INVESTIGATING SELECTED EGOCENTRIC USERS ON SOCIAL MEDIA PLATFORMS USING SOCIAL NETWORK ANALYSIS IN MINING FORENSIC EVIDENCE FOR LAW ENFORCEMENT IN KENYA

LAMEK KIPRUTTO RONOH

A Thesis Submitted to the Institute of Postgraduate Studies of Kabarak University in Partial Fulfilment of the Requirements for the Conferment of the Doctor of Philosophy Degree in Information Technology Security and Audit

KABARAK UNIVERSITY

OCTOBER 2017

DECLARATION

The research thesis is my own work and to the best of my knowledge it has not been presented for the award of a degree in any university or college.

Signed: ______________________________________Date:_____________________________

Lamek Kiprutto Ronoh

GDI/M/1198/09/15

RECOMMENDATION

To the Institute of Postgraduate Studies:

The research thesis entitled "**Investigating Selected Egocentric Users on Social Media Platforms using Social Network Analysis in Mining Forensic Evidence for Law Enforcement in Kenya"** and written by **Lamek Kiprutto Ronoh** is presented to the Institute of Postgraduate Studies of Kabarak University. We have reviewed the research thesis and recommend it be accepted in partial fulfilment of the requirement for award of the degree of Doctor of Philosophy in IT Security and Audit.

Signed: ______________________________________ Date:____________________________

Prof. Kefa Rabah

Professor

Faculty of Computer Science and Bioinformatics

Kabarak University

Signed: ______________________________________ Date:____________________________

Dr. Nickson Menza Karie

Senior Lecturer

Faculty of Computer Science and Bioinformatics

Kabarak University

COPYRIGHT

ACKNOWLEDGEMENT

I wish to acknowledge a number of people who made it possible for this thesis to be completed. I wish to thank Kabarak University for granting me a chance to study in the institution. I sincerely wish to thank my able supervisors Prof. Kefa Rabah and Dr. Nickson Karie for the time and energy devoted to this work, their invaluable analysis of the work, patience and encouragement brought hope when the situation almost looked hopeless. All those whose names are not mentioned and made contributions to this study, I wish to thank you. May Almighty Father bless you abundantly!

DEDICATION

I dedicate this thesis to my wife Elseba and our sons Reuel, Levy and Elimelech for the entire support they accorded me while I was working on this scholarly task. To all my friends and colleagues for their invaluable ideas without which this thesis would not have been completed successfully, to my critics for making life worth living.

ABSTRACT

Investigation of social media using social network theory is a new powerful tool that will aid and ease law enforcement agencies in multi-faceted ways in this ever evolving digital landscape. It is against this backdrop that this study focused on identifying and investigating selected individuals on Facebook and Twitter social media platforms. In particular, selected respondents from University of Eldoret, Kibabii, Moi, Kisii and Rongo Universities were involved in the study. The objective of the study was to demonstrate how Social Network Analysis (SNA) can be employed as an investigate tool to mine, analyse data from selected online social media users and present digital forensic evidence to aid law enforcement in Kenya. Particularly, the study aimed at identifying high degree nodes in the network and their behavioural patterns and profiles using visualizations, network metrics and user profile/demographic information. Social network analysis experimental research design was employed in this study. The sample size of the respondents was arrived at by employing Yamane's formula of calculating sample size. The respondents were guided to create pseudo-online parody accounts in various social media platforms which was used to carry out the online data mining from the selected respondents to aid in social network analysis. The significance of the study was to fill the knowledge gap that hitherto not been researched by previous scholars yet it is imperative area as far as cyber-security and law enforcement is concerned in Kenya. Data mining and analysis was done using NodeXL, an Add-in tool in Ms-Excel for social network analysis. Computation of centrality measures, network clusters, cliques were presented using both infographic visualizations and centrality metrics of the respondents on egocentric networks Focal communication paths through which information flows in the network were also depicted. The findings demonstrated that Social Network Analysis can be effectively used on social media platforms to mine, analyse and present digital forensic evidence of individuals under investigation. The outcome of this study gives a new insight and techniques that can help law enforcement agencies and related stakeholders to identify or detect important individuals, subgroups, interaction patterns between subgroups and roles they play in a given network. The findings presented in this research illustrates how social network analysis can be used to determine the interpersonal connections, importance of actors in a given social network and detect communities of people and principally how law enforcement agencies can utilize this technique in identifying and tracking suspicious characters and ultimately help in maintaining law and order. SNA ought to be embraced as a supplement of conventional investigation, not necessarily replacing it.

Key words: Social Media, Social Network, Social Network Analysis, Digital forensic evidence, law enforcement

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

API:	Application Programming Interface
BAKE:	Bloggers Association of Kenya
CLEDS:	Commissioner for Law Enforcement Data Security
CSV:	Computer Separated values
GPS:	Geographical Positioning System
KCA:	Kenya Communications Authority
IP:	Internet Protocol
NIS:	National Intelligence Agency
NODEXL:	Network Overview, Discovery and Exploration for Excel
OCG:	Organized Crime Group
SNA:	Social Network Analysis
SNS:	Social Network Structure
TOS:	Terms of Reference

OPERATIONAL DEFINITION OF TERMS

Node (Vertex/Actors): Refers to people in one’s social network. Nodes are represented by the circles/dots in the image.

Edges (Links): Refers to the relationships between people, shown as lines connecting the nodes.

An Undirected Edge*:* It is a tie that represents co-occurrence, co-presence, or a bonded-tie between the pair of actors . Generally, it means that the people connected by the edge know one another.

A Directed Edge: Refers to a tie that originates in a source actor and reaches a target actor is usually drawn with an arrow head pointing in the direction of the relationship.

Graph (Social Graph): Refers to a visual representation of a group of nodes and edges make up a social network.

Clique: A cohesive subgroup of network nodes where every node is connected directly to one another

A path: Refers to a series of edges connecting two nodes. Usually, path’s length are measured in the number of edges one can traverse from one node to the next.

Relational ties: It refers to link actors within a network where these ties can be informal or formal

Socio-centric (*Complete networks*): It consist of the relational ties among members of a single, bounded community.

Ego-centric (*Personal networks*): This refers to the ties directly connecting the focal actor (ego) to others (ego’s alters) in the network, plus ego’s views on the ties among his or her alters.

Geodesic distance: Refers to the least number of ties that must be passed to arrive between two actors

Degree/Node degree: Refers to the number of connections a node.

Egocentric Networks: This is a social network focused around one individual.

Density: The degree of interconnection between actors of network. Low density network implies few connections whereas high density network means highly interconnected actors

Cluster*:* Clusters are groups of nodes that have many connections between them and are more tightly grouped than others.

Homophily: Refers to a method where nodes who have related attributes are probable of establishing a relationship

Centrality*:* It describes the collection of measures that indicate how important a node is in a social network.

Degree centrality: shows people with many social connections. A node with 10 social connections would have a degree centrality of 10. A node with 1 edge would have a degree centrality of 1.

Closeness centrality: indicates who is at the heart of a social network by looking for the node that is closest to all other nodes. Closeness centrality for a node is the average length of all the shortest paths from that one node to every other node in the network.

Betweenness Centrality: Describes people who connect social circles as well as measure that captures a person's role in allowing information to pass from one part of the network to the other.

Eigenvector Centrality: It measures the influence that a node has in a network. It is high among influential people in the network.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Social network research have gained significant acknowledgment in terms of both theory and method in contemporary (Freeman, 2004). Passmore (2011) defines a *social network* as a social formation constituting of individuals hereby called “nodes", which are linked together for some reason such as companionship, same interest, monetary exchange, dislike, romantic relationships, or associates of a particular faith, understanding or status. This definition was echoed by Mincera and Niewiadomska-Szynkiewicza (2012) where they concurred that a social network is formally defined as a set of actors or social groups, and relationships such as friendship, collaboration, business or political affiliations. The first approach to capture the global properties of such systems is to model them as graphs which nodes represent the actors and links the relationships between them.

In quest of defining further social network and Social Network Analysis, Passmore (2011) noted that Social Network Analysis focuses on the structure of relationships within a set of social nodes in a given social media network. That is, social network analysis regards social interactions from the perspective of network theory comprising of nodes and links (also called ties, edges, or connections). Nodes in this case are the persons within the social networks, whereas the ties are the associations between these persons. The resultant graphical structures are mostly quite intricate. The definition was reaffirmed by Granovetter (1973) that there can be many kinds of ties between the nodes and that social networks functions at different categories spanning from one’s kin up to the category of a state. He concluded that in any social network, ties between nodes serve a vital position in influencing how difficulties are worked out, institutions are managed, and the extent to which folks thrive in realizing their aspirations.

Golbeck (2015) succinctly defines Social network analysis as the analysis of social connections a person has with others. It involves studying the structure of people's connections—especially things like who is most important or influential in the network and which groups of people are closely connected. In a nutshell, Gunnell, Hillier and Blakeborough (2016) summarized that the social network analysis approach allows different types of links between individuals to be explored. Social Network Analysis is to understand a community by mapping the relationships that connect them as a network, and then trying to draw out key individuals, groups within the network (‘components’), and or associations between the individuals. The increasing availability of large-scale, real-world sociographic data derived from social media, web pages and datasets has led, among other things, to a renaissance of Social Network Analysis and its application in new fields of enquiry. Social network analysis allows one to measure, map and explain everything pertaining social network and its elements (Gupta & Brooks, 2015).

Social media constitute all the hardware and software that facilitate the end-users worldwide to essentially create and share information with each other. In the context of this, *software* are the implicit spaces that allow users to interact, make and circulate information whereas the *hardware* refers to all the computing technologies that enable users to utilize the aforementioned hardware platform (Gupta & Brooks, 2013). Hansen *et al* (2011) argued that social media emerged as a pervasive platform for peoples’ communication, the latent ties that connect one another and has become more visible and machine readable. The net effect is a new prospect to depict social networks in detail and scale never before seen.

1.2 Background to the study

Drawing from Golbeck (2015), social media platforms such as twitter, facebook, instagram, pinterest, email, discussion forums, blogs and foursquare are used by myriad of users globally. As they interact online by means of the aforementioned social media platforms using various applications on terrestrial and mobile devices the result is the creation of multiple intricate social network structures. The active communications and networks of relationships resulting from these technologies is crucial to persons, institutions, and society. Comprehending how these social media networks spread, transform, fall short, or thrive is an interesting concern to researchers and professionals. The field of social network analysis provides a set of concepts and metrics to systematically study these dynamic processes. The different techniques of depicting information have also turned out to be important in assisting users to discern patterns, trends, clusters, and outliers, even in an intricate social networks.

According to a report by PewResearch Center (2014), social media include the various ways and means people hook with one another through online interactions. Mobile devices, social networks, email, texting, micro-blogging and location sharing are just a few of the many ways people engage in computer-mediated collective action. As people connect, like, follow, friend, reply, retweet, comment, tag, rate, review, edit, update, and text one another they form collections of connections. These set of connections develops into network formations that can be mined, investigated and the results depicted using various ways and techniques. The result can give a new understanding of into the structure, size, and key positions in these networks. Social Network Analysis employs the concepts from mathematics of graph theory to examine to visualize complicated array of interrelations between actors to create visual maps and present centrality metric scores (Freeman, 2004). In fact, Social network analysis focuses on relationships between the members of a given network rather than concentrating on a single node characterises of those associates (Giuffre, 2013).

1.2.1 Social Networking and Media Information Security Concerns

According to Wüest (2010) social networks are an inherent part of today's Internet and used by more than a billion people worldwide. They allow people to share ideas and interact with other people, from old friends to strangers. This interaction reveals a lot of information, often including personal information visible to anyone who wants to view it. Hence privacy is often a key concern by the users since millions of people are willing to interact with others. Though a social media user may not have populated the Internet with any of his/her personal information, it is highly likely that third-party associates have. That is to say that family and friends who are active in social media may have unwittingly created or contributed to the cyber footprint and consequently placed one's privacy in jeopardy.

In a rejoinder, Zambri (2015) concurred that social media and social networks have changed how criminal and criminal organizations conduct business against law enforcement. Moreover, in connecting social media and criminal investigations, Kerschbaum and Schaad (2008) pin pointed that not only social media used to organize and carry out felony but also information that is exchanged on different platforms is becoming increasingly important in undertaking its law enforcement activities. Thus social media sites and resources can be effective means for law enforcement personnel in the prevention, identification, investigation, and prosecution of crimes.

Wright (2010) underscored that social media platforms such as Facebook, Twitter, or other similar networks can be a rich source for forensics investigations. The ever growing ocean of data in those networks is appealing to investigators. Law enforcement agencies have already discovered that criminals socialize online with criminals ring members, and other similar felons network with their cohorts. A simple investigation might view just the publicly-available text and images posted on a suspect's social page. Intense investigations may require the investigator to be issued a permit of authority to do so. Nouh and Nurse (2015) corroborated these findings by observing that not only does social media platforms provide a new unexploited fountain of mining intelligence for law enforcement community, but it also gives an insight of understanding behavioural patterns of covert groups.

Social networks gives a good way of understanding the abstract concept of the ever-changing and various connections or interactions between individuals in a network (Mincer & Niewiadomska-Szynkiewicz, 2012). A highly complex investigation will look at more than just the data appearing on the face of social media platform. The research continued to give an instance of a successful use of social media evidence in investigations and trials revolving around a child pornography case in which a suspect lured a minor. The charges came from a covert police investigation carried out online with an officer posing as a 13-year-old boy. The law enforcer achieved this by creating a pseudo-account profile purporting to be a 13-year-old boy and then sending a friend request to the suspect. Based on this preliminary investigation, a written order was served on the various internet service providers, which resulted in identification of the litigant, the IP addresses associated his accounts and residential address. This was subsequently used to apply for a search warrant of the suspect’s house.

1.2.2 Information Sharing and Privacy on the Social Media

In his comprehensive analysis of security and privacy issues on various social media platforms, Hudaib (2014) underscored that no one is compelled to sign up on an online social network although a lot of networks persuade, but do not oblige users to disclose their vital primary information such their dates of birth, contacts, or place of residence. In spite of that, it is quite amazing to realize the magnitude and particulars of the individual’s information a number of social media users give, and wonder about how clued-up this information divulging is. Radical philosophies can propagate throughout the various social media platforms and probable criminals can interrelate with persons with whom they have the same school of thought whether they live in the same locality or across the globe (Kunkle, 2012).

The ever shifting ethos or societal patterns, getting accustomed and having trust in variety of computing platforms, being inexperienced or mere abuse of private data by unscrupulous felons could contribute in this unparalleled trends of information disclosure. Nonetheless, the security of various social media platforms and access controls poses vulnerability by design - to entice their significance as network products and help increase their expansion by enhancing registration, gaining access, and sharing of information simple. In harmony, Witnov (2011) agreed that some information on social networks is publicly available, although most of it is restricted. The easiest means of gaining entry to a one’s online confidential data is to "*befriend*” him/her on that social media platform, which typically permits access to more restricted information.

According to Global Justice Initiative report (2013), various social network platforms are progressively exploited to initiate or carry out unlawful acts and therefore law enforcement agencies ought to comprehend the idea and purpose of these platforms. They also need to understand the way social media forensic apparatus and resources can be used to thwart, lessen, act in response to, and probe illegal actions. Although the advent of social media has created new investigatory opportunities for law enforcement, it also presents ethical, legal, and technical challenges. Depending on the country in which the investigation takes place, it may be illegal to gather information on social media if a user's profile is not public (Rice & Parkin, 2016).

According to Everett and Borgatti (1999) online social networks might be employed to organize a unlawful-connected burst gang or arrange a burglary, or radical groupings that might utilize social media platforms to enrol fresh recruits to ascribe heinous objectives. Information stored on various social network platforms could immensely help law enforcement agencies in obtaining appropriate data in continuance of barring criminal activities, maintenance of civic order, and the investigation of illegal acts, as well as suspicious terror movements. These rules are disseminated partially to inculcate the appropriate poise linking the undercover potential of social networks and confidentiality requirements. This report concurred with CLEDS (2013), which indicated that information from social network sites is largely used to corroborate other intelligence where law enforcement agencies undertake the active use of social media for covert operations.

1.2.3 Social Network analysis in law enforcement

A study on social network analysis for anti-terrorism, Choudhary and Singh (2015), established that Social Network Analysis has widely been applied by the investigators and law enforcement agencies in quest of comprehending the setup of terrorist networks and coming up with well- schemed plans to interrupt them by identifying influential leadership and latent patterns in the illegal and terror networks. In a study conducted at North Eastern University (Boston, USA), use of Social Network Analysis in latent pattern recognition and employing appropriate data mining software, researchers found it that it is likely to establish an precise profile of a target being investigated not using what that individual has shared about his or herself on social media but by investigating what his or her friends have posted publicly (Russell, 2013). As stated by Johnson and Reitzel (2011) Social Network Analysis is a technique that can give investigators a set of powerful visualizations and centrality metric scores upon which they can swiftly unearth, examine and depict crime deeds committed online and ultimately help in coming with plans of intercepting such vices. Centrality metrics of a network are complimented with network visualizations which help in understanding the patterns that may not be observed by just examining the metrics (Everton, 2008).

In his quest to find out how law enforcers utilize social media platforms to covertly examine unlawful movements, Wyllie (2015) established that law enforcement agencies are employing the use of popular social media platforms such as Facebook and Twitter in various ways to assist in combating offences and give services to their societies. More significantly is the fast growing application of social media investigating means to uncover evidence of criminal activity which the lawless individuals on their own volition frequently place to the online network habitually paving way for swift apprehension of themselves and impending likely sentence. Detectives can use social media platforms as an investigative tool by creating undercover pseudo accounts in order to use to harvest intelligence on crimes and suspects or get the identity and movements of suspects (Murphy & Fontecilla, 2013).

Piett (2012) reaffirmed that main civil security units have incorporated probe of social media users into their ways of forensic analysis, singling out conspirators who may exchange online information with each other or even gathering correspondence online communication session logs which evidently maps to an incident and incriminating the accomplices. Lawbreakers have more often made work easier for law enforcers by boasting on the internet (also occasionally sharing still and motion pictures) of the offences they have committed. Cockbain *et al* (2011) outlined that law enforcement agencies have now adopted Social Network Analysis in thoroughly probing to unravel the manner in which the accomplices are connected. Notably, social network analysis is a crucial means in aiding law enforcement using various approaches. Initially, it can be able to aid probing schemes and give a hint on suspects for interruption or arresting the culprit(s). Next, Social Network Analysis can give an unbiased ways and counteract individual prejudices or suppositions depending on past knowledge of persons or of other related undertakings. Last but not least, is that Social Network Analysis is capable of finding areas of significance. The central scientific essence and aim of Social Network Analysis is basically about people's relations when defined as who we are and how do things. Personalities of different people, ethnicity, tribe or education interact creating a particular pattern of relationships. It this these patterns that Social Network Analysis endeavours to study and answer numerous questions about people's sociality in a social media network.

Fatih and Bekir (2015) succinctly puts it that social media as a source of intelligence has the capacity to provide police with access to significant volumes of material, posted by all manner of people, and divulging astonishingly candid information to a public audience. This information is often contemporaneous with events of interest to police and can be documented and traced. Such information can prove to be a source of valuable intelligence, at times even capable of assisting in the pre-emption of crime. Increasingly, social media is used as a tool for gathering community intelligence that could be used to inform operational decision-making. Analysis of social media content has been used to assist with offender apprehension and the location of missing persons (CLEDS, 2013).

Krebs (2002), studied and mapped the 9/11 Al-Qaeda terrorist network by gathering publicly available information on 19 hijackers of Al-Qaeda and applying basic Social Network Analysis centrality and community measures with the help of Social Network Analysis tools to identify the key players and leaders in the network. This research gives some vision for further work and research into the terrorist networks analysis. Visualization methodologies ease the understanding of a complex inter-connection of nodes in a network (Basu, 2005).

Elsewhere across the globe, Mateescu *et al* (2015) observed that the relationship between security intelligence force and online communication technologies has changed significantly in the last ten years. More often than not, law enforcement officers are doing their work manually, perusing public profiles, doing searches on various sites, or creating profiles to connect with targets of interest. But as Internet penetration, social media usage, and mobile device usage have all increased, law enforcement agencies and technology vendors alike have begun focusing on new forms of training and technology, including systems that would automate social media surveillance activities.

In USA, law enforcement agencies employ Social Network Analysis tools to investigate users on social media for a wide range of reasons; which include strategies such as discovering criminal activity and obtaining probable cause for a search warrant, collecting evidence for court hearings, pinpointing the location of criminals, managing volatile situations, witness identification, and broadcasting information or soliciting tips from the public. Investigative uses of social media are either targeted – focusing on individuals and their networks – or general concentrating on monitoring a delimited geographic area – either for identifying specific incidents or producing predictions of criminal risk (Rodríguez & Rodríguez, 2005). Social Network Analysis can be used to investigate specific targeted nodes or networks or concentrate the monitoring to a given geographical area whenever an incident is identified or predicted Mateescu *et al* (2015). Sources of intelligence can include publicly accessible posts shared by users who have not limited their privacy settings, information obtained by accessing a user's social network (e.g. adding a criminal suspect as a "Friend" on Facebook to view private posts), or the use of a search warrant to obtain a user's private communications from social media platforms themselves.

Similarly, Keenan, Diedrich and Martin (2013) resonated that law enforcement agencies need to be able to respond to criminals who are using technology for their criminal enterprises. Today, law enforcement agencies have numerous modern high-technology tools including license plate readers, digital voice recorders, mobile data terminals, electronic control weapons with high-quality digital cameras, and rapid identification devices. While some agencies have successfully used these tools, many have been caught off guard by criminals' growing use of online social network platforms.

1.2.4 Social Media and Law Enforcement in East Africa

According to CIPESA (2015) internet access report, Kenya's is far ahead with 69% penetration rate representing 29.6 million of its populace whereas Rwanda came second at a distance of 31%, Tanzania (22%) and Uganda at fourth position with penetration of internet access at 20%. With regards with to mobile access rates, the same pattern was depicted with Kenya maintaining the lead at 84%, Rwanda (74%), Tanzania (71%) and last but not least was Burundi at 31%. Thus as the number internet continues to grow exponentially, the content created also grows tremendously. However, most East African governments have enacted laws that imperil the right to freedom of expression which consequently abuses internet freedom including surveillance and interception of communicating devices. For instance, towards the end of 2014, Ugandan government enacted a bill that tampered with privacy rights of her citizens. Almost at the same time, Tanzanian government passed a cybercrime act that left loop holes for violating internet users' rights of citizenry.

A comparative analysis reported by CIPESA (2015) on the perception of whether the East African nations are monitoring and carrying out surveillance on its populace, an overall 61% of respondents believed that law enforcement agencies were monitoring their online communications, Kenya was the highest at 91% followed Tanzania at 80%. At least 75% of respondents in Rwanda believed that their government did not intercept or monitor its citizen's communications.

1.2.5 Social Media and Law Enforcement in Kenya

Social media web pages can be a rich repository of crucial data for investigating workforce as they accomplish their civil security enhancing assignments. Waters (2012) advises that law implementing officers can utilize social media to communicate with the civil society, crime investigators can have their ways into various online social networks to aid in recognizing and arresting of suspects, forensic crime investigators can use online networks tools piecing together digital evidence pertaining crimes committed and big data analysts can also make use of social media platforms to help in designing and creation of diagnostic appraisals. Social media evidence includes, among other things, photographs, status updates, a person's location at a certain time, and direct communications to or from a litigant's social media account (Murphy & Fontecilla, 2013).

According to the report released by the Kenya Communications Authority (2016), there were an approximately 30 million online users in Kenya by mid year of 2015, with an estimate of 70% of the Kenya's populace having access to the internet connectivity. Social media is widely used in Kenya. As cited by KCA, Bloggers Association of Kenya (BAKE) in their June 2015 report indicated that popular social network platforms are becoming a valuable resource in which Kenyans can express any subject of interest to them and also performing their freedom of speech. The report further indicated that the most preferred platforms such as Twitter and Facebook were 10% of its population with almost one million established periodical loyal users on Twitter in which the mainstream part of it are daily Twitter members.

In Kenya, access of online data by law enforcement is still unclear because there is no proper law that supports it. According to a report by Kenya Communications Authority (2016), the state has not so far enacted a law relating particularly with cyber crime. For instance, section thirty one (31) of the Kenya Information and Communication Act, stipulates that licensed telecommunication practioners are lawfully forbidden from carrying out any technological requests necessary to permit legal interception of users data on transit or in storage. Furthermore, section 15(1) of the same Act but a subsidiary of Consumer Protection under regulations of 2010, loosely translates that "any license holder should not be obliged to scrutinize, reveal or permit any individual to probe or divulge the subject matter of any information of any service consumer communicated through the licensed systems by eavesdropping, storage, or other type of tapping or investigation of communications and related information." However, the recently adopted Kenya Information and Communications (Registration of Subscribers of Telecommunication Services) Regulations 201427 allows access to personal or classified information on clients without a court order. The report gives the current state of affairs by Security and law enforcement agencies. It gives an account of two intelligence agencies in Kenya that deals with cybercrime.

Currently, the key intelligence body in Kenya is NIS, which was instituted by the 2012 NIS Act. It is both the internal and external intelligence body of Kenya which as has an extensive authorization. Its core task is to collect, assemble, examine and convey or disseminate with the applicable state bodies, security intelligence and counter intelligence with an aim of discovering and recognizing threats or possible threats to the state security. The Kenya police also has a unit vested with powers to carry out surveillance. The unit was setup in the National Police Service Act 201127 and the National Police Service Commission Act of 2011. According to Apantech (2017) however, NIS detectives employ various techniques to retrieve both call logs and call data records and even intercepting mobile calls. In fact, Privacy International reported that they (NIS) are illegally and covertly stationed at telecommunication operators' precincts.

Jambo news spot (2015) recounted of how the East African Data Handlers (a city data company) conducted an independent forensic investigation and revealed the social media user who leaked the photos of the school girl arrested with bhang hidden in her underwear. The investigation was independently carried out with the hope that it will aid the law enforcement officers to be able to track the original sources of the indecent photos and that to inform Kenyans that all their actions online can be tracked and they should use technology responsibly. Regardless of above aforementioned constitutional and legal developments on privacy protections, Kenya's most telecommunication operators often gives out their clients data both law enforcement agencies and detectives (Apantech, 2017). According to Mutung'u (2012), over 17 million Kenyans are internet users creating voluminous content across the various social media platforms. However, it is clear many countries notably the developing countries like Kenya still use conventional real-world laws and adhoc policies to deal with social media criminality concerns.

However, it is important to take cognizance that other social media platforms are more willing to surveillance than others. Features such as default privacy settings play a role in the revelation of the individual posted or shared online content to law enforcement bodies. Moreover, social media platforms such as Facebook make social networks discernible beyond their contiguous cohorts. These attributes pose fresh queries about how social media monitoring and law enforcement investigative procedures on the basis intersect and shape each other. Towards the end of 2016, the communication Authority of Kenya contracted an Israeli firm to be carrying out real time surveillance on social media users and map ties between them Apantech (2017). The tactics and techniques used by Kenya's law enforcement agencies to monitor and carry out surveillance on her populace included but not limited to use of FinFisher (surveillance software), cracking of passwords, use of Geographic Position Systems technology to monitor and track the locations and movements notably using Subscriber Identity Module (SIM) cards, scrutinizing communications by use of Internet Protocols and call logs among others (CIPESA, 2015).

It is against this backdrop that there was need to study and bring to light the tools and techniques for investigating and gathering intelligence from social media platforms to aid Kenya's law enforcement community.

1.3 Statement of the Problem

This study sought to present the application of Social Network Analysis techniques in investigating and retrieving meaningful information from Facebook and Twitter social media platforms from the selected individuals in order to analyse mined data and present digital forensic evidence and intelligence that can aid in law enforcement agencies in Kenya.

The research revolved around an ego. The egocentric analysis entails the examination of a single node within a given social network platform including all the actors the node is linked to. Social theorist postulates that influence between social network friends goes up to three degrees or three intermediary levels (Campbell, Dagli & Weinstein, 2013). In criminology and law enforcement, Social network analysis has been proved to be a powerful tool to learn the structure of a criminal network notably in social media platforms. It allows researchers to understand the structural relevance of single actor and his/her connections amongst members of a given social network by defining the key concepts to characterize network structure and roles Ferrara *et al* (2014).

Social media has become an indispensable tool of communication in contemporary society. Rather than meeting and communicating in person, conversations are increasingly occurring on social networking sites like Facebook, Twitter, MySpace and YouTube. Individuals are spending more and more of their time online, and establishing their public and private identities through cyberspace mediums. The revolutionary increase of Kenya’s populace in embracing and using online social media platforms to interact and communicate has posed a serious challenge to law enforcers in obtaining the digital forensic evidence of the cyber criminals.

In essence, the evident increase in the sophistication of cyber criminals has a significant impact that can threaten the national security if it goes unabated. Presently, use of social media in mining crucial digital or forensic evidence by law enforcement bodies in Kenya is a novel idea that needs to be explored and implemented.

1.4 Purpose of the Study

To perform experiments on real life social networks available in commonly used types of popular social services such as Facebook and Twitter in quest to demonstrate how Social Network Analysis as an invaluable tool can be employed to extract latent knowledge or information from networks encountered in nature, especially networks formed by people.

1.5 Specific Objectives of the Study

The objective of the study was to demonstrate how to investigate social media in quest of acquiring forensic evidence to aid law enforcement agencies in Kenya.

The Specific objectives of the study were:

i) To visualize social networks and clusters to uncover the patterns of the social relationships of people in investigating crimes committed over selected popular social media platforms in Kenya.

ii) To demonstrate the use of social network analysis tools over selected popular social media platforms in analysing and identifying the most important actors in a network using known metrics derived from the user's network data in Kenya.

iii) To determine how demographic and other related information of social media users can aid in tracking the online offenders in Kenya.

1.6 Research Questions

The questions for the study were stated as follows:

i) How can social network analysis aid in depicting and visualizing patterns and connections of people in investigating intelligence crimes committed over selected popular social media platforms in Kenya?

ii) How can social network analysis tools help law enforcement agencies in analysing and identifying the most important actors in a network using known metrics derived from the users' network data in Kenya?

iii) What can demographic and other related information posted by social media users' aid in tracking the online offenders in Kenya?

1.7 Significance of the Study

There is significant increase in the number of cyberspace profiles, making social media platforms important research repository for studying about almost everybody's deeds, likes, and way of thinking.

Therefore, this study is hoped to benefit the Kenyan law enforcement agencies in adopting or intending to embrace social network analysis tools by providing them with a new perspective of looking for network forensic evidence in the various social media platforms. This way, the law enforcement agencies will be able to make sound decision based on the findings of this research alongside the existing investigating techniques. This is because the findings obtained from this study is expected to provide the stakeholders in the field of social media and other related technologies a concrete understanding of how intrinsically mine and analyse data from the various social media platforms and use it as court evidence against the subject or target under investigation.

Thus the justification of the study lies in the fact that there is a paradigm shift of Kenya's populace of all ages embracing the use of various social media platforms. When appropriate tools are employed in social network analysis, it will empower the demanding needs of intelligence analysts, law enforcement, investigators, researchers, and information workers because it provides insights into patterns and trends hidden in the social media data. For instance, NodeXL data visualization platform provided and depicted quick to see multi-level links among entities and model different relationship types. Furthermore, Social Network Analysis metrics revealed the most interesting “suspects” in complex webs in the context of this study.

Since Social Network Analysis methodologies are aimed to determine the pattern of relationships between online network social media users, they are well endowed to be used by law enforcement agencies in detecting clusters, unearthing their patterns of relations, isolating important and central individuals in the entire network structure.

This study is therefore significant in that it highlighted the pertinent underlying issues that could immensely help in mitigating crime rates and related vices that are committed over online social media platforms. Based on the findings of this study, recommendations was made which is hoped to help the Kenya’s law enforcement agencies adopt social network analysis in investigating and even intercepting suspicious online characters.

1.8 Scope of the Study

This study only focused on popular selected social media platforms in Kenya using egocentric network of selected respondents from five Universities in Migori, Uasin Gishu and Kakamega counties. Consequently, analysing a large socio-centric network of users of social media platforms in Kenya was beyond the scope of this study.

1.9 Limitations of the Study

More specifically, the challenge for social network analysis Social Network Analysis is that personal Twitter networks can only be mined to a limited extent. Currently, the Twitter application programming interface (API) imposes severe limits on this activity. Let's say, for instance, that you are interested in who is following a particular account rather than in what is being tweeted. Twitter imposes strict limits on how many links you may download in a 15-minute time span. In practice, that would mean that downloading a meaningful Twitter account network would take days—and often the API will cut off your access in the middle. Efforts to streamline the Twitter API are underway. However, at this time, the limits are still in place (Kwak, Park & Moon, 2010).

Other limitations of this research are core legal issues that emerges while mining data from social media platforms . These include a person's right to privacy and the investigator violating a site's terms of service (TOS). Furthermore, the TOS of a website provide a set of rules that users agree to follow when they use a site. A site might also use them to offer disclaimers or deny legal responsibility for some actions. Each social media site has its own TOS, and so, there is no single rule an investigator can follow in order to ensure he or she is within the bounds of allowable activity on every site—except that it is important to read the terms for each site.

This study created pseudo-accounts online accounts for investigative research purposes only. However, the main issue that arises with respect to TOS is that of creating profiles with fake identities. If you are investigating someone, it is obvious that you would not want to use details from your personal account. You may not want to personally add someone as a friend or otherwise reveal your true identity as someone who is looking at what the target posted. Thus, creating a fake or anonymous account can be a solution (Narayanan & Shmatikov, 2009).

Facebook registration and security of account explicitly exclude fake accounts. Aside from the explicit requirements, the requirements of keeping contact information upto-date, not sharing passwords, and not transferring accounts may prohibit many activities investigators may take with fake accounts. However, Twitter has no rules about using a true identity. Google, and its social media services including YouTube and Google+ used to have a real name policy, but they eliminated that policy in 2014. Moreover, Social network analysis as a methodology essentially deals with mined data that has already been exchanged online - not real time data mining and analysis (Kirchner & Gade, 2011).

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Since social media has now been integrated in everyday life to a myriad number of online users across the globe, the internet of today is by far a different place from the online world of yesteryears. In particular, social networking has tremendously turned out be a trendy means of communication on various social media platfotms. Gone are the days of invariable information on a social media site that was modified gradually. Nowadays, social media has transformed the cyberspace into a thriving, 'always on' environment of continuous acts, almost at the same time interactive communications and myriad of motion pictures uploads to popular platforms every minute (Gudaitis, 2015).

Wasserman and Faust (1994) define 'social media' as a construct from two areas of research, namely communication science and sociology. A medium, in the context of communication, is simply a means for storing or delivering information or data. In the realm of sociology, and in particular social network theory, Social Network Analysis and social networks are social setup that constitute of an array of social actors such as individuals, groups or organizations with a complex set of dyadic ties among them. Giuffre (2013) highlighted that by employing network analysis, one is able discover crucial information pertaining the network communities because it revolves around the relationships between nodes or subgroups of a network.

In recent years, social media sites have emerged as a useful tool for friends, coworkers, and families to keep in touch and interact with one another. Persons and groups can share still and motion pictures, organize meetings or arrangements for the particular days of the week or provision of information on newsworthy events to their friends, family, or customer base.

2.1.1 Social Media

Gupta and Brooks (2013) defined social media that it involves the creation and sharing of information. Information on social media is meant to be promiscuous—it can be and often is created by numerous people at different times, and consumed by numerous people at different times. Specifically, Williamson and Ruming (2015) asserted that regardless of whether sharing is broadened or focused, every social media technology allows for the spontaneous creation and sharing of information. This ability has led to the creation and sharing of petabytes of data—more digital information is now created in a day than existed in the entire written works of mankind from the beginning of recorded history. Figure 2.1 below depicts the definition of social media.

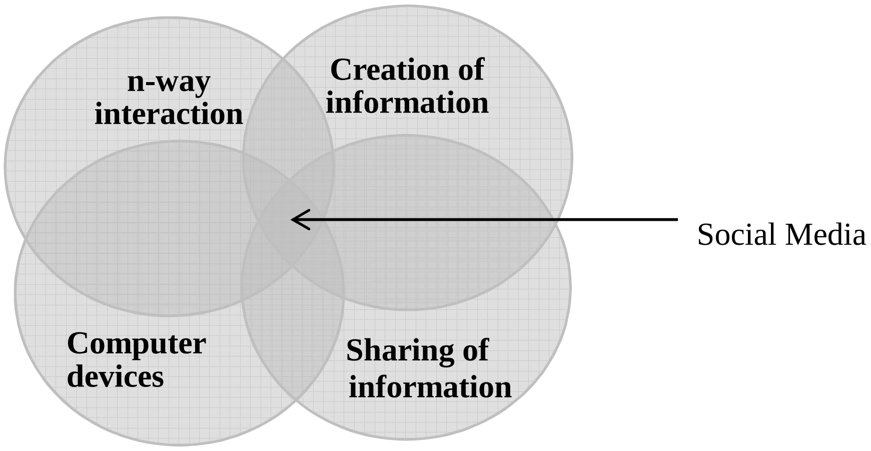


Figure 2.1: Definition of Social Media

Source: *Gupta and Brooks (2013)*

Given this colossal amount of data on social media, Blomberg (2012) advises that law enforcement agencies can capitalize on some of the procedures, tools and methods already employed by business firms to investigate what individuals are saying about criminal activity and local services. With the right technology, the forums lawbreakers exploit to disseminate their information can also be investigated to avert their next offence. With proliferation of smartphones, easy access and quick internet connectivity, virtually any information is within reach or fingertips of every possible criminal-minded individual. Law enforcement wants to know how suspects are talking to each other, but they need an enabling technology so that they don't have to sift through individual social media sites and pages. The fact is, social media can be a great repository of information and medium of communication, but can also contain crucial evidentiary information for computer digital forensics, social network interactions, and cybercrime issues. Social media platforms forms an online social network that enables individuals to interact and exchange niceties as well as facilitating the users to search for one another then establish a link as online friends.

2.1.2 Social Network Structure

The Internet has become a central point for information-sharing in today's world. A strong part of the so-called Web 2.0 is represented by social networks. A social network is an interconnected network of individual entities which share a mutual interest and gain a method of interaction or information sharing through the service. Social networks come in many different facets. Some are strong in a particular geographic location like Orkut in Brazil, VKontakte in Russia, or Mixi in Japan. Others are well known globally, like Facebook and Twitter (Wüest, 2010).

Kirchner and Gade (2011) gave a synoptic basic summary of a social network that it consists of nodes (vertices) that are connected to other related nodes by links (relationships). The connection between two nodes is called an edge. If all the nodes in a social network are connected to each other, it is called a fully-connected network. A path refers to a collection of nodes that are connected by a link. De Nooy *et al* (2005) defines social network as a finite set or sets of actors that share ties with one another while social network analysis involves the detection and interpretation of the patterns of social ties among actors. Some actors are generally embedded relatively deeply within a subgroup, while others sit more on the periphery, serving as bridges between subgroups (Everton, 2008). Network scholars have studied a wide array of categories of ties. These include communication ties (such as who talks to whom, or who gives information or advice to whom), formal ties (such as who reports to whom), affective ties (such as who likes whom, or who trusts whom), material or work flow ties (such as who gives money or other resources to whom), proximity ties (who is spatially or electronically close to whom), and cognitive ties (such as who knows who knows whom). Figure 2.2 depicts a theoretical social network.

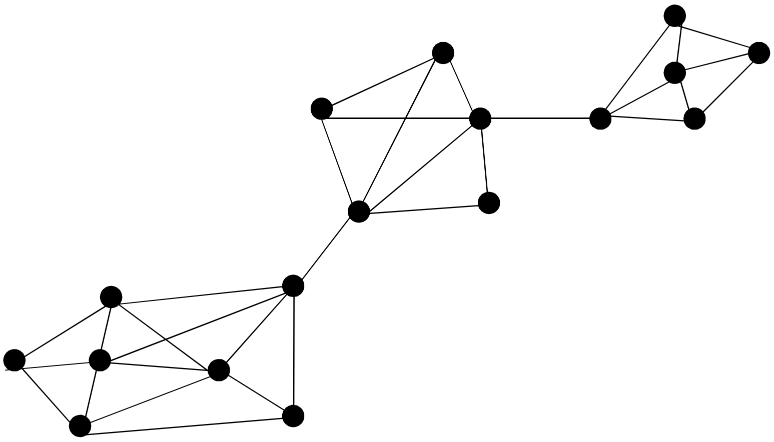


Figure 2.2: Sample Social Network
Source: *Author*

Gupta and Brooks (2013) noted that social networking platforms emphasize and enable users to create relationships and foster their personal and/or business networks—to meet and get to know others. They put the social in social media. The most popular social media platforms employ some sort of social networking functionality because humans crave the ability to network with others. This craving and tendency to maintain relationships seduces users and keeps them coming back regularly over a long period of time. Social media evidence includes, among other things, photographs, status updates, a person's location at a certain time, and direct communications to or from a defendant's social media account (Murphy & Fontecilla, 2013).

With reference to social media usage, Murphy and Fontecilla (2013) described that there were hundreds of social networking websites with each catering to a different demographic and providing a different type of content. Moreover, their users are constantly creating massive amounts of data. Twitter users send one billion tweets every two and a half days, Instagram users upload forty million images every day, 10 Facebook users share 684,478 pieces of content every minute and YouTube users upload forty-eight hours of new video every minute. Social media users create more than just photos, videos, and tweets. They share other information, such as their location as well. "As of 2012, seventeen billion location-tagged posts and check-ins were logged." The myriad and continually changing ways to share information via social media has resulted in a digital goldmine of potential evidence, such as profiles, lists of friends, group memberships, messages, chat logs, tweets, photos, videos, tags, GPS locations, likes, check-ins, and login timetables.

According to Jamali and Abolhassani (2006) a social network is a social structure between actors, mostly individuals or organizations. It indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familiar bonds. Social network analysis is the mapping and measuring of relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The nodes in the network are the people and groups, while the links show relationships or flows between the nodes. Mapping and examining social networks can unravel the identity of the significant nodes and relations as well as monitor the propagation of ideas (Gupta & Brooks, 2015). Whenever behaviours of observed nodes are analysed on social media, detectives can classify these behaviours as either group or node-level behaviour. The Node level behaviour belongs to a particular actor whereas group level behaviour is exhibited by a network sub group or cluster (Zafarani & Liu, 2014).

Actors in a network can possibly become online friends if they happen to be close geographically (Kadushin, 2004). Social network analysis provides both a visual and a mathematical analysis of human relationships. Social Network Analysis either requires data on the ‘whole network’, in which case boundaries of the population of interest must be drawn, or upon ‘personal networks’, where all the ties of an individual ‘ego’ are recorded along with the ties between their ‘alters’ which are called ‘egonetworks’.

Since a social network consists of actors and binary relations between them, it may be modelled as a complex network, or graph, in which actors of the social network would correspond to nodes in the graph and relations would correspond to the edges. Then, applying the methods from graph theory would become possible in social network data (Semenov, 2013). Whenever all the individuals and relations are established, the numerous statistics generated gives crucial information about the network (Golbeck, 2013).

The proliferation of social media is an indisputable fact. Never before has so much sharing occurred between people of every walk of life, all facilitated through this modern communication medium. Over the past few years, social media has become an integral part of the social, cultural, business and political process in most parts of the world, affecting almost every facet of life – including criminality (Blomberg, 2012).

2.1.3 Social Networks: Privacy Analysis

In analysing revelation of information on social networks, Gross and Acquisti (2005), concluded that fears on confidentiality on various online social media sites have been an issue to worry about because the development and growing reputation of various social media platforms. Issues relating to trailing, stealing and using one's information, sexual offenders and employment constantly give on increasing, as well as the moral values pertaining storage of information and the handling and dissemination of the aforesaid data than an individual who only utilizes the web page resources. The findings supported theory of privacy by Moor (1997) which highlighted that privacy concerns doesn't actually have to entail security violations. The probable damage to a single online user actually zeroes down to the extent in which that user connects and interact on social network and also the extent and measure of the information that they are voluntarily willing to expose or publicize.

Drawing from Witnov (2011) social network sites like Facebook or Twitter are becoming a large databases of self-reported information having various kinds of information including from images of themselves to evidence of fraud to the details of criminal conspiracies. An individual having numerous followers or friends or participating on various clusters has a high likelihood of being hurt by a contravention than an individual who only utilizes that social network platform. Users share private content, such as personal information or photographs, using online social networks applications. Online social network clientele ought to have faith on the platforms they are using pertaining the protecting of their private information although such online networks offers benefits from examining and sharing that information.

In essence, regarding data of interest, the graph data can reveal friendship connections, communication patterns and common group membership. Furthermore, activity data can inform the investigators of time, frequency of log-in in a particular and typical behaviour on social media user.

2.1.4 Social Network Analysis (SNA)

Johnson *et al* (2013), pin pointed that Social Network Analysis is an assessment and evidence provision tool that can be employed to depict, trace and determine social associations. Through quantifiable measures and powerful and informative visualizations, law enforcers can employ Social Network Analysis to uncover, find hidden patterns and depict the social networks of crime offenders. This way, Social Network Analysis acts as crucial tool for law enforcement agencies in apprehending the lawbreakers. Although Social Network Analysis seems to be complex when integrated with technology, it has shown to be simple to use. Thus, utilizing the accessible data, law enforcement bodies can gather and analyse a lawbreaker's social network using different means techniques that hitherto not employed. Johnson and Reitzel (2011) posit that Social Network Analysis is a social science technique that can provide investigators with a series of centrality metrics and powerful visualisations on which they can swiftly discern, analyse and portray crimes committed by network members with the aim of coming up with meticulous barring schemes.

