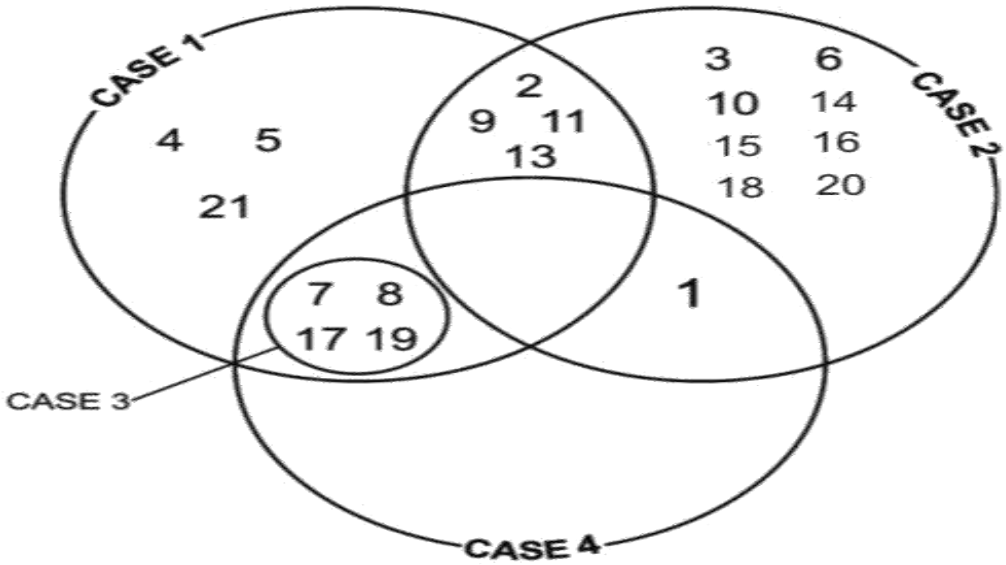

Figure 4.9: Venn Diagram of Evasive Nodes in the N'17 Network

4.1.2 Results of classification of N'17 participants using SNA-Q algorithm

Results presented in this section were obtained from classifying participants in the N'17 network. The participants (nodes) were classified into four categories of criminal profiles using the SNA-Q algorithm. The results are presented according to cases A to D that is, a pair of network attributes used as inputs to the SNA-Q model.

SNA-Q model has four quadrants Q1 to Q4. Each quadrant harbours a set of actors according to a pair of network attributes involved. Points of interest are Q2 and Q4. Q2 is designated as the most prominent, these actors are considered to be the most vulnerable actors, while Q4 is designated as less prominent, and they are regarded as less vulnerable participants or affiliate criminals. Smart participants in terrorist groups or OCGs have high propensity to Q4 properties. Legitimate actors can be described as smart participants.

Case A: Degree and Betweenness Centrality Inputs

Figure 4.10 presents a graph of SNA-Q classification obtained when degree centrality and betweenness centrality were used in SNA-Q model. Actor IDs 2, 9, 10 13, 15, 18, and 21

emerged in Q1; actor IDs 1, 3, 6, 14, and 16 emerged in Q2; actor IDs 4, 7, 8, 11, 12, 17, 19, 20 and 22 emerged in Q3 and actor ID 5 emerged in Q4.

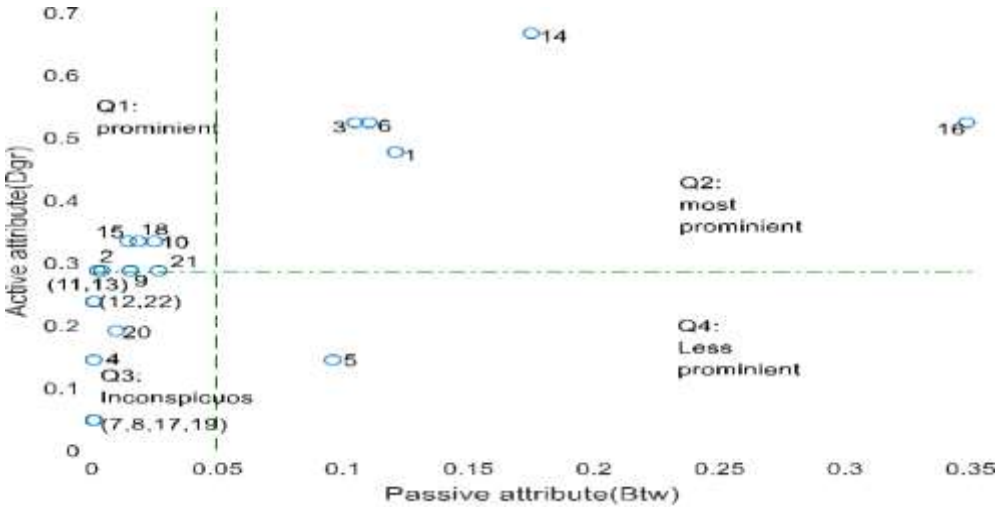

Figure 4.10: SNA-Quadrant Classification of the N'17 Criminal group with Case A

Case B: Degree and Closeness Centrality Inputs

Figure 4.11 presents a graph of SNA-Q classification obtained when degree centrality and closeness centrality were used in SNA-Q model. Q1 has actor IDs 2, 10 13, and 18; Q2 has actor IDs 1, 3, 6, 14, 15 and 16; Q3 has actor IDs 4, 5, 7, 8, 12, 17, 19, and 22; finally, Q4 has actor IDs 9, 11, 20 and 21.

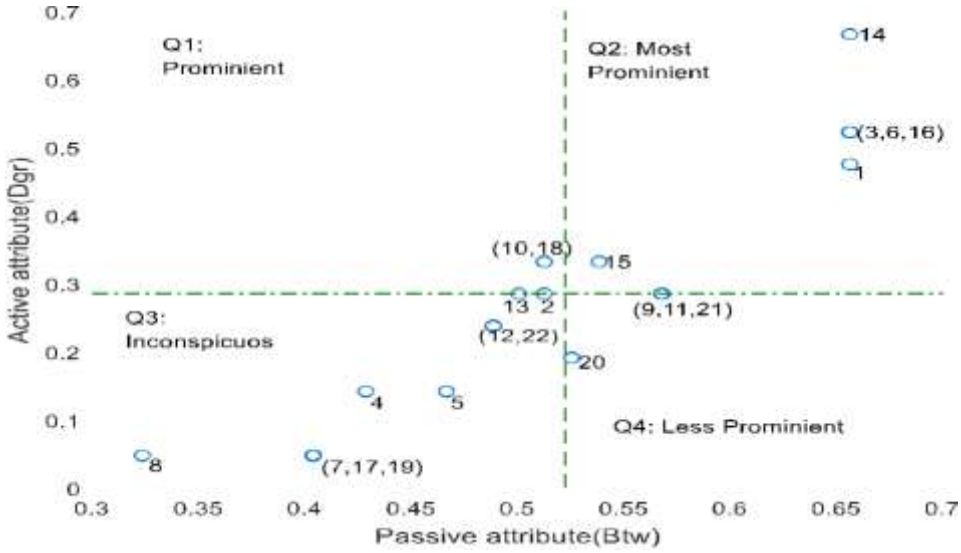

Figure 4.11: SNA-Quadrant Classification of the N'17 Criminal group with Case B

Case C: Eigenvector and Betweenness Centrality Inputs

Figure 4.12 presents a graph of SNA-Q classification obtained when eigenvector centrality and betweenness centrality were used in the SNA-Q model. Q1 has actor IDs 2, 9, 10, 11, 13, 15, 18, and 21; Q2 contains actor IDs 1, 3, 6, 14 and 16; Q3 has actor IDs 4, 7, 8, 12, 17, 19, 20 and 22 and finally Q4 has only actor ID 5.

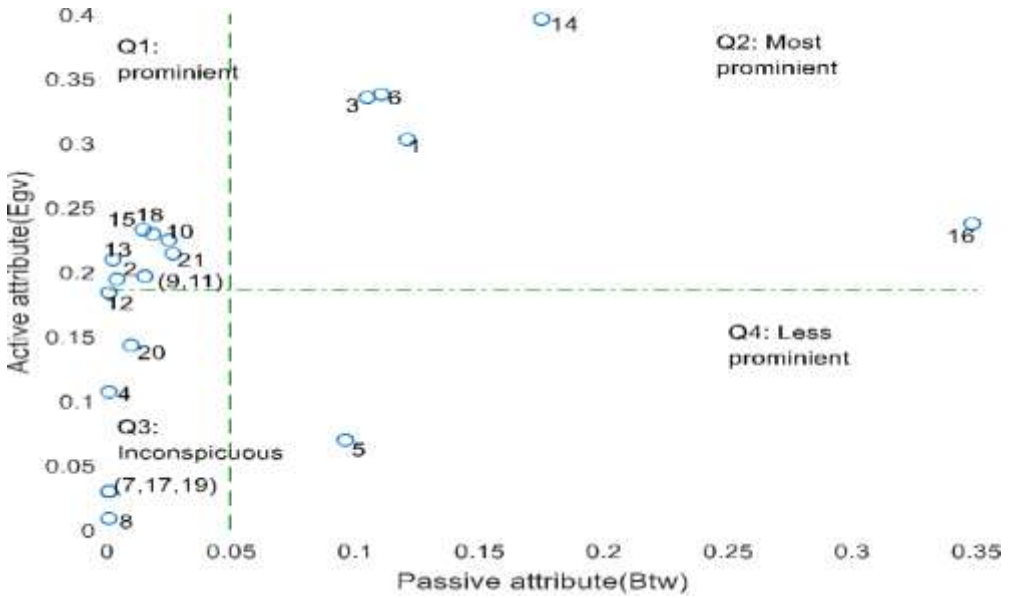

Figure 4.12: SNA-Quadrant Classification of the N'17 Criminal group with Case C

Case D: Eigenvector and Closeness centrality inputs

Figure 4.13 presents a graph of SNA-Q classification obtained when eigenvector centrality and closeness centrality were used in SNA-Q model. Actor IDs 2, 10, 13 and 18 appeared in Q1; actor IDs 1, 3, 6, 9, 11, 14, 15, 16 and 21 appeared in Q2; actor IDs 4, 5, 7, 8, 12, 17, 19 and 22 appeared in Q3 and Q4 has only actor ID 20.

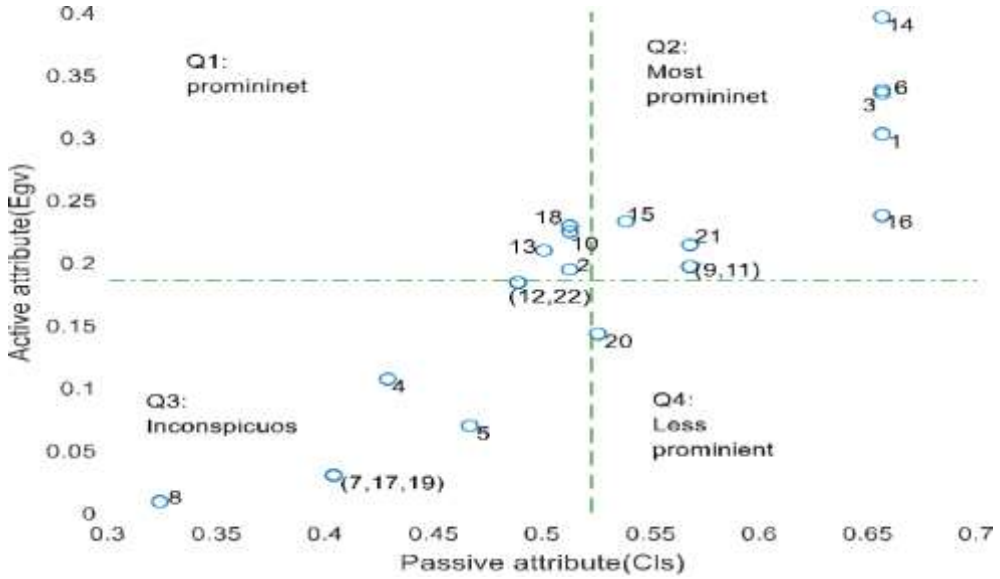

Figure 4.13: SNA-Quadrant Classification of the N'17 Network with Case D

Distributions of N'17 Actors in Quadrants of SNA-Q Model

Table 4. 2 presents summary of nodes distribution in SNA quadrants. These were quantified in percentage. This was done across all Q1 to Q4. The number of nodes in each quadrant was measured and its percentage computed. Recall that the N'17 criminal network has 22 actors.

Table 4.2: Distribution of N'17 Participants in SNA-Q Model

Q-model variables	Distribution				Total
	Q1(%)	Q2(%)	Q3(%)	Q4(%)	
Degree and Betweenness	31.8	22.7	40.9	4.5	100
Degree and Closeness	18.2	27.3	36.4	18.2	100
Eigenvector and Betweenness	36.4	22.7	36.4	4.5	100
Eigenvector and Closeness	18.2	40.9	36.4	4.5	100

Q1 is the quadrant that has the second highest percentage across all the model variables. It has percentage between 36.4 and 18.2. It is designated for prominent actors; those engaged more in indicting activities than in covert roles. Their engagement in such activities make them become vulnerable to security operative detection.

Q2 has percentage between 22.7 and 40.9. This is the quadrant of most prominent criminals; participants that excessively engage both in indicting activities as well as covert roles that is, roles connected with criminal group existence or persistence. They are more vulnerable and more significant than set of actors in Q1.

Q3 has the highest percentage of nodes distribution. Its least percentage in the Table is 36.4 and the highest percentage is 40.9. It was designated as inconspicuous for describing actors whose indicting activities as well as their covert roles were considered to be insignificant within SNA-Q model scope. This description revealed factors that make them less vulnerable or invulnerable.

Q4 has the least percentage. This quadrant is designated as less-prominent. It contains participants who play more vital roles in crime commission than engaging in indicting activities. Only few actor IDs were identified in it. These actors are important to OCGs.

Q2 and Q4 are very important in identifying profiles of participants in OCGs. The set of nodes in the two quadrants compared favourably with those inferred from the BNM model. SNA-Q model helps to examine correlation between properties of nodes inferred as central participants and those inferred as evasive participants.

4.1.3 Verification of BNM inferred nodes in the N'17 network

This section attempts to verify the detection made by developed BN model using the SNA-Q model. Recall that the SNA-Q is a model used to classify actors into four criminal profiles. The focus here is to verify the actor IDs identified as central participants and evasive participants by the BN model.

(i) ***Verification of BNM Inferred Central Participants Using SNA-Q Model***

Table 4.3 presents inferred participants from the N'17 network by BNM and SNA-Q. The Table contains participants' name and ID number, factional groups and those convicted by court as leaders. The factional groups and convicted leaders were used as ground truth. shows that out of thirteen (13) nodes that BNM identified as central participants, the SNA-Q model confirmed seven (7) as actually belonging to the central participants. Moreover, the court convicted three (3) of these actors as leaders based on available evidence. However, there were actually five (5) leaders in the group. The BN model detected all the five leaders in the group and they were confirmed by the SNA-Q model. This shows that the developed BN model has capacity to detect very important central nodes in a criminal group.

Table 4.3 shows that actor IDs 1, 3, and 6 who were convicted leaders were also detected in all cases of SNA-Q and case 3 of BNM. But case 1 detected actor 3 and 6. Actor ID 6 confessed that he was the leader of the group. Actor ID 3 was alleged and convicted as group hitman. Actor ID 1 was served 21 years life terms and a 25-year sentence as he was convicted as chief ideologue (Kassimeris, 2007). This confirms that key players can be vulnerable. And case 3 of BNM has 100 percent detection while case 2 of BNM bas 66.7 percent.

Actor IDs 14 and 16 were identified under SNA-Q cases and BNM's case 1 and 3. The two actors were convicted members but not as leader. It implies that they were central participants but not leaders. Arrest of actor ID 16 led to chain arrest of all participants except actor ID 6-Dimitris Koufontinas. Actor IDs 14 and 15 belonged to two factional groups. Sardanopoulos was named after actor ID 15. Both actors 14 and 15 belong to G-

first generation leadership faction. They were not convicted as leaders. It is evident that they are central participants as they were inferred by the BNM and SNA-Q.

Table 4.3: Comparison of Inferred Central Nodes from the N'17 Network

Inferred participants		Ground truth		BNM Algorithm Cases: (central nodes inferred)				SNA-Q Algorithm Cases: (Q2)			
Actor	Actor ID	Faction name	Court Conviction	1	2	3	4	A	B	C	D
Alexandrous	1	G	Yes			Yes		Yes	Yes	Yes	Yes
Christodulos	3	K	Yes	Yes		Yes		Yes	Yes	Yes	Yes
Karatsolis	4	S			Yes						
D. Koufontinas	6	K	Yes	Yes		Yes		Yes	Yes	Yes	Yes
Georgiadis	7	K			Yes						
Elias Gaglias	8	K			Yes						
Ojurk Harnuz	12	K			Yes						
P Tselentis	13	S					Yes				
Pavlos Serifis	14	G and S		Yes		Yes		Yes	Yes	Yes	Yes
Sardanopouls	15	G and S							Yes		Yes
Savas Xiros	16	K		Yes		Yes		Yes	Yes	Yes	Yes
Kondylis	17	nil			Yes						
V. Tzortzatos	19	K			Yes						
Yianni	21	nil					Yes				Yes

Case 4 and case D detected actor ID 21 as a central participant. Actor ID 13 was among convicted members that received a maximum of 25-year sentence, it means ID 13 played central roles. Actor IDs 7, 8, 17 and 19 are marginal nodes. They are the least susceptible actors with single link connected to each of them. Actor ID 17, Sotirios Kondylis received maximum 25-years sentence. The BNM is able to detect legitimate actors who had been always considered as inconspicuous by the SNA-Q.

(ii)*Verification of BNM Inferred Evasive Participants Using SNA-Q Model*

Table 4.4 compares inferred evasive participants from the N'17 network. List of actors are part of those emanated from the Venn diagram of evasive nodes in the N'17 network that is, Figure 4.9 and those that occurred in Q4 of SNA-Q model. The Table presents only nodes in Q4. Some actor IDs earlier inferred as central participants like IDs 1, 3, 6, 7, 8, 13, 14, 15, 16, 17 and 19 were removed.

From Table 4.4, it can be seen that all the nodes identified by the BN model as evasive nodes were all confirmed by the SNA-Q model. These four nodes were also convicted in court even though they did not play central roles. This again confirms the effectiveness of developed BN model in identifying other key players in OCGs serving as affiliate or legitimate actor.

Table 4.4: Comparison of Inferred Evasive Nodes from the N'17 Network

Participants		Ground truth	BNM Algorithm (Evasion)		SNA-Q Algorithm (Q4)			
Actor	Actor ID	Faction	Case 1	Case 2	Case A	Case B	Case C	Case D
C. Telios	5	K	Yes		Yes		Yes	
Fotis	9	G	Yes	Yes		Yes		
Nikitas	11	G	Yes	Yes		Yes		
Vassilis Xiros	20	K		Yes		Yes		Yes

Actor ID 20 was inferred as evasive actor in case 2. Case B and case D identified it as a smart actor. The ID 20, Vassilis Xiros was a younger brother of Christodulous Xiros and Savas Xiros. He was also in the same factional group that is, Koufontinas (K) faction with his elder brothers (Rhodes and Keefe, 2007). He was among the five convicted members that received the maximum of 25-year sentence.

4.2 Experimental Results of Algorithm Using 9/11 Terrorist Group Dataset

The results of BNM and SNA-Q algorithms evaluated with the 9/11 terrorist network is presented. Section 4.2.1 contains presentation of results from testing the BNM algorithm, section 4.2.2 contains the results of classification of nodes according to profile of SNA-Q algorithm and section 4.2.3 presents verification of actors identified from the 9/11 network using the SNA-Q.

4.2.1 BN model evaluation results using network attributes of the 9/11 terrorist group

The MAP distribution of participants in 9/11 terrorist network and performance of BNM's detection are presented here. Each case that is, input to the BNM has different graphical results.

Case 1: Degree Centrality Input in BNM

Figure 4.14 is the MAP distribution of 9/11 network with Case 1. Degree centrality was used in BNM - eqn. (3.21) for predicting participants' evasion. Actor IDs 14 and 54 were found to be structurally equivalent actors on MAP value of 397.3 MAP. Then actor ID 15 has MAP value of 400.8 behind the first least MAP value. Actor-IDs 3, 25 and 41 too appeared as structurally equivalent actors on 401.2 MAP. These actor IDs were inferred as central participants in the criminal group.

