Using SNA Centrality Metrics to Detect Suspicious Social Media
Users to Aid Law Enforcement Agencies in Kenya

Lamek Ronoh*1, Nickson Karie2 and Kefa Rabah2
1School of Information, Communication & Media Studies
Rongo University, P.O. Box 103-40404, Rongo Kenya
2School of Computer Science and Bioinformatics
Department of Information Technology Security
Kabarak University, Private Bag 20157, Kabarak, Kenya
Email: ronohlamek@gmail.com

* Corresponding author

Received: July 30, 2017
Published: August 15, 2017

Abstract
Investigation of social media using social network theory is a new powerful tool that is expected to aid and easy law enforcement agencies in targeting cybercriminals in multi-faceted ways in this ever evolving digital landscape. It is against this backdrop that this study focused on identifying, investigating and detecting individuals based on selected users of Facebook and Twitter social media platforms. 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 using network metrics. Social network analysis experimental research design was employed in this study. The sample size of the respondents was arrived at by employing snowball sampling procedure and particularly Yamane’s formula of calculating sample size. The respondents were guided to create pseudo-online parody accounts in various social media platforms, which were later used to carry out the online data mining from the selected respondents to aid in social network analysis. Data mining and analysis was done using NodeXL. Results were presented in form of social network centrality metrics or measures on egocentric networks. 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 and roles they play in a given network. The findings presented in this research principally demonstrates how law enforcement agencies can utilize this technique in identifying and tracking suspicious characters and ultimately help in maintaining law and order.

Keywords: Degree centrality, Betweenness, Eigenvector, Closeness Social Network Analysis, law enforcement

© 2016 by the author(s); Mara Research Journals (Nairobi, Kenya)

1. INTRODUCTION
Social network research has gained significant acknowledgment in terms of both theory and method in contemporary research body (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-Szynkiewicz (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. Golbeck
(2015) succinctly defines SNA 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. Granovetter (1973) 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.

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 are 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 turn out to be important in assisting users to discern patterns, trends, clusters, and outliers, even in an intricate social networks.

1.1 Social Network analysis in law enforcement

A study on social network analysis for anti-terrorism, Choudhary and Singh (2015), established that SNA 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, use of SNA in latent pattern recognition and employing appropriate data mining software, researchers found out 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).

Wyllie (2015) noted 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. 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 SNA centrality and community measures with the help of SNA tools, he was able 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). SNA 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).

1.2 Social Media and Law Enforcement in Kenya

According to the report released by the Kenya Communications Authority (KCA) (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.

1.3 Statement of the Problem
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.

This study sought to present the application of SNA techniques in investigating, retrieving meaningful information from the selected individuals on Facebook and Twitter social media platforms, in order to detect and present digital forensic evidence and intelligence that can aid in law enforcement agencies in Kenya.

1.4 Objective of the Study
The Specific objective of the study was:

- 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.

1.5 Research Question
The research question for the study was stated as:

- 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 user’s network data in Kenya

2. 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 signify 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).
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 unveils 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 acts, is a score showing the importance or significance of that node in the entirety position of the network (Johnson & Reitzel, 2011).

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 that the person will be in a good position to hear from most friends of friends.

### 2.1 Degree Centrality

Centrality measure is a network metric used to highlight actors that cover relevant roles inside the analyzed 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 in-degree and out-degree 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 direct or an indirect 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.

### 2.2 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, it’s minimum when an actor is not connected to any others.

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, i.e., it means that it takes fewer steps to get to other nodes. Closeness centrality corresponds to 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.
With regards to actors scoring highest closeness centrality metrics; Hanneman and Riddle (2005) claims 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).

2.3 Degree Betweenness

Betweenness centrality is a network metric that measures 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 or 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 support 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 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) observes that this metric 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).

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) observe that the eigenvalue centrality score suggest that the 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.
2.5 Clustering Coefficient

Clustering coefficient is applicable to both a single node or 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 also true (Johansson & Tenggren, 2015).

2.6 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 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.

