Dissertation on

"Terrorism Hub Detection"

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

Bachelor of Technology
in
Computer Science & Engineering

UE18CS390A – Capstone Project Phase - 1

Submitted by:

Sumukha	**PES2201800282**
Karthik M	**PES2201800410**
Vinyas Vittal	**PES2201800641**
Varsha G P	**PES2201800559**

Under the guidance of

Prof. Sreenath M.V.
Professor
PES University

January - May 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

‘Terrorism Hub Detection’

is a bonafide work carried out by

Sumukha	**PES2201800282**
Karthik M	**PES2201800410**
Vinyas Vittal	**PES2201800641**
Varsha G P	**PES2201800559**

In partial fulfilment for the completion of sixth semester Capstone Project Phase - 1 (UE18CS390A) in the Program of Study -Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period Jan. 2021 – May. 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 6th semester academic requirements in respect of project work.

Signature	Signature	Signature
Sreenath M.V.	Dr. Sandesh B J	Dr. B K Keshavan
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 - 1 entitled **"Terrorism Hub Detection"** has been carried out by us under the guidance of Prof. Sreenath M.V, 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 January – May 2021. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES2201800282	**Sumukha**	
PES2201800410	**Karthik M**	
PES2201800641	**Vinyas Vittal**	Vinyas
PES2201800559	**Varsha G P**	

ACKNOWLEDGEMENT

I would like to express my gratitude to Prof. Sreenath M.V., Department of Computer Science and Engineering, PES University, for his continuous guidance, assistance, and encouragement throughout the development of this UE18CS390A - Capstone Project Phase – 1.

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, Professor, 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 Capstone Project could not have been completed without the continual support and encouragement I have received from my family and friends.

ABSTRACT

Terrorism centers and their vulnerable targets are identified. We are going to analyze the feature of the network by building a dynamic global terrorism network. We investigate the network's resistance to targeted attacks and random failures. We used the disparity filter algorithm that used to identify the back bone of the terrorist network and its vulnerable motifs and also, we used the influence maximization algorithm that enhance the final result.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

Terrorism may be described as a deliberate and orchestrated effort to incite mass hysteria to achieve political or ideological goals by violence or threats of violence. It aims at instilling terror in the general population, including religious and ethnic competing groups, state governments or whole nations. Recently, physicists have been trying to incorporate scientific models, statistical measurements, network observations, as well as possible responses to crime, conflicts and other social issues, compared to social science remedies. This paper provides a network-based study of terrorist centers' identification and vulnerability. The Global Terrorist Database (GTD) analyses terrorist events and collected material from print and streaming media from 1970 to 2016. In reporting on actual events, media sources are believed to be unbiased and truthful. We create a diverse world terrorism network and we look at the antisocial characteristics of the network. We are investigating the sensitivity of the Network to target attacks and alarm disturbances that may contribute to the development of terrorism strategies. We use a method to isolate the disparities and isolate the infrastructure of the network and categories terrorist centers. What key organizations we should target to maximize the spread of influence while the promising seeds are concerned with maximizing influence. It is achieved using the theory of information diffusion to define the influential node within the terrorist network by means of certain algorithms.

CHAPTER-2

PROBLEM DEFINITION

“Identifying terror hubs and vulnerable motifs of covert network using disparity filter method.”

We examine the resistance of the network to intentional and random attacks, which may enable counter-terrorist organizations, and develop terrorist strategies. We use a filter technique to isolate the core of the network and identify terror centers and vulnerable acts of terrorism. Later on, we can find terror hubs that are more involved, in which case the controlling node with the Cost-Effective Lazy Forward (CELF) and the Grade Discount algorithms can be compared.

CHAPTER-3

LITERATURE SURVEY

3.1 “Uncloaking Terrorist Networks by Valdis Krebs”

Introduction

This paper examines the use of data from news outlets on the Internet to chart clandestine networks. We look at the network that surrounded the horrific events of September 11, 2001 in particular. With the help of publicly available data, we build a network based on nineteen hijackers The terrorist group is depicted on this map in some detail. To forecast the network's form, he used network layout algorithms.

Characteristics and Implementation

The best approach for network disruption could be to find potential perpetrators and then map their individual personal networks using snowball sampling to see who else they lead to and where they intersect. Diverse intelligence services aggregating their individual knowledge into a broader emergent map tends to be the best tool for finding these suspects. A more accurate image of potential danger can be drawn by exchanging information and experience. He used different countries data in the form of news report etc. In order to win the terrorism people, need to have well knowledge and indigence about the networks.

Features

Author encounter some problems while handling the and analyzing the network.

1. Missing nodes that make the availability of the data incomplete
2. Lack of measuring the importance to the nodes in the network.

3. These networks are dynamic, meaning they are constantly evolving.

Data source:

News source several newspapers

Network build

He chose to map three strengths' relations (and corresponding thickness). The extent of the ties will be largely determined by the amount of time a pair of terrorists spent together. Strength of the bond between the nodes is measured as closeness, attend the same school, or take the same classes/training. Many who fly together and attend meetings together will have moderately strong and medium-thick links. Weak links those who are far away from each other which are represented as the thin width in the network.

Evaluation

We must define mission and ties of trust between conspirators to provide a plausible image of an overt network. By mapping the same four alliances that we map across several sectors, we understand a lot about criminal categories. For cooperative customers, this information can be difficult to gain. It will take a long time for undercover criminals to be confronted with an enormous challenge.

