Dissertation on

"Community Detection in Dynamic Networks"

Submitted in partial fulfilment of the requirements for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering

UE18CS390B – Capstone Project Phase - 2

Submitted by:

Mahammad Thufail	**PES2201800646**
Purushotham S	**PES2201800480**
Manne Vasanth	**PES2201800425**
Pulle Manikya Sri Manasa	**PES2201800468**

Under the guidance of

Prof. Sreenath MV
Assistant Professor
PES University

June - Nov 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

'Community Detection in Dynamic Networks'

is a bonafide work carried out by

Mahammad Thufail	**PES2201800646**
Purushotham S	**PES2201800480**
Manne Vasanth	**PES2201800425**
Pulle Manikya Sri Manasa	**PES2201800468**

In partial fulfilment for the completion of seventh semester Capstone Project Phase - 2 (UE18CS390B) in the Program of Study -Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period June 2021 – Nov. 2021. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it satisfies the 7th semester academic requirements in respect of project work.

Signature	Signature	Signature
Prof. Sreenath MV	Dr. Sandesh B J	Dr. B K Keshavan
Assistant Professor	Chairperson	Dean of Faculty

External Viva

Name of the Examiners	**Signature with Date**
1. _____	_____
2. _____	_____

DECLARATION

We hereby declare that the Capstone Project Phase - 2 entitled **"Community Detection in Dynamic Networks"** has been carried out by us under the guidance of Prof. Sreenath MV, Assistant Professor and submitted in partial fulfilment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester June – Nov. 2021. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

Mahammad Thufail	**PES2201800646**
Purushotham S	**PES2201800480**
Manne Vasanth	**PES2201800425**
Pulle Manikya Sri Manasa	**PES2201800468**

ACKNOWLEDGEMENT

I would like to express my gratitude to Prof. Sreenath MV, Department of Computer Science and Engineering, PES University, for his continuous guidance, assistance, and encouragement throughout the development of this UE18CS390B -Capstone Project Phase – 2.

I am grateful to the Capstone Project Coordinator, Dr. Sarasvathi V, Associate Professor, for organizing, managing, and helping with the entire process.

I take this opportunity to thank Dr. Sandesh B J, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement I have received from my family and friends.

ABSTRACT

The importance of social network research, both as a theoretical viewpoint and as a methodological toolkit, for understanding and evaluating terror groups, as well as developing counterterror policies and practises to identify and disrupt terror attacks, is a key issue in monitoring transnational terror patterns. Terrorist activities have led to the creation of a number of high-end methodologies for studying terrorist organisations and networks around the world. Social Network Analysis (SNA) is one of the most powerful and predictive methods for combating extremism in social networks, according to existing studies. In terms of a global and regional context, the study examined various SNA steps for predicting the key players/main actors of terrorist networks. The applicability and viability of SNA tools for online and offline social networks were demonstrated in a comparative analysis of SNA tools. Data mining techniques can be used to integrate temporal analysis. It has the potential to improve SNA's ability to handle the complex behaviour of online social networks.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

Terrorism is an organised type of violence that has a direct impact on stability, a country's or community's daily routine, and security, as well as a means of instilling fear in civilians. Terrorism is a fluid phenomenon, so equipping counter-terrorism operators with the resources they need to combat it is important.

The main goal of this research is to identify a method for eliciting information about terrorist attack suspects by examining terror attacks over time. The aim is to create a sociogram, or a network of criminals, in which the nodes represent terrorist groups and the edges represent generic interactions between two groups.

We used a method that allows us to detect clusters of terror groups that are similar in terms of organizational characteristics. Specifically, we built a network of terrorist groups and related information on tactics, weapons, targets, and active regions using open access data from terrorist attacks that occurred worldwide since 1970, from START organization.

Each partition is linked to the terrorist groups in our model. Later, we'll try to avoid attacks by identifying the most powerful party with the greatest number of connections to other networks. Community identification is a technique for identifying groups of nodes in which the connections between nodes within a group are greater than the connections between nodes in other networks.

CHAPTER-2

PROBLEM DEFINATION

Using a database of terrorist occurrences, elicit terrorist group networks and find the most influential node and community detection.

We use rough set approach to understand the composition and evolution of terrorist networks. Later on we will find influential nodes based on neighborhood correlation coefficient. Finally for Community Detection we use Louvain algorithm using locality modularity optimization.

CHAPTER-3

LITERATURE REVIEW

3.1 Paper 1: Understanding the composition and evolution of terrorist group networks: A rough set approach.

3.1.1 Introduction

The key idea behind this research is to use GTD's historical data, which includes information on terrorist attacks that have occurred since 1970, to conceptualize terrorist groups' activities over time intervals. Following that, such conceptualizations are used to explain the similarities between terrorist groups and to evoke relationships to reflect terrorists' networks. Through applying the method to various time points along the continuum and examining variations among the resulting networks, the above networks can be used to research the temporal evolutions of terrorist groups' behaviors. The method is focused on Rough Set Theory and Three-way Decisions Theory, and it generates an original similarity function based on boundary region description.

3.1.2 Characteristics and Implementation

The first step is to choose the correct attributes from the Global Terrorism Database's entire collection of attributes. The precise conceptualizations of terrorist groups and their approximations are given in the second step. The specification of the time interval for considering terrorist activities and the selection of terrorist organizations for which to analyze activity are the preliminary actions for the second level. The third step constructs the boundary regions for the above rough conceptualizations using the three-way decisions theory and the rough approximations given by the second step. The similarity matrix, which reports the similarities between all pairs of considered terrorist groups, is built in the fourth step, using the newly described similarity function. Finally, the

fifth step entails converting the similarity matrix into a network graph that depicts the generic relationships between terrorist groups.

3.1.3 Features

The proposed method was demonstrated and evaluated by comparing the results to expert information obtained from the GTD database's resources and using a Python implementation of rough set operators (created by the authors) (START project). Using a similarity function based on rough set theory, the paper proposes an original method for eliciting terrorist group networks from a database of terrorist incidents. Several illustrative examples have been given to explain and demonstrate the method.

3.1.4 Limitations

The methodology may be included in the research systems used by counter-terrorism operators, based on the promising findings. The following are some of the limitations of the findings presented in this study:

i. the lack of a method for dealing with time intervals automatically,,
ii. the lack of a mechanism for analysing the temporal growth of terrorist organisations' networks automatically,
iii. the inability to determine the semantics of the elicited networks' relations

All of the above issues should be explored, and the authors have already planned several improvements to the proposed approach:

- 135 features were extracted from the original dataset via a systematic extraction process (new relevant features could emerge, perhaps improving the expert-based approach used in this study).
- For terrorist groups, the concept of a semantic similarity feature (to improve the technique, it could be useful to integrate semantics into the similarity function in order to consider portions of existing and possibly incomplete information).

- The assessment of many alternatives to standard rough sets for conceptualizing terrorist group activities, such as tolerance rough sets and neighborhood rough sets.
- The development of a method for analyzing the temporal evolution of terrorist group networks.

3.2 Paper 2: Finding influential nodes in social networks based on neighborhood correlation coefficient.

3.2.1 Introduction

Rapid dissemination of news and advertisements via social networks has created new opportunities for social media platforms to replace traditional forms of advertising such as radio, television, and banners. It is nearly impossible to reach out to all users directly, so a clever solution is to target a small number of influential users and communicate with them in the hope that they can maximize the desired influence on others. If the target users are carefully chosen, they can disseminate the messages more widely by spreading them through their friendship networks. Users are not all equal, and some have more clout because of their personal and social characteristics, as well as their placement in strategic locations. Influential users have a greater impact on the success of information dissemination than others. In recent years, there has been a lot of interest in defining users' influence ranges and ranking them based on their influence ranges. In many cases, the structure of the connection network is the only available information for identifying influential nodes. A number of techniques for locating influential users using network structure have been proposed in the literature.

3.2.2 Characteristics and Implementation

The idea behind this method is that users who share more friends with their Neighbours tend to spread information over shorter distances. Only the number of friends shared by each user and their Neighbours is considered in the cluster rank method. In this paper, we look at the properties

of the users' commonalities with their Neighbours. The proposed method estimates a node's influence score based on the similarity (or correlation) of the node's and its Neighbour connection structures. The proposed method is tested on a variety of synthetic and real-world networks, and it is compared to a variety of state-of-the-art influence ranking algorithms.

The proposed algorithm makes a significant contribution by taking into account the detailed correlation structure of the neighborhood and effectively using it to identify the most influential nodes.Pearson correlation is used for this purpose.To estimate the influence of nodes in effective information propagation, a number of centrality measures have been proposed. The most basic metric for determining influential nodes is degree centrality. However, it is common for nodes with lower degrees to have far greater influence than nodes with higher degrees.
Indeed,a low-degree node may be located in a strategic location of the network, facilitating information spread more than higher-degree nodes. The influence range of a node is determined by the degree distribution of its first and second-order Neighbours in the entropy centrality method.