Rahim, Amalina and Sulaiman (2015) while studying political bloggers, acknowledged that Social Network Analysis offers both visual and mathematical techniques of analysing online users relations. Social Network Analysis methods have been successfully employed to a range of problems to unearth relations that could not have been understood using other conventional techniques. Kriegler (2014) concurred that indeed Social Network Analysis can be used to gather profound information pertaining the structural network formations being mapped. This way, latent patterns in data can be revealed to enable and facilitate law enforcement agencies use the findings to take the next course of action. Thus the main purpose of employing Social Network Analysis techniques is to comprehend network communities by mapping the relations that link them as a network and thereafter determine key players or groups and ties between the nodes (Faust & Fitzhugh, 2012)

Drawing from Ravindran and Garg (2015) *data mining volume*, Social Network Analysis techniques are derived from sociological and social-psychological theories and take into account the whole network (or, in case of very large networks such as Twitter -- a large segment of the network). As stated by Gupta & Brooks (2015), since social media entails the creation and sustainability of social networks and relations between nodes, comprehending Social network analysis is as good as discerning social media. Use of online social media platforms is swiftly taking a centre stage globally. Mining, analysing and disseminating the findings is becoming complex because it transcends geographical boundaries. Conventional modes of investigations are no longer the effective in capturing the nuances of criminalities committed especially by online users. Social network analysis methodology is a prospective technique to address this gap. Therefore, social network analysis proves to be a prospective diagnostic tool for mining and analysing relevant data with respect to patterns of relationships among actors under investigation in a given network.

With a significant growth in the number of online users, the data created has expanded considerably, prompting the necessity of discernment into the unexploited rich source that is social media repository. Social Network Analysis is a powerful methodology when the appropriate tool is applied in research. It helps to tap the tremendous amount of valuable social data in Facebook, Twitter, LinkedIn, and Google+. This agrees with Mincer and Niewiadomska-Szynkiewicz (2012) that Social Network Analysis is a significant and effective tool for extracting knowledge from amorphous big data. It aids one to discover who's making connections with social media, what they're talking about, and where they're located. Furthermore, it enables one to learn how to combine social web data, analysis techniques, and visualization to find what you've been looking for in the social haystack—as well as useful information you didn't know existed.

Choudhary and Singh (2015) portrayed Social Network Analysis as an application of network theory to analyse social networks in terms of social relationships. It comprises of nodes (actors, persons, organizations etc.) within the network and ties representing relationships (friendship, kinship, conversation, financial transaction etc.) among the nodes. Social relationships may be in the form of real world offline social networks (like friendship, kinship, communication, transaction etc.) or it may be online social networks (like Facebook, Twitter etc.). Various Social Network Analysis measures has been used for representing interaction among actors, examining strong or weak ties, identifying key/central players and subgroups in network, finding topology and strength of network (Karthika & Bose, 2011). In his study entitled "*The Strength of Weak Ties*" Granovetter (1973) underscored that important channels of communication to be closely monitored are the ones that are rarely utilized and usually located at the network's periphery and also comprise of quite dense cliques.

A similar illumination was resounded by Mincera *et al* (2012) that social network analysis is a group of graph theory based techniques that can be used to retrieve meaningful knowledge from networks formed by various actors. They argued that data about relationships of people are commonly available like never before, and applying analytical methods to them became a source of unique and valuable Knowledge. In a nutshell, needless to say that a common tool for the criminal investigator is social network analysis. It graphically depicts the suspects and their connections to other people or artefacts and allows the computation of certain metrics. Not all the facts composing the entire picture of a case may be known to one investigator.

2.2 Relating SNA and Social Network in Law Enforcement

The rapid growth and availability of technology associated with social relationships in the cyber world is changing the traditional context of competing oppositional interest between law enforcement and criminality. Johnson *et al* (2013) advises that physical inspection of social networks is likely to be challenging, wasting time and discretionary rendering it highly error-prone. Instead, SNA offers a methodical method for probing voluminous quantity of data on individuals and their relations. It enhances law enforcement usefulness and competence by employing intricate information concerning persons socially associated to suspected criminals. This generally paves way to enhanced quick way of terminating numerous offences and advancement of improved crime deterrence schemes.

The numerous social media platforms have turn out to be a tool that offenders are employing for wicked and unlawful reasons. Such-like criminal activities on social media comprise of persons organizing a criminal-linked random criminals groups or using a social media platform to plot a burglary, online users who exploits others getting connected to a social media website to recognize and communicate with possible sufferers, and terror clusters by means of social media to sign up fresh members and take up criminal acts. Social media might be a crucial probing tool to discover and avert unlawful acts. Social media has been used for society outreach proceedings which comprise availing crime aversion inklings, offering offenders maps, and beseeching clues about puzzling offences. The enormous expansion in the use of social media is echoed in the amount of attention worldwide now being given to its role in law enforcement (Piett, 2012).

Law enforcement has begun to recognize how helpful these websites can be as a source of valuable insight, particularly into public sentiment. In many cases, law enforcement can capitalize on some of the processes, tools and techniques already used by commercial organizations to find out what people are saying about criminal activity and local services (CLEDS, 2013). However, Fatih and Bekir (2015) gave a caveat that the high visibility of young people and their activities on social media also complicates surveillance and interpretation of online speech.

According to a 2015 Pew Research Survey, 90% of young adults (ages 18 to 29) are active on social media, and another Pew study found teens (ages 13 to 17) are active across a wide variety of social media platforms. As a result, one concern is the potential over-criminalization of youth, particularly minority youth. On the one hand, monitoring youth on social media can provide opportunity for intervention (Mateescu *et al,* 2015). Ego-centred network analysis focuses upon an individual agent and his/her relationships with others. This approach allows the researcher to depict the node’s domain of power. It establishes node’s connections and unearths the nature of those intrinsic relations. This approach, concentrating on the specific actor, is worthwhile when frontiers are difficult to delineate in a large network.

Johnson, Reitzel, Norwood, McCoy, Cumming and Tate (2013) summarily underscored that Social Network Analysis has not only been demonstrated be an effective investigative tool for law enforcement community, but also helps bridge the gap between detectives and other organs such as legal fraternity. In Kenya, a court ruling on 6th February, 2017 declared that there will be no charges for online posted offensive messages as the judge termed it unconstitutional the section of the penal code that created criminal defamation (KTN Prime News, 2017).

CLEDS (2013) report indicated that the use of social media for law enforcement purposes is growing worldwide and it is apparent that social media has the capacity to be a wide-reaching and engaging law enforcement tool. Law enforcement agencies recognise that in order to obtain the benefits of social media outcomes, they need to participate in social media and social networking platforms, most of which are decidedly more open and dynamic than they are accustomed to. Thus Social media sites and resources should be viewed as another tool in the law enforcement investigative.

2.2.1 Successful Application of Centrality Measures in Forensic Law Enforcement

In the context of social network, power is measured by means of centrality metrics obtained by use of mathematical relations which computes ratings for every node on the basis of the node’s rank in the social network. When applied effectively, Social Network Analysis can reveal the relative importance of each entity by analysing power as conferred by links to other network members, rather than individual personality traits (Hanneman & Riddle, 2005). This can identify important associations contained inside the criminal setup as main aim of interruptions. In United Kingdom for example, law enforcement officers have benefited from the chronicles of mapping ties between nodes and merchandise implicated in crime specifically planned crime, in conformity with systems such as Organized Crime Group (OCG) mapping (Borgatti *et al.,* 2009).

According to Johnson *et al* (2013), the centrality of actors compromising those representing law-breakers, distinguishes the reputation of individuals to the general working of the network system. It implies their significance to the criminal network, position, degree of activity, be in charge of the flow of information, and associations between nodes. Fundamental centrality metrics gives additional facts. *Degree* measures the number of ties a specific user have, whereas *betweenness* measures how essential it is to the flow and *closeness* signifies how fast the actor accesses data from the social network setup. Actors are sorted and given position according to their centrality, with topmost playing the most important responsibility. These measures do not inform an investigator what the criminal makeup ought to be but nevertheless, they give details on the real composition of the social network. The worth and litigable intelligence of each of these metrics is influenced by the information the investigators needs.

Choudhary and Singh (2015) demonstrated the use of social network analysis and media based evidences in quest of finding the possible leader after Bin Laden by examining the Al-Queda terrorist network. By applying various social network analysis measures they recognized profiles of high rank criminals, various criminal communities and public groups on Facebook. Knowing the persons one associates with can easily help the detectives to know the ambitions the node has (Semitsu, 2011) claims that detectives can slight efforts to gather crucial and non-crucial information that is linked to the individuals under investigation.

Social Network Analysis and centrality measures applies immensely in investigation of various dubious groups in revealing crucial leads that can ultimately aid the law enforcement agencies. As stated by Koschade (2006), the node with the highest level of betweenness will almost certainly be the critical node within the network. For instance, in Krebs's (2002) calculation of members involved in the 9/11 terrorist attacks, Mohammed Atta scored the highest on degree and closeness centrality, but not betweenness centrality, where he scored the second highest, and in Burgert and Senekal (2014), Noordin scored highest on all three measures of centrality. While these three centrality measures provide different perspectives on the roles played by entities, when taken together, an entity that consistently scores high on all three can be considered a key figure in an organization.

Although peripheral entities are not very active in the network under consideration, they are often part of other networks that are not currently considered, and because they provide ties with other networks. Nagl *et al* (2008) reasoned that nodes on the periphery receive very low centrality scores. However, peripheral nodes are often connected to networks that are not currently mapped. The outer nodes may be resource gatherers or individuals with their own network outside their insurgent group. These characteristics make them very important resources for fresh information not available inside their insurgent group.

2.2.2 Analysis: Metrics and Visualization

Social network metrics play an important role in analysing a social network's dynamics. Density and other related measures can help researchers gain an overall understanding the overall "shape" of the network (i.e., its topography); centrality measures can help identify key and peripheral actors within a network; clustering algorithms can help locate various subgroups within the larger network (and also provide additional information on the network as a whole); and brokerage measures can help identify actors and ties between actors that serve as channels for the exchange and flow of information and other resources (e.g., financial, affection). Social network analysts generally use a variety of metrics (rather than just one or two) in their attempts to gain an overall understanding of a network.

In a typical network and information flow, the innermost circle nodes signifies a more stronger social relations of the ego while the outermost circle are typified by fluctuating level of friendliness - also called sympathy or active network groups (Arnaboldi, Conti, Passarella & Pezzoni, 2013). By studying how information flows in a network, one is able to detect nodes that serve a pivotal role in a criminal network or having links to different subgroups (Ferrara, De Meo, Catanese, & Fiumara, 2014). Sharma and Strategy (2008) notes that although nodes at the periphery have less interaction with the entire network, they may be having links beyond the network and as result they can be a reservoir of new information. By studying how information flows in a network, one is able to detect nodes that serve a pivotal role in a criminal network or having links to different subgroups (Ferrara, De Meo, Catanese, & Fiumara, 2014).

2.3 Visualization as Forensic Evidence

According to Mulazzani, Huber and Weippl (2012) visualizations can be a very effective tool in forensic investigations of social networks. Hence visualizations is a crucial tool in Social Network Analysis because it enables the law enforcement agencies understand the behaviour of social media users from analysing and visualizing social networks. Law enforcement agencies can benefit a lot from visualization because they can predict criminal activity by monitoring connections between suspects, understand the dynamics such as discovering the leaders, followers and new individuals being integrated into a group.

The network visualizations are created by analysing dataset containing actors (suspects) and ties (relations) between them. Investigators can use Social network analysis metrics to examine these links to infer and derive crucial information about the nodes and subgroups of a given network. They can also find leads to enquiry by mapping known connections between crimes. This concept was corroborated by Mena (2003) that a social network can be easily understood and explored in a graph format, using people as nodes, relationships as edges and additional information (characteristics, preferences or affiliations) as properties. With Social Network Analysis visualization techniques, there is a possibility to detect clusters, identify the most important actors and their roles and unveil interactions between nodes.

Moreover, Denny (2014) pointed out that while social network theory can be readily applied in theoretical research, there is a general emphasis on the use of software to analyze and visualize network data once they have been collected. This way, visualizing social networks will provide a clear overview of a complex network at a glance and furthermore, furnish law enforcement agencies with an insight into patterns and trends hidden in the data. Visualizations empower Social Network Analysis specialists to discern unseen social network structures and patterns the nodes (Tayebi & Glässer, 2016). Centrality measures of a network that have a small number of hubs with many ties depicts a network free of scales and therefore has the power law of degree distribution (Van der Hulst, 2009). Visualization using social network analysis has proved to be a valuable tool that allows investigators comprehend the structural importance of single nodes and the ties amongst members, when viewed as individuals or actors of one or more cluster(s).

Ferrara *et al* (2014) observed that amongst the more frequently used analytic techniques, there is the mapping of interactions among the members of the organization and their activities by means of a graph which allows the overview the network structure, to identify the cliques, the groups, and the key players. He summarised and underscored that the possibility of mapping the attributes of data and metrics of the network using visual properties of the nodes and edges makes this technique a powerful investigative tool. Thus visualization as technique for investigating social structures on online social media can speed up discovery of digital evidence and help in law enforcement (Freeman, 2004).

Fundamentally therefore, mining and visualizing network users data can help researchers to unearth intrinsic patterns, groups and other related information in both simple and complex networks (Hansen, Shneiderman & Smith, 2011). Visualization functionality aid the law enforcement agencies to unearth hidden intelligence such as clusters, identify central nodes and give better understanding of network structure by utilizing limited information from a large dataset from a complex network (Yang, Liu, & Sageman, 2006). In quest of apprehending or disrupting a suspicious network, the investigators can seek to single out the actors who play pivotal role in network (Sparrow, 1991).

2.3.1 Network Clusters

Hoppe and Reinelt (2010) outlined a cluster as a subgroup of closely connected members of a network having similar attributes. Himelbloim *et al* (2017) posit that networks visualizations that depict low density and few isolated nodes are said to be community clusters. The algorithms for generating clusters aid in identifying a range of subgroups contained in the entire network as well as giving supplementary information pertaining the whole network (Everton, 2008). A dense cluster having numerous interactions is suspicious that warrants investigations to illuminate or discover more information Krebs (2002). Thus clustering networks into communities enables detectives to identify specific subgroups and if an individual belonging to a particular community commits a crime or he/she is a suspect, it will help the investigators to confine themselves to a specific subgroup (Zhu, Watts & Chen, 2010).

According to Zafarani and Liu (2014), nodes tend to connect to clusters if they have several friends in that community. Clustering coefficient is applicable to both a single node and the entire network. If one's alters are familiar to one another, then that actor will have a high clustering coefficient score and the opposite is true (Johansson & Tenggren, 2015). According to Tayebi and Glässer (2016) advises that detectives can establish who main actors of the network are and then start probing profoundly. According to Newman and Givran (2004), the modularity score of a network indicates the qualities of clusters in that network. The main aim of social network analysis is to discover nodes in clusters that highly dense and have strong relations between themselves (Wasserman & Faust, 1994).

2.3.2 Community Detection

Researchers have found that network density is positively related to the likelihood that actors within the network will follow accepted norms and behaviour, which is why a primary basis for moral order is highly-connected social networks. In Himelboim, Smith, Rainie, Shneiderman and Espina (2017), nodes within a densely connected clusters are likely to have same attributes, a concept known as homophily in social theory. The main reason is that in dense networks it is easier for people to monitor the behavior of others and prevent them from engaging in deviant behaviour (Everton, 2008). Thus denser online networks spread behaviours faster. The attributes of clusters exhibiting high density usually comprise of few interrelated or connected ties (Voigt, Hinz & Jansen, 2013).

The detection of network communities is a concept which entails dividing the network into distinct subgroups is a crucial subject in the analysis of networks and investigation landscape (Mincer & Niewiadomska-Szynkiewicz, 2012). The techniques of detecting and identifying communities helps to know groups of nodes densely connected than other in the entire network (Tayebi & Glässer, 2016). There are several algorithms for detecting communities in a large network (Blondel, Guillaume, Lambiotte & Lefebvre, 2008).

2.4 Network Centrality Measures Metrics

Social network researchers measure network activity for a node by using the concept of degrees - the number of direct connections a node has. Freeman (1978) defines centrality as the collection of measures that indicate how important a node is. Degree entails the aggregate sum number of ties a user has to other users. A node with bigger degree value in contrast to its peers signifies more prominence or influence in the network. These centrality metrics gauges the rank and importance of an actor with respect to other actors in that specific network (Johnson & Reitzel, 2011). Thus centrality can be determined by degree, closeness, betweenness and eigenvector metrics (Bonacich, 1987). The objective of analysing the centrality measures of nodes is to establish the key nodes in a given network in order to know their reputation, prominence or power in the entire network (Tayebi & Glässer, 2016). The higher the centrality score of a node, the more vivacious that actor is in the network (Ergün & Usluel , 2016).

Nodes that are positioned at the epicentre of a network structure tend to be more connected and therefore exert more power over others (Hanneman, & Riddle, 2005). Centrality metric measures is a concept which refers to how important a specific node and his /her rank in the whole network. In essence, it is a metric that describes and measure the attributes the specific node’s position in a network (Wasserman & Faust, 1994). The greater the degree centrality score is, the more significant this actor is to the social network because he/she might be representing a centripetal for information and resource flow inside that network (Van der Hulst, 2009). Borgatti, Everett and Johnson (2013) recommend that in general, the effective way of investigation is to target several actors with high centrality metric scores rather than isolating a lone actor.

According to Hopkins(2010), centrality measures of an individual in a network gives knowledge about that node’s position in the network whereas relations between centrality measures of all actors unearths the general structure of the network. As mentioned elsewhere in this study, numerous centrality metrics do exist. With regards to importance of centrality metrics to investigators, Wu, Carleton and Davies (2014) resonates that centrality helps to identify pivotal player in a network because it shows how deep rooted that actor is to the entire network in conjunction with other nodes. The centrality of an actor such as criminal, is a score showing the importance or significance of that node in the entirety position of the network (Johnson & Reitzel, 2011).

2.4.1 Degree Centrality

Centrality measure is a network metric used to highlight actors that cover relevant roles inside the analysed network. It shows people with many social connections. The degree centrality for a node is simply its degree such that a node with five (5) social connections would have a degree centrality of five (5). A node with one (1) edge would have a degree centrality of one (1). Hence this metric is the most basic network measure and captures the number of ties to a given actor. For un-directed ties this is simply a count of the number of ties for every actor. For directed networks, actors can have both *indegree* and *outdegree* centrality scores (Denny, 2014). Hanneman and Riddle (2005) argues that whenever a node exhibits several ties, that actor is said be having high prominence or is prestigious and that several other nodes would one to have a direct link to him/her. In particular, nodes with very high out-degree centrality scores are capable of exchanging or interacting with several other nodes in the network.

In this metric of degree centrality, higher values mean that the node is more central.
These show that while degree centrality accurately tells us who has a lot of social connections, it does not necessarily show who is in the "middle" of the network. Thus centrality measures shows how central or well-connected an actor is in a network. This theoretically signals importance or power and increased access to information or just general activity level and high degree centrality is generally considered to be an asset to an actor.

According to Berzinji, Kaati and Rezine (2012), an actor with the highest degree centrality score is placed at a strategic position with regards to connectivity and therefore plays a focal role of propagating information in the network. According to Mainas (2012), actors with highest degree centrality scores are likely be a broker planners or controllers. The gatekeeping position of an actor with high centrality value position him/her to be the controller of information or resources in the entire network (Wu, Carleton & Davies, 2014).

2.4.2 Closeness Centrality

Closeness centrality is a metric that depicts the node which is closest to all other nodes. It indicates who is at the heart of a social network because it measures how many steps (ties) are required for a particular actor to access every other actor in the network. The measure will reach its maximum for a given network size when an actor is directly connected to all others in the network and its minimum when an actor is not connected to any others. Denny (2014) advises that this measure is sensitive to network size and is decreasing in the number of actors in the network.

Closeness centrality for a node is the average length of all the shortest paths from that one node to every other node in the network. In the case of closeness centrality, unlike with degree centrality, smaller values mean that the node is more central, because it means that it takes fewer steps to get to other nodes. Closeness centrality corresponds the closest to what we see visually. Nodes that are very central by this measure tend to appear in the middle of a network. A node with strong closeness centrality also tends to be close to most people. This measure goes beyond what degree offers and emphasizes the geodesic distance between the nodes. In Wasserman and Faust (1994), connections with nodes having high closeness scores when put together with nodes having low degree centrality values can have indirect impact on the behaviour of the other nodes in that network. With regards to structural similarity of nodes in a network, McPherson, Smith-Lovin and Cook (2001) underscored that interaction between nodes with structural similarity, proliferates network connectivity.

Leadership can usually be identified using the centrality score of an individual in the network. According to Everton (2008), most social networks contain people or organizations that are more central than others and because of their position, they often enjoy better access to information and better opportunities to spread information. In an investigation, that means the person will be in a good position to hear from most friends of friends.

With regards to actors scoring highest closeness centrality metrics, Hanneman and Riddle (2005) claim that such nodes are capable of reaching other nodes at shorter path-lengths, or can easily be reached by other nodes in the network. They will be a good source of second hand information since it can reach him/her quite easily. The likely role that actor(s) scoring high closeness centrality scores is that of being an organizer because he/she can quickly reach many other actors in the network (Kaye, Khatami, Metz, & Proulx, 2014). A node with a bigger closeness centrality value can easily spread information across the network than a node with a smaller value (Johansson, & Tenggren, 2015).

2.4.3 Degree Betweenness

Betweenness centrality is a network metric measure a person's role in allowing information to pass from one part of the network to the other. According to Denny (2014), this metric can be described as the number of shortest paths between alters that go through a particular actor. Betweenness centrality is usually interpreted as a way of finding the most important entities in a network, for without these entities, the network loses coherence and becomes fragmented (Everton, 2008).

It is worth mentioning that betweenness is a measure of how important the node is to the flow of information through a network. In an investigation, a node with high betweenness is likely to be aware of what is going on in multiple social circles. Such a node with high betweenness has great influence over what flows and does not in the network. Thus it describes people who connect social circuits. In fact, Kirchner and Gade (2011) advises that one can use this centrality to identify the nodes that are pivotal to the success of the dubious network and, in turn, focus resources on investigating these suspects with crucial leads in social network.

According to Xu, Marshall, Kaza and Chen (2004), betweenness centrality is helpful measure to law enforcement agencies or detectives. In an investigation, it is important to know who are the individuals that other actors need to connect in order to link to the entire network and gather other valuable leading information. Boundary spanners supports information flow between network members who are either not in contact or have no trust for one another (Long, Cunningham & Braithwaite, 2013). A node with the highest betweenness centrality is a threat if that node somehow ceases to exist from the network because interaction will disappear in that network all of a sudden too (Johansson & Tenggren, 2015).

2.4.4 Eigenvector Centrality

Eigenvector centrality is a centrality metric that measures the influence that a node has in a network. Thus eigenvector centrality measure is high amongst influential people in the network. Denny (2014) defines this metric that it measures the degree to which an actor is connected to other well connected actors because it captures the value of having a lot of friends in high places. A node with high eigenvector metric score is linked to numerous nodes who are themselves tied several nodes (Tsvetovat & Kouznetsov, 2011). This was corroborated by Nouh and Nurse (2015) that an actor with high eigenvector centrality metric is one that is closest to other actors that have high eigenvector scores too.

A node may have a low-degree centrality and or perhaps weak closeness centrality as well as betweenness centrality but notwithstanding that, it can still be influential in a network. This intuitively implies that, even though a node that is central by one measure is often central by several other measures, this is not necessarily always the case. Wu, Carleton and Davies, (2014) that eigenvalue centrality score of an suggest that actor is placed at strategic position in the network and therefore can be an influential person to his/her neighbouring actors as well as being the focal player in the network. According to Hansen, Shneiderman and Smith (2011) a node connected to a few number of other nodes who are ranked highly with respect with eigenvector scores could himself register a very high eigenvector value.

2.4.5 PageRank

PageRank is a quality metric measure which identifies key actors in a network by determining its importance based on the number of in-coming connections to the node (Kwak, Lee, Park & Moon, 2010). In essence, PageRank is said to be an algorithm of link analysis that gives number weights to each actor with an aim of computing and evaluating their significance in the network. PageRank is a regarded as a centrality metric because it combines both direct and indirect connections between the nodes of the network (Heidemann, Klier & Probst, 2010). The main concept regarding PageRank is to permit the spread of influence amongst the nodes. Wang *et al* (2013) underscored that PageRank regards users as website whereas connections between nodes is considered as links. The significance of employing PageRank metric is because it considers not only the quantity of ties between actors as it is in Eigenvector centrality, but it also takes into consideration the quality of such ties between actors (Nouh & Nurse, 2015).

2.4.6 Social Network Density

Wassermann and Faust (1995) articulated that density is the common position of connection point between the social network actors. It is termed as the number of edges in a part of a social network to the highest number of edges that hypothetically constitute the social network. Thus network density is a characteristic of a network as a whole. Formally it is defined as the total number of ties within a network divided by the total possible number of ties.

Network density is a useful property that is used to assess the general potency or power of activities and interactions within that network. Its formulation can be given as follows:

$$\textbf{Density}_{\textbf{(network)}} = \frac{\textbf{t}}{\textbf{t(t-1)}}$$

Source: Faust(2006).

Where **t** is the number of *ties* between nodes of a given social network

Tsvetovat and Kouznetsov (2011) defines density as connections present in a social network possibly to all other connections in the network and shows the rate at which information streams from or between the actors. This was corroborated by Faust (2006) who succinctly summed up the definition of density of a network that it is proportional to the probable number of connections in that network. Network density is proportional to the probable number of connections in that network (Faust, 2006). This concurs with Wölfer, Faber and Hewstone (2015) which expressed that network density mirrors the general connections in a network by associating the current number of links against the theoretical probable figure of connections between the overall members of the network.

.

Density metrics helps to predict the flow of information between nodes of a given network and it indicates homogeneousness of the community and nodes' interactions with one another (Martino & Spoto, 2006). Hansen, Shneiderman and Smith (2011) summarizes network density as a cumulative measure that describes interrelationships of nodes and that it is a quantifiable way of capturing cohesive concepts of sociology and network affiliation. The authors concluded that the essence of network densities is to make a comparative analysis of communities in order to establish the communities that are either more densely interlinked or sparsely connected. According to Campbell, Dagli and Weinstein (2013) a community refers a sub network cluster of actors that have strong ties within that cluster and weak ties across clusters of a network.

In Waskiewicz (2012) majority of community detection algorithms for determining and depicting the density between actors helps to identify individual nodes with high degree centrality scores. The speed at which information flows within a network is dictated by the density of that network (Himelboim, Smith, Rainie, Shneiderman & Espina, 2017). According to Everton (2008), research has shown that the density of a network is strongly correlated to the possibility that nodes inside a particular network will go along with the established standards and conduct hence forming an ethical order that is greatly linked in social networks. Granovetter (2005) expounded this concept by indicating that in a network with high density, it is easy for the nodes to monitor the behavioural patterns of each other and hinder them from indulging in unusual demeanour. Faust (2006) concurred that networks or subgroups of network with high density are more likely to establish connections or relations that they do not want to let go. When the density of a network is sparsely formulated, it is bound be deficient of social relationships that will hinder the members from being unruly.

2.5 Demographic and Related Information

Nearly every social media site has some profile page for its users, and that page has some essential demographic information such as age, gender, location, and a short personal description. Some sites have very long personal profiles, while others have very brief personal profiles. With regards to pseudo-accounts, Robbins (2012) pointed out that not all social media platforms are strict that a person’s real name or handle reflect their real name. In such circumstances, he advises that there is need for law enforcement to gather more evidence that supports their claims that an account provides useful verification in solving a case and is owned and operated by the individual in question. In order to have a concrete proof against an individual under investigation, law enforcement officers should collect enough evidence that can really confirm that the social media account in various platforms belongs to the suspect (Robbins, 2012).

People can use social media to reveal their online identity such as sharing life events or their lifestyles. This way, they leave crucial traces about their personality, friends, activities they like, patterns of behaviour and actions. For instance, on Facebook, one is likely to find a user's current location and a list of all the other places they have lived and other demographic information such as education history, work history, contact information, family members, political preferences, relationship status, religion, and more. On the other hand, the microblogging site Twitter has very limited demographic information. It can be a name, city, and a short self-written biography. That makes social media a powerful tool for investigators (Sparrow, 1991). Nodes whose attributes such as date of birth, tribe or religious background are similar are focal in creating network relations and co-perpetration (Nouh & Nurse, 2015).

Fundamentally therefore, individuals and groups can now be easily identified using demographic information consisting of names, dates of birth, phone numbers, relationships status, level of education and other crucial information even if they use pseudonyms or aliases (Flynn, 2002). Rice and Parkin (2016) recommends that inaccessible social media user accounts can logged onto by obtaining a court order. Terms of Service (TOS) of most social media platforms prohibit individuals who creates fake accounts or impersonate other users. Despite such rule, there is no effective means to check whether one’s profile is fake or not, unless a complaint is lodged by someone to a particular social media platform (Robbins, 2011).

2.5.1 Social Links and Acquaintances

Social media profiles can be a crucial tool for identifying a person's friends, family members, and acquaintances. Most sites support the creation of explicit social connections with other people which come in two forms either as *friending* which implies a mutual relationship that requires both parties to acknowledge the relationship and the second type of connection is *following* which can be a one-way relationship. A person follows someone on social media when they are interested in what that person posts. The person being followed does not have to approve the relationship in most cases, nor is there a requirement or expectation that the follow is reciprocated. Certainly, two people may choose to follow one another, but unlike “friending,” it is not required.

A person's social connections, regardless of how they are created, are often visible on social media. Usually, a list of friends or followers is linked from a user's profile. This list tends to have a profile photo and name for the person. Besides social connections, other crucial information such as photos or *likes* or *comments* posted on social media sites that allow such can reveal much about the person being investigated as well as his/her close associates. According to Miller (2011), study on social media interactive patterns shows that phatic gestures such as *likes, pokes* and *comments* on status updates gives a means of some evidential online presence, even when those communications tend to be lean in the amount of information they offer. Ego-centric networks are relevant vis-à-vis social media because they comprise of social units such as individuals or groups and embedded social relations such as family or friends and assist an investigator to comprehend the structure, closeness, and density of individual-to- individual or group-to-group social worlds (Freeman, 2004).

2.5.2 Mapping and Time Stamping Location Data.

Hansen, Shneiderman and Smith (2011) underscored that there are innumerable ways users of social media and social networks leave behind traces of their identities that form an intricate web connecting them with other people, locations, and digital objects around them.

According to Fusco *et al* (2010) technology has made it possible to develop geographic profiles of individuals by tracking and aggregating their location information and depicting it on map. Location-based social network applications such as GeoSocial Footprint or Google latitudes, Loopt, and Brightkite, employ technology that pinpoint an individual’s location and tracks his or her movement. It is worth mentioning that digital cameras and smartphones automatically geotag pictures with the exact locations where the pictures were taken.

Zambri (2015) concurs that when law enforcement agencies are armed with these profiles and using the appropriate tools, they are likely to trace social media user’s geographic footprint and potentially predict future movements. For instance, investigators can map out the time when an individual is at home or not, locations the person frequents and times the he/she frequents them, preferred types of food, and games he/ she plays. Buccafurri, Fotia and Lax (2013) advices that if a suspected person claims to have been in some place when a particular took place, the investigators can harvest information from his social media accounts and analyse the location of the device at the time of the incident to be used as an alibi.

It is important to note that several social media sites supports geotagging which is a technology that allows one to associate Geographical Positioning System (GPS) coordinates or other location data with their posts. Such geotagged information will help to verify and fix a suspect in arresting or arraigning him/her in court of law (Vicente, Freni, Bettini & Jensen, 2011). Geotagging can help in corroborating alibis or linking a person to a criminality scene. There is so much that can be done with location information. It can tell you where a person claimed to be, when, and what they were doing there. It can also help you establish patterns of behaviour. For instance, with an individual's Twitter username, information can be collected about GPS coordinates form every available post and use it to plot on a map to depict where the person had visited (Hanson, 2011). Some uploaded or tagged photographs associated to a particular node can give hints on what that individual loves, places he has visited, people he has been with and activities they did together (Vicente *et al*, 2011).

2.5.3 Behavioural Patterns

Besides collecting and analysing demographic information, investigations can still be carried out to discover behavioural patterns of social media users. Behaviour pattern information can be obtained by scrutinizing and investigating what people do online, when and with whom. For instance, a series of posted photos can be analysed to reveal the conduct and activities of people online as well as how they interact with others.

Furthermore an individual's personal and behavioural preference such as health status, favourite movies, online gaming systems and favourite online games, reading preference, locations frequented and even military or technical experience can also play a big in deducing one's behavioural patterns (Zambri, 2015). According to Vercellone-Smith, Jablokow and

Friedel (2012), one's descriptions online social media profile account as well as provision of location related information can be plotted on a map to disclose more about the behaviour of a particular actor. With regards to evaluating and carrying out mental assessment status of some suspects, De Choudhury, Gamon, Counts and Horvitz (2013) proposed the use of information they posted on their social media accounts for crucial leads and conclusion on their mental soundness.

Moreover, biographical descriptions from one's profile and location data can be depicted on a map to divulge more about the demeanour of social media users. People are fond of posting their daily activities unknowingly reveal their behaviour and movements to others notably the investigators (Bradbury, 2011). In situations, where people use pseudonyms or alias, stylometric techniques can be applied to identify the author based on the characteristics of the textual content (Vercellone-Smith, Jablokow & Friedel, 2012).

Zambri (2015) concurs that uploaded and tagged images associated with social media and social networking outlets document what users look like, places they have been, and things they do. An image also connects those pictures together and connects the people in the image with the user who uploaded the image. Fraser (2008) summarized that besides use of logs of activities, preferences and favourites, timeline matching which entails the timestamps can be used to match the timelines of different users, and to create an exact timeline for an entire cluster of friends or even a larger group which indicate where a person was and when can invaluable lead for investigators in gathering crucial forensic evidence. Klerks (2001) advices investigators to target actors in the network with a particular expertise in a given discipline.

2.5.4 Shared and Posted Information

The content of people's posts such as the text they write, what it says or the content of their photos and videos, and the ratings they assign can be a crucial lead investigators can find on social media. Thus the content of the posts alone, where people detail their thoughts, feelings and ideas reveals what they are doing, what they care about, who they interact with, and why. The contents shared or posted on social media user accounts can be used by investigator to detect crime-related behaviour, help in singling out the eye-witnesses, confirming alibi, presenting evidential proof in an investigation and can be utilized during court proceeding to confirm or disapprove witness (Rice & Parkin, 2016).

By looking at the content of the posts people are making, one can uncover a lot of leading intelligence information about their actions. This coincided with Zambri (2015) findings that it is possible for an individual to reveal personal information or interests by association, a social profile signature, rather than directly posting or establishing a social media or social networking profile or footprint. Robbins (2011) commended that most online users are nonsensical whenever they post their items online because they leave tracks which can seriously incriminate them. This susceptibility affects not only the common populace but also the other law enforcement officers.

Rice and Parkin (2016) gave an exemplar of a case in which detectives in USA apprehended a teenager who posted information with the intention of joining some radical Islam in the Middle East. More particularly, was a scenario in which a policeman was killed in New York City immediately after the criminal posted what he had intended to do on social media account. Sometimes, unscrupulous social media user can post information online that act as an harbinger of his/her intention, which if captured on time by investigators, they can intercept and prevent the act from taking place. In a rejoinder, Blomberg (2012) underscores that by harvesting and analysing data from social media accounts of individuals, investigator can detect signs and perceive any atrocious activity that is about to be committed. Nouh and Nurse (2015) however cautions that a node that actively post the most contents on online social media platform may not be essentially the most influential in the network.

2.6 Empirical Literature Review

In their study of analysing and visualizing criminal network networks, Xu, Marshall, Kaza and Chen (2004), applied cluster analysis in quest of detecting network sugroups as well as detect the central nodes and interaction patterns between the communities. Their findings established the general network structures which can provide lead information for investigators. The study however failed to report the metrics or visualizations that can be used by detectives to zero in to the suspects.

Basu (2005) employed Social Network Analysis to generate a linkage terror map in India. The author employed only the centrality and betweenness indices to unearth the influential nodes in the terror network. The shortcomings of this study however, is that it did not employ other centrality metric scores in identifying the key actors in the network.

L'huillier, Ríos, Alvarez and Aguilera (2010) employed a social network analysis methodology to carry out a topic –based text mining in quest of pinpointing important or influential members in virtual network communities. The research yielded an insight comprehension findings that helped law enforcement community in coming up with a counter-terror framework. The study however did no employ the visualizations and centrality metrics to isolate individual influential nodes in a network.

Perliger and Pedahzur (2011) underscored that whenever Social network analysis is employed, we evaluate the investigated social phenomena as a product of a (social) framework, which includes actors (groups) and ties (relations or interactions between actors). Hence, in contrast to other quantitative methods which tend to focus on the description and aggregate analysis of the attributes of those actors who make up the research population, Social network analysis assumes that in order to comprehend the social phenomenon, it is more conducive to map out and analyse the system of ties among the various actors and the ways in which these relational patterns shape actors' activity, decision making and group dynamics and eventually, the outcome of the group's collective action.

In their study of attempting to identify the most influential actors in a network, Ilyas and Radha (2011) used the technique of principal component analysis in comparison with eigenvector centrality metrics. Although principal component analysis is a new technique of measuring the centrality of a node, it failed to pinpoint individual nodes that are influential in that network.

Acquisti, Gross and Stutzman (2011) did an experiment that involved aggregating the anonymous network with another network by recognizing nodes that correspond to the same individual. In a similar experiment at Carnegie Mellon University, using publically available online social network data, researchers were able to overlay virtual information with a person's real identity to break down anonymity layers. In their hybrid study approach, Huber *et al* (2011) employed a web crawler to harvest social network data from users as snapshots. However, the study failed to clearly show the centrality scores for each actor in the network and their respective visualizations of communication channels.

Mincera and Niewiadomska-Szynkiewicza (2012) undertook a simulated experimental research in quest of applying social network analysis to investigate interpersonal connections. They performed multiple experiments using data acquired from widely used social network platforms and simulated the findings by comparing two social network algorithms namely Clauset & Newman and Blondel Algorithms. More specifically, the experiments were performed for data about interpersonal connections, acquired from two commonly used platforms Facebook and Twitter. The results discovered that both platforms are typical social networks, that is, scale-free and small-world networks. The study gave a new perspective of how to investigate interpersonal connections on social media platforms but it failed to highlight the network metrics of main actors in the network.

In quest of demonstrating how to detect network communities, Staudt, Marrakchi and Meyerhenke (2014) underscored that unearthing network community structures reveals remarkable realistic patterns of the real world interactions. According their study, detection of a community starts with a modest number of seed actors as input which ultimately generates other actors revolving around and communicating with the seed node. In Resonance, Yang, Liu and Sageman (2006) advises that detecting and discovering network subgroups helps to comprehend the function of each network community and that identifying groups such as this assists a detective to quickly unearth the associated criminals when a small number of suspicious characters are known. According to Xu, Marshall, Kaza, and Chen (2004), a highly dense group is more susceptible and exposed to law enforcement officers for further scrutiny and identification of the main actors who are most likely to be the leaders of a particular cohort of felons. They concluded that if the density of various groups keeps on fluctuating from high to low or vice versa, then it implies that the groups are competing for leadership positions.

Rice and Parkin (2016) highlighted a scenario in which detectives utilized the posted information to apprehend and charge the nodes under investigation, in USA. They underscored that sometimes, unscrupulous social media user can post information online that act as a harbinger of his/her intention, which if captured on time by investigators; they can intercept and prevent the act from taking place. Such information cannot only be used for identification the suspect and or witness(s), but also they can be used to substantiate alibi or used as evidence in court of law. In a terror network, investigators usually go for actors with the highest degree centrality scores because they are the most connected and possibly the most influential nodes in the entire network. One can easily identify a subgroup of a network communities that particular belongs by visually assessing the links amongst network actors (Wu, Carleton & Davies, 2014).

Gunnell, Hillier and Blakeborough (2016) conducted an examination on how to apply Social Network Analysis techniques by utilizing the available police intelligence data, as a tool to more systematically understand gangs and to help direct law enforcement activities. Using five individuals chosen as focus group, the objectives of study was undertaken to unravel what Social Network Analysis can reveal about criminals and establish how useful the social network analysis outputs were to the police. The research findings revealed an overall network of 137 individuals were identified, from the starting point of five (5) individuals identified as having gang links. The relatively large network identified contained only a small number of people explicitly linked together by the police as being in a gang. This demonstrated the importance of understanding how other kinds of connections inside a social network may work and how aversion meant to interrupt a criminal network might be more useful if they take these relationships into consideration.

2.7 Theoretical Framework

Few network theories was highlighted in order to know how scholars explain why nodes establish, keep, break and probably re-establish network relations and who is probable to establish relations with who. Network theory's development in the twentieth century occurred in parallel with developments in intelligence analysis. The roots of Social Network Analysis (the social branch of network theory) are usually traced to Jacob Moreno, whose publication of *Who shall survive*? (1934) was "a signal event in the history of social network analysis" Borgatti *et al* (2009).

All these early studies somehow neglected the importance of network visualization and network metrics, stressing aspects related more to statistical network characterization, or interpretation of individuals' roles rooted in social theory. Kadushin (2004) argues that social network theory is the only social science theory that is non-reductionist because it can be applied to numerous types of network sizes. De Nooy, Mrvar and Batagelj (2005) opined that Social Network Analysis is both confirmatory and exploratory. In particular, exploratory analysis entails use of quantitative metrics and visualizations while confirmatory analysis involves the test of hypothesis. Social network analysis is fundamentally a multi-theoretical approach. It is a unique methodology since it is mostly concerned in the connection between network nodes.

According to Raab and Milward (2003) Social Network Analysis is a set of techniques and theories the gives pragmatic information as far as social structures are concerned. It has been fruitfully applied to various scenarios where social relations are involved. Social network analysis has its mathematical origin in graph theory and matrix algebra however it has been applied to various disciplines.

According to Jamali and Abolhassani (2006) social networks open up a whole new world of information, because at least as much value is contained in the relationships between entities as in the entities themselves. This concurred with Raab and Milward (2003) social network analysis, then, is a collection of theories and techniques that provide empirical content to social context which explain varieties of behaviour in terms of constraints and options that are inherent in the way social relations are organized. Borrowing from the sociology theory, Waskiewicz (2012) stated that in the ego-centric friend of a friend dichotomous relations and influence can be felt up to 3 degrees. Hanneman and Riddle (2005) postulates that the boundaries of a given network are bound for expansion from 1 degree ego-network around a single actor whenever ego selects or connect those that are deemed to be members of that network and this ultimately diffuses to a higher degree ego-network.

Wasserman and Faust (1994) made a rejoinder that the procedure of scaling and rhyming the social network setup is summarized using twofold vital procedural resolutions. The initial methodology gives answers to the question of which nodes to incorporate whereas the subsequent methodology or procedure deals with establishing the way connections between nodes should be measured or classified. Two necessary methods are general in Social Network Analysis examinations for unravelling the initial problem. The initial one is centred on the node's oneself of the limits of the network cluster. Those that are regarded by the associates as fitting to the cluster are incorporated. Although this method can be useful in situations in which the research worker has access to the clusters affiliates, it can be challenging when handling undercover networks.

The largely general methods for assessing the node’s responsibility or influence inside the social network are measures of importance (loosely similar to centrality measures). A number of the important appropriate measures, which can be helpful in the learning of brutal clusters, are degree of centrality, closeness and betweenness (Perliger & Pedahzur, 2011). An analogous measure was initiated by Brams *et al* (2006) who came up with the idea of power as a function of the node significance within the social network which consequently is established by the extent of their linkages and direction comparatively to other nodes).

Brams *et al* (2006) gave a more complex method utilizing directional links in a bid to unravel the patterns of information within the social network and ultimately the level of power of every node on other nodes. Seigel (2009) underscored that fundamentals features of the social network are connected with the group results, inside social means and procedure as well as members’ conduct such as the choice of whether or not to take part in its actions.

A resounding observation was made by Koschade (2006). He summarized that the Social Network Analysis literature gives an array of means and ways which facilitate the investigator to unravel these attributes. Firstly are measures that offer data on the level of unity and degree chain of command within the social network such as group degree centrality, density and inclusiveness. These measures are crucial for comprehending the inside influence and workings within the social network, patterns of interaction, and the kind of decision-making process within similar clusters. They are also vital in proofing some fundamental premises concerning the attributes of covert social networks.

Generally, the majority of social networks ought to find an equilibrium point linking competence and robustness/survivability (high density, high number of redundant ties). Somewhat paradoxically, however, in the case of clandestine network it seems that high numbers of redundant ties lessens the chances of network survival. Clandestine networks are interested in secrecy, hence high levels of density and group centrality increase the chances that the group is exposed. Then again, high density facilitates effective indoctrination, a crucial element in the radicalization process of the network.

As showed by Krebs (2002), one of the ways to bypass these contradicting needs of the clandestine network is by deactivating strong ties while the network operates in hostile environments. In this case the density of the network is being lowered when the network becomes active and prepares to act. Another important set of tools refers to the uncovering of internal cohesive subgroups within the whole network. The theoretical importance of such subgroups stems from the causality found in various political and sociological studies between cohesiveness and the tendency for group uniformity, intensive socialization and radicalization. The concept most often used in this context is clique - when each actor in the subgroup is tied to all other actors and there are no actors who are tied to all the clique members. Duijn (2016) defines a clique as a subgroup of a network in which nodes are tightly knitted and connected together.

However, sometimes there are less cohesive subgroups in the network; hence, they do not fit the definition of a clique. That is, in order to identify these other types of subgroups, other methodological concepts are often used, such as n-cliques (n geodesic distance between members) or k-core groups (every actor has ties to at least K actors within the subgroup).

According to Duijn (2016) such a geodesic distance scores between actors implies that information travels swiftly and it also indicates that largely, information goes via the central nodes so as be availed to other nodes in the network. Identifying the subgroups allows us to detect different functions of the network (founders, collaborators, passers-by), network recruitments paths, operational characteristics and patterns of flow of information. Moreover, by looking at the attributes of the subgroups, we can evaluate ideological homogeneity and level of solidarity within the network and how this influences the activities and development of the unlawful network (Bonacich, 1972).