Benefits:

The paper's authors split network borders into three categories based on their importance or power. In terms of network expansion, this is highly advantageous.

3.2 “Measuring Link Importance in Terrorist Networks by Uffe Kock Wiil, Jolanta Gniadek, Nasrullah Memon”

Introduction

This paper provides a novel way of evaluating the relevance of ties in terrorist networks focused on transport network studies. The attach value is used to evaluate known terrorist networks in the Crimefighter Assistant.

Summary:

Before moving on on to particular approaches relating to the study of a terrorist network, the author used various techniques for analysis of the social network and the terrorist network. We are presenting and evaluating a novel approach to identify linkages in terrorist networks. Terrorist networks share knowledge between nodes and links. If a suitable node is deleted, the network will be destabilized. Any tied that binds a node is also removed when the node is removed. The importance of a node and connection must be taken into consideration when determining which nodes to delete in order to destabilize a network. By emphasizing on the importance of connections, a node linked by critical links to other nodes becomes important. The objective of this document is to decide which interconnections are essential to the security and performance of the network and how the removal of one will affect the secrecy and efficiency of the network.

The graphic theory is used to measure social networks with size, density, nodal degree, cluster and central node. Node Closeness Centrality, Node Betweenness Centrality and Eigenvector centrality.

Characteristics and Implementation

Characteristic of terrorist network:

1. Terrorist networks operate in the shadows. The main distinction between terrorist and normal social networks is their secrecy.
2. Terrorist networks have close relations between members, but they are not clear or apparent in everyday life.
3. Long-term relationships are required;
4. They often "asleep," which means they are planning but not participating. It would be more difficult to locate them this way.

Methods used:

Secrecy of a network:

1. Exposure probability is the probability of nodes in the network which is identified as the terrorist group.
2. Link detection probability: This is the likelihood of a part of the network being exposed if a member is identified.

Find the efficiency:

1. To find the efficiency of the network we have to calculate the sum of the shortest path.
2. Then we have to take the inverse of it.
3. To find the network's average performance, we have to divide this result by number of pair nodes.

Features

Importance of Links in Transportation Networks
The sum of travel costs from node I to node k is known as performance in transportation networks. Based on this, a definition of link value for transportation networks has been proposed, which takes into account three considerations. (1) travel demand between nodes I and j; (2) the generalized cost of travel between nodes I and j in the initial, undamaged network; and (3) the generalized cost of travel between nodes I and j when link k is closed

Importance of Links in Terrorist Networks:

The calculation of association importance in terrorist networks was affected by transportation networks. The total production is used to determine how powerful a terrorist network is. P0 refers to the network's original performance, while Pk refers to the network's performance since link k has been removed. As a consequence, when link k is removed, the output change can be described as Pk is equal to P0 – Pk.

Evaluation

Advantages:
This paper provides the detail about

1. How to analyze the terrorist network.
2. How to measure the secrecy and efficiency of covert networks.
3. How the transport network and terrorist network can be connected.

Research on information management is essential before doing a network study. Data obtained and analyzed in order to produce useful information delivered in a network (Filtered, mined, etc.). This paper does not focus on the necessary mechanisms of information management.

3.3 “Identifying the global terror hubs and vulnerable motifs using complex network dynamics By Syed Shariq Husain, Kiran Sharma, Vishwas Kukreti , Anirban Chakraborti Link”

Introduction

Construct a complex global terrorism network and investigate its growth mechanisms as well as the anti-social network's statistical properties. Investigate the terrorist attacks found in the Global Terrorism Database.

Characteristics and Implementation

1. They use the open-access Global Terrorism Database to investigate terrorist activity attacks from around the world over the last half-century in order to create a picture of their spatial-temporal dynamics.
2. They create a dynamic network of global terrorism and investigate its development processes, as well as the anti-social network's mathematical properties, which are interesting. In most cases, each government pursues its own vision of international security based on its mission and specific political and security policies to combat terrorism, which can require tactical interventions, negotiating arrangements, or even physical action.
3. They investigate network tolerance for targeted threats and random errors, which could aid counter-terrorism organizations in developing strategies. The foundation of the network is

then isolated, and a disparity filter system is used to classify terror centers and global terrorism prone motifs.

4. The complexities of terror centers and delicate motifs discovered in the network backbone are fascinating, and they may provide profound insight into their creation and dissemination, assisting in the battle against terrorism or the formulation of public policy that can stop it.

Features

Construction of a network

1. Whenever actor a1, actor a1, attacks a target, actor a2, an event E1 is recorded.
2. The source is connected to the target by a rim or a led link by an arrow of the unit weight (could be generalized by weighing the edge by impact of attack, etc.)
3. If source a3 and objective a2 are engaged in a different event E2, a3 is connected to a2 by a directed relation of unit weight.
4. As a result, sources a1 and a3 are both linked to the same goal, a2.
5. By summing all such events in the time window t, connected components are generated.
6. Any time the same pair of actors is engaged, the edge mention (weight) increases by one. As a consequence, the frequency of attacks between the source and target pairs determines the edge's weight.

Evaluation

The disparity filter algorithm removes the backbone by taking the required edges into account at all scales of the structure and by benefiting from local heterogeneity and local weight correlations.

