Towards Mining of Stakeholders in Criminal Organizations from Telecommunication Metadata: Analytic Approach to Latent Feature Extraction

Abideen A. Ismail, E. N. Onwuka, B. A. Salihu, and C.O. Ubadike
Telecommunication Engineering Department,
School of Electrical Engineering and Technology,
Federal University of Technology, Minna, Niger State.

ABSTRACT

Mining techniques extract implicit, previously unknown and potentially useful information from data towards providing a solution to future challenges. Mining involves analysis of links in existing datasets. For fighting crime, its efficiency and robustness are on active covert criminals and not on silent key-players like criminal stakeholders (CS). The analytic approach attempts to differentiate a CS from well-known covert nodes in a social network structure (SNS). A CS is a two-face-attribute covert node; a bottle-neck for mining techniques. Some conceptually general attributes often used to describe covert nodes are examined and analyzed to figure out a unique latent feature for a CS. It is found that forceful use of influence and vulnerability are bogus for description and detection of a CS. Our approach arrived at a new feature for the description and identification of a CS. The new feature is dynamic and it is not self-sufficient like vulnerability and influence that is general to the covert nodes, or network leader (NL). Discovered attribute for mining CSs depends on NL’s position in the structure.

INTRODUCTION

Metadata summarizes activities of electronic communication channels users. It is underlying information related to electronic communication devices usage [1] containing digital footprint about online/offline relationships, transactions, location and attitude of users. Metadata’s framework excludes the content of conversations or transactions reportages; thus aids relationships’ formation[2]. As an informative label obtainable from electronic communication channels about phone calls, short messages services, social media, and email about their users, it denies security operatives access to substantial exhibits for crackdown investigation or prosecution purposes[3]. Parts of the labels in metadata include caller, called, calling time, the period of calls, locations, and frequency of calling [4]. The caller and called are primarily being used for forming the network of relationships of mobile phone users [5], [6].

Network structures of different groups can be obtained from metadata to establish elements and links. The structural relationships are called social network structure (SNS) or complex network (CN)[7], [8]. A complex network describes a system that its components cannot be exhausted [9]–[11]. Defective data subscribe to criminal organizations to a CN due to missing information about memberships and links [12]–[14]. Criminal organizations are unique CN as rigorous methods for collecting data do not assure comprehensive data reportage through the deliberate habit of criminal participants in submerging evidence about their relationships and transactions[7], [15]–[17]. Confirming a fundamental attribute of CN through missing information and data defectiveness. A consequence of missing information transcends properties of networks investigated in [17] but it undermines the efficiency of detective techniques and security agencies in fighting organized criminal groups[18].

Having stated this, and reviewed of various intelligence techniques for identifying network leader (NL) in criminal organizations and other active participants whose their removal will have a drastic effect in destabilizing organized crime groups (OCG) [19], [20]. It had been observed that common models for studying OCG gives less attention to external participation and it paves way for less susceptibility of affiliate participants (or stakeholders) in crime. The next section presents a few works where telecommunication metadata served the sources for analyzing criminal organizations. Secondly, it shows how less susceptible members in the OCGs were indirectly inferred within criminal networks. Use of multi-relation network analysis is being resourceful for the task and challenges about the identification of
low profile social-capital actors which is almost impossible from analyzing a single-layer social network (SN) characterized some OCG like a terrorist group. Analysis in the next section is on drug trafficking organization (DTO) [21] as their data are rich in multi-layer network form point of sharing resources and unavoidable inter-relationships [10].

RELEVANT LITERATURE

Multiple-source datasets or multilayer data-network analysis had been suggested as a mitigating mechanism for a criminal investigation to lower the effect of missing information [22], [23] and data defectiveness [24]. According to Butt et al [24], more key-players become more vulnerable and detectable if they are being searched for in more than a layer of the network. Butt’s technique works on the principle that a low participating member in a network layer might be active and be a key player in another layer, which might raise its vulnerability. The technique is accomplishable using a number of datasets like bank accounts transactions, phone conversations, short message services (SMS) and air-flight traveling.