Social Network Analysis could be of high value for understanding the relations between different target groups worldwide especially by using theories of structural balance (which assume that actors will forge ties in cases of sheered interests such as positive ties to other third actor), homophily and hetrophily (Freeman,2004). Making use of Social Network Analysis in order to add a social dimension to the basic socio-demographic profile by looking at their patterns of social interactions, as well as by distinguishing between different roles within the network, which in turn are reflected in different profiles (Gunnell, Hillier & Blakeborough, 2016).

Psychological theories focusing on social learning process could also benefit from the use of social network analysis focusing on ego-networks. Naturally, the growing literature emphasizing the role of social processes within small informal groups and involvement in violent activities is well suited to the use of Social Network Analysis (Klandermans, Bert & Oegema, 1987).

In summation Wasserman and Faust (1994) outlines that Social Network Analysis is both theoretical and methodological in itself. Theoretically, Social Network Analysis assumes the rules of sociology that all nodes are arranged and affected by greater social networks. Methodologically, Social Network Analysis provides graphs and arithmetic techniques that be used to map and measure relationship patterns between actors in a social network.

2.7.1 Application of SNA and Security Issues in Social Media

There are an assortment of tools are employed on social media to generate a colossal quantity of information which is subsequently subjected to analysis to better unravel the patterns of the individuals, groups and societies that utilize them. More specifically, they create relational data: information about who knows, or is friends with, whom; who talks to whom; who hangs out in the same places; and who enjoys the same things (Hansen, Shneiderman & Smith, 2011). Social media sites have emerged as a method for instantaneous connection among people and groups; information obtained from these sites can also be a valuable resource for law enforcement in the prevention, identification, investigation, and prosecution of crimes (Global Justice, 2013).

The pervasiveness of both social media and social networking and the high pace with which it keeps on advancing cannot be played down. It is a societal and technical trend that permeates virtually all facets of human every day existence. Therefore it is important to emphasize the integration of social media and social networking in criminal activities and the sophisticated threat that they pose (Zambri, 2015). In the age of information prevalence, individual data cross over an astounding array of computer systems and networks. Moreover, there is also increasing security challenges and risks to the users of social media. Most of these threats linked with social networking are privacy concerns and spreading of false information (Punjabi, 2014).

According to Hudaib (2014) majority of Facebook account holders have been known to easily consent to friendship invitations from other strange users just because each has numerous friends in a circle. Whenever they accept these requests from friends of friends, users unsuspectingly reveal their confidential data to unfamiliar persons. Upon obtaining the user's data, strangers can harmfully utilize the aforesaid data both in the World Wide Web and in the real world. These dangers rise when the users happen to young people who are by nature more predisposed and susceptible than mature people. Dinerman (2011) affirms that as long we continue using online social media and increasingly get entrenched into the daily lives of users, private data will be prone to exposure and abuse.

Despite the fact that the application of social media and Social Network Analysis is still undeveloped in law enforcement landscape, it is quickly gaining popularity considering that many users are utilizing variety of social media platforms. For example, in the Los Angeles, investigators use social media and social networks to single out and eventually arrest sex predators, drug barons, gangsters, thieves and criminals of different types heinous crimes committed over the internet. Crime investigators in the children department have utilized social media and social networking to instigate covertly cyber operations, using non-existent personalities and parody accounts, to pretend as young children in order to isolate, trail, and finally apprehend risky sex predators. Narcotics detectives have also used the same technique to identify and contact drug dealers, set up drug transactions, conduct a "buy and bust" operation, and ultimately have used the open source information as evidence in subsequent court proceedings (Zambri, 2015).

Gupta and Brooks (2015) underscored that security analysts should know how to use social media to improve security locally, nationally and globally in a cost-effective manner. Social media data can be analysed to map the social networks of various types of offenders. In the cyber world all information is interconnected and individuals can be located via indirect non-linear links. Social media, social networks and Social Network Analysis techniques are just as accessible to criminals and criminal organizations as they are to police. Law enforcement has embraced the use of social media and social networking in a number of areas, including recovering evidence, locating and apprehending suspects, conducting intelligence collections using social networking to conduct crime analysis and intelligence trend analysis (Phillips, Nurse, Goldsmith & Creese, 2015). As a tool for analysis, Social network analysis regards nodes and actions as interdependent entities, recognizes that connections between nodes provide channels for transfer of information between the actors.

Numerous scholars researching on security issues have applied Social Network Analysis and social network theories to analyse variety of scenarios notably the terrorist communication. Borgatti *et al* (2009) note that of all the disciplines that now incorporate network theory, security is probably the field that incorporate it the most. For instance, soon after the invasion of Afghanistan, studies using Social Network Analysis to map terrorist networks emerged. Krebs (2002) initiated the application of Social Network Analysis to map the terrorist networks by applying its centrality measures mentioned elsewhere in this study. Similarly, Koschade (2007) used open-source software to depict the ties between the USA September 2011 hijackers, in which the results showed that Mohammed Atta was the mastermind and instigator of the conspiracy. Rodriguez (2005) mapped the network responsible for the March 2004 Madrid bombings, while Nagl *et al* (2008) gave crucial recommendations on how to apply Social Network Analysis for military intelligence investigation, where he praised Social Network Analysis as a tool for understanding the organizational dynamics of an insurgency and how best to exploit it.

2.8 Conceptual Framework

Golbeck (2013) provides a framework for the analysis of public data currently available and being generated by social networks and social media, like Facebook, Twitter, and Foursquare. Access and analysis of this public data about people and their connections to one another allows for new applications of traditional social network analysis techniques that let us identify things like who are the most important or influential people in a network, how things will spread through the network, and the nature of peoples' relationships. As a tool for analysis, Social network analysis regards nodes and actions as interdependent entities, recognizes that connections between nodes provide channels for transfer of information between the actors.

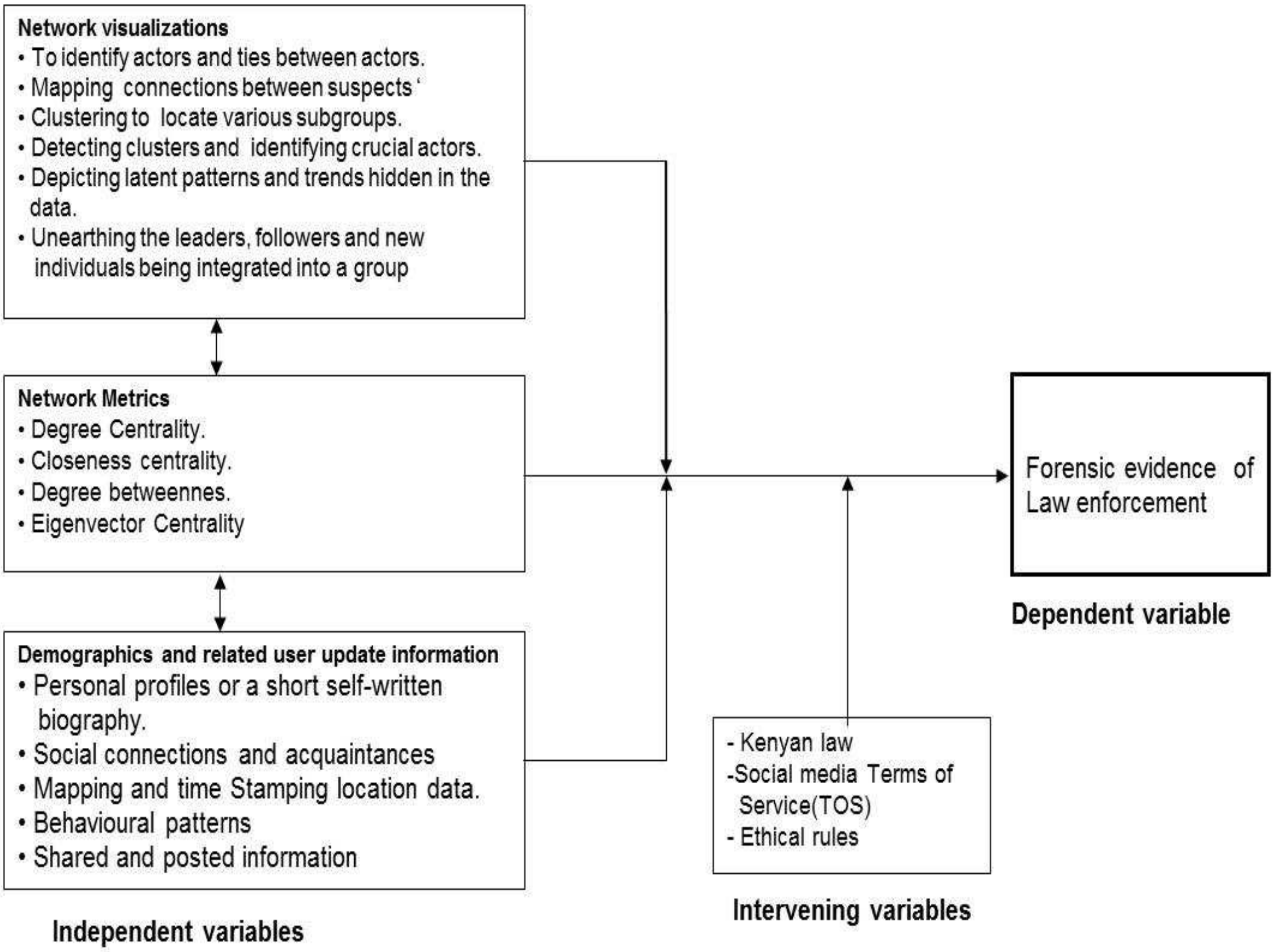
2.8.1 Research Gap

In essence, the aforementioned research studies demonstrates the possible application of the Social Network Analysis technique in building a more nuanced and a methodical comprehension of a neighbourhood crime nuisance using detectives or intelligence database which can be employed or augment police investigative activities. This could also be employed to other similar kind of crimes in which social networks are readily available. The technique gave crucial information about persons most likely endangered from criminal association. The added value of using intelligence data, rather than just focusing on crime data, was highlighted by the wealth of non-criminal links in the overall network. This shows how important it is to understand these non-criminal links if disruption activities are to work, and to identify people at risk of gang association to enable preventative policing. However, the limitations of this study is that the exactness and understanding of detectives or intelligence data is indeterminate as well as the geographic location of individuals in the network was not included.

Nevertheless, the aforementioned previous studies in social network analysis on social media emanates to from developed countries. In essence, their studies are based on countries with established cyber security legal framework with advanced technologies. There is a conspicuous failure by Kenya's detectives to adopt Social Network Analysis methodologies or techniques to investigate crimes committed over the internet, notably the social network platforms. Therefore, this study was carried out not only to fill this gap of knowledge but also to demonstrate how law enforcement agencies can successfully adopt this technique to their advantage

As depicted in Figure 2.3, three key variables namely network visualizations, network metrics and user's profile/demographic and related information are the key attributes that can aid an investigator to systematically establish presence or lack of criminality activities of the individual(s) under investigation. It is important to note that the findings of network infographic visuals are consistent with computed values of network centrality metrics. Demographic and related information only act as further levels of investigation after isolating suspicious online characters. In order to arrive at a sound conclusion, it is advisable to employ a combination of different social network analysis centrality metrics (Williamson & Ruming, 2015).

The study was informed by the conceptual framework shown in Figure 2.2 below.



Network visualizations
• To identify actors and ties between actors.
• Mapping connections between suspects '
• Clustering to locate various subgroups.
• Detecting clusters and identifying crucial actors.
• Depicting latent patterns and trends hidden in the data.
• Unearthing the leaders, followers and new individuals being integrated into a group
Network Metrics
• Degree Centrality.
• Closeness centrality.
• Degree betweennes.
• Eigenvector Centrality
Demographics and related user update information
• Personal profiles or a short self-written biography.
• Social connections and acquaintances
• Mapping and time Stamping location data.
• Behavioural patterns
• Shared and posted information
Independent variables
- Kenyan law
-Social media Terms of Service(TOS)
- Ethical rules
Intervening variables
Forensic evidence of Law enforcement
Dependent variable

Figure 2.3: Conceptual Framework.
Source: *Researcher*

With regards to intervening variables, ethical rules were observed by obtaining consent from the respondents and permit from National Commission for Science, Technology and Innovation (NACOSTI). However, the researcher had no control over specific social media's terms of service. Terms of service for the social media platforms employed in this study was obeyed mutually when creating accounts as well as during mining of data from the respondents.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This chapter presents a detailed description of procedures that was used to answer the research specific objectives. Emphasis is placed on the methodology which includes research design, sampling techniques, data collection procedures and instrumentation and data analysis techniques. It highlights the how the pilot study was undertaken and its outcome as well as ethical considerations adopted in this study.

3.2 Research Design

Social Network Analysis experimental research design was employed in this study. Initially, selected respondents treated as focus groups were subjected to a brief interview and thereafter persuaded to create pseudo-online accounts in specified social media platforms which were used to perform online mining of the selected respondents to obtain data that ultimately aided in social network analysis. Social network analysis is the study of the social structure known as social networks comprised of individuals and their relationships. Social Network Analysis employs mathematical and graphical techniques that utilizes online relations between nodes to map out their individual roles in the entire network and ultimately present those who are highly connected or and more vital in the network.

A social network can consist of the relationship between two people or of the relationships between everyone globally. Because social media is all about creating and sustaining social networks and relationships between people, understanding Social Network Analysis is essential to understanding social media. Social Network Analysis enables you to map, measure, and describe almost anything about a social network and its components. Social

Network Analysis can provide information about individuals, a few relationships, or large-scale networks.

One can use Social Network Analysis to understand the ideas of interest of social networks, how individuals gain influence in social networks, how individuals form relationships with others, how the relationships evolve over time, and how the relationships affect the behaviour of individuals in the social network. Social Network Analysis is very useful for understanding how law breakers use social media to develop relationships with at-risk populations, forecasting how the social networks of they evolve over time. It is a method that reveals unseen network patterns in an unlawful network and present related information as set of nodes.

Thus, Social Network Analysis has evolved as a popular, standard method for modelling meaningful, often hidden structural relationships in communities. Existing Social Network Analysis tools often involve extensive pre-processing or intensive programming skills that can be challenging. NodeXL, an open-source template for Microsoft Excel, integrates metrics/visualization tools to spark insight of activity of online users on social media and shed light on individual behavior, social relationships, and community efficacy (Bradbury, 2011).

3.3 Location of the Study

The study was carried out in Uasin Gishu, Kericho, Kakamega and Migori Counties. More specifically, the study area which the respondents were drawn was selected using purposive sampling technique. The rationale of choosing the aforementioned study areas was due to financial cost implication and time limit of this research. The power of purposive sampling lies in selecting information which is in depth, rich and related to the central issues being studied (Kombo & Tromp, 2006).

3.4 Population of the Study

Target population refers to the population to be studied to which the investigator wants to generalize his results. In research, it is important that the researcher finds out as much as possible about the study population. It is believed that the greater the diversity and differences that exist in the population the larger the researcher's sample size should be. Capturing the variety in population allows for more reliability of the study (Kombo & Tromp, 2006). Thus putting all these factors into consideration, the population for the study was drawn from five selected universities in Migori, Uasin Gishu and Kakamega counties in Kenya. The study targeted a population of 124 selected students from Rongo University, Moi University (annex – school of law), Kisii university (Eldoret Campus), University of Eldoret and Kibabii University.

3.5 Sampling Procedure and Sample Size

As cited in Passmore (2011), Dunbar (2002) suggested that the typical size of an egocentric network is constrained to about 150 members due to possible limits in the capacity of the human communication channel. The rule arises from cross-cultural studies in sociology and especially anthropology of the maximum size of a village (in modern parlance most reasonably understood as an ecovillage). It is theorized in evolutionary psychology that the number may be some kind of limit of average human ability to recognize members and track emotional facts about all members of a group. However, it may be due to economics and the need to track "free riders", as it may be easier in larger groups to take advantage of the benefits of living in a community without contributing to those benefits.

3.5.1 Sampling Procedure

While developing a sample design a researcher must pay attention to the type of universal sampling unit, size sample, parameters of interest and budgetary constraints (Kothari, 1990). An effective population must also take into consideration representation. It is important for the researcher to identify and select respondents that fulfil the questions that research is addressing. It is important that the majority of the population came from same environment. An effective population is the one that is accessible to the researcher (Kombo & Tromp, 2006). All these factors were put into consideration when the researcher was developing the sample design. Accordingly, this study employed purposive sampling technique to select a representative sample from the respondents in each university which was later subjected to a sampling formula. This was ensuring that each member of the target population has equal and independent chance of being included in the sample.

3.5.2 Sampling Size

In order to obtain the subjects for the sample for the five selected universities, Yamane's formula for calculating the sample size was employed in this study (Yamane, 1967).

$$n = \frac{N}{[1+Ne^2]}$$

where:

n= Sample size

N=Population size

e=Sampling error (0.10 was adopted in this study)

The above formula was chosen because it fits in situations for sampling various groups when one wants to analyse and compare especially when sample frame is wide. Table 3.1 shows how respondents were obtained from each selected university.

Table 3.1 shows how respondents were obtained from each selected university.

Table 3.1: Sample Frame

University Name	**Population size**	**Sample size**
Rongo University 4th Year Students	18*	15
Moi University 3rd Year law students	42*	30
Kisii University (*Eldoret Campus*) 2nd Year Information science Students	18*	15
Kibabii University 2nd Year Criminology Students	24*	19
University of Eldoret 1st Year tourism Students	22*	18
Total	**124**	**97**

Source: *Researcher*

* Population per faculty/school of a given university.

It is important to note that, the choice of selecting respondents pursuing specific degree programmes, was purposefully chosen because the findings and recommendations made by this study is include Social Network Analysis in the curriculum of the aforementioned programmes. This way, criminology students for instance will appreciate, embrace and know how to investigate, mine, analyse and presents forensic evidence of criminalities committed by online users.

3.6 Instrumentation

Initially, the researcher used interview schedule on focus group and later NodeXL software was used to mine data from the respondents' social media accounts. The data was mined three months retrospectively. The management of the five selected universities were notified of the intended research and also assured of confidentiality of the information obtained. Upon consent, the researcher engaged the respondents to create or open social media accounts using the given pseudo-names and passwords.

3.6.1 Pilot Study

The main aim of carrying out a pilot study was to check for the reliability and validity of research instrument. A total of 72 respondents were used in the pilot study.

3.6.2 Validity of the Instrument

Hence in order to evaluate the efficacy of Social Network Analysis approach, interview schedule for focus group was carried out and thereafter the demonstration of NodeXL software to mine social media forensic evidence was carried out. Further validation of the NodeXL instrument was done by consulting experts from Codeplex.com, a social media and network consultants from California, USA.

3.6.3 Reliability of the Instrument

Reliability is a measure of the degree of which a research instrument yields consistent results after repeated trials. Reliability ensures that there is a precision with which data is collected. If the same results are gained time after time, no matter how many times you conduct a piece of research, this suggests that the data collected is reliable (Hesse-Biber, 2010). Hence in order to establish the reliability of the NodeXL software as research instrument, pilot study was carried out a using his workmates social media data. The findings the pilot study were affirmative and supported the effectiveness, reliability and validity of this methodology. An excerpt of pilot study findings are shown in appendices III to V.

3.7 Data Collection Procedure

NodeXL is standard Social Network Analysis software that helps in analysing and outlining a general social network at a glance, detection and further examination. The software employs automated algorithms that iterates on a number of steps commencing from mining data from selected social media site, generates results in form of visualizations and other related metrics. It also lets novel users to speedily produce important network statistics such as degree and cluster metrics as well as use of special visual charts using the Excel spreadsheet environment. Easy to comprehend separating and adaptable display characteristics can be employed to draw attention to crucial structures in networks easily.

NodeXL also assist in the investigation of the selected social media with import properties that mine social network data from a variety of data repositories such as individual email archive or database or from popular social media site such as Twitter or Facebook. Moreover, other NodeXL can also obtain data imported as text, CSV, or GraphML files.

3.8 Data Analysis

The data from the selected popular social media platforms were collected, cleaned and analysed using NodeXL software and findings such network statistics metrics and visualizations generated. Many of these standardized measures were mapped to a variety of network display characteristics. Furthermore, NodeXL query was configured to request an "edge" to describe the connection between social media nodes which is created whenever they follow, reply or mention one another. The Clauset-Newman-Moore clustering algorithm was employed to create sub-groups from the larger population.

An exclusive report in where all the actors in the social network come into view as soon as it is generated too in a separate "vertices". In this context, a "vertex" refers to a node that constitutes a network structure. The network metrics reported a variety of attributes concerning the place and link pattern of very node inside the bigger network. "Degree" indicated the count of all unique connections each user has, whereas in-degree or out-degree reports the number of connections to and from every node. Other metrics such as betweenness, Closeness, Eigenvector Centralities and page rank was also is generated. Nevertheless, clustering coefficient which determines how densely associated every node’s associates are related to each other was depicted by visualizing it in the using other modest sized network graphs that reports only the particular node and the ties in the midst other immediate connections.

3.9 Ethical Considerations

All respondents involved in the study were assured of confidentiality of the information they gave. The researcher duly informed them that information gathered during the study was to be used exclusively for academic purposes only.

CHAPTER FOUR

DATA ANALYSIS, PRESENTATION AND DISCUSSION

4.1 Introduction

This chapter gives an analysis and interpretation of the research findings of the egocentric online social media users. From the initial seed of 94 respondents selected for the study, the social network expanded exponentially to 29295 over the three months of study. The analysis was done with particular reference to the three specific aforementioned research objectives of the study using Social network analysis. The findings were used in the explanation of the results and in suggestion of the recommendations.

As mentioned earlier, the social network analysis done on the data mined from the selected study participants was done using NodeXL Version: 1.0.1.373. It computes statistics for individual vertices which are then used for visualization and also calculates graph metrics that help explains the visualizations results. In essence therefore, NodeXL was chosen because it provides a variety of display preferences to specify the outlook of individual edges (*connections or relationships*) and nodes (*respondents*) and ultimately the entire network layout.

4.2 Unearthing Key Social Network Actors Using Visualizations and Clusters

NodeXL was employed by the researcher in the analysis to visualize and identify subgroups (clusters), generate set of graph metrics using various actors' interactions using either force directed algorithms such as Harel- Koren fast multiscale or Fruchterman –Reingold or using geometrical algorithms such as spiral, vertical, grid, horizontal or circle. Therefore, social networks was depicted differently over time due to structural changes as a result of increase or decrease of network membership

In essence, ties between actors was analysed and visualized using various social network analysis metrics.

The researcher employed the use of visualization and clusters so as to aid in focusing and identifying individual nodes that exhibit important network properties to the rest of the network. Use of network visualizations is crucial because it reveals patterns that are otherwise invisible by other means of analysis or investigations of a network setup.

4.2.1 Egocentric Seed Network

Figure 4.1 depicts the egocentric network of the seed actor whose pseudo-name was "samsonpeter9252" (*this name has been by truncated as "sams~"in this study*) is visualized and positioned at the epicentre of the social network. The seed actor's initial connections to the selected respondents are illustrated by the arrows pointing outwards and inwards the main or seed node. Hence it is a directed graph. The graph's vertices were grouped by cluster using the Clauset-Newman-Moore cluster algorithm while the graph was laid out using the Harel-Koren Fast Multiscale layout algorithm. The findings illustrates that at a 1-degree egocentric network comprising of the initial 55 seed correspondences out of the targeted study sample size of 94 from different selected universities.

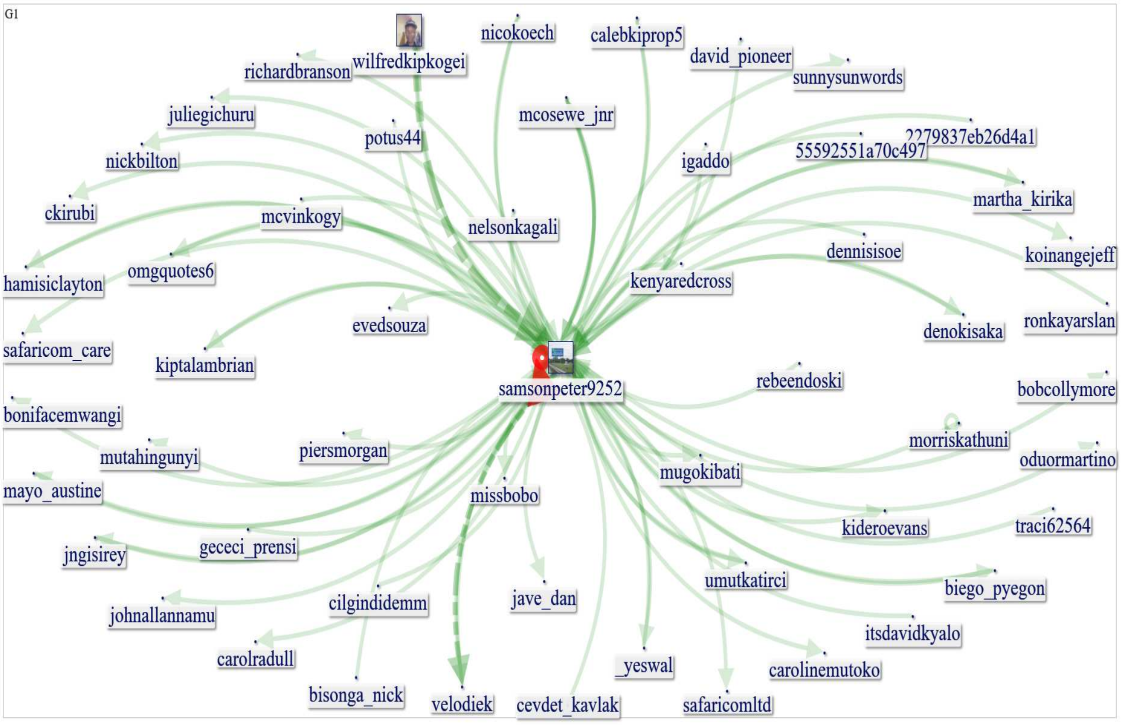


Figure 4.1: Initial Seed 1-Degree Egocentric Directed Network

Source: *Research data (2017)*

4.2.2 Identifying Communities

The visualizations result in Figure 4.2 shows how respondents have regrouped themselves into three distinct communities (labelled C1, C2 and C3). This is an important findings which was achieved by employing the force directed algorithm in the visualization so that the connected actors draw to every other while non-connected actors are separated. This means that the highly linked actors are drawn towards the epicentre of the graph. Evidentially, the findings clearly portrays that in the three communities detected, there is one main influential node in each community. Their pseudo-names are velo~(*velodiek*), wilf~(*wilfredkipkos)* and deno~(*denokisaka*). The results are closely related to study done by Staudt, Marrakchi and Meyerhenke (2014) on detecting communities. In each community, the primary distributors of information were identified. In conformity, Yang, Liu and Sageman (2006) underscores that identifying groups such as this assists a detective to quickly unearth the associated criminals when a small number of suspicious characters are known.

Markedly, in these aforementioned detected communities, respondent velo~ has the highest degree centrality, followed by respondents wilf~ and deno~ in that order. This means that the three actors are not only powerful in the network, but it also shows that have great access to information in their respective communities. These findings are in conformity with that of Wu, Carleton and Davies (2014) which advises that in a terror network, investigators usually go for actors with the highest degree centrality scores because they are the most connected and possibly the most influential nodes in the entire network and that one can easily identify a subgroup of a network communities that particular nodes belongs to by visually assessing the links amongst network actors. In a rejoinder, Perliger and Pedahzur (2011) stressed that discovering the communities enables the investigators to identify the various roles of the nodes in the network such as leadership or brokers and how information flows in the entire network.

Besides examining the roles of several members of the network, investigators ought to concentrate on specific subgroups in order tell their particular duty. Usually, network members come together with an obligation of accomplishing their heinous acts and therefore identifying the subgroups who are interconnected could enhance the chance of detectives comprehending the intention of the entire network. In an investigation scenario, the objective of identifying communities in a suspicious network is to detect their groups and social structures they belong. In Faustand Fitzhugh (2012) Social Network Analysis techniques helps to comprehend network communities by mapping the relations that link them as a network and thereafter determine key players or groups and ties between the nodes.

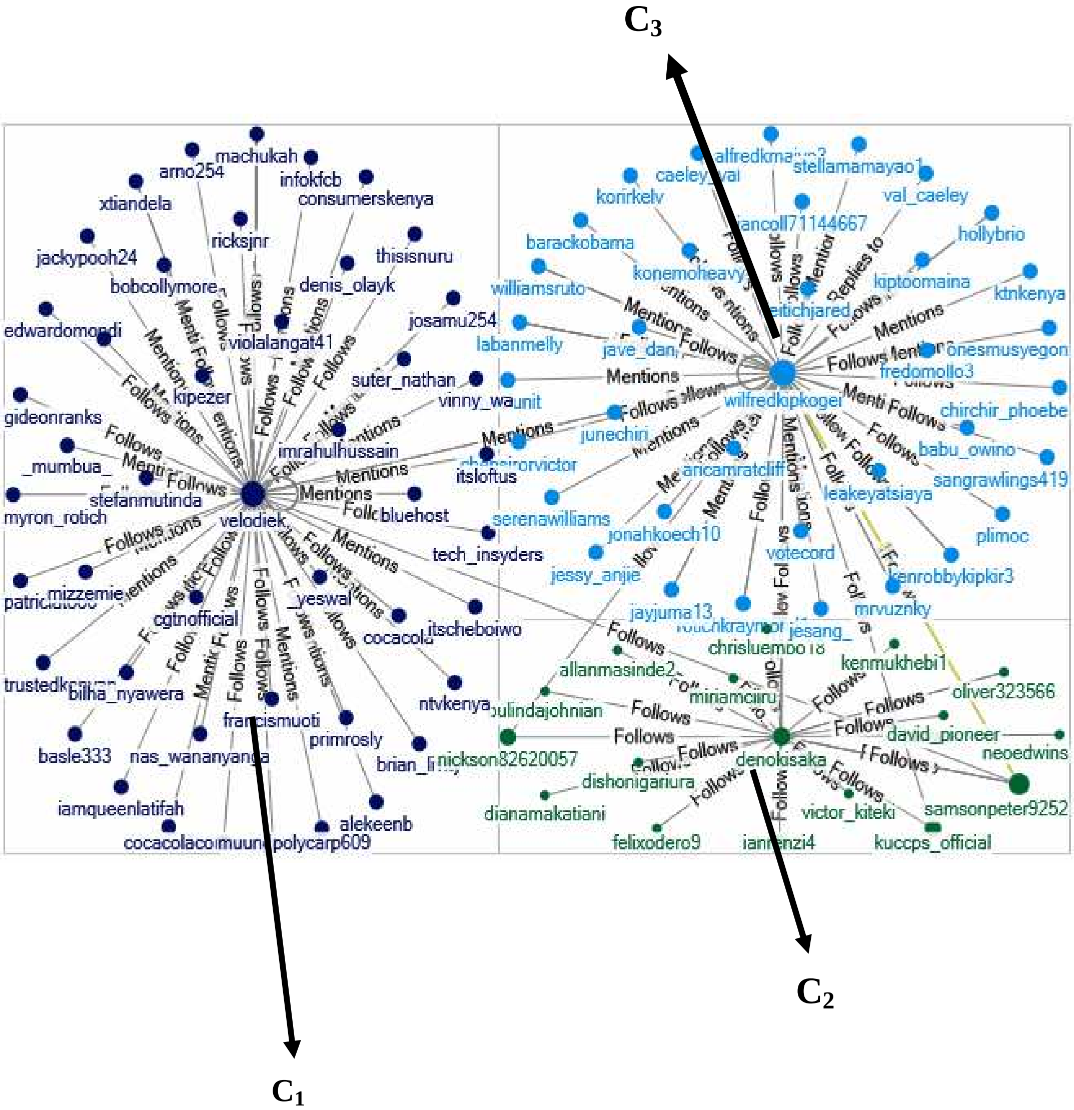


Figure 4.2: Identifying Communities and their Main Actors

Source: *Research data (2017)*

4.2.3 Community Densities

The findings in Figure 4.3 shows varying densities of six communities or subgroups densely connected in one network, each identified by label codings D1 to D6. Necessitated by the need to portray the insight phenomenal of how information visually diffuses amongst the actors in the network, researcher employed the Wakita-Tsurumi algorithm to generate the network densities of the detected communities. A similar approach was employed by Waskiewicz (2012). The research findings indicates that as the size of the network exponentially increased over time, the density of the network expanded but not uniformly for each community.

More specifically, the study results in Figure 4.3 revealed that relationship or links exist between the detected communities. Notably, the densities of D2, D3 and D6 communities are visually more or less equal which implies that members of these three aforementioned communities communicate more frequently about an issue(s) they are all familiar with and probably know one another most. However, the density is slightly higher for the D1 community in the entire network and visibly has a node with the high degree centrality score.

Borrowing from Krebs (2002), a dense cluster having numerous interactions is suspicious and warrants investigations to unearth or discover more information about the group. Last but not least, the D4 and D5 community members are scattered and also depicted the least density in the whole network. These results closely relates to the study done by Staudt, Marrakchi and Meyerhenke (2014) on community detection in quest of revealing structures or patterns of interactions between nodes in a network. Furthermore, the findings conforms with (Hansen, Shneiderman & Smith, 2011) comparative analysis on determining communities that are highly related or sparsely connected. Density metrics helps to predict the flow of information between nodes of a given network and it indicates homogeneousness of a community and nodes' interactions with one another (Martino & Spoto, 2006). By employ the density measure, detectives are able to a holistic understanding structure of the entire network under scrutiny.

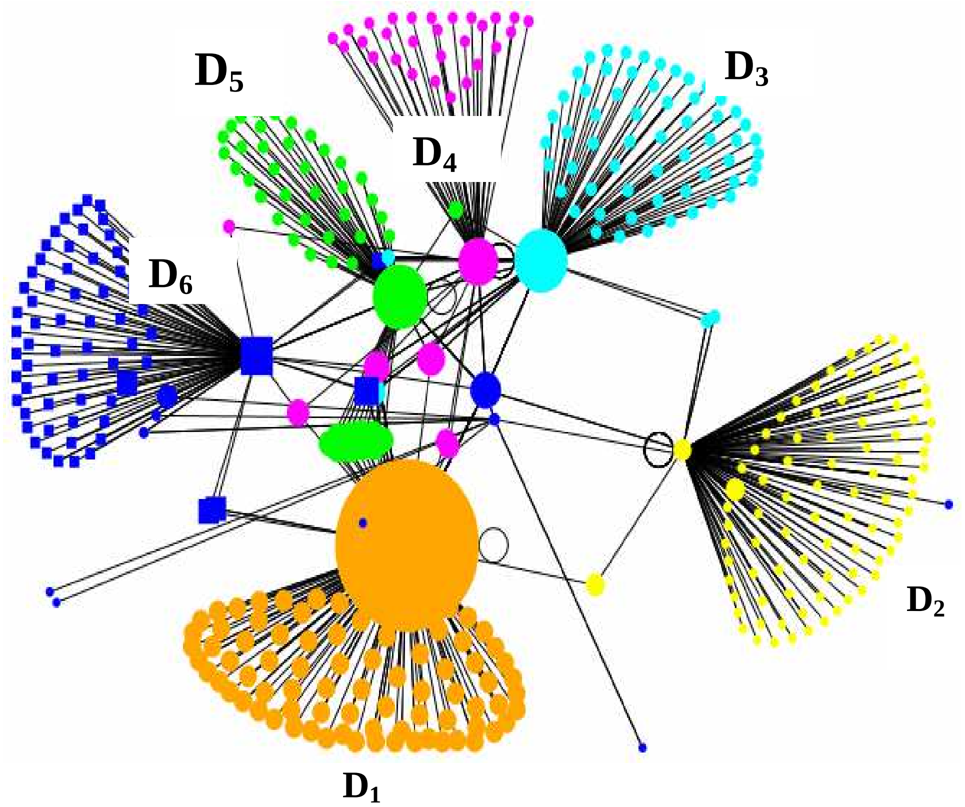


Figure 4.3: Network Density Isolation

Source: *Research data (2017)*

Therefore, the visualization of network community densities has not only helped to portray the interconnections of individuals and subgroups in a social network, but also it has also aided to expose communities who possibly dominate several ranks in the network are likely to be influential or close and bonded than others in the network. These illustrations concurs with Mulazzani, Huber and Weippl (2012) that visualizations can be a very effective tool in law enforcement agencies   investigating social networks because it enables them understand the behaviour of social media users and they can predict criminal activities by monitoring connections between suspects, understand the dynamics such as discovering the leaders, followers and new individuals being integrated into a group. The techniques of detecting and identifying communities helps to know groups of nodes densely connected than other in the entire network (Tayebi & Glässer, 2016). In a rejoinder, Faust, K. (2006) outlined that the density of a network is proportional to the probable number of connections in that network. A conspicuously dense community is susceptible and exposed to law enforcement officers for further scrutiny and identification of the main actors who are most likely to be the leaders of a particular cohort of felons (Xu, Marshall, Kaza & Chen, 2004).

4.2.4 Detecting Clusters of Communities

Figure 4.4 shows the generated visualizations results of six distinct interconnected clusters equivalent to the number of network communities (*the clusters are labelled S1 to S6*). The size of the visuals indicates how active a node is in the cluster or the entire network. Evidently, S4 community cluster density is the highest and its members are seemingly well connected and active too, hence the tight bonding visuals. Voigt, Hinz and Jansen (2013) outlined that the characteristics of clusters presenting high density scores are usually attributed to few interrelated or connected nodes. The cluster labelled S6 is the second highly active group in the network although it slightly scattered to vaguely overlap S5 cluster in the network. Except for one isolated member, all members of S3 cluster are dependent on one actor who connects them to the rest of others in the network. By employing such Social Network Analysis visualization techniques, there is a possibility to detect clusters, identify the most important actors and their roles and unveil interactions between nodes (Mena, 2003). The findings can be related to Hoppe and Reinelt (2010) observation that clustering helps to unearth important communities of a network that were not known previously. This was corroborated by Xu *et al* (2004) in cluster analysis as a way of detecting not only network subgroups but also the central actors and how they interact with other communities.

If a network community depicts a strong connection between its members, then it can aid to know the associates that belong to that network community Krebs (2002). This way, detectives can narrow down the list of suspicious characters under investigation. Moreover, identification of gatekeepers connecting to a particular subgroup (community) is also important in unearthing specific dubious characters. In Zhu, Watts and Chen (2010), network clustering helps detectives to narrow down their investigations to a specific subgroup or community.

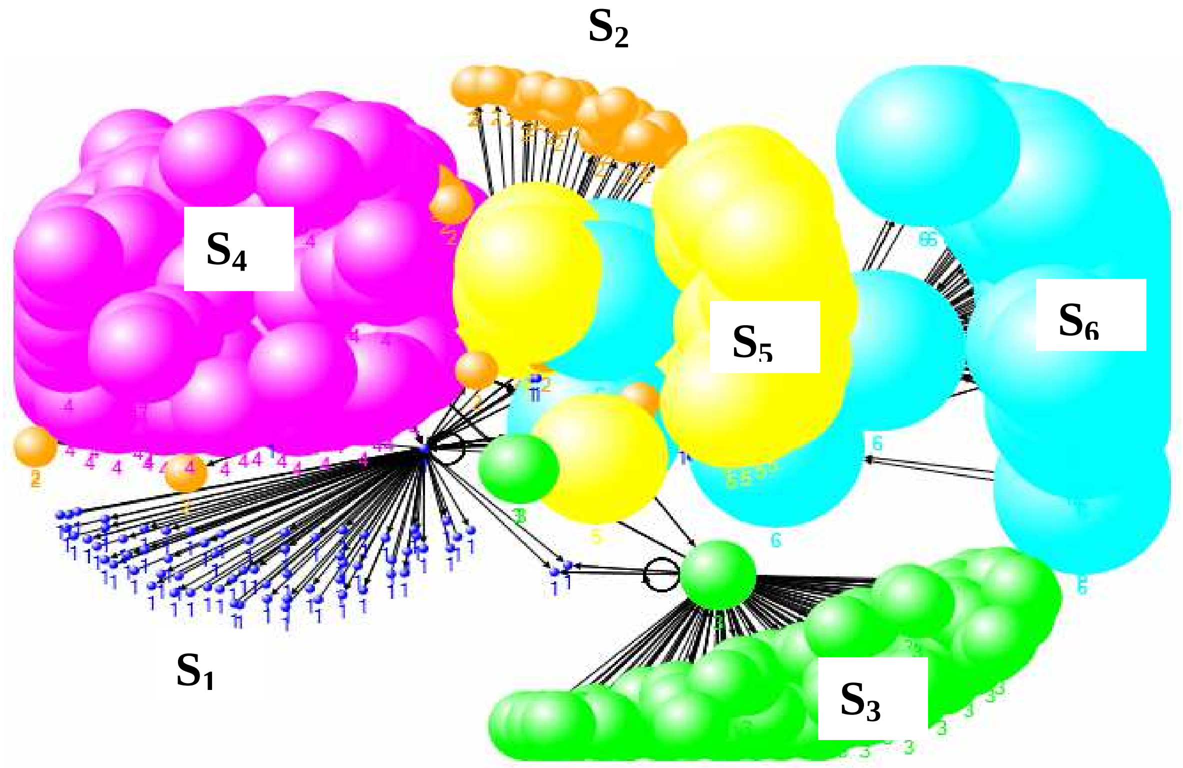


Figure 4.4: Clustering Network Communities

Source: *Research data (2017)*

4.2.5 Degree Centrality Distributions

<u>Betweenness Centrality Visualizations</u>

Figure 4.5 shows the corresponding degree distributions of the network actors under study. It clearly illustrates that actor wilf~ has highest degree and betweenness centrality and therefore is the most influential person in the entire network. This implies that respondent wilf~ probably knows what is going on in multiples social clusters of the entire network.

Identifying the actors with the highest betweenness centrality in a suspicious network helps detectives to focus their attention and resources in profoundly investigating those nodes in the entire network (Kirchner & Gade, 2011).

The research findings further portrays that respondent wilf~ act as gatekeeper by connecting the cluster that he belongs to the entire network because communications emanating from other clusters from the rest of entire network must pass through him. This means that actor wilf~ is capable of influencing the entire network but he is more susceptible to detection. Besides node wilf~ other actors namely 2279837eb26d4a1, velo~, nico~, deno~ and kiptal~ scored considerably higher degree and betweenness centralities after wilf~ from second to sixth positions respectively in the entire network but first in their clusters. This implies that these main actors are leaders or hubs of their respective subgroups in the social network. Intuitively therefore, the six actors act as intermediaries in their network subgroups because information must flow through them. The findings concurs with advice of Xu, Marshall, Kaza and Chen (2004) that while carrying out an investigation, one needs to know which nodes other actors have to connect to in order to link to the entire network and gather other valuable leading information. Nodes acting as bridges to their subgroups create structural holes which help investigators to easily detect friends of the influential nodes (Hanneman & Riddle, 2005).

Boundary Spanners Visualizations

It was also necessary to establish the boundary spanners in the network. By doing so, the researcher was able to know the actors that connect several other clusters as this will imply that they are more central in the entire network. Accordingly, Figure 4.5 further reveals the boundary spanners as actors wilf~, 2279837eb26d4a1, velo~, nico~, deno~ and kiptal~ who bridge their respective clusters and therefore are strategically placed to get information from other clusters. Furthermore, these actors are able to integrate concepts and information from other clusters. The results are closely consistent with Long, Cunningham and Braithwaite (2013) that the boundary spanners act as conduit of information flow between network nodes or individuals who cannot communicate directly or have no or little trust to each other.

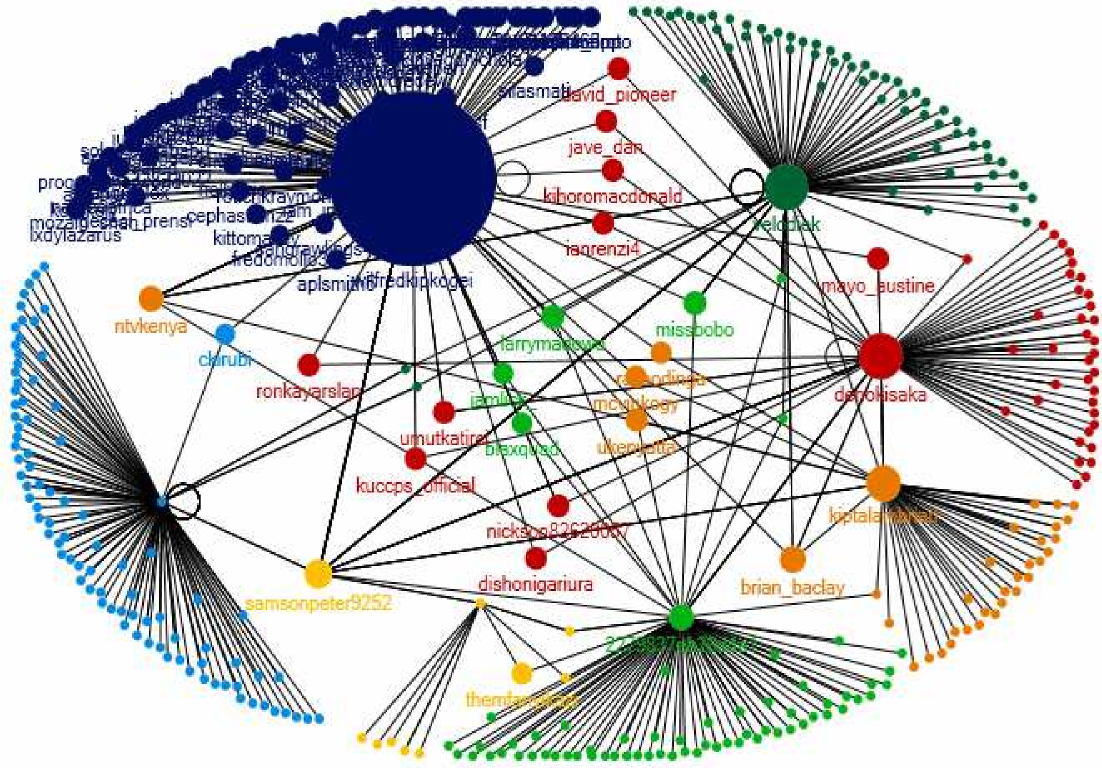


Figure 4.5: Degree Distributions

Source: *Research data (2017)*

Closeness Centrality Visualizations

Figure 4.5 visualization results also indicates that actors sams~,wilf~, deno~, velo~, kipatal~, ntvk~ and 2279837eb26d4a1 had almost equal and similar pattern of closeness centrality measures in the entire network. Connections with nodes having high closeness scores when put together with nodes having low degree centrality values can have indirectly impact on the behaviour of the other nodes in that network (Wasserman & Faust, 1994). This implies that the aforementioned respondents were highly connected to other individuals in the network. It is important to note that the thicker the edge the higher the frequency interaction between any given actors in a network.

