# Criminal Network Community Detection in Social Media Forensics

Ogerta Elezaj[1], Sule Yildirim Yayilgan[1], Edlira Kalemi[2]
[1]Department of Information Security and Communication Technology
Norwegian University of Science and Technology (NTNU), Norway
{ogerta.elezaj, sule.yildirim}@ntnu.no
[2]University of Tirana, Albania
edlira.kalemi@unit.edu.al

**Abstract.** Nowadays, Online Social Networks (OSNs) has created a breeding ground for criminals to engage in cyber–crime activities, and the legal enforcement agencies (LEAs) are facing significant challenges since there is no consistent and generalized framework built specifically to analyse users' misbehaviour and their social activity on these platforms. Data exchanged over these platforms represent an important source of information, even their characteristics such as unstructured nature, high volumes, velocity, and data inter–connectivity, become an obstacle for LEAs to analyse these data using traditional methods in order to provide it to the legal domain. Although numerous researches have been carried out on digital forensics, little focus has been employed on developing appropriate tools to exhaustively meet all the requirements of crime investigation targeting data integration, information sharing, collection and preservation of digital evidences. To bridge this gap, in our preliminary work we presented a generic digital evidence framework, called CISMO as a semantic tool that is able to support LEAs in detecting and preventing different type of crimes happening on OSNs. This paper gives details of the knowledge extraction layer of the framework. Specially, we mainly focus on analyses criminal social graph structures proving the effectiveness of CISMO in a case study with real criminal dataset. Experimental results reveal that applying appropriate Social Network Analyses (SNA), CISMO framework should be able to query and discover the criminal networks, empowering the criminal investigator to see the connections between people.

**Keywords:** Criminal networks, digital forensics, knowledge graph, online social networks, social network analyses, community detection.

## 1 Introduction

In recent years, we have seen a sharply increase on the usage of online social networks (OSNs) by billions of people around the world and these platforms are becoming an indispensable part of their life. People use this platform to easily express and share their day-to-day activities and sentiments. The number of worldwide users reported for January 2020 is 3.8 billion users, with this number increasing by more than 9% since this time last year [1]. It has been alleged that they have the power to energize collective action in social movements like Arab Spring [2].

In UK, police officers reported 32,451 Facebook-related crimes happening during 2017-2018, showing an increase in crime of 19%[1], since the time last year. Of major concern to LEAs is the fact that social media has become a useful tool for terrorism organisation used to recruit and radicalize new members [3], [9]. On the other hand, it is noted that 59% of teenagers have been target of cyberbullying or harassing on OSNs, so this type of crime becomes a major problem for police investigator to identify and manage such cases as often it goes unreported, and thus unpunished[2]. As a result, the exploitation of technology, with the internet and social media at its core, is one of, if not the, most important challenge faced by Law Enforcement Agencies (LEAs) within the EU, and worldwide, today[3]. The paring of virtual marketplaces on the dark web with cryptocurrencies such as bitcoin are increasingly being used as a means to avert authorities ‘efforts to surveillance and trace the exchange of illegal goods and services [4].

A common problem for LEAs during investigation is to analyse people involved in organized crime and to identify groups and key actors [5], using clusters of correlated entities based on information about the connections between the given entities [6]. In this research, the patterns of interactions of the hacker forum can be represented as a network, the individual parts of the forum being denoted by nodes and their private interaction by edges. SNA is employed to detect influencers and communities, such as finding these leaders in such networks and removing them may defragment the criminal network or disrupt it.

The contribution of this paper is twofold. First, this paper introduces the knowledge extraction layer of CISMO framework [25], which is a knowledge graph- based framework developed at our lab originally for the purpose of providing LEAs with the possibility to process unstructured data and identify hidden patterns and relationships in crime datasets with the focus on crime

[1] https://www.infosecurity-magazine.com/news/facebook-crime-rises-19-per-cent/

[2] http://www.bullyingstatistics.org/content/cyber-bullying-statistics.html

[3] https://www.europol.europa.eu/sites/default/files/documents/iocta2017.pdf

investigation and prevention. Second, the research is focused on scalability and usability challenges posed by large criminal graphs to discover communities. In the experimental part we apply some traditional community detection algorithms over information from the Nulled.io5[4] forum, a recently leaked dataset collected for distributing cracked software forum, showing an effective way of processing the information aiming to detect groups with similar characteristics.