3.2.3 Features

Social network analysis and mining have recently gained prominence, with numerous potential applications in a wide range of industries. One of the topics that has received a lot of attention in this field is influence maximisation.The proposed method is based on the local clustering coefficient and takes advantage of the similarity of connections between neighboring nodes. The method is based on the k-shell decomposition approach, in which a node's influence is determined by how many shared connections it has with its Neighbours. In the proposed method, two nodes with Neighbours in the same parts of the networks are considered correlated nodes. Correlated nodes have minimal spreading capacities because they cover a smaller area of the network as compared to uncorrelated nodes. As a result, a node with a low correlation with its Neighbours could be a better candidate to be a powerful node because of its ability to disperse messages around the network with the aid of its Neighbours.

3.2.4 Limitations

The proposed method's main flaw is that it requires the processing of all data in order to locate the most influential node.

3.3 Paper 3: Community detection in large-scale social networks: state-of-the-art and future directions.

3.3.1 Introduction

The discovery of the structure of a social network is an important research area in social network analysis, and community detection is one of the most important research areas in this field. In disciplines like sociology, biology, and computer science, where processes are often described as graphs, detecting societies is critical. This is an NP-hard problem that has yet to be solved satisfactorily. This is an NP-hard problem that has yet to be solved satisfactorily. Two major factors impede this computational complexity. The first aspect is the massive scale of today's social networks, such as Facebook and Twitter, which have billions of nodes. The second reason has to do with the complex existence of social networks, which change their structure over time. As a result, group detection in social network analysis is gaining a lot of scientific interest, and a lot of research has been done in this field. This paper's main aim is to provide a concise overview of group detection algorithms in social graphs. We include a taxonomy of current models based on their computational existence (either centralized or distributed) and thus in static and dynamic social networks for this purpose. We also have a detailed overview of current community identification applications in social networks. Finally, we discuss future research paths as well as some unresolved issues.

3.3.2 Characteristics and Implementation

This study will predict actor characteristics, the appearance of any connections, the diffusion arrangements in the network, and so on. Despite the fact that the graphs have no apparent structural properties, they all share one that defines them regardless of their specific material. Identifying communities helps one to get a macro view of complex structures, which is a useful method for understanding and analyzing them. Group detection is an NP-complete graph partitioning problem from a theoretical standpoint (Fortunato 2010; Xie and Szymanski 2013; Ben Romdhane et al. 2013; Rhouma and Romdhane 2014). Communities are a key feature of the network that can be divided into two types: structure-based communities (a cluster of nodes with more connections to each other than other nodes in a graph) and semantic-based communities (a cluster of nodes with similar semantic meaning or a group of nodes with similar interests). Since both the meaning and the relationship of the nodes are used to detect semantic communities, the result may more effectively reflect community cohesion. The diverse characteristics of topological structures and semantic context make it difficult to characterize both structures and semantic context at the same time.

3.3.3 Features

This survey's main purpose is to present, organize, and evaluate current models for evaluating the communities of a given large-scale network. The following are the main points of the paper:

- The aim of this project is to categories and compare theoretic group detection algorithms.
- Using two strategies: centralized and dispersed, combine techniques for characterizing, identifying, and extracting populations.
- To examine the detection technique in relation to the dynamic of networks : static or evolving over time.
- To include partition detection metrics such as structural-based partition and semantic-and-structural-based partition.

3.3.4 Limitations

Because of the heterogeneity of data and systems developed in them, as well as their scale, group detection is becoming increasingly difficult. When we factor in the complexities in this massive amount of data, the problem becomes much more complex. These limitations sparked a surge in interest in large-scale network distributed algorithms. Two broad definitions can be drawn from these approaches. The first, and most well-studied, is a distributed detection system based on shared-memory parallelism, also known as "distributed computing." The social network is divided into non-overlapping subnetworks that are handled in parallel using these parallelization techniques. The high cost of shared memory multiprocessor systems and the high degree of data dependence, on the other hand, are seen as roadblocks to this strategy. Attempts have been made to detect many groups in a distributed manner using an undistributed social graph in this direction. We perform a portion of our survey here on the most up-to-date models for detecting parallel group structure. In recent years, the quest for such communities has gotten a lot of attention, and several studies have been done on group structure in graphs. These researchers are attempting to deal with massive amounts of complex data as well as the combined use of relational and semantic data.

3.4 Paper 4: Hidden community detection in social networks.

3.4.1 Introduction

Community identification has become an important role in network analysis in recent decades, providing insight into the underlying structure and possible roles of networks. Early research centered on finding disjoint communities that divide a network's collection of nodes. Researchers have recently noticed a multiplicity of interwoven community memberships and built algorithms to find overlapping groups. To deal with the overlapping case, certain partitioning techniques are also extended. Within these two groups, a hierarchical dendrogram centered on the granularity of

the detected communities can be constructed. Despite the progress made, there is a new form of group structure known as the secret community structure that has received little coverage in the literature. The sparse community structure found in real-world networks, such as hidden societies or temporary communities, is significantly weaker than the dense community structure found in families, colleagues, close friends as determined by widely used group scoring metrics When the majority of the members of a less modular group are also members of denser groups, the community is often ignored.

3.4.2 Characteristics and Implementation

The goal of this paper is to reveal the hidden structure. We characterize a community's hiddenness value as the proportion of nodes in stronger communities, and we present Hidden Community Detection (HICODE), a meta-approach for identifying both the dominant and hidden structure in networks. HICODE starts by applying an existing algorithm to a network as a base algorithm, then weakening the structure of the network's detected communities. In this way, the community's weaker, secret structure is shown. This step is repeated until no further significant structures are discovered. HICODE then weakens the secret group structure, resulting in a more reliable version of the dominant community structure.

3.4.3 Features

We assume the information we gleaned about secret group structure would be useful in future investigations. The following are the paper's key contributions:

- **Conception on Hidden Community:** We define a group's hiddenness and propose the concept of hidden community structure, which is prevalent in social networks.
- **Methods on Hidden Community Detection:**HICODE is a technique for finding both the dominant and secret structure. HICODE is implemented using numerous group detection algorithms as the basis algorithm, as well as the structure weakening methods Remove Edge, Reduce Edge, and Reduce Weight.

- **Validation on real world datasets:** We show that the higher a community's hiddenness value, the more difficult it is for an algorithm to discover that community; in testing on a range of real-world networks, HICODE beats various state-of-the-art community detection algorithms at uncovering hidden communities.

3.4.4 Limitations

In studies on a range of real-world networks, we show that the higher the hiddenness value of a community, the more difficult it is for an algorithm to discover that community; HICODE beats numerous state-of-the-art community detection approaches in uncovering hidden communities.

3.5 Paper 5: TI-SC: top-k influential nodes selection based on community detection and scoring criteria in social networks.

3.5.1 Introduction

The investigation of identifying powerful nodes is one of the most critical topics in social networks. An influential spreader detection problem occurs when a single influential node is detected without taking into account the positions of other influential nodes in the network. The influence maximization problem, on the other hand, is described as identifying the set of influential nodes in terms of their topological effects on each other. The aim of influence maximization is to identify the most powerful people who can spread the most information. Diffusion is the mechanism by which network knowledge spreads from one node to another. Total influence time is critical in the diffusion process.

3.5.2 Characteristics and Implementation

By evaluating the relationships between the core nodes and the scoring capacity of other nodes, the TI-SC algorithm selects the powerful nodes. The scores are modified after each seed node is selected to minimise overlap in seed node selection. This algorithm performs well in networks with a lot of Rich-Club members.

In this paper, they propose an effective community-based approach with a scoring metric for identifying the top-K influential nodes. The scoring criterion in the TI-SC algorithm eliminates seed node overlap, resulting in the selection of a K-node with the best impact distribution.

3.5.3 Features

To summarize, the following are the main characteristics of this algorithm:

1. We investigate the impact maximization issue in the context of group structure, with the goal of reducing the search space for seed node selection.

2. We suggest using the scoring capacity of other nodes to minimize seed node overlap.

3. We use the relationship between core nodes to integrate communities with similar knowledge diffusion structures.

4. Experiments on synthetic and real-world networks reveal that the proposed algorithm TI-SC outperforms the base algorithms in terms of influence distribution. Traditional algorithms are slower than the TI-SC algorithm.

3.5.4 Limitations

Influence maximisation is a classic optimization problem that seeks to identify a subset of seed nodes in a social network with the greatest influence on a propagation model. The problem is exacerbated by seed node overlap and a lack of optimal seed node range.

