JJEE
Volume 7, Number 4, December 2021
Pages 378-389
Jordan Journal of Electrical Engineering
ISSN (Print): 2409-9600, ISSN (Online): 2409-9619

Criminals Detection in Social Networks Using Centrality Measures Algorithm

Zeinab Alebouyeh[1], Amir Jalaly Bidgoly[2*]

[1, 2] Department of Computer Engineering, University of Qom, Qom, Iran
E-mail: jalaly@qom.ac.ir

Received: March 21, 2021 *Revised: May 01, 2021* *Accepted: May 06, 2021*

***Abstract*—** Despite the advantages of social networks, they may be a good platform for some crimes such as drug marketing. Unfortunately, due to the large amount of information on social networks as well as the anonymity of users, it is very difficult to identify and detect these crimes; so it is necessary to provide automatic tools to search and report criminals on these networks. So far, several methods are introduced with the capability of automatically detecting criminals. However, all of these methods require access to content published by the users. In this paper, a new method is proposed. It is capable of identifying criminals in social networks based not on their published content but only on their social relationships. The proposed method is based on the assumption that criminals, indirectly, have strong connections with each other. It includes two algorithms, the first algorithm is utilized for crawling the social network, and the second algorithm for detecting criminal users among the users collected in the first phase. Having an initial set of criminals, the method first crawls the network starting from this set. The crawler is configured to collect users who are more likely to be criminals. Then, these users and their relationships form a graph, and users are ranked based on five centrality measures (namely degree, betweenness, closeness, hubs and authority centralities) that have a strong correlation with the likelihood of users being criminals. The obtained results show that the proposed method can well identify criminals and rank them. The degree and closeness centralities showed the best results while betweenness centrality showed the worst results. For instances, the closeness centrality has been able to correctly identify criminals with 90% accuracy.

***Keywords*—** Criminal detection; Social network analysis; centrality measures algorithm; crawling algorithm; Criminals detection algorithm.

1. INTRODUCTION

According to a senior UN official, crime generates billions of dollars a year worldwide making it one of the top twenty economies in the world. He said that the revenue from "illegal trade" each year is around 7% of the size of the global economy [1]. Today, cyberspace has become a new market for criminals due to the inherent anonymity of the environment and the fact that users are hardly traceable, which gives them a sense of security in these networks. Thus, the social network is considered a very good platform for crimes, such as the selling of illicit goods, illegal drug trafficking, narcotics and cold arms. Crime, on the other hand, imposes significant costs on society at the individual, social and national levels, and thus, programs that directly or indirectly prevent crimes can have significant economic benefits [2]. Therefore, one of the most significant police challenges in dealing with crime is to detect criminals and then prevent them from working in cyberspace.

Criminal investigations also take a great deal of human and technical resources to identify the criminals and to track down those responsible for the crime [3]. Given that resources are always scarce, it seems necessary to provide solutions that can make the most of existing resources and produce satisfactory results. Researches have shown that the use of social networks and the study of data and relationships in these networks to detect crimes and criminals are one approach that is useful in the criminal sector [4]. Social networks

* Corresponding author
Article's DOI: 10.5455/jjee.204-1616319844

contain a large number of very valuable data; this vast repository of information - which has the potential to identify and detect crimes and criminals - can be used by security and law enforcement agencies to pave the way for a more safe society [5].

In this paper, we have proposed a novel idea to detect criminals in social networks. For this purpose, the specification of certain criminals (buying and selling drugs) on the social network was obtained from one of the legal entities (police). These accounts are considered as the initial set of criminals and then their relationships are obtained through crawling the social network. Based on the information obtained, a graph of user's relationships is constructed. Then, various graph evaluation measures such as centrality measures including degree centrality, betweenness centrality, closeness centrality, and hub and authority are examined on all graph nodes. Each of the measures gives a different ranking of the likelihood of users' criminality. Finally, this article examines which of these measures best identifies criminals on the network. It is notable that - in this investigation - the detection of criminals is done without analyzing the content (without examining the likes, text of the comments, and profiles of the people) of social networks and is based only on the existence of communication between people. Only the information of public accounts is crawled; however, criminals whose accounts are private can also be identified by crawling and collecting information from the public accounts of other users in the network.

The rest of the paper is organized as follows: in section 2, related works are reviewed. In section 3, the proposed method is explained in three parts (overview of the method, measures used, and suggested algorithms). The evaluation and results of the proposed method are discussed in section 4.