3.4 “Modelling the re-emergence of information diffusion in social network by Dingda Yang a,d , Xiangwen Liao b,d, Huawei Shen c,e , Xueqi Cheng c,e , Guolong Chen a”

Introduction

It suggested a new knowledgeable state to be included in the conventionally sensitive–infected–removed paradigm, an enhanced model of information diffusion in this study. They test the model suggested in real-world social networks and the findings demonstrate that information reappear during the dissemination process will correctly be replicated.

Characteristics and Implementation

Method used:

The algorithm of forest fire and its contribution to the modelling of information diffusion. In the traditional forest fire model, there are three states for cell automatics: vacuum, tree and fire. The model has two likelihoods of p and f, where p represents the probability of new forest trees, and f represents the probability of burning. To begin with, an algorithm is used to model the degree of the dissemination of information or the total number of infected persons using an online social forest fire model. It proposes adding a Burnt node to the traditional model of forest fire to disperse and spread the polluted population. This helps to identify those who distribute the information. It sets out the parameters that underlie the distribution of information.

For online social networks, the modified forest fire algorithm (MFF) is used. As Algorithm 1 is grabbed by a tree, it stretches and extends the fire to its neighbors. In social networks, not all help to disseminate content. This will divide the two groups of infectious or competent individuals – distributors and non-distributors. We connect a new Burnt Node to the model non-spreader. This is related to forest fires where a large number of trees may be burned completely without further spreading flames.

Evaluation

Advantage:

This paper seeks to shape the evolution of the dissemination of information in social networks. According to the report, the educated state played a major part in the distribution process. We find the curious process of re-emergence of diffusion, as we use comparatively small values for parameters.

Negative aspects:

The population variation has not been taken into account, and some other important factors have been overlooked, weakening the position of the informed state.

3.5 “Information Diffusion Models and Algorithms for Influence Maximization in Social Networks- A Survey by Lal Vikram Singh and S.Siva Sathya”

Introduction

Diffusion is a process that transmits creativity over time across specific networks to participants of a social system. This concept introduces three essential elements: individual members, mutual relationships, and communication networks, which serve as the foundation for a potential analytical system.

Various diffusion models have been proposed to research the exchange of an attitude or emotional state among a large group of people, in areas as diverse as viral marketing knowledge dissemination on blogs and infectious disease transmissions in epidemiology.

The algorithms for the Influence Maximization problem, as well as recent developments in theoretical propagation models or diffusion models of online social networks, are discussed in this paper. A overview of existing models of diffusion is given. And it uses this tool to examine the various methods to maximize effect with high scalability. The last part of the article concludes with applications for social management and dissemination of information.

Characteristics and Implementation

Social impact and maximization of influence: Rashetta describes social influence as the change in the thinking, feelings, attitudes and behavior of a person resulting from interaction with other individuals or groups. Social influence takes multiple forms in online social networks and is visible everywhere (OSNs). Many technologies, such as viral marketing, recommendation systems and information dissemination, struggle with the societal effects in the area of data processing and big data research. In the analysis of social control one of the most critical topics is maximizing leverage (IM).

Models for the dissemination of information: the framework for disseminating knowledge through the use of intermediaries amongst individuals in a social network. At present, a variety of diffusion models have been established by economic and sociological societies. The Linear threshold model and the Independent Cascade model are the most popular methods used to analyze social effect issues.

In addition to these two well-known models, numerous versions and extensions represent more complex real-world scenarios. This section has analysed the latest literature on theoretical models of the diffusion of power.

Maximizing Influence Algorithms

The enhanced and submodular effect characteristic f(.) is monotone. The glamorous algorithm of Kether and others addresses the problem at a ratio of 1 to 1 / e- > 0. Yet O(n2(m+n), forbidden in big networks, is the worst running time for the naïve greedy algorithm. This has contributed to a lot of change.

The algorithm used to describe impact node maximization is:

The high scalability of effect maximization topics, such as CELF+, MPINS and LDAG algorithms, have been demonstrated in recent algorithm experiments.

Feature

Study and practice will continue to grow as social networking becomes more common in millions of people's everyday lives. Moreover, the networks to which the systems underlying the network have to be run increase over time. As a consequence, reliable and active social effect solutions are highly requested. The paper offers a thorough analysis of models and algorithms for influencing social networks.

Evaluation

A major area of study in the future will be the advancement of flexible, precise and quantifiable social impact models for different uses of social networks and social media. This field borders on informatics, sociology and physics. The way sociologists conceive about this issue has evolved in terms of scalable, parallel data mining algorithms, scalable database and network technology. The use of large-scale data mining algorithms in analysing social network data is growing, rather than designing numerical modelles and conducting simulation and user studies.

3.6 “Information Diffusion in Social Networks by Mohammed Zuhair Al-Taiecorresponding author and Seifedine Kadr”

Introduction

Knowledge is transmitted from one location to another by interactions, which is known as diffusion. It is a discipline that incorporates techniques from a variety of sciences and areas. No one want to become infected with a disease that is infectious.

On a wider scale, and with the interconnections between cultures, it is very likely that the technology could jump over border crossings (or bridges) from community to community and then start to spread again. The more people an individual is connected to, the more likely they are to follow the innovation. Early adopters are often too creative to have an impact in a local network. Stopping product production, minimizing product distribution, lowering product visibility, declining product interests or reducing group interactions are both examples of measures.

Characteristics and Implementation.