3.6 Paper 6: Detection of Influential Nodes Using Social Networks Analysis Based On Network Metrics.

3.6.1 Introduction

Social media is a form of communication and interaction between people in which they develop, share, exchange, and access data and ideas, forming groups (small nets) and networks in the process (net of individuals and groups). A social network is a reflection of social network research that is made up of different social actors. The study of social issues. The study of social network metrics, which involves a variety of criteria that define social media analysis, is becoming increasingly important in the modern age. Various approaches are used for visualizing and analyzing the patterns of complex social networks. A mathematical model that connects various dots to evaluate nodes and their relationships is known as social network analysis. When it comes to analyzing complex networks, finding the main player in an online social network is the most crucial aspect. The main player is the person who has the most clout in the social network. The identification of powerful nodes from a social network has gotten a lot of attention in the online social culture in recent decades. The main hypotheses in the detailed analysis of social networks are network measures. The most critical metric for studying organizational and team behavior is centrality.

3.6.2 Characteristics and Implementation

The identification of powerful nodes from a social network has gotten a lot of attention in the online social culture in recent decades. The main hypotheses in the detailed analysis of social networks are network measures. The importance of the centrality measure in analyzing organizational and team actions cannot be overstated. The suggested procedure is as follows:

1) **Degree Centrality (DC):** The DC (Degree Centrality) is a network exposure index that counts the number of direct contacts a node has to determine how well it is connected to the network.

2) Centrality: The degree of interaction and contact between individuals in a social network is often measured by centrality.

a) Closeness Centrality (CC): The closeness or normalized closeness centrality (NCC) of a node in a complex network graph is the length that is called the average length of the shortest path between two nodes in the network.

b) Betweenness Centrality (BC): Betweenness Centrality (BC) is a quantifiable measure that gives a node complete control over its behavioral interaction with two other nodes in a social network.

c) Eigenvector Centrality (EC): Eigenvector centrality refers to the measurement of a node's influence in a social network.

d) Clustering Coefficient: The Clustering Coefficient determines how likely nodes are to be associated with one another.

3.6.3 Features

The paper is divided into four sections.

- Explain the related work that focuses on the key features that we used in social network analysis to dentify the key players (influential nodes).
- Our proposed scheme is Centrality
- The outcomes of the experiments and statistics used to evaluate the output
- Validity of the proposed schema.

3.6.4 Limitations

With an increasing number of people entering social networks every day, identifying prominent nodes in such a network is a difficult challenge.

3.7 Paper 7: Influential Nodes Detection in Dynamic Social Networks.

3.7.1 Introduction

With the rise in popularity of social media platforms such as Facebook and Twitter in recent years, more scientists studying the influence maximization problem have turned their attention to this area. This issue has gotten a lot of attention, and several studies have suggested different algorithms for detecting influential nodes, the majority of which are focused on static social networks. True social networks, on the other hand, evolve over time. New contacts are made between users, and some users lose touch. Since several powerful nodes can be identified by carefully studying the relationships among the links,complex social network analysis relies heavily on them.

3.7.2 Characteristics and Implementation

SNDUpdate's main goal is to find the most powerful nodes in a complex social network.It takes advantage of the network's structural and semantic features. As a result, the key concept is to suggest a two-phased approach. Indeed, the first phase of SNDUpdate focuses on the network's structural aspect, while the second phase focuses on the semantic aspect. Each consumer is defined in the next section of this paper by a collection of interests that are represented as an attributes vector.The weight of the relation between two nodes remains unchanged in the previous work because the powerful nodes are detected in a static social network.The network in this paper is dynamic, which means that the structure of the network changes, and thus the relation between two nodes belonging to asnapshot graph Gt can be removed in the snapshot graph Gt+1, resulting in a change in the meaning of the sem antic similarity between two nodes as well as a change in the set of leader nodes.

1) Phase 1: Community Detection

2) Phase 2: Influential Nodes Detection

3.7.3 Features

A social network resemble a graph structure made up of nodes and edges, with nodes representing members and edges representing interactions between them. Users form bonds with one another as time passes, and their interactions change. A changing social network is made up of social networks observations at various time stamps (G1,G2,...,Gn,) and includes notionly a collection of node relationships, but also information about how these relationships evolve overitime.

This segment addresses a study of various complex social network research studies. In this paper, we focus on the problem of detecting influential node in social networks with shifting edges.

1) Methods Based on a Non-linear Model
2) Methods Based on Metrics

3.7.4 Limitations

The aim of the influence maximization issue is to find powerful nodes that will help you achieve your viral marketing goals on social media. Previous research has primarily focused on static social network analysis and algorithm creation in this context. As the network evolves, however, certain algorithms must be modified.

3.8 Paper 8 : Performance Evaluation of Modularity Based Community Detection Algorithm in Large Scale Networks.

3.8.1 Introduction

In the field of complex networks, community detection is a hot topic, and several studies have been done on it. A widely accepted definition of a group in a network is a subset of nodes with high internal density but low external density. Several studies evaluating and comparing various measures of partition efficiency can be found in the literature. For example, Yang and Leskovec examine the suitability of several measures to classify ground-truth based communities in their paper. Moradi etal. In an email network, compares the capacity of various quality functions to distinguish useful and spam messages. Modularity, as suggested by Newman and Girvan, is currently the most commonly used criterion for assessing the consistency of communities in networks. Modularity can be thought of as the difference between the fraction of edges within a group and the fraction of edges predicted by a random version of the network while maintaining the degree distribution of the nodes inia general sense.

3.8.2 Characteristics and Implementation

The aim of this research is to look into the computational issues of group detection methods that can deal with large networks. The fine-tuning stage proposed in this paper is implemented with a reduction in the number of swapped nodes, allowing for up to 50% faster execution without compromising the consistency of the partitions obtained. The methods were chosen with the aim of defining a collection of computational tools capable of dealing with group detection modularity optimization under various approaches: divisive agglomerative; heuristic solution relaxed solution. The approaches were all implemented using free software in order to allow for a reasonable comparison. To do so, a collection of suitable data structures and analytical methods were used to the extent possible.

3.8.3 Features

The computational complexity of the researched approaches is investigated in terms of their computational implementation. The built algorithms are used to compare the Newman spectral approach with the CNM method on a qualitative and quantitative level, with a focus on their applicability to large-scale networks.

3.8.4 Limitations

When the number of moving nodes is increased to 20% (Newman-FT20%), the obtained modularity is almost identical to the value obtained when 100% of the nodes are moved in almost all of the tested networks. In other words, permuting more than 80% of the nodes in the fine-tuning stage almost always results in a massive waste of time and computational resources.

3.9 Paper 9: A comprehensive literature review on Community detection: Approaches and applications.

3.9.1 Introduction

Since it allows to expose coherent and meaningful sub-graphs, recognize the features, functions, structure, and dynamics of such complex networks, community identification has been designed as an axial field in Complex Network Analysis (CNA). In this regard, over the years, a variety of methods and approaches have been established to provide appropriate solutions to complex network paradigms, especially group detection problems. Meanwhile, scientists face a significant challenge in defining populations in a given complex network, which necessitates a significant amount of literature and survey. We outline literatures on community detection for complex networks in this paper because researchers need to conduct reviews on the major papers related to community recognition in complex networks and to point out their major

strength and limitations.We assume that this literature contribution, which does not include all current contributions, may be a valuable source of knowledge for practitioners in the field of group detection. As a result, we were more concerned with the importance of the chosen approaches' contributions than with the publication's chronological order.

3.9.2 Characteristics and Implementation

In this paper, we include a thorough analysis and categorization of various group detection approaches and methods in order to highlight their main strengths and limitations. In addition, examine studies that deal with implementation in various domains.

Finally, we concentrate on journal papers written in English that provide useful information for practitioners interested in the group detection issue, as well as the most recent findings, which are accessible via online databases.

1. Approach based static non-overlapping communities;
2. Approach based static overlapping communities;
3. Approach based static hierarchical communities;
4. Approach based dynamic communities.

3.9.3 Features

1. **Approach**: For group identification, a technique was used.
2. **Technical principal of approach**: The technical principle or algorithm of the implemented method is defined in this column.
3. **Network type**: This attribute includes "Weighted" if the edges connecting the studied network's nodes are weighted, and "Unweighted" if the edges aren't weighted.
4. **Directionality of the network**: Indicates "Guided" if hyperlinks between network nodes are directed, and "Undirected" if the hyperlink's directionality is ignored. If the solution can accommodate both structures, “all” is stated.
5. **Network nature**: If the network is static or dynamic is indicated by this attribute.

6. **Network size**: The network size is an essential metric for the approach's computation efficiency. The size of the sponsored network is listed in this column.
7. **Implementation datasets**: For each approach exploration, we include a standard abbreviation based on Table 1 of the used network datasets.

3.9.4 Limitations

The area of group detection is still evolving, and the categorization of an exhaustive list of methods and approaches appears to be a work in progress.

