Faz Mohammad[1],

Dr. Rakesh Kumar Yadav[2]

Proposed Multicriteria Model For Community Detection Using Modified Cluster Walktrap On Social Media Network For Cybercrime

JES Journal of Electrical Systems

Abstract: The study begins by examining the unique characteristics of online social media platforms as fertile grounds for cybercriminal activities, including the dissemination of malware, phishing schemes, and illicit transactions. Drawing from the fields of computational intelligence, social network analysis, and cybersecurity, the research develops a framework for detecting and mapping out the intricate web of connections among cybercriminal actors within these platforms. Central to the proposed approach is the utilization of advanced data mining algorithms to extract valuable insights from vast amounts of social media data. By analyzing patterns of communication, user interactions, and behavioral attributes, computational intelligence tools can identify suspicious activities indicative of cybercriminal involvement. In conclusion, this study underscores the potential of computational intelligence tools in augmenting traditional cybersecurity approaches by harnessing the vast troves of data available on online social media platforms.

Keywords: Social Media Data, cybersecurity, cybercriminal activities, Modified Cluster Walktrap, Closed cycle approach

INTRODUCTION

SOCIAL MEDIA & CYBER CRIME

The exponential rise in social media usage has revolutionized communication, but it has also amplified exposure to cybercrimes. Studies by Agara et al. (University of Calabar) highlight that while social media platforms offer students opportunities for networking and knowledge sharing, they also increase vulnerability to cyber threats such as identity theft, online harassment, and phishing scams. Predators exploit these platforms to manipulate users, particularly students who often lack adequate cybersecurity awareness.

Similarly, research by Almadhoor et al. underscores the prevalence of cybercrime on social media. According to their survey, platforms like Facebook, Instagram, and Twitter have become hotbeds for malicious activities, including data breaches and financial fraud. The anonymity provided by these platforms enables offenders to engage in illegal activities without immediate detection. The authors emphasize the need for stringent cyber laws and user education to mitigate these risks.

Both studies advocate for enhanced digital literacy to empower users against cyber threats. Institutions and governments must collaborate to develop robust cybersecurity policies and conduct awareness campaigns. As social media continues to grow, fostering a safe online environment is crucial to balancing the benefits of connectivity with the challenges of cybercrime.

SOCIAL MEDIA GRAPH

Social media graphs leverage graph theory to represent and analyze the intricate relationships between users on platforms like Facebook, Twitter, and LinkedIn. According to Kumar et al., graph embedding and graph neural networks play a pivotal role in influence maximization, identifying key nodes (users) that can spread information or trends effectively across the network. This technique is invaluable for targeted marketing and optimizing information dissemination.

The work of Chakraborty et al. highlights the broad applications of graph theory in social media, including community detection, friend recommendations, and sentiment analysis. Nodes in a graph represent users, while edges symbolize connections or interactions, making graph structures ideal for visualizing social dynamics. For

[1]Ph. D. Scholar, Department of Computer Science and Engineering

Maharishi University of Information Technology, Lucknow

[2]Associate Professor, Department of Computer Science and Engineering

Maharishi University of Information Technology, Lucknow

example, detecting strongly connected components helps identify tight-knit user groups, while shortest-path algorithms optimize communication efficiency.

Majeed and Rauf emphasize that advancements in graph algorithms allow for predictive analysis, such as forecasting user behavior and network evolution. Additionally, Newman et al. explore random graph models, which provide insights into network properties like clustering and degree distribution. These models help understand phenomena such as viral content spread and network resilience.

SINGLE CRITERIA CLUSTERING ALGORITHM

When doing analysis, the Single Criteria Model (SCM) focuses on only one criteria or property at a time. Analysis techniques based on graph theory may be used directly to it. Graphs may be hierarchically divided into communities using a divisive clustering technique that removes edges with the highest edge betweenness value. It is restricted to networks with a size of 103 nodes and was suggested by Girvan and Newman. This technique takes modularity into account as a quality function with a complexity of O(n3). To improve computational complexity and the limit of network size, upgraded the Girvan and Newman method. This method disregards the betweenness value and instead eliminates graph edges with a high clustering coefficient. This method has an O(n2) computational complexity.