2. RELATED WORKS

So far, a variety of studies has been done to detect crimes and criminals on social networks. Analyzing the criminal network in order to detect important people or network leaders, discovering patterns of crime, early detection and prevention of crime, predicting law offenders and potential criminals and predicting the location of crime are some of the examples that have been discussed in recent years by social network researches. The most effective approach used in recent years to detect criminals is to analyze the content of people's communications through various networks. Jie and Huadong suggested a framework for detecting criminal gangs, using data from different resources (including bank account data, operator contact data, social network activity data, etc.) and analyzing them to focus on community detection in a criminal network [6]. After discovering the community, the members of each group are ranked in terms of importance. In another study, considering six different features, namely economic status, family background, educational level, alcohol, and drug abuse and human criminal records, the committing of a crime by individuals is predicted by the use of fuzzy techniques. In this method, by defining different fuzzy rules, a value is defined for each person, which helps to identify the individual's criminal psychology [7].

In recent years, the approach of several studies has been to investigate criminal networks. The members of the network are all criminals in this way, and the aim is to identify the group leaders and key individuals in the network. In this way, several methods have been proposed that use centrality measures and graph analysis parameters for

understanding the hierarchy within criminal organizations, identification of central members and those who have a key role in the network. In some articles, closeness and betweenness centrality have been used to analyze the activities of a Russian mafia group [8] and to predict criminal leaders in an Italian mafia group [9]. In another study, a solution was proposed using the degree, closeness, betweenness centrality, and eigenvector to categorize the subjects of the messages exchanged, and prioritize the likelihood of individuals being criminal [10]. In [11], a platform is presented that - using the measures of centrality and graph analysis - provides an in-depth understanding of the hierarchy within criminal organizations and identifies central members and those who play a key role in communication between members as well as communities in the network.

Many studies have been conducted to identify suspicious profiles on social networks. In [12], a framework has been proposed to identify suspicious profiles on social networks. The purpose of this work is to identify suspicious profiles located in the close circle of contacts (first-level contacts who are often trusted by the security policies of the operating system) of a particular person. This framework is based on three indicators: balance, energy, and anomaly resulting from the daily activities of users. The solution presented in [13] is based on creating a "honey profile" on social networks. This profile is designed to capture and collect data on malicious activity. In this method, features such as URL ratio used in messages, number of messages sent, the similarity of messages, number of friends, followers, etc. are collected to identify malicious people. Another article provides a way to evaluate messages posted on Facebook. This study categorizes messages into three categories: legitimate, spam and malicious. In this method, machine learning algorithms are used to classify the messages [14]. Gayo-Avello and Brenes have used graph centrality algorithms to identify and evaluate spammers in Twitter [15]. Halim and Gul provided a way to identify people involved in malicious communications on Facebook. Their method consists of two steps: i) semantic analysis to identify malicious posts and ii) temporal-spatial position tracking analysis of activities performed among malicious users [16]. Other articles provide methods for identifying fake and spam profiles on LinkedIn and Twitter that use analysis of people's general characteristics [17, 18].

Using text analysis to identify suspicious users on social media also presents an important challenge. There are several ways to identify the meaning of phrases, and much work has been done in this area. Alami and Beqqali provide an automated system for identifying suspicious profiles on social networks using text similarity metrics. The proposed idea is to calculate the similarity distance to detect suspicious posts to identify suspicious profiles [19]. Another paper presents a way to identify cyberbullying using text analysis and machine learning algorithms [20]. Another article hypothesizes that words used on social media can help identify suspects. It identifies criminals by analyzing messages exchanged on social networks based on a series of controlled words in specific areas such as terrorism, cyberbullying, etc. [21]. Another article presents a method that selects social media posts with criminal slang terms and automatically classifies these posts according to illocutionary classes and using machine learning methods. This method is used to select suspicious posts and decrypt them, and criminal intent is automatically classified in posts written on social networks based on a trained model [22]. In that article, semantic analysis of social networks has been used to identify the central actor of crimes. In [23], different measures of graph

centrality have been used. Relationships between people are based on friendship, family, cooperation, etc.