3.10 Paper 10: Community detection in social networks.

3.10.1 Introduction:

Individuals' social networks are formed by their experiences and personal relationships with other members of society. Personisocial links are represented and modeled by social networks. The rapid expansion of the internet has resulted in a massive increase in user engagement online. Many social networking sites, such as Facebook and Twitter, have sprung up to help users connect. It is becoming increasingly difficult to keep track of these messages as the number of interactions has increased dramatically. Humans are drawn to people who share their interests and preferences. People can expand their social lives in unprecedented ways thanks to easy-to-use social media. It is difficult to meet friends in the real world, but it is much easier to find friends with similar interests online. These real-world social networks have fascinating patterns and properties that can be studied for a variety of purposes.

3.10.2 Characteristics and Implementation

Communities are sections of the graph with fewer connections to the rest of the graph and denser connections inside them. Unsupervised learning aims to group related objects together without any previous knowledge of them. The clustering problem in networks refers to the grouping of nodes based on their similarity computed using topological features and/or other graph characteristics. In the literature, network partitioning and clustering are two widely used for the approaches of locating groups in a social network graph. In the following subsections, these methods are briefly defined.

1. **Graph Partitioning :** The method of partitioning a graph into a predefined number of smaller components with specific properties is known as graph partitioning.
2. **Clustering :** Clustering is the method of putting together a set of related objects into structures called clusters.

3.10.3 Features

i. **Trend Analysis in Citation Networks:** Citation networks iniacademiaiare made up of citation relationships between papers and researchers.
ii. **Improving Recommender Systems with Community Detection**: Recommender systems (RS) produce recommendations based on data from related users or products.
iii. **Media** : The focus and reach of sites are widening and diversifying as the number of SNS grows.

3.10.4 Limitations

In social networks, communities are increasing at an exponential rate. Each algorithm for various types of group detection has its own set of drawbacks.

3.11 Paper 11: Analysis of the Dynamic Influence of Social Network Nodes.

3.11.1 Introduction

People's social interactions have changed dramatically in recent decades as a result of revolutionary advances in communication technologies. Milgram's small-world experiment in the 1960s demonstrated that the average distance between any two individuals on Earth is six, a phenomenon known as six degrees of separation. The average distance between Facebook network nodes was just 4.74 degrees in 2011, according to the results of a study of the friend networks of 750 million active users on Facebook. Finding out or mining which node has the greatest effect is crucial in social network research. As a result, a variety of metrics, such as degree centrality, betweenness centrality, closeness centrality, k-shell centrality, eigenvector centrality, and the PageRank algorithm, have been proposed to quantify the value of a node from various perspectives.

3.11.2 Characteristics and Implementation

Most current measurements are based on statistical properties of network topology and do not account for the impact of changes in mutual confidence among nodes during information dissemination. A new scheme for measuring the complex power of nodes in a social network is introduced in this paper. The modification of node trust value during information propagation is essential in this new scheme. The new scheme also takes into account the accumulated shift in node confidence value.

3.11.3 Features

A new decision scheme for the complex power of social network nodes is introduced in this paper. A new calculation of node dynamic impact is proposed, taking into account the effect of changes in the information distribution process on confidence values. It is a step forward from previous algorithms. Finally, we examine the power of nodes based on network topology or statistical properties, and compare it to other classical algorithms to ensure the algorithm's validity and accuracy.

CHAPTER-4

DATA

4.1 Overview

The Global Terrorism Database (GTD)™ is the world's largest unclassified terrorist attacks database. The National Consortium for the Study of Terrorism and Responses to Terrorism (START) has made the GTD available on this platform in an effort to raise awareness of terrorist behaviour so that it may be researched and defeated more readily. The GTD is created by a dedicated team of researchers and technical specialists.

Since 1970, the GTD has collected data on domestic and international terrorist threats, and it presently comprises over 200,000 instances. Each instance has information on the date and location of the occurrence, the weapons used, the nature of the target, the number of injuries, and – when known – the organisation or individual responsible.

4.2 Dataset

Characteristics of the GTD :

- About 200,000 terrorist acts are recorded in this database.
- The world's most extensive unclassified database on terrorist attacks is currently available.
- Over 95,000 bombings, 20,000 killings, and 15,000 kidnappings and hostage cases have occurred since 1970.
- Each case contains information on at least 45 variables, with more recent incidents containing data on over 120ivariables.
- To collect event data, more than 4,000,000 news stories and 25,000 news sources were reviewed from 1998 to 2019.

CHAPTER-5

PROJECT REQUIREMENT SPECIFICATION

5.1 Product Perspective

By fusing international and domestic CT intelligence, providing terrorism research, exchanging information with stakeholders around the CT enterprise, and guiding whole-of-government action to protect our national CT priorities, we lead and integrate the national counterterrorism (CT) initiative.

5.1.1 Product Features

Selecting relevant features, constructing rough conceptualizations of terrorist groups, constructing terrorist group boundary areas, and developing terrorist group networks are all steps in eliciting terrorist group networks.

Finding the most powerful nodes: For intelligence and security informatics, predicting terrorist networks and recognizing key players is critical. Using machine learning methods, we suggest a framework for analyzing social networks. To delete unnecessary and passive nodes from the entire network, the proposed technique employs the k-core principle. It then uses a hybrid classifier to classify the key actors by extracting multiple features. The proposed technique is put to the test on a publicly accessible dataset, and the results show that the method is efficient.

In various fields, communities are referred to as classes, clusters, coherent subgroups, or modules; community identification in a social network entails recognizing sets of nodes where the connections between nodes within a set are greater than the connections between nodes in other networks.

5.1.2 User Classes and Characteristics

The proposed methodology, which elicits terrorist groups' networks, finds the most powerful nodes, and tracks the temporal evolution of terrorist networks, is not an open-source model; rather, these data are shared with national intelligence management in order to protect the country from terrorist threats and carry out counter-terrorism operations. This has been put in place with the help of college professors. Data scientists, National Intelligent Management, and Government are among the users who can change the dataset and improve the algorithms.

5.1.3 General Constraints, Assumptions and Dependencies

- Regulatory policies

We use the START project's Global Terrorism Database (GTD), which can be analyzed to include, for example, prediction models. The key idea behind this research is to use GTD's historical data, which includes information on terrorist attacks that have occurred since 1970, to conceptualize terrorist groups' activities over time intervals.

- Hardware limitations.

Depending on the hardware we use, analyzing the dataset and implementing the algorithm we implement takes time. When we use any modern CPU, the task is completed faster.The dataset needs the least amount of storage possible.

- Safety and security consideration

GTD research and algorithm implementation should not be used for anything other than educational and development (counter-terrorism) purposes. This processed data is only available to national intelligent management in order to defend the country from terrorist attacks and carry out counter-terrorism operations.

- Assumption: Terrorism cannot be defeated

Terrorism is, without a doubt, one of the defining characteristics of our day. It hits the news regularly, threatening or targeting states, private businesses, and ordinary people. It has also become one of the most serious challenges to peace, security, and stability in many parts of the world.

5.1.4 Risks

Data leakage occurs when sensitive or otherwise confidential information leaves an organization's infrastructure, leaving it vulnerable to unauthorized disclosure or malicious use. Mitigating the risks of such data handling and leakage may be a costly endeavor.

5.2 Functional Requirements:

Data is gathered from the Global Terrorism Database (GTD) as well as other sources. Steps for pre-processing have been completed. Rough Sets are used to approximate conceptualizations of terrorist group activities. Later, relevant features will be selected, rough conceptualizations of terrorist groups will be created, terrorist group boundary regions will be created, and terrorist group networks will be designed.

Identify more influential nodes in social networks using the neighbourhood correlation coefficient. The proposed method is based on the local clustering coefficient and exploits the similarity of connections between surrounding nodes. The method is based on a k-shell decomposition strategy, which determines a node's effect based on how it shares relationships with its neighbours.

The aim of this project is to categorize and compare theoretic group detect on algorithms. Using two strategies: centralized and dispersed, combine techniques for characterizing, identifying, and extracting populations. To examine the detection technique in relation to the dynamic of networks: static or

Evolving over time. To include metrics for partition detection: structural based partition or semantic and structural-based partition.

5.3 External Interface Requirements

5.3.1 User Interfaces

On the top of the project UI (web page), there are options such as checkout (gather the information given in the text field), view (display the network's visual output), and so on. In this project, we create a web page that displays the outcome of our work. Essentially, the user interface is a simple HTML page that allows the user to communicate with the project's outcome. The working area is in the middle of the tab, where the user can fill in the information and see the results. To communicate with the server, the project will use python and json. An error message will appear on the website if the user sends incorrect input or input format.

5.3.2 Hardware Requirements

Any Intel(7th gen or higher) or AMD(2nd gen or higher) processor with a base clock speed of at least 3.5 GHz.On the computer, all of the project's results are shown. The TCP protocol is used to retrieve the result from the server. The TCP protocol is used for all XML requests and responses between the client and the server.