The remaining part of the paper is organized as follows. In the background section some preliminaries and notation are summarized. Section 3 describes the architecture, and steps applied in the knowledge extraction layer of the CISMO framework developed by authors for crime detection on OSNs. Section 4 represents the results obtained from applying the proposed algorithms to a real crime data set. Finally, conclusions are presented in section 5.

## 2 Background

Social networks can be modelled as a graph G = (V, E). In OSNs, the nodes represent actors and the edges represent the relationships among actors. Each network represented as graph is characterised by a list of properties which provide information about the structure of the network as a whole. These properties do not provide any information related to the specific actors in the network. Here are definitions of some of popular properties which are used in this research.

- **Size:** the number of nodes within the graph. This property is important as it provides information to classify a graph as a big graph or not. When the size is big the analysing and processing of it it's a challenge.
- **Diameter:** the length of the longest shortest path among all vertices in a given graph. Diameter affects the speed of the diffusion of information within the network.
- **Average Clustering Coefficient:** the mean of local clustering of each node in a given graph calculated as a fraction of triangles that actually exist over all possible triangles in its neighbourhood.
- **Average Path Length:** the average number of steps along the shortest paths for all possible pairs of network nodes, used to measure the efficiency of information or mass transport on a network.

In order to analyse the importance of different actors in social graphs, centralization degrees are calculated. Here, in this paper we focus our analyses different

[4] https://archive.org/details/nulled.io_database_dump_06052016

centrality measures, namely, degree, weighted degree, closeness centrality, harmonic closeness centrality, betweenness centrality [15] and eigen centrality [14], given in Table 1.

**Table 1.** Graph based centrality measures

| Centralities | Definition | Formula |
|---|---|---|
| Degree | number of direct ties that involve a given node | $C_d(i) = \sum_{i=1}^{N} A_{ij}$ (1)<br>N-number of nodes<br>A- the adjacency matrix Aij = 1 if there is a link between the nodes i and j and Aij = 0 if there is not a link between these nodes |
| Closeness | estimates how fast the flow of information would be through a given node to other nodes | $C_c(i) = \sum_{j=1}^{N} \frac{1}{d(i,j)}$ (2)<br>N-number of nodes<br>d (i, j)- the distance between node i and other nodes |
| Betweenness | captures how much a given node is in-between others | $C_c(i) = \sum_{j \neq k} \frac{g_{jk}(i)}{g_{jk}}$ (3)<br>$g_{jk}(i)$ - the number of shortest paths between j and k passing through i<br>$g_{jk}$ -the total number of shortest paths between j and k where $\neq k$. |
| Eigenvector | measures a node's importance while giving consideration to the importance of its neighbors | $C_e(i) = \frac{1}{\omega} \sum_{j=1}^{N} A_{ij}\, C_e(j)$(4)<br>N-number of nodes<br>A- the adjacency matrix |

During the last decade, there has been a considerable interest in community detection in social graphs. There are different definitions of community concept in graphs. The common definition is that a community is a group of nodes densely interconnected compared to the other nodes for a given network.

For a given social network, represented by a graph G = (V, E) where V is the set of nodes and E the set of edges, the community detection is a partition of the nodes in V of the form C = $C_1$, . . ., $C_k$ such that each $C_i$, 1 ≤ i ≤ k exhibits the community structure that presents groups of nodes so called communities [19]. There are two types of community detection, overlapping and non-overlapping (disjoint) communities. In this paper, we focus on applying some well-known non-overlapping community detection, used to find a community structure that any actor in a social network can be member of only one community. Here, we will introduce a set of algorithms we have applied in the forum graph we have created. We ignored some of algorithms that are very slow as the graph we are conducting our experiments is big.

In this research, we have used R software and the igraph library to compare community detection algorithms. This library provides mostly used community detection algorithms ie. Infomap, Louvain, Fast greedy and Walktrap.

*Walktrap*

In [16] author proposed the random-walk concept to find community in a network. This method is based on node similarity and it uses the hierarchical agglomerative clustering, where random walks tend to be confined to denser region of a graph (ie. communities). This algorithm starts from a non-clustered area and calculates distance between adjacent nodes, where two adjacent communities are chosen and merged into one updating the distance between communities. This process is repeated (N-1) times.