Along with all the methods proposed to identify criminals, the question that arises is whether or not having the specifications of certain criminals in a social network will identify other criminals in the network by analyzing relationships? Xu et al. [24] proposed a relevant work in this regard, with content analysis and using a clustering algorithm; they calculated the weight of people's communication with criminals and innocents and then compared these two degrees to determine whether a person is criminal. Communication weight between people is obtained based on the topic of discussion exchanged, analysis of the content of the messages exchanged and the number of communications.

3. THE PROPOSED METHOD

3.1. An Overview of the Method

Overview of the proposed method is represented in Fig. 1 which shows that the proposed method has two main modules: i) the *crawling* module that is responsible for collecting information from the network and ii) the *detection* module that ranks users based on the probability of being criminal. In both modules, some measures - which will be discussed in the next section - are used to identify users and calculate the likelihood of being a criminal. In the first module, with an initial set of known criminals, the social network crawls. This module collects information of users from the network and at the same time tries to find users who are more likely to be criminals. For this reason, crawling is started from the users who are directly connected to the initial set of criminals since it is assumed that people who are in direct contact with criminals are more likely to be criminals. In the second module (detection module), the probability of users being criminal is estimated by different measures.

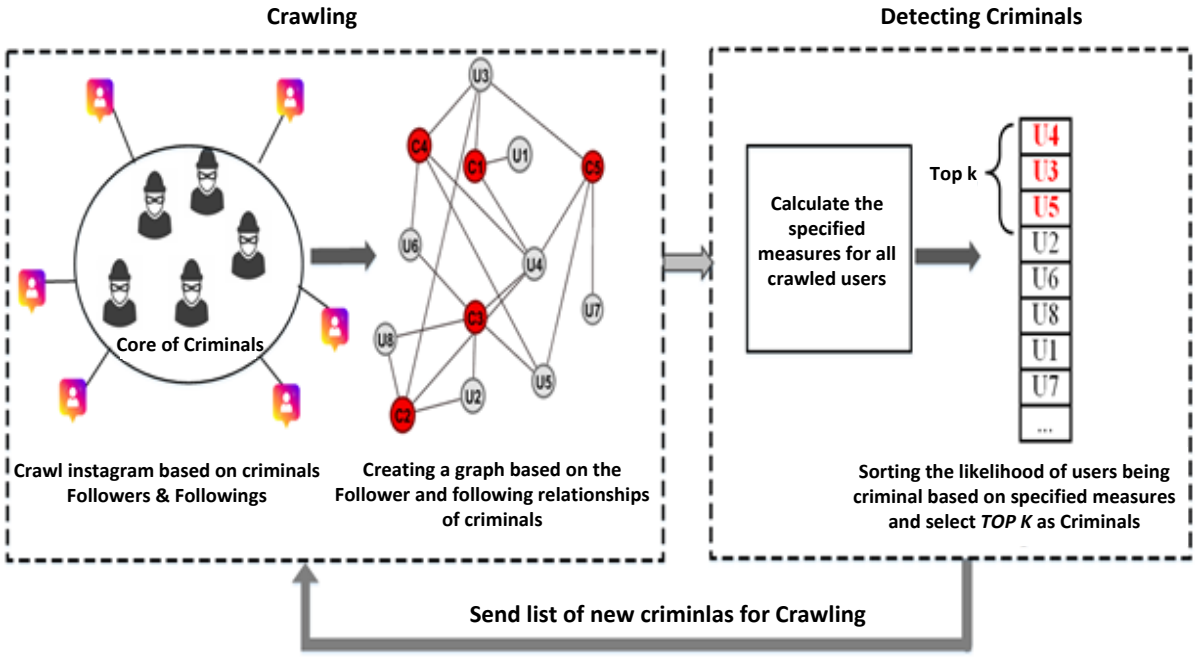

Fig. 1. An overview of the proposed method.

This estimation is only based on the edge relationship between users. Then, by determining a threshold for each measure, users who have more likelihood of being criminals are introduced as criminals. To crawl more information from the network, these criminals are sent back to the crawling module as the new set of criminals. In fact, the proposed module has a recursive structure and can be run as many repetitions as required to extend the set of criminals. In each run, according to the defined measures, the criminal set is revised to include the newly selected users who are more likely to be criminals and the rest of the users who are less likely to be criminals are excluded. In this way, the criminal set is updated at every run, and then the process continues based on the new set of criminal users.