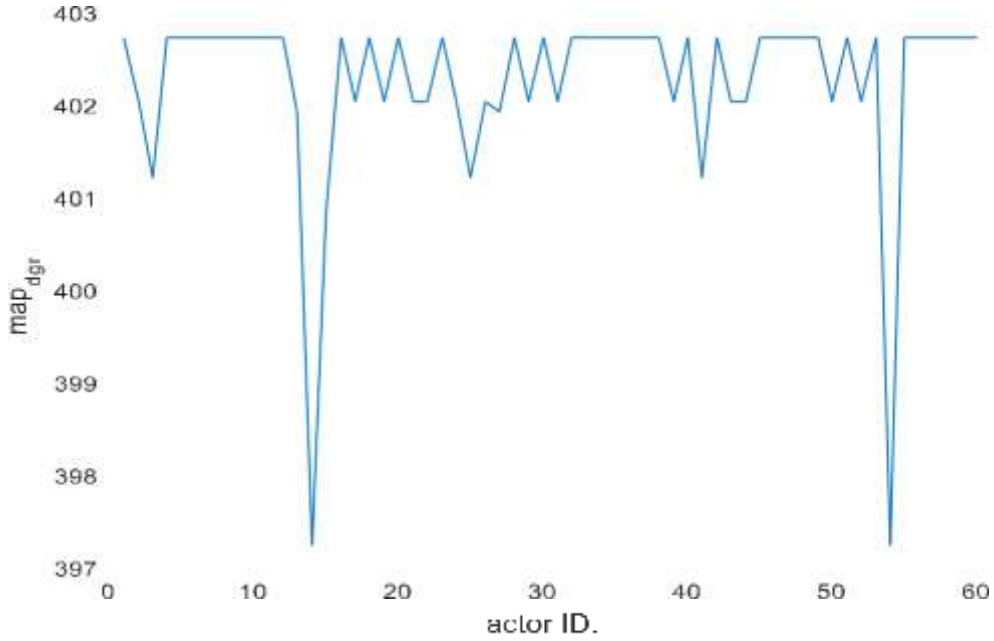

Figure 4.14: MAP Distribution of the 9/11 Network with Case 1

Thirty-eight (38) actors emerged as structurally equivalent on the highest MAP value of 402 with these IDs: 1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 18, 20, 23, 28, 30, 32, 33, 34, 35, 36, 37, 38, 40, 42, 45, 46, 47, 48, 49, 51, 53, 55, 56, 57, 58, 59 and 60. These actors were predicted as those having high propensity to evasion that is the least susceptible actors to ascribe importance in the group. They could become susceptible under different inputs or conditions.

Figure 4.15 presents assessment on the BNM's detection when degree centrality was used as input. The Figure shows that BNM has 26 TP, 12 FP, 15 FN and 7 TN. This means that BNM has 0.6341 TPR that is, detection probability, 0.3658 FNR that is, miss rate; 0.6315 FPR that is, probability of false alarm (Pfa), 0.3684 TNR for specificity; 0.6842 precision and 0.55 accuracy in detection using degree centrality network attribute.

		True Condition (9/11 network)			
	Total population (60)	Condition positive (Conspirators)	Condition negative (Attackers)	$Prevalence =$ 0.6833	$ACC =$ 0.55
Predicted condition (degree)	Predicted condition positive (High MAP)	TP 26	FP 12	$Precision =$ 0.6842	$FOD =$ 0.3157
	Predicted condition negative (Low MAP)	FN 15	TN 7	$FOR =$ 0.6818	$NPV =$ 0.3181
		$TPR =$ 0.6341	$FPR =$ 0.6315	$LR+=$ 1.0040	$DOR =$ 1.0111
		$FNR =$ 0.3658	$TNR,=$ 0.3684	$LR-=$ 0.9930	$F1\ score =$ 0.6582

Figure 4.15: Performance **Evaluation** of BNM's Detection with Case 1

Case 2: Betweenness Centrality Input in BNM

Figure 4.16 is the MAP distribution of 9/11 network with Case 2. Betweenness centrality was used in BNM eqn. (3.22) to predict participants' evasion. The MAP distribution pattern for the 9/11 network is similar with that of the N'17 network under Case 2. Nodes were spin into either the lowest MAP or the highest MAP. Majority of nodes become structurally equivalent in either of the direction.

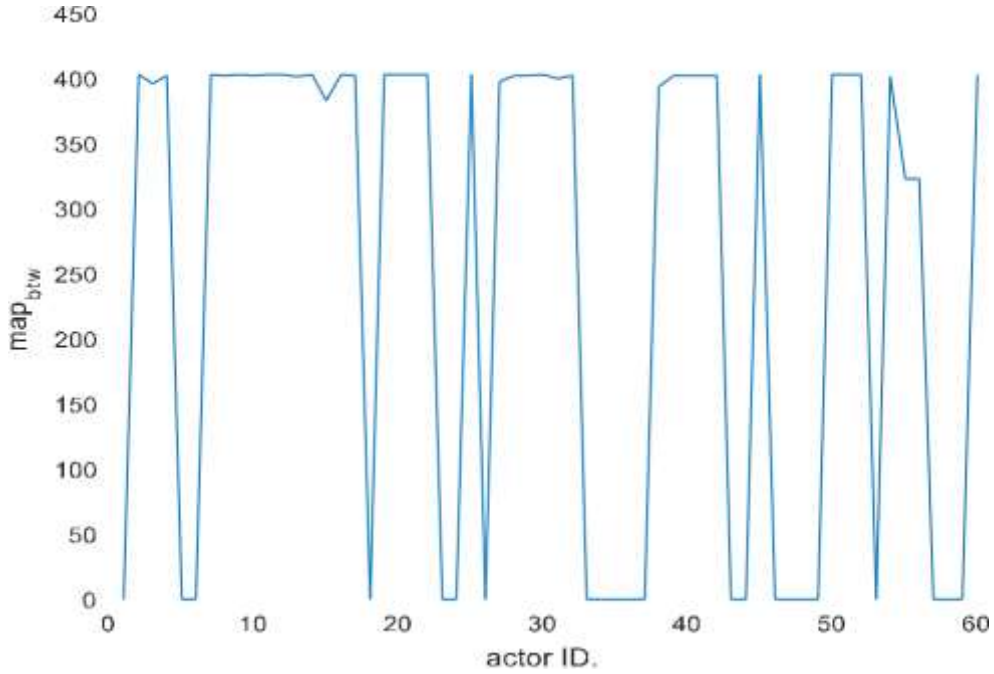

Figure 4.16: MAP Distribution of the 9/11 Network with Case 2

Twenty-two (22) actors emerged on MAP of zero as structurally equivalent actors. They are identified with IDs 1, 5, 6, 18, 23, 24, 26, 33, 34, 35, 36, 37, 43, 44, 46, 47, 48, 49, 53, 57, 58, and 59. They are inferred as central participants that is, most susceptible to detection. This means that their roles are expository by prediction. Some of these inferred actors include those having low betweenness metrics; those could be regarded as legitimate actors.

Twenty-nine (29) actors emerged on the highest MAP value of 402.7 with the following IDs: 2, 4, 7, 8, 9, 10, 11, 12, 14, 16, 19, 20, 21, 22, 25, 28, 29, 30, 32, 39, 40, 41, 42, 45, 50, 51, 52, 54, and 60 as structurally equivalent. This MAP value is for actors regarded as less susceptibility actors to detection. It was observed from case 2 that BNM was able to predict some actors that have similar structural attribute with legitimate actors that is, those considered to be unapproachable in terms of vulnerability and key players. Deterministic approach does not take nodes that have the least metrics as vulnerable nodes

or key player. But through prediction, this had been achieved that is, legitimate actors that have low betweenness metrics were predicted to be vulnerable.

Figure 4.17 presents assessment on the BNM's detection when betweenness centrality was used as input. The Figure shows that BNM had 23 TP, 15 FP, 18 FN and 4 TN. This scores BNM 0.5609 TPR that is, detection probability, 0.4390 FNR that is, miss rate; 0.7894 FPR that is, probability of false alarm (Pfa), 0.2105 TNR for specificity; 0.6052 precision and 0.45 accuracy.

		True Condition (9/11 network)			
	Total population (60)	Condition positive (Conspirators)	Condition negative (Attackers)	$Prevalence =$ 0.6833	$ACC =$ 0.45
Predicted condition (btw)	Predicted condition positive (High MAP)	TP 23	FP 15	$Precision =$ 0.6052	$FOD =$ 0.3947
	Predicted condition negative (Low MAP)	FN 18	TN 4	$FOR =$ 0.8181	$NPV =$ 0.1818
		$TPR =$ 0.5609	$FPR =$ 0.7894	$LR+=$ 0.7105	$DOR =$ 0.3407
		$FNR =$ 0.4390	$TNR =$ 0.2105	$LR-=$ 2.0853	$F1\ score =$ 0.5822

Figure 4.17: Performance Evaluation of BNM's Detection with Case 2

Case 3: Closeness Centrality Input in BNM

Figure 4.18 is the MAP distribution of 9/11 network with Case 3. The MAP represents participants' propensity to evasion predicted by BNM using closeness centrality - eqn. (3.23). Actor ID 15 has the least MAP value of 392.2. And it did not form structural equivalent with any actor. Actor IDs 5 and 37 formed structural equivalents on MAP

value of 397.5. Actor ID 35 emerged on MAP value of 397.7. IDs 47, 48 and 49 formed a structurally equivalent layer on MAP value of 397.8.

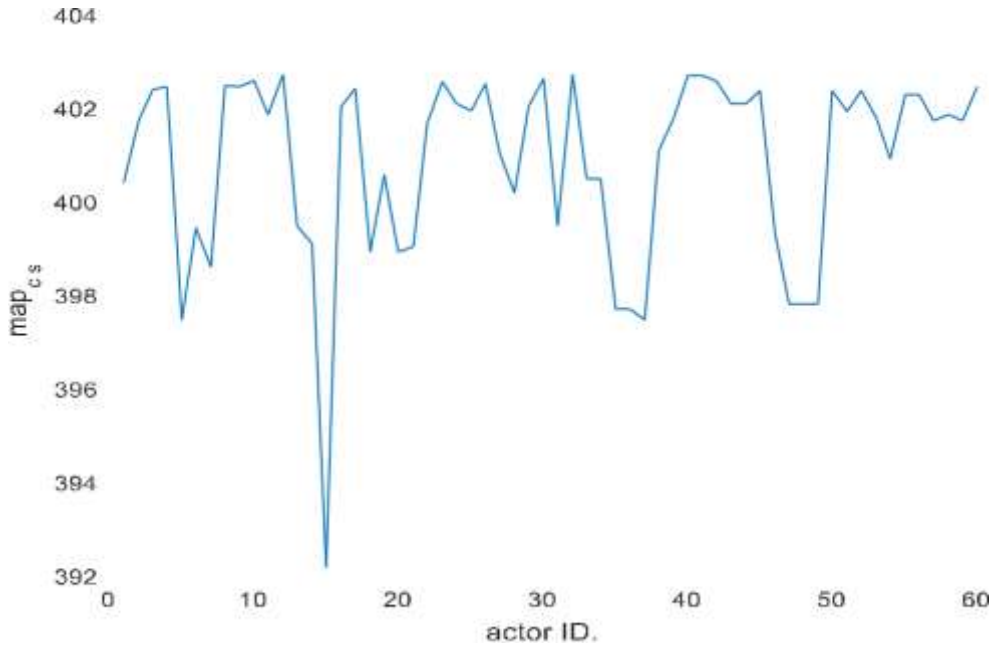

Figure 4.18: MAP Distribution of the 9/11 Network with Case 3

Four (4) actors: 12, 32, 40 and 41 are outliers. They emerged on the highest MAP value of 402.7 in which they become structurally equivalent. These actors were identified in Case 1 but only three re-occurred in Case 2 that is Figures 4.14 and 4.16. Actor IDs 32, 40, and 41 are conspirators and conspirators are bound to be evasive.

Inferred central participants represent vulnerable participants, those engaged in indicting activities. An indicting activity encompasses interacting with criminals too, that is what closeness centrality strives to depict. Those inferred as central participants also have high closeness metrics, which implies that prediction by BNM agrees with deterministic approach that base vulnerability on nodes that have high closeness metric value.

Figure 4.19 presents assessment on the BNM's detection when closeness centrality was the input to BNM. The Figure shows that BNM had 22 TP, 9 FP, 19 FN and 10 TN. This gives BNM 0.5366 TPR that is, detection probability, 0.4634 FNR that is, miss rate; 105

0.4737 FPR that is, probability of false alarm (Pfa), 0.5263 TNR for specificity; 0.7097 precision and 0.5333 accuracy.

		True Condition (9/11 network)			
	Total population (60)	Condition positive (Conspirators)	Condition negative (Attackers)	*Prevalence* = 0.6833	*ACC* = 0.5333
Predicted condition (cls)	Predicted condition positive (High MAP)	TP 22	FP 9	*Precision* = 0.7097	*FOD* = 0.2903
	Predicted condition negative (Low MAP)	FN 19	TN 10	*FOR* = 0.6552	*NPV* = 0.3448
		TPR = 0.5366	*FPR* = 0.4737	*LR+*= 1.1327	*DOR* = 1.2866
		FNR = 0.4634	*TNR* = 0.5263	*LR−*= 0.8805	*F1 score* = 0.6111

Figure 4.19: Performance Evaluation of BNM's Detection with Case 3

Case 4: Eigenvector Centrality Input in BNM

Figure 4.13 is the MAP distribution of the 9/11 network with Case 4. It presents prediction on participants' tendency to evasion using eigenvector centrality - a tool designed for measuring influence of actors from communication social network. It was the virtual influence used for predicting evasiveness of participants in terrorism. Some criminal activities and roles are inherently covert, like influence and leadership roles. Such activities cannot be assessed and used directly to determine vulnerable actors. The BNM was invoked to identify participants like legitimate actors and conspirators that direct deterministic approach could fail to identify.

Actor ID 15 has the least MAP value of 392.9. It implies that actor ID 15 is the most vulnerable actor by prediction using eigenvector metrics. Actor IDs 26, 29 and 14 have

MAP values of 395.8, 396.9, and 398.4 respectively. These IDs queued behind actor ID 15. Prediction was based on metrics quantifying criminal covert roles. The inclusion of conspirators' IDs 26 and 29 those expected to be less-vulnerable reflect merit of BNM over deterministic approach, as some of conspirators could not be detected.

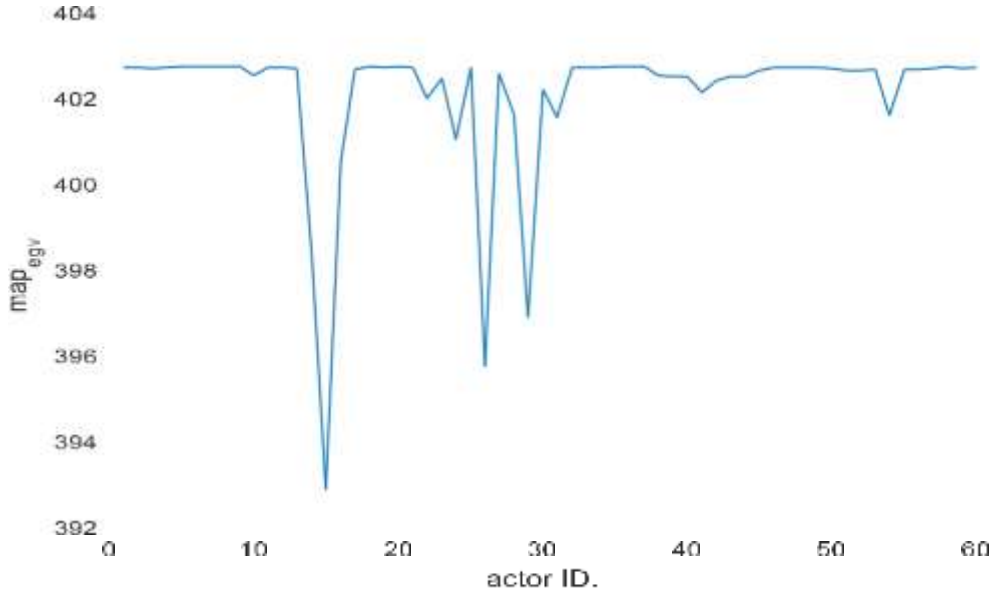

Figure 4.20: MAP Distribution of the 9/11 Network with Case 4

Thirty-five (35) actors with IDs 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 17, 18, 19, 20, 25, 32, 33, 34, 35, 36, 37, 47, 48, 49, 50, 51, 53, 55, 56, 57, 58, 59 and 60 emerged as outliers and also being structurally equivalent on MAP value 402.7. They were predicted as less vulnerable; that is, actors that their covert activities were insignificant. Some of the inferred actors in case 4 had also been inferred as evasive nodes in case 1 and case 2 that is, Figure 4.14 and Figure 4.16 respectively.

Figure 4.21 presents performance evaluation of the BNM's detection when eigenvector centrality was used as input. The Figure shows that BNM had 20 TP, 15 FP, 21 FN and 4 TN. This gives BNM 0.4878 TPR that is, detection probability, 0.5122 FNR that is, miss rate; 0.7894 FPR that is, probability of false alarm (Pfa), 0.2105 TNR for specificity; 0.5714 precision and 0.4 accuracy.

		True Condition (9/11 network)			
	Total population (60)	Condition positive (Conspirators)	Condition negative (Attackers)	*Prevalence* = 0.6833	*ACC* = 0.4
Predicted condition (eigen)	Predicted condition positive (High MAP)	TP 20	FP 15	*Precision* = 0.5714	*FOD* = 0.4286
	Predicted condition negative (Low MAP)	FN 21	TN 4	*FOR* = 0.84	*NPV* = 0.16
		TPR = 0.4878	*FP R* = 0.7894	*LR+*= 0.6179	*DOR* = 0.2539
		FNR = 0.5122	*TNR* = 0.2105	*LR−*= 2.4332	*F1 score* = 0.5263

Figure 4.21: Performance Evaluation of BNM's Detection with Case 4

Summary of Inferred Nodes from the 9/11 Network

Table 4.5 presents the list of actor IDs considered for status verification among participants of the 9/11 network. They are categorised into central and evasive participants. The central participants consist of actor IDs with low MAP values while the evasive participants are those with the highest MAP values.

For central participants, six (6) actor IDs is listed for Case 1 - Figure 4.14, twenty-two (22) actor IDs is listed for Case 2 - Figure 4.16, Seven (7) actor IDs is listed for Case 3- Figure 4.18 and finally, four (4) actor IDs is listed for Case 4 - Figure 4.20. For Case 1 - Figure 4.14, actor IDs is from the first three lowest MAP values. The list in Case 2 - Figure 4.16 contains actor IDs that are structurally equivalent on a single MAP of zero. The list of actor IDs in Case 3 - Figure 4.18 are IDs drawn from the first four lowest MAP values. And Case 4 contains actor IDs that made the first four least MAP values. 108

Table 4.5: Nodes Detected by Inference from the 9/11 Network Using BNM Algorithm

Figure Number	Central participants	Evasive participants
CASE 1 - Figure 4.14	14, 54, 15, 3, 25, 41	1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 18, 20, 23, 28, 30, 32, 33, 34, 35, 36, 37, 38, 40, 42, 45, 46, 47, 48, 49, 51, 53, 55, 56, 57, 58, 59, 60
CASE 2 - Figure 4.16	1, 5, 6, 18, 23, 24, 26, 33, 34, 35, 36, 37, 43, 44, 46, 47, 48, 49, 53, 57, 58, 59	2, 4, 7, 8, 9, 10, 11, 12, 14, 16, 19, 20, 21, 22, 25, 28, 29, 30, 32, 39, 40, 41, 42, 45, 50, 51, 52, 54, 60
CASE 3 - Figure 4.18	15, 5, 37, 35, 47, 48, 49	12, 32, 40, 57
CASE 4 - Figure 4.20	15, 26, 29, 14	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 17, 18, 19, 20, 21, 25, 32, 33, 34, 35, 36,37, 47, 48, 49, 50, 51, 53, 55, 56, 58, 59, 60

Under central participants, actor ID 15 occurs thrice, actor IDs 5, 14, 26, 35, 37, 47, 48, and 49 occur twice while the remaining IDs appear once. Mohamed Attah becomes the most conspicuous. His re-occurrence buttresses that he is a significant actor among all central participants. ‘‘The most important role was played by Mohamed Atta who was on the flight AA11 that crashed into the World Trade Centre North…’’ (Latora and Marchiori, 2004). He was also described as ring leader of this conspiracy (Kreb, 2002).

Actor IDs 5, 14, 26, 35, 37, 47, 48, and 49 are Salem Alhazmi, Marwan Al-Shehhi, Mounir Moutassadiq, Osama Awadallah, Mohamed Abdi, Jean-Marc Grandvisir, Abu Zubeida, and Nizar Trabelis respectively. Some actors inferred as central participants through Case 2 - Figure 4.16 were found to be marginal actors. And marginal actors are potential legitimate actors. Salem and Marwan were hijackers. Salem was on the flight AA77 and Marwan was on the flight AA175. The remaining six actors were conspirators.

Actor IDs 1, 3, 6 and 18 are another set of inferred central participants. They are Majed Moqed, Hani Hanjour, Ahmed Alnami and Wail Alshehri. Majed and Hani were on flight AA77, Ahmed was on AU93 and Wail was in AA11. They were flight hijackers. Hani Hanjour, actor ID 3 has relatively high links and was inferred through Case 1 - Figure

4.14. But Majed, Ahmed and Wail are marginal actors identified as central participants through Case 2 - Figure 4.16. This indicates that marginal nodes were inferred as central participants through Case 2 - Figure 4.16 that deterministic approach has failed to detect them.