*Infomap*

Infomap, introduced by Martin Rosvall et al. [17], it is based on the map equation to find community structure in network, which represents description length of a random walker in a network. It is based on the rule that the partitions with good modular structure have smaller description length. The algorithm first starts with by considering each node as a separate module and then, nodes are selected randomly and are combined resulting in largest decrease in map equation. Then, modules formed in previous steps are considered as nodes and the same process is repeated until there is no further decrease in map equation.

*Louvain*

This algorithm, originally introduced by Blondel et al. in 2008 [20], it is considered as one of the most powerful community detection algorithms, due to the high modularity community partitions in a fast and memory-efficient manner. This algorithm has multiple phases and each phase is characterised by multiple iterations, that are running until the stopping criteria is met. This process stops when there is no change in modularity value. At the beginning of the process, each node i is going to be assigned to a unique community. In the situation of adjacent nodes, if the merging results ends up in a higher modularity gain, these nodes are merged in the same group. Once these calculations are done, the algorithms consider the communities as nodes while total of weights of inter-communities' edges are taken as weight assigned to edges among new nodes. Generally, based on the results presented in literature the method needs only tens of iterations and fewer phases to terminate on real world data, showing significant improvement in terms of computational speed.

*Fast Greedy*

This algorithm is an agglomerative hierarchical clustering method proposed by Clauset et al. [21]. It is recommended to use this algorithm for community detection in networks which have sparse adjacency matrix. This method maximizes the modularity function Q and starts with assigning a different community to each node in a given graph.  Then, the pair of clusters that reach the maximum increase or minimum decrease of ΔQ are combined which results in higher modularity gain, until one cluster remains with all nodes in the network. As an output of this algorithm, a dendrogram, showing the order of merges is produced. The optimal community cluster can be found by cutting the dendrogram at the level of maximum Q.

*Girvan-Newman*

This algorithm [22] detects communities by progressively removing edges from the original network. It is a hierarchical method, based on the edge betweenness. The edges groups that are loosely connected by a few edges are removed. In this way, the groups are separated from each other and reveal the structure of communities, until the connected components of the remaining network are the communities. Instead of basing on the edges are the most central to communities, the Girvan–Newman algorithm focuses on edges that are most likely "between" communities.

*Leading Eigenvector*

This algorithm tries to find densely connected subgraph by moving the maximization process to the eigenspectrum to maximize modularity by using a matrix known as the modularity matrix [24]. The elements of the leading eigenvector measure how firmly each vertex belongs to its assigned community. Thus, large vector elements represent central members of their communities and small vector elements shows more ambivalent results.

In this section, we introduced some of the classic community detection algorithms that are originally designed to be generally applied to any information network. All these algorithms are recursive of high polynomial computational complexity [23]. Thus, their application in big social media networks is limited due in terms of scalability, outcome consistency, and overall reliability. Thus, their application could doubtlessly be considered infeasible

*Evaluation Metrics*

In this paper, we have used modularity [26] and number of communities as the evaluation factors for community detection algorithms. Modularity (Q) is the most widely used and accepted metric, which is used for measuring the quality of community's detection. Let assume that the graph has been partitioned into k communities. Define a $k \times k$ symmetric matrix e whose element $e_{ij}$is the fraction of all edges in the network connecting nodes in community i to those in community j. Let $a_i$=$\sum_j e_{ij}$ be the fraction of edges that connect to nodes in community i. Then modularity is defined as:

$$Q = \sum_i (e_{ij} - a_i^2) \; (4)$$

For practical purspose, a value ranging from about 0.3 to 0.7 ususally appearce to indicate a stong community structure.

## 3 Knowledge Extraction Layer

To understand the criminal behaviour of various actors, the groups they belong to, and to analyse the information shared by them on social media, the knowledge extraction layer of CISMO framework, uses the combination of machine learning, SNA and community detection on OSN to unveil the communication patterns of online users. The steps of the knowledge extraction are outlined in Figure 1. After pre-processing the messages sent in a specific OSNs, each message is converted into feature vectors that are learnable for the machine learning models.