2.7 Social Network Density

Wassermann and Faust (1995) articulated that density is the common position of connection point between the social network actors. 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) who 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.

Researchers have found that network density is positively related to the likelihood that actors within the network will follow accepted norms and behavior, which is why a primary basis for moral order is highly-connected social networks (Himelboim, Smith, Rainie, Shneiderman & Espina, 2017).

2.8 Application of SNA in Social Media - Information Security Issues

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).

Numerous scholars researching on security issues have applied SNA 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 SNA to map terrorist networks emerged. Krebs (2002) initiated the application of SNA to map the terrorist networks by applying SNA’s centrality measures mentioned elsewhere in this study. Similarly, Koschade (2006) used an 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.
2.9 Conceptual Framework

The study was guided by the conceptual framework shown in Fig. 1.

![Conceptual Framework Diagram]

**Figure 1:** Conceptual Framework.
*Source: Researcher*

### 3. RESEARCH DESIGN AND METHODOLOGY

Social Network Analysis (SNA) 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 platform, which were used to perform online mining of the selected respondents to obtain data that ultimately aided in social network analysis. Social network analysis (SNA) is the study of the social structure known as social networks comprised of individuals and their relationships. SNA 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.

Thus, Social Network Analysis has evolved as a popular, standard method for modeling meaningful, often hidden structural relationships in communities. Existing SNA tools often involve extensive pre-processing or intensive programming skills that can be challenging to a non expert user. 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).

### 4. DISCUSSION AND EVALUATION

#### 4.1 Summary Statistics of One Degree Ego-centric 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. Table 1 shows global information that summarises the main actor’s initial seed network. The findings indicate that at 1-degree egocentric network, there were 55 vertices having 66 unique connections (edges).
Table 1: Summary Statistics of Initial seed 1-Degree Egocentric Network

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertices</td>
<td>55</td>
</tr>
<tr>
<td>Unique edges</td>
<td>66</td>
</tr>
<tr>
<td>Graph density</td>
<td>0.0232323232323232</td>
</tr>
<tr>
<td>Average Geodesic Distance</td>
<td>1.927934</td>
</tr>
</tbody>
</table>

Source: Research data (2017)

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 to communicate. 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).

4.1.1 Summary Statistics of One Degree Ego-centric Network

Table 2 demonstrates the overall results of an expanded network from the initial seed of 55 vertices to 483 vertices in a directed graph. 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.

Table 2: Network Statistics of 1.5 Degree Egocentric Network

<table>
<thead>
<tr>
<th>Graph Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph Type</td>
<td>Directed</td>
</tr>
<tr>
<td>Vertices</td>
<td>483</td>
</tr>
<tr>
<td>Unique Edges</td>
<td>565</td>
</tr>
<tr>
<td>Edges With Duplicates</td>
<td>198</td>
</tr>
<tr>
<td>Total Edges</td>
<td>763</td>
</tr>
<tr>
<td>Self-Loops</td>
<td>22</td>
</tr>
<tr>
<td>Reciprocated Vertex Pair Ratio</td>
<td>0.15819209</td>
</tr>
<tr>
<td>Reciprocated Edge Ratio</td>
<td>0.273170732</td>
</tr>
<tr>
<td>Connected Components</td>
<td>1</td>
</tr>
<tr>
<td>Single-Vertex Connected</td>
<td>0</td>
</tr>
<tr>
<td>Components</td>
<td></td>
</tr>
</tbody>
</table>
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 linking 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. From Table 2, the results indicate the modularity value of approximately 0.6, which implies that the network clusters were fairly separated from each other in the network.

The average geodesic distance score 3.2 in Table 2 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 density score of 0.003 implies that vertices were loosely interconnected, hence, low density, but high density at cluster level.

However, in measuring the strength partition and robustness of the network, the modularity score was found to be 0.5. The score closely relates to Newman and Givran (2004), who noted 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.

**Centrality Metrics**

In quest of demonstrating how SNA 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 later used to identify the importance of an actor in network. Centrality measures of an individual in a network provide an idea about the actor’s role in that network and the connections between nodes reveals the general structure of the network (Hopkins, 2010). 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.