The following are conspirators; actor IDs 23, 24, 25, 29, 33, 34, 36, 41, 43, 44, 46, 53, 54, 57, 58 and 59. Actor ID 29 was detected in Case 4 - Figure 4.20 and actor ID 54 was detected in Case 1 - Figure 4.14. Case 2 - Figure 4.16 detected the rest of actor IDs. Agu Budiman and Essid Sami Ben Khemail are actor IDs 29 and 54. Both are high nodal degree participants. Both were inferred from Case 4 - Figure 4.20 and Case 1-Figure 4.14 respectively.

Mamoun Darkazanli, Zakariya Essabar, Said Bahaji, are actor IDs 23, 24 and 25 respectively. Bandar Alhazmi, Faisal Al Salmi and Abdussattar Shaikh are actor IDs 33, 34 and 36. Abu Qatada, Jerome Courtaillier, David Courtaillier and Abu Walid are actor IDs 41, 43, 44 and 46. Lased Ben Heni, Fahid al Shakri, Madjid Sahoune, and Samir Kishk are actor IDs 53, 57, 58 and 59 respectively. Said Bahaji and Abu Qatada have seven links, Zakariya Essabar has 5 links, Jerome and David Courtaillier have four links. The number of links from three upward could make actors become significant in the network as well as making them emerge as central participants or key players.

Mamoun Darkazanli, and Abu Sattar Sheik have three links. Faisal Al Salmi, Bandar Alhazmi, Abu Walid, and Lased Ben Heni have two links. Finally, Fahid, al Shakri, and Sakir Kishk have a link each. Although three links downward may be as insignificant or relatively low for placing an actor as a key player under deterministic approach. Such actors are always regarded as marginal actors. But through the BNM inference, marginal participants were inferred among the central participants. In a nutshell, this shows more

actors acceded to central participants. It has included conspirators those less expected to be key players. Secondly, inferring of marginal actors - those that SNA-based detections failed to ascribe significance or importance. Through prediction, marginal actors become vulnerable, and there is no longer a shield for legitimate actors.

The list of IDs under evasive participants of Table 4.5 were drawn from the highest MAP values. Case 3 - Figure 4.18 has the least number of evasive IDs which is four (4); Case 1– Figure 4.14 has thirty-eight (38) nodes, Case 2 – Figure 4.16 has twenty-nine (29) nodes, and Case 4 -Figure 4.20 has thirty-five (35) nodes. They were predicted by inference as less-susceptible actors.

Figure 4.22 presents a Venn diagram showing places of occurrence of actor IDs presented under evasive participants in Table 4.5. The Venn diagram includes list of inferred actors in Case 3 - Figure 4.18. But the Case's number was not shown due to scattering of those actors in other Cases' list. The Venn diagram summarize re-occurrence of the actor IDs in Case 1 - Figure 4.14, Case 2 - Figure 4.16 and Case 4 - Figure 4.20:

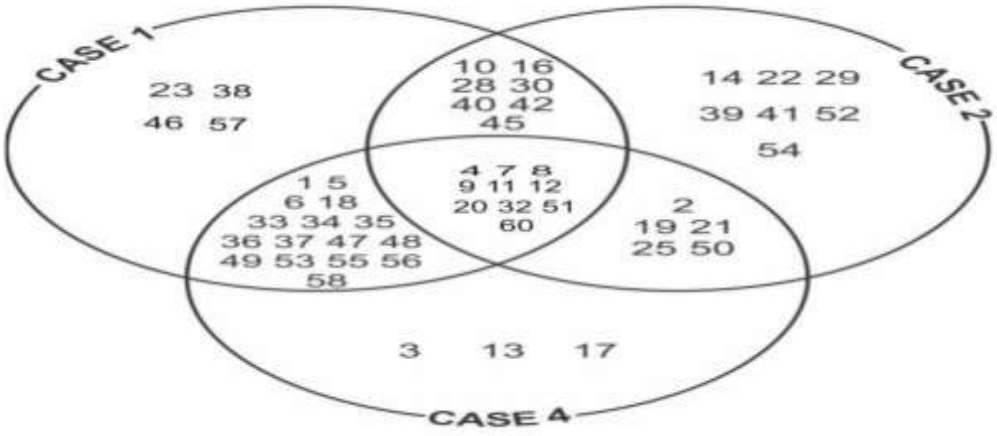

Figure 4.22: Venn Diagram of Evasive Nodes in the 9/11 Network

Actor IDs: 4, 7, 8, 9, 11, 12, 20, 32, 51 and 60 re-occurred in all the three cases. These represent Nawaf Alhazmi, Ahmed Alghamdi, Saeed Alghamdi, Hamza Alghamdi,

Mohand Alshehri, Fayez Ahmed, Raed Hijazi, Rayed Mohammed Abdullah, Mohammed Bensakhria, and Kamel Daoudi respectively. The first six actors are flight hijackers. But only actor IDs 12 and 32 appeared in Case 3 - Figure 4.12 which is Fayez Ahmed and Rayed Mohammed Abdullah.

The following actors re-occurred twice: Those that appeared in Case 1- Figure 4.14 and Case 4 - Figure 4.20 are IDs: 1, 5, 6, 18, 33, 34, 35, 36, 37, 47, 48, 49, 53, 55, 56, 58, and 59. And these are Majed Moqed, Salem Alhazmi, Wail Alshehri, Bandar Alhazmi, Faisal Al Salmi, Osama Awadallah, Abdussattar Shaikh, Mohamed Abdi, Jean-Marc Grandvisir, Abu Zubeida, Nizar Trabelsi, Lased Ben Heni, Seifallah ben Hassine, Essoussi Laaroussi, Madjid Sahoune and Samir Kishk respectively.

Actor IDs 10, 16, 28, 30, 40, 42 and 45 appeared in Case 1-Figure 4.14 and Case 2-Figure 4.16. They are Ahmed Al Haznawi, Abdul Aziz Al-Omari, Ramzi Bin al-Shibh, Ahed Khalil Ibrahim Samir Al-Ani, Tarek Maaroufi, Djamal Benghal and Ahmen Ressam.

Another set of actors appeared twice in Case 2- Figure 4.16 and Case 4 - Figure 4.20 with the following IDs: 2, 19, 21, 25, and 50. These represent Khalid Al-Mihdhar (KAM), Satam Suqami, Nabil al-Marabh, Said Bahaji and Haydar Abu Doha. The first two actors are hijackers. Khalid Al-Mihhar (KAM) was identified as a facilitator (Karthika and Bose, 2011).

The following actor IDs only occurred once. Actor IDs 23, 38, 46, 57 occurred in Case 1 - Figure 4.14. The IDs represent Mamoun Darkazanli, Mohamed Belfas, Abu Walid and Fahid al Shakri. Actor IDs 14, 22, 29, 39, 41, 52 and 54 occurred in Case 2 - Figure 4.16. They are identified as Marwan Al-Shehhi, Mustafa Ahamend al-Hisawi, Agus Budiman, Imad Eddin Baraat Yarkas, Abu Qatada, Mohammed Bensakhria, and Essid Sami Ben Khemail. Finally, actor IDs 3, 13 and 17 only occurred in CASE 4-Figure 4.20 alone.

actors are Hani Hanjour, Zaid Jarrah, and Waleed Alshehri. Mustafa Ahamend al-Hisawi was said to be sponsor of the 9/11 while Zaid Jarrah was a pilot. The two were actors identified as being evasive.

Notice that Mustapha Ahamend al Hishawi appeared less important in the network, yet he was a sponsor, which make him very relevant to the survival of the group. Again, Zaid Jarrah appeared in inconspicuous in the network but he was a pilot. This made him close to important member of the group - hence his detection with eigenvector attribute. All this demonstrate the strength of the BN model.

4.2.2 Results of classification of the 9/11 participants using SNA-Q algorithm

Results of classification of participants in 9/11 criminal group using SNA-Q algorithm is presented here. Four cases of paired network attributes were used. The results are presented according to cases A to D that is, a pair of network attributes were used as inputs to SNA-Q model. Quadrants ae designated as following as: Q1 for prominent, Q2 for the most prominent, Q3 for inconspicuous and Q4 for less-prominent. The point of interests is Q2 and Q4. Outcome of the classification in Q2 and Q4 were used to validate the actor IDs detected in BN model.

Case A: Degree and Betweenness Centrality Inputs

Figure 4.23 presents the classification of participants in the 9/11 criminal group. The SNA-Quadrant model was tested using degree and betweenness centrality of participants in the 9/11. Q1 has thirteen (13) nodes: 28, 14, 25, 41, 60, 40, 32, 9, 8, 24, 52, 29 and 39. This denotes vulnerable actors on account of having high degree centrality but low covert roles – that is, relevant actors with respect to participation in heinous activities denoted by high degree centrality but less-important in roles considered to be covert roles. These participants could be regarded as gateway terrorists as they have potential to lead or

champion the course of terrorisms. There is high propensity for actors in Q1 to replace any ex-communicated terrorists in Q2.

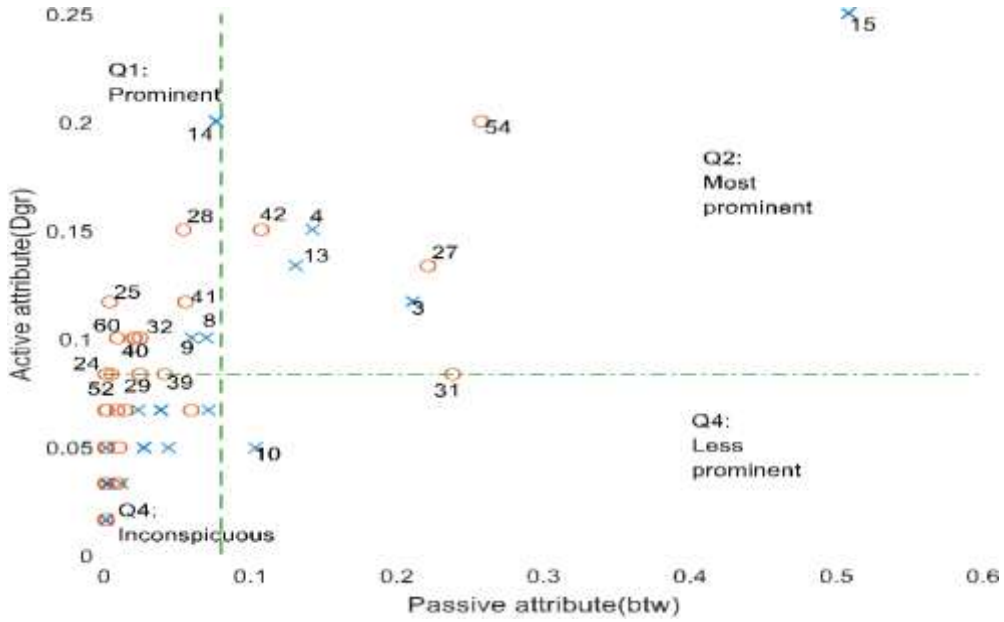

Figure 4.23: SNA-Quadrant Classification of the 9/11 Criminal group with Case A

Q2 the second quadrant, contains actors that are grossly involved in indicting activities as well as covert activities. Some factional leaders, financial managers and experts in special skills are likely to be in this class. Factional leaders, for instance, need frequent relation with subordinates, which gives him prominence. Distribution of resources could make a participant become prominent. Prominence could also come through communications. But the provision of logistic, tactics, motivation and orders could be seldom that is, once in a while and unnoticed. For instance, an informant might be infrequent. Such a role is regarded as passive. When non-frequent activities become noticeable, it makes participants become vulnerable. But when it is below a noticeable threshold, the actor becomes less susceptible to security operatives.

The significance of Q2 is that actors' vulnerabilities lie with prominence in both two attributes. That is the vulnerability of these actors depends on their exorbitant activities.

Q2 has seven (8) nodes: 15, 54, 42, 4, 13, 27, 3 and 31. Out of the actors in Q2, four (4) are attackers: 3, 4, 13, 15 and four (4) conspirators: 27, 31, 42, and 54.

Q3 is for errand nodes and few fugitives. Based on the description of active and passive attributes, both attributes usually fall below average. And the implication is that they are less indictable.

The last quadrant Q4 is the quadrant of smart criminals. Actors here are less vulnerable as an assessment of active attribute will be insignificant. It falls below average of all frequent relationships, and activities. They are different from Q3, their passive attributes are noticeable and significant. A fugitive is less suspected. Q4 has only actor ID 10 that is, Ahmed Haznawi who was an attacker.

Case B: Degree and Closeness Centrality Inputs

Figure 4.24 is the node classification of participants in the 9/11 criminal group using degree and closeness centrality as inputs to SNA Quadrant model. Here, degree centrality was retained as an activity that make participants become vulnerable, while closeness was used as one that hide or lower actor's susceptibility.

Excessive relationships, sharing of ideas or resources with criminals could raise actors' susceptibility. But due to vast relationships among criminals and unavailability of data about criminal relationships, this may prevent security operatives from veracity of criminal closeness. Closeness constitutes parts of activities that conceal participants. For instance, proximity could be obstructed when exact number of overt criminals an actor relate with is unknown. There could be discrepancy between the proximity in a network graph and that of geographical proximity. Two actors could be geographically close, but they could be far apart in network graph.

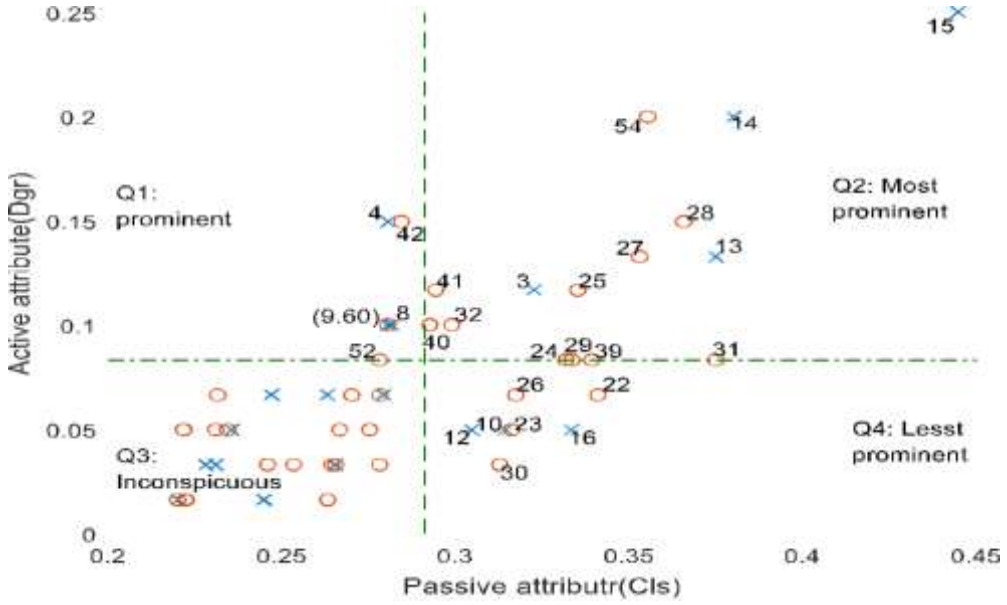

Figure 4.24: SNA-Quadrant Classification of the 9/11 Criminal group with Case B

Sharing resources using weak ties (kinships) conceal actors who are really close. There was a report of a drug cartel where the network leader had one of his associates as his son. It was found that the son had a higher proximity than his father. And the father had low proximity because he did not engage in a phone conversation with his associates including the son as his son did with others (Bright, Greenhill, Reynolds, *et al.*, 2015).

Q1 has six (6) actors consisting of three (3) attackers and three (3) conspirators. The attackers: 4, 8, 9 are Nawaf Alhazmi, Saeed Alghamdi and Hamza Alghamdi. The conspirators: 42, 52 and 60 are Djamal Benghal, Mohammed Bensakhria, and Kamil Daoudi respectively.

Q2 has fourteen (15) actors: 15, 54, 14, 28, 27, 13, 41, 3, 25, 32, 40, 24, 29, 39 and 31. Four (4) are attackers:3, 13, 14 and 15; and the remaining ten are conspirators. Q3 has remaining thirty-five (35) entrapped actors These actors were regarded as less indictable or inconspicuous actors.

Q4 has seven actors: 30, 26, 23, 22, 16, 12 and 10. They were identified as Ahed Khalil Ibrahim Samir Al-Ani, Mounir El Motassadeq, Mamoun Darkazanli, Mustapha Ahamend al-Hisawi, Abdul Aziz Al-Omari, Fayez Ahmed and Ahmed Al Haznawi respectively. These actors include both attackers and conspirators. The conspirators are 30, 26, 23 and 22 while the attackers are 10, 12 and 16.

Case C: Eigenvector and Betweenness Centrality Inputs

Figure 4.25 presents classification of participants in the 9/11 using eigenvector and betweenness centralities as inputs to SNA-Quadrant model. Eigenvector centrality represents active attribute while betweenness represent passive attributes. The peculiarity of the combination lies with leaderships or influence as it affects key players. Influence or leadership roles at disposal of individual vary. Its possession and usage sometime indict actor's vulnerability.

High prominence could be ascribed to actors who exorbitantly exhibit his leadership roles than those who are not. Between centrality was used for passive attribute here to represent activities that key players - conspirators and leaders - enjoy most because of being less indicting. The combination is to show that some of high-profile criminals possess some attributes that lower their susceptibilities. An affiliate terrorist may avoid leaderships roles that is, indicting activities, but they cannot do without partaking in some activities considered as lessen to security operative attention.

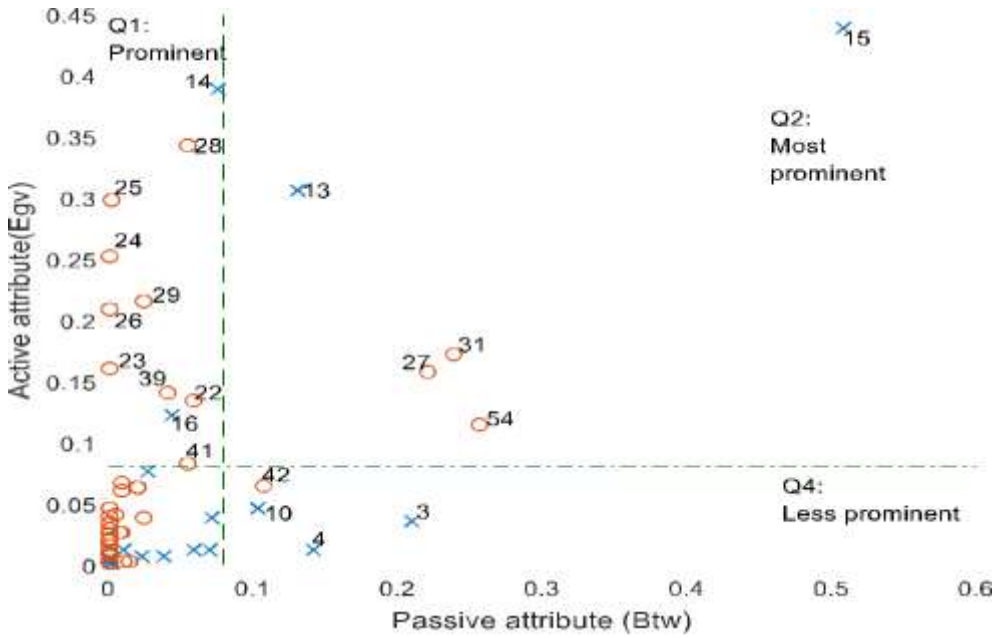

Figure 4.25: SNA-Quadrant Classification of the 9/11 Criminal group with Case C

Q1 has eleven (11) actors: 14, 16, 22, 23, 24, 25, 26, 28, 29, 39, and 41. There are two (2) attackers, and nine (9) are conspirators. Q2 has five (5) actors: 13, 15, 27, 31 and 54. Two (2) are attackers and three (3) are conspirators. Q4 has four (4) actors: 3, 4, 10 and 42. Three (3) out of these actors are attackers.

Case D: Eigenvector and Closeness Centrality Inputs

Figure 4.26 presents node classification for 9/11 using eigenvector and betweenness centralities. Q1 has no actors within it. Q2 has sixteen (16) actors: 13, 14, 15, 16, 22, 23, 24, 25, 26, 27, 28, 29, 31, 39, 41 and 54. Only (4) out of this list are attackers, the remaining twelve actors are conspirators. Q4 has six (6) actors: 3, 10, 30, 32 and 40.