Other mitigating strategies are research works on link-prediction [16], [25], missing node information [23], [26], [27] and node discovery [13], [28]. Link prediction was proposed and implemented on missing node information. Devised techniques get links that are not officially reported in the dataset and the structure [12] which connect the set of nodes that were not initially connected. This is known as Link discovery or prediction [16], [25], [29]–[31]. Contrary to link prediction, node discovery was designed for identification of nodes in the midst of existing nodes, but with a particular attribute [13]. ‘Latent structure’ was defined in ‘node discovery’ as a place within a social network where the node may finally emerge [13]. Link discovery is about inserting new and unknown links while node discovery does not insert new or unknown node(s) into the structure. Link prediction could be useful as it might increase the vulnerability of a node (participant) through the number of new links discovered/predicted [12], [16]. Calderoni [32], [33], and Bright [34] exploring multi-layer network with betweenness and degree centrality tools to ameliorate the effect of missing information on drug business OCG. Synergy from multi-relatin network analysis and degree centrality is about raising the vulnerability of low social-capital key players [24]. But the incorporation of betweenness centrality in multi-layer network search for hidden players that connect two layers within network i.e. actors with high betweenness centrality as a result of connecting layers in a network [35].

Analyzing multi-layer networks lower tendency of key players from being elusive. Calderoni [33] still reported fugitive of a key member in the drug business. The key player still escaped detection despite the deployment of ameliorating strategies - as a result of not participating in telephone conversations with other members. Criminal stakeholders (CS) as participants/affiliates whose roles and contributions in crime data record could be significantly low in order to cover their vulnerabilities and influential status [2]. High human-capital profile actor is expected to be more susceptible in the multi-layer network analysis. Combination of betweenness and degree centralities deployed for studying strategic positioning and vulnerability of high human-capital actors show that strategic positioning in mafia network and high social status have limited association with network centrality [36]. But strategic positioning is expected to minimize the vulnerability of high human-capital actors as they were not actively involved in criminal activities.

It is obvious that affiliate could be perpetually having a low profile-social status or abruptly omitted in a number of network layers in which multiple sources of data on crime will not yield an effect on missing key player. It becomes necessary to find feature to identify an external participant in a criminal network or a CS. As it is noticed that an outsider could be at the edge of taking advantage of data defectiveness than an NL. It can be said that missing information held efficiency of detective techniques to ransom [17] as well as criminal intelligence investigation [33], [37].

Social network analysts and criminal investigators found SNA being powerful for mining of hidden attributes as it offers indications on level of participation [37]–[39]. The SNA tools are recognized for classification of structural components (criminal members) [40], [41]. Also being employed for compartmental visualization [1], hierarchies [4], [42] and influential status of members, primarily to identify, distinguish, predict and confirm structural attributes of some nodes. It can be used for decision purpose i.e. determine the most vulnerable actors, or influential members [15], [34]. However, SNA lack capacity in identifying a CS due to incompatibility of attributes because the influence of a CS does not determine its vulnerability.
STAKEHOLDER’S ATTRIBUTE AND VULNERABILITY CHALLENGES

Attributes define behavior, characteristic upon which an entity can be identified. Stakeholders are individuals who have an interest in a business, or system. Level of involvement is depicted in a network structure as a number of links incident on a node. Quantum of links incident on a node might be too boggis to quantify roles played by an individual. Use of links in an SN serves the purpose of informing analysts about the internal structure of a system. Sometimes in a CN, unweight links are used in the system model to save study from running into the complex evaluation. Use of unweighting links streamlined network structures to only relationships by cutting off participants’ status which could be awkward to depict and evaluate. It is cleared that a CS could be an external member of the organization. And contributions of those members from outside to criminal activities will always be infinitesimal in social profile status compared to other active internal criminals [12], [13]. A CS can be simply described as a key member with relatively low social status as a result of low participation and links reported in the dataset [37].

Vulnerability measures how reachable or accessible a node is ‘within a structure’ [10], [15]. The vulnerability had been infatuated and infused into the attribute of covert nodes. It assumes attribute for mining hidden nodes. A node is said to be covert or hidden if it is well-embedded in network structure; one that is inconspicuous to observers and investigators as being important [12], [43]. Technically, it denotes a node that controls the resources of an organization [38], [44] or a node that occupies a central position of a structure[2]. Based on controlling resource, it emphasizes communication power. A node with high communication power might be well embedded and then become completely inconspicuous. It had experimented that the vulnerability of a covert entity increases as it’s in the ks increases or lie close to the center of the structure [2], [45], [46]. Depicting participation level with link will definitely lower social-profile of a CS status in the structure and in turns lower vulnerability for participants who at the edges of evading detection. This reveal discrepancy between a SNA and a CS from participation level and links representation.