4.2 Degree Centrality

Table 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. 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. 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). 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.

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilfredkipkogei</td>
<td>196</td>
</tr>
<tr>
<td>Nicokoech</td>
<td>93</td>
</tr>
<tr>
<td>Velodiek</td>
<td>82</td>
</tr>
<tr>
<td>2279837eb26d4a1</td>
<td>73</td>
</tr>
<tr>
<td>Denokisaka</td>
<td>54</td>
</tr>
<tr>
<td>Kiptalambran</td>
<td>39</td>
</tr>
<tr>
<td>martha_kirika</td>
<td>9</td>
</tr>
<tr>
<td>Samsonpeter9252</td>
<td>7</td>
</tr>
</tbody>
</table>

Source: Research data (2017)

4.2.1 Respondents Betweenness Centrality Scores

Table 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 order to determine, which actor in the network act as bridge between subgroups of the entire network. As can be seen from Table 4, actor wilf~ has highest betweenness centrality score of 69302.912, whereas actor kiptal~ 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.

Judging from these results, therefore, actor wilf~ is the most important node in this network as far as 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 conforms with Xu, Marshall, Kaza and Chen (2004) that betweenness
centralities 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.

**Table 4:** Top Seven Scores Actors for Betweenness Centralities

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Betweenness centralities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilfredkipkogei</td>
<td>69302.912</td>
</tr>
<tr>
<td>Nicokoech</td>
<td>37752.733</td>
</tr>
<tr>
<td>Velodieck</td>
<td>31198.671</td>
</tr>
<tr>
<td>2279837eb26d4a1</td>
<td>27026.598</td>
</tr>
<tr>
<td>samsonpeter9252</td>
<td>22823.611</td>
</tr>
<tr>
<td>Denokisaka</td>
<td>19777.960</td>
</tr>
<tr>
<td>Kiptalambrian</td>
<td>13802.640</td>
</tr>
</tbody>
</table>

Source: *Research data (2017)*

Betweeness 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). Nodes scoring highest betweenness centrality values are powerful in the network (Berzinji, Kaati & Rezine, 2012). Moreover, the research results agrees with that of Krebs (2002), who noted 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.

**4.2.2 Respondents Closeness Centrality Scores**

Table 5 demonstrates the closeness centralities of the top actors from the entire network. Interestingly, several of the top actors scored a similar closeness centrality score of 0.001. The results also suggested that majority of network actors, notably the actors with various high centrality metrics were able to reach one another fast across the network. These scores conform to that of Denny (2014), who noted that closeness measure is sensitive to network size and is decreasing in the number of actors in the network.

**Table 5:** Top Seven Scores Actors for Closeness Centralities

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Closeness Centralities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilfredkipkogei</td>
<td>0.001</td>
</tr>
<tr>
<td>samsonpeter9252</td>
<td>0.001</td>
</tr>
<tr>
<td>Denokisaka</td>
<td>0.001</td>
</tr>
<tr>
<td>Velodieck</td>
<td>0.001</td>
</tr>
<tr>
<td>Kiptalambrian</td>
<td>0.001</td>
</tr>
<tr>
<td>Ntvkenya</td>
<td>0.001</td>
</tr>
<tr>
<td>brian_baclay</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Source: *Research data (2017)*

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. 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.

4.3 Respondents Eigenvector Centrality Scores

Eigenvector centrality was computed to measure the influences of an actor in the entire network. Table 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 the findings of Wu, Carleton and Davies, (2014) that a higher eigenvalue centrality score of a node suggest that the 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 they are less influential and less popular amongst the top network important actors. This coincides with Wasserman and Faust (1994) findings 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. Eigenvector score of an actor indicates the significant roles and status of his/her network neighbours (Arnaboldi, Conti, Passarella & Pezzoni, 2013).