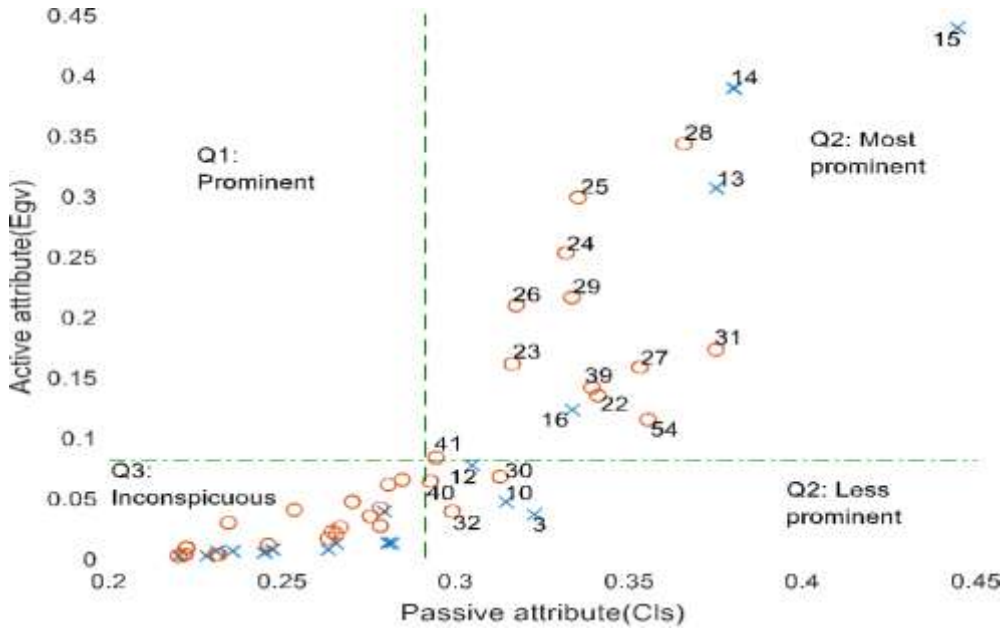

Figure 4.26: SNA-Quadrant Classification of the 9/11 Criminal group with Case D

Distributions of Actors in Quadrants of SNA-Q models

Table 4.6 presents distribution of nodes in SNA-Q models. These were quantified in percentage. Percentage of nodes in each quadrant was given. This was done across all Q1 to Q4. Number of nodes under Q1 to Q4 was divided by sum of total nodes in SNA-Q model.

Table 4.6: Distribution of 9/11 participants in the SNA-Quadrant Model

Q-model variables	Distribution				Total
	Q1(%)	Q2(%)	Q3(%)	Q4(%)	
Case A - Figure 4.23	18.3	13.3	66.7	1.7	100
Case B - Figure 4.24	8.3	23.3	58.3	10.0	100
Case C - Figure 4.25	18.3	8.3	66.7	6.7	100
Case D - Figure 4.26	0.0	25.0	65.0	10.0	100

Q1 is the quadrant that has the least percentage of zero and the highest percentage of 18.3. It was designated as prominent for describing participants that engaged more in indicting activities than covert activities. The actors in Q1 are dispensable.

Q2 has percentage between 8.3 and 25.0. Actors in Q2 and Q1 are equal vulnerability level. But in terms of relevancy to the criminal group, actors in Q2 are more important than Q1 because, Q2 actors had more covert roles that Q1 actors. Detection and removal of actors in Q2 may be effective for disrupting a criminal network.

Q3 has the highest percentage of node distribution. It has 66.7 percent as the highest and 58.3 as lowest percentage. Actors in Q3 were assumed to be less conspicuous and highly elusory.

Q4 has the least percentage across all the model variables. Its highest percentage is 10.0 and the lowest is 1.7 percent. The quality of actors in Q4 is more significant than their quantity. Actors in Q4 are less-vulnerable due to low or incessantly participation in indicting activities. But they are important to the group because they played covert roles too. High profile members aiding resistance to security perturbation are few compared to the entire group. And the importance of high-profile members lies in the ability to evade security detection. This can contribute to crime persistence.

Finally, the following are actor IDs that reoccurred in quadrant of Case A through Case D. There are two types of participants involved: attackers and conspirators. The two types of participants were found in Q4. An attacker with ID 10 is the only one consistently appeared in the Q4 from Case A - Figure 4.23 to case D-Figures 4.26. A conspirator with ID 30 emerged in Q4 of Case B - Figure 4.24 and Case D - Figure 4.26 only. An attacker with ID 3 appeared in Q4 of Case C – Figure 4.25 and Case D -Figure 4.26. It was also observed in Q2 of Case A – Figure 4.23 and Case B - Figure 4.24. Actor ID 4 occurred

in Q4 of Case C - Figure 4.25. A conspirator with ID 40 featured in Q4 of Case D - Figure 4.26. A conspirator with ID 42 featured in Q4 of Case C - Figure 4.25.

4.2.3 Verification of BNM inferred nodes in the 9/11 network

This section used the SNA-Q algorithm to verify and validate set of nodes detected through the BNM algorithm for the 9/11 criminal group. It also validates the performance of BNM. There is a strong correlation between set of nodes inferred by the BNM and SNA-Quadrant. SNA-Q served as criminal profiles in absence of real data about participants in OCGs. The verification was carried on set of actors inferred as central participants and evasive participants.

(i) ***Verification of BNM Inferred Central Participants Using SNA-Q Model***

Table 4.7 presents nodes inferred as central participants in 9/11 criminal network by BNM and SNA-Q. The Table contains forty (40) inferred actors. Twenty-eight (28) out of forty (40) are inferred by the BNM. Twenty-one (21) was inferred by the SNA-Q. It shows that BNM provided additional nineteen (19) actors that did not fall within Q2 of SNA-Q. The actors are identified as 1, 5, 6, 18, 33, 34, 35, 36, 37, 43, 44, 46, 47, 48, 49, 53, 57, 58, and 59.

Ten (10) out of forty (40) listed actors are inferred by both the BNM and SNA-Q. The IDs are 3, 14, 15, 23, 24, 25, 26, 27, 41 and 54. Actor ID 15 has the highest re-occurrence. It occurred in seven (7) cases: thrice under BNM cases and four times under SNA-Q which implies that actor ID 15 is an important central participant. Its emergence through case A to D confirm that, it is an actor that has influence, participated in indicting activities and also play covert roles for the group.

Table 4.7: Comparison of Inferred Central Nodes from the 9/11 Network

9/11 Participants			BNM Algorithm				SNA-Q Algorithm (Q2)			
Actor Name	Actor-ID	Type	Case 1	Case 2	Case 3	Case 4	Case A	Case B	Case C	Case D
Majed Moqed	1	A		YES						
Hani Hanjour	3	A	YES				YES	YES		
Nawaf Alhazmi	4	A					YES			
Salem Alhazmi	5	A		YES	YES					
Ahmed Alnami	6	A		YES						
Ziad Jarrah	13	A					YES	YES	YES	YES
Marwan Al-Shehhi	14	A	YES			YES		YES		YES
Mohamed Atta	15	A	YES		YES	YES	YES	YES	YES	YES
Abdul Aziz Al-Omari	16	A								YES
Wail Alshehri	18	A		YES						
Mustafa Ahamend al-Hisawi	22	C								YES
Mamoun Darkazanli	23	C		YES						YES
Zakariya Essabar	24	C		YES				YES		YES
Said Bahaji	25	C	YES					YES		YES
Mounir El Motassadeq	26	C		YES		YES				YES
Zacarias Moussaoui	27	C				YES	YES	YES	YES	YES
Ramzi Bin al-Shibh	28	C						YES		YES
Agus Budiman	29	C						YES		YES
Lofti Raissi	31	C						YES	YES	YES
Rayed Mohammed Abdullah	32	C						YES		
Bandar Alhazmi	33	C		YES						
Faisal Al Salmi	34	C		YES						
Osama Awadallah	35	C		YES	YES					
Abdussattar Shaikh	36	C		YES						
Mohamed Abdi	37	C		YES	YES					
Imad Eddin Baraat Yarkas	39	C						YES		YES
Tarek Maaroufi	40	C						YES		
Abu Qatada	41	C	YES					YES		YES
Djamal Benghal	42	C					YES			
Jerome Courtaillier	43	C		YES						
David Courtaillier	44	C		YES						
Abu Walid	46	C		YES						
Jean-Marc Grandvisir	47	C		YES	YES					
Abu Zubeida	48	C		YES	YES					
Nizar Trabelsi	49	C		YES	YES					
Lased Ben Heni	53	C		YES						
Essid Sami Ben Khemail	54	C	YES				YES	YES	YES	YES
Fahid al Shakri	57	C		YES						
Madjid Sahoune	58	C		YES						
Samir Kishk	59	C		YES						
			6	22	7	4	7	15	5	16

Actor ID 14 and 26 re-occurred twice under BNM's cases. Actor ID 14 has its profile description in case B and D while actor ID 26 has its profile in case D only. It means that both actors are significant. The case D denotes participants that have significant influence and also have significant covert role offered to the group. Case B provide additional feature that actor ID 14 has. Case B denotes feature for being involved in indicting activities and covert roles.

The remaining five (5) actor IDs: 3, 23, 24, 25 and 41 occurred once either in case 1 or 2. Actor IDs 24, 25 and 41 have their profiles description in case B and D. The similar digital profile description with actor ID 14. This trio are conspirators. Actor ID 3 has its profile description in case A and case B. The case A is description of participant that is actively participated in indicting activities and covert roles – intermediary. Actor ID 13 is an attacker. Actor ID 23 has profile under case D which denotes being prominent through influence and proximity combined.

Case 2 of BNM has the highest number of inferred central participants given as twenty-two (22). Some of the actors include those that SNA-Q regarded their profile as inconspicuous. They are also regarded as legitimate actors. It is obvious that roles of legitimate actors may be difficult to quantify. That is the reason they are missing out in Q2 – simply because they have low participation; they interact less with criminal members or they have inconspicuous relationship with criminal group member. Nevertheless, they are facilitators and sometimes aid crime commission.

Out of the twenty-two (22) inferred actor IDs 23 and 24 occurred once under BNM. Both are confirmed by SNA-Q. Significance of actor ID 23 was given in case D and that of actor ID 24 was obtained in both case B and case D. The two actors are conspirators.

Actor ID 26 is another that inferred as a central participant through case 2 and its importance was confirmed through case D.

Actor IDs 5, 35, 37, 47, 48 and 49 were inferred under case 2 and case 3 but they did not fall within case 2 of SNA-Q which implies that SNA-Q profile is insufficient in providing profile status for nodes inferred by BNM. SNA-Q criminal profile classification covers all six (6) actors inferred under case 1 and four (4) actors inferred under case 4. But profile of only one actor ID out of seven in case 3 and three (3) actor IDs in case 2 out of twenty-two (22).

(ii) ***Verification of BNM Inferred Evasive Participants Using SNA-Q Model***

Table 4.8 presents comparison of evasive nodes detected by the BNM algorithm with SNA-Q. The Table shows actors inferred as being evasive with those inferred in Q4 of SNA-Q. Recall that quadrant Q4 designates important participants who are less vulnerable to security operative detection. This is to identify potential fugitive in the 9/11 network. The Table has fourteen (12) actors.

Actor IDs 12 and 32 occurred as evasive nodes in all cases of BNM. Profiles of actor ID 12 is obtained in case B and D while profile of actor ID 32 is obtained in case D. Actor IDs 4 and 40 occurred in three cases of BNM. Their profiles are given in C and D respectively. The description of participants in Q4 of case C and case D is for those partake less in indicting activities but active in furnishing the group with covert role.

Actor IDs 10, 16, 30 and 42 occurred twice in cases of BNM. SNA-Q confirmed their profile status as smart actors under case A to for actor ID 10; case B and case D for actor ID 30; case B for actor ID 16 and case C for actor ID 42.

Actor IDs 3, 22 and 23 occurred once in BNM's cases. Both actor IDs 3 and 23 were confirmed as smart participants. They were also inferred and confirmed as central participants. Actor ID 22 was inferred as evasive participant confirmed as smart actor. It was found that as financial aider to the group. This implies that not all key players occupy central position of criminal structure.

Table 4.8: Comparison of Inferred Evasive Nodes from the 9/11 Network

9/11 Participants			BNM Algorithm				SNA-Q Algorithm: Q4			
Actor Name	Actor-ID	Type	Case 1	Case 2	Case 3	Case 4	Case A	Case B	Case C	Case D
Hani Hanjour	3	A				YES			YES	YES
Nawaf Alhazmi	4	A	YES	YES		YES			YES	
Ahmed Al Haznawi	10	A	YES	YES			YES	YES	YES	YES
Fayez Ahmed	12	A	YES	YES	YES	YES		YES		YES
Abdul Aziz Al-Omari	16	A	YES	YES				YES		
Mustafa Ahamend al-Hisawi	22	C		YES				YES		
Mamoun Darkazanli	23	C	YES					YES		
Mounir El Motassadeq	26	C						YES		
Ahed Khalil Ibrahim Samir Al-Ani	30	C	YES	YES				YES		YES
Rayed Mohammed Abdullah	32	C	YES	YES	YES	YES				YES
Tarek Maaroufi	40	C	YES	YES	YES					YES
Djamal Benghal	42	C	YES	YES					YES	
			9	9	3	4	1	7	4	6

4.3 Analysis of BNM Algorithm's Performance

The BNM algorithm was tested with two criminal datasets. Its performance was collected and examined with respect to defined cases of inputs used in the BN model. Variety of analysis carried out on of BNM permits to evaluate its achievement with datasets on two criminal groups. The analysis was divided into three subsections. The first sub section presents summary of direct evaluation metrics about BNM's detection performance. The second sub section presents indirect evaluation on BNM's detection performance using

Receiver Operating Characteristic (ROC) plotted as Signal to Noise Ratio (SNR). Finally, BNM algorithm was compared with the entropy variation algorithm.

4.3.1 Summary of direct assessment metrics on BNM's performance

Table 4.9 presents summary of metrics obtained from direct assessment of BNM's performance. The value of prevalence is the same for all the inputs under each network. There is prevalence of 0.6833 and 0.2727 for 9/11 and N'17 respectively. The precision was low under N'17 and it was significant under 9/11 network.

Table 4.9: Summary of BNM's Performance Metrics

	N17 network				Sept 11 network			
	BNM (dgr)	BNM (btw)	BNM (cls)	BNM (egv)	BNM (dgr)	BNM (btw)	BNM (cls)	BNM (egv)
TPR	0.8333	1	0.6667	0.5	0.6341	0.5609	0.5366	0.488
FNR	0.1667	0	0.3333	0.5	0.3659	0.439	0.4634	0.512
FPR	0.8125	0.5625	0.8125	0.875	0.6316	0.7895	0.4737	0.79
TNR	0.1875	0.4375	0.1875	0.125	0.3684	0.2105	0.5263	0.211
Prevalence	0.2727	0.2727	0.2727	0.2727	0.6833	0.6833	0.6833	0.683
Precision (PPV)	0.2778	0.4	0.2352	0.1764	0.6842	0.6053	0.7097	0.571
FOR	0.25	0	0.4	0.6	0.6818	0.8182	0.6552	0.84
LR+	1.0256	1.7778	0.8205	0.5714	1.0041	0.7106	1.1328	0.618
LR-	0.8889	0	1.7778	4	0.993	2.0854	0.8805	2.433
Accuracy (ACC)	0.3636	0.5909	0.3182	0.2273	0.55	0.45	0.5333	0.4
FDR	0.7222	0.6	0.7647	0.8235	0.3158	0.3947	0.2903	0.429
NPV	0.75	1	0.6	0.4	0.3182	0.1818	0.3448	0.16
DOR	1.1538	0	0.4615	0.1429	1.0111	0.3407	1.2865	0.254
F1 score	0.4167	0.5714	0.3478	0.2609	0.6582	0.5823	0.6111	0.526

Table 4.9 shows that the highest TPR 1 is obtained under N17 network with BNM (btw) and followed by 0.833 and 0.6341 obtained for BNM (dgr) under N17 and the 9/11 network respectively, while the least TPR (probability of detection) is 0.4878 obtained by BNM (egv) under the 9/11 network. The highest probability of false alarm (Pfa) FPR is 0.875 obtained by BNM (egv) under the N17 network, follow by 0.812 recorded for

both BNM (dgr) and BNM (cls) under the N17 network. And 0.4736 is the least FPR recorded under 9/11 for BNM (cls). And BNM (cls) has the highest TNR of 0.5267 under the 9/11 network.

Precision values are very low under N17 network and the highest value of 0.7096 is recorded under 9/11 network with BNM (cls) follow by 0.6818 obtained by BNM (dgr). BNM (egv) has the highest false omission rate of 0.84 and followed by BNM (btw) recorded 0.81981 FOR under the 9/11 network. And BNM (btw) has 0 false omission rate under N17 network. The lowest record of FOR implies that all nodes selected as conspirators are true conspirators. And the BNM (btw) had 0 for those missed detection while BNM (dgr) has 0.25 of those missed detection.

The highest accuracy recorded is 0.5909 by BNM (btw) under N17 network. BNM (dgr) and BNM (cls) has 0.55 and 0.533 scores respectively under 9/11 network. And the BNM (egv) has the least accuracy value of 0.2272. The highest false discovery rate (FDR) is 0.8235 recorded by BNM (egv) and the least is 0.6 by BNM (btw) under N17 network. But 9/11 has low FDR values. The least is 0.2903 and the highest is 0.4285 for BNM (cls) and BNM (egv) respectively.

Finally, 9/11 network has the highest metric of F1 score of 0.6582 by BNM (dgr), 0.6111 for BNM (cls), 0.5822 for BNM (btw) and 0.5263 for BNM (egv). And N17 network recorded the least as 0.2608 by BNM (egv), 0.3478 by BNM (cls), 0.416 value by BNM (dgr) and its highest as 0.5714. All show that there are slight differences from the direct evaluation carried out on the BNM detector using different attributes as inputs. It simply implies that, some attribute yet unknown may have significant impact BNM detection.

The variation of the algorithm's performance under different network attributes could be due to the different criminal nodes profiles (e.g., prominent, inconspicuous and evasive).

It could be that certain profiles are better described by certain network node attribute. Again, the nature of data the datasets used could be a significant factor in the overall performance evaluation. Datasets of criminals are prone to missing data, outliers, sundry inconsistency, all of which is attributed to hidden nature of criminal activities. Nevertheless, it is a good consolidation that the developed BNM performed appreciably well in the detection of covert nodes with ambiguous characteristics.

4.3.2 Performance assessment on attributes used in enhanced BNM

This section provides probability of detection ( ) against probability of false alarm for each attribute used in detection ( ). This is presented in terms of Signal-to-Noise Ratio (SNR) curves obtained through Receiver Operating Characteristic (ROC). A detector's performance is measured from its ability to achieve a certain probability of detection, and probability of false alarm . A given SNR value, 'ROC_SNR' function automatically calculates and values that is linear or square law detector can achieve using a single pulse. The in Table 4.9 was used as signal pulse to generate the ROC_SNR curves. Evaluation was carried on the two criminal networks and different attributes used for their detection.

(i) ***EnBNM's Performance on N'17 Network***

Probability of detection against probability of false alarm was carried on four attributes used as inputs when experimenting EnBNM's algorithm.

Performance on Case 1: Degree Centrality

Figure 4.27 presents performance of EnBNM using degree centrality attribute with of 0.833. Probability of false alarm increases as probability of detection increases. of 0.2, is obtained at of 10^{-2} and of 0.6 is obtained at of 10^{-1}.

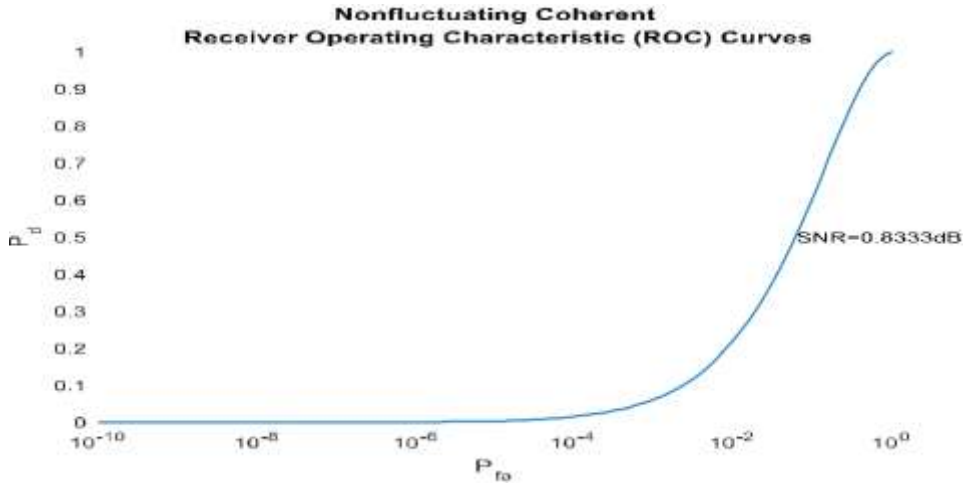

Figure 4.27: Detection Probability against False Alarm Detection Case 1of N'17

Performance on Case 2: Betweenness Centrality

Figure 4.28 presents performance of EnBNM using betweenness centrality attribute with of 1 meaning all key actors were detected using the betweenness centrality as input. The probability of false alarm started emerged around from 10^{-4}, and increases as the probability of detection increases. of 0.22, is obtained at of 10^{-2} and of 0.65 is obtained at of 10^{-1}.