METHOD FOR CIMMUNITY DETECTION

For community detection, merging the vertices arises instead of separating the network depending on specified characteristics. Starting with n communities, Newman's greedy method optimizes a quality function known as modularity as a partition quality, and then merges the vertices (Newman, 2004). It takes O(mn) calculations to complete this method. Donetti and Munoz measured vertices' similarity using the Eigen vectors of the Laplacian matrix of the network. Computation of Eigen values for sparse matrices is shown to be O(n3) time difficult. The walk trap method determines a distance that assesses the structural similarity of vertices and communities by using random walks. The dense community is trapped at a given point on the graph by the random walk. Because the nodes in the graphs are interconnected, the walks that are generated will recur inside the set of them. Nodes form communities or clusters based on the relationships between them that cause the walk to recur inside that set. Only a subset of nodes will remain outside the community after n steps of the random walk. Starting from the bottom up, new communities will be created by combining the outcomes of the random walk steps. As a rule, random walks use a divide-and-conquer strategy, which simplifies community discovery in terms of time.

LITERATURE REVIEW

Jewkes (2016) examines cybercrime classifications in depth. He delves into topics such as online victimization, the social construction and policy implications of Internet crime, the duality of cyberspace, the virtual universe's impenetrable anonymity, the difficulties of regulation and control, and offers solutions to these problems. Given the ever-changing landscape of cybercrime, the provided cases and data are rather out of date. When it comes to VR and social media, it likewise falls short. Because of his Western bias, Jewkes has ignored the growing number of Asian internet users in tandem with their distinctive online environments, a factor that has given birth to novel forms of cybercrime.

Geetha, S. et.al. (2020). Despite several verification measures, information housed in public forms has been abused due to recent changes in the digital world. To better understand how hackers disseminate valid data, this research makes advantage of Facebook, a widely used social media platform. Facebook provides digital evidence of illicit conduct, which may help identify and prosecute cybercriminals. Criminals attempt to sexually attack women on this site by posting hate speech and spreading falsehoods. We train Pyspark as a forensic tool to analyze a Facebook data set, differentiating between legitimate and wrong posts, in order to provide digital evidence for criminal investigations.

Rawat, Romil et.al (2021). Concurrently with the expansion of the online infrastructure, In recent years, sentiment analysis has emerged as an indispensable tool in many domains, such as cyber-vulnerability assessments, online diaries that track harmful movements, more targeted websites, and informal networks pertaining to the investigation of cyber-criminals' actions. In order to appeal to other people's sentiments and worldviews, the choices and tactics utilized to spark interest in the present are often modified. Thus, in order to make a final decision, carry out research, and evaluate the work of others, the traditional looking-through method is used. All the way from the individual to the larger society, this can be shown. Given the comprehensive nature of this bias evaluation, we are requesting that natural language processing (NLP) determine whether a given piece of content contains theoretical information, the kind of enthusiastic information it transmits via cyber-malicious post overview, and, if so, whether the source of the content is positive or negative (). Modern technology has opened up new avenues for fraud, such as the ability to check users' blockages, host cyber-occasions via online social groups and security offices, and much more. Understanding the emotions underlying user-generated content

and instructional materials is, however, very helpful for business and personal usage, among other things. Multiple layers of information processing may be used to organize the work, which means that the most comprehensive vocabulary, phrases, or whole educational sets may be necessary. Using an AI technique, this methodology examines a typical system for cyber-weakness overviews.

Emmanuel, Etuhe, et al., (2022) The world as we know it today is really a computer simulation. People are understandably worried about the security of their personal information because of how reliant they are on social media. One of the main reasons why social media has become so popular is the ease with which people from all over the world can connect with one another and discuss subjects that they both find interesting. The impetus for these extraordinary developments is the fear of cybercrime, which is analogous to physical crime. Many malicious actions are taking place on social media platforms (SMP) due to groups with criminal purpose and hackers. With the development of increasingly advanced detection technology, hackers are constantly adapting their plans and approaches. It is a challenging effort for organizations and security managers to develop and execute the necessary policies and procedures to prevent the attacks. Hackers are always coming up with new techniques and tools to hack in order to break security and take over social media networks. The increasing frequency of cyberattacks on social media platforms necessitates the implementation of more stringent and urgent security protocols. In an effort to build a safe social cyberspace and improve virtual socializing, this study investigates the method and strategy of hacker assaults on social media platforms and offers solutions to this problem. The last step in preventing cybercrime on social media networks is to implement an intrusion detection mechanism. This will be done after the platforms have been briefly categorized, the different types of attacks have been highlighted, and the research has suggested current state-of-the-art preventive mechanisms to overcome these attacks.