3.2. The Used Criteria

In this research, network crawling has been performed with the initial assumption that people who have a direct relationship with criminals on the social network are more likely to be criminals than others. After crawling information about users in the network, according to various measures, the probability of users being criminal is calculated. To this end, different centrality measures have been used as a measure for detecting criminals. In this section, five important centrality measures used in this study are summarized below.

Degree centrality: The simplest centrality measure is degree centrality. This measure may contain important information for many applications [14-25]. The measure indicates the extent to which the node is active in a network [26]. This measure is defined as the number of direct links of a node in the network [27]. Based on the crawling method performed in this study, in the graph obtained from crawling, users with a higher degree represent users who have more direct communications with criminals and this raises the likelihood that these users being criminals. The degree centrality measure for each user is calculated according to Eq. (1) based on degree centrality in [26]:

$$Cent_{Deg}(v) = \boldsymbol{F_r} + \boldsymbol{F_g} \quad (1)$$

where F_r: number of criminals that follows v and F_g: number of criminals that v is one of their followers.

Betweenness centrality: Betweenness centrality determines the extent to which a particular node in a network is in the communication path of other nodes [28]. Regarding the fact that the graph in this study is composed of people with a higher likelihood of criminality than others, this measure can somehow show the intermediates on the network. This measure is calculated as following [27]:

$$Cent_{Bet}(v) = \sum_{\substack{u,w \in V \\ u \neq w \neq v}} \frac{\sigma_{uw}(v)}{\sigma_{uw}} \quad (2)$$

$\sigma_{uv}(v)$ = total number of shortest paths between each pair of users like u and w that pass through user v and σ_{uv} = total number of shortest paths from u to w.

Closeness Centrality: This measure indicates the average distance of one node from the other nodes of the network. The closeness of the node is the inverse of the total distance of this node to all other nodes [28]. The higher this measure for a node, the better the access to information in other nodes or the more direct effect on the rest of the network nodes [26]. In our network, it is assumed that the higher closeness centrality for users, the closer these

users are to other criminals, so the more likelihood of being criminal. The measure is calculated as follows [27]:

$$Cent_{Close}(v) = \frac{N-1}{\sum_{u \in V} d(u,v)} \quad (3)$$

where $d(u,v)$ is the distance of user u from user v and N is the total number of users.

Hubs and Authorities: There are two important types of nodes in networks: authorities,which are nodes that contain useful information about a topic of interest, and hubs, which are nodes that tell us where the best authorities can be found [26]. According to these two measures, the network is divided into two parts: nodes that are inherently criminal (authorities) and nodes that are customers of criminals or somehow have an interest in criminals (hubs). Authority and hub are calculated according to Eq. (4) and Eq. (5), respectively according to [29].

$$auth(v) = \sum_{u \in V_{to}} hub(u) \quad (4)$$

where V_{to} is all users which follows user v.

$$hub(v) = \sum_{u \in V_{from}} auth(u) \quad (5)$$

where V_{from} is all users which user v follows them.

3.3. The Proposed Algorithm

The proposed method includes two algorithms, the first one for crawling the social network, and the second one for detecting criminal users among the users collected in the first phase. These algorithms are shown in Figs. 2 and 3, respectively. As depicted in Fig. 2, the crawling algorithm takes the initial list of criminals and the number N as input, and then crawls the network based on the followers and followings of the initial list. The output of this algorithm is a graph with a set of V nodes and a set of E edges. If the number of crawled users reaches N, the crawling process will be terminated and the *DetectCriminals* function is called for identifying criminals according to defined measures. Otherwise, the *DetectCriminals* function is called for crawled users from this step ($V1$) to identify new criminal users, and the crawling function is re-called with the new criminal user list. This will continue until an acceptable number (N) of users has crawled.

Algorithm I: Crawl Instagram

```
Function: Crawler
Input: CriminalList[],N
Output: G(V,E)  ➔ A Graph constructed from Instagram users who may be criminal
Begin
(V1,E1)=crawl followers & followings of CriminalList[]
Add (V1,E1) to (V,E)
Build G (V,E)
 if (len(V[]>=N))
        DetectCriminals (G(V,E))
else
        newCriminalList[]=DetectCriminals (G(V1,E1))
        Crawler(newCriminalList[])
end
```

Fig. 2. The crawling algorithm.