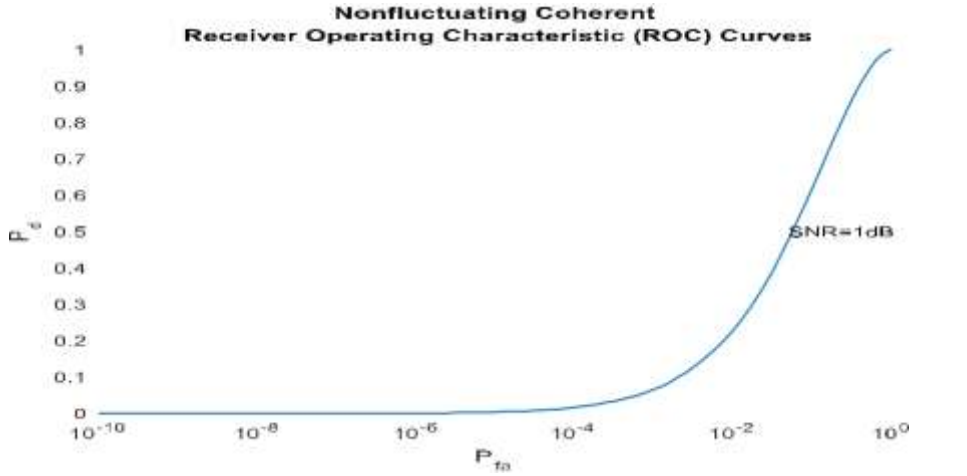

Figure 4.28: Detection Probability against False Alarm Detection Case 2 of N'17

Performance on Case 3: Closeness Centrality

Figure 4.29 presents performance of EnBNM under closeness centrality attribute with of 0.666 shown in the graph. With the same trend of increases with . Probability of

detection, of 0.2, with probability of false alarm of 10^{-2} and of 0.6 is obtained at of 10^{-1}.

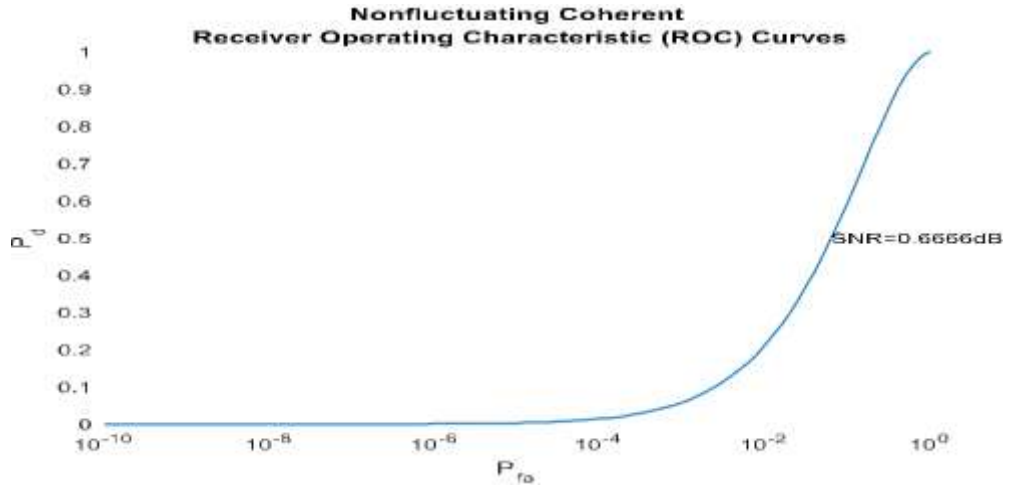

Figure 4.29: Detection Probability against False Alarm Detection Case 3 of N'17

Performance on Case 4: Eigenvector Centrality

Figure 4.30 presents performance of EnBNM under eigenvector centrality attribute with of 0.5 shown in the graph. Probability of detection, of 0.2 was obtained when probability of false alarm is 10^{-2} and of 0.6 is obtained at of 10^{-1}.

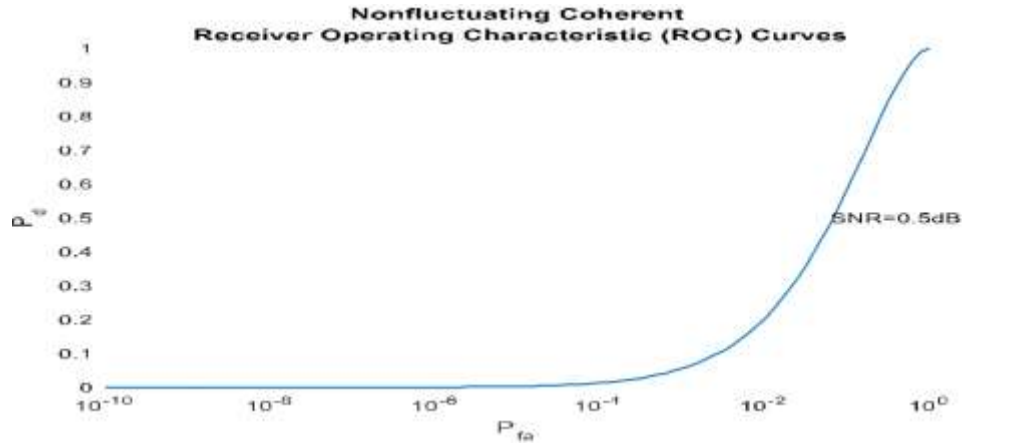

Figure 4.30: Detection Probability against False Alarm Detection Case 4 of N'17

(ii) *EnBNM's Performance on 9/11 Network*

Graphs of probability of detection to probability of false alarm over detected members of 9/11 terrorist group were used in examine EnBNM's performance. The slight differences in each graph is contribution of each attribute in the algorithm.

Performance on Case 1: Degree Centrality

Figure 4.31 presents performance of EnBNM using degree centrality attribute with of 0.6341. The Figure presents probability of detection over probability of false alarm over detection of participants in N'17 network. of 0.2 was obtained at of 10^{-2}
and of 0.6 at of 10^{-1}.

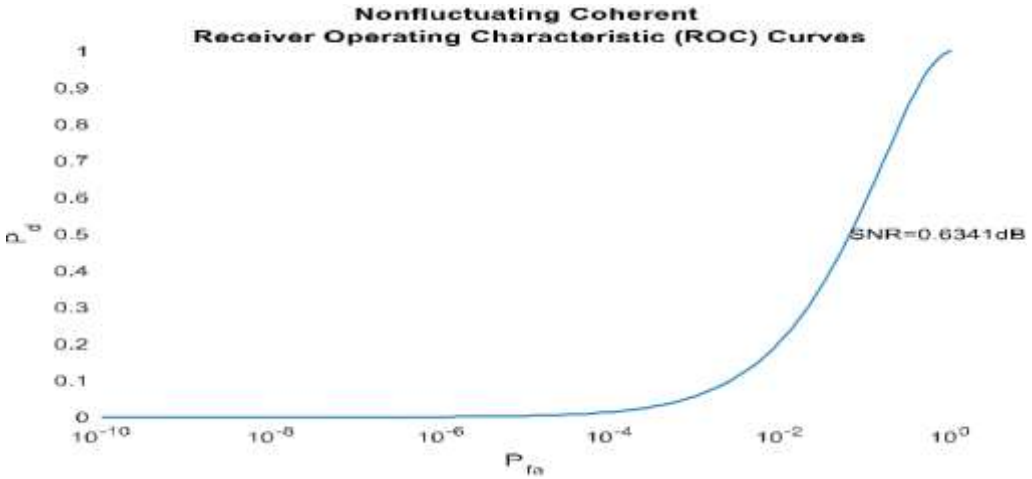

Figure 4.31: Detection Probability against False Alarm Detection Case 1 of 9/11

Performance on Case 2: Betweenness Centrality

Figure 4.32 presents performance of EnBNM using betweenness centrality attribute with of 0.5609. The curve is for against under testing of 9/11 terrorist network.

of 0.2 was obtained at of 10^{-2} and of 0.6 at of 10^{-1}.

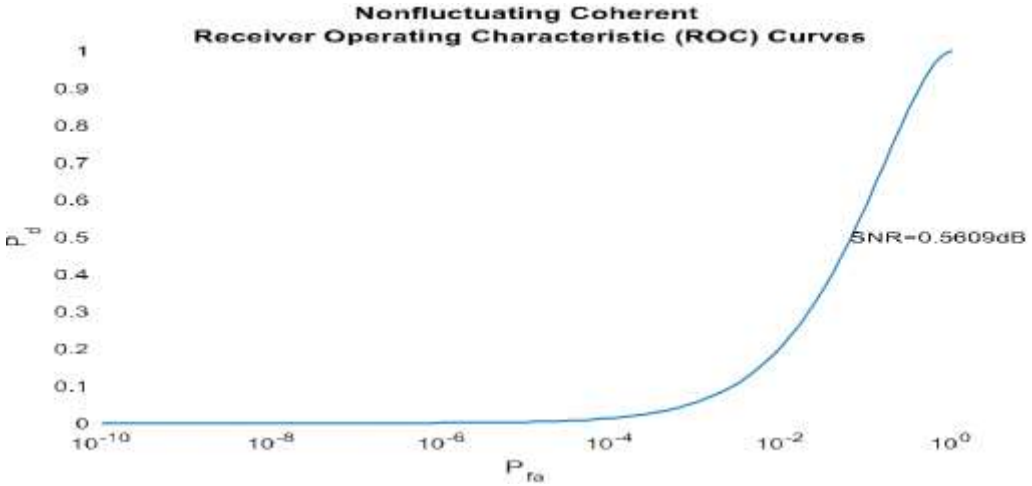

Figure 4.32: Detection Probability against False Alarm Detection Case 2 of 9/11

Performance on Case 3: Closeness Centrality

Figure 4.33 presents performance of EnBNM using closeness centrality attribute with of 0.5366. It shows that of 0.2 was obtained at of 10^{-2} and of 0.6 at of 10^{-1}.

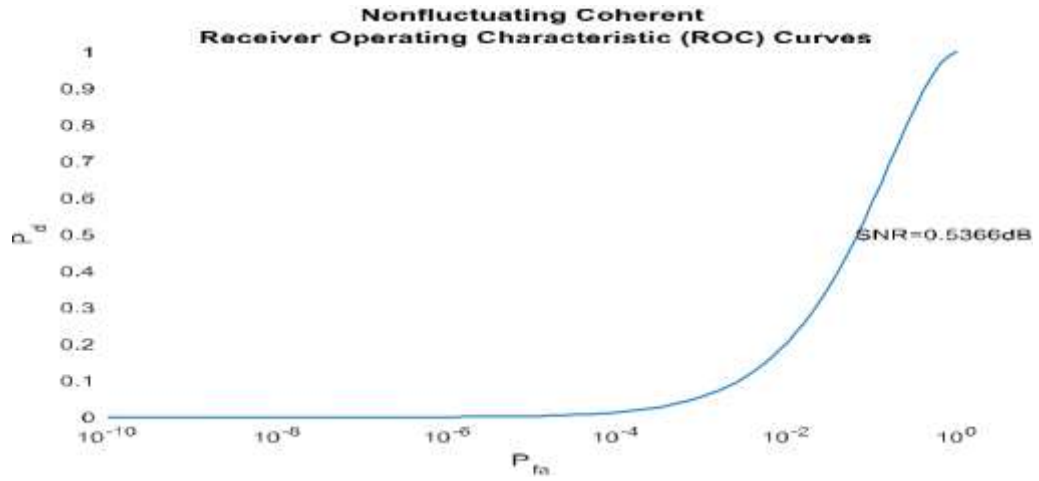

Figure 4.33: Detection Probability against False Alarm Detection Case 3 of 9/11

Performance on Case 4: Eigenvector Centrality

Figure 4.34 presents performance of EnBNM using eigenvector centrality attribute with of 0.4878. This is the least detection probability from all attributes used in EnBNM. The probability of detection shows that 0.2 of was obtained at of 10^{-2} and of 0.6 at of 10^{-1}.

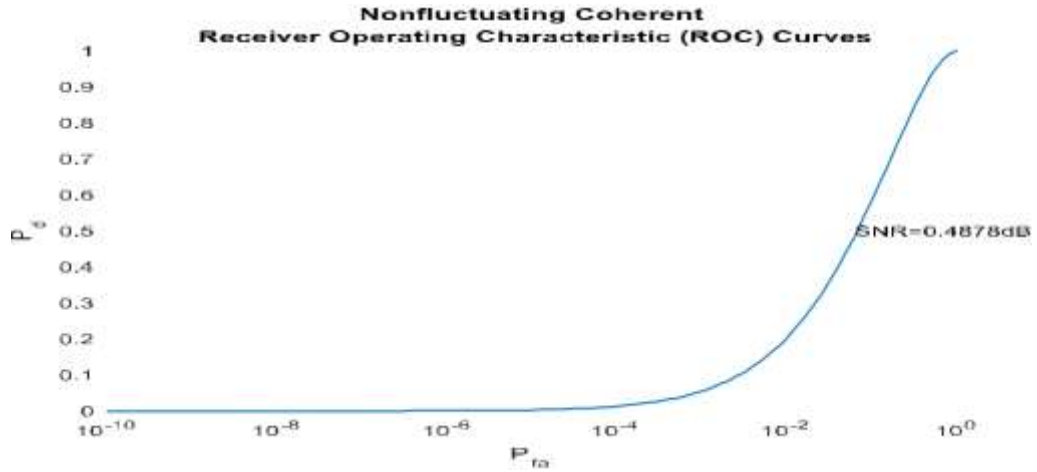

Figure 4.34: Detection Probability against False Alarm Detection Case 4 of 9/11 132

These values are the same with values obtained in Figure 4.29 which is curve for 9/11 using eigenvector centrality. It implies that eigenvector centrality is a poor attribute so far from the four attributes used in predicting covert members.

Probability of detection against probability of false alarm presented through Figures 4.27 to 4.34 reveal property of attributes used in predicting covert nodes. Eigenvector centrality has the least TPR of 0.5 and 0.4878 from Table 4.9 under N'17 and 9/11 network respectively. These produce of 0.2 at $= 10^{-2}$ and of 0.6 obtained at $= 10^{-1}$. Degree centrality that had 0.8333 and 0.6341 TPR under N'17

and 9/11 networks also had of 0.2 at $= 10^{-2}$ and of 0.6 at $= 10^{-1}$. Betweennss centrality is only attribute that had 100 percent TPR, which was obtained under N'17 network. This gives of 0.22 at $= 10^{-2}$ and of 0.65 at $= 10^{-1}$. It implies that input has little impact in influncing detection of EnBNM.

4.3.3 Comparison of BNM algorithm with entropy variation algorithm

The EnBNM algorithm was compared with the entropy variation algorithm using three networks of different sizes: kite network, the N17 network and the 9/11 network. Algorithm were compared on indication to identify detected node using only degree centrality as input. The results were presented separately to observe the trade off in the two algorithms over the dataset sizes.

(i) Comparison of Algorithms of BNM and Entropy Variation Using the Kite Network Dataset

Figure 4.35 presents entropy variations of the Kite network (Appendix F). It shows that actor ID 8 has the lowest depression of entropy connectivity and entropy centrality. This is a node that its removal can cause the network to disintegrate into two sub networks. Removal of actor ID 7, a central node cannot cause breaking up of the network. Even

though it has a high centrality value it is not regarded as key player under entropy connectivity and entropy centrality. Note that from entropy curves, the lowest point of depression depicts a node that its absence in the network will lower entropy of the network.

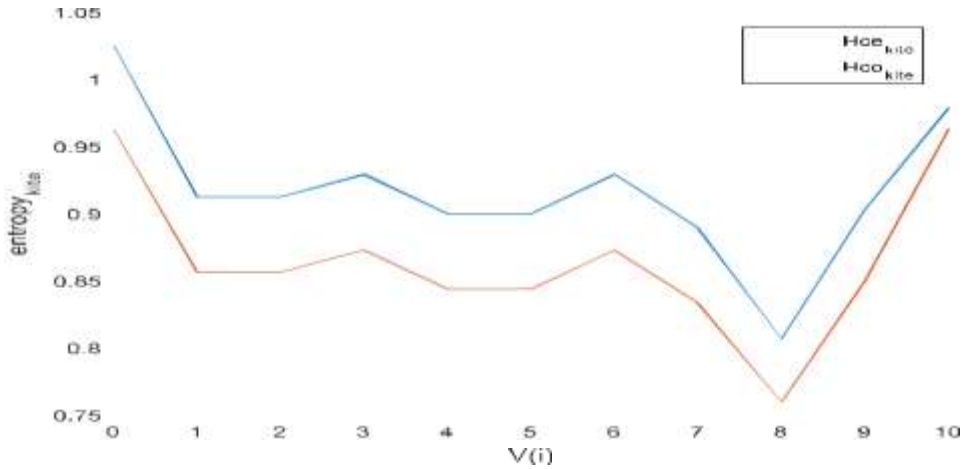

Figure 4.35: Entropy variation of the kite network

Figure 4.36 is MAP distribution of the kite network. The Figure identified ID 7 as a central node with lowest depression. From Figures 3.35 and 3.36 show that different actor IDs were identified by entropy variation algorithm and BNM. It is obvious and conspicuous depression of BNM - Figure 3.36 is sharper than one in entropy variation curves - Figure 3.35.

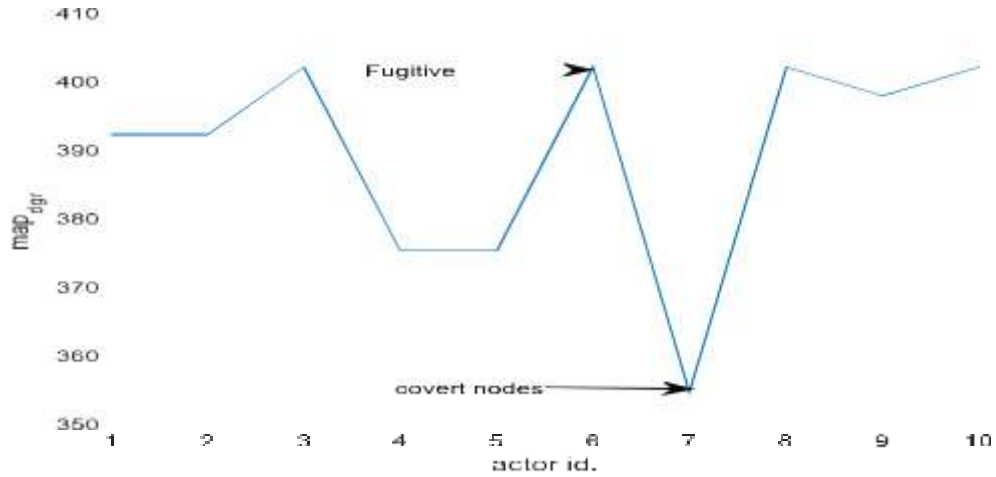

Figure 4.36: MAP Distribution of the Kite Network

(ii) Comparison of BNM and Entropy Variation Algorithm Using the N'17 Network Dataset

Figure 4.37 presents entropy variation of N17 network. Actors with lowest depression are assumed to be key players in entropy algorithm. According to the entropy algorithm, lowest depression nodes will cause serious disruption in the network structure when removed. The Figure contains both entropy centrality and entropy connectivity .

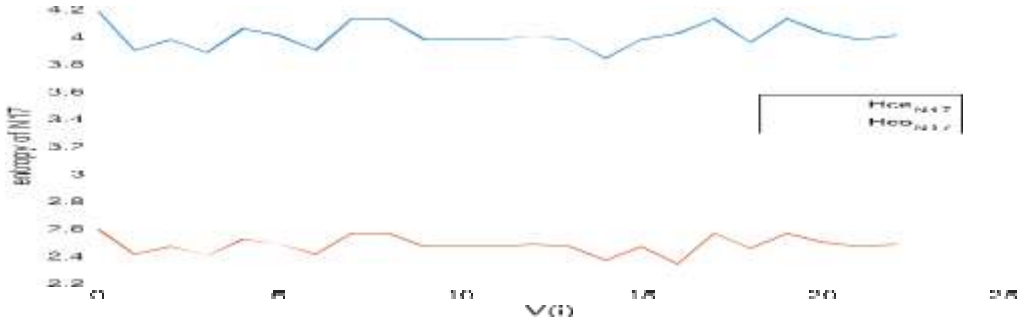

Figure 4.37: Entropy variation of N17 network

Four actors are noticed in entropy centrality and five (5) actors are noticed in the entropy connectivity considering depressions in the graph. The actors are 1, 3, 6, 14, and 16 for but include all except 16. Actor 16 is Pavlos Serifis. These are the same set of actors inferred as central participants in Case 1-Figure 4.1 and Case 3-Figure 4.3. Figure 4.38 and Figure 4.39 present snapshots of data points and prevalence of central actors. Figure 4.38 and Figure 4.39 have actor ID 1, 3, 6, 14 and 16 as key players. Both algorithms identified the same set of actors.