Regarding structural similarity, the researcher attempted to depict and find actors who are linked to more other nodes connected to the influential actor in the network. This implies that if two or more actors have similar friends, then this implies that all of them are friends in real world. Actors 2279837eb26d4a1, velo~, nico~, deno~ and kiptal~ was found to have structural similarity because they largely share a number of friends as shown by the edges in the diagram and also they are leaders of their respective clusters too. This agrees with McPherson, Smith-Lovin & Cook (2001) similarity yields interconnections between nodes with structural similarity.

4.2.6 Information Flow between Network Actors

It was also imperative for the researcher to visually depict how information flows in a network between actors/respondents under study.

Figure 4.6 was generated using geometric spiral algorithm. It clearly illustrates the flow of information in the entire network between actors. At the epicentre of the information flow is the subgroup members labelled A3, spreading to the second layer. It is densely surrounded by the A1 subgroup members. This implication here is that the network members of both A1 and A3 could be sharing similar information or have same interests with each other. The A2 subgroup members are somewhat spreading to the hub of the communication flow of the network and also found at the periphery of the communication labyrinth. The results resonates with Nagl, Amos, Sewall and Petraeus (2008) reasoning that nodes on the periphery receive very low centrality scores and are often connected to networks that are not currently mapped but they are important links since they may be resource gatherers or individuals with their own network outside their isolated group. These characteristics make them very important resources for fresh information not available inside their isolated group.

Equally significant is the A6 and A4 subgroups, though a little blurred, they are also similarly positioned at the hub of the communication flow and spread heavily to the second layer after epicentre to the periphery. Cluster A3 is diminutively encircled halfway by subgroup A6 which spreads heavily to periphery. This conforms with Arnaboldi, Conti, Passarella and Pezzoni(2013) observation that the innermost circle signifies a more stronger social relations of the ego while the outermost circle are typified by fluctuating level of friendliness (also called sympathy or active network groups).

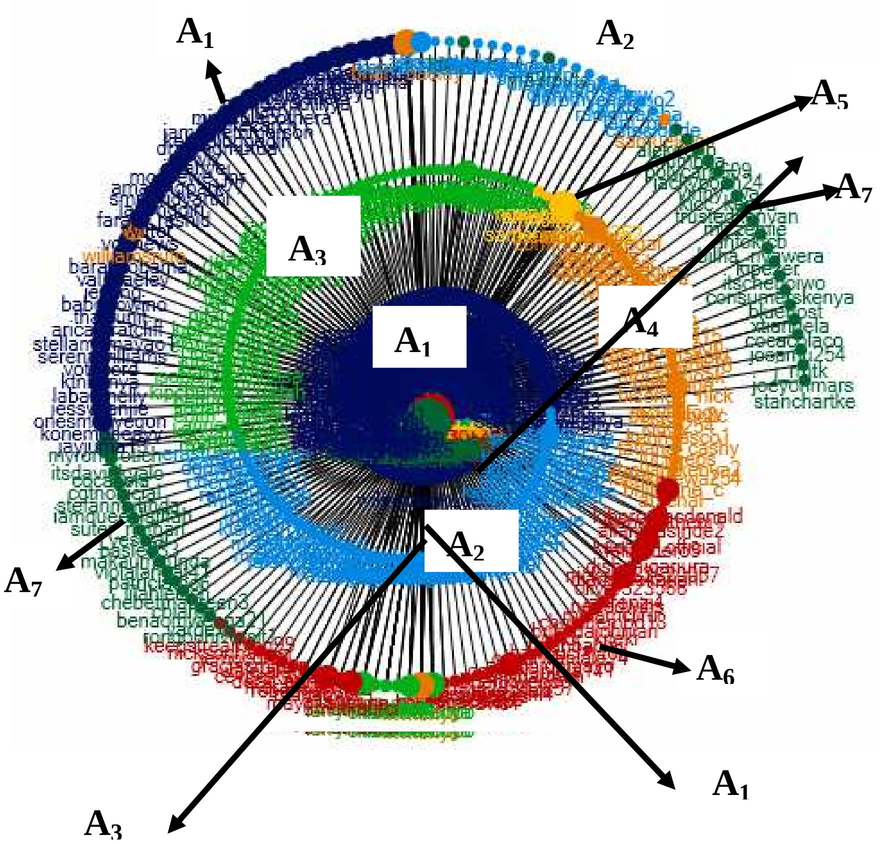


Figure 4.6: Network Concentric Information Flow

Source: *Research data (2017)*

The A7 clusters, on the periphery of these visualisations are the isolates in the communication flow. Heeding the advice of Granovetter (1973), that the important channels of communication to be closely monitored are the ones that are rarely utilized and usually located at the network's periphery, cluster A7 members elicits more scrutiny.

They seem to be recipient of the information from the entire network or they only share information specific to themselves and their interests. This visualization results correlate with Sharma and Strategy (2008) that although nodes at the periphery have less interactions with the entire network, they may be having links beyond the network and as result they can be a reservoir of new information.

Drawing and concluding from Figure 4.6 therefore, it is apparently evident that A1, A3, A6 and A5 subgroups are the most influential actors over others in this network. This concurs with Ferrara, De Meo, Catanese and Fiumara (2014) that by analysing the flow of information pattern in a given network, one can unearth actors that play key role in a criminal network or have connections to different clusters. Visualizing and identifying subgroups of a network enables investigators to unravel rich information pertaining the nodes, network recruitments paths, operational characteristics and patterns of flow of information (Bonacich, 1972). Hence, an investigator can narrow down his/he probe to those aforementioned subgroups to gather more valuable information and reveal their activities.

4.2.7 Network Communication Channels

The visualization results in Figure 4.7 shows the channels of communication between the various subgroups of the entire network under study. Notice that group one (denoted G1) and some members of group 4 (G4) have two major common communication links between themselves. This implies that some members in G4 who may not be prominent in that cluster knows each with the most influential member of G1. The two groups (G1 and G4) also share some influential or common actor in cluster G5. However, the influential node in G1 is conspicuously in touch with periphery members of cluster G6. Noticeably, G1 has the most members while G6 has the least. Nodes that are located at the terminals of the communication channels are likely to influence others whereas those found in-between the channels of communication paths are likely to be information conveyance belts (Waskiewicz, 2012).

In Figure 4.7 the findings also showed that nodes in various clusters frequently communicate with actors within their clusters than with those outside their clusters. By deriving such information, investigators can easily identify cohesive clusters and ultimately establish part of the network where information moves faster and also which cluster(s) closely keeps information to themselves.

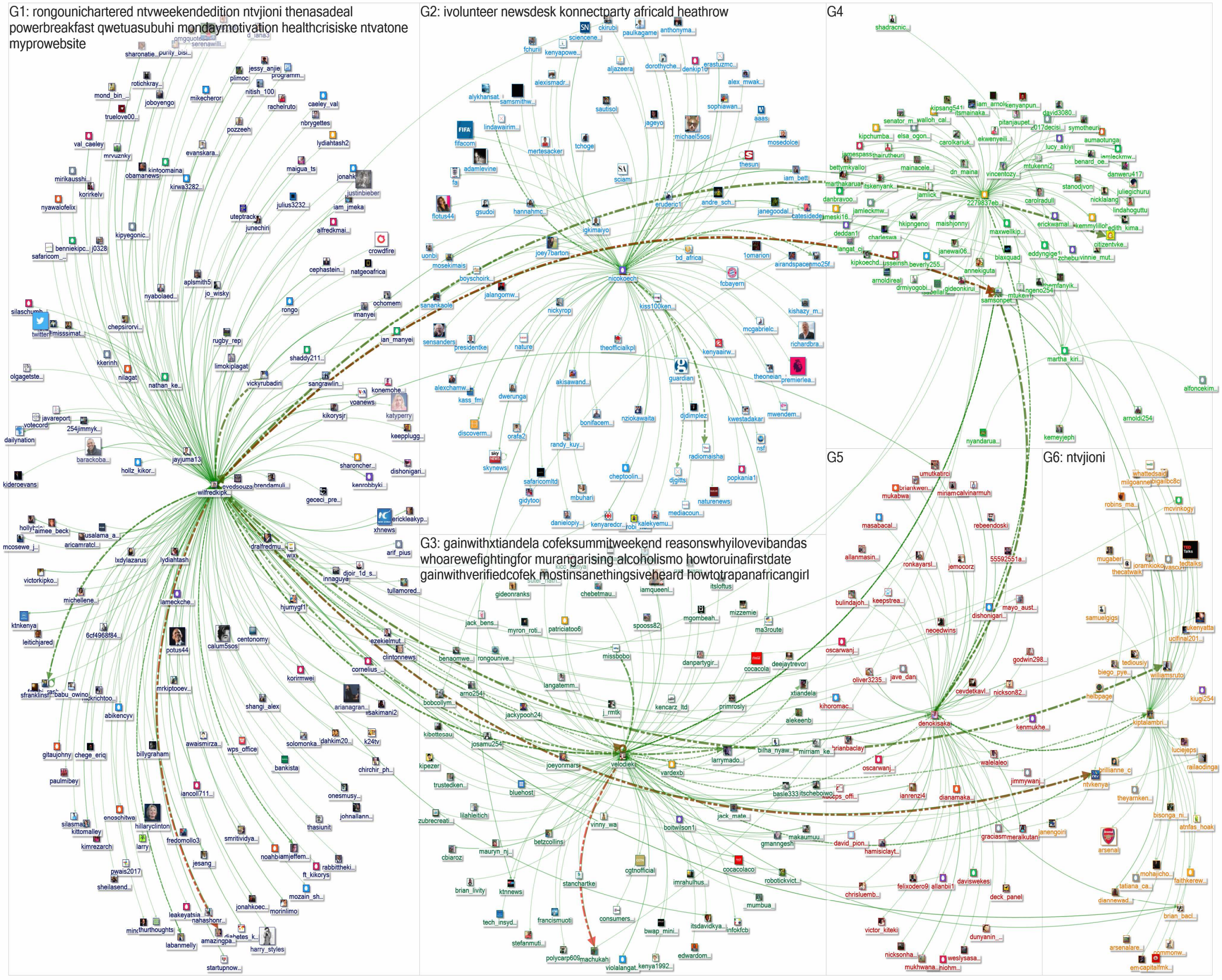

G1: rongounichartered ntvweekendedition ntvjioni thenasadeal powerbreakfast qwetuasubuhi mondaymotivation healthcrisiske ntvatone myprowebsite
G2: ivolunteer newsdesk konnectparty africald heathrow
G4
G3: gainwithxtiandela cofeksummitweekend reasonswhyilovevibandas whoarewefightingfor murangarising alcoholismo howtoruinafirstdate gainwithverifiedcofek mostinsanethingsiveheard howtotrapanafricangirl
G5
G6: ntvjioni

Figure 4.7: Network Communication Channels

Source: *Research data (2017)*

4.2.8 Network Cliques

In quest of gaining deeper insight into the network, the researcher generated a complex network visualization shown in Figure 4.8 depicting how actors of a network over time, can ultimately fragment into interesting groups called cliques. Thus Figure 4.8 shows the visualization findings of network cliques generated from the now complex network. Notice that the network has now fragmented into subgroups of cliques (*denoted as G1, G2, G3,……up to G23*) and each clique is labelled with the topics they frequently discuss. The Clauset-Newman-Moore grouping algorithm identified 23 cliques within this network. However, some cliques such as G1, G2, G3, G4 and G19 have further but minute fragments within the respective cliques. These findings indicate the nodes in the entire network are highly diverse. This implies that as time elapses, network users tend to slowly degenerate into fragmented interactions and eventually form their own cliques according the nature of information exchanged between themselves as well as interests.

Stemming from these findings, the results are consistent with Bonacich (1972) on the painstaking examination of characteristics of such cliques for homogeneous ideologies and the strength of their cohesion with that network, as well as how this influences the activities and development of the unlawful network. This way, such information can assist an investigator to tell if they are still an healthy communication or leadership roles have been changed and splinter groups emerged. Thus, the graph was generated not only to depict the number of cliques in the network but also to utilize the visual properties to map the attributes of the network showing the interaction of the actors.

The network visualizations depicted in the findings so far discussed above, underscored how important Social Network Analysis visualizations is to law enforcement agencies in unearthing leading information from a large set of data, which could otherwise been difficult or impossible to tell using conventional methods of investigations. While backing up the use of Social Network Analysis, Rahim, Amalina and Sulaiman (2015) emphasized that visualisation methods are crucial because it helps scholars to comprehend social interaction or patterns of online relations and who communicates with who more or less frequently. Hence cannot comprehend the trends and concepts of social networks without employing the use of computerized visualizations as presented in this thesis.

In an investigation scenario, the detectives ought to concentrate their probe efforts to specific actor under scrutiny then traverse the network as they examine for crucial leads. Visualizations makes this investigation process much forthright because it enables the discovery of unknown interactions and relationships that exists between actors.

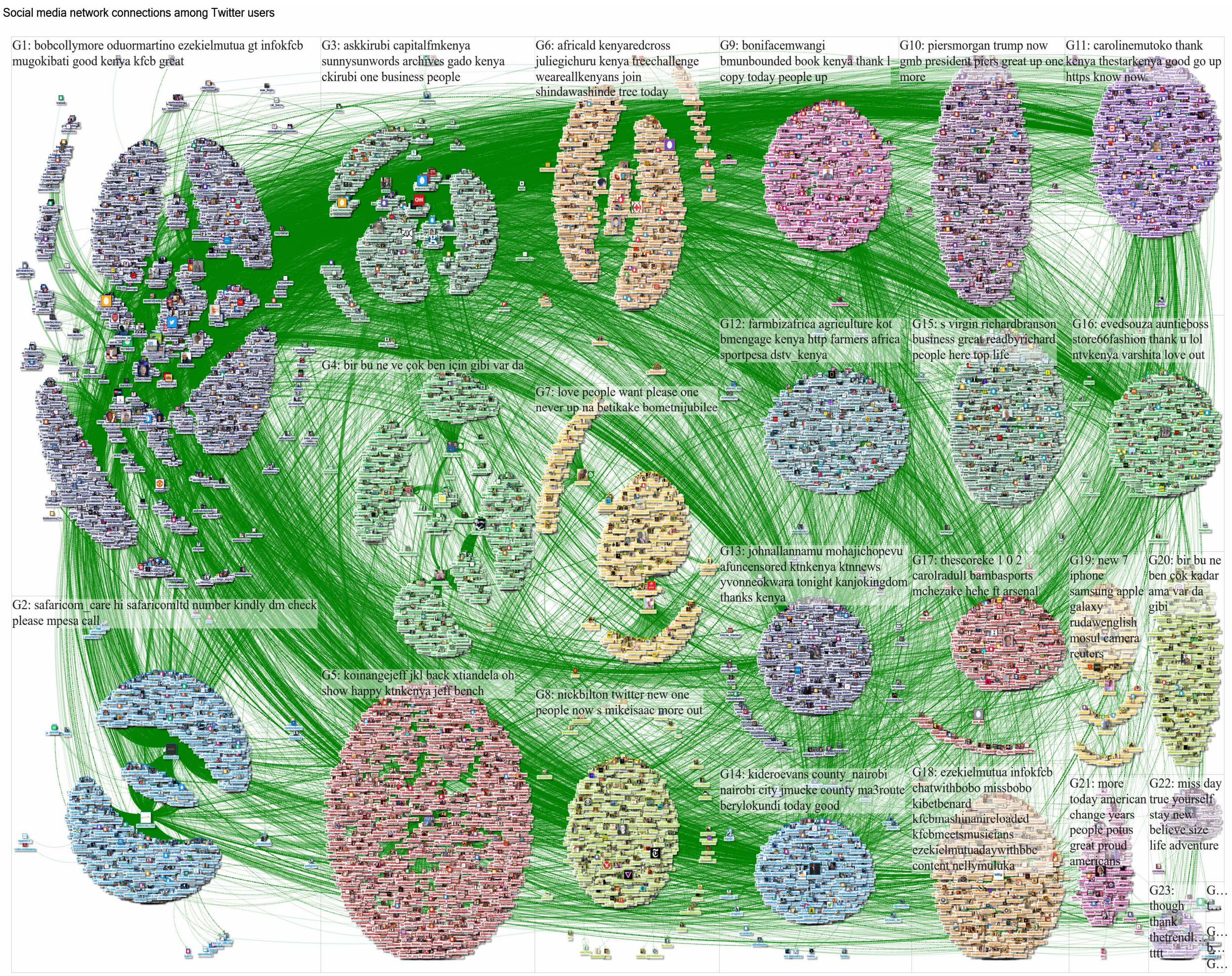

Social media network connections among Twitter users
G1: bobcollymore oduormartino ezekielmutua gt infokfcb mugokibati good kenya kfcb great
G3: askkirubi capitalfmkenya sunnysunwords archives gado kenya ckirubi one business people
G6: africald kenyaredcross juliegichuru kenya treechallenge weareallkenyans join shindawashinde tree today
G9: bonifacemwangi bmunbounded book kenya thank l copy today people up
G10: piersmorgan trump now gmb president piers great up one more
G11: carolinemutoko thank kenya thestarkenya good go up https know now
G4: bir bu ne ve çok ben için gibi var da
G7: love people want please one never up na betikake bometnijubilee
G12: farmbizafrica agriculture kot bmengage kenya http farmers africa sportpesa dstv_kenya
G15: s virgin richardbranson business great readbyrichard people here top life
G16: evedsouza auntieboss store66fashion thank u lol ntvkenya varshita love out
G2: safaricom_care hi safaricomltd number kindly dm check please mpesa call
G13: johnallannamu mohajichopevu afuncensored ktnkenya ktnnews yvonneokwara tonight kanjokingdom thanks kenya
G17: thescoreke 1 0 2 carolradull bambasports mchezake hehe ft arsenal
G19: new 7 iphone samsung apple galaxy rudawenglish mosul camera reuters
G20: bir bu ne ben çok kadar ama var da gibi
G5: koinangejeff jkl back xtiandela oh show happy ktnkenya jeff bench
G8: nickbilton twitter new one people now s mikeisaac more out
G14: kideroevans county_nairobi nairobi city jmueke county ma3route berylokundi today good
G18: ezekielmutua infokfcb chatwithbobo missbobo kibetbenard_ kfcbmashinanireloaded kfcbmeetsmusicians ezekielmutuadaywithbbc content nellymuluka
G21: more today american change years people potus great proud americans
G22: miss day true yourself stay new believe size life adventure
G23: though thank thetrendl... tttt
G...
t...
G...
b...
G...

Figure 4.8: A 2.5 - degree Network Cliques

Source: *Research data (2017)*

4.3 Identifying Significant Actors in a Network Using Centrality Metrics

Besides using visualisations to depict interesting patterns of interactions between respondents (actors) of the study, the researcher also employed use Social Network Analysis metrics as buttress of visualizations findings. In this section therefore, centrality measures which are be used to identify crucial actors with the values of closeness, betweenness, degree and eigenvector centralities were computed and tabulated. The network metrics generated has been used to describe either whole network or specific nodes within the network. It is also important to note that the number of vertices and edges kept on growing over time.

4.3.1 Summary Statistics of One Degree Ego-centric Network

Table 4.1 shows global information that summarises the main actor's initial seed network mapped to Figure 4.1 in previous section. The findings indicate that at 1-degree egocentric network, there were 55 vertices having 66 unique connections (edges). The graph density of the initial 1 degree egocentric network was 0.023 units implying that the initial dyadic connection of initial seed respondents (actors) to the then main actor in this chapter , was 2.3% links to the rest of actors in that network.

This percentage does not only demonstrate the presence of ties between actors, albeit low, but it also shows that few actors in the entire network were in communication. It is also important to note that the density values lies between 0 and 1, where 0 means there was no communication between actors. Thus network density has crucial ramifications for the actors communicate.

Table 4.1: Graph Statistics of Initial seed **1- Degree** Egocentric Network

Metrics	Value
Vertices	55
Unique edges	66
Graph density	0.0232323232323232
Average Geodesic Distance	1.927934

Source: *Research data (2017)*

4.3.2 Summary Statistics of One Degree Ego-centric Network

Table 4.2 demonstrates the overall results of an expanded network from the initial seed of 55 vertices to 483 vertices in a directed graph. Needless, to pinpoint that the results clearly indicates that there was no isolated vertices in the entire network as exhibited by a score of zero(0) of single vertex connected components. The findings indicates that the shortest path between actors of this network had the value of 4 as given by the maximum geodesic distance (diameter) score, which is 0.8 shy above the average geodesic distance score of 3.2. For instance, the shortest path between actor wilf~ and actor deno~ is 4 hops. This implies that connecting any two furthest actors in the network would need 4 links. Other geodesic distances between the rest of the actors in the network is small. Just like density scores range, modularity values also lies between 0 and 1. In Table 4.2, the results indicates the modularity value of approximately 0.6, which implies that the network clusters were fairly separated from each other in the network as shown by the visualizations in the previous section.

The average geodesic distance score 3.2 in Table 4.2 above implies that the whole community membership was slightly detached suggesting that the nodes in this network did not know one another directly. This can be explained by the fact that the actors (*respondents*) in this network were from different geographical locations and perhaps initial connections were through acquaintances or friend of a friend basis. Nevertheless, the overall network graph density score of 0.003 implies that vertices were loosely interconnected hence low density, but high density at cluster level. Needless, to say that the graph metrics was calculated using NodeXL version 1.0.1.373.

Table 4.2: Graph Statistics of **1.5 Degree** Egocentric Network

Graph Metric	**Value**
Graph Type	Directed
Vertices	483
Unique Edges	565
Edges With Duplicates	198
Total Edges	763
Self-Loops	22
Reciprocated Vertex Pair Ratio	0.15819209
Reciprocated Edge Ratio	0.273170732
Connected Components	1
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	483
Maximum Edges in a Connected Component	763
Maximum Geodesic Distance (Diameter)	4
Average Geodesic Distance	3.15864
Graph Density	0.002641684
Modularity	0.52856
NodeXL Version	1.0.1.373

Source: *Research data (2017)*

Centrality Metrics

In quest of demonstrating how Social Network Analysis can be used to identify the most important actors in the network, the centrality metrics were computed for all actors in the network. The centrality metric score was used to identify the importance of an actor in network. centrality measures of an individual in a network provides an idea the actor's role in that network and the connections between nodes reveals the general structure of the network (Hopkins, 2010).

As mentioned in chapter two of this study, numerous centrality metrics do exist. Wu, Carleton and Davies (2014) underscored the importance of centrality metrics to investigators that it helps to identify key actors in a network since it portrays how connected is that node is to the whole network alongside other actors. Tayebi and Glässer (2016) succinctly highlighted that the objective of analysing the centrality measures of nodes is to establish the key nodes in a given network in order to know their reputation, prominence or power in the entire network. Generally the effective approach for detectives is to investigate numerous nodes with high centrality metric scores rather than isolating a lone actor (Borgatti, Everett, & Johnson, 2013).

4.3.3. Degree Centrality

Table 4.3 shows the findings of degree centrality scores of the top seven actors in the network under study. Degree centrality was calculated to determine the most popular actor in the entire network having the most links. Evidently, actor wilf~ has many direct contacts than other actors hence he is the most connected actor in the entire network with a degree centrality score of 196 which is almost twice to the centrality score of nico~ who is second prominent actor in the network with a score of 93. Thus wilf~ has the highest influence in this network. The findings resonates with Berzinji, Kaati and Rezine (2012) that an actor with the highest degree centrality score is placed at a strategic position and plays a focal role of propagating information in the network. The possible roles of nodes with highest degree centrality scores include controllers, planners or brokers (Mainas, 2012).

Table 4.3: Top Eight Scores Actors for Degree Centralities

Vertex	Degree
Wilfredkipkogei	196
Nicokoech	93
Velodiek	82
2279837eb26d4a1	73
Denokisaka	54
Kiptalambrian	39
martha_kirika	9
Samsonpeter9252	7

Source: *Research data (2017)*

4.3.4 Respondents Betweenness Centrality Scores

Table 4.4 shows the computed results of betweenness centrality scores of the first seven(7) respondents or actors in the network under study. These scores were computed in in order to determine which actor in the network act as bridge between subgroups of the entire network. As can be seen from Table 4.4, actor wilf~ has highest betweenness centrality score of 69302.912, whereas actor Kiptala~ has the least score of 13802.640. Other important bridge and gatekeeper actors in this network worth mentioning from the results include actors nico~ and velo~ among others because they form the shortest pathways of communication in the entire network. The findings therefore signifies that wilf~ is the most central actor with respect to the communication to all other actors in the entire network. This conforms with Xu, Marshall, Kaza and Chen (2004) that betweenness centrality is useful for investigators to understand the crucial nodes that other actors ought to link in order to connect to the rest of the network in quest of gathering invaluable leading information. Betweenness centrality can be used to identify the actors that are critical to the success of the suspicious network and, in turn, redirect one’s attention and resources on investigating these suspects with crucial leads in social network (Kirchner & Gade,2011).

Table 4.4: Top Seven Scores Actors for Betweenness Centralities

Vertex	Betweenness centralities
Wilfredkipkogei	69302.912
Nicokoech	37752.733
Velodiek	31198.671
2279837eb26d4a1	27026.598
samsonpeter9252	22823.611
Denokisaka	19777.960
Kiptalambrian	13802.640

Source: *Research data (2017)*

4.3.5 Respondents Closeness Centrality Scores

Table 4.5 demonstrates the closeness centralities of the top actors from the entire network. The study results tabulated shows that the value of closeness centralities of a node to other nodes is a uniform score of 0.001. These scores conforms with Denny (2014) that closeness measure is sensitive to network size and is decreasing in the number of actors in the network. There is a trade-off whenever nodes with high closeness scores are connected with nodes having low degree centrality scores (Wasserman & Faust, 1994). This study outcome specifies that network actors were not only closely linked to each other but they were also able reach or access one another in the network in equal steps. These closeness scores also mean that the efficiency of broadcasting information in the entire network was relatively low. Kaye, Khatami, Metz and Proulx, (2014) observed that the probable role that node(s) scoring high closeness centrality scores is that of being an organizer because he/she can quickly reach many other actors in the network.

Table 4.5: Top Seven Scores Actors for Closeness Centralities

Vertex	**Closeness Centralities**
Wilfredkipkogei	0.001
samsonpeter9252	0.001
Denokisaka	0.001
Velodiek	0.001
Kiptalambrian	0.001
Ntvkenya	0.001
brian_baclay	0.001

Source: *Research data (2017)*

4.3.6 Respondents Eigenvector Centrality Scores

Eigenvector centrality was computed to measure the influences of an actor in the entire network. Table 4.6 shows the results of eigenvector centralities for top seven actors. Respondent Wilf~ tops the rank having a maximum eigenvector value of 0.057. This implies wilf~ is the most influential and most popular actor because he has the highest eigenvector score in comparison to other actors in the network. This concurs with Wu, Carleton and Davies, (2014) that eigenvalue centrality score of an suggest that actor is placed at strategic position in the network and therefore can be an influential person to his/her neighbouring actors as well as being the focal player in the network. Notice that as opposed to the previous centrality scores, actor deno~ has jumped up the ladder to second position with a score of a score 0.013 at par with actor velo~. The actors brian_ba~ and 2279837eb26d4a1 both have the lowest score of 0.06, implying that are less influential and less popular in amongst the top network important actors. This coincides with Wasserman and Faust (1994) that connections with actors having high closeness scores when put together with nodes having low degree centrality values can have indirect impact on the behaviour of the other nodes in that network.

Table 4.6: Top Seven Scores Actors for Eigenvector Centralities

Vertex	Eigenvector Centralities
Wilfredkipkogei	0.057
Denokisaka	0.013
Velodiek	0.013
Kiptalambrian	0.009
samsonpeter9252	0.007
brian_baclay	0.006
2279837eb26d4a1	0.006

Source: *Research data (2017)*

Centrality Metric Scores Summary

In summary, node wilf~ consistently scored high in most centrality values whereas most actors swapped positions in most categories. By ranking position overall in all centrality measure values, respondent wilf~ implies that he is the strongest, popular and influential actor in the entire network in comparison to the rest of the network actors. In Wu, Carleton and Davies (2014) stronger actors possess a lot of liberty, power and influence, but their redundant connections cannot make them good brokers. However, actors having fewer ties are the most secure, although their capacities have limits in accessing information.

4.3.7 Respondents PageRank Scores

In order to establish the actors' centrality scores using their connectivity in the weighted subgroups of the network, the researcher employed the use of PageRank, which is a variant of Eigenvector centrality measure. Presentation of results in Table 4.7 hitherto gives respondent wilf~ a leading edge with a score of 81.686, followed at a far distant with nico~ with a score of 40.995. The results are clearly different from that of Eigenvector centralities. These concurred with Nouh and Nurse (2015) recommendations that the importance of employing PageRank metric is because it considers not only the quantity of ties between actors as it is in Eigenvector centrality, but it also takes into consideration the quality of such ties between actors. It also confirmed Kwak, Lee, Park and Moon (2010) findings that PageRank is a quality metric measure which identifies key actors in a network by determining its importance based on the number of in-coming connections to the node. In essence, PageRank is said to be an algorithm of link analysis that gives number weights to each actor with an aim of computing and evaluating their significance in the network.

Table 4.7: PageRank

Vertex	PageRank
Wilfredkipkogei	81.686
Nicokoech	40.995
Velodiek	33.346
2279837eb26d4a1	29.795
Denokisaka	20.860
Kiptalambrian	15.584
martha_kirika	3.594
samsonpeter9252	2.595

Source: *Research data (2017)*

4.3.8 Respondents Clustering Coefficient Scores

In contrast to measures of centralities, the researcher computed the clustering coefficient to show how connected friends of main actors were to the ego-neighbourhood. Interestingly, unlike in previous centrality scores, the results in Table 4.8 placed actors brian_ba~ and itsdavid~ as top actors with a clustering coefficient of 1.00 apiece amongst many other actors. This signifies that friends of actors brian_ba~ and itsdavid~ and other actors with similar clustering coefficient scores knew each other well. Surprisingly, the main actors of centrality measures were nowhere to be seen near the top of the clustering coefficient results. More specifically, actor wilf~ scored clustering coefficient value of 0.001, actor velo~ registered 0.05 and actor deno~ recored 0.016. The implication of these findings is that friends of actors like wilf~, velo~ or denok~ and others having low clustering coefficient scores, were most likely not acquainted each other.

Table 4.8: Clustering Coefficient

Vertex	Clustering Coefficient
brian_baclay	1.000
Itsdavidkyalo	1.000
Railaodinga	1.000
jave_dan	1.000
Umutkatirci	1.000
Mayo_austine	1.000
allanbii1	1.000
Ntvkenya	0.500
samsonpeter9252	0.333
Missbobo	0.333
Larrymadowo	0.333
Ukenyatta	0.333
Kiptalambrian	0.020
Denokisaka	0.016
Velodiek	0.005
Wilfredkipkogei	0.001
2279837eb26d4a1	0.000

Source: *Research data (2017)*

4.3.9 Respondents Reciprocated Vertex Pair Ratio Scores

Table 4.9 shows an excerpt of the reciprocated vertex pair ratio results. Recall that in Table 4.2, the overall network score for reciprocated vertex pair ratio was 0.15819209. It also important to note that these results always oscillate between 0 and 1 such that if all edges are connected, the score is 1 and 0 if the network is not connected.

Table 4.9: Reciprocated Vertex Pair Ratio

Vertex	*Reciprocated vertex pair Ratio*
Umutkatirci	1.000
mtuken1	1.000
sangrawlings419	1.000
Rugby_rep	1.000
Cephasteinzz	1.000
………………………	
samsonpeter9252	0.857
brian_baclay	0.750
Itsdavidkyalo	0.500
Missbobo	0.333
martha_kirika	0.333
Denokisaka	0.269
Wilfredkipkogei	0.211
Kiptalambrian	0.205
Velodiek	0.113
2279837eb26d4a1	0.096
Nicokoech	0.055

Source: *Research data (2017)*

4.3.10 Overall Graph Statistics of 2.5 -Degree Egocentric Network

Table 4.10 above demonstrates results of 2.5 -degree network cliques as depicted previously in Figure 4.8 visualization findings. Once again, the vertices and edges have tremendously increased from 483 to 29,295. These numbers denotes the relationships between actors around the ego. There were a lot of isolated networks as explained by single vertex connected components score of 15 as supported graphically different cliques. The maximum geodesic distance or diameter of the network increased to 7 with a slight increase of its average to of approximately 3.5. The results closely relates to Duijn (2016) observation that a high geodesic distance scores between nodes means that information flows faster in the network. Moreover, the overall network graph density score of 6.112E-05 indicated that actors' interconnections were quite scattered in the entirety of the network but there was a high density at the network cliques. It also denotes that in that network, each actor has almost 6 other actors connected to him/her.

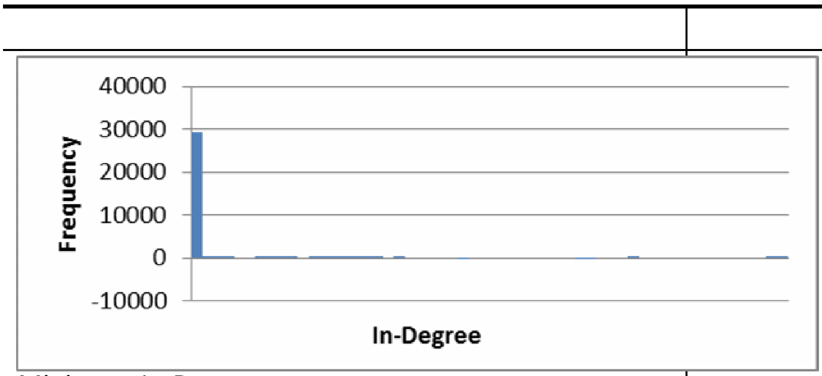
Table 4.10: Overall Graph Statistics Final **2.5 -Degree** Egocentric Network

Graph Type	Directed
Vertices	29295
Unique Edges	39202
Edges With Duplicates	129975
Total Edges	169177
Number of Edge Types	4
Mentions	90361
Tweet	32817
Replies to	31406
Follows	14593
Self-Loops	32817
Reciprocated Vertex Pair Ratio	0.174827524
Reciprocated Edge Ratio	0.297622452
Connected Components	16
Single-Vertex Connected Components	15
Maximum Vertices in a Connected Component	29280
Maximum Edges in a Connected Component	168997
Maximum Geodesic Distance (Diameter)	7
Average Geodesic Distance	3.453219
Graph Density	6.11174E-05
Modularity	0.203981
NodeXL Version	1.0.1.373

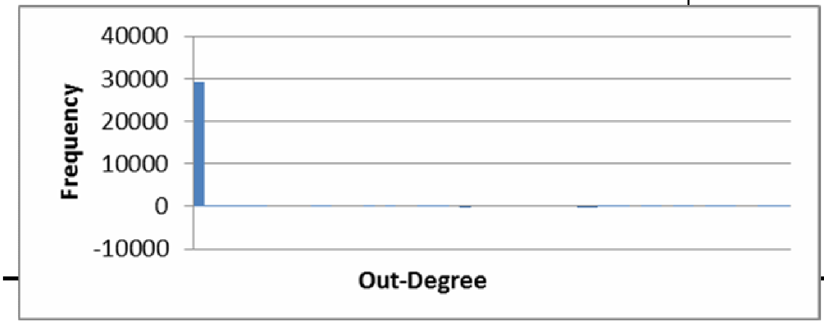
Source: *Research data (2017)*

4.3.11 Respondents In and Out –Degree Centrality Scores

Figure 4.9 shows the results of degree centralities. More specifically, it depicts and tabulates the findings of in-degree and out-degree centralities of the network actors for 2.5 degree egocentric network shown in Figure 4.8. The maximum in-degree score was 1828 recorded against actor safari~. This implies that the actor received the highest attention from other actors. On the other hand, the maximum out-degree value was 2121 recorded against actor koinang~ suggesting that this actor interacted most by sending out the most information to other actors in the network. Nodes scoring high in-degree centrality scores signifies high reputation and on the other hand, nodes with high out-degree centrality scores implies that they are influential (Mainas, 2012). In an investigative scenario, Sparrow (1991), advices detectives to identify nodes that play important roles and remove them so as to immobilize a suspected criminal network.



Minimum In-Degree	0
Maximum In-Degree	1828
Average In-Degree	1.825
Median In-Degree	1.000



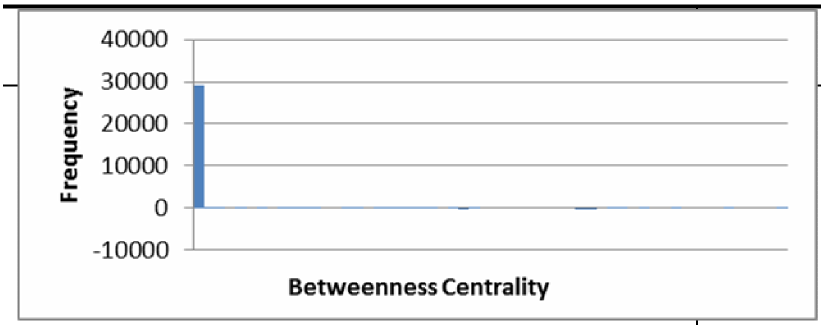
Minimum Out-Degree	0
Maximum Out-Degree	2121
Average Out-Degree	1.825
Median Out-Degree	0.000

Figure 4.9: In and Out Degree Centralities of **2.5 -Degree** Egocentric Network

Source: *Research data (2017)*

4.3.12 Respondents Summary Betweenness Centrality Scores

Figure 4.10 demonstrates the findings of 2.5 degree egocentric betweenness centrality. The maximum score 148196529.709, a value recorded against actor sams~ in the network. Therefore, the research concluded that actor sams~ high betweenness centrality score made him a controller and an information bridge of the social links between actors of the entire network. Nevertheless, the maximum score of betweenness centrality gave, study hinted a possible existence of a connection on geodesic between actor sams~ and any other node in the network. According to Johansson and Tenggren (2015) a node with the highest betweenness centrality is a threat if that node somehow cease to exist from the network because interaction will disappear in that network all of a sudden too.



Minimum Betweenness Centrality	0.000
Maximum Betweenness Centrality	148196529.709
Average Betweenness Centrality	71794.481
Median Betweenness Centrality	0.000

Figure 4.10: Betweeness Centrality of **2.5 -Degree** Egocentric Network

Source: *Research data (2017)*

4.3.13 Respondents Summary Closeness Centrality Scores

The closeness centrality results in Figure 4.11 registered an average closeness of an actor to other actors was a score of 0.00, a similar results to its maximum and minimum closeness centralities. The inference drawn from these results is that there was slow communication between actors of this network and probably all the nodes were not in a position to observe information flow in the network. Another implication for the results is that since the maximum or average closeness score is the same across, then it was somehow difficult to propagate information. This agrees with (Johansson and Tenggren, 2015) that an actor with a highest closeness centrality score can easily circulate information throughout the network than a node with a smaller score. In an investigation scenario, a node with high betweenness is likely to be aware of what is going on in multiple social circles and has as great influence over what flows and does not in the network.

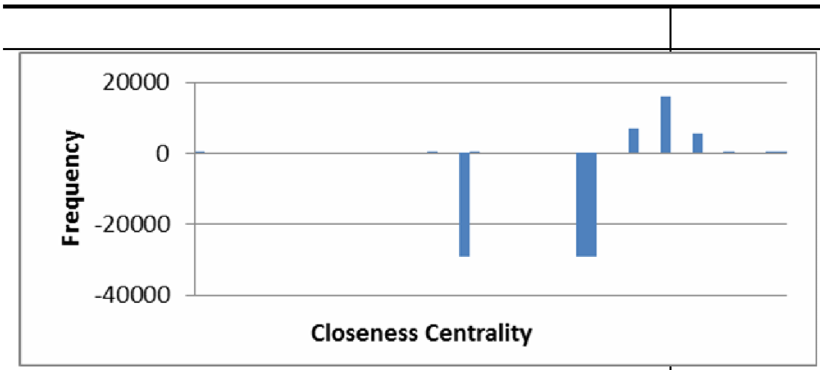
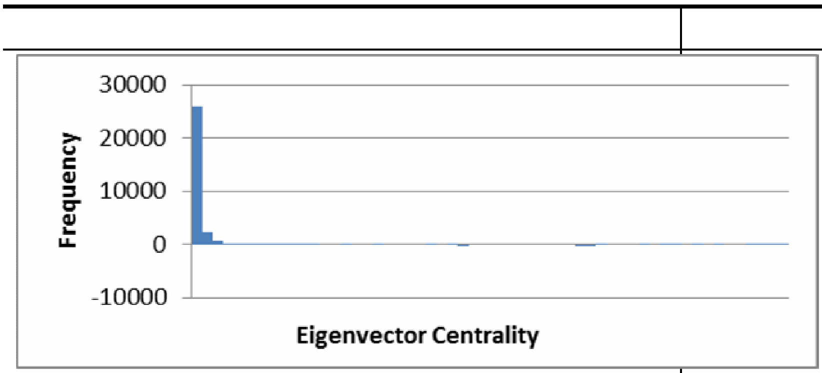
	0.000
Maximum Closeness Centrality	0.000
Average Closeness Centrality	0.000
Median Closeness Centrality	0.000

Figure 4.11: Closeness Centrality of **2.5 -Degree** Egocentric Network

Source: *Research data (2017)*

4.3.14 Respondents Summary Eigenvector Centrality Scores

The findings in Figure 4.12 shows the maximum eigenvector centrality value for the 2.5-degree egocentric network was 0.03 recorded against respondent mugoki~. This means that actor mugoki~ is connected with other well connected nodes of the network because the eigenvector centrality metrics put into consideration the degree centralities of the node together with those that the node connects to. Getting connected to a prominent node is far much better than getting connected to a lonely actor (Hansen, Shneiderman & Smith, 2011)



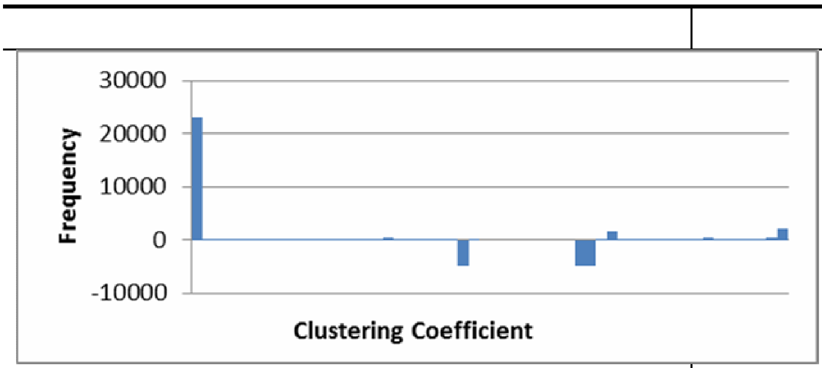
Minimum Eigenvector Centrality	0.000
Maximum Eigenvector Centrality	0.003
Average Eigenvector Centrality	0.000
Median Eigenvector Centrality	0.000

Figure 4.12: Eigenvector Centrality of **2.5 -Degree** Egocentric Network

Source: *Research data (2017)*

4.3.15 Respondents Summary Clustering Coefficient Scores

Figure 4.13 shows the clustering coefficient score of ego- neighbourhood of the entire network. The findings reported a maximum clustering coefficient of 1.00 and minimum clustering coefficient of 0.00. The implication of the results indicate that whereas some friends of other actors knew one another, other actors' friends were totally not acquainted with one another. This concurs with Johansson and Tenggren (2015) that if one's alters are familiar to one another, then that actor will have a high clustering coefficient score and the opposite is true (Johansson & Tenggren, 2015). Clustering coefficient is applicable to both a single node or the entire network. According to Tayebi and Glässer (2016) advises that detectives can establish who main actors of the network are and then start probing profoundly.



Minimum Clustering Coefficient	0.000
Maximum Clustering Coefficient	1.000
Average Clustering Coefficient	0.139
Median Clustering Coefficient	0.000

Figure 4.13: Clustering Coefficient of **2.5 -Degree** Egocentric Network

Source: *Research data (2017)*

4.4 Demographic and Other Related Information of Social Media Users

In this section, the researcher used the background information of the actors who scored high centralities scores to demonstrate how different range of intelligence information can be harvested from social media users. This information can be harvested directly or using traces the actors leave as they traverse the various social media platforms.

4.4.1 Demographic and Related Information

In order to demonstrate how social media user's background information can be invaluable tools that are able to give a leading intelligence to investigators, the researcher sampled the social media account details of actors who were amongst the top seven (7) in network centrality scores.

Figure 4.14 depicts the facebook profile of the node whose pseudoname was Nicoko~. The picture reveals that the actor has an attachment to a particular institution of higher learning as shown by the inset picture. Another point we can deduce about this actor is that the use of an image similar to that of virgin Mary as used by the Roman catholic church. The presumption here is that the actor is more likely to be subscribing to Christian doctrines, notably a Catholic by faith. This corroborates with Hudaib (2014) that when creating a social account a lot of social networks persuade users to disclose their vital primary information such their dates of birth, contacts, or place of residence and is quite amazing to realize the magnitude and particulars of the individual's information a number of social media users give, and wonder about how clued-up this information divulging are.