In the previous research, we trained multiple classifiers with the labelled data, including Bayesian network, support vector machine, neural networks and k-nearest neighbours. As the data are unlabelled, we manually labelled 5% of the data in order to build up predictive models for labelling the whole dataset. Linear SVM achieved the highest mean accuracy. Thus, we used linear SVM with the tuned parameter to machine label the rest of the corpus. Thus far, the focus has been on identifying each user's private message (i.e., as a criminal profile, or non-criminal profile), we then constructed a forum network to understand how in-group and cross-group communicate in the structural communities detected in the forum networks.

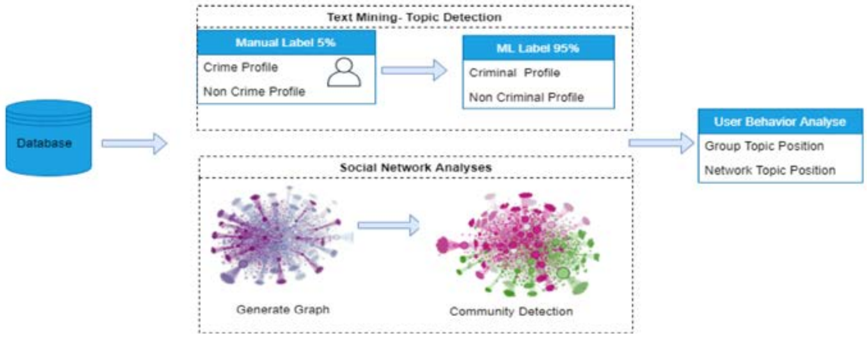

**Fig. 1**. Knwoledge extraction layer of CISMO framework

# 4 RESULTS AND DISUCSSIONS

## 4.1 Data source and data pre-processing

In this research, we use a dataset from Nulled.io, a popular dark web forum which has been hacked and its data leaked. The main reason for using this dataset in our experiments is the real life characteristics and the large number of records in it. However, we do not claim that the data found in this forum represent all diffrent categories of crimes happening in OSNs, but this data is a treasure trove of information for investigators that could yeild powerfull follow-up research in the social media digital forencis, and not only. As this data contains confidential and sensitive information, the research is done after deep consideration about research ethics, and as a consequence in our results we do not provide any data that can directly or indirectly identify the users. Moreover, in legal proceedings we can find out many attemps to analyse the growth and membership of the involved communties in these networks [10], [11]. This database contains a wealth of information, 599,085 user profiles and their private and public communication, but we will limit our research on the private communication among users, where the relevant information is stored in the table message_topics, as shown in Table 2.

During the data preprocessing, the messages are processed in order to remove HTML tags. For this task an HTML parser, Beautiful Soup is used first and then to convert nouns and verbs to their lemma we applied lemmatizers in NLTK. The text messages contains special characters, punctation marks and stop word which are removed using NLTK.

**Table 2.** Database information

| Database | Table | Number of instances |
|---|---|---|
| **Nulled.io** | members<br>message_topics<br>message_posts | 599,085<br>404,355<br>800,593 |

## 4.2 Data Graph description

A communication network can be modeled as a connected undirected graph, where the nodes respresent users and the edges reprsent the communication line between them. In the forum, a user communicates with another user by sending a private message. The graph we created is a weighted graph,considering the frequency of messages exchanged be- tween users using weights. Mt_starter_id field is used as *source vertex, mt_to member_id as target vertex, and mt_to count + mt_replies as edge weight*, as ilustrated in table 3.

**Table 3.** Weighting unit determination

| Table | Interaction Kind | Source Vertex | Target Vertex | Weighting |
|---|---|---|---|---|
| **message topics** | User A send a private message to user B | starter_id (User A) | member_id (User B) | count replies |

In a connected graph, the normalized closeness centrality (or closeness) of a node is the average length of the shortest path between the node and all other nodes in the graph. This adjustment allows comparisons between nodes of graphs of different sizes. In table 4 shows only the centrality indices of the moxt 10 influential nodes of the network, due to page limitiations.