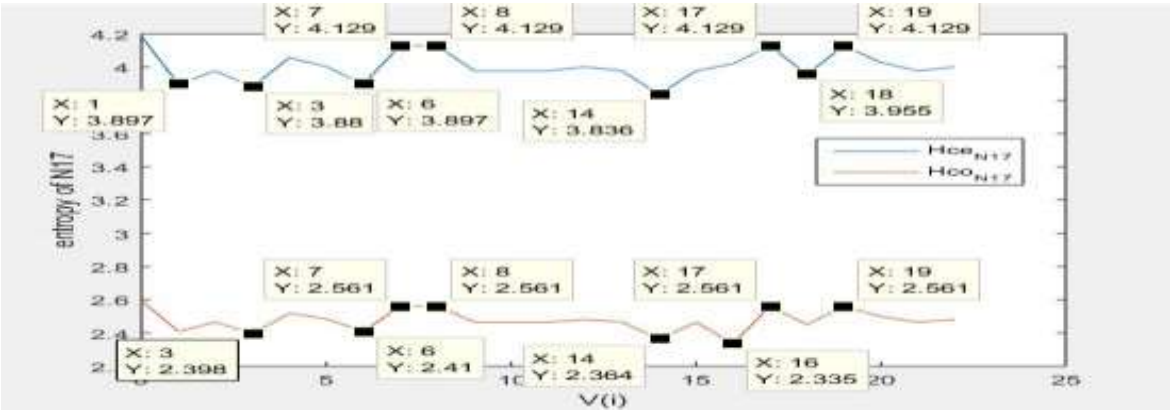

Figure 4.38: Snapshot of cursor data for the N17 entropy variation

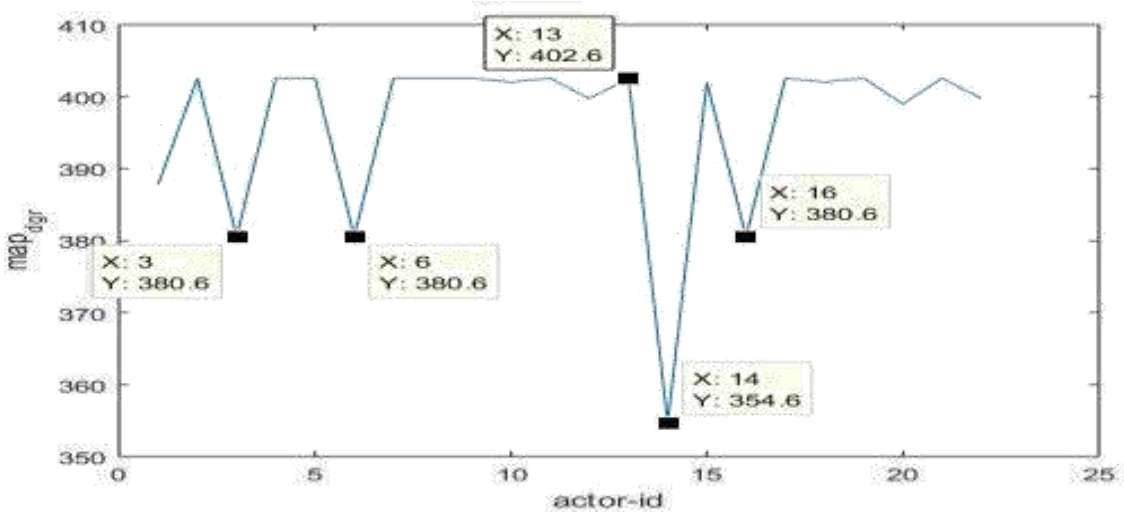

Figure 4.39: Snapshot of cursor data for MAP Distribution of the N'17

(iii) Comparison of BNM and Entropy Algorithms using the 9/11 Network Dataset

Figure 4.40 presents graph of entropy variation of the 9/11 network. Actor ID 54 has the lowest entropy value, followed by actor IDs 42 and 28. They were identified as Essid Sami Ben Khemail, Gjamal Bengal and Ramzi Bin al-Shibh respectively. By entropy connectivity and entropy centrality, these are relevant actors that their removal can disrupt the 9/11 network structure. Secondly, the most central actors become faded away as the size of the network increases. Figure 4.41 and 4.42 present snapshots of the data points under entropy variation and MAP distribution respectively.

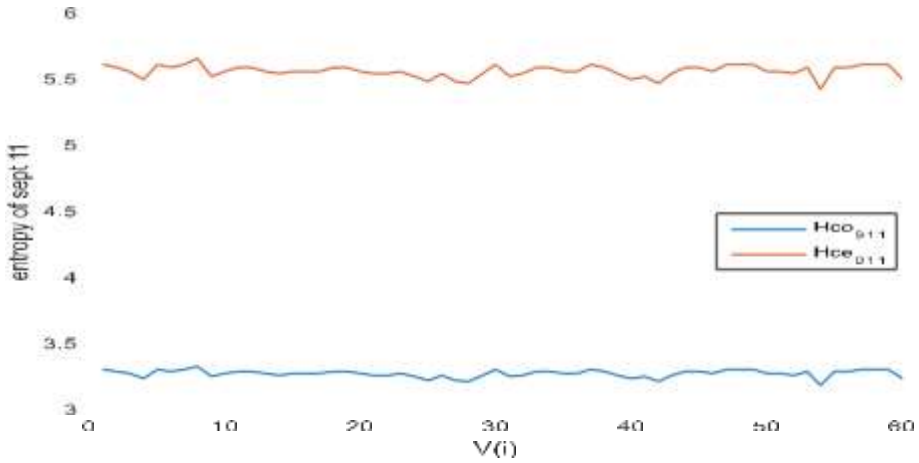

Figure 4.40: Entropy variation of the 911 network

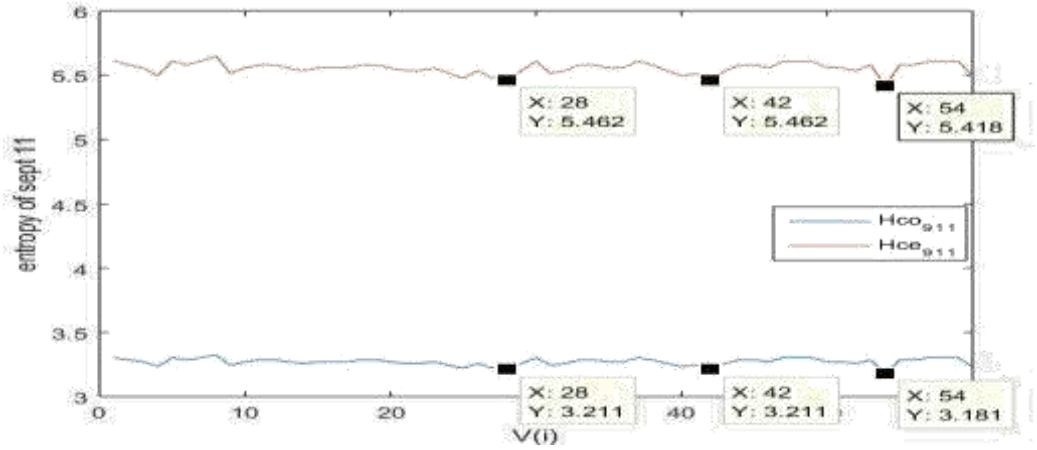

Figure 4.41: Snapshot of cursor data for 9/11 entropy variation

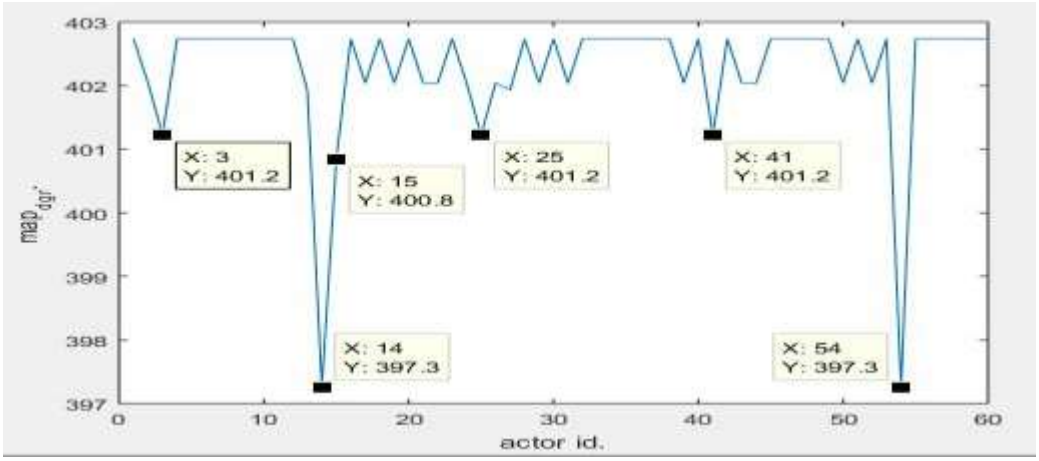

Figure 4.42: Snapshot of cursor data for 9/11 MAP Distribution

The EnBNM compared favourably well with the entropy variation algorithm. Both detect some set of nodes in common. However, a close look at Figures 4.35 and 4.37 of entropy variation curves compared with the EnBNM curve of Figure 4.36 and 4.39 showed that depressions that indicate important points are not sharp under entropy variation. This situation become more obvious in the 9/11network dataset which is a larger dataset than the N'17 datasets. With even larger set, no points will be detectable in the entropy variation curve. This simply means that the BNM can handle larger datasets than the formers. This is a significance performance indicator.

4.4 Summary

This research work presents empirical evidence to support the phenomenon that not all key players in OCGs are central nodes (section 2.6). The developed EnBNM algorithm was able to predict set of nodes using the network attributes of nodes as inputs to the EnBNM. The first inference was made on central participants - nodes vulnerable to detection. The second inference was on set of nodes that are less vulnerable. The set of nodes in the two inferences were analysed to identify key players using SNA-Q classification as ground truth information about participators.

The results shows that there is strong correlation between the EnBNM and SNA-Q; some nodes were detected in common. However, the EnBNM inferred more central participants than the SNA-Q. Case 2 has large portion of inferred central participants. Majority of this actors were classified as inconspicuous by SNA-Q including nodes that having legitimate actors' attributes. Detection of this class participants is unprecedented in covert nodes detection. Finally, detection of financial aiders among evasive confirm that that some key players are not centric actors.

CHAPTER FIVE

5.0 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The development of Bayesian Network Model (BNM) for detection of covert nodes was achieved. The set nodes identified are from central as well as peripheral of a network structure that is, detection was not limited to central nodes alone.

Algorithm for detection of covert nodes based on Bayesian Network was developed. The algorithm presents procedural steps for enhancing Bayesian Network Model for detecting covert nodes.

Evaluation of the Enhanced Bayesian Network Model (EnBNM) algorithm was carried out on two terrorist groups datasets. The Algorithm detected both known and unknown key players. Unknown key players referred to new nodes detected by EnBNM while known key players referred to actors that previous works or literature identified as key players.

The validation was carried out on detected nodes using ground truth data and SNA-Q. The ground truth validates SNA-Q; all actors convicted as leaders by court were also identified as key players by SNA-Q. This SNA-Q detection was also used for validation of detection on 9/11 dataset. The validation was achieved.

Finally, comparison analysis was successfully carried out between EnBNM and entropy variation algorithms using kite, N'17 and 9/11 networks. The comparative analysis revealed strength of EnBNM over entropy over data size. The trade-off is that EnBNM detect more covert nodes with different inputs

5.2 Recommendation

The following are recommended:

1. Extraction of covert networks from telecommunication network is a research area that detection of covert nodes hinged on. Much are still needed to be done.

2. Investigating features and developing techniques that would aid detection of legitimate actors because of their pertinent to crime commission.

3. Finding attributes apart from social network attribute is a research area for mitigating challenges associated with profile status verification of participants in OCGs.

5.3 Contribution to body of knowledge

Through the development of BNM and implementation with terrorist datasets, the following were extracted as contribution to the knowledge:

1. Bayesian network inference identified low degree actors as key players that deterministic approach cannot detect; all key players are not bound to central of a network.
2. Level of participation in criminal activities is incongruent to participants' level of conversation; some actors evade detection due to low or inconspicuous relationship with overt criminals;
3. Identify position of fugitive; some actors that are important to the OCGs but have low propensity to detection;
4. Finally, prediction through BNM inference identify more central participants - key players than SNA-based approaches.

REFERENCES

Abazia, F. (2017) Mapping crime: network analysis of the mala del brenta criminal organization, *American Political Science Association*, 15(3), pp. 45–67.

Agreste, S., Catanese, S., De Meo, P., Ferrara, E. & Fiumara, G. (2016) 'Network structure and resilience of mafia syndicates', *Information Sciences*. Elsevier Inc., 351, pp. 30–47.

Ahsan, M., Singh, T. & Kumari, M. (2015) 'Influential node detection in social network dring community detection', *IEEE Cognitive Computing and Information Processing (CCIP), 2015 International Conference*.

Alvarez, A. J., Herrera, G. C. & Gonz, L. A. (2015) 'Eigencentrality based on dissimilarity measures reveals central nodes in complex networks', *Nature Publishing Group*. Nature Publishing Group, pp. 1–10.

Ashby, M. (2016) Using crime science for understanding and preventing theft of metal from the British railway network. *University of London*.

Barnes, N. (2017) 'Criminal politics: an integrated approach and violence', *American Political Science Association*, 15(4), pp. 967–987.

Basaras, P. (2013) 'Detecting influential spreaders in complex, dynamic networks', *Lincoln Laboratory Journal*, 20(1), pp1-10.

Basaras, P., Iosifidis, G., Katsaros, D. & Tassiulas, L. (2017) 'Identifying influential spreaders in complex multilayer networks: a centrality perspective', *IEEE Transaction on Network Science and Engineering*, (06), pp. 1–8.

Basu, A. (2014) 'Social network analysis: a methodology', *Springer International Publishing Switzerland*, pp. 215–242.

Behzadan, V. (2016) Real-time inference of topological structure and vulnerabilities for adaptive. The University of Nevada Reno.

Behzadan, V., Nourmohammadi, A., Gunes, M. & Yuksel, M. (2017) 'On fighting fire with fire: strategic destabilization of terrorist networks', *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, (iv), pp. 1120–1127.

Belinda, C. (2010) 'Group-based social network characterisation of hidden terrorist networks', *the 1st International Cyber Resilience Conference*, (08), pp. 11–21.

Berlusconi, G. (2013) 'Do all the pieces matter? assessing the reliability of law enforcement data sources for the network analysis of wiretaps', *Global Crime*, 14(January 2015), pp. 61–81.

Berlusconi, G., Calderoni, F., Parolini, N., Verani, M. & Piccardi, C. (2016) 'Link prediction in criminal networks: A Tool for Criminal Intelligence Analysis', *PLOS ONE*, pp. 1–21.

Berzinj, A., Kaati, L. & Rezine, A. (2012) 'Detecting key players in terrorist networks', *European Intelligence and Security Informatics Conference*, pp. 297–302.

Bichler, G., Bernardino, S., Malm, A., Beach, L., Cooper, T. & Bernardino, S. (2017) 'Drug supply networkss: a systematic review of the organizational structure of illicit drug trade', *Global Crime, Routledge Taylor & Francis Group*. Routledge, 06(05), pp. 14–35.

Bliss, N. T. & Schmidt, M. C. (2013) 'Confronting the challenges of graphs and networks', *Lincoln Laboratory Journal*, 20(1), pp1-20.

Blondel, V. D., Decuyper, A. & Krings, G. (2015) 'A survey of results on mobile phone datasets analysis', *EPJ Data Science*, 4(1), pp. 1–55.

Bonacich, P. & Lloyd, P. (2001) 'Eigenvector-like measures of centrality for asymmetric relations', *Social Networks*, 23(3), pp. 191–201.

Borgatti (2006) 'Identifying sets of key players in a social network', *Springer*, pp. 21–34.

Borgatti, S. P., Everett, M. G. & Freeman, L. C. (2012) *UCINET 6 for Windows Software for Social Network Analysis*. pp 1-9.

Bright, D. (2015) 'Disrupting and dismantling dark networks: lessons from social network analysis and law enforcement simulations', in *Illuminating Dark Networks*. pp1-6.

Bright, D., Greenhill, C., Britz, T. & Ritter, A. (2017) 'Criminal network vulnerabilities and adaptations', *Global Crime, Routledge Taylor & Francis Group*. Routledge, 07(2), pp. 1–18.

Bright, D., Greenhill, C., Reynolds, M., Ritter, A. & Morselli, C. (2015a) 'The use of actor-level attributes and centrality measures to identify key actors: a case study of an Australian drug trafficking network', *Journal of Contemporary Criminal Justice*, 31(3), pp. 262–278.

Bright, D., Greenhill, C. & Ritter, A. (2015b) 'Networks within networks : using multiple link types to examine network structure and identify key actors in a drug trafficking operation', *Routledge taylor and Francis Group*, (5), pp. 37–41.

Brunetto, D., Calderoni, F. & Piccardi, C. (2016) 'Communities in criminal networks: a case study', *MOX, Dipartimento di Matematica Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)*, (26).

Burcher, M. & Whelan, C. (2017) 'Social network analysis as a tool for criminal intelligence: understanding its potential from the perspectives of intelligence analysts', *ResearchGate*. Trends in Organized Crime, (May), pp. 1–18.

Butt, W. H., Qamar, U. & Khan, S. A. (2014) 'Hidden members and key players detection in covert networks using multiple heterogeneous layers', *Journal of Industrial and Intelligent Information*, 2(2), pp. 142–146.

Calderoni, F. (2010) 'Strategic positioning in mafia networks', *Joint Research Centre on Transitional Crime, Universita Cattolics del Sacro Cuore di Milano*, pp. 198–199.

Calderoni, F. (2012) 'The structure of drug trafficking mafias : the ' Ndrangheta and cocaine', *Springer Science +Business Media*, 58(9), pp. 321–349.

Campana, P. & Varese, F. (2012) ‘Listening to the wire: criteria and techniques for the quantitative analysis of phone intercepts’, *Springer Science +Business Media*, 15, pp. 13–30.

Carley, K. M., Reminga, J., Kamneva, N. & Carley, K. M. (1998) ‘Destabilizing terrorist networks’, in *Institute for Software Research , Carnegie Mellon University*.pp1-7.

Carter, K. M., Idika, N., Streilein, W. W. & Member, S. (2014) ‘Probabilistic threat propagation for network security’, *IEEE Transactions on Information Forensics and Security*, 9(9), pp. 1394–1405.

Catanese, S., Ferrara, E. & Fiumara, G. (2013) ‘Forensic analysis of phone call networks’, *Social Network Analysis and Mining*, 3(1), pp. 15–33.

Chatterjee, J. (2005) 'The changing structure of organized crime groups' *International Conference on Advances in Social Network Analysis and Mining*, pp. 149–164.

Clauset, A., Moore, C. & Newman, M. E. J. (2008) ‘Hierarchical structure and the prediction of missing links in networks’, *Letters*, 453(5), pp. 98–101.

Clauset, A. & Woodard, R. (2013) ‘Estimating the historical and future probabilities of large terrorist events’, *Annals of Applied Statistics*, 7(4).

Costa, L. da F., Rodrigues, F. A., Travieso, G. & Boas, P. R. V. (2006) Characterization of complex networks: A survey of measurements.pp1-11.

Course, C. & Hill, A. (2014) ‘The case of the terrorist organization november 17 (17n) media and political power’, 17.pp1-23.

Dawoud, K., Alhajj, R. & Rokne, J. (2010) ‘A global measure for estimating the degree of organization of terrorist networks’, *International Conference on Advances in Social Network Analysis and Mining*.

Du, Y., Gao, C., Hu, Y., Mahadevan, S. & Deng, Y. (2014) ‘A new method of identifying influential nodes in complex networks based on TOPSIS’, *Physica A: Statistical Mechanics and its Applications*. Elsevier B.V., 399, pp. 57–69.

Duch, J. & Arenas, A. (2005) ‘Community detection in complex networks using extremal optimization’, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 72(2), pp. 1–4.

Duijn, P. A. C., Kashirin, V. & Sloot, P. M. A. (2014) ‘The relative ineffectiveness of criminal network disruption’, *Scientific Reports*.

Eilstrup-Sangiovanni, M. & Jones, C. (2008) ‘Assessing the dangers of illicit networks: Why al-Qaida may be less threatening than many think’, *International Security*, 33(2), pp. 7–44.

Eiselt, H. A. & Bhadury, J. (2015) ‘The use of structures in communication networks to track membership in terrorist groups’, *Journal of Terrorism Research*, 6(1), pp. 1–18.

Everton, S. F. (2009). Disrupting dark network with social networks analysis. *Expert Systems with Applications*. pp1-13.

Eyal, R., Kraus, S. & Avi, R. (2011) 'Identifying missing node information in social networks', *Twenty-Fifth AAAI Conference on Artificial Intelligence Identifying*, pp. 1166–1172.

Ferrara, E., De Meo, P., Catanese, S. & Fiumara, G. (2014) 'Detecting criminal organizations in mobile phone networks', *Expert Systems with Applications*. Elsevier Ltd, 41(13), pp. 5733–5750.

Fortunato, S. (2010) Community detection in graphs, physics reports. *Complex Networks and Systems Lagrange Laboratory, ISI Foundation*, Viale S. Severo 65, 10133, Torino, I-ITALY. pp1-34.

Freeman, L. (1978) 'Centrality in social networks: conceptual clarification', *Social Networks*, 1, pp. 215–239.

Ghasemi, M., Seidkhani, H., Tamimi, F., Rahgozar, M. and Masoudi-nejad, A. (2014) 'Centrality measures in biological networks', pp. 1–17.

Gliwa, B., Zygmunt, A. & Byrski, A. (2012) 'Graphical analysis of social group dynamics', *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)*, 1, pp. 41–46.