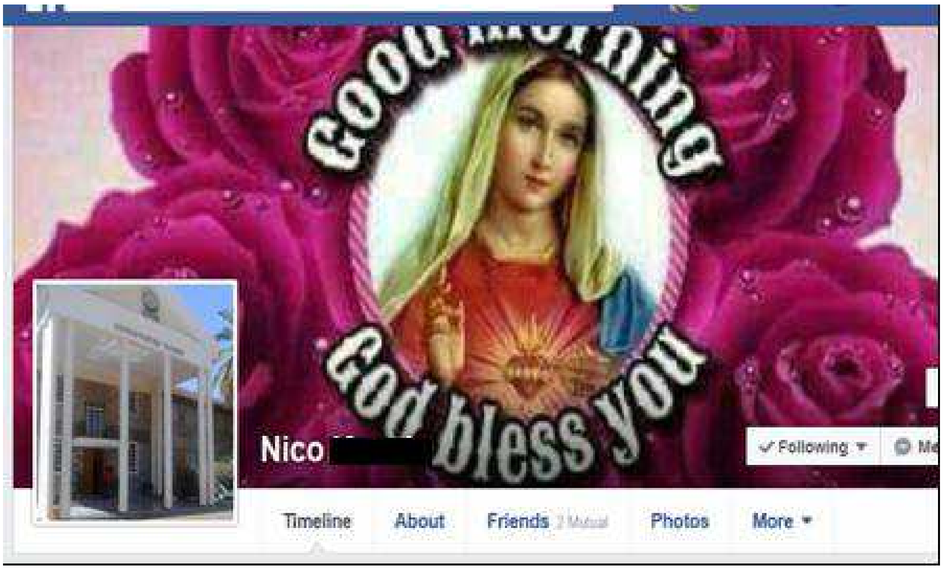
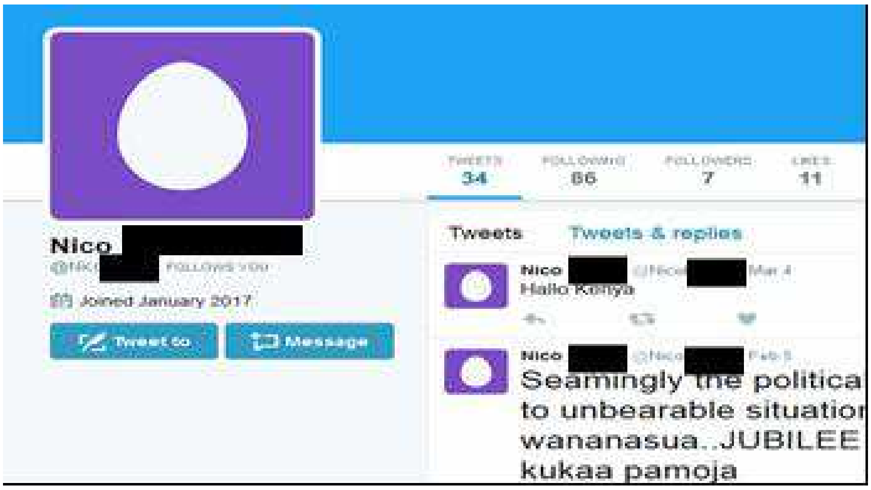


Figure 4.14: Nicoko~ Facebook Profile Picture

Source: *Research data (2017)*

Moreover, the researcher also took an extra level and studied actor Nicoko~ twitter profile picture and other details. As shown in Figure 4.15, twitter micro-blog did not reveal a lot about the node in question. Nonetheless, Figure 4.15 displays crucial information about the actor. The information that was extracted include what this actor likes discussing about as exhibited in his tweet dated February 5th, 2017. The same actor also has tweeted 34 times and few *follows* but he *follows* a lot of other social media users. This means that the actor is less likely to be a leader of many as supported by the number of his *followers* and *likes*.



Nico [redacted]
@NK [redacted] FOLLOWS YOU
Joined January 2017

TWEETS	FOLLOWING	FOLLOWERS	LIKES
34	86	7	11

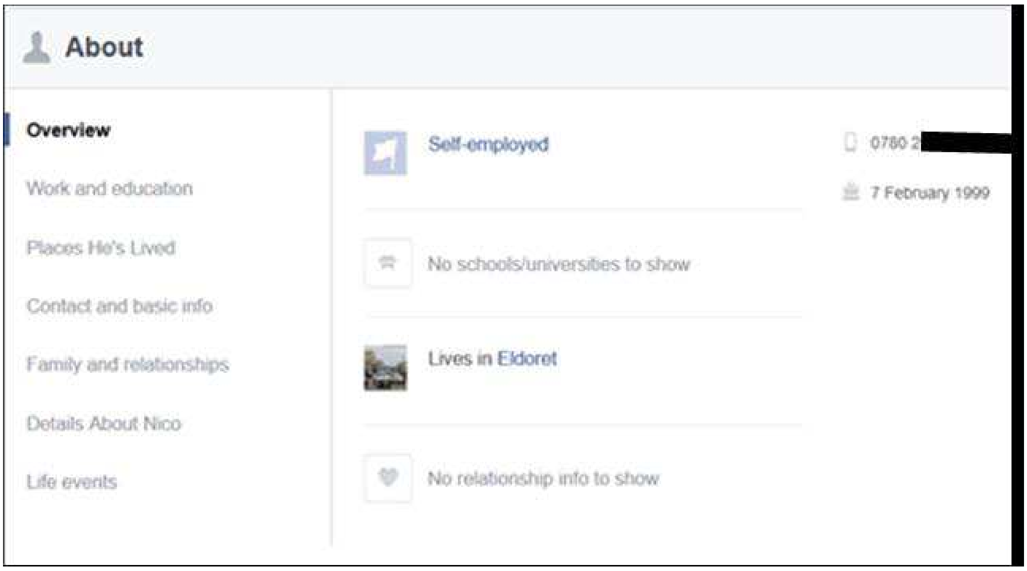
Nico [redacted] @Nic [redacted] Mar 4
Hallo Kenya

Nico [redacted] @Nico [redacted] Feb 5
Seamingly the politica
to unbearable situatio
wananasua..JUBILEE
kukaa pamoja

Figure 4.15: Nicoko~ Twitter Profile Picture

Source: *Research data (2017)*

Figure 4.15 portrays an overview profile information belonging to Nicoko~. The profile evidently reveals the actor's important information which comprise of phone number, date of birth and the city he lives in. From the date of birth, we can infer that the actor is a teenager. The findings closely relates to Golbeck (2015) that individuals using online social media platforms create and paint their online identification.



Self-employed

No schools/universities to show

Lives in Eldoret

No relationship info to show

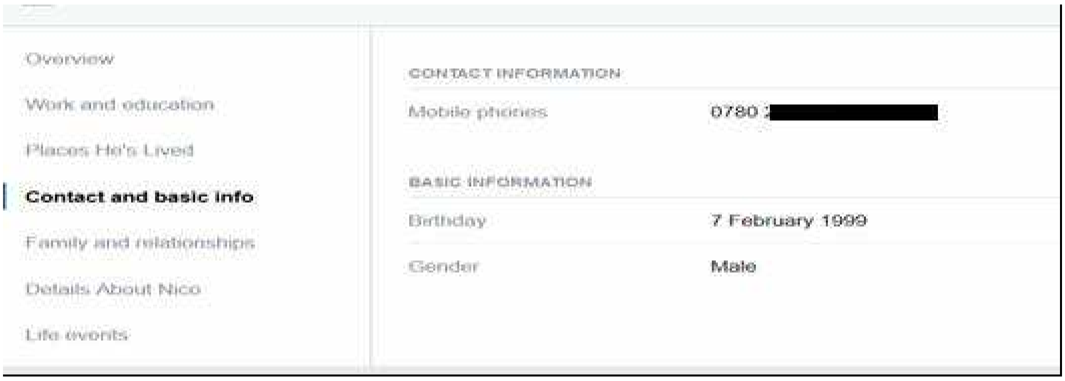
0780 [redacted]

7 February 1999

Figure 4.16: Nicoko~ About Facebook Information

Source: *Research data (2017)*

As a furtherance of actor Nicoko~ 's profile, the contact and basic information revealed his gender as male teenager as depicted in Figure 4.17.



CONTACT INFORMATION

Mobile phones 0780 [redacted]

BASIC INFORMATION

Birthday 7 February 1999

Gender Male

Figure 4.17: Nicoko~ Facebook Contacts and Basic Information

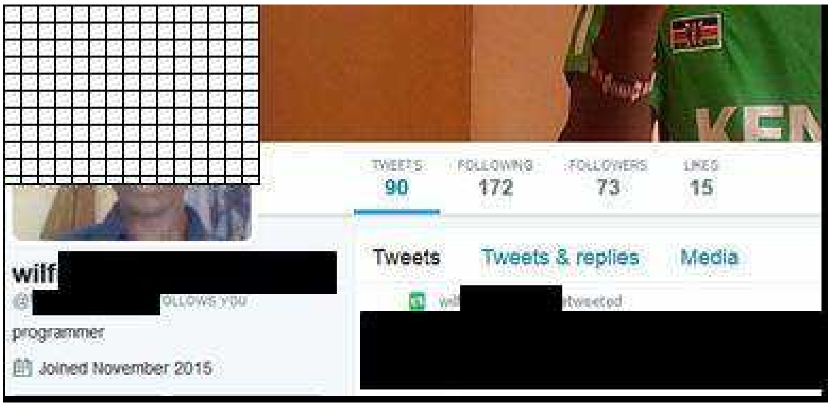
Source: *Research data (2017)*

Figure 4.18 shows the actor wilf~ profile picture (*the pictures has been concealed security reasons*). Nevertheless, the profile picture of this actor is a milestone intelligence lead for investigators as this will enable them know whom they are probing.

Figure 4.18: Wilf~ Facebook Profile Picture

Source: *Research data (2017)*

From Figure 4.19 shows the Twitter profile page of actor wilf~ which was extracted and examined by the researcher. The Profile page reveals important information that can aid investigators about the actor. More specifically, the actor revealed much information such as his real photograph, his professional inclination on daily basis and the date he joined the micro-blogging site twitter. Moreover, the number of followers and tweets also explains why this actor was quite influential as demonstrated by the highest centrality metric scores he had elsewhere in this thesis. Yang, Liu and Sageman (2006) advises that a good investigation process is to start from some known actor, scrutinize the links he/she has with other actors and of some crucial relationship is revealed, and then a detective can follow that lead and continue with expansion of the associates until a crucial connection is unearthed between actors who seemed unrelated in the first instance.



TWEETS 90 FOLLOWING 172 FOLLOWERS 73 LIKES 15

wilf [redacted]

@[redacted] OLLOWS YOU

programmer

Joined November 2015

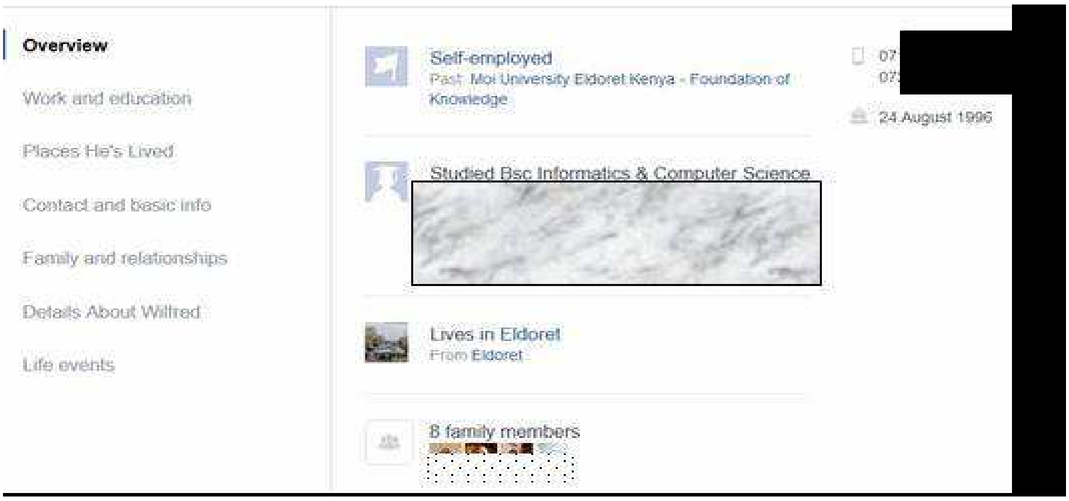
Tweets Tweets & replies Media

wil [redacted] etweeted

[redacted]

Figure 4.19: Wilf~ Twitter Profile Picture
Source: *Research data (2017)*

Figure 4.20 displays more profile information about actor wilf~. On examining profile overview of this actor, the researcher discovered invaluable information such as date of birth, where university he attended, the course he did, the city he lives, two sets of mobile numbers and the number of his family members.



Overview
Work and education
Places He's Lived
Contact and basic info
Family and relationships
Details About Wilfred
Life events

Self-employed
Past: Moi University Eldoret Kenya - Foundation of Knowledge

Studied Bsc Informatics & Computer Science

Lives in Eldoret
From Eldoret

8 family members

07
07 [redacted]

24 August 1996

Figure 4.20: Wilf~ Facebook Overview Information
Source: *Research data (2017)*

Last but not least, the researcher scrutinized the demographic details of the third important actor whose pseudo-name was velo~. From Figure 4.21, unravelling the profile picture alone reveals that the actor is most likely to be a female with a taste of fashion trends as depicted by display of shoes and some clothes.

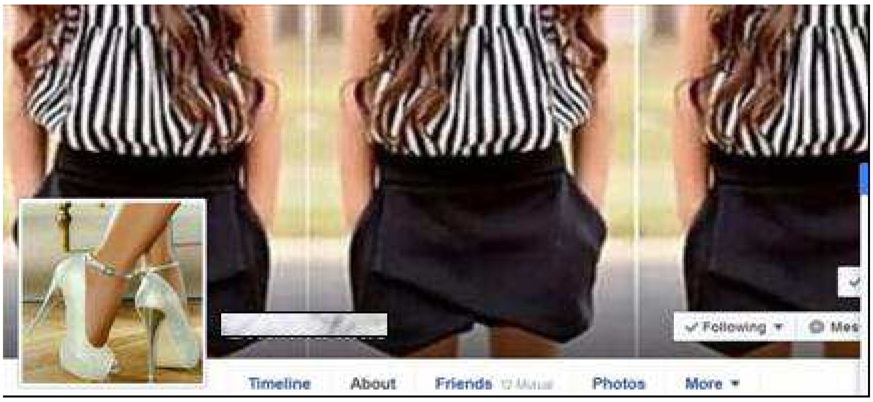
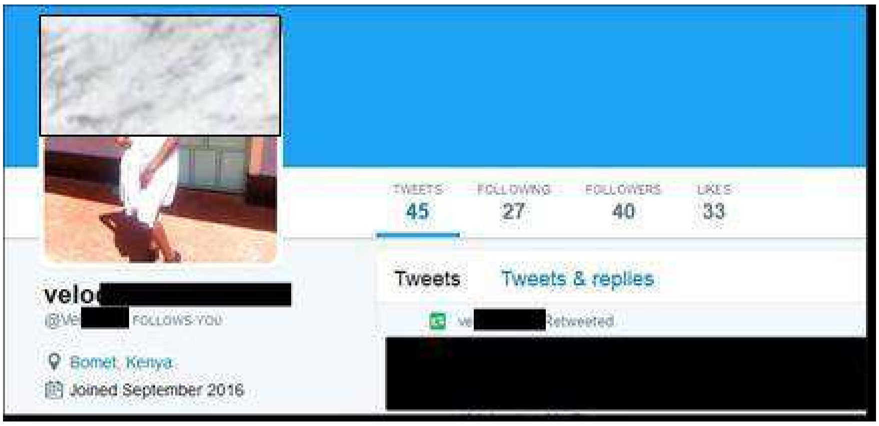

[redacted]

Figure 4.21: Velo~ Facebook Profile Picture
Source: *Research data (2017)*

Actor velo~ twitter profile page continued eliciting the researcher's attention in obtaining more information about her. As can be shown in Figure 4.22, the actor reveals her likely true picture (*concealed for security and study ethics reasons*) and her likely real name is also posted here. Rice and Parkin (2016) advices that detectives can use the social media for investigating criminal activities either as reactive or proactive means.


[redacted]

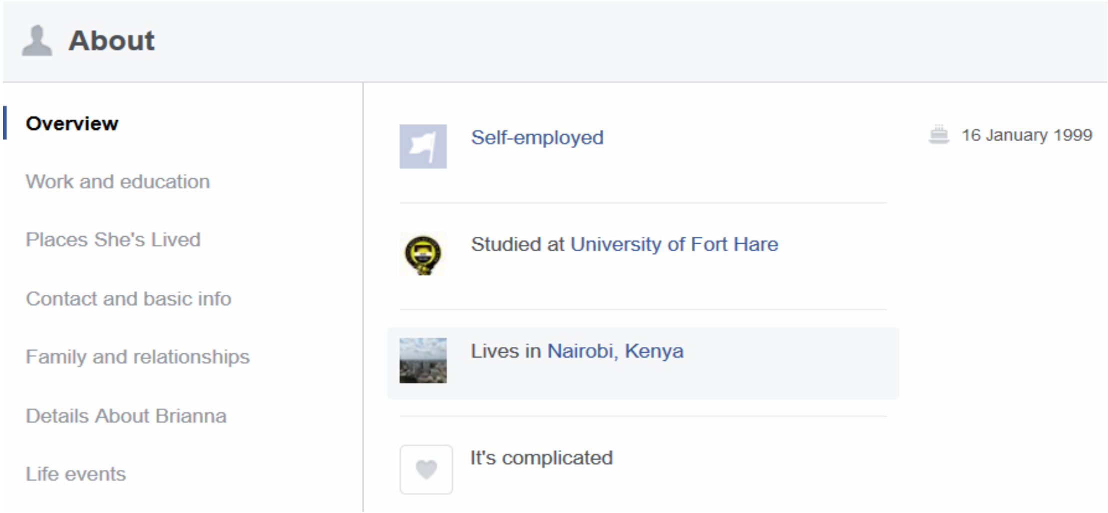
velo[redacted]
@Ve[redacted] FOLLOWS YOU
Bomet, Kenya
Joined September 2016

TWEETS	FOLLOWING	FOLLOWERS	LIKES
45	27	40	33

Ve[redacted] Retweeted
[redacted]

Figure 4.22: Velo~ Twitter Profile Picture
Source: *Research data (2017)*

A background check of actor velo~ revealed additional information about her. The date of birth, the current city where she lives in and the university she attended are freely harvested from Figure 4.23. Nouh and Nurse (2015) advises that actors whose having similar attributes such as date of birth, tribe or religious background are likely to be vital creating network ties and co-offending.


About

Overview

Work and education

Places She's Lived

Contact and basic info

Family and relationships

Details About Brianna

Life events

Self-employed

16 January 1999

Studied at University of Fort Hare

Lives in Nairobi, Kenya

It's complicated

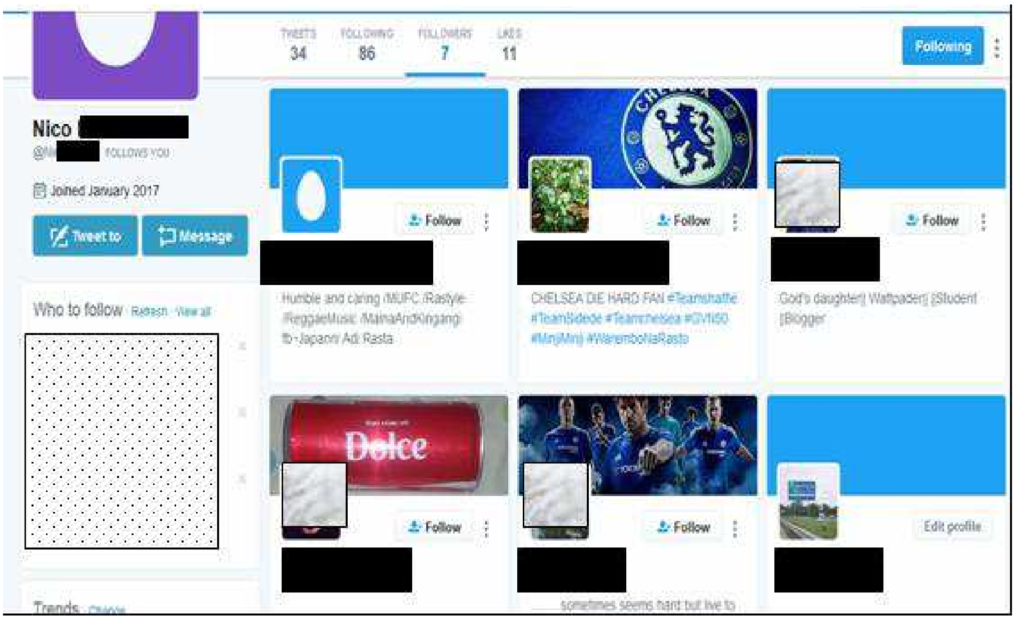
Figure 4.23: Velo~ Facebook Profile Information
Source: *Research data (2017)*

The findings above demonstrate how myriad of information about specific targeted online actors can be harvested to a point of revealing the true identify. The above results concurs with Flynn (2002) reflection that nodes or clusters can easily be identified using profile information such as date of birth, names, phone numbers, education or work history among other crucial leads. In a rejoinder, Sparrow (1991) states that individuals on various social media platforms can divulge their true online identify by leaving vital leads or traces about themselves either knowingly or unknowingly.

4.4.2 Social Links and Acquaintances

In this section, the researcher extracted vital information from actors that can be used help investigators know how to probe and understand the nodes being investigated and whom the node associates with. Waskiewicz (2012) concurs with the research topic of this thesis that the investigation of ego-network entails the process of probing of a particular node in that specific network as well as all the actors the node is connected to. This way, the study analysed the connections and associations of few actors who sored highly in both centrality metrics and visualizations. Using the advice of Hudaib (2014) that Facebook account holders have been known to easily consent to friendship invitations from other strange users just because each has numerous friends in a circle, the researcher circumvented with this idea and obtain wealthy information for investigators.

The researcher practically tested the old adage "*Show me your friends and I will tell you who you are*", to reveal crucial information the selected actors of high centrality measures. Figure 4.24 exhibits twitter followers of actor Nicoko~. His followers are likely to be football fans implying that the actor is also a football fanatic. One of his friends seems to be a reggae music fan too. The attributes of a particular node or a cluster or not so crucial than their connections or the associations between the nodes within the network (Berzinji, Kaati & Rezine, 2012).



TWEETS 34 FOLLOWING 86 FOLLOWERS 7 LIKES 11

Nico [redacted]
@Ni[redacted] FOLLOWS YOU
Joined January 2017

[redacted]
Humble and caring /MUFC /Rastyle /ReggaeMusic /MainaAndKingang fb-Japanni Ad Rasta

[redacted]
CHELSEA DIE HARD FAN #Teamshaffie #TeamSidede #Teamchelsea #GV1650 #MinjMinj #WaremboNaRasta

[redacted]
God's daughter| Wallpaper| |Student |Blogger

[redacted]

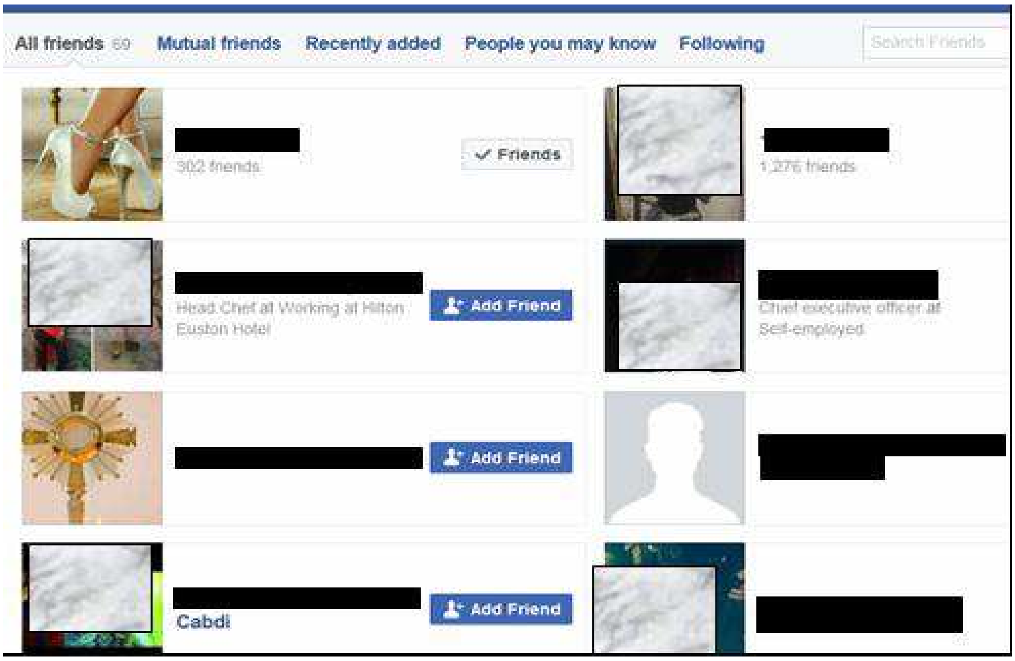
[redacted]
sometimes seems hard but live to

[redacted]

Figure 4.24 : Nicoko~ Twitter Followers

Source: *Research data (2017)*

From Figure 4.25, Nicoko~ has 69 friends. A cursory look at these friends divulges more about the kind of individuals that Nicoko~ associates with. A good number of his friends seems to be influential given the number of friends they have. For instance, two of his friends have 302 and 1,276 friends respectively. Intuitively therefore, Nicoko~ seems to be associated or has good rapport with celebrities and senior people of various institutions. Semitsu (2011) opined that knowing the individuals that a node in question associates with can easily help the detectives to know the ambitions the node has and this ease their investigative efforts in gathering important and trivial information that is linked to the individuals under investigation. This agrees with Sageman (2004) that when investigating a particular node on social media, one ought to examine relationships which comprise of friends, relatives, clerics, acquaintance(s) and teacher(s).



[redacted]
302 friends

[redacted]
1,276 friends

[redacted]
Head Chef at Working at Hilton Euston Hotel

[redacted]
Chief executive officer at Self-employed

[redacted]

[redacted]

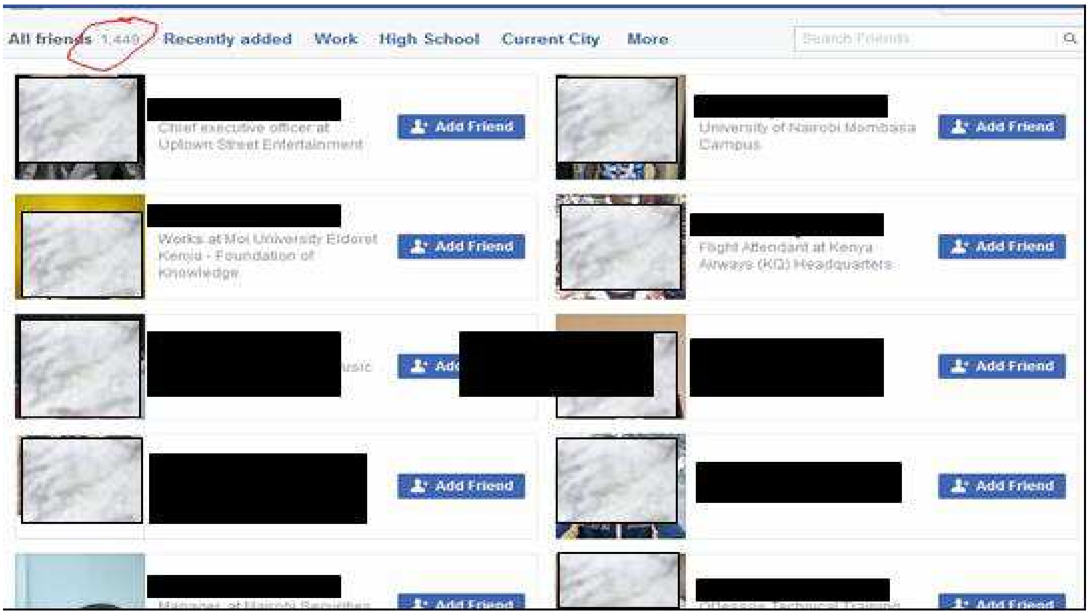
[redacted]
Cabdi

[redacted]

Figure 4.25 : Nicoko~ Facebook Friends

Source: *Research data (2017)*

Figure 4.26 shows the number of Facebook on actor wilf~ profile friends page. Respondent wilf~ whose centrality metrics and visualizations scores were majorly the highest amongst the actors/ respondents of the network, apparently still emerges to maintain the same status. Evidently actor wilf~ has 1449 friends to his Facebook account, almost 30% of the maximum 5,000 friends that Facebook allows. This actor certainly has a lot of individuals interested in him. Some category of his friends seems to be holding senior positions in various institutions as exhibited by profile briefs. Knowing the persons one associates with can easily help the detectives to know the ambitions the node has and ultimately use little efforts to gather crucial information that is linked to the nodes under investigation (Semitsu, 2011). Kunkle (2012) accentuated that extreme ideologies can propagate throughout the various social media platforms and probable criminals can interrelate with persons with whom they have the same school of thought whether they live in the same locality or across the globe.



[redacted]
Chief executive officer at Uptown Street Entertainment

[redacted]
University of Nairobi Mombasa Campus

[redacted]
Works at Moi University Eldoret Kenya - Foundation of Knowledge

[redacted]
Flight Attendant at Kenya Airways (KQ) Headquarters

[redacted]
usic

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]
Manager at Nairobi Securities

[redacted]
[illegible] Technical Training

Figure 4.26: Wilf~ Facebook friends
Source: *Research data (2017)*

Nevertheless, more important traces continued to be harvested from actor wilf~. In particular, Figure 4.27 has a wealthy of pictorial leads to any investigator after this actor. He has uploaded 28 photos from his mobile phone and 24 photos posted on his profile gallery among other sections of his facebook account. An investigator would examine these photos and deduce so many information about this actor. For instance, these photos can tell us what kind of people actor wilf~ associates and hang around with. The detective can also get to know and isolate individuals that appear most in these photos. Furthermore, investigators can also take keen interest on other actors who *like* or *comment* on the posted photos and the frequency they do it. This way, they will get to know his true associates. This agrees with advice of Miller (2011) that patterns of interactions between actors can be harvested from phatic expressions such as likes, comments and pokes one's social media account updates.

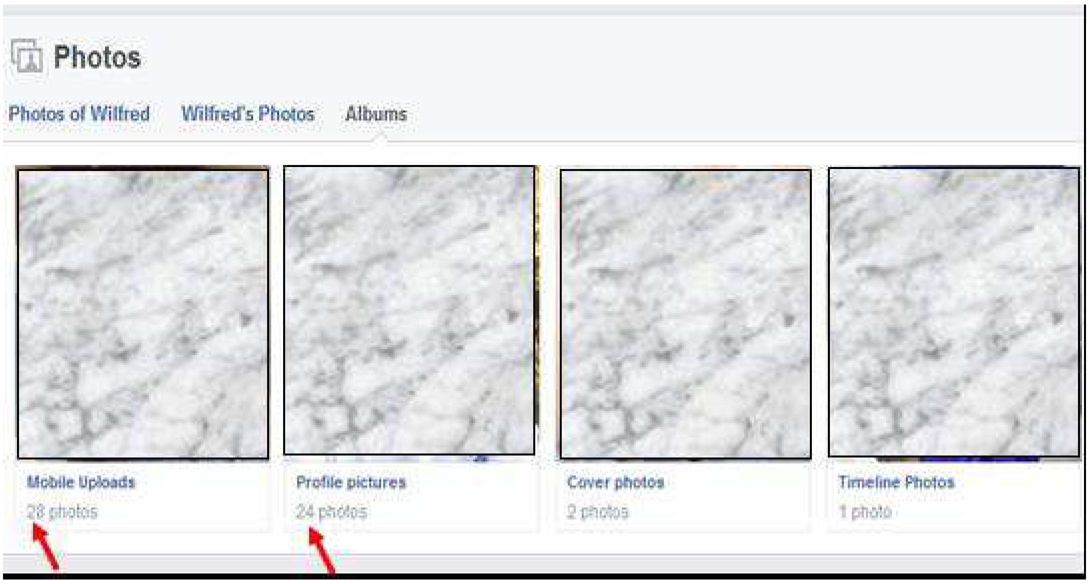


Figure 4.27: Wilf~ Facebook Friends in Photos
Source: *Research data (2017)*

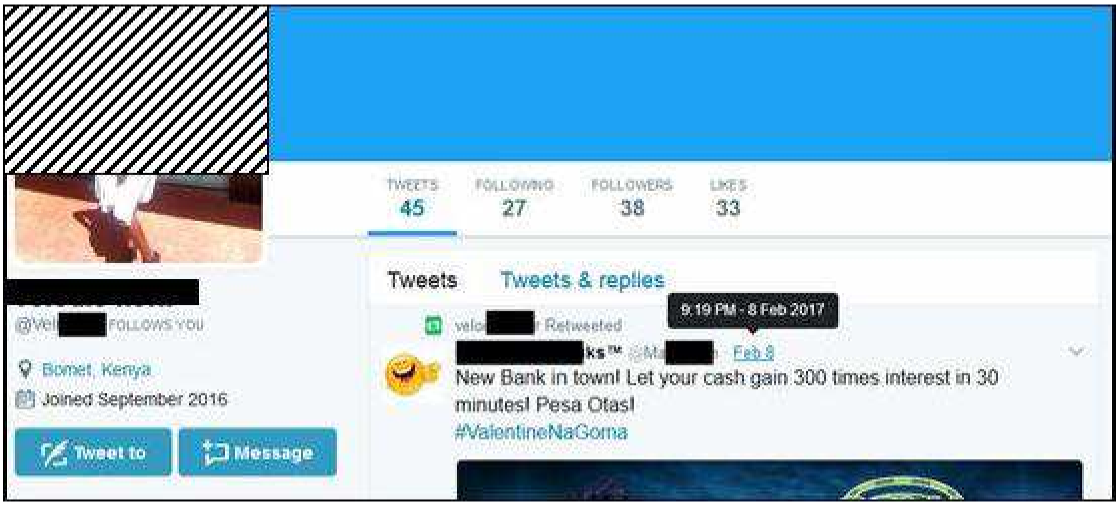
Golbeck (2015) succinctly states that any individual's social links and their online network accounts irrespective how they were created will ultimately be visible on various social media platforms either through *followers* or *friends* connected to one's profile. However, Krebs (2002) cautions that criminals usually do no create various new connections that is external to their network and rarely activate the his/her associates within the network

4.4.3 Mapping and Time Stamping Location Data.

In quest of demonstrating how more court ready social media information can be gathered pertaining the actor under investigation, the researcher typified this concept by use of various available techniques of mapping and showing time stamps as well as locations data of the actor(s). Therefore, the researcher illustrated how to harvest information of the actors' by either checking and analysing their posts online or by extracting embedded information that are usually automatically fixed on photos that one can post.

Location Mapping

Figure 4.28 shows twitter profile page of actor velo~. When creating an account with various social media platforms, one is prompted to supply the location information (*though not compulsory in some platforms*). One way of getting such information is that most social media users knowingly or unknowingly supply location information, making it easier to trace where that actor hails from. Scrutinizing actor velo~ page carefully, one would clearly see location on the bottom left, underneath the actor's profile picture. The actor most likely hails from Bomet, Kenya.



TWEETS 45 FOLLOWING 27 FOLLOWERS 38 LIKES 33

[redacted]

@Vel[redacted] FOLLOWS YOU

Bomet, Kenya

Joined September 2016

Tweets Tweets & replies

velo[redacted]r Retweeted

9:19 PM - 8 Feb 2017

[redacted]ks™ @Ma[redacted] Feb 8

New Bank in town! Let your cash gain 300 times interest in 30 minutes! Pesa Otas! #ValentineNaGoma

Figure 4.28: Velo~ Twitter Profile Page

Source: *Research data (2017)*

Time Stamping

Besides location mapping, time stamp of posted messages was also be extracted from the actor's tweets/retweets or timeline. For instance, in Figure 4.28, the actor retweeted a message on February 8th 2017 at 9:19PM. In Twitter, such time stamp is obtained by pointing at the date the message was tweeted or retweeted. This concurs with Fraser (2008) that timeline matching which entails the timestamps can be used to match the timelines of different users, and to create an exact timeline for an entire cluster of friends or even a larger group which indicate where a person was and when can invaluable lead for investigators in gathering crucial forensic evidence.

EXIF Metadata

In situations where actors failed to supply their location information on the social media platforms they used to create their online accounts, the researcher checked if they have ever posted photos or similar graphical objects on their online social media platform accounts. Figure 4.29 shows a concealed photo extracted from actor samso~ facebook account. This way, photos will still provide its Exchangeable Image File Format (EXIF) metadata showing various important information such as Global Position coordinates indicating the location where the photo was taken as well as time and date among other numerous crucial information pertaining the gadget that was used to capture the said photo(s). Digital cameras and smartphones automatically geotag pictures with the exact locations where the pictures were taken (Fusco *et al*, 2010).

Figure 4.29: Sams~ Facebook Photo
Source: *Research data (2017)*

GPS Coordinates

Table 4.11 shows the GPS coordinates and other valuable information extracted from the picture in Figure 4.5. The picture depicted was then uploaded to a free Online [1]Exif viewer which generated important results shown in Table 4.5. The result gives auto-encoded GPS location of where the picture was taken, the date and time of when it was captured using Apple iPad device. Particularly, the GPS coordinates generated comprised of latitude, longitude and altitude (i.e. *GPS Latitude = S 0º 10.21' 0'', GPS Longitude = E 35º 57.88' 0'' and GPS Altitude =1898.00m*). When interpreted, it means that Latitude is zero degrees and 10.21 minutes to the South of the equator (denoted by Zero), longitude is 35 degrees and 57.88 minutes to the East of the Greenwich Meridian(GMT) or longitude zero and altitude is 1898 meters above sea level. The findings concurs with Hanson (2011) that harvesting and using embedded information contained in a graphical object like a photo or video can provide crucial information about the location of that individual at that time as given by GPS coordinates, same information will also serve as evidence of where the individual was, versus where he/she claims to have been at that time and date and possibly what he/she was doing there. This concurred with Vicente, Freni, Bettini and Jensen (2011) that geotagging can help in validating alibis or linking a node to a crime scene.

Table 4.11: Extracted Photo Exif Data

Exif data	
Camera make	Apple
Camera model	iPad
Date/Time	2015/07/17 16:05:20
Resolution	1936 x 2592
Flash used	No
Focal length	4.3mm (35mm equivalent 35mm
Exposure time	0.0034 s (1/296)
Aperture	f/2.4
ISO equiv.	80

[1] http://www.verexif.com/en/index.php

Whitebalance	Auto
Metering Mode	Pattern
Exposure	program (auto)
GPS Latitude	S 0º 10.21' 0"
GPS Longitude	E 35º 57.88' 0"
GPS Altitude	1898.00m

Source: *Research data (2017)*

Figure 4.30 depicts the location map generated form the Exif viewer to show the place where the picture was taken from and other crucial details. The Exif viewer results clearly shows that the picture was taken in Nakuru, Kenya. These findings concurs with Fusco *et al* (2010) observation that nowadays, smartphones and digital cameras are becoming helpful tools because they automatically geotag and embed metadata of location information where the photos were capture, including time and date. Thus the above location related findings is support by Zambri (2015) assertion that any uploaded or tagged photographs associated to a particular individual suggests what that person loves, places he has visited, people he has been with and activities they did together.

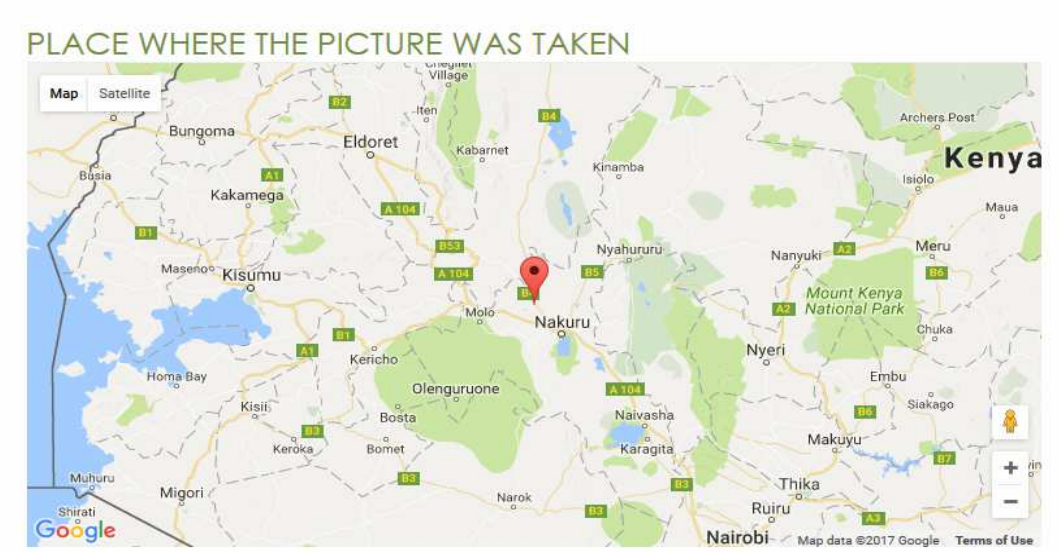
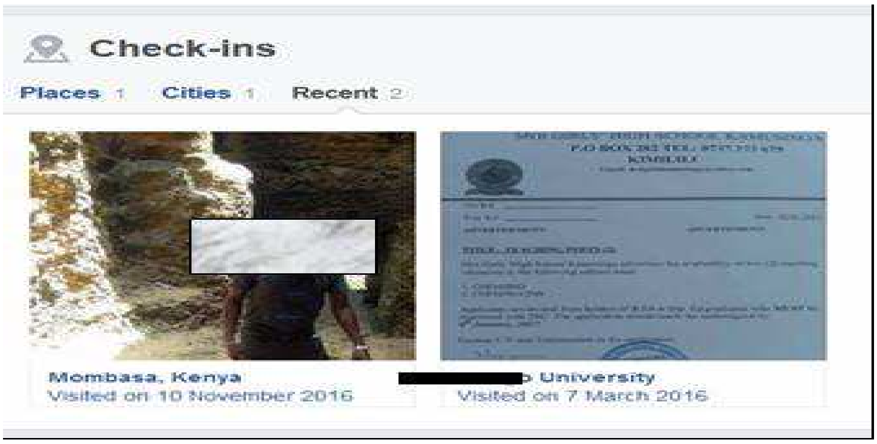


Figure 4.30: Sams~ Place Where Picture was Taken

Source: *Research data (2017)*

Extracting intelligence from Check-in details

Figure 4.31 shows actor wilf~ Facebook check-ins details. The actor's check-in details indicates to have visited Mombasa city on November 10th, 2016. In Investigating a particular actor, a detective can also use *Check- in*, if any, to locate exact location where that actor was while posting his/her message(s). This way, can easily know where the actor in frequents and his/her tastes. In gathering important forensic evidence about a particular suspect, timestamps can be used to match timelines where the person was and at what time (Fraser, 2008). Similarly, the second check-in information gives a lead onto about the institution this actor is affiliated with. An individual's locations frequented plays a big role in judging and determining one's behavioural patterns (Zambri, 2015).



Check-ins

Places 1 Cities 1 Recent 2

Mombasa, Kenya
Visited on 10 November 2016

[redacted] University
Visited on 7 March 2016

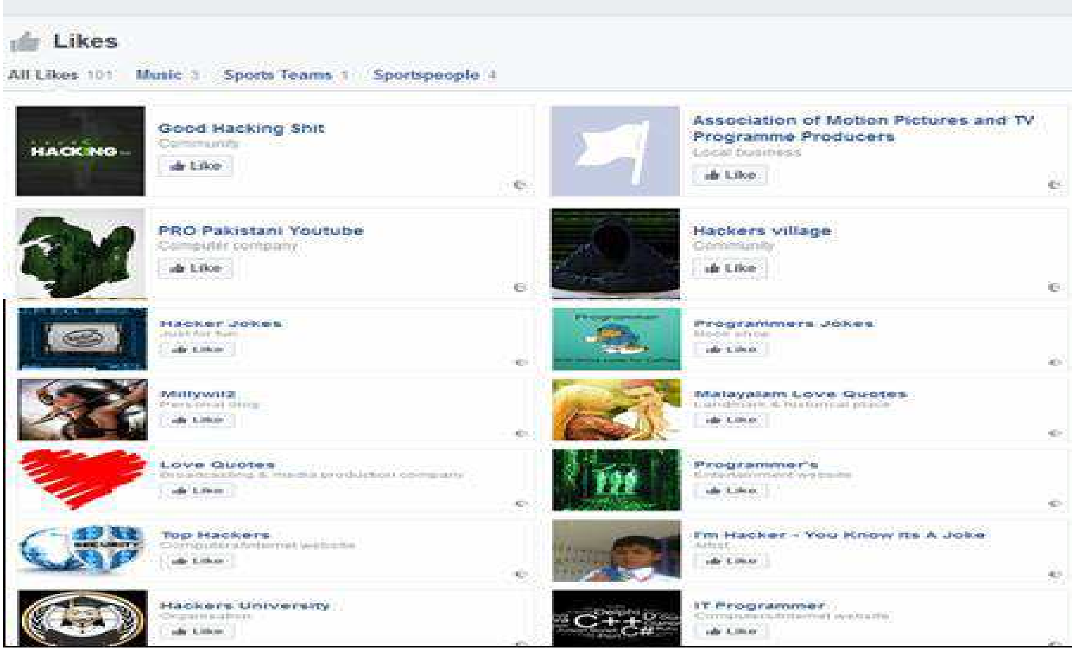
Figure 4.31: wilf~ Facebook check-ins

Source: *Research data (2017)*

The findings in section so far strongly confirms Hansen, Schneiderman and Smith (2011) reflections that actors on social media indeed leave behind innumerable traces of trails as they traverse and connect with other individuals, move from one place to another as they become active online and using their digital devices to capture images or any related graphical objects. Similarly Murphy and Fontecilla (2013) postulated that a court –ready evidence harvested from social media includes but not limited to items such photographs, status updates, a person’s location at a certain time, and direct communications to or from a suspect’s social media account.