Two-Step Flow Model

Since decision leaders who control those in the media have influence, the model is also known as the multi-stage flow model.

In contrast to the previous assumption that individuals were actively affected by the media, the model proposed that media influences were induced by interpersonal impacts. By media communications, opinion makers were able to persuade voters to follow their views to support their arguments so they were more vulnerable to press and more knowledgeable of current trends.

Rate of adoption

The acceptance rate is a measure to determine the dissemination rate at a certain time. The network structure clearly has a major influence on the dissemination process, in order to reach a significant portion of a target population in a shorter period of time through a dissemination process which begins with a node in a central position. The flow of adoption is also affected by the way individuals are inventive and intimate.

Adoption Categories and Thresholds

People may be grouped into adoption groups depending on when they were adopted and how they compare to other adopters. Adopters are divided into four groups of marketing: Early adopters representing 16%; an early majority representing 34%; a late majority representing 34%; and late adopters and later adopters representing 16% of all adopters. Marketers benefit from this definition because it enables them to identify early adopters' social and demographic characteristics.

Intake amounts

The part of a neighbor who introduced the breakthrough before then is that of a single entity in a network at a given time. If the requisite exposure level has been achieved, the individual may authorize the innovation and start to infect others. This model is easier to handle because it shapes what takes place at the microlevel each time it is diffused. Some researchers also refer to the amount of exposure as a network exposure, referring to the effects of the social network of an individual.

3.7 “Practical Issues and Algorithms for Analyzing Terrorist Networks by Tami Carpenter, George Karakostas, and David Shallcross”

Introduction

Graphs are used to model relationships within social network analysis between actors or participants. Each node or vertex in the graph shows a participant or actor. The bond or edge is expressed for each interaction or connection between two participants. A variety of graphic algorithms have been developed to analyze the dynamics of social networks and the location or importance of each player. There are many ways to determine each actor in a social network's "importance" or "centrality." For the most common of these centrality measures, the calculation or listing of the shortest paths between all the pairs of nodes is needed. Such calculations will take a long time in large graphs. They may also become a problem in moderate size networks when shifting data or when routine precomputations are used for "what if" scenario analyses.

Characteristics and Implementation

We should emphasized the need to calculate the length of the shortest path between all node pairs in G to calculate closeness centrality. The solution of the shortest all pair paths (APSP) problem needs to be found. In unweighted graphs, the best algorithms for APSP on sparse graphs need O(nm) operations, and in weighted graphs, (log) 2 O nm + n n operations. These limits asymptotically become ( ) 3 O n in dense graphs in the worst case, but recent (more complicated) algorithms provide slightly better worst-case bounds.

This paper provides an approach based on the study of transport networks motivated by the role of linkages in terrorism networks. The relational meaning calculation for identified terrorist networks is extended and analysed in Crimefighter Assistant.

Feature

1. This paper goes into how to analyze a terrorist network in depth.
2. Clandestine networks' confidentiality and effectiveness can be measured.
3. The relevance of the transportation network and the terrorist network be determined.

Evaluation

Limitation: Work on information management is essential before doing a network study. Data must be gathered and analyzed to produce useful knowledge structure information (filtered, mined, etc.). This paper does not focus on the necessary processes of information management.

3.8 “Network Topography, Key Players and Terrorist Networks By Sean F Everton, PhD is an Assistant Professor at the Naval Postgraduate School in Monterey, California, in the Defense Analysis Department.”

Introduction

In recent years we've strengthened our understanding of how extremist networks function and our study of social networks (SNAs). To date, however, SNA analysis has continued to focus on main players within a highly central network, or whose structural status (i.e. their location in the overall network) makes it possible for them, within the network, to share information and/or services. While this focus can seem intuitive and provide short-term excitement, it can put your chariot in front of your horse.

Before we classify key players, we must first study the overall topography of a network. Research show that networks too provincial (i.e. dense, high clustering levels, overflows of strong ties), too cosmopolitan (i.e., low clustering levels, overflows of small ties), too hierarchically (i.e. clustered, low variance levels), or too heterarchical (i.e., decentralizing, high variance levels), are poorly functioning. According to the research. In some cases, the main players' approach may be efficient but may have detrimental implications in others, if these dynamisms extend to terrorism networks too. In particular, it shows that before developing interruption techniques, analysts should think about the total topography of a network.

Characteristics and Implementation

A number of steps may be used to calculate network topography. Regarding heterarchical hierarchy, degrees, proximity and centralization, we must be mindful of how a network is organised. The higher the coefficient of centralization, the greater are the probability that one player is actually central and the other not such that the unequal distribution of individual actor values can be taken into account. Therefore, we need to look at the different indices in terms of rough types of centralities.

Network is the most often used metric for measures of the provincial-coastal scale density. Unfortunately, as social networks grow wider, network density continues to decrease. As future lines increase exponentially, as the number of possible lines increases, the number of actors exceeds the number of connections between and actor. It is normally small you we keep up with. In certain cases, it is therefore just helpful. As a measurement It can be used for comparing similar-scale networks. However, that's what it has to do.

An alternative method for calculating the average of a network the degree of centrality Although it is positively correlated with network "provincials," it is unaffected by network size. It helps analysts to compare networks of various sizes.