Grassi, R., Calderoni, F., Bianchi, M. & Torriero, A. (2019) 'Betweenness to assess leaders in criminal networks: new evidence using the dual projection approach', *Social Networks*. Elsevier, 56, pp. 23–32.

Gregory, S. (2007) 'An algorithm to find overlapping community structure in networks', 4702 LNAI, pp. 91–102.

Guillaume, J. & Latapy, M. (2006) 'Bipartite structure of all complex networks', *HAL Id : hal-00016855 Bipartite Structure of all Complex Networks*, (5), pp. 215–221.

Gunnell, D., Hillier, J. & Blakeborough, L. (2016) Social network analysis of an urban street gang using police intelligence *Data*. pp 1-16.

Hasan, M. Al, Chaoji, V., Salem, S., Zaki, M. & York, N. (2006) 'Link prediction using supervised learning', *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security*.pp1-14.

Hensen (2011a) 'Calculating and visualizing network metrics', in *NodeXL Tutorial-5*, pp. 69–78.

Hensen (2011b) 'Social network analysis measuring mapping and modeling collections of connections', in *NodeXL Tutorial-3*, pp. 31–50.

Holme, P. (2003) 'Congestion and centrality in traffic flow on complex networks', pp. 1–9.

Hu, P. & Mei, T. (2017) 'Ranking influential nodes in complex networks with structural holes', *Physica A*. Elsevier B.V.pp 1-16.

Huang, D.-W. & Yu, Z.-G. (2017) 'Dynamic-sensitive centrality of nodes in temporal networks.', *Scientific reports*. Nature Publishing Group, 7(41454), pp. 1–11.

Hulst, R. (2009) 'Introduction to social network analysis (SNA) as an investigative tool', *Springer*, 12, pp. 101–121.

Hussain, D. M. A. (2009) 'Predicting hierarchical structure in small world', *ResearchGate*, (012), pp. 1–2.

Hussain, D. M. A. & Ortiz-arroyo, D. (2008) 'Locating key actors in social networks using bayes' posterior probability framework', *Springer-Verlag*, pp. 27–38.

Husslage, B. G. M., Lindelauf, R. & Hamers, H. J. M. (2012) 'Leaderless covert networks: a quantitative approach', *CentER Discussion Paper;2012 Tilburg: Econometrics. General*, 2012–0057, pp. 1–15.

Ilachinski, A. (2005) Self-organized terrorist- counterterrorist adaptive coevolution. *Expert Systems with Applications*.pp1-15.

Interpol (2018) *Organized crime underpins major conflicts and terrorism globally, of illicit flow*. pp1-13. Retrived on 25th of September, 2019.

Ismail, Abideen, Onwuka, E. N., Salihu, B. A. & Ubadike, C. O. (2019a) 'Towards mining of stakeholders in criminal organizations from telecommunication metadata : analytic approach to latent feature extraction', *Journal of science technology and education*, 7(3), pp. 42–48.

Ismail, A., Onwuka, E. N., Salihu, B. A. & Ubadike, O. C. (2017) 'Survey of techniques for detecting covert members of dark networks', in *Proceeding of 2nd International Engineering Conference (IEC2017) Federal University of Technology, Minna, Nigeria*. Minna, pp. 226–233.

Ismail, A, Onwuka, N., Salihu, A. & Ubadike, O. C. (2019b) 'Detecting covert members: quadrant approach for classification and identification of smart criminals', *International Journal of Information Processing and Communication (IJIPC)*, 7(2), pp. 71–82.

Jalayer, M., Azheian, M. & Kermani, Mehrdad Agha Mohammad, A. (2018) 'A hybrid algorithm based on community detection and multi-attribute decision making for influence maximization', *Computers & Industrial Engineering. Elsevier* pp1-18.

Jones, N. P., Dittmann, W. L., Wu, J. & Reese, T. (2018) 'A mixed-methods social network analysis of a cross-border drug network: the Fernando Sanchez organization (FSO)', *Springer Science+Business Media. Trends in Organized Crime.* pp 1-23.

Karthika, S. & Bose, S. (2011) 'A comparative study of social networking approaches in identifying the covert nodes', *International Journal on Web Service Computing (IJWSC)*, 2(3), pp. 65–78.
[

Kassimeris, G. (2007) 'For a place in history: explaining greece's revolutionary organization 17 november by', *The Journal of Conflict Studies*, (September 2002), pp. 129–145.

Kasture, P. (2012) 'Cluster based Outlier Detection', *International Journal of Computer Applications*, 58(10), pp. 11–15.

Keller, F. B. (2015) *Net*works of power: using social network analysis to understand who will rule and who is really in charge of the chinese communist party. *Expert Systems with Applications*.2(6), pp1-23.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E. & Makse, H. A. (2010) 'Identification of influential spreaders in complex networks', *Nature Physics*. Nature Publishing Group, 6(11), pp. 888–893.

Klemm, K., Serrano, M. A., Eguiluz, V. M. & Miguel, M. S. (2012) 'A measure of individual role in collective dynamics: spreading at criticality', *Scientific Reports*, 2(2), pp. 292 -301.

Klerks, P. (2001) 'The network paradigm applied to criminal organisations: theoretical nitpicking or a relevant doctrine for investigators on recent developments in the netherlands', *Connection*, 24(3), pp. 53–65.

Kossinets, G. (2006) 'Effects of missing data in social networks', *Social Networks Elsevier*, 28, pp. 247–268.

Kramer, S. (2007) 'A new method for detecting and tracking covert terrorist networks', *Paragon Science, Inc.*, pp1-46.

Kreb, V. E. (2002) 'Mapping networks of terrorist cells', *Connections 2002*, 24(3), pp. 43–52.

Kriegler, A. (2014) 'Using social network analysis to profile organised crime', *Institute for Security Studies*, (09), pp. 1–8.

Lampe, K. Von (2009) 'Human capital and social capital in criminal networks: introduction to the special issue on the 7th bankensee colloquium', *Springer Science+Business Media*, 12, pp. 93–100.

Latora, V. & Marchiori, M. (2004) 'How the science of complex networks can help developing strategies against terrorism', *Chaos,Solutions and Fractal*, 20, pp. 69– 75.

Le, V. (2012) 'Organised crime typologies: structure, activities, and conditions', *International Journal of Criminology and Sociology*, 1, pp. 121–131.

Lee, M., Lee, J. & Park, J. (2012) 'QUBE: a quick algorithm for updating betweenness centrality', *Proceedings of the 21st international conference on World Wide Web*, pp. 351–360.

Leuprecht, C., Aulthouse, A. & Walther, O. (2016) 'The puzzling resilience of transnational organized criminal networks', *Police Practice and Research*, 17(4).

Levi, M. (2007) 'Organized crime and terrorism'. *Oxford University* Press.pp-45.

Lin, S. & Chalupsky, H. (2003) 'Unsupervised link discovery in multi-relational data via rarity analysis', *Third IEEE International Conference on Data Mining (ICDM'03)*. pp1-23.

Liu, J. J., Lin, J., Guo, Q. & Zhou, T. (2016) 'Locating influential nodes via dynamics-sensitivity centrality', *Scientific Report*. Elsevier B.V., 43(xxxx), pp. 600–614.

Lü, L.& Zhou, T. (2011) 'Link prediction in complex networks: a survey', *Physica A*. Elsevier B.V., 390(6), pp. 1150–1170.

Madeira, M. & Joshi, A. (2013) 'Analyzing close friend interactions in social media', *Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013*, pp. 932–935.

Maeno, Y. (2007) 'Node discovery problem for a social network' *Expert Systems with Applications*. pp. 62–76.

Maeno, Y. (2009) 'Node discovery in a networked organization', *IEEE International Conference on Systems, Man, and Cybernetics*, (October), pp. 3522–3527.

Maeno, Y. & Ohsawa, Y. (2007a) 'Analyzing the covert social network foundation behind terrorism disaster', *International Journal of Services Sciences*, 2(x), p. pp.125-141.

Maeno, Y. & Ohsawa, Y. (2007b) 'Human – computer interactive annealing for discovering invisible dark events', *IEEE Transactions on Industrial Electronics*, 54(2), pp. 1184–1192.

Maksim, T. & Carley, K. M. (2003) 'Bouncing back: recovery mechanisms of covert networks', *casos*, pp. 1–7.

Malm, A. & Bichler, G. (2011) 'Networks of collaborating criminals: assessing the structural vulnerability of drug Markets', *Journal of Research in Crime and Delinquency*, 48(2), pp. 271–297.

Malm, A., Bichler, G. & Nash, R. (2016) 'Co-offending between criminal enterprise groups', *Routledge talour and Francis Gruop*, 0572(06), pp. 112–128.

Manning, J. D. (2010) Dark Networks, *U.S. Army College, Carlisle Barrack*. pp1-78.

Matous, P. & Wang, P. (2019) 'External exposure, boundary-spanning, and opinion leadership in remote communities: A network experiment', *Social Networks*. Elsevier, 56, pp. 10–22.

Memon, N., Wiil, U. K., Alhajj, R., Atzenbeck, C. & Harkiolakis, N. (2011) 'Harvesting covert networks: a case study of the iminer database', *International Journal of Networking and Virtual Organisations*, 8(1/2), p. 52.

Minor, T. (2012) 'Attacking the nodes of terrorist networks.', *Global Security Studies*, 3(2), pp. 1–13.

Molinero, X., Riquelme, F. & Serna, M. (2018) 'Satisfaction and power in unanimous majority influence decision models', *Electronic Notes in Discrete Mathematics, Science Direct*, 68, pp. 197–202.

Morselli, C. (2009) 'Inside criminal networks', *Springer Science +Business Media*. series/656. Edited by F. Bovenkerk. Springer, pp. 1-12.

Morselli, C. (2010) ‘Assessing vulnerable and strategic positions in a criminal network’, *Journal of Contemporary Justice*, 26(4), pp. 382–392.

Namtirtha, A., Dutta, A. & Dutta, B. (2018) ‘Identifying influential spreaders in complex networks based on kshell hybrid method’, *Physica A: Statistical Mechanics and its Applications*, pp 499-521.

Onwuka;, E. N., Bala;, A. S. & Murtala, S. (2016) ‘A Survey of influential nodes detection methods in mobile phone network’. *International Conference on Information and Communication Technology and its Applications,(ICTA) 2016, Federal University of Technology*, pp. 213–219.

Ortiz-arroyo, D. (2010) ‘Discovering sets of key players in social networks’, (11), pp. 1–20.

Ouellet, F., Boivin, R. & Leclerc, C. (2013) ‘Friends with (out) benefits: co-offending and re-arrest’, *Routledge taylor and Francis Group*, 14(4), pp. 141–154.

Ozgul, F. (2016) ‘Analysis of topologies and key players in terrorist networks’, *Socio-Economic Planning Sciences*. Elsevier Ltd. pp 1-23.

Palla, G., Der´eny, I., Farkas, I. & Vicsek, T. (2005) ‘Uncovering the overlapping community structure of complex networks in nature and society’, *physics.soc-ph*, 1(6), pp. 1–10.

Parisi, F., Caldarelli, G. & Squartini, T. (2018) ‘Entropy-based approach to missing-links prediction’, *Applied Network Science*. Applied Network Science, pp. 1–15.

Park, O. (2018) ‘Social network analysis for law enforcement’, *Interanational Association of Crime Analysts iaca*, 02, pp. 1–19.

Paul, A. (2012) ‘Detecting Covert members of terrorist networks’. *Expert Systems with Applications*. pp 1-15.

Pei, S., Morone, F. & Makse, H. A. (2017) ‘Theories for influencer identification in complex networks’, *Expert Systems with Applications*.pp1-12.

Petta, D. L. (2018) ‘Why there is no real difference between a terrorist organization and an organized crime faction, just a matter of interaction towards the state’, *Contemporary Voice*, 1(5), pp. 39–48.

Piraveenan, M. R. (2010) ‘Topological analysis of complex networks using assortativity’. *Expert Systems with Applications*. pp 1-23.

Qiao, T., Shan, W. & Zhou, C. (2017) ‘How to identify the most powerful node in complex’, *Entropy*, 19(4), pp. 1–24.

Reingen, P. H. & Zinkhan, G. M. (1994) ‘Structural holes: the social structure of competition’, *American Marketing Association*, pp. 152–155.

Ren, G., Wang, X., Ren, G. & Wang, X. (2014) ‘Epidemic spreading in time-varying community networks’, *An interdisciplinary journal of Nonlinear Science*, 023116, pp. 24–32.

Ren, J., Wang, C., Liu, Q., Wang, G. & Dong, J. (2016) 'Identify influential spreaders in complex networks based on potential edge weights', *Internal Journal of Innovative Computing, Information, and Control*, 12(2), pp. 581–590.

Reserved, A. R., Pdf, T. & Datasets, M. (2019) 'Learn to conduct descriptive whole social network analysis within an educational setting in ucinet with data from the inclusive education project learn to conduct descriptive whole social network analysis within an educational setting in ucinet with data', *SAGE Publications*.

Rhodes, C. J. and Keefe, E. M. J. (2007) 'Social network topology: a bayesian approach', *Journal of the Operational Research Society*, 58(12), pp. 1605–1611.

Roberts, N. and Everton, S. F. (2011) 'Strategies for combating dark networks', *Journal of Social Structure*, 12, pp 1-22.

Robinson, D. & Scogings, C. (2018) 'The detection of criminal groups in real-world fused data: using the graph-mining algorithm "graphextract"', *Security Informatics*. Springer Berlin Heidelberg, 7(2), pp. 1–16.

Rodrigues, E. M. & Milic-Frayling, N. (2011) 'Flickr linking people, photos, and tags', in *NodeXL Tutorial-13*, pp. 201–223.

Rostami, A. & Mondani, H. (2015) 'The complexity of crime network data: a case study of its consequences for crime control and the study of networks', *PLOS ONE*. Edited by T. Niederkrotenthaler. Public Library of Science, 10(3), pp 1-34.

Rotman, D. & Golbeck, J. (2011) 'YouTube contrasting patterns of content, interaction, and prominence', in *NodeXL Tutorial-14*, pp. 225–246.

Salvatore, C., Pasquale, D. M. & Giacomo, F. (2016) 'Resilience in criminal networks', *AAPP | Atti della Accademia Peloritana dei Pericolanti Classe di Scienze Fisiche, Matematiche e Naturali*, 94(2), pp. 1–19.

Saxena, C., Doja, M. N. & Ahmad, T. (2018) 'Group based centrality for immunization of complex networks', *Physica A*. Elsevier B.V. pp 1-18.

Serin, E., Adali, S. & Balcisoy, S. (2009) 'Entropy-based sensitivity analysis and visualization of social networks'. *Expert Systems with Applications*. pp 1-9.

Sharma, S. & Singh, A. (2016) 'An efficient method for link prediction in weighted multiplex networks', *Computational Social Networks*. pp 1-14.

Sina, S., Rosenfeld, A. & Kraus, S. (2013) 'Solving the missing node problem using the structure and attribute information', *Expert Systems with Applications*. 2(07), pp. 744–751.

Smith, M., Shneiderman, B., Milic-frayling, N., Rodrigues, E. M., Barash, V., Dunne, C., Capone, T., Perer, A. & Gleave, E. (2009) 'Analyzing (social media) networks with nodeXL'. *Expert Systems with Applications*.pp 1-15.

Smith, S. T., Kao, E. K., Senne, K. D. & Bernstein, G. (2014) 'Bayesian network detection using absorbing markov chains', *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 3435– 3439.

Smith, S. T., Kao, E. K., Senne, K. D., Bernstein, G. & Philips, S. (2014) 'Bayesian discovery of threat networks', *IEEE Transactions on Signal Processing*, 62(2), pp. 5324–5338.

Smith, S. T., Philips, S. & Kao, E. K. (2012) 'Harmonic space-time threat propagation for graph detection', *ICASSP*, (1), pp. 3933–3936.

Smith, S. T., Senne, K. D., Philips, S., Kao, E. K. & Bernstein, G. (2013) 'Covert network detection', *Lincoln Laboratory Journal*, 20(1), pp. 47–61.

Sparrow, M. K. (1991) 'The application of network analysis to criminal intelligence: An assessment of the prospects', *Social Networks*, 13(3), pp. 251–274.

Sun, Q., Qiao, Y., Wang, J. & Shen, S. (2016) 'Node importance evaluation method in wireless sensor networks based on an energy field model', *Eurasip Journal on Wireless Communications and Networking*. EURASIP Journal on Wireless Communications and Networking, 2016(1), pp. 1–9.

Tayebi, M. A. (2015) Predictive models for public safety using social network analysis.PhD Thesis, Simon Fraser University Summer.pp1-189.

Thangaraj, M. and Amutha, S. (2018) 'Mgephi: modified gphi for effective social network analysis', *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3(1), pp. 39–50.

UCINET Software - 17 November Greece Bombing (2017).

UCINET Software - 9/11 Hijackers (2017).

United Nations (2014) 'World crime trends and emerging issues and responses in the field of crime prevention and criminal justice', *Commission on Crime Prevention and Criminal Justice, 23rd session*, 00885(2), pp. 1–30.

Varese, F. (2013) 'The structure and the content of criminal connections: The Russian Mafia in Italy', *oxford journals*, 29(5), pp. 899–909.

Wang, S., Du, Y. & Deng, Y. (2016) 'A new measure of identifying influential nodes: Efficiency centrality', *Communications in Nonlinear Science and Numerical Simulation*. Elsevier B.V. pp 1-22.

Wang, S., Du, Y. & Deng, Y. (2017) 'A new measure of identifying influential nodes: efficiency centrality', *Communications in Nonlinear Science and Numerical Simulation*, 47, pp. 151–163.

Wang, S. and Zhao, J. (2015) 'Multi-attribute integrated measurement of node importance in complex networks', *Chaos*, 25(11). pp 1-19.

Welser, H. T., Underwood, P., Cosley, D., Hansen, D. & Black, L. W. (2011) 'Connections of creativity and collaboration', in *NodeXL Tutorial-15: Analyzing Social Media Networks with Nodexl*. Elsevier Inc., pp. 247–271.

Xu, J. & Chen, H. (2008) 'The topology of dark networks', *Communication of the ACM*, 51(10), pp. 58–65.

Yang, A., Tang, Y., Wang, J. & Chen, J. (2014) 'Covert nodes mining in social networks based on games theory', *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2014*, pp. 541–545.

Yao, L., Wang, L., Pan, L. & Yao, K. (2016) 'Link prediction based on common-neighbors for dynamic social network', *Procedia - Procedia Computer Science*. Elsevier Masson SAS, 83(Ant), pp. 82–89.

Ying, L., Tang, M., Do, Y. & Hui, P. M. (2017) 'Accurate ranking of influential spreaders in networks based on dynamically asymmetric link-impact', pp. 1–9.

Zejun, S., Bin, W., Jinfang, S., Yixiang, H., Yihan, W. & Junming, S. (2017) 'Identifying influential nodes in complex networks based on the expansion factor', *International Journal of Modern Physics C*, 27(09), p. 1650105.

Zhang, J., Chen, D., Dong, Q. & Zhao, Z. (2016) 'Identifying a set of influential spreaders in complex networks', pp. 1–13.

Zhao, B., Sen, P. & Getoor, L. (2006) 'Entity and relationship labeling in affiliation networks', *In proceedings of the 23rd International Conferen on MAchine Learning, Pittsburgh,*.pp 1-21.

Zhao, J., Miao, L., Yang, J., Fang, H., Zhang, Q. & Nie, M. (2015) 'Prediction of links and weights in networks by reliable routes', *Nature Publishing Group*. Nature Publishing Group, pp. 1–15.

Zignani, M., Quadri, C., Bernardinello, S., Gaito, S. & Rossi, G. P. (2015) 'Calling and texting: social interactions in a multidimensional telecom graph', *Proceedings - 10th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2014*, pp. 408–415.