4.4.4 Behavioural Patterns

Figure 4.32 shows the *Likes* page of actor wilf~. There are 101 things that are of interests to this actor. More specifically, examining the excerpt of his facebook *likes* page reveals that node wilf~ is fascinated with hacking as shown by number of items on mentioning hacking related stuff. Thus, wilf~ is either a white or black hat hacker of some sort. For an investigator to reveal the insight information pertaining the behavioural patterns of actor(s) being probed, several actors related activities were harvested from the selected social media platforms. Knowing what an individual likes, whom he/she interacts or frequently associates with or places one visits at a given time of day or on a given day of the week will definitely help an investigator to unearth more information about the behaviour of the actor(s) in question. The findings agrees with Klerks (2001) that detectives ought to target actors in the network with a particular expertise in a given discipline.



Likes

All Likes 101 Music 3 Sports Teams 1 Sportspeople 4

Good Hacking Shit
Community

Association of Motion Pictures and TV Programme Producers
Local business

PRO Pakistani Youtube
Computer company

Hackers village
Community

Hacker Jokes
Just for fun

Programmers Jokes
Movie show

Millywil2
Personal blog

Malayalam Love Quotes
Landmark & historical place

Love Quotes
Broadcasting & media production company

Programmer's
Entertainment website

Top Hackers
Computers/internet website

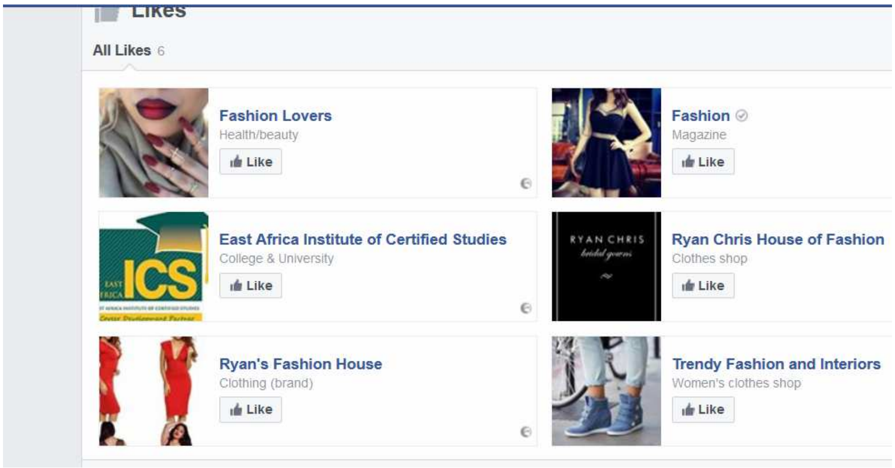
I'm Hacker - You Know Its A Joke
Artist

Hackers University
Organisation

IT Programmer
Computers/internet website

Figure 4.32: Wilf~ Facebook *Likes*
Source: *Research data (2017)*

A similar Facebook page of respondent velo~ is depicted in Figure 4.33. This seemingly indicates that this actor is a fashion fanatic. All the items on this page are linked to vogue related stuff. Therefore, this actor can easily be found in fashion and design shops sampling the latest arrivals. The findings closely relates to Bradbury (2011) that a sequence of posted contents, especially images can be analysed to reveal the behaviour and activities of individuals online as well as how they interact with others.



Likes

All Likes 6

Fashion Lovers
Health/beauty
Like

Fashion
Magazine
Like

East Africa Institute of Certified Studies
College & University
Like

Ryan Chris House of Fashion
Clothes shop
Like

Ryan's Fashion House
Clothing (brand)
Like

Trendy Fashion and Interiors
Women's clothes shop
Like

Figure 4.33: velo~ Facebook *likes*
Source: *Research data (2017)*

Figure 4.34 is a Facebook music page of actor deno~ . On a cursory look of this page, one would quickly understand that deno~ love hip hop and rhythm and blues musical genres. This findings agrees with Zambri (2015) observations that most online users’ behavioural patterns can be inferred from what they like as a person, their preferred music, movies or games.

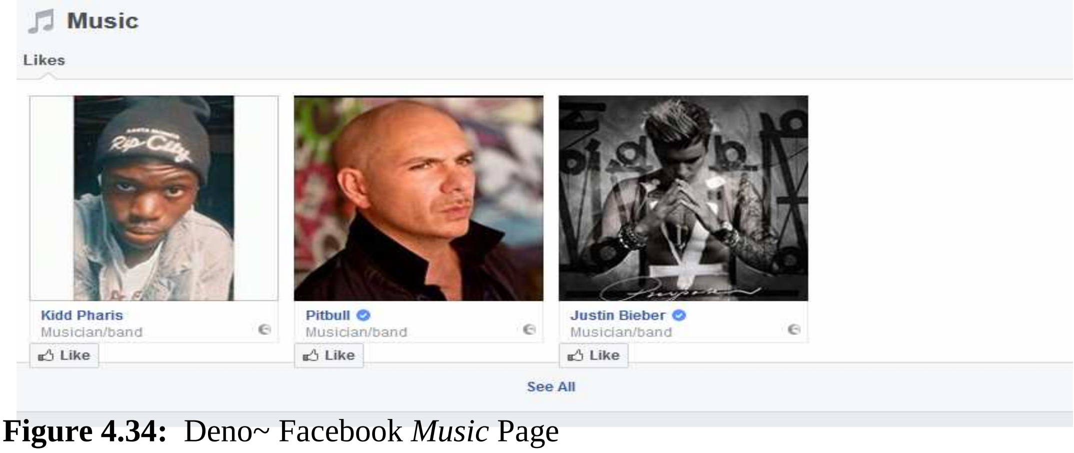


Figure 4.34: Deno~ Facebook *Music* Page

Source: *Research data (2017)*

In an attempt to demonstrate other side of some actor's multi-faceted behavioural inclinations, the researcher once more harvested more information about actor wilf~. As can be seen in Figure 4.35, besides being a hacker, actor wilf~ also likes sports, notably athletics. This implies that perhaps he was once an athlete, or he is a part-time athlete, or he mingles in social places with these sports men and women. This actor is also a football fan of Arsenal team. These results agrees with PewResearch Center (2014) that as people connect, like, follow, friend, reply, retweet, comment, tag, rate, review, edit, update, and text one another they form collections of connections. These set of connections develops into network formations that can be mined, investigated and the results depicted using various ways and techniques. The result can give a new understanding of into the structure, size, and key positions in these networks. Sometimes, individuals share photos or videos which may identify other hitherto unknown users whenever they tag (Robbins, 2011).

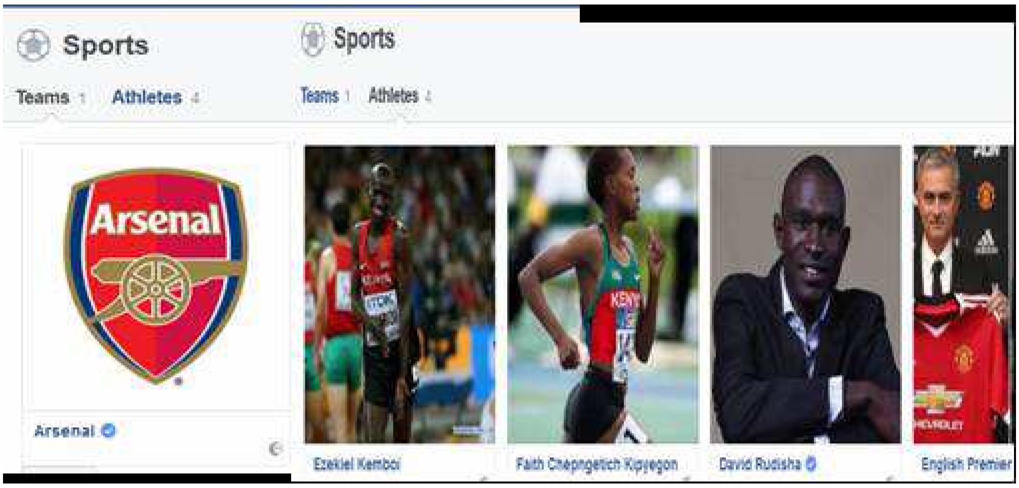
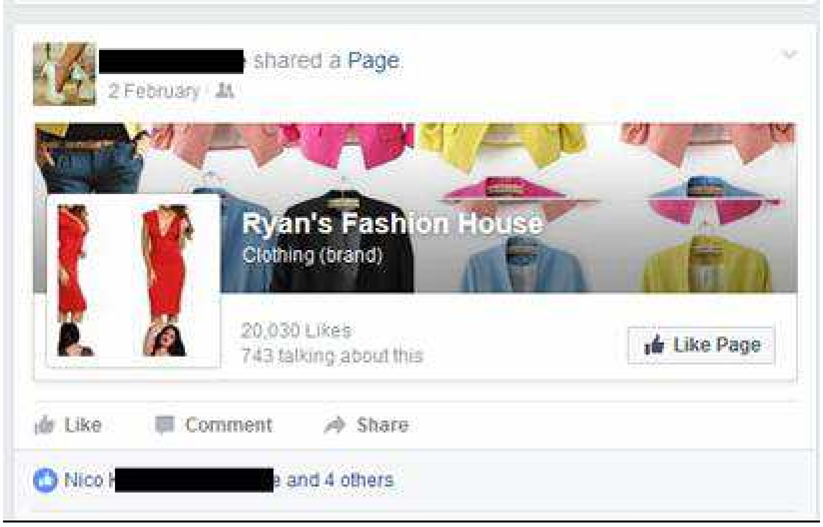


Figure 4.35 : wilf~ facebook sports
Source: *Research data (2017)*

4.4.5 Shared and Posted information

Figure 4.36 shows actor velo~ shared page to her friends. The page evidently displays brand of clothes in a particular fashion design shop. This post attracted a whooping 20,030 likes and other 743 followers talking about it. Rice and Parkin (2016) underscored that sometimes, unscrupulous social media user can post information online that act as a harbinger of his/her intention, which if captured on time by investigators; they can intercept and prevent the act from taking place. As a golden rule, speed is imperative in gathering evidence immediately after a crime has been committed to resolving a crime. In a social media platform for instance, this information may comprise of harvesting what is posted or shared.

Thus besides knowing the behavioural pattern of this actor, the posted information can enable the investigator to better know what she does online and with whom so as to reveal their conducts on how they interact. The content that people posts on their social media accounts either inform of text, pictures or similar stuff, can assist the investigators to know the actor question better. This way, a detective will be in a position to know the activities. This agrees with Wright (2010) who underscored that posted or shared information can be a rich source for forensics investigations because a simple investigation might entail viewing just the publicly-available text and images posted on a suspect's social page. The contents posted on social media bears a seriously legal risk, regardless of whether a poster is a genuine owner of the account or a pseudo-account. Thus as many individuals increases in usage of online social media, the content posted for the public domain plays a vital function in investigations and lawsuits. For instance, the content posted such as photographs can be used as evidence in a court of law.



[redacted] shared a Page

2 February

Ryan's Fashion House

Clothing (brand)

20,030 Likes

743 talking about this

Nico [redacted] and 4 others

Figure 4.36: Velo~ Facebook shared Page

Source: *Research data (2017)*

Besides being a clothing fashion fanatic, actor velo~ noticeably seems to be a reader of some genre of particular story books. Such information about the actor in question is crucial to an investigator in having intelligence of what type of literature this actor reads.

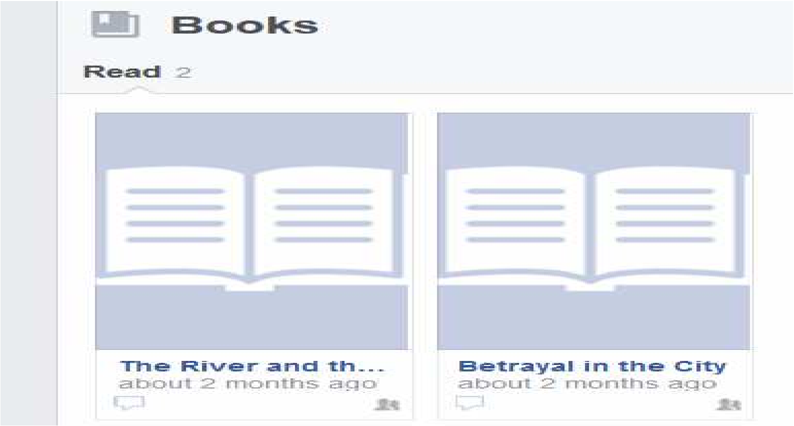
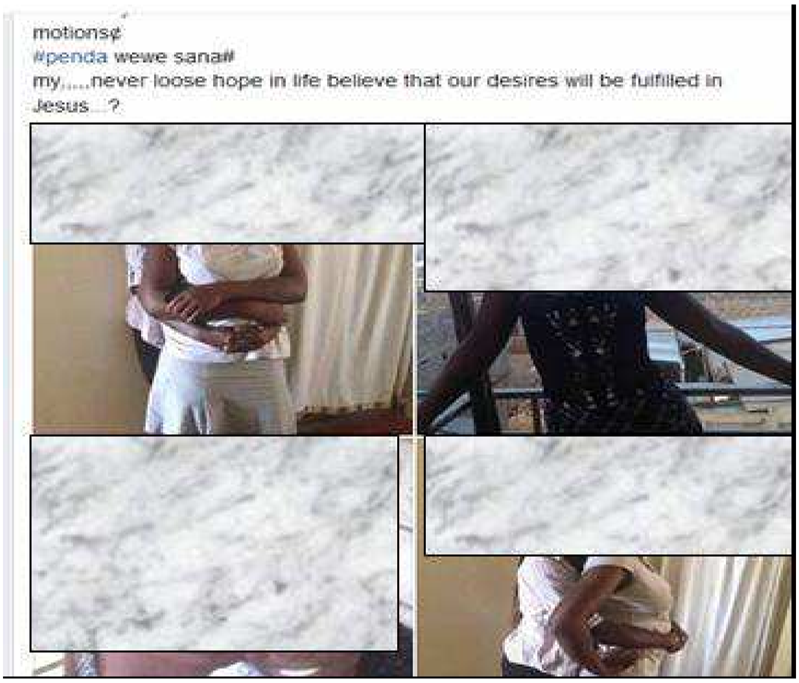


Figure 4.37: Velo~ Facebook Books Page
Source: *Research data (2017)*

Figure 4.38 shows Nico~ facebook timeline post of some words accompanied with photos. The posted message and its associated photos indicates that the kind of relationship existing between the actor and one of the characters in these pictures is probably a romantic one. Clearly, one individual appears in all of these photos. Some uploaded or tagged photographs associated to a particular node can give hints on what that individual loves, places he has visited, people he has been with and activities they did together (Vicente *et al*, 2011).

In his insightful advice, Payne as (*cited in Rice & Parkin, 2016*) underscored that investigators ought to move with swiftness in harvesting information posted by the suspect preferably earlier than when the crime is committed. This reflects with Witnov (2011) claim that most social media users knowingly or unknowingly share private content, such as personal information or photographs, leaving a trail for investigators Social network platforms is a fertile ground for all sorts of criminal acts given that platform is easy to use and provide anonymity to the criminals. Most social media postings have been used in litigation as a court ready evidence against the suspected individuals who posted the contents (Robbins, 2011)

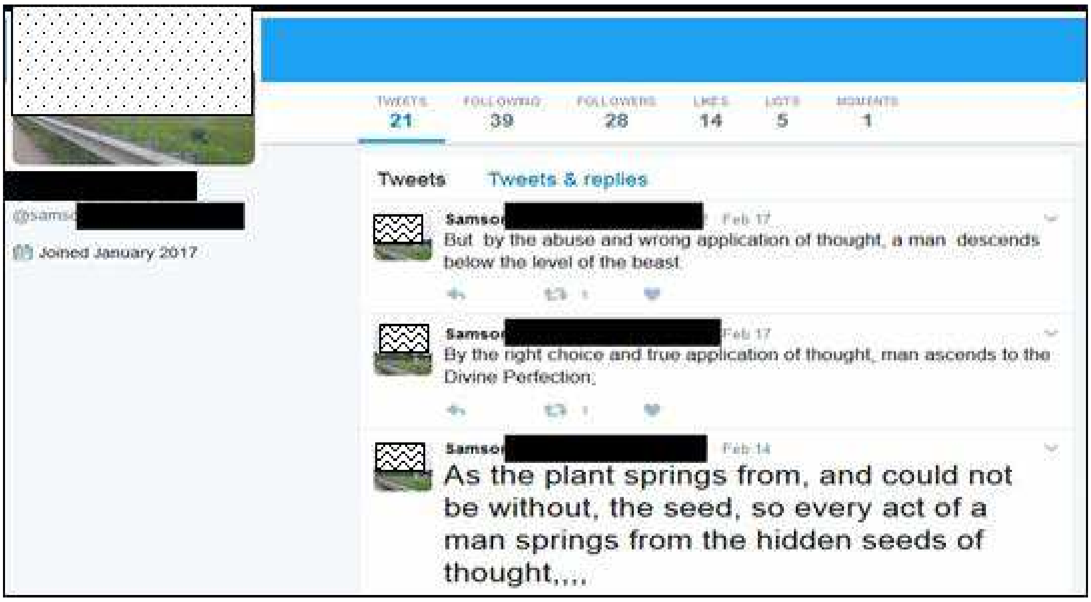


motions¢
#penda wewe sana#
my,....never loose hope in life believe that our desires will be fulfilled in Jesus...?

Figure 4.38: Nico~ facebook timeline post

Source: *Research data (2017)*

Last but not least, Figure 4.39 displays a section of tweets by actor sams~. The tweets plainly appears to be inspirational in nature. Accordingly, this actor could be either a motivational speaker a preacher of some sort and more likely to be attached to some denominational movement imbuing doctrines of virtue to its members. This confirms Vercellone-Smith, Jablokow and Friedel (2012) observation that most individuals like posting their daily activities unwittingly which ultimately disclose their behaviour and movements online. The content of people's posts such as the text they write, what it says or the content of their photos and videos, and the ratings they assign can be a crucial lead investigators can find on social media. Thus the content of the posts alone, where people detail their thoughts, feelings and ideas reveals what they are doing, what they care about, who they interact with, and why. By looking at the content of the posts people are making, one can uncover a lot of leading intelligence information about their actions. For instance, by harvesting and analysing what one is posting, it can help law enforcement to concentrate their investigation on pivotal nodes with negative influence on the network and be able to envisage their next moves.



[redacted]

@sams[redacted]

Joined January 2017

TWEETS	FOLLOWING	FOLLOWERS	LIKES	LISTS	MOMENTS
21	39	28	14	5	1

Tweets Tweets & replies

Samso[redacted] Feb 17
But by the abuse and wrong application of thought, a man descends below the level of the beast

Samso[redacted] Feb 17
By the right choice and true application of thought, man ascends to the Divine Perfection;

Samso[redacted] Feb 14
As the plant springs from, and could not be without, the seed, so every act of a man springs from the hidden seeds of thought,,,,

Figure 4.39: sams~ tweets
Source: *Research data (2017)*

In essence therefore, the information harvested above demonstrates how helpful information posted or shared on social media platform can aid law enforcement officers to track and apprehend or disrupt an individual or group of individuals planning or already executing heinous acts. Regardless of whether sharing is broadened or focused, every social media technology allows for the spontaneous creation and sharing of information (Williamson & Ruming, 2015). In multi-faceted ways, the information mined and analysed so far concurs with Blomberg (2012) that by harvesting and analysing data from social media accounts of individuals, investigator can detect signs and perceive any atrocious activity that is about to be committed. A similar scenario was highlighted Rice and Parkin (2016) in which detectives utilized the posted information to apprehend and charge the individual under investigation, in USA.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS.

5.1 Introduction

The study was carried out on respondents of five selected universities in Kenya namely Rongo university, Moi university (annex – school of law), Kisii university (Eldoret Campus), University of Eldoret and Kibabii university. The purpose of this study was concerned with investigation of selected egocentric users (*up to 2.5 degree*) on social media platforms using Social Network Analysis in mining forensic evidence that can be used by law enforcement agencies in Kenya.

Chapter one provided the background information to the study, an overview of social media usage issues or factors that have considerable impact in investigating and mining invaluable forensic evidence that have hitherto not been researched here in Kenya. The chapter also expressed the view that the researcher pin-pointed as the security gap in Kenya's law enforcement landscape notably on applying Social Network Analysis techniques and the appropriate software tools to harvest important information pertaining individuals under investigation . It was on the basis of this background that the statement of the problem was stated, objectives and significance of the study outlined. Research questions, justification of the study and the delimitations of the study were also highlighted.

Chapter two highlighted literature reviews from authors who have researched on various methods of applying Social Network Analysis on social media to map and visualize as well as harvest other important information from the identified actors in a given social media platform. The chapter singled out a researchable niche that has hitherto not been studied in

Kenya as far as application of Social Network Analysis on social media to mine forensic evidence to aid law enforcement agencies.

Chapter three was concerned with the methodology employed in this study. Social Network Analysis was used to design how selected online users data was to be mined and analysed to derive important information for law enforcement.

In chapter four, online primary data was mined from respondents (including their aliases) and analysed using NodeXL software version 1.0.1.373. Social Network Analysis descriptive statistics consisting of centrality metrics, visualizations and mapping of influential nodes/actors and harvesting of demographic information related to one's online profile, time and locations data, shared or posted information were used in the analysis to report and present the findings.

5.2 Summary

As stated earlier in this thesis, the main objective of the study was to demonstrate how investigation can be carried out to mine forensic evidence that can aid law enforcement officers to take appropriate action based on the findings. The study used respondents from five selected universities in Kenya. In particular, the study sought to attain the specific objectives as discussed below.

The first specific objective of the study was to answer how *to visualize social networks and clusters to uncover the patterns of the social relationships of people in investigating crimes committed over selected popular social media platforms in Kenya.* In quest of answering the above objective, data from the respondents were mined and analysed to generate visualizations and thereafter examine the graphs for nodes with conspicuous and high graph metrics as well as presence or absence of clusters in the network.

To begin with, the study began when the "trigger" actor sams~ created a pseudo-account in twitter and Facebook social media platforms and thereafter invited the respondents to join by either following or friending him. After a month, 55 respondents out of 94 had joined the social media platforms as directed to them by the researcher. The social media data of the aforementioned respondents was mined and used to generate a 1- degree egocentric network using Harel-Koren Fast Multiscale layout algorithm. The seed respondents used in the study comprised of a section of respondents from the aforementioned universities and the rest of the respondents joining the network were not be controlled by researcher. As mentioned earlier in this thesis, the research began when the seed actor creating a pseudo-account on selected social media platforms and thereafter invite the respondents of this study to connect with him.

As time elapsed, subsequent mining of respondents data was done and the visualizations generated using Fruchterman-Reingold force algorithm depicted interesting findings. Visualizations empowers Social Network Analysis specialists to discern unseen social network structures and patterns the nodes (Tayebi & Glässer, 2016). Generally, the network generated was 1.5 degree egocentric network. The network had been fragmented into subgroups or communities. The detection and visualizations of communities helps the speedy understanding of the functionalities of the network as well as illustrate the connections between the actors (Johnson & Reitzel, 2011). Visualization functionality aid the law enforcement agencies to unearth hidden intelligence such as clusters, identify central nodes and give better understanding of network structure by utilizing limited information from a large dataset from a complex network (Yang, Liu, & Sageman, 2006).

More specifically, there were three main communities. In each community, the visualizations evidently displayed three main actors as velo~, deno~ and wilf~. This implies that the three above mentioned actors scored the highest centrality measures than other actors in the network. In particular, respondent velo~ noticeably scored the highest in degree and closeness centrality measures in both the entire and community network. The findings corresponded with Sharma and Strategy (2008) that the closeness centrality score shows nodes are contiguous to the many other nodes in the network either using direct or indirect links, and moreover indicates the node who has access to classified first-hand information the network. Thus, the actor (velo~) is evidently the most influential and very important node in this network so far, followed closely by wilf~ and deno~ respectively in that order. These three actors are capable to influence the rest of the actors in the network. This conforms with Freeman (2004) observation that the patterns of social connections between actors rooted in a particular network or community, has significant impact for those nodes.

In the 1.5 degree egocentric network, there were no peripheral nodes. All of them were linked to their most central actors in their respective communities. It is also important to mention that the findings in 1.5 degree egocentric network, actor sams~ has the highest closeness and eigenvector centralities because he is connected to the three most influential actors in the network. He also has high betweenness centrality by virtue of connecting other actors to very important actors. A node with highest betweenness suggest not only the crucial position the actors has in the network but also the great power that actor has in controlling the dissemination of information as a bridge to other communities of the network (Berzinji, Kaati

& Rezine, 2012). This concurred with Rice and Parkin (2016) that when investigating egocentric users on the selected social media platforms is pertinent because they comprise of social components such as individuals or clusters and underlying social ties such as friendships, or kinfolks and assist in understanding the structure, nearness and density of individual to individual or cluster to cluster social spheres. With regards to such findings, Golbeck (2015) advises that when carrying out an investigation, actors who are in strategic positions as seen by their highest centrality scores, tend to hear more information from friends of friends and thus act as crucial informants to the detectives. In a rejoinder, Xu, Marshall, Kaza and Chen (2004), highlighted that with such measures and techniques on is able to detect and tell the transformations in the attributes of a particular actor in the network, his/her position and ultimately comprehend the dynamism of structures amongst members of the network.

Besides visualizing and identifying centrality measures for crucial individuals in the network, there was need to also depict characteristics of every group/community. Visualizations reveals unobserved structures within social network and relations as well as patterns that exist amongst the nodes of that network (Blomberg, 2012). As a result, Wakita-Tsurumi algorithm was employed to depict the density of each cluster. In total five clusters were generated. The findings indicated, the cluster of actor wilf~ swelled in its membership significantly and therefore its density was the highest in the entire network. These implies that the members of this cluster do communicate or interact more frequently under influence or leadership of actor wilf~. Moreover, the members of this cluster could be knowing one another or they are talking about a familiar topic that excites the entire camaraderie.

Overall, the least denseness was exhibited by the cluster belonging to actor deno~ According to Xu, Marshall, Kaza, and Chen (2004), a highly dense group is more susceptible and exposed to law enforcement officers for further scrutiny and identification of the main actors who are most likely to be the leaders of a particular cohort of felons. They reckoned that if the density of various groups keeps on fluctuating from high to low or vice versa, then it implies that the groups are competing for leadership positions. Moreover, identification of gatekeepers connecting to a particular subgroup (community) is also important in unearthing specific dubious characters.

In endeavouring to clearly visualize more cohesive groups of nodes that are closely interconnected, the researcher visualizations of clusters grouped according their clusters. The results generated six distinct interconnected clusters. Evidently, cluster with highest degree centrality and density were respondents under the leadership or influence of actor wilf~. The higher the centrality score of a node, the more vivacious that actor is in the network (Ergün & Usluel , 2016). Ferrara, De Meo, P., Catanese & Fiumara (2014) succinctly highlighted that visualizations showing simulated features of a criminal network is the use of clustering. In Himelboim, Smith, Rainie, Shneiderman and Espina (2017), nodes within a densely connected clusters are likely to have same attributes, a concept known as homophily in social theory. Apparently, the membership of the network has now increased more significantly.

Worth mentioning is the fact that few members of actor wilf~ cluster fragmented and formed a new cluster. In this, we deduce that there could have been some misunderstanding or leadership position struggle in the original cluster, which made this new cluster to secede and reorganize their community. One cluster was visibly disconnected from the entire network but was getting information flow using two members only. These two members therefore are said to be brokers or gatekeepers because they have high betweenness centrality. If they are deleted, the network will lose an isolated cluster. This agrees with Hansen, Shneiderman and Smith (2011) observation that in any given network, members tend to cultivate multifaceted connections and therefore such individuals are potential members of many other groups(clusters) in that network.

The findings above synchronizes with Zhu, Watts and Chen (2010) sequels that clustering network communities is important as it facilitates detectives to know specific persons who belongs to a particular cluster. Armed with such information, a law enforcement officer can narrow down the list of suspicious characters under investigation. Thus clustering networks into communities enables detectives to identify specific subgroups and if an individual belong to a certain subgroup commits a crime or he/she is a suspect, it will help the investigators to limit their probe to that community.

The study recognized a need to depict and present the relationships between actors in order to show their synergy and distinct properties in the entire network. It against this backdrop that centrality analysis of visualized nodes was employed in this study to depict the most central nodes in the network so as to discover their rank, importance or influence they have in the network. Accordingly, the researcher made an effort to visual the main actor’s degree distributions in order to exhibit their possible roles in the network. Thus, the degree, betweenness, closeness and eigenvector centrality metrics were computed. The results revealed that actor wilf~ was the most significant actor in the entire network having evidently scored highest betweenness centrality. This indicated actor wilf~ was not only the most influential person in the network, but also he seemingly knew the on-goings in various clusters because he is well connected and therefore acts as a gatekeeper of his cluster linking them to the rest of the network.

As pointed by the centrality metric of a network, existence of a small number of hubs having many connections portrays a network free of scales and has the power law degree distribution (Van der Hulst, 2009). Nonetheless, Zhu, Watts and Chen (2010) cautions that degree centrality might not be an actual pointer to the real ring-leader position of a given actor in a network and advises that law enforcement officers need to go a notch higher gather more intelligence regarding the degree centrality of the actor under investigation.

Actors that closely followed suit in scoring high centrality measures pictorially included actors nico~, velo~, 2279837eb26d4a1, denokisaka and kiptal~ in that order hierarchically. This implies that these main actors are leaders or hubs of their respective subgroups in the social network. The visualizations also depicted other important actors in the network who may not score higher in other conspicuous metrics. For instance, actor samso~ clearly has high betweenness centrality metric since he is connected to all the influential or important actors in the network. This implies that actor samso~ is informed of the on-goings in various clusters of the network he is connected to. Thus, these findings suggest that a actor with high degree centrality signifies leadership role in the network whereas an actor with high betweenness centrality is a connotation of a gatekeeping role in one’s cluster linking the rest of the network. This conforms with Xu, Marshall, Kaza and Chen (2004) that actors with higher betweenness centrality metrics are valuable to law enforcement officers because they provide a smooth and subtle link to the important actors and also in collecting helpful information from the entire network.

Besides identifying those who hold the echelons of power and influence in the network, the study also sought to find out actors with structural similarity. This was achieved by determining actors connected to other more influential actor(s) in the network. The inference made from this type of relationship is that if two or more actors have same friends, then it suggests that all of them are also friends in the real world. The results indicated that except for actor wilf~, the other main actors namely nico~, velo~, 2279837eb26d4a1, deno~ and kiptal~ depicted similar structure formations in their respective clusters. Nodes in a social network can be classified on the grounds of having same position or roles in that cluster or the entire network (Tayebi, & Glässer, 2016).

However, social network analysis experts advice that degree centrality may not be a true indicator of ring-leader role of an actor in a given network. Identifying a person who has high degree centrality, hence many connections to other individuals in the network is crucial to law enforcement agencies but detectives need to be cautious about this measure. Perhaps the actor in question has least intelligence in the cluster. Thus more intelligence regarding the nature of degree centrality of a particular actor under investigation need to be explored to assist the investigators (Zhu, Watts & Chen, 2010).

It was also essential to depict the pattern of information flow between actors in the network. By use of a geometrical spiral algorithm, the research clearly demonstrated the overlapping relationship of flow of information between two clusters at the nucleus of the network. One of the two central clusters had its members fragmented and spreading the outer ring of the network. The findings also showed one different group seemingly isolated from the rest of the network. This group looked like it was either recipient of the information from other clusters in the network or perhaps they knew some information only classified to themselves. The overall results revealed that blue, cyan green, red and orange- coloured subgroups were the most dominant communication actors over others in this network. For an investigator therefore, an investigator can narrow down his/he probe to these prominent clusters to gather more important intelligence and reveal their activities. Diffusion of information flow within the network was therefore imperative to be depicted to show how actors communicate as information cascades from one cluster to another. In any network communiqué and movement of information of a covert network group, the chain of command is always confined the clusters (Hopkins, 2010).

As a furtherance of depicting network information flow in a concentric manner, the researcher endeavoured to portray luminous visualizations of communication channels existing between different clusters/subgroups of the network. The visualizations findings showed significant channels of communications between various clusters, in which influential actors with high centrality measures conspicuously becoming important “bridges” between the subgroups. Several clusters’ channels of communication denoted that there were various “leaders” or ‘hubs” in their respective subgroups. The findings concurred with Granovetter (1973) that specific actors are crucial when analysing organized clusters or sub groups of a network in order to detect trust ties between pairs of nodes and also detect the strong relations linked to the influential actors in that network.

Interestingly, the research outcome showed a considerable number of “isolate” actors in various clusters of the network. The communication channels depicts interaction patterns and how the information flows amongst the actors of the network, and it also shows the nodes dominating the communication in the entire network. The complexity of the network became evidence with time as membership increased exponentially resulting in 2.5 or more degree network. Valuable information are usually inferred from the visualizations whenever there is a clear depiction of information flow and the relations between network actors (Berzinji, Kaati & Rezine, 2012).

However, at cluster level relationship, the network fragmented into elite networks or clique of actors. The 2.5 degree egocentric network that was generated was employed to establish whether the prominence of the main actors identified by the results still remains static or have changed. The findings showed that roles of some actors have changed as new actors emerge with time. This research outcome was similar with that of Gunnell, Hillier and Blakeborough (2016) who conducted study by utilizing the available police intelligence to unravel what Social Network Analysis can reveal about criminals and establish how useful the social network analysis outputs were to the police. The research findings revealed an overall network of 137 individuals were identified, from the starting point of five (5) individuals identified as having gang links. This information was sought by the research to demonstrate its importance to an investigator as it can be used to determine if there is still a healthy communication amongst the network members or leadership roles have been changed and rebel groups evolving. The results reported a total of 22 cliques in number.

In order to evade the problem that comes with having several cliques in the network, Mainas (2012) advices that there is need to completely do identification of the linked cliques in the network. The implication here is that in each clique of the network, there must be some classified information that is only privy to each member of these exclusive subgroups. Such vital intelligence is important for an investigator while carrying out a covert analysis to unearth vital criminal –information for the actor(s) under investigation. A similar tone was highlighted by Xu *et al* (2004) that a criminal network is always dynamic and continually undergoes metamorphosis as time elapses whose effect is this the nodes turnover in the network, where other new nodes join while others are leaving the network; whereas other clusters could fragment further, others also amalgamate resulting into a complex network.

Moreover, it is important for an investigator to monitor and profile the changes of actor(s) in question by analysing the personal attributes and network roles they have, and in doing so, a detective can comprehend the fluid patterns of actors or clusters under investigation. This way, such information can assist an investigator to tell if there is still a healthy communication or leadership roles have been changed and splinter groups emerged.

Borrowing from Granovetter (1973) principle of weak ties, the study concludes that the crucial channels of communication to be closely monitored are the ones that are rarely utilized and usually located at the network's periphery and also comprise of quite dense cliques. Krebs (2002) resonates that there is a strong relationship in weak connections as demonstrated in a secret social network memberships where there is little activation rate amongst them.

In a nutshell, the findings of network visualizations showing the dynamics, patterns and other attributes of the individual actors as well as the overall network, underscores how valuable Social Network Analysis is to law enforcers in investigating, mining and visualizing leading information that which could otherwise have been difficult or impossible to interpret using conventional methods of investigations in Kenya.

The second specific objective of the study was *to compute centrality metrics, PageRank and clustering coefficients in quest of identifying the most important actors using the selected* *network user's data in Kenya.* Accordingly, the researcher endeavoured to compute degree centrality measures of all actors in the network to determine the actors connected to many friends. The more the connections an actor have, the higher the degree centrality score metric that actor has over others in the network. Betweenness centrality was computed to ascertain the actors that act as bridge connecting one cluster to another in the network whereas closeness centrality was used to show actors at the epicentre or nucleoids of the network. Eigenvector centrality, just like PageRank, was likewise computed to establish the highly influential actors in the network. Moreover, clustering coefficient metric score was also generated to determine how closely related each actor's connected to others were in the network.

At the primary stage of this study, the researcher computed the initial centrality metric scores of the seed respondents corresponding to 1-degree egocentric network visualizations findings. Thus, the preliminary results of the seed network actors consisting of 55 vertices registered a graph density of 0.023 implying existence of dyadic connection between the seed actors to the main actor. It is important to note that network density scores range between 0 to 1. Himelboim *et al* (2017) observed that the higher the interconnection between the actors , the higher density of the network and vice versa. Network density reflects the overall ties in a network by associating the current number of connections against the theoretical probable figure of connections between the overall members of the network (Wölfer, Faber & Hewstone, 2015) .

As a way of detecting patterns of change in both actors and the entire network, the researcher allowed a considerable amount of time before mining data from the actors. After two months, data harvested from the network actors reported 483 vertices up from the initial seed actors of

55 vertices. The graph was directed since data was harvested from twitter accounts of actors under study. The ultimate complex network structure obtained is closely related to the study undertaken by Gunnell, Hillier and Blakeborough (2016) where they used five nodes chosen as focus group and ended up with an overall network of 137 nodes.

The metric reciprocated edge ratio of approximately 0.3 hinted the presence of important network structural properties considering the social interactions between actors are asymmetrical and not essentially reciprocal. Thus reciprocated edge score above suggests the degree of importance of various nodes in this network because they are possibly connected to influential persons in the network. Similarly, the results also indicated that all actors were inherently connected as indicated by a single vertex score of zero (0) of connected components.

The existence of clusters as reported by visualizations results was supported by the average geodesic distance metric score of 3.2. The implication of this score is that, the entire network was not fragmented and also the nodes in this network did not know one another directly. They only did so through an average of three (3) friends as the shortest path between actors in the network. General linkage amongst the nodes in the entire network indicated a slightly loose interaction between all actors of the network as indicated by the density score of 0.003. This score implied that the information in the entire network was not effectively disseminated. Hopkins (2010) reinforces these findings as he affirmed that regarding the communication and flow of information in clandestine network, the chain of command is always knitted to their clusters which are exhibited by an average geodesic length and a highest clustering coefficient. Hence, the nodes who depended on other nodes to get information may not have been connected well. Duijn (2016) states that the network density score paints a picture on rate at which information is circulated between the nodes of the network. However, Demiroz and Kapucu (2012) observed that a quite low density score characterises the main attribute of a secret network.

Suspicious characters do link with any member of a network but only through proxies making the network to be sparsely connected with low tie density hence mitigating risk of being detected and at the same time enhance improved communication in the network. Therefore, the higher the clustering coefficient the higher the local communication efficiency is in a given cluster and the entire network in general.

However, in measuring the strength partition and robustness of the network, the modularity score of 0.5. The score closely relates to Newman and Givran (2004) that the modularity score of a network indicates the qualities of clusters in that network. This indicated that the connection between nodes in clusters (modules) of the entire network was slightly high, despite the low the overall density score. Drawing from the above modularity score therefore, the study concludes that the dynamics and structures that existed in different clusters of the network allowed a fair rate of spreading information amongst the loosely connected members of the entire network. Himelboim *et al* (2017) highlighted that networks exhibiting low modularity scores but have higher densities signifies a cohesive cluster.

In order demonstrate the actor with more network connections than others, degree centrality was computed by the researcher. Top on the popularity measure was actor wilf~ with a degree centrality score of 196 links, far much ahead of the second and third most popular actors nico~ and velo~ with 93 and 82 connections respectively. Evidently, the results indicate that actor wilf~ was the most popular actor in entire network and draws a lot of attention as far as communication is concerned. Thus wilf~'s focal point in the network communication is attributed to the many number of ties indicate his degree centrality score. Hence, actor wilf~ served as a hub in the network and so the most influential node. Ferrara *et al* (2014) pinpointed that an actor with the highest degree centrality score in a suspicious network usually takes the pivotal position of leadership, giving commands, rules or essentially ensuring that the information flows effectively well in that dubious covert network. In a similar tone, Van der Hulst (2009) underscored that the higher the degree centrality score of a node is, the more significant that node is to the entire social network since he/she likely to be signifying a focal for information and resource flow inside that network.

Betweenness centrality metric measure was employed by the researcher in order determine the node who acted as the most direct path or bridge between different clusters of the entire network. Analogous results to degree centrality was generated whereby actor wilf~ scored the highest betweenness centrality score of 69302.912, followed by actor nico~ and velo~ with 37752.733 and 31198.671 respectively. Nodes scoring highest betweenness centrality values are powerful in the network (Berzinji, Kaati & Rezine, 2012). Judging from these results therefore, actor wilf~ is the most important node in this network as far information flow is concerned. Hence actor wilf~ is a crucial bridge of reaching other actors he is connected to. This implies that if actor wilf~ is cut from this network, a lot of disruptions is bound to happen in the entire network. This findings agrees with Krebs (2002) that a node with high betweenness centrality will markedly have a lot of influence over other nodes regardless of whether it is centrally or peripherally positioned in the network.

The researcher sought to establish the actor(s) who were more nearer to all other actors in the network. Interestingly, several of the top actors scored a similar closeness centrality score of 0.001 implying that the members of the network were almost directly connected to one another. The results also suggested that majority of network actors, notably the actors with various high centrality metrics were able to reach another fast across the network. An actor with the highest closeness centrality score is quite familiar and aware of events or happenings within his/her network (Hoppe & Reinelt, 2010). In a rejoinder, Ferrara *et al* (2014) contextualize closeness centrality in a covert and suspicious network that this score exposes nodes that are closer to many other actors in the network and can quickly pass information to anyone in that network.

In order to ascertain the most influential actor in the network, eigenvector centrality was computed and once again, actor wilf~ scored highly with an eigenvector value of 0.057. Two actors deno~ and velo~ tied at second position with an eigenvector score of 0.013. Thus despite registering diminutive score in degree centrality, deno~ proved to be an influential actor in the network nonetheless. Hansen, Shneiderman and Smith (2011) advises that nodes that have a small number of connections can still score a high eigenvector centrality score if they are connected to other well connected or influential persons in the network. In this regard therefore, deno~ must be connected to influential people in this network, hence the highest eigenvector score. Other actors such as nico~ who featured prominently in other centrality scores scored very low eigenvector centrality values indicating he was not connected to other influential nodes in the network. These findings resonated with Nouh and Nurse (2015) supposition that a node with high eigenvector centrality metric is one that is adjacent to other nodes having high eigenvector scores in that network. Eigenvector score of an actor indicates the significant roles and status of his/her network neighbours (Arnaboldi, Conti, Passarella & Pezzoni, 2013).

Nevertheless, PageRank quality metric measure was computed by the researcher as a summative measure of evaluating and comparing values eigenvector for different actors. Intuitively, PageRank is employed to measure the node(s) connectivity in the weighted clusters of the network. Actor wilf~ still retained his top position with a PageRank score of 81.686 and followed by actors nico~ and velo~ with 40.995 and 33.346 in that order. These results are indicative measure of the large number of in-coming connections that aforementioned actors cultivated in the network.

As a way of determining actors whose friends and possibly friends of friends knew each other, a clustering coefficient was computed. The analysis results showed that unlike the previous scores of centrality metrics, the highest clustering coefficient values of 1.00 were scored by new actors namely actors brian_ba , itsdavid~ and other four actors. This implies that these first six actors scoring high clustering coefficient values were acquainted to each other almost at personal level as in real world.

Moreover, there friends also knew one another too. Strangely, the hitherto top ranking actors, such as actors wilf~, velo~ and deno~ scored dismal clustering coefficient values implying these actors were most likely not acquainted to each other and so was their friends to one another. Waskiewicz (2012) opined that despite the fact that a particular has no direct link him/her and the friends of his/her friends, nonetheless, there exist some crucial connections whereby a ripple effect of influence can be felt as it is disseminated across the network. In Ferrara *et al* (2014) higher clustering coefficient scores implies that there is a nodes interaction between and their neighbours is higher and a significant volume of information are exchanged.

Last but not least, the researcher once again generated the graph statistics of the network after significant length of time yet again. By this time, the network complexity had grown tremendously from 483 to 29,2985 vertices. The visualized 2.5 - degree egocentric network generated in the previous chapter conformed with graph statistics. More specifically, the single vertex connected components of 15 concurred with the many cliques and isolates depicted by Figure 4.2(h). Himelboim *et al* (2017) defined isolates as nodes without connections to other ties in the network and consequently impacting the network density score. The results was also supported by the overall graph low density score which suggested network fragmentations. However, there was no significant increase in the average geodesic distance. The number of messages tweeted was a record of 32,817, having 31,406 replies and 14,593 *follows*. The number of mentions was a whooping 90,361.

Corresponding to the overall graph metrics, the overall centrality measures metrics once more were computed and generated for 2.5 degree egocentric network. The findings reported a maximum in-degree score of 1828 recorded against safari~ meaning that the actor attracted (received) a lot of attention from other nodes in the network. In a similar way, the study findings also reported a maximum out-degree value scored by koinang~ implying that this actor was the most active node to send out many types of information to other actors in the network. Interestingly, the average in and out degree scores were similar values of 1.825. Wu, Carleton and Davies (2014), higher in and out degree centrality scores implies that the nodes are not only prestigious and influential in the network but also they are likely to be coordinators or overt machinists in the network.

Nevertheless, the maximum betweenness centrality score was a whopping 148,196,529.709 recorded against sams~ in the network. The findings confirm the visualizations results which depicted this actor to have higher in betweenness centrality score. In other words, actor sams~ is acting as bridge or gatekeeper between the nodes of the network and to some extent manages the information flow in the network. It is worth mentioning that betweenness is a measure of how important the node is to the flow of information through a network. In an investigation, a node with high betweenness is likely to be aware of what is going on in multiple social circles. A node with high betweenness has great influence over what flows and does not in the network. Thus it describes people who connect social circuits.