Tayebi, Mohammad et.al (2020) Connecting local and global patterns to better understand the linked nature of criminal behavior, this book explains how open-source information may be used as a formidable weapon against crime. We cover all the latest developments in data mining, machine learning, natural language processing, predictive analytics, and social network analysis that might aid in the detection, analysis, and mitigation of cyber and physical threats. Chapters dedicated to open-source intelligence and social media analytics detail the most recent results in both fields. Scholars, students, and professionals in the domains of open-source intelligence, cybercrime, and social network analytics will find this comprehensive collection very valuable.

MULTICRITERIA MODEL (MCM) FOR COMMUNITY DETECTION

A bipartite graph BG= {P, C, R} is used to represent the crime data in the suggested technique. Here, P stands for the set of people, C for the collection of crimes, and R for the relationship between the two sets. The planned work's schematic is shown in Figure 1 In order to conduct analysis, the analyst collects data from the scene of the crime and prepares it into the necessary format.

A criminal event matrix, with rows for people and columns for crimes created by the analyst, is used as input to the suggested task. Each matrix value stands for a relationship between a criminal. The criminal event matrix must be transformed into a bipartite graph as the first stage.

Finding the communities that exist in the graph is the second stage. The suggested cluster walk trap method for community discovery takes the built bipartite graph as input. The method produces a clustering set as its output. Groups of persons linked to a certain crime are shown by the clusters, and vice versa.

Figure 1 shows the process flow of proposed MCM model.

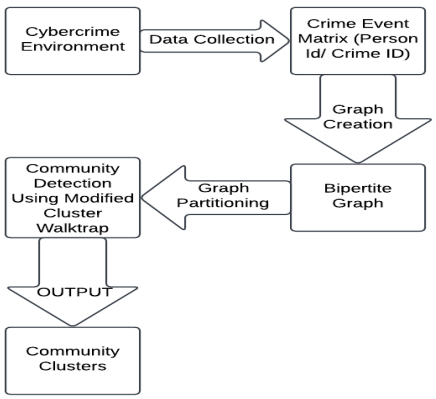

Figure 1 Process Flow Diagram of MCM for Community Detection

In order to determine the distance between communities and vertices using a similarity metric, the MCM method employs random walks. The method iteratively merges communities based on distance measure until it reaches a point where it traps the denser portion of the network. Once the maximum modularity value is attained, the method terminates. There are five categories with a modularity score of 0.668 when the algorithm is run.

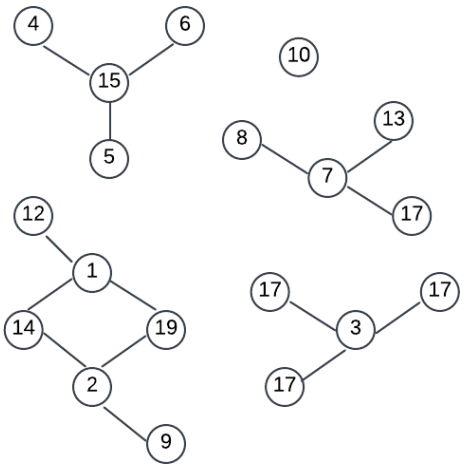

Figure 2 Visualization of Groups

After the clustering was done, the groups that formed are shown in Figure 2. Individual 1 has ties to 5, 7, and 12 crimes; Person 2 has ties to 2, 7, and 12 crimes; Person 3 has ties to 4, 9, and 11 crimes; Persons 4,5, and 6 are associated with 8, and Person 7 has ties to 1, 6, and 10. Because no individual is associated with Crime 3, it will not be further investigated. Thus, as seen in Figure 2, the two-mode data is projected to uncover the hidden linkages via the use of incidence matrix creation. Despite the fact that person 2 is not directly associated with crime 12, a hidden link between 2 and 12 is revealed during network analysis, which reveals a relationship between 2 and 1 with common crimes 9 and 14, where person 1 is tied to crime 12. Incidence matrix production is a part of data pre-processing that prepares the data for using the suggested method of community discovery, which entails identifying such hidden relationships.

DATA SET FOR COMMUNITY DETECTION

The Crime Dataset, which stands for a Crime Event Matrix, is used to execute the suggested task. The set of persons is shown in the row, while the set of criminal occurrences is represented in the column. The relational tie values in the matrix relate to the specific person's involvement in the criminal occurrence for which they are represented. In this context, one of the most important criteria for community identification is the level of public engagement.

The KONECT Crime Data collection is being examined for the proposed work's implementation using the Single Criterion Model (Freeman, 2004). The dataset's foundation was laid by the St. Louis police records pertaining to criminal incidents. Norm White and Rick Rosenfeld analyzed these police reports and entered their findings into a person’s crime event matrix.