5.3.3 Software Requirements

Python Version : 3.7 or higher
Operating Systems : Ubuntu 16.04 or higher, Windows 7 or higher, Mac OS 10 or higher

Tools and Libraries (Open-source):

igraph - The program for network analysis. igraph is a collection of network analysis tools focused on performance, portability, and simplicity of use.

NetworkX - NetworkX is a Python module that allows you to create, manipulate, and investigate the structure, dynamics, and functions of complicated networks.

5.3.4 Communication Interfaces

To obtain the result from the server, as well as all XML requests and responses between the client and the server, we use the TCP protocol. To load a few image format outputs, the line speed should be at least 10 kbps. The application's predefined functions will handle the network's entire buffer size.

5.4 Non-Functional Requirements

5.4.1 Performance Requirement

Efficiency: The number of nodes that can access a huge number of distinct nodes at the same time. – sourcesiof information, status, and so on – through a relatively limited number of connections is a measure of network performance. Nonredundant contacts are applied to these nodes.

Effectiveness: The purpose of effectiveness is to reach a cluster of nodes using non-redundant contacts. Performance, on the other hand, helps to reduce the amount of time and resources expended on redundant contacts. Each group of contacts is a self-contained source of information. Since people linked to one another want to know about the same things at about the same time, one cluster around this non-redundant node, no matter how large it is, is just one source of knowledge.

5.4.2 Safety Requirements

In order to have a stable and healthy working atmosphere, we in the safety profession have had to reconsider our positions. Engineering protections, procedural methodologies, and technological obstacles, as well as ways to eliminate risks and even weapons of mass destruction, have all been considered.
If a catastrophic failure, such as a disc failure, results in significant damage to a large portion of the database, the recovery method restores a previous copy of the database that was backed up to archival storage(typically) and reconstructs a more current state by reapplying or redoing committed transaction operations from the backed-up log, up until the time of failure.

5.4.3 Security Requirements

Since the model relies on the data for learning, it should not be tampered with or poisoned, since this might bring the system to a halt. Database storage is often needed by security systems.

5.5 Other Requirements

Scalability: We believe that the scope of this research is not limited to the Global Terrorism Database (GTD), but that it can be applied to other social media platforms as well.
Maintainability: The framework should be designed in such a way that it can be expanded in the future. It should be simple to add new feature specifications accommodate changes to existing requirements.

CHAPTER-6

SYSTEM DESIGN

6.1 Novelty

For intelligence and security informatics, predicting terrorist networks and identifying key players is critical, and few studies address this topic. As a result, we suggest a framework for analysing social networks that makes use of machine learning techniques (ie.,K-Core Concepts). After the networks have been clustered appropriately, we use group identification methodology to evaluate the relationships between terrorist nodes within the same cluster as well as between terrorist nodes from different clusters.

6.2 Innovativeness

The methodology we suggest elicits terrorist group networks, identifies the most powerful nodes, and tracks the temporal evolution of terrorist networks, but it is not an open-source model; rather, these data are shared with national intelligent management in order to protect the country from terrorist threats and carry out counter-terrorism operations.

6.3 Performance

The method we suggested has a higher efficiency since we calculate it by the number of nodes that can immediately reach a large number of different nodes by a relatively limited number of connections. Nonredundant contacts are used to handle the nodes.

The effectiveness of some approaches is targeted at a cluster of nodes that can be reached through non-redundant contacts. In our case, however, reliability means reducing the amount of time and resources

expended on redundant contacts. Each group of contacts is a self-contained source of information. Since people linked to one another want to know about the same things at about the same time, one cluster around this non-redundant node, no matter how large it is, is just one source of knowledge.

6.4 Reliability

The methodology we use is capable of conducting operations using data from the terrorist database and generating results in a time frame that allows us to focus on other tasks. It's dependable for the data we use and the effective outcomes we get at the end.

6.5 Maintainability

To ensure that the users see the correct results, we need to use good ranking methods and algorithms. We'd also give the nodes weights so that the value of defining the group changes over time. To ensure that the tool works properly, these ranking methods must be checked and revised on a regular basis. The underlying search engines results are retrieved using their respective APIs, which are all free of charge. If their respective policies change, maintenance would be needed.

6.6 Legacy to modernization

To improve operational efficiency as part of the legacy modernization, we're upgrading and optimising business processes by giving government agencies graphical access so they can see each community's development and powerful nodes in a single graphical view. As a result, users are able to meet their needs in terms of their experience and are more readily adapted to newer technology platforms.

6.7 Application compatibility

Since our project can run on a variety of operating systems and has a user-friendly environment, we can simplify the testing process, ensuring that all of the applications are checked for compatibility at the same time. To a certain degree, auto-removal of features that aren't enabled by the operating system is possible.

6.8 Resource Utilization

There is a lot of data to process, and much of it isn't necessary for the results we want. As a result, we only use the results that have a significant impact on terrorist growth and are also needed for future connection prediction among the groups. We can almost see the relationship between different groups and their development over time thanks to group detection. As a result, the data is used to meet our intermediate needs and forecasts.

CHAPTER-7

PROPOSED METHODOLOGY

7.1 Preprocessing, Feature Selection and Network Building

7.1.1 Selecting features from GTD

The suggested strategy is based on the notion of conceptualising terrorist groups using information about the attacks they have carried out. As a result, it's essential to pick a relevant subset of the GTD's features. Such a subset must be appropriate for describing terrorist groups behaviour.

To determine the typical conduct of a perpetrator, we must summarise the behaviours expressed by the same perpetrator in a given series of events. As a result, we have Mutual Information based feature selection. Mutual information is a measurement of the reduction in uncertainty for one variable when the other variable's value is known.

7.1.2 Designing the terrorist groups network

The first move is to construct the terrorist group's network. The network-building algorithm is fairly straightforward.Assume W = (V, E), with W representing the terrorist group's network, V representing the network's nodes, and E = V X V representing the network's borders.

7.2 Finding influential nodes .

The amount of links incident upon a node is known as degree centrality (i.e., the number of ties that a node has). If the network is directed (i.e., ties have a direction), two independent metrics of degree centrality, indegree and outdegree, are specified.

Degree

In graph theory, the degree (or valency) of a vertex is the number of edges incident to it, with loops counted twice. So the degree of a vertex v is denoted by $\deg(v)$ or $\deg v$. The maximum degree of a graph G is denoted by Δ(G), and the minimum degree of a graph is denoted by δ(G), these two are the maximum and minimum degree of its vertices.

Degree Centrality

Degree centrality, which is defined as the number of linkages that go to a node, is the oldest and potentially the most straightforward (i.e., the number of ties that a node has). The degree is measured by a node's immediate risk of catching whatever is travelling through the network (such as a virus, or some information). We commonly create two distinct measures of degree centrality in the case of a directed network (ties with direction), namely indegree and outdegree. As a result, indegree is the number of ties directed to the node, while outdegree is the number of ties directed to other nodes. When positive characteristics such as friendliness are present in a relationship, it is more likely to be successful.

For a given graph G:= (V, E) with | V | vertices and | E | edges, the degree centrality of a vertex V is defined as :

$$C_D(v) = \deg(v)$$

In order to Calculate the degree centrality for the given nodes in a graph takes $\Theta(V^2)$ if the graph is represented by a dense adjacency matrix , and for the edges takes $\Theta(E)$ in a sparse matrix representation.

The concept of node centrality may be extended to the entire network, which is referred to as graph centralization.

Let v* be the node in G with the highest degree centrality.

Let X:= (Y, Z) be the | Y | node linked network that maximises the following quantity (y* being the node in X with the highest degree centrality):

$$H = \sum_{j=1}^{|Y|} [C_D(y*) - C_D(y_j)]$$

In turn, the degree centralization of the graph G is as follows:

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v*) - C_D(v_i)]}{H}$$

When the network X comprises one central node to which all other nodes are linked (a star graph), the value of H is maximal, therefore in this situation :

$$H = (n-1) \cdot ((n-1) - 1) = n^2 - 3n + 2.$$

A high edge betweenness centrality score denotes a bridge-like link between two regions of a network, and its removal may disrupt communication between many pairs of nodes via the shortest pathways between them.

Betweenness centrality

In graph theory, betweenness centrality is a measure of centrality in a graph based on shortest routes. In a connected graph, there is at least one shortest path between any two vertices that minimises either

the number of edges the path travels through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) (for weighted graphs). The betweenness centrality of a vertex is the number of these shortest routes that pass through it.

In network theory, betweenness centrality is used ti indicate the degree to which nodes are separated from one another.In a telecommunications network, for example, a node with a greater betweenness centrality would have more control over the network since more data would travel through it. Betweenness centrality was developed as a broad measure of centrality that may be used to a wide range of network challenges, including social networks, biology, transportation, and scientific collaboration.

The expression: gives the betweenness centrality of a node v:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is basically the total number of shortest paths from a node S to a node t and $\sigma_{st}(v)$ is the total number of these paths that pass through v .