**Table 4.** The details of the most influential nodes selected by different methods

| Name | Degree | Weighted Degree | Closnesscentrality | Harmonicclosnesscentrality | Betweenesscentrality | Eigencentrality |
|---|---|---|---|---|---|---|
| 1 | 2814 | 284984 | 0,934919 | 0,971208 | 52,237939 | 1 |
| 1471 | 1587 | 2263 | 0,494116 | 0,498154 | 0,055225 | 0,006038 |
| 1337 | 1504 | 2121 | 0,49395 | 0,497933 | 0,10866 | 0,005725 |
| 334 | 1321 | 2111 | 0,494808 | 0,498262 | 0,127523 | 0,005077 |
| 8 | 1260 | 1612 | 0,493564 | 0,497414 | 0,055841 | 0,004869 |
| 0 | 1259 | 1662 | 0,492575 | 0,496835 | 0,216206 | 0,004817 |
| 1539 | 1229 | 1819 | 0,49326 | 0,497189 | 0,134325 | 0,00473 |
| 6 | 1049 | 1289 | 0,33721 | 0,34098 | 0,102374 | 0,003722 |
| 4481 | 840 | 1237 | 0,493606 | 0,496945 | 0,059106 | 0,003372 |

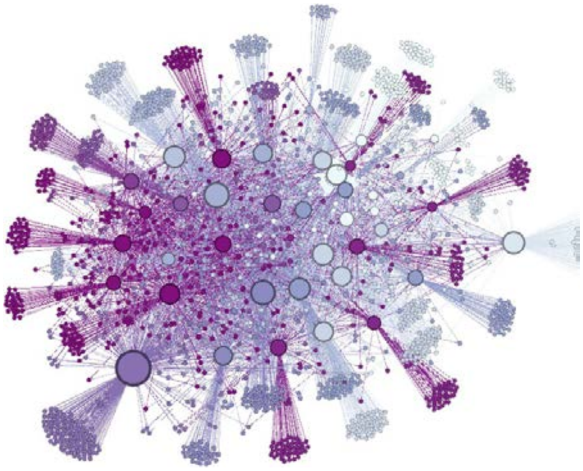

**Fig. 2**. Network for users in Nulled.io forum with private communication. Deep colour and big size of nodes represent users that are having many connections.

Based on the graph presentation, it is evident that one of the nodes has more connections compared to all the others, the node with bigger size belonging to the user 1. After manual checking of the private messages send and received by this node, it is evident that most of the messages are welcome messages and for this reason it can be concluded that this user is the administrator of the network. In order to define relevant criminal community, it has been deleted all the connections where the sender or receiver is user 1 and the connection weigh is equal to one. When the weight is one, it has been shared only a welcome message between the user 1 and any other user in the forum. After deleting all these welcome messages, and some other irrelevant messages, it was obtained a graph with the properties presented in Table 5.

**Table 5.** Graph properties

| **Property** | **Value** |
|---|---|
| Nodes | 25983 |
| Edges | 80671 |
| Diameter | 14 |
| Average clustering coefficient | 0.144 |
| Average Path Length | 4.5 |

Looking at Table 5, as the network of interest of Nulled.io is large, the graph has a small average path length and low clustering coefficient. Investigation done in social networks concluded a short path length between individuals, the so-called "six degrees of separation" [12], which is seen in Nulled.io. This graph has an average clustering coefficient of 0.1444, in the same range with other studies carried out for OSNs data such as Facebook. The range of this property for Facebook data varies between 0.133 and 0.211 with an average of 0.167 [13].

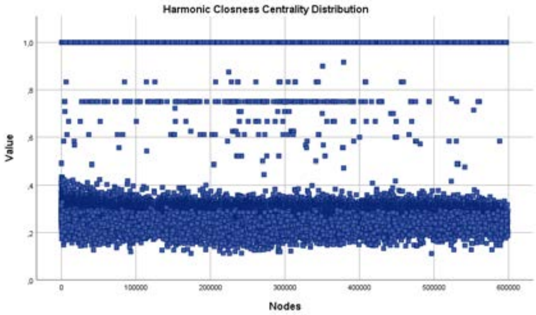

**Fig. 3** Harmonic Closeness Centrality Histogram

In order to analyse and gain a better insight of importance of individuals and their influence in the forum, we analysed the distribution of graph centralities, Between Centrality and Harmonized Closeness Centrality respectively.  As illustrated in figure 3, in total there are 1335 nodes with a centrality over 0.8. These nodes are considered as central nodes as they have the shortest path length to other nodes. These nodes give an idea about the number of communities that can be discovered in this graph. The distribution of the Betweenness Centrality, illustrated in figure 4, shows that more than 13000 nodes has a value close to zero. These nodes belong to one community as they are far away from other nodes in the graph.  From the graph it is evident that there are some nodes with centrality value over 500000, which means that those nodes play a central role in the spreading process in their local neighbourhood.