EXPERIMENTAL RESULTS

Here we discuss the experiments that were conducted utilizing the publicly accessible KONECT Crime Dataset (KONECT, 2017). This method is tested in R Studio 3.2.5 (64-bit) on a PC with an Intel Core i5 7th Gen CPU, 8 GB of RAM, and a 2.70 GHz speed. We do experiments to show that the suggested method works in criminal network community detection.

Dataset

Training Data

Using a two-mode valued crime matrix of persons participating in the criminal event, the suggested MCM methodology was trained. Being a two-mode matrix, the dataset itself cannot be separated into training and testing data. There is a critical function for each row-wise person value and each column-wise crime value in the community detection process.

For example, separating the information into training and testing sets row-wise ignores the fact that individuals are often interconnected across crimes, which might lead to discrepancies in identifying important figures. On the

other side, dividing the dataset column-wise ignores the interconnected nature of the incidents, which masks any variation in community crime detection. To get around this, we used the KONECT crime data as a reference and produced the sample training data using the data wrangling concept in the R program. Data manipulation software called grammar is used to create the example data.

Testing Data

An 870 x 557 two-mode valued matrix of persons participating in criminal occurrences constitutes the data set. Victims (item 1), suspects (item 2), witnesses (item 3), and duals (victims + suspects) (item 4) make up the criminal event matrix. The individuals listed in Table 1 are members of a bipartite network that this data creates.

Table 1 Properties of the KONECT crime dataset

Properties	Details
Number of Actors	1380 (Persons + Crimes)
Number of Links	1476 (Involvements)
Average Degree (overall)	2.1391 edges/vertex
Average Person Degree	1.7805 edges/vertex
Average Crime Degree	2.6788 edges/vertex
Diameter	32 edges
Mean Shortest Path Length	13.37 edges

Performance Analysis

Results from other graph clustering approaches, such as Cluster Edge Betweenness (CEB) and Cluster Walktrap (CWT), are also used to prove that the suggested Multi Criteria Model (MCM) method is successful. Using the R tool, we analyzed the performance of the suggested strategy. Table 2 shows the results of comparing the suggested MCM method to other graph clustering methods in terms of efficacy, number of clusters, iterations, and CPU execution time.

Table 2 Comparison of Community Detection using KONECT CRIME DATASET with other approaches

Algorithm	Number of Clusters		Iterations	Effectiveness	CPU Execution Time
	NRS	RS			
MCM Model	146	13	14	10.43	3.81 secs
Cluster_Edge_Betweenness	126	20	13	9.69	3.25 secs
Cluster_Walktrap	130	10	18	7.22	4.86 secs

A two-mode criminal network has to be projected into two one-mode projections in order to locate clusters using techniques other than modified cluster walk trap. Cluster formation using both repetitive (RS) and non-repetitive (NR) sets of nodes in the network, as well as the total number of iterations used to monitor the network, are used to quantify performance. The efficiency of the methods is evaluated by dividing the number of iterations by the NR. Despite edge betweenness's low iteration count in comparison to modified walk trap, it generates a large number of non-repetitive groupings of nodes.

In comparison to other ways, the suggested strategy yields more effective outcomes. While other graph clustering methods yield about the same amount of clusters and iterations, MCM generates much fewer overlapping communities. While existing graph clustering methods fail to take some network nodes into account while forming communities, the suggested MCM method does just the opposite. Also, compared to the CEB method, the suggested MCM method uses a lot more computer time. Proposed MCM takes more time than CEB approach for community clustering as it computes the node-wise similarity score for all nodes to be considered, while CEB just calculates the relational similarity measure for links in the community. When compared to previous methods, the suggested MCM methodology with node-wise similarity computation is more effective and accurate since it takes into account all of the information about the investigated network's structural features.

SCM: PROMINENT PERSON IDENTIFICATION USINGCLOSED CYCLE APPROACH

Simply said, identifying important actors in a criminal network is what the "Prominent Person" refers to. It is quite difficult to find important nodes in the network that are community-structured. Recently, there has been a lot of interest in research that aims to identify critical nodes in a network that have a community structure. In a unipartite network, the most important nodes are those that are either highly coupled with other nodes and their removal impacts the entire network or those that are straightforward and simple, identified using centrality measures (Brandes, 2001).

The two-mode structure of a bipartite network makes it challenging to identify notable nodes. Recognizing important nodes in both sets of vertex information is no easy feat. The most basic way to examine a bipartite network and find its notable nodes is to split it into two single-mode networks by removing one vertex at a time from the original two-mode network. Afterwards, the nodes that stand out are identified by computing the centrality metric.