The summation indices imply that the betweenness centrality of a node grows with the number of pairs of nodes.As a result, by dividing through by the number of pairs of nodes that do not include v, the computation may be rescaled. , so that $g \in [0, 1]$. The division is done by $(N-1)(N-2)$ for directed graphs and $(N-1)(N-2)/2$ for undirected graphs, The number of nodes in the enormous component is denoted by N. Note that this scales to the greatest possible value, when every single shortest path crosses one node.This isn't always the case, and a normalisation can be done without sacrificing precision.

$$\text{normal}(g(v)) = \frac{g(v) - \min(g)}{\max(g) - \min(g)}$$

Which in-turn results in:

$$\max(normal) = 1$$
$$\min(normal) = 0$$

Weighted Networks

The linkages linking the nodes in a weighted network are no longer viewed as binary interactions, but are instead weighted in proportion to their capacity, influence, frequency, and other factors, which adds another layer of heterogeneity to the network beyond the topological effects.The sum of the weights of a node's neighbouring edges determines its strength in a weighted network.

$s_i = \sum_{j=1}^{N} a_{ij} w_{ij}$

With a_{ij} and w_{ij} being adjacency and weight matrices between nodes i and j , respectively. The strength of a particular node follows a power law distribution, similar to the power law distribution of degree seen in scale free networks.

$s(k) \approx k^{\beta}$

According to a study of the average value S(b) of the strength for vertices with betweenness b, the average value S(b) of the strength for vertices with betweenness b can be approximated by a scaling form.

$s(b) \approx b^{\alpha}$

Eigenvector centrality

Eigenvector centrality is a statistic for determining the amount of influence a node has in a network. A score or value will be awarded to each node in the network, with the higher the score, the greater the network's effect. The number of connections a node will have to other nodes determines its ranking. Connections to high-scoring eigenvector centrality nodes contribute more to the node's score than connections to low-scoring nodes. In graph theory, eigenvector centrality is a measure of a node's influence in a network. It assigns comparable ratings to all nodes in the network based on the assumption that connections to high-scoring nodes contribute more to the node's score than identical connections to low-scoring nodes.

Using the adjacency matrix to find eigenvector centrality

Let A be the adjacency matrix for a graph G : = (V, E) with | V | vertices. , i.e. $a_{v,t} = 1$ if vertex v is linked to vertex t , and $a_{v,t} = 0$ otherwise. Vertex v's relative centrality score is calculated as follows:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

where M(v) is a set of the neighbors of v and λ is a constant. This may be represented in vector notation as the eigenvector equation with a minor rearranging.

$$\mathbf{Ax} = \lambda \mathbf{x}$$

The extra constraint that all items in the eigenvector be non-negative, however, indicates that only the highest eigenvalue results in the required centrality measure (under the Perron–Frobenius theorem).The component of the corresponding eigenvector then yields the network vertex v's relative centrality score.Only the ratios of the vertices' centralities are well specified since the eigenvector is only defined up to a common factor.To determine an absolute score, the eigen vector must be normalised, for example, so that the sum of all vertices is 1 or the total number of vertices is n. One of several eigen value techniques that may be used to discover this dominating eigenvector is power iteration.

Furthermore, as with a stochastic matrix, this may be extended so that the elements in A are real values denoting connection strengths.

7.3 Community Detection:

7.3.1 Louvain algorithm using locality modularity optimization:

We use the Louvain algorithm with local modularity optimization to detect communities. This algorithm employs a greedy optimization method that iteratively attempts to improve the modularity of a network partition.

The strength of division in modules is measured by modularity, which is a measure of network structure. Dense connections exist between nodes in networks with high modularity, but sparse connections exist between nodes in different modules. It's often used in optimization methods for detecting network group structure.

In each iteration, the objective function is maximised to quantify the communities. Small communities are created in the first step (step 1) by optimising modularity locally. In this stage, only local community improvements are permitted. Nodes belonging to the same group are aggregated into a single node that represents community in a new aggregated network of communities in the next step (step 2). These steps are repeated iteratively until no further increases in modularity are possible with the creation of a hierarchy of groups.

When the original algorithm prevents the addition or removal of new nodes and edges after acquiring the group structure, the communities must be re-computation from the beginning.

7.3.2 Adding the edges/nodes:

Adding edges/nodes results in four types of effects at the community structure level in this method.

Cross-community edge:

When we try to join two nodes in a Cross-community edge that are already linked to other nodes, two things can happen. The community structure remains unchanged if the linking nodes belong to the same community. If the linking nodes belong to separate communities, the two communities are merged into one.

Inner community edge:

If the two nodes incident to the edge exist and belong to the same group, adding a new edge between these two nodes strengthens the community's inner connections while keeping the inter-community connections unchanged. As a result, the group structure remains unchanged.

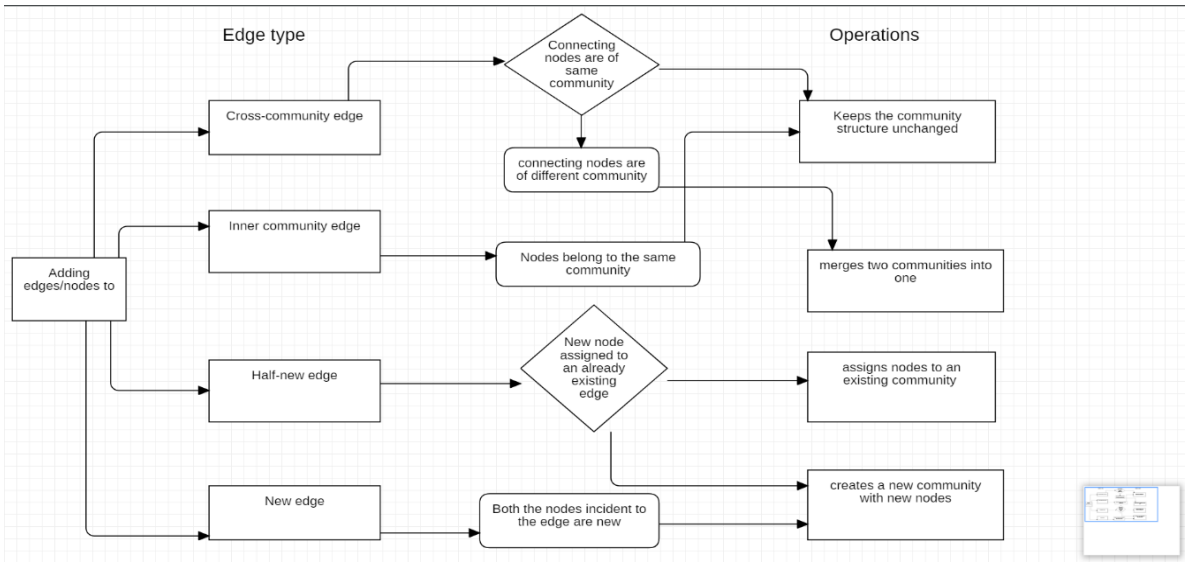

Fig.7.1 Adding edges/nodes

Half-new edge:

The increased edge in this one is a half-new edge, which means that one of the nodes is already in the network and the other is brand new. The community structure remains unchanged if the node is allocated to an established community. Otherwise, a new group with new nodes is created.

New edge:

Both nodes incident to the edge are fresh in the new edge. As nodes are added to a new edge, they are either assigned to the same new community or two separate communities are created for each node.

7.3.3 Removing the edges/nodes:

At the group structure stage, removing edges/nodes results in four types of effects:

Cross-community edge:

Two nodes incident to the removed edge belong to separate groups in a cross-community edge. Through removing these types of edges, the community's inner ties are preserved while intercommunity connections are reduced. This operation does not result in the merger of existing communities, nor does it disband any of the communities in which the removed edge is a member.

Inner community edge:

The two nodes incident to the edge in the inner-city edge belong to the same community. Removing these types of edges reduces the community's inner ties while maintaining the intercommunity connections. So, if the nodes connected to the removed edge are connected to other edges, the disbanding has no effect on the community structure; otherwise, the community is divided into smaller groups, and the sections which join other existing communities.