In this graph, we applied some community detection algorithms in order to define communities and to discover possible criminal communities.

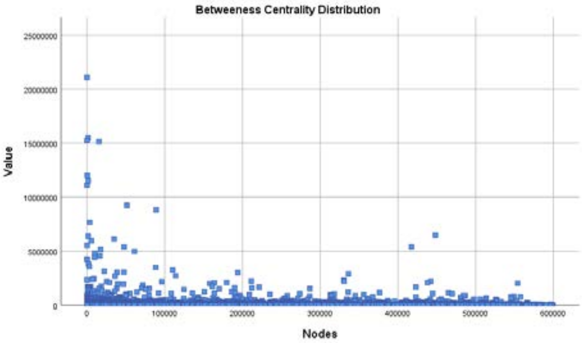

**Fig. 4**. Betweenness Centrality Histogram

### 4.3 Community Detection

In this section, we are evaluating some of the existing algorithms used for community detection in order to compare them. The results are compared based on the two metrics, the modularity $Q$ and number of communities discovered, presented in table 6.

**Table 6.** Modularity of the network when partitioned by each algorithm.

| **Algorithm** | **Modularity (Q)** | **No of communities** |
|---|---|---|
| Louvain algorithm | 0.58 | 861 |
| Girvan-Newman algorithm | 0.47 | 986 |
| Fast Greedy | 0.48 | 1137 |
| Leading_eigen | 0.35 | 730 |
| Imfomap | 0.37 | 2741 |
| Walktrap | 0.39 | 3079 |

Modularity reported in Table 6 varies from 0.35 (Leading Eigenvector) to 0.58 (Louvain). Regarding to the identified communities, the Walktrap algorithm obtained the highest number of communities. However, it also got low modularity; this is due to the principle of random walks that tend to fall into isolated groups of nodes. Based on the results shown in Table 6, we can conclude that partitions obtained by Louvain have consistently high modularity scores, indicating that the network partitions are more community-like.  Fast Greedy, Infomap and Walktrap algorithms also have high modularity scores. These algorithms also differ in terms of the number of communities being detected. Infomap, Walktrap and Fast Greedy detect a large number of communities, result that is not surprising due to the propagation methods behind these algorithms.

Based on the achieved results, we can conclude that the Louvain algorithms for this graph model generates 861 communities with $Q$ = 0.49, the highest modularity. On real world networks, Louvain algorithm achieves the detection of communities which are densely connected inside communities and sparsely connected between communities, detecting a lower number of communtities compared to other algorithms.. Louvain algorithm remains both effective and efficient also when the probability of edges between communities increases (results on artificial networks). On the other hand, Infomap, Leading Eingenvector and Walktrap are weak on modularity metrics.

By using graph analysis techniques. LEAs can identify key members of diffrent criminal communities that might be targeted to disrupt these communities. It was observed that by extracting relevant knowldge, a broad overview of some criminal activities can be obtained; however, due to the heterogeneity of private messages, it is difficult to obtain further details on different crime categories.

## 5 Conclusions

In this paper we presented some challenges faced by LEAS during their daily activities to fight crime happening on social media. We elaborated the knowledge extraction layer of the CISMO framework, a framework developed to semantically detect and prevent crime happening on OSNs. We focus on methodical and analytical aspects of graph analyses of criminal data in big data environments on large datasets with thousands of nodes and edges. Experimental results reveal that applying appropriate Social Network Analyses (SNA), CISMO framework should be able to query and discover the criminal networks, empowering the criminal being capable to identify key members of criminal communities and the communities they belong to. Based on the modularity used as a metric to quantitatively compare the selected community detection algorithm, we conclude that Louvain algorithm appears to be robust in terms of higher modularity and lower number of discovered communities. Our study shows that modeling the data coming from OSNs into a knowledge graph and applying SNA and community detection algorithms, LEAs can gain valuable insights into how criminal communities are organized. Future work will consist in testing the framework with real data of OSNs covering a broader range of crimes, considering both more algorithms and more networks for testing.