In order to demonstrate the global distance between nodes of a network, the maximum, minimum and average closeness centrality scores were compute. An overall score of 0.00 for the maximum and other closeness scores were peculiarly generated. This score implies that in the last 2.5 degree egocentric network, many nodes were isolated into cliques and consequently the actors were not close to one another in the overall network. This results may not actually imply that there is 0.00 distance between nodes, instead it suggest that actors were immensely separated in their respective clusters but merely an average geodesic distance. On the contrary, the 0.00 closeness metric score hinted the presence of numerous nodes whose information were able to quickly reach other nodes in the network (Kaye, Khatami, Metz & Proulx, 2014). Johnson and Reitzel (2011) recommended that actors scoring higher closeness centrality metrics are good for propaganda, de-campaigning or act as a means of accessing a clandestine network. In a rejoinder, Hanneman and Riddle (2005) assented that such nodes scoring high closeness centrality metrics are capable of reaching other nodes at shorter path-lengths or can easily be reached by other nodes in the network.

A probably new entrant and influential actor in the network were shown to be actor mugoki~ with a maximum eigenvector value of 0.03, implying that actor wilf~ was replaced ultimately. This indicates at “global” network, as opposed to clusters, actor mugoki~ surreptitiously has the overall influence in the network. In other words, actor mugoki~ has an average power radius of 0.03 between him and other actors in the entire network. To sum it up, ego-neighbourhood tie-strength of actors were determined using clustering coefficient. The findings reported a maximum and minimum clustering coefficient of 1.00 and 0.00 respectively. This meant that whereas other others perfectly knew other despite the increase in complexity of the network, other actors did not know one another to the end since the actors(respondents) came from different geographical regions.

The third and final objective also sought *to demonstrate how demographic/background and related information of social media users can help in tracking and zeroing to specific online lawbreakers in Kenya.* Needless to reiterate that this research objective was carried out in order to demonstrate how social media user’s background information can be invaluable standpoint for law enforcement agencies. It is also to note that in this section, the researcher utilized the advice of Vercellone-Smith, Jablokow and Friedel (2012) that in case a node under investigation cannot be easily known because he/she used pseudonyms or aliases, the sytlometric techniques was employed to identify the author based on the characteristics of the textual content.

A thorough probe was carried out to painstakingly extract demographic or background information of the selected actors who toped in various centrality metrics scores. First, the Facebook profile picture of actor nico~ revealed crucial information about him. The information deduced from the picture implied that the actor has is a Roman catholic church believer and he is affiliated to unnamed institution of higher learning here in Kenya. Despite his twitter profile picture not disclosing any crucial information about this actor, his tweets nevertheless gave an insight of what the actor likes discussing with others. Seemingly, the actor is more of a follower than a tweeter and this means he is less of leader in this network. An overview profile information of nico~ divulged more demographic data about him. The crucial information to be extracted comprised of his gender, phone number, date of birth and the current city that the actor lives in. Rice and Parkin (2016) advices detectives undertaking investigations on social media platforms to utilize their own accounts and be cautious of wrongly identifying the targeted node(s) being probed.

To cement and reaffirm the importance of gathering demographic information that can immensely help investigation in getting evidence of the characters in question, the second actor's wilf~ demographic information, was harvested by the researcher. The profile picture of this actor was very rich that it can easily help the investigators to know his real image. His overview profile disclosed the name of the university that the actor attended, the undergraduate degree course he pursued, date of birth, the current city he lives in and two sets of mobile phone numbers. Moreover, the Facebook continued to display the number of his family members. Similarly, the twitter profile of this actor still unearthed crucial lead information for investigators such as his actual picture consistent with the one in his Facebook account and his profession as well as the many number of *followers* he has. The results concurred with Golbeck (2015) research findings that social connections of any actor one's account eventually becomes detectable whenever followers, friends, likes, comments or poke information are harvested and analysed for the relationships trends and patterns.

To synopsize on how to extract actor’s demographic information that can aid law enforcement officers, the researcher summarized it with the third actor, velo~. A quick analysis of the profile picture of this actor indicated that it(she) is likely to be a female node, drawing from the feminine display of fashion trending stuff, shoes and other related paraphernalia. Her Facebook account continued to expose more background information such as date of birth, the current city (town) she lives in as well as the name of the university she attended. Her twitter account revealed wealthy of information about this actor. The profile picture could reveal her real identity, posted alongside her name. The findings above demonstrate how numerous details pertaining to a particular node under investigation can be harvested to a point of revealing their true identify. This corroborates with Flynn (2002) that actors in a network can easily be identified by use of bounty demographic information consisting of names, dates of birth, phone numbers, relationships status, level of education and other crucial information even if they use pseudonyms or aliases

The second approach employed by the researcher to demonstrate how to harvest crucial information that can aid law enforcement officers to generate forensic evidence against social media criminals was achieved by the use of social media connections and related friends. Once again, this was accomplished by critically analysing and extracting such information from actors who scored highly in both visualizations and centrality metrics.

An analytical look of nico~ twitter account revealed that this actor has followers that are football fanatics indicating that he is likely to a football fan too. One of his great follower appears to be a reggae music fan. Actor nico~’s Facebook profile noticeably shows the he has 69 friends. Digging deeper into a section of these friends suggest that they are influential people, going by the number friends of these friends. For instance, two of his friends have

302 and 1,276 friends respectively. These results closely relates to Waskiewicz (2012) that in the ego-centric friend of a friend dichotomous relations and influence can be felt up to 3 degrees. As an investigator, such actor can be deemed to be either in possession of crucial information or he simply likes associating with celebrity figures in the society.

.

The social connections of actor wilf~ yielded bountiful lead information that can that can used by investigators to discern the character traits of this actor. First, the actor boasts of 1449 Facebook friends some of them seemingly are holding ranking positions in various institutions. Actor wilf~ has so far uploaded a total of 55 photos in his Facebook account, some of which were uploaded using the mobile phone. A thorough analysis of such photos can reveal the calibre of persons this actor relates to, especially those that appear many times in these photos. In emphasising the concept of homophily, Himelboim *et al* (2017) paraphrased that nodes with the same interests and background information tends to form network connections. Moreover, another way detectives can get to know the loyal friends of a particular is to check those who *liked* or *commented* the posted photos and frequency they did.

The third perspective of investigating online suspect(s) and gathering forensic evidence from them was demonstrated by use various techniques or tools for mapping and revealing time stamps and location data of users under probe. The researcher demonstrated this concept by harvesting and analysing the actor's online posts or using specific software to obtain embedded information that are automatically fixed whenever photos are captured using a particular device.

More often, users are prompted to supply location information when creating their online social media accounts. However, this is an optional requirement and may not be demanded by all social media platforms. In view of that, when actor velo~ twitter profile page was scrutinized, it revealed her location information which was harvested showing the town/city she hailed from in Kenya. These revealing outcomes matches Buccafurri, Fotia and Lax (2013) that if a suspected person claims to have been in some place when a particular took place, the investigators can harvest information from his social media accounts and analyse the location of the device at the time of the incident to be used as an alibi. Apart from extracting location information of the actor, if at all it exists, the researcher demonstrated how time stamp of information posted online can be used to get the exact time it was posted. On a twitter account of any actor, one can point (*using a pointing input device*) over the date of the message was tweeted or retweeted and it will pop the exact time the message was posted. It is important to note that any content posted a social media user contains time-stamps and this information can be used to reconstruct the time line of such posts to show itinerary of a person under investigation against his/her alibi. Thus to an investigator, such information can reliably act as forensic evidence against a particular offender in a court of law.

In circumstances where an investigator cannot obtain location or time stamp information, he/she can try other avenues of obtaining similar type of information. To demonstrate this concept, the researcher downloaded a photo posted by an actor and then went ahead to obtain Exchangeable Image File Format (EXIF) metadata of that photo. The metadata information of the photo includes GPS coordinates which can then be used to know the exact location where the photo was taken as well as the device used to capture that picture. The ensuing process entailed the use of a free online Exif viewer by the researcher to extract the photo metadata. A wealth of intrinsic and interesting information was generated. The results comprised of auto-encoded GPS location of where the picture was captured, the date and time when it was also taken. The device used was an Ipad from Apple Company. The same Exif viewer software generated the location map of the place where the picture was captured.

Some Photos in posted in one's social media account are likely to reveal location information of where such photos were captured if at all the photos were captured using smart phones or digital device like cameras or ipad. A similar study was carried out by Vicente, Freni, Bettini and Jensen (2011) who advised that geotagged information can be used to verify and fix a suspect in arresting or arraigning him/her in court of law. In a rejoinder, Fusco (2011) concurs that digital cameras or smartphones are enhancing investigative forensic evidence because they automatically geotag and embed metadata of location where photos were captured and date as well as time it was taken. The myriad and continually changing ways to share information via social media has resulted in a digital goldmine of potential evidence, such as profiles, lists of friends, group memberships, messages, chat logs, tweets, photos, videos, tags, GPS locations, likes, check-ins, and login timetables.

Sometimes, other social media users update their whereabouts either knowingly or unknowingly. For instance, if a node keeps on updating the *check-in* section in his/her Facebook, detectives will easily use this person's itinerary to tell exact location this node was in while posting a particular message or other stuff. In Harmony, Murphy and Fontecilla (2013) assented that investigators can use social media platforms as an investigative tool by creating undercover pseudo accounts in order to use to harvest intelligence on crimes and suspects or get the identity and movements of suspects. In addition, the investigators can get to know where a specific actor loves visiting mostly and when. Utilizing this intelligence, the researcher established that actor velo~ had visited the city of Mombasa sometimes in

November, 2016. Some uploaded or tagged photographs associated to a particular node can give hints on what that individual loves, places he has visited, people he has been with and activities they did together (Vicente *et al*, 2011). Moreover, the second *check-in* information helped the researcher to know which institution this actor is affiliated with. In some social media platforms, the whereabouts of a particular node can be harvested from posts made in one's account as provided by temporal and geographic data (Rice & Parkin, 2016).

A fourth way in which an investigator can profile the attributes of a particular online user under probe entails harvesting his/her online behavioural activities or patterns. This can be achieved by analysing whom the node interacts with most, places he/she visits at what time of the day or which day of the week among other crucial information.

In mining the activities of actor wilf~, the researcher discovered that this actor is attracted to hacking has evidenced by his Facebook *mentions* and *likes*. Therefore, the study concluded that actor wilf~ is either a white or black hat hacker in one way or another. Moreover, the researcher also discovered that besides hacking, actor wilf~ also fancies athletic sports as well Arsenal football team fan. In a similar way, actor velo~ Facebook page also revealed that she is a fashion trendy person as exhibited by the numerous fashion page links she has *liked*. On actor deno~'s Facebook music page, the researcher discovered that this node love rhythm and blues as well as hip hop music. Findings agrees with Zambri (2015) observations that most online users' behavioural patterns can be drawn from what they *like* as an individual, their preferred music, movies or games among other stuff.

Last but not least, the researcher finally demonstrated how detectives can extract the contents posted or shared by the online user(s) under probe and use it to gain more insight about him/her. For instance, a content shared by actor velo~ was *liked* by over twenty thousand individuals and other almost eight hundred *followers* discussing the shared the content. Furthermore, actor velo~ shared the title of some story books on her Facebook account. Once such information are obtained, they can be utilized to confirm the exact location of the defendant or plaintiff or even help to establish other hitherto unknown accomplices. Again, a detective can tell the mindset of such a person based on the literature books she reads.

In Thompson (2011) information can be harvested from the suspect's social media account in order to predict what that node is contemplating or intending to do. An investigator can therefore take such data as lead information to determine the activities of this person and get to know the kind of people one interacts with. For instance, in his Facebook timeline, actor nico~ posted photos accompanied by some romantic-like messages. Similarly, a few extracts of tweets from actor sams~ suggested that he is either a motivational speaker with some religious attachment. This confirms Vercellone-Smith, Jablokow and Friedel (2012) observation that most individuals like posting their daily activities unwittingly which ultimately disclose their behaviour and movements online.

The content of people's posts such as the text they write, what it says or the content of their photos and videos, and the ratings they assign can be a crucial lead investigators can find on social media. Thus the content of the posts alone, where people detail their thoughts, feelings and ideas reveals what they are doing, what they care about, who they interact with, and why. Social media is a source of intelligence that has the capacity to provide law enforcement officers with access to large volumes of material, posted by all manner of people, and divulging astonishingly candid information to a public audience (Fatih & Bekir, 2015). In a press release from national police service, the inspector general warned the public about a particular post circulating on various social media platforms whose content was to the effect of a plot to rig elections. The circular however, seemed not have tangible evidence against the suspect (Internet Freedom, 2017).

By looking at the content of the posts people are making, one can uncover a lot of leading intelligence information about their actions. This agrees with De Choudhury, Gamon, Counts and Horvitz (2013) that if there is need to assess mental status of a particular suspect, it is wise to utilize his/her online posted information to establish the mental soundness. Nouh and Nurse (2015) nevertheless gave a caveat to investigators that active actors that post the most items on online social media platform may not be necessarily the most influential in the network. The contents shared or posted on social media user accounts can be used by investigator to detect crime-related behaviour, help in singling out the eye-witnesses, confirming alibi, presenting evidential proof in an investigation and can be utilized during court proceeding to confirm or disapprove witness (Rice & Parkin, 2016). Mateescu *et al* (2015) synopsized the above ways of harvesting crucial forensic information that as we keep on sharing or exchanging information via social media platforms, we continue creating a digital goldmine of potential evidence, such as profiles, lists of friends, group memberships, messages, chat logs, tweets, photos, videos, tags, GPS locations, likes, check-ins, and login timetables.

In situation where a suspected individual is using pseudo-names for his/her account, the investigators should strive to gather more evidence in defence of their claim and alibi. Individuals are deemed to be linked or related on Facebook if any of them post, tag or mention the other individual in one’s wall status update (Blomberg, 2012). However, in many countries including Kenya, the legality of using information collected from online social media platforms including blogs are deemed as illegal. According to KTN Prime News

(2017) in Kenya, a court ruling on 6th February, 2017 declared that there will be no charges for online posted offensive messages as the judge termed it unconstitutional the section of the penal code that created criminal defamation. In conclusion Dinerman (2011) law enforcement agencies should know that as long we continue using online social media and increasingly get entrenched into the daily lives of users, private data will be prone to exposure and abuse. Social media platforms provide not only a new unexploited fountain of mining intelligence for law enforcement community, but it also provides an insight of comprehending behavioural patterns of clandestine sub groups of a given social network (Nouh & Nurse, 2015).

5.3 Conclusions

Based on the findings and discussions, it is possible to draw the following conclusions.

The study sought to answer and demonstrate how data obtained from individuals from specific social media can visualized or graph metrics computed as well as harvesting their demographic and related information can aid law enforcement agencies in mining forensic evidence that ultimately lead to arrest of the suspects or arraigning them before a court of law. Through analysis, study noted that there exist numerous ways and measures of determining people that have particular online prominence or influence. Visualizations and graph metric scores supplement one another in the study findings.

The visualizations were used to show fluid relations between nodes in the network and establish the structures of social network connections that exist instead of depending on theoretic or numeric values only. The use visual displays also aided in knowing that a community or network breaks into subgroups based on their interests and other information that captivates specific cohort(s).Thus, Social Network Analysis helps not only to investigate the suspicious characters, but also assist to unearth other dubious nodes that not under probe.

The findings of the study have indicated that the visualizations employed by social network analysis alongside its appropriate software can depict interesting information about the social media users interaction. It is believed that ways and means in which patterns of particular nodes were discovered using visualization will aid the law enforcement officers with ways and techniques of investigating and possibly apprehending online criminal.

Thus, when law enforcement officers employ Social Network Analysis automated tools or techniques to visualize and expose the nature of interactions or structures of suspected criminals using reliable online information, they can remarkably help to stop them from unleashing their heinous acts to the unsuspecting populace. Thus, graph properties of the visualizations employed in this study helped the researcher to know not only the most central nodes in the network but also nodes that were most influential, popular and those who acted as bridges between subgroups of a network.

With regards to information flow, the study founded out that there exist a network that has few nodes drawing attention to a considerable number of connections and these nodes depended on the main actors for information dissemination. The density scores help to establish the frequency of information flow in a network.

This study established that analysis of a network using Social Network Analysis, aids in mapping interactions between nodes and identify structural levels that exist between them. Law enforcement agencies or detectives can use structural attributes of suspects to know their roles, their subgroups and the overall network not forgetting identification of peripheral network members. Hence, mining and analysing social media platforms of suspects can help in gaining more insight into such intrinsic properties of the actors in question.

Drawing from the initial 55 seed nodes and ending with 29,295 nodes, the study concluded that interactions and connections between specific groups of persons in a social network is not static. Instead, it always dynamic as it keeps on changing with time. Individuals may join while others leave the network. The roles of specific nodes can be altered where others are displaced from the top chain of command as other leaders emerge. As relationship keep on changing, some clusters having some interests emerge while other subgroups fragment themselves from the larger network. When detectives are armed with such insight of knowledge, they can use it to investigate dynamism specific to a given network membership and comprehend their attributes and roles in that network. Nevertheless, investigations in real world can supplement their forensic evidence by physically exploring connections between suspected criminals and establish routes that are crucial for creating leads and to detectives.

A node that scores highly in a particular centrality measure is most likely to score highly other related centrality metrics. It was established that degree centrality scores is good for identifying popular nodes that have many connections with other nodes. The most popular individuals in the network draws a lot of attention with regards to communication and therefore serve a network hub. Particularly, users who scored highly in in-degree metrics indicated that received or attracted a lot of attention from other users whereas individuals who scored highly in out-degree centrality was established to be the most active nodes who keep on sending many types of information to others in the network. The study also established that individuals who had direct path or acted as a bridge between subgroups of a network scored highly in betweenness centrality metrics.

In general, the study established that in most situations and research, the degree and betweenness centrality scores are used determine the leaders of a particular covert network group. The study established that the power of an actor is not his/her trait but come as a result of establishing relationships with other powerful nodes in the network. The study established that with Social network analysis, an investigator cannot only identify sub groups within a network but also can gauge the strength of ties between nodes in that network.

The actor with the highest degree centrality is considered to be the most strongly (or most frequently) connected node in the network. Such a node holds an advantaged position in the network in terms of connectivity with other nodes which gives it a key role to propagate information. In other words, degree centrality of an ego node is a measure of immediate influence, that is, what proportion of the nodes in the network are influenced by the ego if the latter influences its neighbours with a piece of information and none of the influenced nodes is allowed to further spread the information. The higher the proportion of nodes influenced, higher will be the degree centrality of the ego node.

If investigators wants to know the person who is quite familiar with the almost all the happenings in a particular, the study recommends that they look for nodes scoring the highest closeness centrality values because they are deemed to be close or more nearer to many other nodes. However, if the detectives are interested to know who the most influential individual in the network is, then a node scoring high eigenvector metric values is their answer. In fact, the study found out that top eigenvector scorers must always be connected to other influential persons. As a confirmatory measure on eigenvector scores, PageRank can be employed to the check for the quality and weighted links between individuals. In case detectives wants to establish if individuals in a suspicious network knows another, then the clustering coefficient scores will effectively direct them to the ego-neighbourhood tie-strength and determine if friends of friends of friends are acquainted with each other.

It is also important to note that actors that score high centrality metrics are not necessarily ring leaders of some felons. Thus if Kenya's law enforcement community embrace and effectively utilize Social Network Analysis practices to extract data from suspected criminal network, it can create big impact as far investigations of criminals is concerned.

People leave traces on social media that can be used by investigators to know their personalities, friends, activities they like, patterns of behaviour and actions. The study also found out that the profiles of social media users has wealthy of information for investigators. These information can be gained from one's demographic or profile information, social connections and relations, using posted or shared contents to map locations one visits frequently and establishing the behavioural patterns based what they post, like, follow or mention among other means. Once the detectives have established a particular suspect, they can initiate ego-network investigation from numerous social media platforms affiliated to him or her.

The study established that most online social media users like posting their daily activities unwittingly which ultimately disclose their behaviour and movements online. The content of people's posts such as the text they write, what it says or the content of their photos and videos, and the ratings they assign can be a crucial lead investigators can find on social media. Thus the content of the posts alone, where people detail their thoughts, feelings and ideas reveals what they are doing, what they care about, who they interact with, and why. Hence, investigators can harvest and analyse the content of the posts people are making to unearth lead intelligence information about their actions. The findings underscored the importance getting justified and accurate evidence in the verdict of a court proceeding based on the contents that people post wittingly or unwittingly.

The harvested information such as photos and GPS information can be circulated and used by the law enforcement officers to trace, investigate and apprehend suspects. Besides harvesting metadata from pictures uploaded by the persons under investigation, the detectives use location and time data of that node as provided by the GPS coordinates which will be ultimately used as evidence or alibi in a court of law against the suspect. It was found out that, in a criminal network, the suspicious characters do not establish a direct link with their subjects. Instead, they get connected using friends of a friend network chain.

In situation where a suspected individual is using pseudo-names for his/her account, the investigators should strive to gather more evidence in defence of their claim and alibi. In Kenya however, there is no clear framework that gives the law enforcement officers an authority to access one’s social media account.

When armed with this information, the law enforcement officers can appropriate techniques and tools to trace the suspect’s social media geographic footprints and ultimately provide sound forensic evidence against him/her. Investigators can create covert social media accounts to help them mine intelligence information pertaining the suspects on the list. In order to effectively employ Social Network Analysis techniques in their investigation, law enforcement agencies ought to keep on monitoring social media activities so as to isolate nodes with criminal mindsets.

It is more evident that social network analysis techniques provides valuable tools that can be used to mine, study and comprehend the functionalities of complex social network structures, the source and power distribution. The study therefore concludes that Social Network Analysis together with social media platforms plays a crucial role as unexploited tool for law enforcement agencies to employ it in investigating all sorts of crimes notably in matters identifying key suspects, their location and time as compared to when a particular event happened hence verification of alibi and predicting likely future acts that the suspect may be contemplating to commit. From the findings, it is apparently evident that social network analysis can be employed to understand how nodes become influential in a social network, how the actors establish relations with other nodes and these relationships advances and grow with time.

A cursory look on court cases done in Kenya based on forensic evidence painted a bad picture. Evidence mined and taken to court of law was rejected by the sitting judge terming it unconstitutional. It established also that some platforms cannot allow mining of data from user's accounts. For instance, as from December 2016, Facebook API that blocks mining of data from accounts its accounts was introduced, which is a great setback for Social Network Analysis. End to end encrypted messages such as Whatsapp running in other platforms are also difficult to mine. Nevertheless, it is hoped that skills and knowledge obtained from this study could help Kenya's law enforcement officers to identify and mine crucial lead information from the suspected criminals and ultimately aid in arresting them as well as presenting forensic evidence in a court of law.

Kenya’s security and law enforcement agencies should know how to use social media to improve security locally, nationally and globally in a cost-effective manner. Social media data can be analysed to map the social networks of various types of offenders. Social media, social networks and Social Network Analysis techniques are just as accessible to criminals and criminal organizations as they are to police.

Recommendations made in the study are hoped to be of great help to the law enforcement agencies in understanding how to can mine, analyse and unearth concealed network elements and patterns between individuals in question. The study has broadened the knowledge on how to apply some Social Network Analysis techniques that is hoped to be of great help to the law enforcement agencies. Kenya’s law enforcement community ought to be challenged to keep abreast both procedurally and legally by the findings of this study. Advanced degree of mining or harvesting data is significant with regards to the forensics evidence from social media users

In Kenya, the majority of members of law enforcement seems to unfamiliar with Social Network Analysis techniques and its associated tools for investigating online suspects. Kenya’s law enforcement agencies should embrace the use of social media and social networking in various ways or applications, including recovering evidence, locating and apprehending suspects, conducting intelligence collections using social networking to conduct crime analysis and intelligence trend analysis. Hence, an enabling technology and trained law enforcement officers will help mitigate or thwart crimes about to be committed in real world. The outcomes of this thesis could influence law enforcement community by providing them with a new insight of investigation and analysing crimes from a large dataset.

The study also established the limitations of Social network Analysis which comprised of incomplete datasets, not knowing in advance whom to include or exclude and the fact that social network is dynamic and transcends geographical boundaries.

5.4 Recommendations

Having looked at the findings of how social network analysis can be used to mine, analyse and present forensic evidence for law enforcement agencies in Kenya, the study gave its recommendations.

5.4.1 Policy Recommendations

(i) Social Network Analysis is an effective tool in mining, analysing and investigating criminal activities committed on various social media platforms. Social Network Analysis views criminal networks as social structures, emphasizing relationships between nodes.

(ii) Kenya’s law enforcement agencies ought to embrace Social Network Analysis and social media as an investigative tool and crime analysis.

(iii) Intelligence obtained from the centrality metrics and visualizations of the nodes under probe should be used in conjunction with real world intelligence.

(iv) The Social Network Analysis design approach of mining and analysing data exchanged between social media platforms is hoped to provide the Kenya’s law enforcement community with knowledge, novelty and insight on ways and means of carrying investigation on suspected criminals.

(v) Kenyan government to impart or facilitate her law enforcement officers with technical skills of graph theory and its associated arithmetic techniques as well as Social Network Analysis knowledge.

(vi) To avoid limitations of getting large and crystal clear visualizations, it is recommended that one should use a computer system with preferably 16GB RAM or Or one might hire IaaS Amazon's cloud (AWS EC2 VM) for 16 GB System is Kshs 3.60 per hour (or $0.36/hour).

(vii) Some NodeXL visualization features are lost whenever the graph is exported to another platform. Is such a situation, this study recommends the use of R language to do visualizations.

(viii) For better visualizations and if adopted, law enforcement community should consider using other commercial Social Network Analysis software such as Sentinel visualizer or Pajek.

(ix) In situation where suspect information is scanty, investigators can utilize search engines to build their profiles.

(x) Subpoena or court orders may be required to access some blocked or private accounts.

5.4.2 Recommendations for Further Research

This study dwelt on demonstrating how Social Network Analysis can be used to mine, analyse and harvest forensic evidence from selected Social media users. However, there are other numerous areas of research (almost infinite) and investigation that needs to explored. Hence, the following areas require further research.

1. Link analysis using K-means clustering on various social media
2. Application of cluster analysis in detecting subgroups within the network
3. Real time Sentiment analysis for various social media platforms
4. Mining and Visualizing large dataset from suspects accounts using Fisheye and Fractal views – Kenya's perspective
5. Mapping criminals using graph theory and scientometric tools

.

REFERENCES

Acquisti, A., Gross, R., & Stutzman, F. (2011). Faces of facebook: Privacy in the age of augmented reality.

Al-Taie, M., & Kadry, S. (2012). Applying Social Network Analysis to Analyze a Web-Based Community. *arXiv preprint arXiv:1212.6050*.

Apantech (2017, March 17). How CA procured technology from Israeli firm to monitor Kenyans' social media activity [Blog post]. Retrieved from http://aptantech.com/2017/03/how-ca-procured-technology-from-israeli-firm-to-monitor-kenyans-social-media-activity/

Arnaboldi, V., Conti, M., Passarella, A., & Pezzoni, F. (2013, April). Ego networks in twitter: an experimental analysis. In *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on* (pp. 229-234). IEEE.

Basu, A. (2005, June). Social network analysis of terrorist organizations in India. In *North American Association for Computational Social and Organizational Science (NAACSOS) Conference* (pp. 26-28). NAACSOS.

Bates, A. (2015). *Using Term Statistics to Aide in Clustering Twitter Posts* (Doctoral dissertation, University of Colorado at Colorado Springs).

Berzinji, A., Kaati, L., & Rezine, A. (2012, August). Detecting key players in terrorist networks. In *Intelligence and Security Informatics Conference (EISIC), 2012 European* (pp. 297-302). IEEE.

Blomberg, J.(2012). *Fighting crime through social media and social network analysis*. Power to Know Sas- White paper. Cary- North Carolina, SAS Institute Inc.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of .sociology*, 1170-1182.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113-120.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE Publications Limited

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *science*, *323*(5916), 892-895.

Bradbury, D. (2011). In plain view: open source intelligence. *Computer Fraud & Security, 2011*(4), 5-9.

Brams, S. J., Mutlu, H., & Ramirez, S. L. (2006). Influence in terrorist networks: From undirected to directed graphs. *Studies in Conflict & Terrorism, 29*(7), 703-718.

Buccafurri, F., Fotia, L., & Lax, G. (2013, July). Allowing privacy-preserving analysis of social network likes. In *Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on* (pp. 36-43). IEEE.

Campbell, W. M., Dagli, C. K., & Weinstein, C. J. (2013). Social network analysis with content and graphs. *Lincoln Laboratory Journal, 20*(1), 61-81.

Carley, M, K.(2003). *Dynamic Network Analysis.* National Research Council workshop on Social Network Modelling and Analysis. Retreived on October 5th, 2016 from http://www.casos.cs.cmu.edu/publications/ protected/2000-2004/2003-2004/carley_2003_dynamicnetwork.pdf

Choudhary, P., & Singh, U. (2015). A Survey on Social Network Analysis for Counter-Terrorism. *International Journal of Computer Applications, 112*(9), 24- 28.

CIPESA (2015). State of Internet Freedom in East Africa 2015. *Survey on Access, Privacy and Security Online.*

Commissioner for Law Enforcement Data Security (CLEDS*).*(2013, July). *Social Media and Law Enforcement*. Retrieved from https:// www.cpdp.vic.gov.au/ images/content/pdf/cleds_special_reports/ CLEDS-Social-Media-and-Law-Enforcement-11-11.pdf

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. *ICWSM, 13*, 1-10.

De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek, Structural Analysis in the Social Sciences*. Cambridge University Press.

Denny, M.(2014). *Social Network Analysis*. Institute for Social Science Research, University of Massachusetts, Amherst.

Demiroz, F., & Kapucu, N. (2012). Anatomy of a dark network: the case of the Turkish Ergenekon terrorist organization. *Trends in organized crime, 15*(4), 271-295.

Dinerman, B. (2011). Social networking and security risks. *GFI White Paper*, 1-7.

Duijn, P. A. C. (2016). Detecting and disrupting criminal networks: A data driven approach.

Dunbar, R. (2002). Social Network Size in Humans. *Human Nature*, 14(1), 53-72.

Ergün, E., & Usluel, Y. K. (2016). An Analysis of Density and Degree-Centrality According to the Social Networking Structure Formed in an Online Learning Environment. *Educational Technology & Society, 19 (4), 34–46.*

Everett, M. G., & Borgatti, S. P. (1999). The centrality of groups and classes. *The Journal of mathematical sociology*, *23*(3), 181-201.

Everton, S. S. (2008). Tracking, destabilizing and disrupting dark networks with social networks analysis.

Fatih, T., & Bekir, C. (2015). Police use of technology to fight against crime. *European Scientific Journal, 11(10), 1857 – 7881.*

Faust, K. (2006). Comparing social networks: size, density, and local structure. *Metodoloski zvezki*, *3*(2), 185.

Faust, K., & Fitzhugh, S. (2012). Social Network Analysis: An Introduction. *Recuperado de: https://www. icpsr. umich. edu/icpsrweb/sumprog/ syllabi/97573 [Consulta: 2014, 12 de agosto]*.

Ferrara, E., De Meo, P., Catanese, S., & Fiumara, G. (2014). Visualizing criminal networks reconstructed from mobile phone records. *arXiv preprint arXiv:1407.2837*.

Flynn, S. E. (2002). America the vulnerable. *Foreign Affairs*, 60-74.

Fraser, L.(2008). Social Networks – Problems of Security and Data Privacy Background. *LSI SIN*, 8(03).

Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*. Vancouver Empirical Press.

Freeman, L. C. (2004). *Graphic Techniques for Exploring Social Network Data, in Models and Methods in Social Network Analysis*. Cambridge University Press, Cambridge, UK.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.

Fusco, S. J., Michael, K., & Michael, M. G. (2010). Using a social informatics framework to study the effects of location-based social networking on relationships between people: A review of literature. International Conference on Mobile Business . Greece: IEEE.

Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., & Snasel, V. (2012). Social networks analysis: Tools, measures and visualization. In *Computational Social Networks* (pp. 3-23). Springer London.

Giuffre, K. (2013). *Communities and networks: using social network analysis to rethink urban and community studies*. John Wiley & Sons.

Global Justice Information Sharing Initiative.(2013). D*eveloping a Policy on the Use of Social Media in Intelligence and Investigative Activities*. Guidance and Recommendations.

Golbeck, J. (2015). *Introduction to Social Media Investigation*. A Hands-on Approach. Syngress Publishers, Massachusetts.

Golbeck, J. (2013). *Analyzing the web.* Elsevier Publishers, Massachusetts.

Granovetter, M. (2005). The impact of social structure on economic outcomes. *The Journal of economic perspectives*, *19*(1), 33-50.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 73 (6), 1360-1380.

Gross, R., & Acquisti, A. (2005, November). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society* (pp. 71-80). ACM.

Gudaitis, T. (2010). The Impact of Social Media on Corporate Security: What Every Company Needs to Know. *Cyveillance, Inc: Virginia*.

Gunnell, D., Hillier, J., & Blakeborough, L. (2016). *Social Network Analysis of an Urban Street Gang Using Police Intelligence Data* . Report 89. Home Office Research. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/491578/horr89.pdf

Gupta, R.,. & Brooks, H. (2015). *Using Social media for global Security*. John Wiley & Sons Inc., Indianapolis.

Hambrick, M. E., & Pegoraro, A. (2014). Social Sochi: using social network analysis to investigate electronic word-of-mouth transmitted through social media communities. *International Journal of Sport Management and Marketing*, *15*(3-4), 120-140.

Hanneman, R. A., & Riddle, M. (2005). Ego networks. *Introduction to Social Network Methods. CA: Riverside. Analytictech. com*.

Hansen, D., Shneiderman, B. & Smith, A,M. (2011). *Analyzing Social Media Networks With Nodexl*: *Insights From A Connected World*. Elsevier Inc., Massachusetts

Hanson, W. (2011). How Social Media is Changing Law Enforcement. *Justice and Public Saftey. GovTech. Last modified December, 2*.

Heidemann, J., Klier, M., & Probst, F. (2010). Identifying key users in online social networks: A pagerank based approach.

Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter Topic-Networks Using Social Network Analysis. *Social Media+ Society*, *3*(1), 2056305117691545.

Hopkins, A. (2010). Graph Theory, Social Networks and Counter Terrorism. *Social Network and Counterterrorism Analysis*.

Hoppe, B., & Reinelt, C. (2010). Social network analysis and the evaluation of leadership networks. *The Leadership Quarterly*, *21*(4), 600-619.

Huber, M., Mulazzani, M., Leithner, M., Schrittwieser, S., Wondracek, G., & Weippl, E. (2011, December). Social snapshots: Digital forensics for online social networks. In *Proceedings of the 27th annual computer security applications conference* (pp. 113-122). ACM.

Hudaib, A. A. Z. (2014). Comprehensive Social Media Security Analysis & XKeyscore Espionage Technology. *International Journal of Computer Science and Security (IJCSS)*, *8*(4), 97.

Internet Freedom(2017, January 23). Kenyan Police threaten to arrest anyone posting ‘malicious’ Social Media updates [Blog post]. Retrieved from https://www.ifree.co.ke/2017/01/kenyan-police-threaten-to-arrest-anyone-posting-malicious-social-media-updates/

Ilyas, M. U., & Radha, H. (2011, June). Identifying influential nodes in online social networks using principal component centrality. In *Communications (ICC), 2011 IEEE International Conference on* (pp. 1-5). IEEE.

Jamali, M., & Abolhassani, H. (2006, December). Different aspects of social network analysis. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)* (pp. 66-72). IEEE.

Jambo News spot. (2015, August 11). *Kenyan data company “reveals” user who leaked photos of bhang-in-underwear girl*. Retrived from http://www.jambonewspot.com/kenyan-data-company-reveals-user-who-leaked-photos-of-bhang-in-underwear-girl/

Johansson, N., & Tenggren, C. (2015). Social Network Analysis of Open Source Projects. *LU-CS-EX 2015-30*.

Johnson, J., Reitzel, J. D., Norwood, B., McCoy, D., Cumming, B., & Tate, R. (2013). Social network analysis: A systematic approach for investigating. *FBI Law Enforcement Bulletin. Available at https://leb. fbi. gov/2013/march/social-network-analysis-a-systematic-approach-for-investigating, 350*.

Johnson, J. A., & Reitzel, J. D. (2011, November). Social network analysis in an operational environment: Defining the utility of a network approach for crime analysis using the Richmond City Police Department as a case study. In *International Police Executive Symposium*.

Kadushin, C. (2004). Introduction to social network theory. *Boston, MA.*

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

Kaye, T., Khatami, D., Metz, D., & Proulx, E. (2014). Quantifying and comparing centrality measures for network individuals as applied to the enron corpus. *SIAM Undergraduate Research Online. submitted*.

Keenan, V., Diedrich, D., & Martin, B.(2013, March). Developing Policy on Using Social Media for Intelligence and Investigations. *FBI Law enforcement Bulletin.* Retrieved from http://www.policechiefmagazine.org/ magazine/index.cfm?fuseaction =display_arch&article_id=2951&issue_id=62013

Kenya Communications Authority (2016, March). *State of Surveillance: Kenya| Privacy International.* Retrieved from http:// www. Privacyinternational .org/ node/735

Kerschbaum, F., & Schaad, A. (2008, October). Privacy-preserving social network analysis for criminal investigations. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society* (pp. 9-14). ACM.

Kirchner, C., & Gade, J. (2011). Implementing social network analysis for fraud prevention. *CGI Group Ind*.

Klandermans, B., & Oegema, D. (1987). Potentials, networks, motivations, and barriers: Steps towards participation in social movements. *American sociological review*, 52(4), 519-531.

Klerks, P. (2001). The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands. *Connections*, *24*(3), 53-65.

Kombo, D. K., & Tromp, D. L. (2006). Proposal and thesis writing: An introduction. *Nairobi: Paulines Publications Africa*, 10-45.

Koschade, S. (2006). A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence. *Studies in Conflict & Terrorism*, *29*(6), 559-575.

Koschade, S. A. (2007). The internal dynamics of terrorist cells: a social network analysis of terrorist cells in an Australian context.

Kunkle, J. (2012). Social media and the homegrown terrorist threat. *The Police Chief, 79*(22).

Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web* (pp. 591-600). ACM.

L'huillier, G., Ríos, S. A., Alvarez, H., & Aguilera, F. (2010, July). Topic-based social network analysis for virtual communities of interests in the dark web. In *ACM SIGKDD Workshop on Intelligence and Security Informatics* (p. 9). ACM.

Long, J. C., Cunningham, F. C., & Braithwaite, J. (2013). Bridges, brokers and boundary spanners in collaborative networks: a systematic review. *BMC health services research*, *13*(1), 158.

Miller, G. (2011). Social scientists wade into the tweet stream. *Science, 333*(6051), 1814-1815.

Mutung'u, G. (2012). Kenyan lawyer examines regulations governing new online media. *New media in Kenya: Time for regulation*? Retrived from https://www.ifex.org/kenya/2012/09/18/article_19_weekly/

Passmore, D. L. (2011). Social network analysis: Theory and applications. *Institute for Research in Training & Development–IRTD.*

Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.

Krebs, V. E. (2002). Mapping networks of terrorist cells. *Connections*, *24*(3), 43-52.

Kriegler, A. (2014, September). Using social network analysis to profile organised crime. *Institute for Security Studies*. Policy Brief 57.

KTN Prime News(2017, February) https://www.standardmedia.co.ke/ktnnews/video/2000121705/joy-for-media-as-high-court-rules-that-there-will-be-no-charges-for-online-offensive-messages

Mainas, E. D. (2012). The analysis of criminal and terrorist organisations as social network structures: a quasi-experimental study. *International Journal of Police Science & Management*, *14*(3), 264-282.

Martino, F., & Spoto, A. (2006). Social network analysis: A brief theoretical review and further perspectives in the study of information technology. *PsychNology Journal*, *4*(1), 53-86.

Mateescu, A., Brunton,D., Rosenblat, A., Patton, D., Gold, Z., & Boyd, D. (2015, June 27). *Social Media Surveillance and Law Enforcement.* Data & Civil Rights: A New era of Policing and Justice. Retrieved from http://www. datacivilrights. org/pubs/2015-1027/Social_Media_ Surveillance_and_Law_Enforcement.pdf

Mena, J. (2003). *Investigative data mining for security and criminal detection*. Butterworth-Heinemann.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415-444.

Mincer, M., & Niewiadomska-Szynkiewicz, E. (2012). Application of social network analysis to the investigation of interpersonal connections. *Journal of Telecommunications and Information Technology*, 83-91.

Moor, J. H. (1997). Towards a theory of privacy I1', in the information age. *Computers and Society*, *27*(3), 27-32.

Mulazzani, M., Huber, M., & Weippl, E. (2012, January). Data visualization for social network forensics. In *IFIP International Conference on Digital Forensics* (pp. 115-126). Springer Berlin Heidelberg.

Murphy, J. P., & Fontecilla, A. (2013). Social media evidence in government investigations and criminal proceedings: A frontier of new legal issues. *Rich. JL & Tech.*, *19*, 11-14.

Nagl, J. A., Amos, J. F., Sewall, S., & Petraeus, D. H. (2008). *The US Army/Marine Corps Counterinsurgency Field Manual*. University of Chicago Press.

Narayanan, A., & Shmatikov, V. (2009, May). De-anonymizing social networks. In *2009 30th IEEE symposium on security and privacy* (pp. 173-187). IEEE.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, *69*(2), 026113.

Nouh, M., & Nurse, J. R. (2015, November). Identifying Key-Players in Online Activist Groups on the Facebook Social Network. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on* (pp. 969-978). IEEE.

Piett, T.(2012, February). How Law Enforcement Uses Social Media for Forensic Investigation.*Mashable*. Retrieved frommashable.com/2012/02/13/ social- media-forensics/#zOAl4_BbqZqh

Perliger, A., & Pedahzur, A. (2011). Social network analysis in the study of terrorism and political violence. *PS: Political Science & Politics*, *44*(01), 45-50.

Petraeus, H,D. & . Mattis, N, J. (2006). *The U.S. Army and Marine Corps Counterinsurgency Field Manual*. Washington: Department of the Army and Department of the Navy Retrieved on September 30th, 2016 from http://fas.org/irp/doddir/ army/fm3-24fd.pdf

PewResearch Center (2014, February*)*. How we analyzed Twitter social media networks with NodeXL**.** Numbers, Facts and Trends Shaping the World. Pew Research Center. Retrieved from http://www.pewinternet.org/files/ 2014/02/How-we-analyzed-Twitter-social-media-networks.pdf

Phillips, E., Nurse, J. R., Goldsmith, M., & Creese, S. (2015). Applying social network analysis to security. In *International Conference on Cyber Security for Sustainable Society* (pp. 11-27).

Prell, C., Hubacek, K., & Reed, M. (2009). *Stakeholder Analysis and Social Network Analysis in Natural Resource Management, Society and Natural Resources*, 22, 501-518.

Punjabi, V. (2014). *Security Risks/Threats & Rewards in Social Media*. Unpublished Masters thesis, University of Oulu

Raab, J., & Milward, H. B. (2003). Dark networks as problems. *Journal of public administration research and theory*, *13*(4), 413-439.

Rahim, A., Amalina, N., & Sulaiman, S. (2015). Social Network Analysis for Political Blogosphere dataset. *International Journal of Advances in Soft Computing & Its Applications*, *7*(3).

Ravindran, S. K., & Garg, V. (2015). *Mastering Social Media Mining with R*. Packt Publishing Ltd, Mumbai.

Rice, S. K., & Parkin, W. S. (2016).Social Media and Law Enforcement Investigations. *Criminology and Criminal Justice, Police and Policing*, Oxford University Press, 2015.

Robbins, I. P. (2011). Writings on the Wall: The Need for an Authorship-Centric Approach to the Authentication of Social-Networking Evidence.

Rodríguez, J. A., & Rodríguez, J. A. (2005). The March 11th terrorist network: In its weakness lies its strength. XXV International Sunbelt Conference, Los Angeles

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. " O'Reilly Media, Inc.".

Sageman, M. (2004). *Understanding terror networks*. University of Pennsylvania Press.

Semitsu, J. P. (2011). From Facebook to mug shot: How the dearth of social networking privacy rights revolutionized online government surveillance.

Senekal, B. A. (2014). Mapping a dark network with Social Network Analysis (SNA): the right wing Vaal Dam bomb plot. *Journal for Contemporary History, 39*(1), 95-114.

Sharma, N., & Strategy, D. G. P. (2008). Sphere of Influence. *The Importance of Social Network Analysis, Pitney Bowes Software*.

Semenov, A. (2013). Principles of social media monitoring and analysis software. Unpublished thesis, University of Jyväskylä.

Siegel, A, D. (2009). Social Network and Collective Action. American Journal of Political Science. 53(1): 122–138.

Shneiderman, B., & Aris, A. (2006). Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 733-740.

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

Staudt, C. L., Marrakchi, Y., & Meyerhenke, H. (2014, October). Detecting communities around seed nodes in complex networks. In *Big Data (Big Data), 2014 IEEE International Conference on* (pp. 62-69). IEEE.

Tayebi, M. A., & Glässer, U. (2016). Social Network Analysis in Predictive Policing. In *Social Network Analysis in Predictive Policing* (pp. 7-14). Springer International Publishing.

Thompson, R. L. (2011). Radicalization and the use of social media. *Journal of strategic security*, *4*(4), 167.'

Tsvetovat, M., & Kouznetsov, A. (2011). *Social Network Analysis for Startups: Finding connections on the social web*. " O'Reilly Media, Inc.".

Van der Hulst, R. C. (2009). Introduction to Social Network Analysis (SNA) as an investigative tool. *Trends in Organized Crime*, *12*(2), 101-121.

Vercellone-Smith, P., Jablokow, K., & Friedel, C. (2012). Characterizing communication networks in a web-based classroom: Cognitive styles and linguistic behavior of self-organizing groups in online discussions. *Computers & Education, 59*(2), 222-235.

Vicente, C. R., Freni, D., Bettini, C., & Jensen, C. S. (2011). Location-related privacy in geo-social networks. *IEEE Internet Computing*, *15*(3), 20-27.