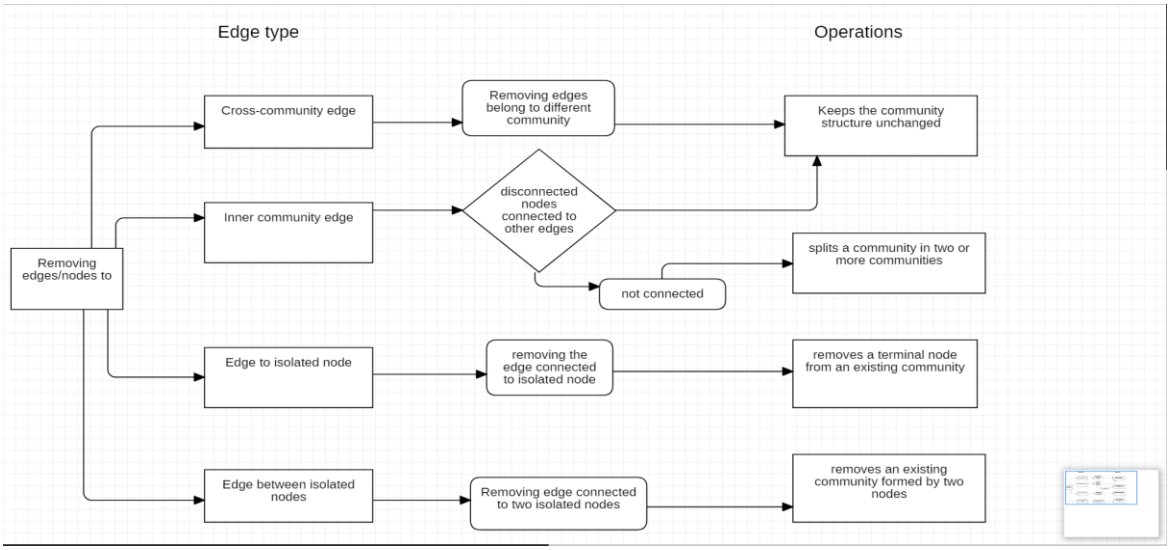

Fig. 7.2 Removing edges/nodes

Edge to isolated node:

One of the nodes incident to the edge is an isolated node, so eliminating this edge also means removing the isolated node. Since the removed node isia terminal node, this action has no effect on the community structure and hence has no effect on the community's inner connections.

Edge between isolated nodes:

The edge to be eliminated in Edge between isolated nodes belongs to two isolated nodes. Removing this edge means removing all nodes, effectively putting an end to the group or groups to which they belong. The rest of the community will be untouched.

Two of the eight resulting operations on adding/removing nodes and edges decrease the number of communities, two increase the number of communities, and four leave the community structure unchanged.

CHAPTER-8

IMPLEMENTATION AND PSEUDOCODE

Dynamic Community Detection Algorithm :

```
V ← {u1, u2, .., uv} , E ← {(i1, j1), (i2, j2), .., (ie, je)}
A ← array{(i1, j1), .., (im, jm)}
R ← array{(i1, j1), .., (in, jn)}
procedure Main(G ← (V,E), A, R)
Cll ← {C1, C2, .., Cn}, Cul ← {}, Caux ← Cll
InitPartition(Caux)
mod ← Modularity(Caux), old mod ← 0
m ← 1, n ← 1
while (mod ≥ old mod ∨ m ≤ |A| ∨ v ≤ |R|) do
Caux ← OneLevel(Caux)
n, c CommunityChangedNodes(Cll, Caux)
Cll ← UpdateCommunities(Cll, n, c)
old mod ← mod, mod ← Modularity(Cll)
Cul ← PartitionToGraph(Cll)
if m ≤ |A| then 16: src, dest A[m]
anodes ← AffectedByAddition(src, dest, Cll)
Cll ← AddEdge(src, dest, Cll)
Cll ← DisbandCommunities(Cll, anodes)
 Cul ← SyncCommunities(Cll, Cul, anodes)
end if
 if n ≤ |R| then
src, dest R[n]
anodes ← AffectedByRemoval(src, dest, Cll)
```

```
Cll ← RemoveEdge(src, dest, Cll)
Cll ← DisbandCommunities(Cll, anodes)
Cul ← SyncCommunities(Cll, Cul, anodes)
end if
Caux ← Cul, m ← m + 1, n ← n + 1
end while
end procedure
```

The algorithm's task procedures are logical and aggregate subprocedures for completing particular tasks (i.e. adding edge to the Cll). They are replicated until a modularity increase or the removal or addition of edges is necessary.

Procedure P1a: The addition of an edge to Cll results inithe retrieval of ailist ofiaffected nodesiand populations, as well as the addition of the edge itself to Cll.

Procedure P1b: In the Cll, the edge has been removed. When an edge is removed, this process involves retrieving a list ofiaffected nodesiand their populations, as well as the removal of the edge itself to Cll.

Procedure P2: Affected Communities in Cll. should be disbanded. Affected communities will be disbanded in Cll based on the list of affected nodes and respective communities obtained by AffectedByAddition() or AffectedByRemoval().

Procedure P3: Cul should be updated to reflect the changes in Cll. The AffectedByAddition() or AffectedByRemoval() lists of affected nodes and groups can also be used to reproduce the changes in group composition to the Cul. It's worth noting that the Cul will be changed as well as the added or disabled edges in this procedure.

Procedure P4: Cul will be used to carry out Step 1 of the Louvain Algorithm and quantify the changes in group structure that may lead to locally optimised modularity.

Procedure P5: By applying the Louvain Algorithm Step 1 to Cul, update Cll with the communities that have modified.

Procedure P6: To perform Phase 2 of the Louvain Algorithm and community aggregation, use the Cul.

Network Building Algorithm :

$V \leftarrow \{g1, g2, \ldots, gn\}$
$E \leftarrow \phi$
$W \leftarrow (V,E)$
for i=1.....m do
for j=1.....m do
  if $si,j \geq \gamma$ and $i \neq j$ and $(gj,gi) \notin E$ then
$E \leftarrow (gi, gj)$
    end if
  end for
end for

Adding Cross Community edges:

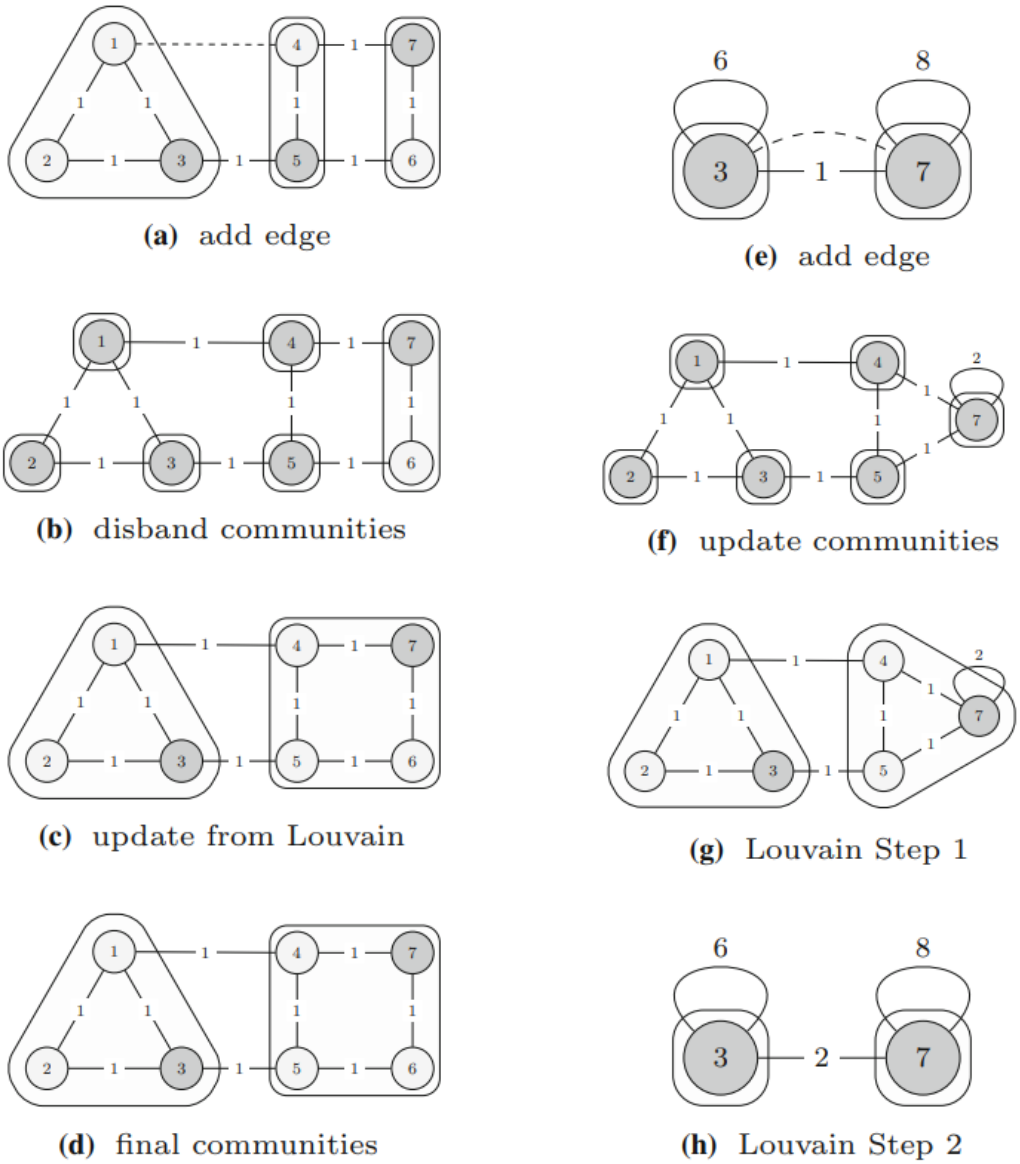

Fig. 8.1 Adding Cross Community edges

Remove Edge to terminal nodes:

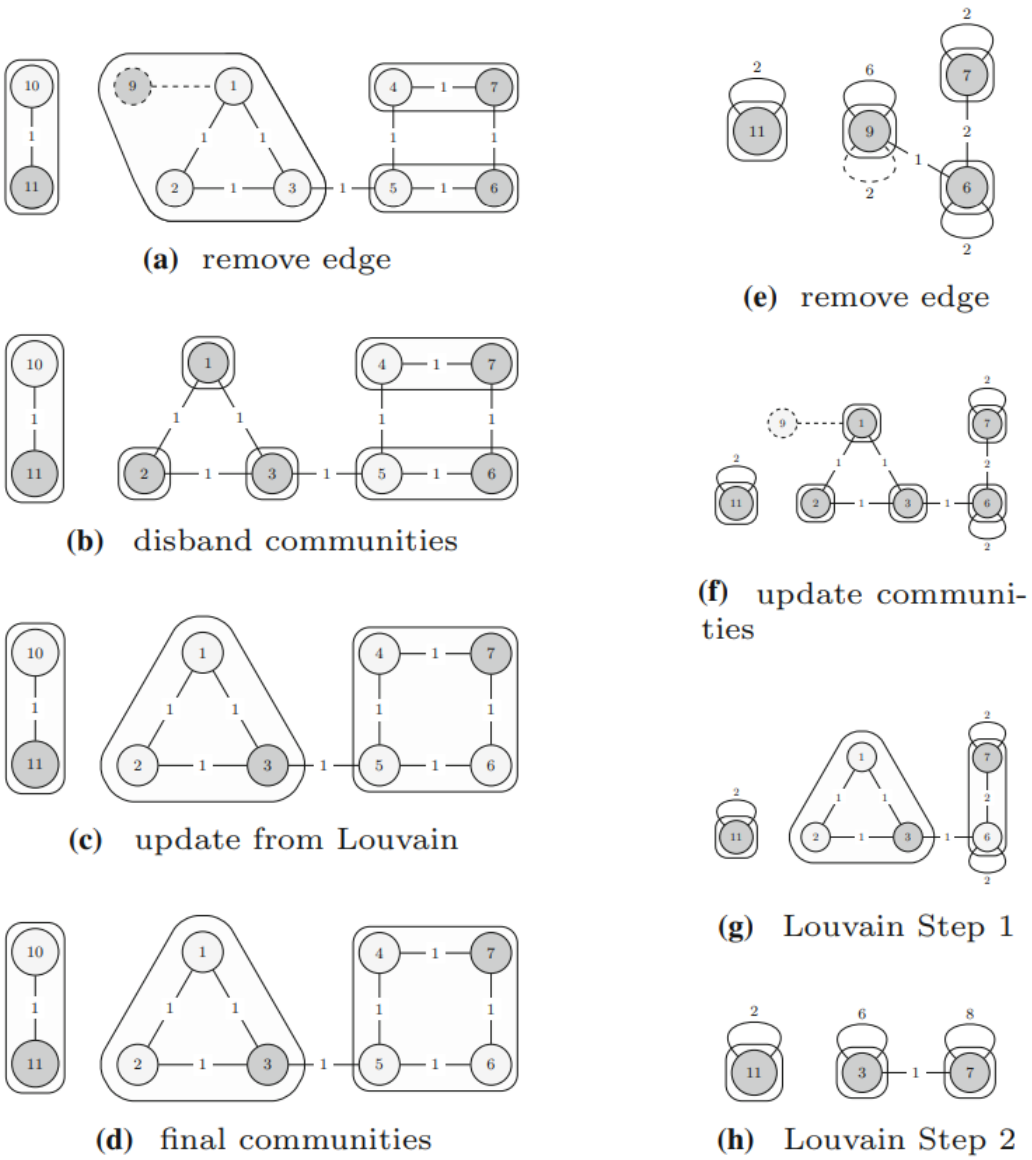

Fig. 8.2 Remove Edge to terminal nodes

CHAPTER-9

RESULTS AND DISCUSSION

Several algorithms have been proposed in order to solve unique problems in this area. We proposed a method in this study that has a low computational complexity and can be applied to large-scale networks. When it came to the consistency of the group structure, the results showed that thereiwas no penalty on the measured global modularity when just considering the locally modularity optimization for the regions where the network had nodes or edges added or removed. As predicted, the findings show that the number of edges added or removed should be adjusted to the network's scale. The likelihood of groups disbanding rises in direct proportion to the number of edges shifted in each iteration. Since the Proposed Methodology / Approach have algorithms with lesser time and space complexity which performs well, It is expected to provide better results.

CHAPTER-10

CONCLUSION AND FUTURE WORK

To understand the composition and evolution of terrorist networks we built the network using networkx, then we will find influential nodes based on centrality measures.

The number of links occurring upon a node that finds the significant node in the network is known as degree centrality. A bridge-like connector between two regions of a network with a high edge betweenness centrality score calculates the best link in the network. Eigenvector centrality is a metric for determining a node's amount of impact in a network, with the most influential node at the top. Finally, we apply the Louvain algorithm to find communities in the network utilising locality modularity optimization.

REFERENCES/BIBLIOGRAPHY

[1] Vinceno Loia, Francesco, "Understanding the composition and evolution of terrorist group networks: A rough set approach." Paper 2019, https://www.sciencedirect.com/science/article/pii/S0167739X19307757

[2] Ahmad Zareie, Amir Sheikhahmadi, Mahdi Jalili, Mohammad Sajjad Khaksar Fasaei, "Finding influential nodes in social networks based on neighborhood correlation coefficient." Paper 2020, https://www.sciencedirect.com/science/article/pii/S0950705120300630

[3] Mehdi Azaouzi, Delel Rhouma, Lotf Ben Romdhane, "Community detection in largescale social networks: stateoftheart and future directions." Paper 2019, https://www.sciencedirect.com/science/article/pii/S0020025517310101

[4] Kun Hea, Yingru Li a, Sucheta Soundarajanc, John E. Hopcroft , "Hidden community detection in social networks." Paper 2019, https://link.springer.com/article/10.1007/s12652-020-01760-2

[5] Hamid Ahmadi Beni, Asgarali Bouyer, "TI-SC: top-k infuential nodes selection based on community detection and scoring criteria in social networks." Paper 2020, https://www.researchgate.net/publication/333197917

[6] Aftab Farooq,Muhammad Uzair,Gulraiz Javaid Joyia,Usman Akram , "Detection of Influential Nodes Using Social Networks Analysis Based On Network Metrics" Paper 2018, https://ieeexplore.ieee.org/abstract/document/8346372

[7] Hong-Jian Yin, Hai Yu, Yu-Li Zhao, Zhi-Liang Zhu, Wei Zhang, "Analysis of the Dynamic Influence of Social Network Nodes." Paper 2019 https://www.hindawi.com/journals/sp/2017/5046905/

[8] Nesrine Hafiene,Wafa Karoui1, and Lotfi Ben Romdhane, “Influential Nodes Detection in Dynamic Social Networks”. Paper 2019 https://link.springer.com/chapter/10.1007%2F978-3-030-20482-2_6

[9] Mohamed EL-Moussaoui, Tarik Agouti, Abdessadek Tikniouine, Mohamed Eladnani , “Community detection: Approaches and applications.” Paper 2019, https://www.sciencedirect.com/science/article/pii/S1877050919305046

[10] Punam Bedi and Chhavi Sharma, “ Community detection in social networks ". Paper 2016, https://www.researchgate.net/publication/295395520_Community_detection_in_social_networks

[11] Vinicius da Fonseca Vieira, Carolina Ribeiro Xavier, Nelson Francisco Favilla Ebecken and Alexandre Goncalves Evsukoff , “Performance Evaluation of Modularity Based Community Detection Algorithms in Large Scale Networks.” Paper 2014, https://www.hindawi.com/journals/mpe/2014/502809/

APPENDIX DEFINITIONS, ACRONYMS AND ABBREVIATIONS

GTD - Global Terrorism Database

The GTD is an open-source database, which provides information on domestic and international terrorist attacks around the world since 1970, and now includes more than 200,000 events. For each event, a wide range of information is available, including the date and location of the incident, the weapons used, nature of the target, the number of casualties, and when identifiable the group or individual responsible.

START - Study of Terrorism and Responses To Terrorism

The National Consortium for the Study of Terrorism and Responses to Terrorism better known as START, it is a university-based research and education center comprised of an international network of scholars committed to the scientific study of the causes and human consequences of terrorism in the United States and around the world.

SNA - Social Network Analysis

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them.