The second algorithm (shown in Fig. 3) is for criminal detection. It takes the output graph of the first algorithm and the number *k* as input, and returns a list of *k* users - which are more likely to be criminal than the others - as output. In this algorithm, the measures mentioned in section 3.2 are calculated for all crawled users, and the result for each measure is arranged in descending order. Then, K users who are at the top of each list (*CriminalList[]*) are identified as criminals. An example of network crawling is shown in Fig. 4 in which the network is crawled in three steps.

Algorithm II: Select Top k users with the highest likelihood of being criminal

Function: DetectCriminals
Input: *G(V,E), k*
Output: *CriminalList[]* *A list of criminals*
begin
foreach *v in V[]*
Compute (CentFunction(v))*
end
Sorted[]=Sort By (V[],CentFunction(v))
CriminalList[]= select first k users from sorted[]
return *CriminalList[]*
end

*CentFunction can be any of the centrality functions ($Cent_{Deg}(v)$, $Cent_{Bet}(v)$, $Cent_{Close}(v)$, auth(v) and hub(v))

Fig. 3. The criminals detection algorithm.

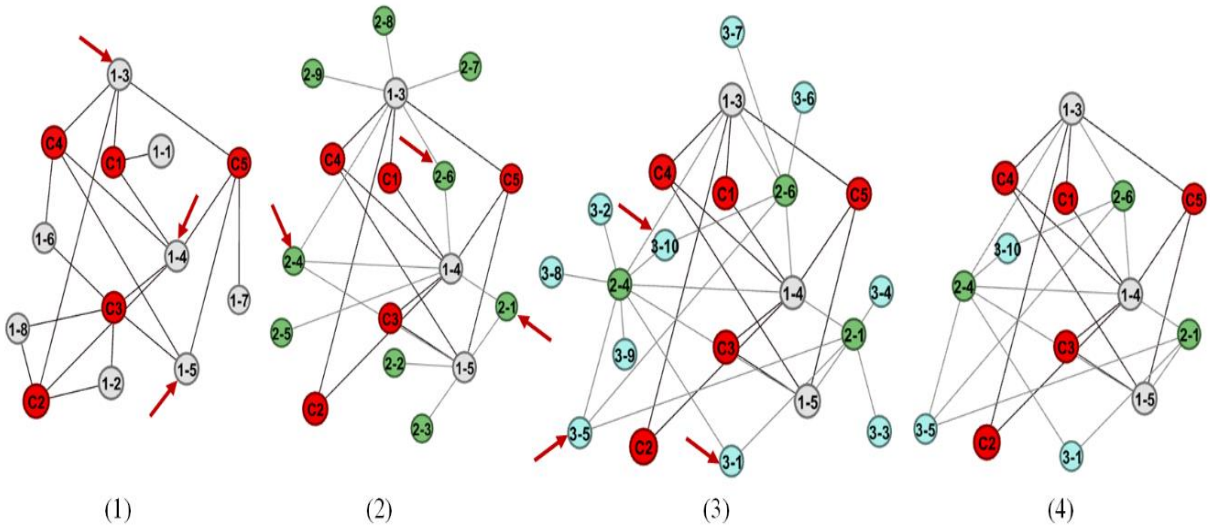

Fig. 4. Example of crawling steps: nodes that start with the letter C are the initial set of the criminals, and the crawl in the first step begins with these nodes. Crawling nodes in step x are denoted by x-. For example, node 3-5 is the node resulting from clawing in step 3. In this example, when users are sorted according to centrality measures, in each step the 3 users at the top of the list (Top3) are kept in the graph and the rest of the nodes are removed, in the next step, the graph crawling continues with the remaining three nodes from the previous step. This example illustrates the three stages of crawling. In the figure, the nodes that are selected in each step for crawling in the next step are shown with an arrow. The graph obtained in step 4 is the result of three crawling steps. At this stage, different centrality measures are applied to the graph and potential criminals are identified.

4. EXPERIMENTS

In this section, the method of data collection is explained and the results are analyzed. For this aim, we have evaluated the proposed algorithm on Instagram social network with an initial set of criminals (in drug context). The proposed method has been evaluated in this environment to determine whether in practice it can identify other criminals only from their social relations in the social network or not.

4.1. Data Collection

To collect users data from the Instagram network, our scheme is implemented in Python 3. For this crawling process, Instagram mobile APIs have been used and only the followers and followings APIs have been used. This code is executed on an internet server and the data were collected over a period of about two months. The collected data only contains the usernames of users in the Instagram network and the list of their followers and followings.