The vulnerability of a node is also conceived from an influence perspective [47], [48]. The most influential covert node in a structure is an NL or criminal leader. a NL is the most influential in the structure because very important links go to him, he has the highest number of important neighbors (criminals), for controlling network resources through communication, approves submissions and operations, or actively involved in the affairs. The list of activities, number of cohorts and availability of metadata might increase the influence of participants in a crime. Also, the number of relationships, transactions, and other inimical activities engaged in by an actor could increase the chance of surveillance monitoring on suspected social-capital influential criminals before he could become an influential member in a criminal group. Ranking of influential members will finally determine most vulnerable influential criminals. And an NL has a high probability of falling within the list of influential members due to relative activities and cohort than a CS. This feature is being measured by SNA metrics [38], [49].

Metrics for measuring node’s influence take cognizance of the global effect of a particular node on entire network nodes [47], [50]. Assessment involves an indirect approach like shortest paths used in-betweenness, closeness and eigenvector centralities or combined [39]. With links, a node with relatively high links often emerges, while relatively insignificant links become unimportant. It signifies that an important, real-life influential criminal will be rendered as unimportant and unimportant. SNA evaluates only ‘structural’ influence and not real personality based influence. A node becomes vulnerable if it has a high score in any centrality metric deployed for monitoring influence. Criminal stakeholders concede to influence because their contributions fall short of values. This signifies that if SNA is deployed for mining stakeholder on the conception of influence, the outcome will be filled with false-alarm because SNA’s detection appeals to nodes with a relatively high number of links which criminal stakeholders might not subscribe to. There is a high need for analytical concept suitable for identifying important members who are not vulnerable to SNA.

Analytic Filtering Feature/ Separating Technique

Filtering is a technical term for removing noise from the main signal. The unwanted signal here is a controversial feature on CSs in order to identify appropriate class where a CS belongs. Analytic Filtering Feature-AFF is conceived from the phenomenon used in separating two immiscible liquids that have close physical properties like water and paraffin (kerosene). The separating funnel is
laboratory equipment for separation of two immiscible liquids. The point is that affiliate criminal will remain undetectable and difficult to distinguish if techniques fail to consider fundamental about criminal participants. A CS shares attribute it does not possess making it camouflage as a social-capital influence but also clone among a set of nodes it does not have a membership.

Previous sections revealed that influence can be classified into human-capital and social-capital. It is also drawn that social-capital profile of nodes (actors) is assessable using SNA metrics while individual personal profile known as human-capital based influence cannot be subjected to evaluation. In addition, influence can be used to categorize the level of vulnerability of criminal participants against strategic positioning in the structure [36]. The type of influence being used for vulnerability and strategic positioning is social-capital based [21], [37]. Part of filtering steps is a selection of influential-based profile for classification of actors.

**Classification**

Classification informs analysts about distributions or groups. The two major classes that criminal nodes are being grouped into overt and covert members; active and passive members; influential and non-influential. Overt members are referred to as known or apprehended criminals, while covert members are unknown criminals. This classification does not require technique. The classification into active and passive requires measurement. The outstanding technique for measuring activeness is SNA. Activeness is relative to the number of links incident on a node or proximity to the center of the structure, while passive nodes are not.

Classification of influential status often involves rigorous mathematical assessment of all nodes in the structure. At times, optimization algorithms are also deployed. Outstanding results from the influence assessment still maintain the status quo with centrality. Eigenvector, Betweenness, and Closeness centralities are an example of influence measuring tools. A highly influential node is found to be very close to the center.

**Filtering Of Ascribed Attributes**

This section tries to obtain attributes that CSs have in common with other stakeholders that had been mining in the name of covert nodes. While some attributes are ascribed to a CS in the pretense of the influential nodes, they are also being filtered to identify exclusive feature(s). Although criminal stakeholders can be treated as influential covert individuals to their criminal groups, and network leader is the most highly influential member in the network structure. Some of the major difference is pointed out below.

**a. Leadership:**

A leader in community network has cordial relationships with ‘majority’ of overt members. Stakeholder- Lack of cordial relationships with ‘majority’; denies a CS leadership

**b. Participation:**

Participation in inimical activities, nefarious acts, and conversation with overt members (raises activeness status).