The data collected by the Crime Network is classified, therefore any study of it must take this into account. It is not worthwhile to focus on only one network mode in order to determine which nodes are most important. Here, in order to identify the significant nodes in the network under study, we take into account the network's structure, the communities it forms, and the clustering coefficient measure for each node in the community and the overall network.

CONCLUSION

In the bipartite criminal network, a method called Multi Criteria Model(MCM) is suggested for Community Detection. This method uses a single criterion to identify dense communities, and it uses precise measurements of node structural similarities to group them into clusters. The suggested Closed Cycle Approach (CCA), an extension to the MCM based on transitivity, measures the clustering coefficient of the nodes in communities and the complete network to identify the set of important nodes. It sorts the most important people into three groups: victim, suspect, and witness, depending on the restriction. In order to identify the set of prominent persons in a bipartite crime network, the Multi Criteria Multi Constraint based Perception Grading (MC2PG) method uses evaluation matrices generated using SNA metrics for the network to focus on multiple criteria. Subsequently, the suggested method sorts the nodes or individuals in the network from most prominent to least prominent according to their similarity scores.

REFERENCE

[1] Jewkes , Yvonne (2016), Crime Online, William Publishing,Canada

[2] Geetha, S. & P., Dineshkumar& Velan, Senthil & Syed, D & Varathan, Kanya. (2020). Big Data Analysis - Cybercrime Detection in Social Network. Journal of Advanced Research in Dynamical and Control Systems. 12. 2020.

[3] Rawat, Romil & Mahor, Vinod &Chirgaiya, Sachin & Shaw, Rabindra & Ghosh, Ankush. (2021). Sentiment Analysis at Online Social Network for Cyber-Malicious Post Reviews Using Machine Learning Techniques. 10.1007/978-981-16-0407-2_9.

[4] Emmanuel, Etuh&Bakpo, Francis & H, Eneh. (2022). Social Media Networks Attacks and their Preventive Mechanisms: A Review.

[5] Tayebi, Mohammad &Glässer, Uwe & Skillicorn, David. (2020). Open-Source Intelligence and Cyber Crime Social Media Analytics: Social Media Analytics. 10.1007/978-3-030-41251-7.

[6] Soomro, Tariq & Hussain, Mumtaz. (2019). Social Media-Related Cybercrimes and Techniques for Their Prevention. Applied Computer Systems. 24. 9-17. 10.2478/acss-2019-0002.

[7] Sabillon, Regner &Cavaller, Víctor &Cano M., Jeimy & Serra-Ruiz, Jordi. (2016). Cybercriminals, cyberattacks and cybercrime. 1-9. 10.1109/ICCCF.2016.7740434.

[8] Injadat, Mohammad Noor& Salo, Fadi & Nassif, Ali. (2016). Data Mining Techniques in social media: A Survey. Neurocomputing. 214. 10.1016/j.neucom.2016.06.045.

[9] Rasel, Risul Islam & Sultana, Nasrin & Akhter, Sharna &Meesad, Phayung. (2018). Detection of Cyber-Aggressive Comments on Social Media Networks: A Machine Learning and Text mining approach. 37-41. 10.1145/3278293.3278303.

[10] Khan, Mohiuddin & Pradhan, Sateesh & Fatima, Huda. (2017). Applying Data Mining techniques in Cyber Crimes. 213-216. 10.1109/Anti-Cybercrime.2017.7905293.

[11] Yeboah-Ofori, Abel. (2018). Cyber Intelligence and OSINT: Developing Mitigation Techniques Against Cybercrime Threats on social media. International Journal of Cyber-Security and Digital Forensics. 7. 87-98. 10.17781/P002378.

[12] Z. Xu and S. Zhu, "Filtering offensive language in online communities using grammatical relations," in Proceedings of The Seventh Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS'10), 2020.

[13] C. Zhang, D. Zeng, J. Li, F. Y. Wang, and W. Zuo, "Sentiment analysis of Chinese documents: from sentence to document level," Journal of the American Society for Information Science and Technology, vol. 60, pp. 2474-2487, 2019.

[14] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," In EMNLP'02: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, pp. 79-86, 2012.

[15] A. Mahmud, Ahmed, Kazi Zubair, and Khan, Mumit "Detecting flames and insults in text," in Proc. of 6th International Conference on Natural Language Processing (ICON' 08), 2018.