Wang, R., Zhang, W., Deng, H., Wang, N., Miao, Q., & Zhao, X. (2013, June). Discover community leader in social network with pageRank. In *International Conference in Swarm Intelligence* (pp. 154-162). Springer Berlin Heidelberg.

Waskiewicz, T. (2012, January). Friend of a Friend Influence in Terrorist Social Networks. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Waters, G. (2012). Social media and law enforcement. *FBI Law Enforcement Bulletin, 81*(11), 1-5.

Williamson, W., & Ruming, K. (2015). Who's talking, who's listening: exploring social media use by community groups using social network analysis. *14th International Conference on Computers in Urban Planning and Urban Management(CUPUM 2015)*, Cambridge, MA USA , July 7-10, 2015,

Witnov, S. (2011). Investigating Facebook: The Ethics of Using Social Networking Websites in Legal Investigations. *The Santa Clara Computer and High Technology Law Journal, Forthcoming, 28(1), 2*

Wölfer, R., Faber, N. S., & Hewstone, M. (2015). Social network analysis in the science of groups: Cross-sectional and longitudinal applications for studying intra-and intergroup behavior. *Group Dynamics: Theory, Research, and Practice, 19*(1), 45.

Wondracek, G., Holz, T., Kirda, E., & Kruegel, C. (2010, May). A practical attack to de-anonymize social network users. In *Security and Privacy (SP), 2010 IEEE Symposium on* (pp. 223-238). IEEE.

Wyllie, D.(2015, March). How do investigators use social media tools to monitor criminal activity? *PoliceOne.* Retrieved from https://www.policeone.com/ investigations/articles/8502656-How-do-investigators-use-social-media-tools-to-monitor-criminal-activity/

Wright, B.(2010, April 20). Digital Forensics and Social media . *SANS Digital Forensics and Incident Response Blog.* Retrieved from https://digital-forensics.sans.org/ blog/ 2010/04/20/digital-social-media/

Wu, E., Carleton, R., & Davies, G. (2014). Discovering bin-Laden's replacement in al-Qaeda, using social network analysis: a methodological investigation. *Perspectives on Terrorism, 8*(1).

Wüest, C. (2010). The Risks of Social Networking. *Symantec Corporation*.

Voigt, S., Hinz, O., & Jansen, N. (2013). Law Enforcement 2.0-The Potential And The (Legal) Restrictions Of Facebook Data For Police Tracing And Investigation. In *ECIS* (p. 5).

Xu, J., Marshall, B., Kaza, S., & Chen, H. (2004, June). Analyzing and visualizing criminal network dynamics: A case study. In *International Conference on .Intelligence and Security Informatics* (pp. 359-377). Springer Berlin Heidelberg.

Yang, C., Liu, N., & Sageman, M. (2006). Analyzing the terrorist social networks with visualization tools. *Intelligence and security informatics*, 331-342.

Zafarani, R., & Liu, H. (2014). Behavior analysis in social media. *IEEE Intell. Sys*, *29*(4), 9-11.

Zambri, J. (2015, September 30). *Undercover Does Not Mean Out of Sight: Social Media, Social Networking, Facial Recognition Technology and the Future of Law Enforcement Undercover Operations*. Retrieved from http://www.iacpsocialmedia.org/Portals/1/documents/ Undercover%20Does%20Not%20Mean%20Out%20of%20Sight.pdf

Zhu, B., Watts, S., & Chen, H. (2010). Visualizing social network concepts. *Decision Support Systems*, *49*(2), 151-161.

APPENDICES

Appendix I: Interview Schedule for Focus Groups

1. Kindly mention the social media site that you regularly use
2. Describe the first time you log in above mentioned social media site

   i) Who introduced this site to you

   ii) Why do you think person introduced you to this site

   iii) How did he/she convince you to try this site

   iv) What role they play in making you join this site
3. What activities did you perform during your first interaction with the site?

   i) Did you reveal real identity while writing your profile?

   ii) Did you indicate other biographical information such as relationship status, date of birth, phone number, email or other crucial information

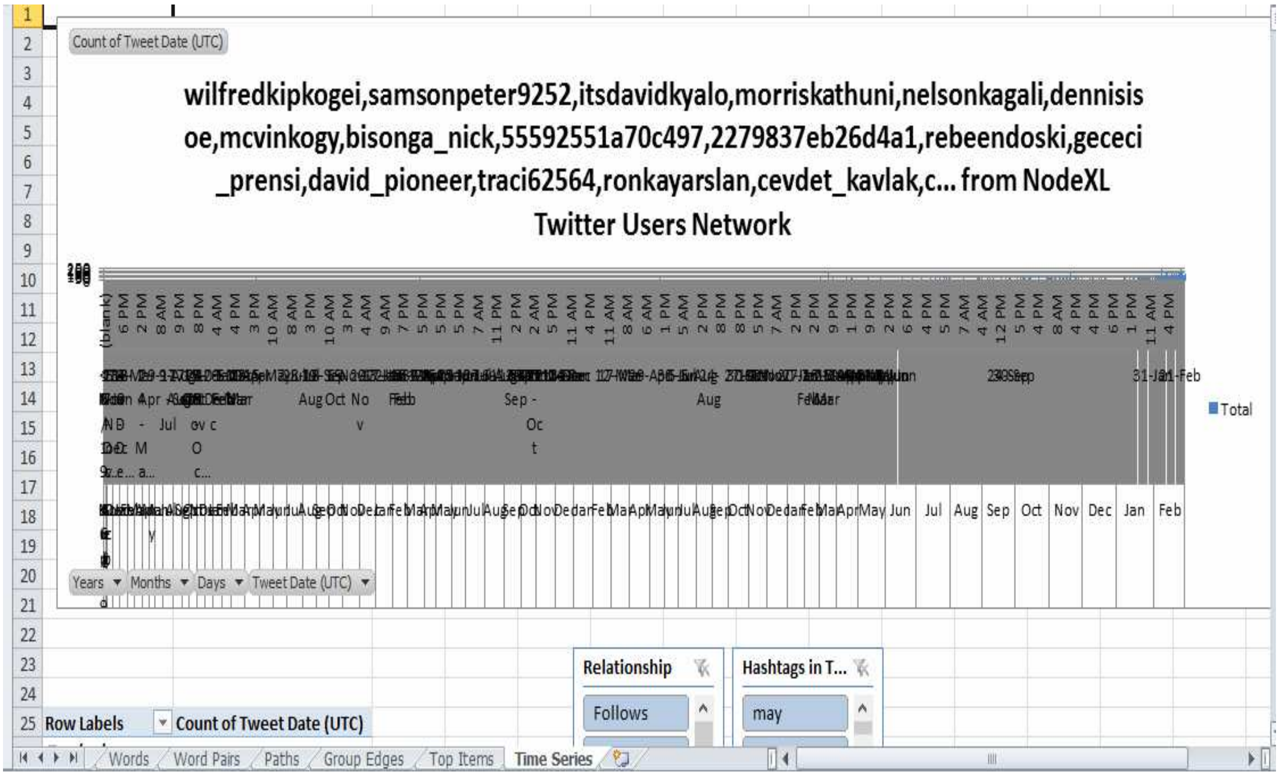
   iii) Upload photos?

   iv) Find friends?

   v) Send messages to friends?

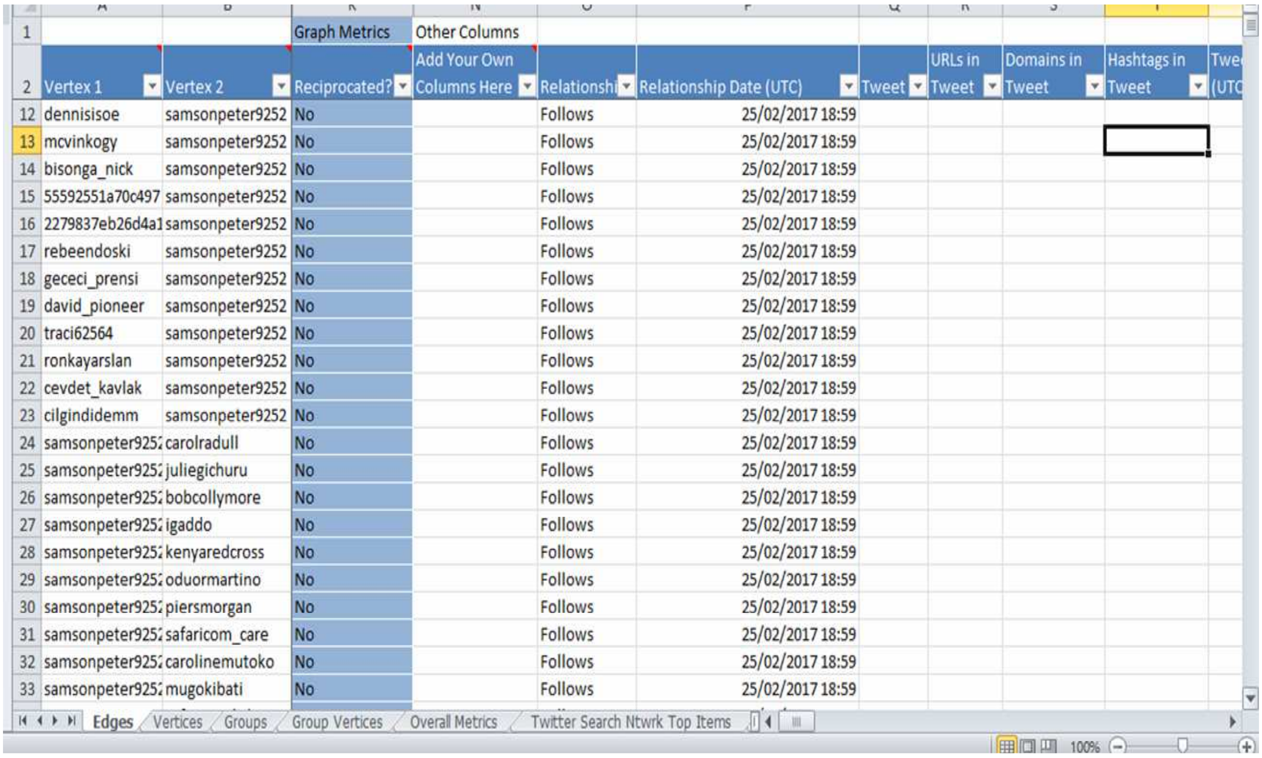
   vi)
4. Have you shared life events that you attend with your friends
5. Do you maintain or have a list of friends or followers linked to your profile
6. Do you perform *likes* or *comments* on various social media
7. Do you maintain other group pages such as online gaming, favourite movies
8. Do you frequently update locations you visit and upload pictures

Appendix II: 2.5 Degree Ego-Centric Time Series Excerpts of Nodes on Twitter Account



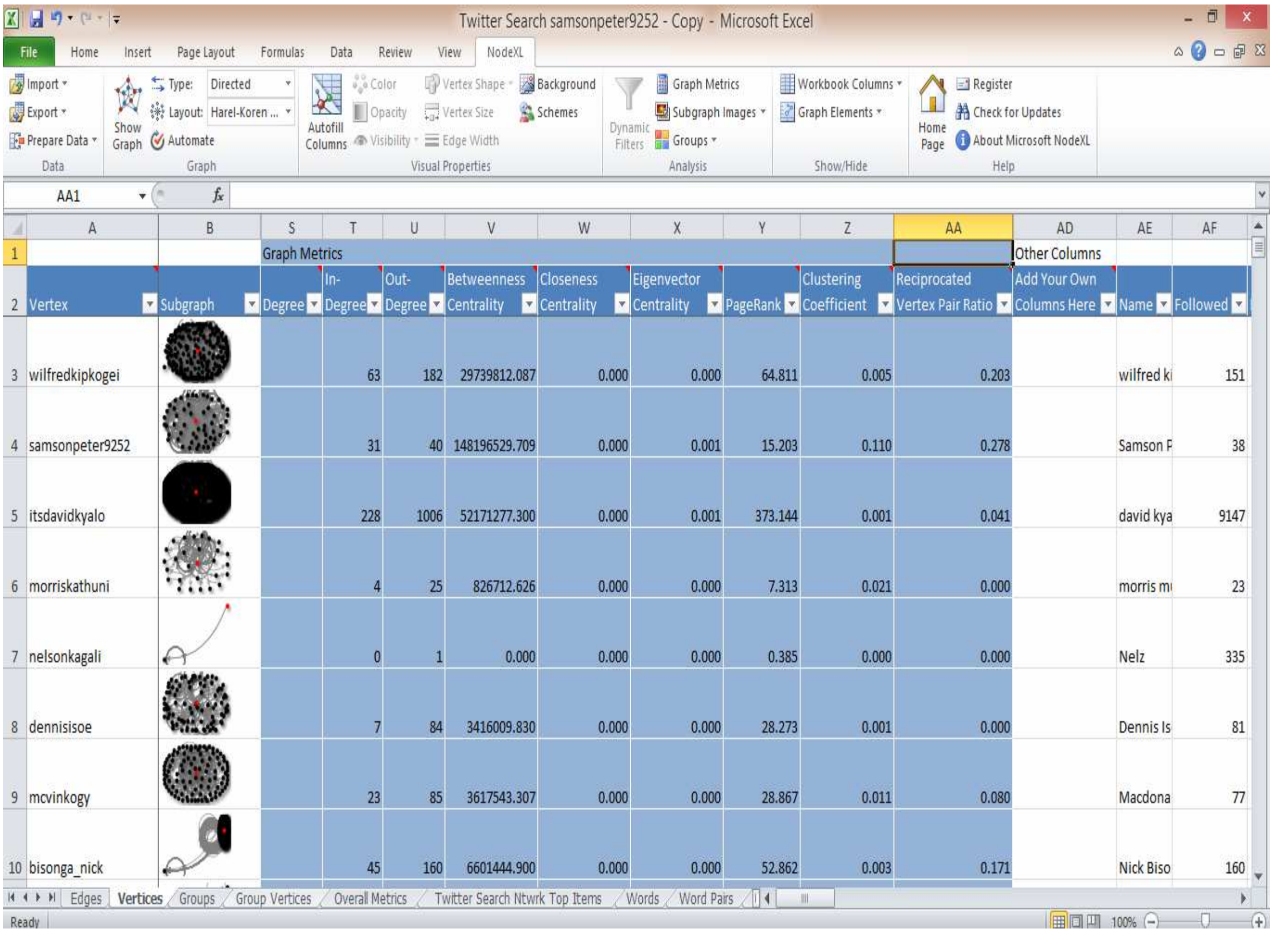
wilfredkipkogei,samsonpeter9252,itsdavidkyalo,morriskathuni,nelsonkagali,dennisisoe,mcvinkogy,bisonga_nick,55592551a70c497,2279837eb26d4a1,rebeendoski,gececi_prensi,david_pioneer,traci62564,ronkayarslan,cevdet_kavlak,c... from NodeXL Twitter Users Network

Appendix III: 2.5 Degree Ego-Centric Edges Excerpts on Twitter Account



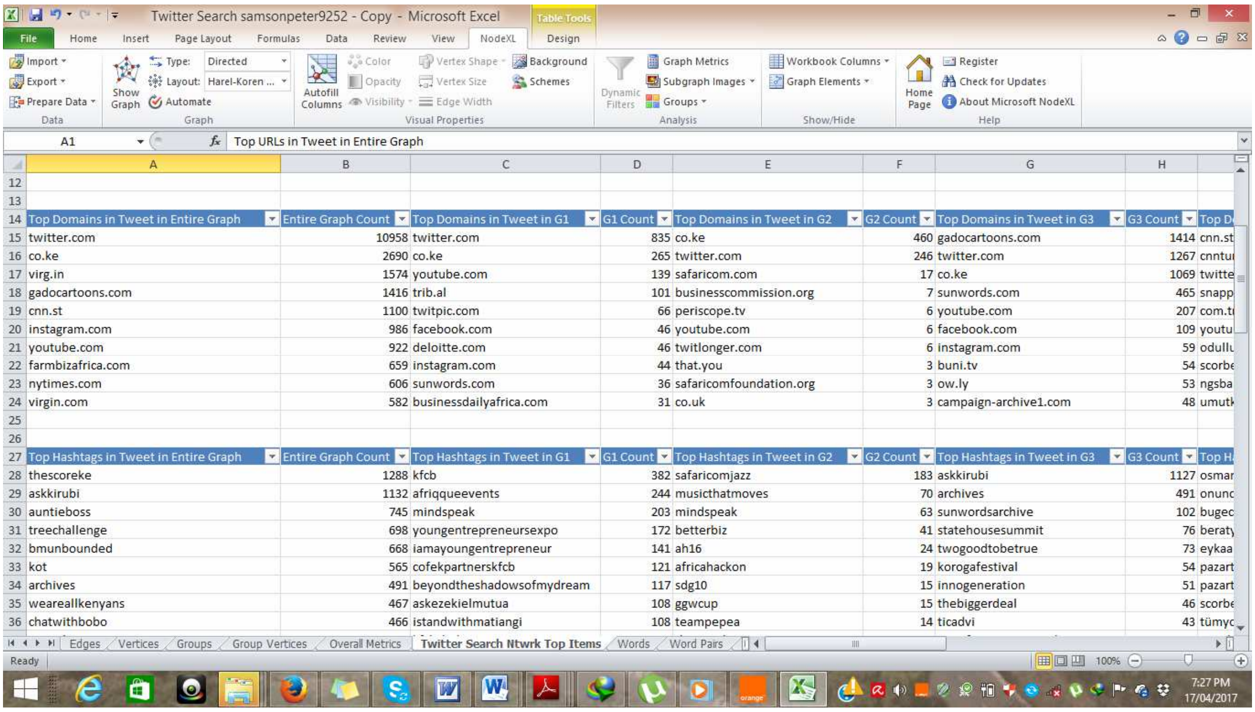
1			Graph Metrics	Other Columns							
2	Vertex 1	Vertex 2	Reciprocated?	Add Your Own Columns Here	Relationshi	Relationship Date (UTC)	Tweet	URLs in Tweet	Domains in Tweet	Hashtags in Tweet	Twe (UTC
12	dennisisoe	samsonpeter9252	No		Follows	25/02/2017 18:59					
13	mcvinkogy	samsonpeter9252	No		Follows	25/02/2017 18:59					
14	bisonga_nick	samsonpeter9252	No		Follows	25/02/2017 18:59					
15	55592551a70c497	samsonpeter9252	No		Follows	25/02/2017 18:59					
16	2279837eb26d4a1	samsonpeter9252	No		Follows	25/02/2017 18:59					
17	rebeendoski	samsonpeter9252	No		Follows	25/02/2017 18:59					
18	gececi_prensi	samsonpeter9252	No		Follows	25/02/2017 18:59					
19	david_pioneer	samsonpeter9252	No		Follows	25/02/2017 18:59					
20	traci62564	samsonpeter9252	No		Follows	25/02/2017 18:59					
21	ronkayarslan	samsonpeter9252	No		Follows	25/02/2017 18:59					
22	cevdet_kavlak	samsonpeter9252	No		Follows	25/02/2017 18:59					
23	cilgindidemm	samsonpeter9252	No		Follows	25/02/2017 18:59					
24	samsonpeter925	carolradull	No		Follows	25/02/2017 18:59					
25	samsonpeter925	juliegichuru	No		Follows	25/02/2017 18:59					
26	samsonpeter925	bobcollymore	No		Follows	25/02/2017 18:59					
27	samsonpeter925	igaddo	No		Follows	25/02/2017 18:59					
28	samsonpeter925	kenyaredcross	No		Follows	25/02/2017 18:59					
29	samsonpeter925	oduormartino	No		Follows	25/02/2017 18:59					
30	samsonpeter925	piersmorgan	No		Follows	25/02/2017 18:59					
31	samsonpeter925	safaricom_care	No		Follows	25/02/2017 18:59					
32	samsonpeter925	carolinemutoko	No		Follows	25/02/2017 18:59					
33	samsonpeter925	mugokibati	No		Follows	25/02/2017 18:59					

Appendix IV: 2.5 Degree Ego-Centric Excerpts Twitter Graph Metrics



	A	B	S	T	U	V	W	X	Y	Z	AA	AD	AE	AF
1			Graph Metrics									Other Columns		
2	Vertex	Subgraph	Degree	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Reciprocated Vertex Pair Ratio	Add Your Own Columns Here	Name	Followed
3	wilfredkipkogei			63	182	29739812.087	0.000	0.000	64.811	0.005	0.203		wilfred ki	151
4	samsonpeter9252			31	40	148196529.709	0.000	0.001	15.203	0.110	0.278		Samson P	38
5	itsdavidkyalo			228	1006	52171277.300	0.000	0.001	373.144	0.001	0.041		david kya	9147
6	morriskathuni			4	25	826712.626	0.000	0.000	7.313	0.021	0.000		morris m	23
7	nelsonkagali			0	1	0.000	0.000	0.000	0.385	0.000	0.000		Nelz	335
8	dennisisoe			7	84	3416009.830	0.000	0.000	28.273	0.001	0.000		Dennis Is	81
9	mcvinkogy			23	85	3617543.307	0.000	0.000	28.867	0.011	0.080		Macdona	77
10	bisonga_nick			45	160	6601444.900	0.000	0.000	52.862	0.003	0.171		Nick Biso	160

Appendix V: 2.5 Degree Ego-centric excerpts Top Hashtags and Domains on Twitter

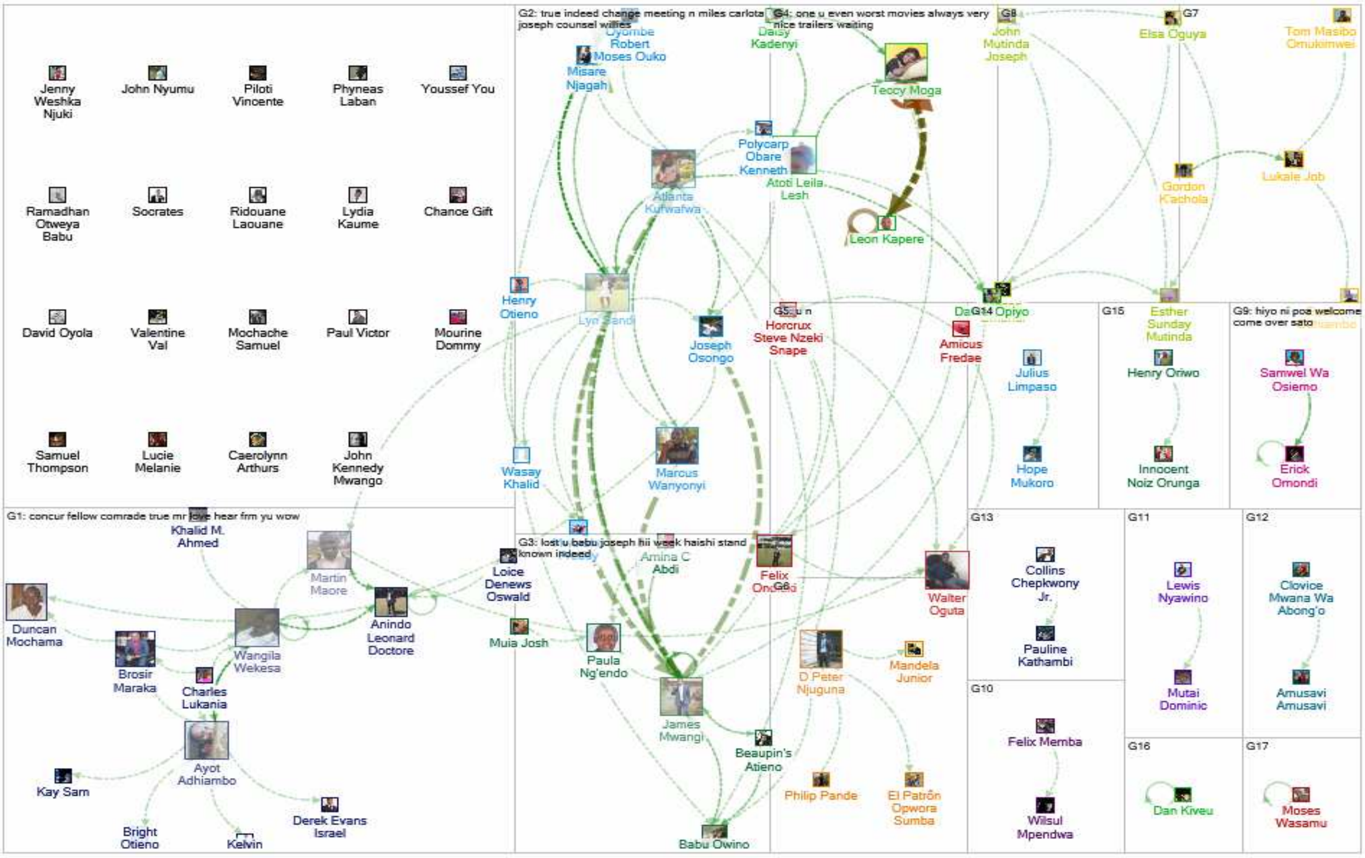


Type: Directed
Layout: Harel-Koren ...
Automate

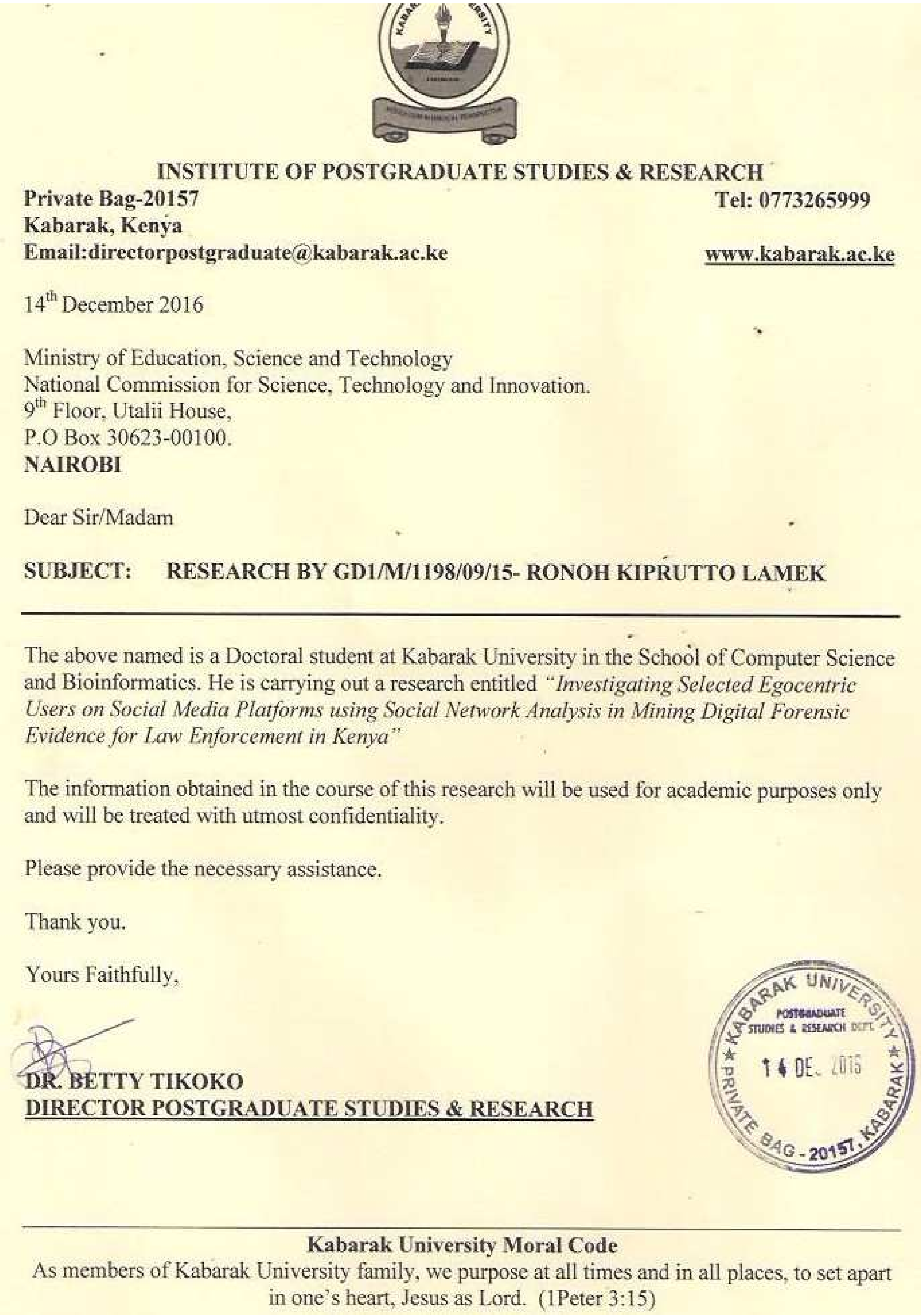
A1 fx Top URLs in Tweet in Entire Graph

	A	B	C	D	E	F	G	H	
12									
13									
14	Top Domains in Tweet in Entire Graph	Entire Graph Count	Top Domains in Tweet in G1	G1 Count	Top Domains in Tweet in G2	G2 Count	Top Domains in Tweet in G3	G3 Count	Top D
15	twitter.com	10958	twitter.com	835	co.ke	460	gadocartoons.com	1414	cnn.st
16	co.ke	2690	co.ke	265	twitter.com	246	twitter.com	1267	cnntu
17	virg.in	1574	youtube.com	139	safaricom.com	17	co.ke	1069	twitte
18	gadocartoons.com	1416	trib.al	101	businesscommission.org	7	sunwords.com	465	snapp
19	cnn.st	1100	twitpic.com	66	periscope.tv	6	youtube.com	207	com.t
20	instagram.com	986	facebook.com	46	youtube.com	6	facebook.com	109	youtu
21	youtube.com	922	deloitte.com	46	twitlonger.com	6	instagram.com	59	odullu
22	farmbizafrica.com	659	instagram.com	44	that.you	3	buni.tv	54	scorbe
23	nytimes.com	606	sunwords.com	36	safaricomfoundation.org	3	ow.ly	53	ngsba
24	virgin.com	582	businessdailyafrica.com	31	co.uk	3	campaign-archive1.com	48	umutk
25									
26									
27	Top Hashtags in Tweet in Entire Graph	Entire Graph Count	Top Hashtags in Tweet in G1	G1 Count	Top Hashtags in Tweet in G2	G2 Count	Top Hashtags in Tweet in G3	G3 Count	Top H
28	thescoreke	1288	kfcb	382	safaricomjazz	183	askkirubi	1127	osmar
29	askkirubi	1132	afriqqueevents	244	musicthatmoves	70	archives	491	onunc
30	auntieboss	745	mindspeak	203	mindspeak	63	sunwordsarchive	102	bugec
31	treechallenge	698	youngentrepreneursexpo	172	betterbiz	41	statehousesummit	76	beraty
32	bmunbounded	668	iamayoungentrepreneur	141	ah16	24	twogoodtobetrue	73	eykaa
33	kot	565	cofekpartnerskfcb	121	africahackon	19	korogafestival	54	pazart
34	archives	491	beyondtheshadowsofmydream	117	sdg10	15	innogeneration	51	pazart
35	weareallkenyans	467	askezekielmutua	108	ggwcup	15	thebiggerdeal	46	scorbe
36	chatwithbobo	466	istandwithmatiangi	108	teampepea	14	ticadvi	43	tümyc

Appendix VI: Respondents Group Facebook Account



Appendix VII: Kabarak University Research Consent



INSTITUTE OF POSTGRADUATE STUDIES & RESEARCH

Private Bag-20157 **Tel: 0773265999**
Kabarak, Kenya
Email:directorpostgraduate@kabarak.ac.ke **www.kabarak.ac.ke**

14th December 2016

Ministry of Education, Science and Technology
National Commission for Science, Technology and Innovation.
9th Floor, Utalii House,
P.O Box 30623-00100.
NAIROBI

Dear Sir/Madam

SUBJECT: RESEARCH BY GD1/M/1198/09/15- RONOH KIPRUTTO LAMEK

The above named is a Doctoral student at Kabarak University in the School of Computer Science and Bioinformatics. He is carrying out a research entitled *"Investigating Selected Egocentric Users on Social Media Platforms using Social Network Analysis in Mining Digital Forensic Evidence for Law Enforcement in Kenya"*

The information obtained in the course of this research will be used for academic purposes only and will be treated with utmost confidentiality.

Please provide the necessary assistance.

Thank you.

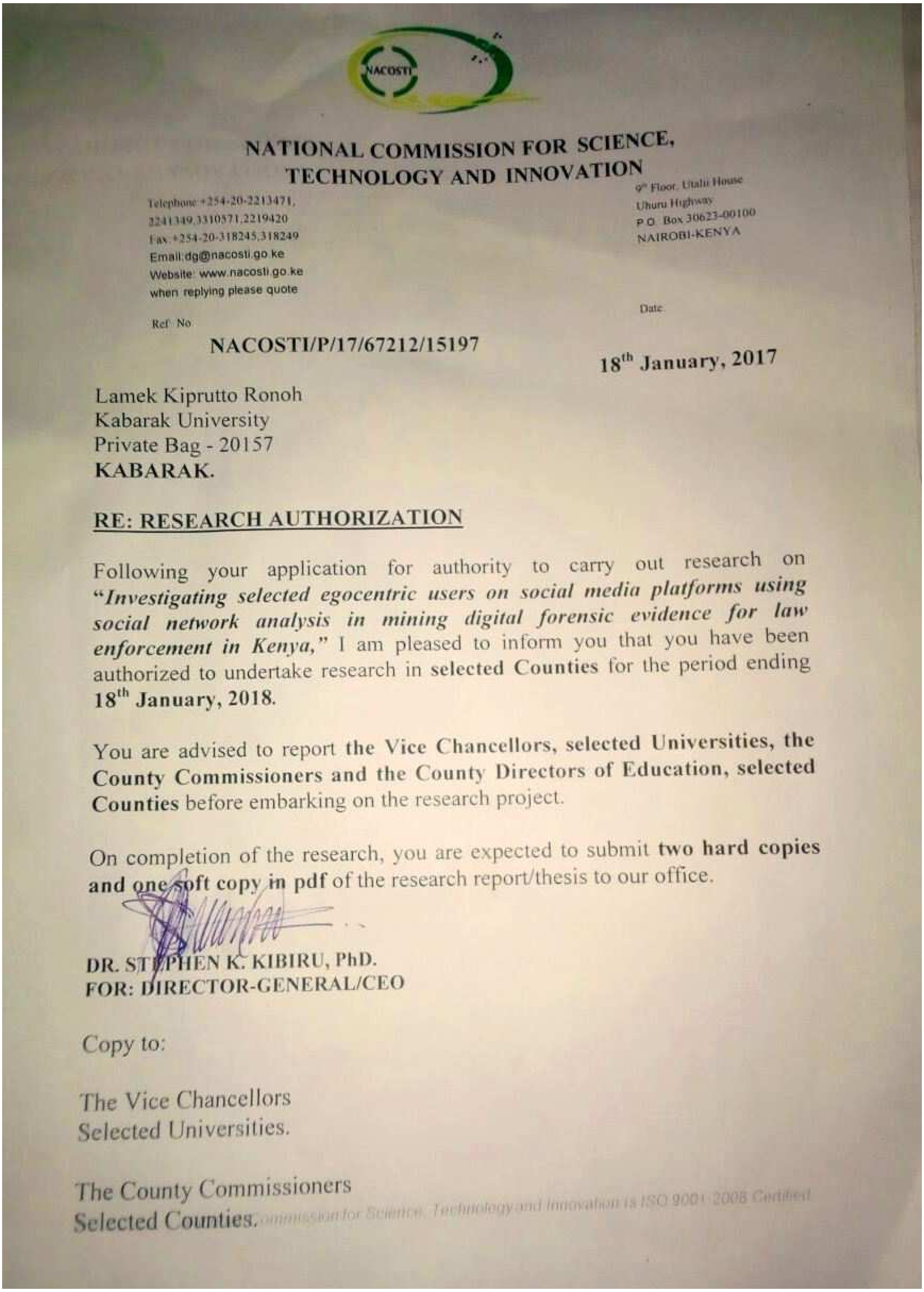
Yours Faithfully,

DR. BETTY TIKOKO
DIRECTOR POSTGRADUATE STUDIES & RESEARCH

KABARAK UNIVERSITY
POSTGRADUATE STUDIES & RESEARCH DEPT.
14 DEC 2016
PRIVATE BAG - 20157, KABARAK

Kabarak University Moral Code
As members of Kabarak University family, we purpose at all times and in all places, to set apart in one's heart, Jesus as Lord. (1Peter 3:15)

Appendix VIII: NACOSTI – Consent Letter



NATIONAL COMMISSION FOR SCIENCE,
TECHNOLOGY AND INNOVATION

Telephone:+254-20-2213471,
2241349,3310571,2219420
Fax:+254-20-318245,318249
Email:dg@nacosti.go.ke
Website: www.nacosti.go.ke
when replying please quote

9th Floor, Utalii House
Uhuru Highway
P.O. Box 30623-00100
NAIROBI-KENYA

Ref: No

NACOSTI/P/17/67212/15197

Date

18th January, 2017

Lamek Kiprutto Ronoh
Kabarak University
Private Bag - 20157
KABARAK.

RE: RESEARCH AUTHORIZATION

Following your application for authority to carry out research on "***Investigating selected egocentric users on social media platforms using social network analysis in mining digital forensic evidence for law enforcement in Kenya,***" I am pleased to inform you that you have been authorized to undertake research in **selected Counties** for the period ending **18th January, 2018.**

You are advised to report **the Vice Chancellors, selected Universities, the County Commissioners and the County Directors of Education, selected Counties** before embarking on the research project.

On completion of the research, you are expected to submit **two hard copies and one soft copy in pdf** of the research report/thesis to our office.

DR. STEPHEN K. KIBIRU, PhD.
FOR: DIRECTOR-GENERAL/CEO

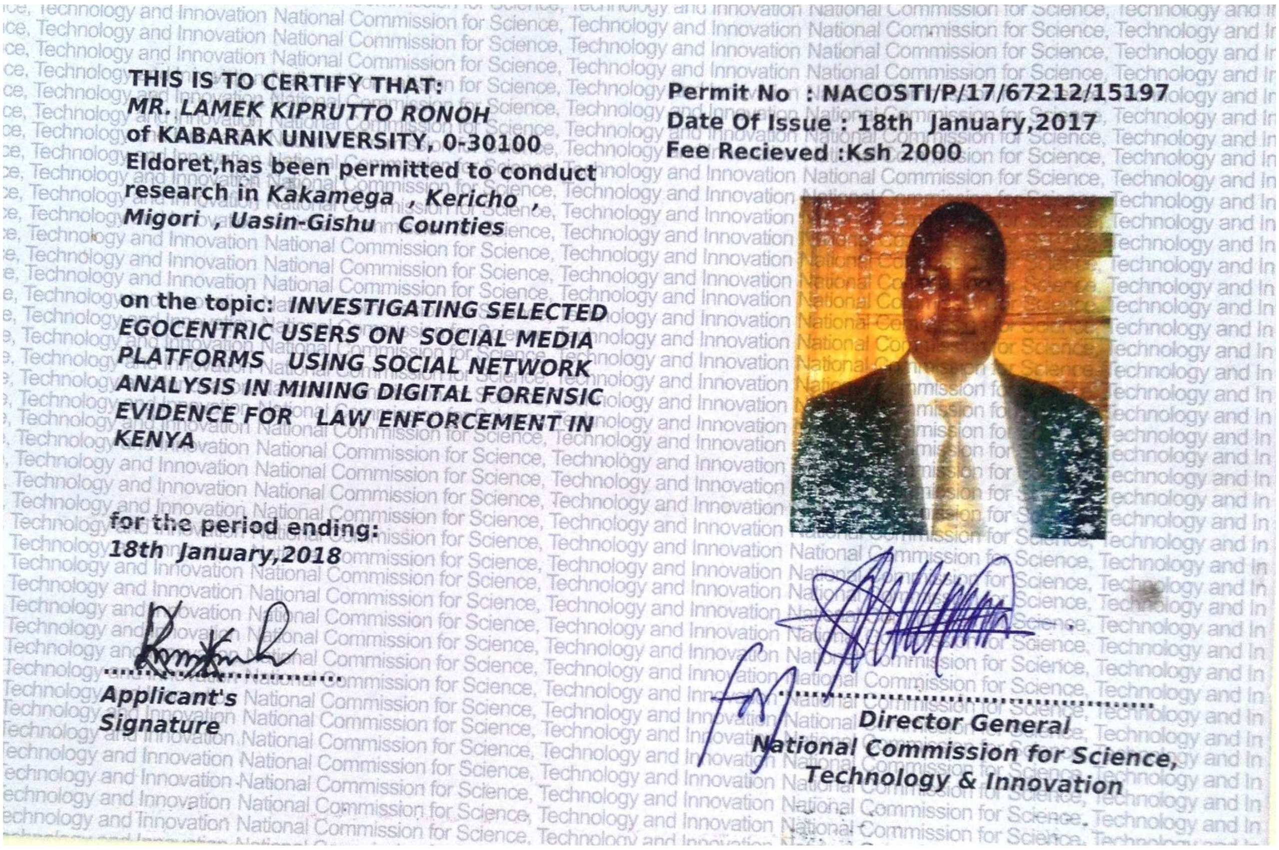
Copy to:

The Vice Chancellors
Selected Universities.

The County Commissioners
Selected Counties.

ommission for Science, Technology and Innovation is ISO 9001: 2008 Certified

Appendix IX: Research Permit



THIS IS TO CERTIFY THAT:
MR. LAMEK KIPRUTTO RONOH
of KABARAK UNIVERSITY, 0-30100
Eldoret,has been permitted to conduct research in *Kakamega , Kericho , Migori , Uasin-Gishu Counties*

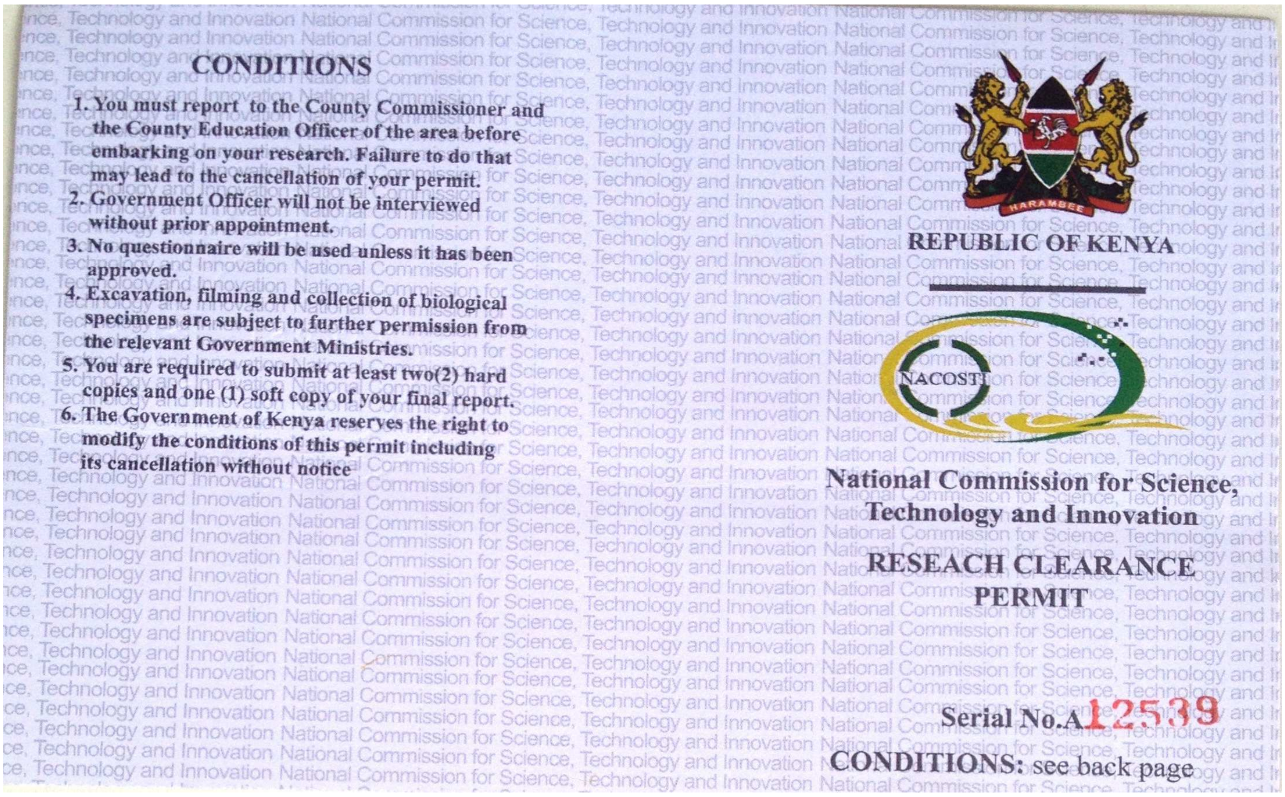
on the topic: *INVESTIGATING SELECTED EGOCENTRIC USERS ON SOCIAL MEDIA PLATFORMS USING SOCIAL NETWORK ANALYSIS IN MINING DIGITAL FORENSIC EVIDENCE FOR LAW ENFORCEMENT IN KENYA*

for the period ending:
18th January,2018

Permit No : NACOSTI/P/17/67212/15197
Date Of Issue : 18th January,2017
Fee Recieved :Ksh 2000

Applicant's Signature

Director General
National Commission for Science, Technology & Innovation



CONDITIONS

1. You must report to the County Commissioner and the County Education Officer of the area before embarking on your research. Failure to do that may lead to the cancellation of your permit.
2. Government Officer will not be interviewed without prior appointment.
3. No questionnaire will be used unless it has been approved.
4. Excavation, filming and collection of biological specimens are subject to further permission from the relevant Government Ministries.
5. You are required to submit at least two(2) hard copies and one (1) soft copy of your final report.
6. The Government of Kenya reserves the right to modify the conditions of this permit including its cancellation without notice

HARAMBEE

REPUBLIC OF KENYA

NACOSTI

National Commission for Science, Technology and Innovation

RESEACH CLEARANCE PERMIT

Serial No.A 12539

CONDITIONS: see back page