Stakeholder – abstain from inimical activities and conversation with overt members (lowering activeness status)

**c. Influence Spreading Capacity:**

A node with relatively high direct relationships has potent for spreading capacity Stakeholders abstain from relationships/transactions with overt members to safeguard detection (lowering spreading capacity)

**Extrapolation of Structural Attributes for CS**

Three properties are presented for re-classification of CS in a structure. These properties are defined based on topology. Leadership in a social network structure (SNS) is mostly accorded to a node that is active; significant participation and spreading influence rate. But influence ascribed to a CS cannot earn him leadership. Therefore, a CS can be adequately described as a silent influencer because his participation and spreading rate cannot aid accessibility or vulnerability. It suggests that CSS should be searched for among passive classified nodes.

**a. Leadership:** CS is passive

**b. Participation:** CS is passive

**c. Spreading rate:** CS is passive

All indicate that a CS will be a passive leader, with passive participation and passive spreading rate. The inference above pointed out that a CS is a member of passive nodes. Therefore, mining a CS through the SNA technique might not be effective or lead to high false-alarm.

**Latent Attributes**

The section within the NS where a CS is lying had been identified, it remains attribute(s) to.

---

*Corresponding author: Abideen, A. Ismail. ismail.pg.610938@st.futminna.edu.ng Telecommunication Engineering Department, Federal University of Technology, Minna. © 2019 Faculty of Technology Education, ATBU Bauchi. All rights reserved*
distinguish a CS from other passive nodes. Passiveness covers all nodes that are not leader i.e. less vulnerable nodes irrespective of a number of links incident on them. It includes participants that Key Player Problem Positive/Negative (KPP-Pos./Neg.) cannot identify[20] because of personality influences the position of some key players. It should be recalled that passiveness is statistical-based classification for nodes[17], [37], [40], [51], [52]. The definition of passiveness cut across all nodes except the one detected as a network leader (NL) or set of nodes with the same highest centrality values which automatically include a CS. One clarification on a CS is that it is a covert node that dares not attempt to be an NL. This real-life attribute of a CS affects structural attributes that increase the number of invulnerability covert nodes.

There is an obvious dichotomy between a CS and an ordinary overt member (OOM) in a criminal organization, which is that OOM is seldom selective in relation. An overt node relates with a large number of members haphazardly in order to earn a high level of participation in turn for structural influence. But a CS applies care to his relationships i.e. avoids a relationship with OOM. A CS prefers the choice of an influential OOM who is an emerged leader than a pool of OOM. Seldom and restricted relations protect a CS better. Motivations behind preference relationship with a network leader over OOMs curtail CS’s level of participation, influence in the structure and vulnerability. Equation (1) and (2) present categories of nodes a CS and OOM relate with. Equation (3) denotes that the influence of a covert node \( j \) \( I_f_j \) is grossly affected by relationships from both NL and OOM.

\[
G_r = \{ r_N, r_O \}
\]

(1)

Where \( r_N \) and \( r_O \) denotes quantum of relation with NL and OOM; overt members. And \( C_r \) is a factor to determine if a node in question is truly an OOM or CS.

\[
C = \{ r_N \}
\]

(2)

Equation 2 shows the model expression for a node that has an ego in relating with OOM

\[
I_f_j = \sum_{r \in r_N} (r_N | r_O)
\]

(3)

A node can annex its influence through interaction with influential nodes; NL and OOM but a node (actor) that decides to interact with only influencers cannot access such influence. Finally, the vulnerability of a covert CS depends on identification of a criminal NL. Extrapolation or filtering feature is able to identify two covert nodes that have similar properties; CS and OOM. And the latest feature is highly appealing.

CONCLUSION

Stakeholders are significant in criminal organizations as their roles reinforces the criminal trend. Scarcity of data on stakeholder undermines detective techniques and security operative efficiency. Analytic approach renders a formidable process to understand factors keeping CSs undetected from metadata structure. The solution offered is the identification of latent attribute of stakeholder from SNS especially obtained through electronic communication metadata. Finally, stakeholder’s position is found vehemently attached to the network leader. This latent feature for mining a CS is logical and it can be subject to changes as a result of transition or external effect that do affect criminal organization structures such as network disruption.

REFERENCES


[8] M. Salehi, H. R. Rabiee, and A. Rajabi,


*Corresponding author: Abdeen, A. Ismail. ismail.pg.610938@st.futminna.edu.ng Telecommunication Engineering Department, Federal University of Technology, Minna. © 2019 Faculty of Technology Education, ATBU Bauchi. All rights reserved