Feature

In this paper they conclude that while research into social networks has increased our awareness of how terrorist network's function, it is largely not accountable for the complexities of a network. Consider the total topography before methods are established for its interruption. Moreover, recent research has continued to focus on the identification of players with a strong central position or structural position to allow them to broker agreements. The results of previous research. Knowledge is transmitted over a network.

However, as they have seen, there is evidence that too provincial, too cosmopolitan, too bureaucratic or too heterarchical networks function poorly than those that match these extremes. If extremist networks have the same dynamics, it is maybe not necessarily the safest solution to recognize key actors in the network.

In the near future, more studies are obviously needed to examine all the sophisticated terrorist networks, not only to define their central players but also to delineate their topographical characteristics.

Evaluation

The creation of metrics that assess the effectiveness of terrorist networks reliably may also be included in more study. One method of measurement to be taken into account is the number and scale of attacks (Milward & Raab, 2008). For our purposes, it is however less important to locate the right metric than to recognize that if one or more relationships have been formed we can only empirically verify (or disprove) the hypothesized relationships.

Limitation: There is a tradition of disregarding the general topography of the network while developing tactics for disruption. If extremist networks are using the same dynamic, it might not always be the safest approach to recognize key actors in a network.

3.9 “Tractable Models for Information Diffusion in Social Networks by Masahiro Kimura and Kazumi Saito”

Introduction

As we advanced the Internet, the World Wide Web and blogging, we have collected genuine massive social networks, a lot of focus has been paid to social networking. A web of links and interactions among social institutions, including individuals and organizations and groups is a social network. All of these are examples of e-mail networks, website link networks, blog tracking systems and networks of science cooperation.

As awareness, ideas and authority can be transmitted through a social network through word-of-mouth contact, it is important in sociology and marketing to seek powerful nodes for information dissemination in the underlying network. The problem of location of nodes that contain a large quantity of data, for example, is essential.

Characteristics and Implementation

When it comes to identifying strong nodes focused on the dissemination of information in a broad social network, we have to use the independent cascade model to quantify the forecast value (ICM). The number of nodes influenced by a node group. Nevertheless, a reasonable estimate of this quantity requires a considerable amount of ICM.

In this article, they propose as natural special cases of ICM two new models of information diffusion to calculate a good approximation of (A). We are examining the properties of the proposed models using huge real information from a blog network and a test network.

Firstly, the proposed model is compared to the ICM, standard methods of social network analyses and the 'PageRank' methodology for ranking methods of removing nodes of influence and shows that the proposed models provide new modular methods of grading, which can remove nontrivial nodes in general, as powerful Nodes. It also shows that, if the probability of dissemination through links is minimal, the proposed model will give the ICM good approximations to locate sets of influence in the social network. On the other hand, when they took into consideration the effect maximization problem at ICM K4-0etal obtained a demonstrated success guarantee for a natural greedy algorithm.

If the odds of spread over links are minimal, you can give strong ICM approximations to find sets of influence nodes.

Feature

The two natural models for the diffusion of information in a social network (SPM and SP1M) were suggested, to evaluate effectively the influence (A) of a practical goal. For the natural greedy algorithm, we have given a proven performance guarantee for maximization of impact problems in the models proposed. We also examined the SPM and SP1M properties across current big social networks.

Initially, they have demonstrated that the models proposed can include new scalable classification methods to exclude strong nodes from a social network.

They showed that, if the probability of dissemination through links is minimal, the ICM can be used to find sets of influence nodes on the social network.

Evaluation

In addition, the models proposed can be scaled and even quicker than the ICM to solve the problem of influence maximization using a greedy algorithm. They therefore assume that the SPM and SP1M can be helpful models to analyze social networks dependent on the dissemination of information.

3.10 “Predicting Information Diffusion in Social Networks Using Content and User’s Profilesby Cedric Lagnier,Ludovic Denoyer, Eric Gaussier, and Patrick Gallinari”

Introduction

Diffusion is an iterative process that relies on whether the information is already disseminated by the incoming neighbors of the recipient. Models influenced by IC or LT- have two key failings on the other hand:

1. They do not believe the information's content has been disseminated.
2. They don't take into account any detail about social network users.

In this article, you propose a new family of diffusion models in which (a) the information input is used; (b) and user's profile is taken into consideration; and (c) you are able to disseminate a specific piece of information. To test the validity of this new model family, experiments with two real, widely used blogosphere datasets are used.

Characteristics and Implementation

Predicting content shared on social networks is a big obstacle for applications such as opinion leader detection, buzz detection and viral marketing. Many new models of dissemination are direct extensions of Cascade and Threshold models, respectively, originally introduced for epidemiology and social sciences. The dissemination process of these models is focused largely on the complexities of interactions between neighboring network nodes (social pressure) and lacks important dimensions, such as the content of the disseminated information.

1. In this article they propose a new family of probabilistic models which aims to predict how an information component diffuses in a network through the incorporation of additional dimensions such as the content of the information component, the profile of the consumer and the readiness to spread.

2. These models are seen and compared to other approaches in two blog datasets.
3. Experimental findings on these datasets show that proper modelling of the diffusion process allows the accuracy of the diffused knowledge to be taken into account.

Feature

The aim of this article is to forecast the increase in traffic across a three-dimensional network: the content, user profiles and the capacity for distribution (user-center models). They also demonstrated how these dimensions can be integrated into fundamental characteristics and a new probabilistic model developed.

