Community Detection in Dynamic Networks*

1st Mahammad Thufail
dept. of Computer Science
PES University, EC Campus
Bangalore, India
thuppa786@gmail.com

2nd Manne Vasanth
dept. of Computer Science
PES University, EC Campus
Bangalore, India
vasanthmanne9482@gmail.com

3rd Purushotham S
dept. of Computer Science
PES University, EC Campus
Bangalore, India
purushotham1103@gmail.com

***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 counter terror policies and practices 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 organizations 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.**

***Index Terms*—Social networks, Terrorist Groups, Influential node, Community Detection.**

I. Introduction

Terrorism is an organized 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.

II. Data

A. 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.

B. 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 atleast 45 variables, with more recent incidents containing data on over 120 variables. To collect event data, more than 4,000,000 news stories and 25,000 news sources were reviewed from 1998 to 2019.

III. Project Requirement Specification

A. 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.

a) 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.

b) 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.

c) 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.

d) 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.

B. 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.

Using the neighborhood correlation coefficient, identify more influential nodes in social networks. The proposed approach uses the similarity of connections between neighboring nodes and is based on the local clustering coefficient. The method is based on a k-shell decomposition approach, in which a node's influence is determined by how it shares relations with its neighbors.

The aim of this project is to categorize 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 metrics for partition detection: structural-based partition or semantic-and-structural-based partition.

IV. System Design

a) 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.

b) 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.

c) 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.

d) 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.

e) 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.

f) 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.

g) 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.

h) 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.

V. Proposed Methodology

A. Preprocessing, Feature Selection and Network Building

a) 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.

b) 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.

Algorithm for building terrorist groups network : V← g1, g2, . . . , gn E ← W ←(V,E) for i=1.....m do for j=1.....m do if si,j and i j and (gj,gi) E then E ← (gi, gj) end if end for end for

B. Finding influential nodes

Degree centrality is defined as the number of links incident upon a node (i.e., the number of ties that a node has). If the network is directed (meaning that ties have direction), then two separate measures of degree centrality are defined, namely, indegree and outdegree.

a) Degree: The number of edges incident to a vertex in a graph is the degree (or valency) of the vertex in graph theory, with loops counted twice. So the degree of a vertex is denoted by or . The maximum degree of a graph G is denoted by (G), and the minimum degree of a graph is denoted by (G), these two are the maximum and minimum degree of its vertices.

b) Degree Centrality: Degree centrality, which is defined as the number of linkages occurring to a node, is historically the earliest and theoretically the simplest (i.e., the number of ties that a node has).The degree can be expressed in terms of a node's immediate danger of catching whatever is moving over the network (such as a virus, or some information).In the case of a directed network (ties with direction), we often construct two independent measures of degree centrality, namely indegree and outdegree.As a result, indegree is a count of the number of links directed to the node, whereas outdegree is the number of ties directed to others. When relationships are connected with good features such as friendship or collaboration, indegree is frequently viewed as a type of popularity, whereas outdegree is interpreted as gregariousness.

For a given graph G:= (V, E) with — V — vertices and — E — edges, the degree centrality of a vertex V.

In order to Calculate the degree centrality for the given nodes in a graph takes if the graph is represented by a dense adjacency matrix , and for the edges takes 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):

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

c) Betweenness Centrality: Betweenness centrality is a measure of centrality in a graph based on shortest routes in graph theory.There exists at least one shortest path between any pair of vertices in a connected graph that minimises either the number of edges that the path travels through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs).The number of these shortest routes that travel through the vertex is the betweenness centrality for that vertex. In network theory, betweenness centrality is used to 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.

Where is basically the total number of shortest paths from a node S to a node t and 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 . The division is done by for directed graphs and 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.

d) 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.

With and 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.

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

e) EigenVector Centrality: Eigenvector centrality is a metric for determining a node's amount of impact in a network.Each node in the network will be assigned a score or value, with the higher the score, the stronger the network's effect.This rating is based on how many connections a node will have to other nodes.Connections to eigenvector centrality nodes with high scores contribute more to the node's score than connections to nodes with low scores. Eigenvector centrality is a measure of a node's influence in a network in graph theory.It gives all nodes in the network comparable ratings based on the idea that connections to high-scoring nodes contribute more to the node's score than equivalent connections to low-scoring nodes.

f) 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. if vertex v is linked to vertex t , and otherwise.

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 eigenvalue 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.

C. Community Detection

a) 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 a 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.

b) 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.

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.

c) 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.

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 is a 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.

Implementation And Psuedocode

Algorithm for building terrorist groups network :

```
  V← g1, g2, . . . , gn
E ← phi
W ←(V,E)
for i=1.....m do
  for j=1.....m do
    if si,j   and i  j and (gj,gi)  E then
      E ← (gi, gj)
    end if
  end for
end for
```

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
```

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

[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