APPENDICES

Appendix A (Meta-Analysis of SNA-based Algorithm)

Authors	Title	Method	Strength	Weakness
Kreb (2002)	Mapping Networks of Terrorist Cells	Applying centrality metrics	Open-source data; unitary relationship network	Data defectiveness; Structurally equivalence attribute aid key players evasion
Borgatti (2006)	Identifying sets of key players in a social network	Key player problem;	Visual-graphic positions to key players; unitary network	Tedious in large data set; erratic and dynamic attributes in OCGs
Lampe (2009)	Human capital and social capital in criminal networks: introduction to the special issue	Social network attribute and participants personal attributes	social network attributes accessibility	inaccessibility to human-capital attribute
Morselli (2010)	Assessing vulnerable and strategic positions in a criminal network	Vulnerability and strategic positions. Combined degree and betweenness	accessibility to social network attribute; partly resolved structural equivalent issue	inability to detect low-nodal degree actor; legitimate actors
Karthika and Bose (2011)	A comparative study of social networking approaches in identifying the covert nodes	Multiple centrality metrics: degree, betweenness, and closeness	nodes with high metric score are vulnerable; detected key actors;	poor detection; unverified personality; challenges of hidden relationship, dynamic behaviour and defective data
Berzinj *et al.* (2012)	Detecting Key Players in Terrorist Networks	combination of different centrality measures to detect key players in decentralized network	integration of centrality metrics; focusing on financial manager	unverified personality; high-profile key actors were ignored
Calderoni (2012)	The structure of drug trafficking mafias: the 'Ndrangheta and cocaine	Use SNA to investigate the relevancy of tasks and hierarchy	Identification of strategic position with criminal leaders	The resilience of mafia to law enforcement action
Husslage *et al.* (2012)	Leaderless Covert Networks: A Quantitative Approach	Social network attribute and correlation	flat organizational structure; unitary network structure; social network attributes	Varying network structures of criminals; varying social network attributes
Campana and Varese (2012)	Listening to the wire: criteria and techniques for the quantitative analysis of phone intercepts	Phone conversation wiretapped, SNA analysis	Access to phone content; unitary network; dislodge FOS	High-profile actor with low phone conversation were less-vulnerable; structurally-equivalent problem

Catanese *et al.*(2013)	Forensic analysis of phone call networks	Phone calls log; link analysis and centrality metrics incorporated.	visualization of participant in criminal activities; hierarchies of community, and key players	Inability to handle hidden relationships; detective data
Varese (2013)	The structure and the content of criminal connections: The Russian Mafia in Italy	Phone call interception; investigating the organizational structure and adaptive to security pressure	Access to the content of phone conversation; obtained hierarchical structure in phone conversation of OCGs.	Complex data processing, Potent failure without intercepting the content
Ferrara *et al.* (2014)	Detecting Criminal Organizations in mobile phone networks	Statistical network analysis, community detection, visual exploration of mobile phone networks	Discovery members who play vital roles, hierarchy and community in criminal organisations	Legitimate actors are evasive
Basu (2014)	Social Network Analysis: A Methodology	Social network analysis adopted for terrorist network structure	Open-source data; unitary relationship network	unconcerned with hidden relationships
Butt *et al.* (2014)	Hidden Members and Key Players Detection in Covert Networks Using Multiple Heterogeneous Layers	Degree centrality and multiple relationships and transaction networks	Integration of database for monitoring criminal activities	Inaccessibility to terrorist data The high nodal degree could be fake
Bright *et al.* (2015)	The use of actor-level attributes and centrality measures to identify key actors: a case study of an Australian drug trafficking network	Combined metrics: degree and weighing attributes	Effective in connecting personalities with social network status	Different rating of resources and biased Maybe biased or overrating resources
Bright *et al.* (2015)	Networks within networks: using multiple link types to examine network structure and identify key actors in a drug trafficking operation	using scatterplot of degree and betweenness centrality; study eight links in DTO;	Detecting key actors in each link; more key actors identified including legitimate actors	Accessibility to multiple relations datasets for OCGS; operation of terrorist of hidden
Ozgul (2016)	Analysis of topologies and key players in terrorist networks	comparative analysis of topologies of terrorist groups with different centrality metrics	key-players in terrorist groups assumed to be central actors irrespective of historical lineage.	no consideration for erratic behaviours; dynamic in structural positions; highly prone to false alarm;
Gunnell (2016)	Social Network Analysis of an Urban Street Gang Using	Combined metrics: degree and weighing attributes	Detecting vulnerable members of street gangs	Arbitrary data; Unorganized data

	Police Intelligence Data			
Grassi *et al.* (2019)	Betweenness to assess leaders in criminal networks: new evidence using the dual projection approach	Eight different types of betweenness centrality	A high correlation of metric; Criminal leaders' adherent to brokerage	Bipartite data; Accessibility to criminal meeting attendance
Ismail *et al.* (2019)	Detecting Covert Members: Quadrant approach for Classification and Identification of Smart Criminals	Social Network Analysis and Quadrant Classification of Criminal Attributes	Social network attributes, detecting evasive members in criminal structure	Inaccessibility to terrorist data, profile of participants

Appendix B (Network Attributes of Actors in Al-Qaeda 9/11 Attack)

Actor Name	Actor ID	Degree Centrality ( )	Closeness. Centrality ( )	Betweenness Centrality ( )	Eigenvector Centrality ( )
Majed Moqed	1	0.0167	0.2449	0.0000	0.0052
Khalid Al-Mihdhar	2	0.0667	0.2632	0.0232	0.0078
Hani Hanjour	3	0.1167	0.3226	0.2095	0.0365
Nawaf Alhazmi	4	0.1500	0.2804	0.1414	0.0120
Salem Alhazmi	5	0.0167	0.2198	0.0000	0.0017
Ahmed Alnami	6	0.0500	0.2353	0.0000	0.0053
Ahmed Alghamdi	7	0.0333	0.2281	0.0014	0.0024
Saeed Alghamdi	8	0.1000	0.2817	0.0695	0.0120
Hamza Alghamdi	9	0.1000	0.2804	0.0594	0.0130
Ahmed Al Haznawi	10	0.0500	0.3141	0.1033	0.0472
Mohand Alshehri	11	0.0333	0.2655	0.0103	0.0127
Fayez Ahmed	12	0.0500	0.3046	0.0268	0.0761
Ziad Jarrah	13	0.1333	0.3750	0.1306	0.3067
Marwan Al-Shehhi	14	0.2000	0.3797	0.0760	0.3883
Mohamed Atta	15	0.2500	0.4444	0.5080	0.4388
Abdul Aziz Al-Omari	16	0.0500	0.3333	0.0430	0.1231
Waleed Alshehri	17	0.0667	0.2791	0.0707	0.0386
Wail Alshehri	18	0.0333	0.2308	0.0000	0.0065
Satam Suqami	19	0.0667	0.2469	0.0377	0.0074
Raed Hijazi	20	0.0500	0.2308	0.0098	0.0033
Nabil al-Marabh	21	0.0667	0.2317	0.0139	0.0036
Mustafa Ahamend al-Hisawi	22	0.0667	0.3409	0.0588	0.1340
Mamoun Darkazanli	23	0.0500	0.3158	0.0000	0.1602
Zakariya Essabar	24	0.0833	0.3315	0.0000	0.2526
Said Bahaji	25	0.1167	0.3352	0.0027	0.2985
Mounir El Motassadeq	26	0.0667	0.3175	0.0000	0.2089
Zacarias Moussaoui	27	0.1333	0.3529	0.2201	0.1579
Ramzi Bin al-Shibh	28	0.1500	0.3659	0.0543	0.3427
Agus Budiman	29	0.0833	0.3333	0.0240	0.2158
Ahed Khalil Ibrahim Samir Al-Ani	30	0.0333	0.3125	0.0090	0.0682
Lofti Raissi	31	0.0833	0.375	0.2380	0.1719
Rayed Mohammed Abdullah	32	0.1000	0.2985	0.0236	0.0381
Bandar Alhazmi	33	0.0333	0.2459	0.0000	0.0106
Faisal Al Salmi	34	0.0333	0.2459	0.0000	0.0106
Osama Awadallah	35	0.0500	0.2214	0.0000	0.0033
Abdussattar Shaikh	36	0.0500	0.2214	0.0000	0.0033
Mohamed Abdi	37	0.0167	0.2198	0.0000	0.0017
Mohamed Belfas	38	0.0333	0.2532	0.0003	0.0404

Imad Eddin Baraat Yarkas	39	0.0833	0.3390	0.0409	0.1411
Tarek Maaroufi	40	0.1000	0.2927	0.0206	0.0638
Abu Qatada	41	0.1167	0.2941	0.0552	0.0827
Djamal Benghal	42	0.1500	0.2844	0.1069	0.0644
Jerome Courtaillier	43	0.0667	0.2703	0.0000	0.0470
David Courtaillier	44	0.0667	0.2703	0.0000	0.0470
Ahmen Ressam	45	0.0333	0.2778	0.0082	0.0262
Abu Walid	46	0.0500	0.2344	0.0000	0.0296
Jean-Marc Grandvisir	47	0.0167	0.2222	0.0000	0.0092
Abu Zubeida	48	0.0167	0.2222	0.0000	0.0092
Nizar Trabelsi	49	0.0167	0.2222	0.0000	0.0092
Haydar Abu Doha	50	0.0667	0.2778	0.0088	0.0265
Mehdi Khammoun	51	0.0500	0.2667	0.0006	0.0259
Mohammed Bensakhria	52	0.0833	0.2778	0.0052	0.0413
Lased Ben Heni	53	0.0333	0.2643	0.0000	0.0221
Essid Sami Ben Khemail	54	0.2000	0.3550	0.2568	0.1143
Seifallah ben Hassine	55	0.0500	0.2752	0.0000	0.0344
Essoussi Laaroussi	56	0.0500	0.2752	0.0000	0.0344
Fahid al Shakri	57	0.0167	0.2632	0.0000	0.0163
Madjid Sahoune	58	0.0333	0.2655	0.0000	0.0200
Samir Kishk	59	0.0167	0.2632	0.0000	0.0163
Kamel Daoudi	60	0.1000	0.2804	0.0086	0.0610

Appendix C (Typical CDR Contents)

Target	Date	Time	Duration	DIR	Dialled Number	Dest. Number	Other Number	Status	Special Feature	Caller ID	Switch	Sector	Tower	Switch	Sector	Tower
5107308760	10/29/2011	19:45:51	0:16	Outgoing call	7319368	5107319368	5107319368	Answered	None		San Francisco2	3	183	3	183
5107308760	10/29/2011	23:32:49	0:37	Outgoing call	9275717	5109275717	5109275717	Answered	None		San Francisco2	1	94	1	94
5107308760	10/29/2011	23:33:32	0:36	Outgoing call	9275717	5109275717	5109275717	Answered	None		San Francisco2	1	94	1	94
5107308760	10/30/2011	0:43:18	0:02	Incoming call		5107308760	5109275717	Not Answered	None	5109275717	San Francisco2	3	94	3	94
5107308760	10/30/2011	11:55:46	0:18	Incoming call		5107308760	4085616930	Not Answered	None	4085616930	San Francisco2	1	94	1	94
5107308760	10/30/2011	11:56:43	0:28	Outgoing call	4085616930	4085616930	4085616930	Not Answered	None		San Francisco2	1	94	1	94
5107308760	10/30/2011	13:18:10	0:24	Outgoing call	4087269200	4087269200	4087269200	Not Answered	None		San Francisco2	1	94	1	94
5107308760	10/30/2011	13:20:54	0:08	Incoming call	3.23E+12	5107308760	5104858523	Not Answered	Call FWD- No REPLY	4104858523	San Francisco2	1	94	1	94
5107308760	10/30/2011	13:32:25	0:34	Outgoing call	5104858523	5104858523	5104858523	Answered	None		San Francisco2	1	94	1	94
5107308760	10/30/2011	16:54:27	0:08	Incoming call	3.23E+12	5107308760	5104858523	Not Answered	Call FWD- No REPLY	5104858523	San Francisco2	3	182	3	182
5107308760	10/30/2011	17:03:47	0:07	Incoming call	3.23E+12	5107308760	5104858523	Not Answered	Call FWD- No REPLY	5104858523	San Francisco2	3	182	3	182
5107308760	10/30/2011	17:17:26	0:10	Outgoing call	5104858523	5104858523	5104858523	Answered	None		San Francisco2	3	182	3	182
5107308760	10/30/2011	17:27:48	0:03	Incoming call	3.23E+12	5107308760	5107319368	Not Answered	Call FWD- No REPLY	5107319368	San Francisco2	3	182	3	182

5107308760	10/30/2011 21:17:27	0:06	Incoming call	3.23E+12	5107308760	5105863731	Not Answered	Call FWD- No REPLY	5105863731	San Francisco2	3	204	3	204
5107308760	10/30/2011 21:29:12	2:35	Outgoing call	5105853731	5105863731	5105863731	Answered	None		San Francisco2	3	204	3	204
5107308760	10/30/2011 22:48:41	1:55	Incoming call		5107308760	5105853731	Answered	None	5105863731	San Francisco2	3	204	3	204
5107308760	10/30/2011 23:27:54	0:04	Incoming call	3.23E+12	5107308760	5107319368	Not Answered	Call FWD- No REPLY	5107319368	San Francisco2	3	217	4	495
5107308760	10/31/2011 0:00:54	0:22	Outgoing call	5107319368	5107319368	5107319368	Answered	None		San Francisco2	3	53	3	53
5107308760	10/31/2011 0:01:23	0:54	Outgoing call	5107319368	5107319368	5107319368	Answered	None		San Francisco2	3	53	3	53
5107308760	10/31/2011 0:02:28	0:25	Incoming call		5107308760	5107319368	Answered	None	5107319368	San Francisco2	2	179	2	179
5107308760	10/31/2011 0:03:03	2:31	Incoming call		5107305760	5107319368	Answered	None	5107319368	San Francisco2	3	53	3	53
5107308760	10/31/2011 0:11:03	0:04	Incoming call	3.22E+12	5107308760	5107319368	Not Answered	Call FWD- No REPLY	5107319368	San Francisco2	2	179	2	179

(Source: Ferrara *et al*., 2014)

Appendix D (Source Codes for BNM Algorithm)

Source Codes for BNM Algorithm

```
%% Beta-Binomial Distribution on Centrality Measures
%% Datasets: centrality metrics of 9/11 Terrorist group
%% By ISMAIL A. Adekunle
%% Date: Rabiul-Awwal 25th,1439/December 14th
2017 clear; close all; clc;
ismail =
xlsread('Sept11data.xls'); clc;
Cd = ismail (:2); %degree centrality measures
Cc = ismail (:3); %closeness centrality measures Cb
= ismail (:4); %betweenness centrality measures Ce =
ismail (:5); %eigenvector centrality measures
Ld = length (Cd); % number of actors' values in the degree
centrality Lc = length (Cc); % number of actors' values in the
closeness centrality
Lb = length (Cb); % number of actors' values in the
betweenness centrality
Le = length (Ce); % number of actors' values in the
eigenvector centrality
% Prior parameters
a = 0.3;
b = 0.7;
N = 700;
% Data
initial_range =
0; end_range = 1;
theta =
linspace(initial_range,end_range,N); X=1;
% This section is for computation of:
P_o = betapdf(theta,a,b);                          % prior (P_o)
L_o = nchoosek(N, X)*(theta.^X).*(1-theta).^(N-X); % Likelihood (L_o)
Post_o = betapdf(theta,a+X,b+N-X);                 % posterior (Post_o)
[HPost_o, iPost_o] = max (Post_o);
figure (4)                          %Jumad-Al-Awwal18th,1441
map_post_Ce= zeros (Le,1);
for c = 1: Le
    endrange_up = Ce(c); %update last part value of theta
    range_update = linspace(initial_range,endrange_up,N);
    posterior_Ce = betapdf(range_update,a+X,b+N-X);
%     plot (theta, posterior_Ce);
    [H,i] = max(posterior_Ce);
    map_post_Ce(c) = H;
end
plot(map_post_Ce) %Jumad-Al-Awwal18th,1441
xlabel('actor        id.')        %Jumad-Al-Awwal18th,1441
ylabel('map_e_g_v.') %Jumad-Al-Awwal18th,1441

figure (3)                          %Jumad-Al-Awwal18th,1441
map_post_Cb= zeros (Lb,1);
for c = 1: Lb
    endrange_up = Cb(c); %update last part value of theta
    range_update = linspace(initial_range,endrange_up,N);
    posterior_Cb = betapdf(range_update,a+X,b+N-X);
%     plot (theta, posterior_Cb);
    [H,i] = max(posterior_Cb);
    map_post_Cb(c) = H;
end
plot(map_post_Cb) %Jumad-Al-Awwal18th,1441
xlabel('actor id.')              %Jumad-Al-Awwal18th,1441
```

```
ylabel('map_b_t_w.')              %Jumad-Al-Awwal18th,1441
figure (2)                         %Jumad-Al-Awwal18th,1441
map_post_Cc= zeros (Lc,1); %Jumad-Al-
Awwal18th,1441 for c = 1: Lc
    endrange_up = Cc(c); %update last part value of theta
    range_update  =  linspace(initial_range,endrange_up,N);
    posterior_Cc = betapdf(range_update,a+X,b+N-X);
%     plot (theta, posterior_Cc);
    [H,i] = max(posterior_Cc);
    map_post_Cc(c) = H;
end
plot(map_post_Cc)                 %Jumad-Al-Awwal18th,1441
xlabel ('actor id.')               %Jumad-Al-Awwal18th,1441
ylabel('map_c_l_s.')               %Jumad-Al-Awwal18th,1441
figure (1)                         %Jumad-Al-Awwal18th,1441
map_post_Cd= zeros (Ld,1);
for c = 1: Ld
    endrange_up = Cd(c); %update last part value of theta
    range_update  =  linspace(initial_range,endrange_up,N);
    posterior_Cd = betapdf(range_update,a+X,b+N-X);
%     plot (theta, posterior_Cd);
    [H,i] = max(posterior_Cd);
    map_post_Cd(c) = H;
end
plot(map_post_Cd)                 %Jumad-Al-Awwal18th,1441
xlabel('actor id.')               %Jumad-Al-Awwal18th,1441
ylabel('map_d_g_r.')              %Jumad-Al-Awwal18th,1441
```

Appendix E (Source Codes for SNA-Quadrant)

```
%% Matlab code for generating graphs of centrality measure
%% and Maximun-a-posetriori-map for N17 Attackers
%% By Engr. ISMAIL A. Adekunle
% Monday Jumudal-Al-Awwal,15th 1440/21th
January,2019 %% Acknowledgement: Alhiamdulillah
clear; close all; clc;
d911data = xlsread('C:\Users\ISMAIL\Desktop\Conspirators.xls');
P_id = d911data (:1);
dgr_f = d911data (:2); % all row values in column 2
dgr_a = d911data (1:19,2); % the first 19 rows in column 2
dgr_c = d911data (20:61,2); % the last row values in the column
2 cls_f = d911data (:3); % all row values in column 3
cls_a = d911data (1:19,3); % the first 19 rows in column 3
cls_c = d911data (20:61,3); % the last row values in the column
3 btw_f = d911data (:4);
btw_a = d911data (1:19,4); % the first 19 rows values in column
4 btw_c = d911data (20:61,4); % the last rows values in the
column4 egv = d911data (:5); % all row values in column 5
egv_a = d911data (1:19,5); % the first 19 rows in column 3
egv_c = d911data (20:61,5); % the last row values in the column
3 figure (1)
scatter (btw_a,dgr_a,'*')
hold on
scatter (btw_c,dgr_c,'o')
hold off
xlabel('Brokerage strategy')% Betweenness strategy lowers
visibility ylabel('Influence(local)') % Degree strategy for actors'
visibility figure (2)
scatter (cls_a,dgr_a,'*')
hold on
scatter (cls_c,dgr_c,'o')
hold off
xlabel('Proximity strategy')% Closeness strategy lowers visibility
ylabel('Influence(local)') % Degree strategy for actors'
visibility figure (3)
scatter (btw_a,egv_a,'*')
hold on
scatter (btw_c,egv_c,'o')
hold off
xlabel('Brokerage strategy')% Betweenness strategy lowers
visibility ylabel('Influence(global)') % Eigenvector strategy for
actors' visibility
figure (4)
scatter (cls_a,egv_a,'*')
hold on
scatter (cls_c,egv_c,'o')
hold off
xlabel ('Proximity strategy') % Closeness as a strategy that
lowers susceptibility
ylabel('Influence(global)') % Eigenvector as a strategy for
actors' visibility
```

Appendix F (Kite Network)

Kite network is a relatively small network - a group in karate. Figure 4.35 presents the Kite network. It is made up of ten members. Degree centrality of actors in the group was extracted and fed as input into both BNM and entropy algorithms.

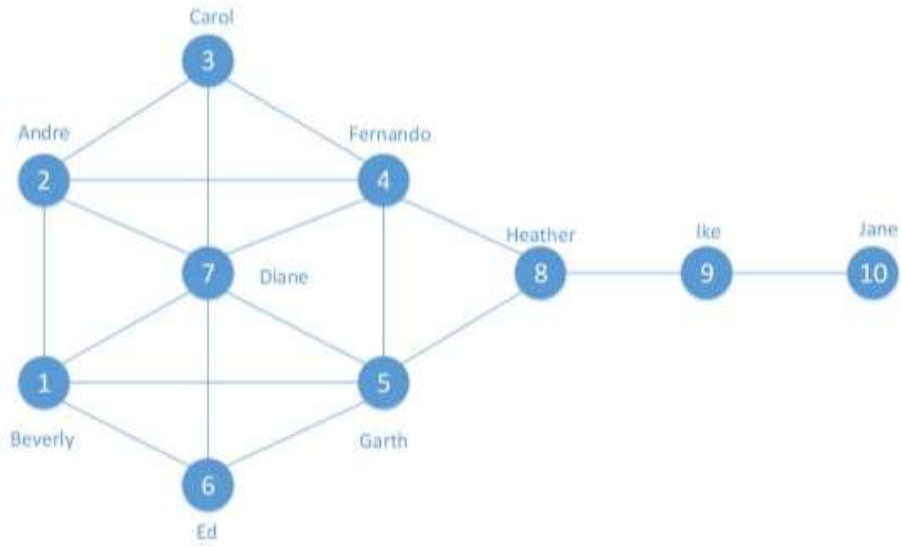

Figure 4.35: The Kite network
(Source: Hensen, 2011)