Two blog datasets were used to describe our models and equate them with other approaches. The experimental findings obtained from these datasets demonstrate that (a) the quality of the disseminated information plays a major role in the dissemination process and that, as previously done, (b) user identities often play a significant role, which have been noted despite not being used extensively in recent studies on dissemination.

CHAPTER-4

DATA

4.1 Overview

The GTD is the largest unclassified terrorist activity archive in the world. Via this website, the GTD is made available by the START National Consortium to raise awareness of Terrorism Activities so that they can be investigated and resolved more quickly. The GTD is founded by a committed team of researchers and technical staff. The GTD is an open access archive that has compiled national and international terrorism threats data since 1970. Every case contains the date and place of the attack, the weapons involved, the purpose of the objective, the number of casualties and – if identifiable – the party or individual responsible.

4.2 Dataset

Characteristics:

- Includes detail on more than 200,000 assaults.
- Today, the world's largest unclassified archive of terrorist assaults.
- This includes details on over 95,000 bombings, 20,000 murders and 15,000 abductions and rehabilitation incidents since 1970
- Indications on at least 45 variables per occurrence, and reports on more than 120 variables, with more recent events.
- More than 4,000,000 headlines and 25,000 news reports alone from 1998 to 2019 have been revised to gather event info.

CHAPTER-5

SYSTEM REQUIREMENT SPECIFICATION

5.1 Product Perspective

We lead and provide a strategic counter-terrorism Strategy (CT) through the integration of foreign and domestic CT intelligence, terrorist analysis, knowledge exchange with partners around the CT company and whole-of-government actions to defend our national CT priorities

5.1.1 Product Features

The development of terrorist group networks is all moves toward selecting characteristics, creating rough conceptualization of terrorist organizations, building terrorist group border zones and developing terrorist group networks.

Search the strongest nodes: Predicting extremist networks and identifying key actors is crucial for intelligence and defense informatics. We propose a methodology for the analysis of social networks using machine learning techniques. The strategy suggested uses the k-core theory to remove redundant and passive nodes from the whole network. It then uses a hybrid classifier to identify the primary actors with multiple characteristics. A freely available dataset is used to validate the proposed methodology and the results demonstrate that the procedure is successful.

Community recognition on a social network involves identifying collections of nodes where the links between nodes are larger than those of the nodes on other networks. In different fields the community is called classes, clusters, coherent subgroups, or modules;

5.1.2 User Classes and Characteristics

The suggested approach that promotes militant organization networks is not an open-source model, but is shared with the national intelligence management in order to defend the country from terrorist attacks and carry out counter-terrorist operations. This was done with the assistance of the university professors.

The users who update the dataset and develop the algorithms include data scientific, national smart management, and government.

5.1.3 General Constraints, Assumptions and Dependencies

Policies governing regulatory compliance:

The global terrorist database (GTD) of the START project can be analyzed for predictive models, for instance. The main theory underpinning the project is to use historical GTD data, including statistics on terrorist acts since 1970, to conceptualize the actions of terrorist groups over time.

Hardware limitations:

The analysis of the dataset and the implementation of the algorithm that we use takes time according to the hardware we are using. The mission is done faster if we use any current CPU. The dataset needs the minimum amount of storage.

Considerations for safety and protection:

For nothing but education and advancement (counter-terrorism), the application of GTD research and algorithms can be used. This stored data is accessible only to smart national management to protect the country from terrorist threats and to conduct counter-terrorism operations.

Assumption:

Terrorism is unbeatable. Terrorism is one of the distinguishing features of our day, beyond doubt. It periodically reports news, threats or targets countries, private companies and ordinary individuals. It also has been one of the greatest obstacles in many parts of the world to peace, prosperity and stability.

5.1.4 Risks

Data leakage happens when classified or otherwise protected information leaves the networks of an entity and leaves it at risk of unwanted exposure or malicious use. It may be expensive to mitigate the costs of such data handling and leakage.

5.2 Functional Requirements

Data is collected from the GTD as well as from other databases. Steps have been completed for pre-processing. Rough sets are used to estimate militant group operations conceptualizations. The characteristics are then chosen, the conceptualization of terrorist groups created, the boundary regions of terrorist groups created and the networks of terrorism groups developed.

Identify more dominant nodes in social Networks with the neighborhood correlation coefficient. The solution is based on the similarity of the links between the adjacent nodes and the local clustering factor. The process is based on a k-shell decomposition technique, which determines how a node interacts with its neighbors. The project aims to compare and category theoretical algorithms for group detection. The combination of methods for characterizing, distinguishing and separating populations uses two strategies: centralized and disperse. To investigate the methodology of identification in relation to network dynamics: steady or time-changing. To have partition recognition metrics: structural or semantic and structural partition.

5.3 External Interface Requirements

5.3.1 User Interfaces

At the end, there are choices like checkout (pick up information provided in the text box), view (show visual performance of the network), etc. At the top of the project UI. We build a web page with the results of our work in this project. The user interface is essentially a plain HTML page that enables the user to convey the project result. The section where the user can fill out the details and display the results is located in the center of the page. The project uses python and json to communicate with the server. If the user sends wrong input or input format, an error message may appear on the website.

5.3.2 Hardware Requirements

An AMD processor with a base claim speed of 3.5GHz or greater, or Intel (7th or higher). Both project reports are shown on the screen. This results from the server can be obtained via the TCP protocol. For all XML questions and answers between the client and the server the TCP protocol is used.