Initially, the Instagram account of 27 known cyber criminals related to buying and selling drugs was obtained from a legal entity (Police), of which 16 accounts were public whose followers and followings information have been crawled. The average number of followers and followings of initial set users is about 2,700. During several stages of the crawl, 125,515 Instagram accounts were collected, but the crawling process was performed only for users whose accounts were public.

4.2. Results

We converted the users' information collected on the Instagram network into a graph using a Python code and NetworkX library (a Python library for studying graphs and networks). The resulting graph nodes represent Instagram users, while the graph edges represent user relationships (follower and following). Then different graph centrality measures in NetworkX were calculated for all graph nodes. Finally, the graph nodes were arranged in descending order based on the results obtained from each of the calculated centrality measures. The lists obtained from the various centrality measures indicate the likelihood of individuals being criminals. This means that users at the top of the list are more likely to be criminals than users at the bottom of the list. Therefore, for different centrality measures, a number of users are labeled criminals (users at the top of the list) and a number of users are labeled non-criminals (users at the bottom of the list).

Using the aforementioned method, a number of criminals and non-criminals have been identified on the Instagram network. The validity of the labels (criminal and non-criminal) assigned to the users has been questioned by a legal center (e-Police) and their responses (the accuracy, precision and recall criterion) have been used to validate the results. Accuracy determines how many users are identified correctly by the system as criminals and non-criminals.

The precision criteria indicate that how many of the users - that the system has identified as criminal - are actually criminal. Precision is a good measure when the costs of false positive are high. In this case, where the detection of criminals is considered positive, false positives are not very important because the goal is to identify as many actual criminals

as possible. Here are some tips about false positives or in other words, about people who are not actually criminals, but have been identified by the system as criminals:

- It is true that these individuals are not drug sellers, but most of them are people who somehow supply drug manufacturers and sellers with goods or services. For instance, drug growth fertilizers or indoor LED growth lights where most of their followers and followings are criminals who buy and sell drugs.
- Some other users are also more likely to be drug users and are actually customers of criminals because a high percentage of their followers and followings are criminals.

Recall is a good metric model when there is a high cost associated with False Negative. It shows how many actual criminals the system has identified as criminals. Since our goal is to identify criminals in the network, this criterion is very important because we need to identify as many criminals as possible and, consequently, stop their activities. The higher the recall, the higher the model's ability to identify the actual criminals. As seen from the results, exhibited Table 1, the value of recall for all centrality measures is one, i.e. measures did not have a false negative, and at the specified threshold (Top10, Top20, and Top30), the model identified all actual criminals. This shows that the model has achieved its goal of maximum detection of criminals and in the specified threshold; there are no real criminals that the system does not correctly identify.

Table 1. The accuracy, precision and recall of different measures.

Measure	Top x	$Cent_{Deg}$	$Cent_{Bet}$	$Cent_{close}$	Hub	Authority
Accuracy	10	0.95	0.9	0.95	0.85	0.9
	20	0.9	0.87	0.9	0.85	0.9
	30	0.93	0.81	0.93	0.86	0.9
Precision	10	0.9	0.8	0.9	0.7	0.8
	20	0.8	0.75	0.8	0.7	0.8
	30	0.86	0.63	0.86	0.73	0.8
Recall	10	1	1	1	1	1
	20	1	1	1	1	1
	30	1	1	1	1	1

Obviously, because five centrality measures (as introduced in section 3.2) have been used to identify criminals, there are various lists of potential criminals. In order to determine which of these measures is best suited to detecting criminals, the accuracy, precision, and recall are calculated for all of them as seen in Table 1. Due to large number of users, the measures are calculated only for the three modes of *Top10*, *Top20* and *Top30* users of each list. For the *Topx* case, we calculate the measures for the top x users on the list that the system has labeled as criminals and the last x users on the list that the system has labeled as non-criminals.

No content analysis was done in the proposed work to identify the criminals, but the analysis of the user biographies, the content of the posts, the review of the likes and the text of the comments will undoubtedly make it easier to identify the criminals. As mentioned earlier, only public account information is collected, but this method is also able to identify the criminal's private accounts without directly collecting the information of their followers and followings; for example, in the degree centrality measure, from the 30 users at the top of

the list who were identified as criminal, 9 accounts were private and after review, it was determined that they were really criminals. In the authority measure, from 30 users at the top of the list, 15 private accounts turned out to be real criminals. As a result, among the five measures introduced, degree and closeness centrality have the highest accuracy and betweenness centrality and hub have the lowest accuracy in detecting criminals.