## References

1. Digital 2020. Retrieved from https://wearesocial.com/digital-2020.
2. G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce, D. Boyd, "The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions", International Journal of Communication, vol. 5, 2011.
3. Eaton, R. (2014). Digital Terrorism and Hate. Simon Wiesenthal Centre. Retrieved 18 March 2020, from http://www.wiesenthal.com/site/apps/nlnet/content.aspx?c=lsKWLbPJLnF&b=8776547&ct=13928897.
4. Janze, C. (2017). Are cryptocurrencies criminals' best friends? Examining the coevolution of Bitcoin and darknet markets. In Proceedings of the Americas Conference on Information Systems (AMCIS) (p. 10). Boston, MA.
5. Décary-Hétu, D., Dupont, B.: The social network of hackers. Global Crime 13(3), 160–175 (2012).
6. S.M. Marcus, M. Moy, T. Coffman, "Social Network Analysis" in Mining Graph Data, D.J. Cook, L.B. Holder, L.B. (Eds.), John Wiley &Sons, Inc. 2007.
7. S. Purohit, S. Choudhury and L. B. Holder, "Applicationspecific graph sampling for frequent subgraph mining and community detection," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1000-1005.
8. J. Leskovec, C. Faloutsos, "Sampling from large graphs", Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 631-636, 2006.
9. Gabriel Weimann (2016) Going Dark: Terrorism on the Dark Web, Studies in Conflict & Terrorism, 39:3, 195-206.
10. Bradbury, D. Unveiling the dark web. Netw. Secur. 2014, 2014, 14–17 [11] Edwards, M.J.; Rashid, A.; Rayson, P. A Service-Indepenent Model for
11. Linking Online User Profile Information. In Proceedings of the 2014 IEEE Joint Intelligence and Security Informatics Conference, The Hague, The Netherlands, 24–26 September 2014; pp. 280–283.
12. Travers J, Milgram S (1969) An experimental study of the small world problem. Sociometry 32: 425-443.
13. Wilson, C., Sala, A., Puttaswamy, K. P. N., & Zhao, B. Y. (2012). Beyond Social Graphs. ACM Transactions on the Web, 6(4), 1–31.
14. Phillip Bonacich. Technique for analyzing overlapping memberships. Sociological methodology, 4:176–185, 1972.
15. Linton C Freeman, Douglas Roeder, and Robert R Mulholland. Centrality in social networks: Ii. experimental results. Social networks, 2(2):119–141, 1979.

16. Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. In International Symposium on Computer and Information Sciences, pages 284–293. Springer, 2005.
17. M. Rosvall and C. T. Bergstrom. ” Maps of information flow reveal
18. community structure in complex networks.” PNAS 105, 1118, 2008.Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10): P10008, 2008.
19. S. Fortunato, and M. Barthelemy: Resolution limit in community detection. Proceedings of the National Academy of Sciences 104(5), 36-41(2007).
20. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre, "Fast unfolding of communities in large networks", Journal of Statistical Mechanics: Theory and Experiment, vol. P10008, no. 10, 2008.
21. A. Clauset, M.E.J. Newman and C. Moore.” Finding community structure in very large networks.” Phys. Rev. E 70, 066111, 2004.
22. M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proc. Natl. Acad. Sci., vol. 99, no. 12, pp. 7821-7826, Jun. 2002.
23. Peel, L.; Larremore, D.B.; Clauset, A. The ground truth about metadata and community detection in networks. Sci. Adv. 2017, 3, e1602548.
24. M. E. J. Newman, “Finding Community Structure in Networks Using the Eigenvectors of Matrices,” Physical Review E Phys. Rev. E, 74.3 2006.
25. O. Elezaj, S. Yildirim, J. Ahmed, E. Kalemi, B. Brichfeldt, C. Haubold, “Crime Intelligence from Social Media Using CISMO“,Fifth International Congress on Information and Communication Technology, London, UK, 20-21 February, 2020.
26. M. E. J. Newman. Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113, 2004.