5.3.3 Software Requirements

Python 3.7 or higher is needed.
Ubuntu 16.04 or higher, Windows 7 or higher, and Mac OS 10 or higher are the supported operating systems.

(Open-source) Tools and Libraries:
igraph is a network analysis program. igraph is a collection of network analysis tools focused on performance, portability, and simplicity of use.
NetworkX is a Python package for creating, manipulating, and studying complex networks' structure, dynamics, and functions.

5.3.4 Communication Interfaces

To obtain the server's response and all XML requests, we use the TCP protocol.
The message and response of the client and the server, the line speed should be at least 10 kbps to load a few images format outputs. The predefined functions of the program will control the whole buffer size of the network.

5.4 Non-Functional Requirements

5.4.1 Performance Requirement

Effectiveness:

Network efficiency is measured by the amount of nodes that can reach a vast number of different nodes immediately - information source, status, etc.. These nodes are not subject to redundant connections.

Affordability:

The cluster of nodes that non-redundant connections can meet is the objective of efficiency. Performance, on the other hand, helps decrease redundant touch time and energy. Each communication community is a source of knowledge in its own right. Because people who are connected at the same time want to know the same things, one cluster around this non-redundant node is only one source of information, however big it may be.

5.4.2 Safety Requirements

We in the safety profession had to rethink our jobs and have a stable and safer working environment. Both considerations were taken on the safety of engineering, methodologies of proceeding, technical barriers and means of eliminating threats and even mass destruction weapons.

If the accident has badly destroyed a significant part of the archive, a prior backup of the files, for example a disc collapse, would be returned by the recovery system in the case of malfunctions.

Based on archival storage (typical tape) and archival storage, it reconstructs a database (typically tape). transactions from backed-up log in a more current state by re-applying or re-doing dedicated operations before the loss.

5.4.3 Security Requirements

The model must be kept up-to-date because of details to be learned. It should not be spoiled or poisoned, so it could lead to the machine knees. Protection devices also need database storage.

5.5 Other Requirements

We assume that the scope of this research is not limited to the Global Terrorism Database (GTD), but can be applied to other social media sites.

Maintainability:

The system should be planned to be extended in the future. Adding new functional specifications or adapting modifications to current requirements should be simple.

CHAPTER-6

DESIGN DETAILS

6.1 Innovativeness

The usage of two algorithms is made in the model namely disparity filter and information diffusion. By doing so, the desired network will get reduced due to the disparity filter algorithm and at the same time nodes which are socially influencing are detected making it a single system powerful tool to find the targets.

6.2 Performance

Since it is a reduced network, for the further analysis of the network including the influence maximization, when it comes to performance wise the model can be said to work pretty well as the burden of vast network is being reduced in the first place.

6.3 Security

Since the model depends on the data for learning purpose which is a trusted source, hence it should not be corrupted or poisoned. Otherwise, it could bring the system down to its knees.

6.4 Maintainability

The model is implemented using python modules like Igraph, NetworkX which is open source and free to use. Maintenance would be required if their respective policies changes.

6.5 Portability

Algorithm used in the projects are very much portable as they are not fixed to a system and not dependent on it to run. Rather they can run on any system that meets the required specifications and features.

6.6 Legacy to modernization

Since the genesis of network analysis there has always been a lot of transformations brought to it. Better algos keeps coming up and developed that in some way improves the overall effectiveness of analysis. So, to implement network analytics consistently in its way to produce good results it always keeps getting modernized through with time.

6.7 Reusability

Using the python, we reuse many modules such as Numpy, NetworkX, Igraph for the implementation.

6.8 Reliability

There are proof of results that the models using Disparity filter method show better insights when used with other algorithms than models working directly with the detection algorithms.

6.9 Application compatibility

The model is compatible with most browsers with up-to-date specifications. It runs independent of the system as it runs in a browser.

CHAPTER-7

IMPLEMENTATION AND PSEUDOCODE

7.1 Guide to understand the procedure involved

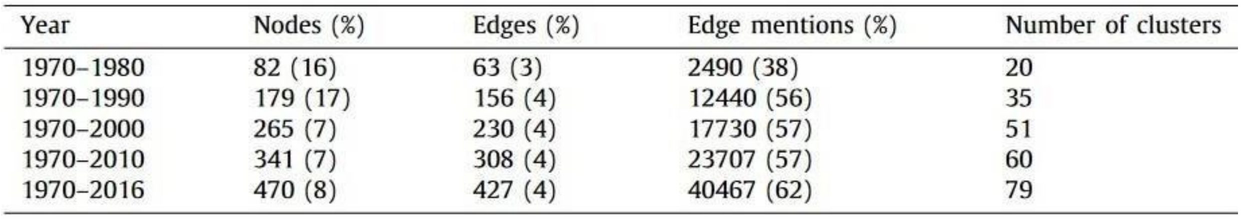
The overall approach’s workflow

- Selection of relevant features
- Constructing the terrorist network by recording the events
- Reducing and extracting the backbone of terrorist network
- Identifying the terror hubs as influential nodes in the network

7.1.1 Selection of relevant features

The open-access Global Terrorism Database (GTD) was provided by the University of Maryland's National Consortium for the Study of Terrorism and Responses to Terrorism (START). News articles from all around the globe are included in the GTD archive. The dataset had to be properly washed before any analysis could be performed. Certain incidents were removed because the dataset showed that they were suspicious. Terrorism attacks by "Unknown" terrorist groups on "Unknown" locations were also filtered out. The dataset's events with no spatial information were also removed. To maintain network modularity at the country level, any attack that attacked the international community rather than a single ethnicity was excluded from the dataset.