5. DISCUSSIONS

In this section, we will discuss the results, limitations, and future works. Regarding the performance of the five different measures, although they work well to measure the criminality of people, there are some points about their performance:

- The degree centrality is one of the measures that has the most accuracy. This means that users who have a higher degree centrality than others are more likely to be criminals and this result is not unexpected because these nodes have more direct relationship with criminals.
- Betweenness was one of the measures with less accuracy; perhaps the reason is that betweenness neglects the intensity of the relationship and focuses only on the shortest paths. On the other hand, this measure sees all nodes as the same type, while in our network, a number of nodes are the initial set of criminals, and the relationship with them is very important to determine criminality. This measure can be very useful for identifying network leaders if all members of the network are criminals.
- For closeness centrality, because the network's crawl was based on the initial set of criminals and in expanding the network, the focus was on communicating with criminals, this measure has good accuracy for detecting criminals. Of course, if we consider the ratio of closeness to criminals compared to closeness to other network nodes, the accuracy of this measure will probably increase.
- Authority centrality is similar to the in-degree centrality. Since the initial crawl of the network was based on the relationships of the criminal nodes. The high value of this measure indicates that many criminals have followed these nodes; so the probability of these nodes being criminal is high and this measure has good accuracy.
- The hub measure has the least accuracy, because hub centrality is similar to the out-degree centrality. The high number indicates that these nodes have followed many criminals and in fact are interested in criminals; for example, these nodes may be the customers of criminals.

Regarding limitations, this method is based on two assumptions: i) we have an initial set of criminals and ii) criminals have relation in the social network. If either of these two assumptions is incorrect, the proposed method may not work properly. If the initial set of criminals is not available or if there are non-criminals in the set, this method cannot crawl other criminals in the network. The results are very dependent on the initial set. Regarding the second assumption, we showed that criminals, although not directly but indirectly, have strong connections with each other. However, this assumption has been just tested and proven in Instagram, and in the case of other social networks, a separate study and research is needed.

For future works, we can look for more complex methods like GCN for detecting criminals, and can look for methods that can be robust to faults and noise in the initial set of

criminals (i.e. the initial set may incorrectly contain some non-criminal nodes) as mentioned in the limitations. On the other hand, we have also planned to investigate the existence of social links between criminals on other social networks such as Twitter and Facebook.

6. CONCLUSIONS

The aim of this article was to identify criminals in a social network just based on social relations in an online social network. For this purpose, with a initial set of criminals, the social network was crawled step by step by considering which users are most likely to be criminals, indicating criminals and finally creating a graph based on the relationships between the crawled users. Then, based on different proposed measures, the probability of a user being criminal was measured. The proposed measures are based on the measurement of centrality in the graph. Finally, according to the results, several users at the top of the list were identified as criminals. It is noteworthy that only public account information is collected, but this method can also identify the criminals' private accounts without collecting the information of their accounts directly. Results show that the value of recall for all centrality measures is one and for accuracy, degree, and closeness centrality had the highest accuracy among the five measures introduced and can detect the criminals with up to 90% accuracy. Finally, as we seek to raise the recall criterion as much as possible, we can skip the accuracy criterion a bit. It should be noted that in addition to detecting criminals, the proposed method can be used in other cases to identify and prioritize nodes with specific characteristics.

REFERENCES

[1] Reuters, *Crime on the World's Top 20 Economies*, 2012. <https://www.reuters.com/article/us-un-crime/crime-one-of-the-worlds-top-20-economies-u-n-idUSBRE83M12P20120423>

[2] K. McCollister, M. French, H. Fang, "The cost of crime to society: new crime-specific estimates for policy and program evaluation," *Drug and Alcohol Dependence*, vol. 108, no. 1–2, pp. 98–109, 2010.

[3] J. Xu, H. Chen, "CrimeNet explorer: a framework for criminal network knowledge discovery," *ACM Transactions on Information Systems*, vol. 23, no. 2, pp. 201–226, 2005.

[4] J. Mena, *Investigative Data Mining for Security and Criminal Detection*, Butterworth Heinemann, 2002.