Table 6: Top Seven Scores Actors for Eigenvector Centralities

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Eigenvector Centralities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilfredkipkogei</td>
<td>0.057</td>
</tr>
<tr>
<td>Denokisaka</td>
<td>0.013</td>
</tr>
<tr>
<td>Velodieck</td>
<td>0.013</td>
</tr>
<tr>
<td>Kiptalambrian</td>
<td>0.009</td>
</tr>
<tr>
<td>samsonpeter9252</td>
<td>0.007</td>
</tr>
<tr>
<td>brian_baclay</td>
<td>0.006</td>
</tr>
<tr>
<td>2279837eb26d4a1</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Source: Research data (2017)

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.

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 less ties are the most secure, although their capacities have limits in accessing information.

### 4.4 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 7 hitherto gives respondent wilf~ a leading edge with a score of 81.686, followed at a far distant with nico~ and velo~ with scores of 40.995 and 33.346, respectively. These results are indicative measure of the large number of incoming connections that aforementioned actors cultivated in the network.

<table>
<thead>
<tr>
<th>Vertex</th>
<th>PageRank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilfredkipkogei</td>
<td>81.686</td>
</tr>
<tr>
<td>Nicokoech</td>
<td>40.995</td>
</tr>
<tr>
<td>Velodiek</td>
<td>33.346</td>
</tr>
<tr>
<td>2279837eb26d4a1</td>
<td>29.795</td>
</tr>
<tr>
<td>Denokisaka</td>
<td>20.860</td>
</tr>
<tr>
<td>Kiptalambrian</td>
<td>15.584</td>
</tr>
<tr>
<td>martha_kirika</td>
<td>3.594</td>
</tr>
<tr>
<td>samsonpeter9252</td>
<td>2.595</td>
</tr>
</tbody>
</table>

Source: Research data (2017)

The results of Table 7 are clearly different from that of Eigenvector centralities. This confirms 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 incoming 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.

### 4.5 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 8 placed new actors namely brian_ba~ and itsdavid~ as top nodes with a clustering coefficient of 1.00 apiece amongst many other 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.

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Clustering Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>brian_baclay</td>
<td>1.000</td>
</tr>
<tr>
<td>Itsdavidkyla</td>
<td>1.000</td>
</tr>
<tr>
<td>Railaodinga</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Surprisingly, the main actors of centrality measures were nowhere to be seen near the top of the clustering coefficient results. 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. More specifically, actor wilf~ scored clustering coefficient value of 0.001, actor velo~ registered 0.05 and actor deno~ recorded 0.016. The implication of these findings is that friends of actors like wilf~, velo~ or deno~ and others having low clustering coefficient scores, were most likely are not acquainted to each other. 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.

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.

### 4.6 Respondents Reciprocated Vertex Pair Ratio Scores

Table 9 above shows an excerpt of the reciprocated vertex pair ratio results. Recall that in Table 2, the overall network score for reciprocated vertex pair ratio was 0.15819209. It is 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 9: Reciprocated Vertex Pair Ratio**

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Reciprocated vertex pair Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Umutkatirci</td>
<td>1.000</td>
</tr>
<tr>
<td>mtuenken1</td>
<td>1.000</td>
</tr>
</tbody>
</table>
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.

5. CONCLUSION

The findings of the study have indicated that through SNA, there exist numerous ways and measures of determining people that have particular online prominence or influence. Thus, SNA helps, not only to investigate the suspicious characters, but also assist to unearth other dubious nodes that are not under probe.

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 are 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. 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 is 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 draw 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 to determine the leaders of a particular covert network group. The study established that the power of an actor is not his/her trait but comes as a result of establishing relationships with other powerful nodes in the 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.

If investigators want 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 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 want to establish if individuals in a suspicious network knows one 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 embraces and effectively utilizes SNA practices to extract data from suspected criminal network, it can create big impact as far investigations of cybercriminals is concerned.

5.1 Recommendations

The study recommends the following:

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

(ii) Kenya’s law enforcement agencies ought to embrace SNA and social media both as an investigative tool and cybercrime analysis

(iii) Intelligence obtained from the centrality metrics 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 agencies with knowledge, novelty and insight on ways and means of carrying out investigation on suspected cybercriminals.

6. REFERENCES


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


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


**Cite this article:**