[5] M. Tayebi, U. Glässer, *Social Network Analysis in Predictive Policing*, Springer International Publishing, 2016.

[6] Y. Jie, W. Huadong, "Discovering gangs of criminals using data fusion with social networks," *International Journal of Database Theory and Application*, vol. 9, no. 6, pp. 33–44, 2016.

[7] S. Gupta, S. Kumar, "Crime detection and prevention using social network analysis," *International Journal of Computer Applications*, vol. 126, no. 6, pp. 14–19, 2015.

[8] F. Varese, "The structure and the content of criminal connections: the russian mafia in Italy," *European Sociological Review*, vol. 29, no. 5, pp. 899–909, 2013.

[9] F. Calderoni, *Networks and Network Analysis for Defence and Security: Identifying Mafia Bosses from Meeting Attendance*, Springer International Publishing, 2014.

[10] Y. Gu, W. Li, L. Zhang, M. Shen, B. Xie, "A prioritization algorithm for crime busting based on centrality analysis," *Proceedings of the 2nd International Conference on Electronic and Mechanical Engineering and Information Technology*, pp. 155–159, 2012.

[11] E. Ferrara, P. De Meo, S. Catanese, G. Fiumara, "Detecting criminal organizations in mobile phone networks," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5733–5750, 2014.

[12] C. Perez, M. Lemercier, B. Birregah, "A dynamic approach to detecting suspicious profiles on social platforms," *2013 IEEE International Conference on Communications Workshops*, pp. 174–178, 2013.

[13] G. Stringhini, C. Kruegel, G. Vigna, "Detecting spammers on social networks," *Proceedings - Annual Computer Security Applications Conference*, pp. 1–9, 2010.

[14] S. Abu-nimeh, "Malicious and spam posts in online socoal networks," *IEEE Computer Society*, pp. 23–28, 2011.

[15] D. Gayo-avello, D. Brenes, "Overcoming spammers in twitter – a tale of five," *Iriiuames*, pp. 41–52, 2010.

[16] Z. Halim, M. Gul, N. Ul Hassan, R. Baig, S. Ur Rehman, F. Naz, "Malicious users' circle detection in social network based on spatio-temporal co-occurrence," *Proceedings - International Conference on Computer Networks and Information Technology*, pp. 35–39, 2011.

[17] A. Al-Zoubi, J. Alqatawna, H. Faris, "Spam profile detection in social networks based on public features," *2017 8th International Conference on Information and Communication Systems*, pp. 130–135, 2017.

[18] S. Adikari, K. Dutta, "Identifying fake profiles in linkedin," *Proceedings - Pacific Asia Conference on Information Systems*, 2014.

[19] S. Alami, O. Beqqali, "Detecting suspicious profiles using text analysis within social media," *Journal of Theoretical and Applied Information Technology*, vol. 73, no. 3, pp. 405–410, 2015.

[20] J. Chen, S. Yan, K. Wong, "Aggressivity detection on social network comments," *ACM International Conference Proceeding Series*, pp. 103–107, 2017.

[21] É. Florentino, R. Goldschmidt, M. Cavalcanti, "Identifying criminal suspects on social networks: a vocabulary-based method," *ACM International Conference Proceeding Series*, pp. 273–276, 2020.

[22] R. Resende de Mendonça, D. Felix de Brito, F. de Franco Rosa, J. dos Reis, R. Bonacin, "A framework for detecting intentions of criminal acts in social media: a case study on twitter," *Information*, vol. 11, no. 3, pp. 154, 2020.

[23] S. Tahalea, S. Azhari, "Central actor identification of crime group using semantic social network analysis," *Indonesian Journal of Information Systems*, vol. 2, no. 1, pp. 24-32, 2019.

[24] Y. Xu, L. Mingyang, A. Ningning, Z. Xinchao, "Criminal detection based on social network analysis," *2012 8th International Conference on Semantics, Knowledge and Grids*, pp. 201–204, 2012.

[25] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications,* Cambridge University Press, 1994.

[26] M. Newman, *Networks: An Introduction*, Oxford University Press, 2010.

[27] L. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[28] M. Wang, W. Pan, "A comparative study of network centrality metrics in identifying key classes in software," *Journal of Computational Information Systems*, vol. 8, no. 24, pp. 10205–10212, 2012.

[29] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.