Table 7.1 Characteristics of alphac=0.01 growing backbone structures



Year	Nodes (%)	Edges (%)	Edge mentions (%)	Number of clusters
1970–1980	82 (16)	63 (3)	2490 (38)	20
1970–1990	179 (17)	156 (4)	12440 (56)	35
1970–2000	265 (7)	230 (4)	17730 (57)	51
1970–2010	341 (7)	308 (4)	23707 (57)	60
1970–2016	470 (8)	427 (4)	40467 (62)	79

7.1.2 Constructing the terrorist network by recording the events

We look at terrorist activity over a 46-year period (1970-2016), with the exception of 1993 (due to a shortage of evidence for that year), but our data reviews are one-day granular. When a terrorist source, actor a1, targets a target, actor a2, an edge or directed attachment ‘connects' the source and the target with a unit-weight arrow (could be generalized by calculating the edge by impact of attack, etc.), where t0 is the initial time in the entire span. If source a3 and target a2 are both involved in event 3 E2 in the same time window, a3 is connected to a2 by a led unit weight relation. As a consequence, sources a1 and a3 are related to the same objective, a2. By aggregating all such events over the time window, linked components are generated. Any time the same pair of actors is engaged, the edge mention (weight) increases by one. As a consequence, the frequency of attacks between the source and target pairs determines the edge's weight.

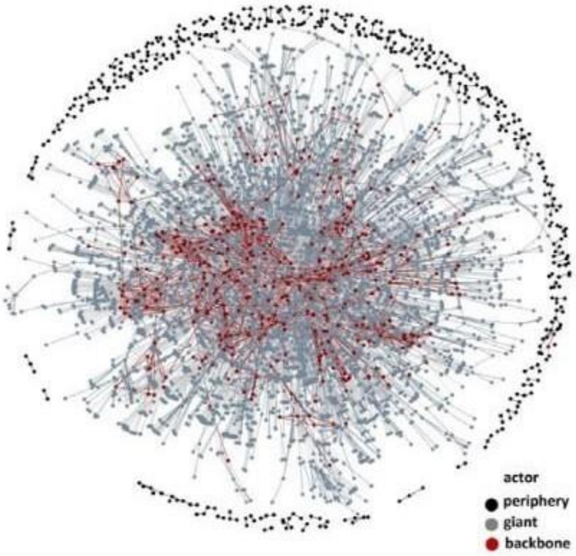


Fig.7.1 For the years 1970–2016, the global terrorism network was compiled.

7.1.3 Reducing and extracting the backbone of terrorist network

The disparity filter algorithm removes the network backbone by considering acceptable edges at all scales in the structure and exploiting local heterogeneity and contextual correlations among the weights. The intensity of a node I, denoted by si, is defined as si=jwij, where wij denotes the weight of the relation between I and j. To implement the disparity filter algorithm without overlooking nodes with low capacity, a normalised weight pij is defined as pij = wij/si. The normalised weights of a node with degree k are created in the null model as follows: Between 0 and 1, k 1 pins are allocated at random. After that, the interval is separated into k subintervals. The length of the subinterval represents the normalised weight of each reference in the null model. For a given normalised weight pij, the p-value ij of pij based on the null model is given by

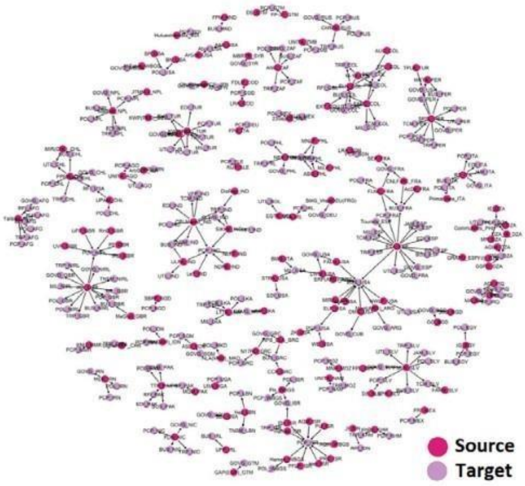
$$\alpha_{ij} = 1 - (k-1)\int_0^{p_{ij}} (1-x)^{k-2}\,dx \text{ which reduces to } \alpha_{ij} = (1-p_{ij})^{k-1}.$$

ij corresponds to the probability of having a normalised weight greater than or equal to pij in the sense of the given null model. Any reference of normalised weight pij will be filtered out if ij is greater than a given significance threshold (between 0 and 1). We can eventually remove unnecessary links by modifying, effectively extracting the backbone structure of the weighted network.

Pseudocode:

```
if(filter=="wls_conf") // filtering with confidence
{
        if(!no_downscale)
        {
                // downscale the views to speed-up the matching stage, as we will need to compute
        both left
                // and right disparity maps for confidence map computation
                //! [downscale]
                max_disp/=2;
                if (max_disp%16! =0)
                        max_disp += 16-(max_disp%16);
                resize (left ,left_for_matcher ,Size(),0.5,0.5);
                resize(right,right_for_matcher,Size(),0.5,0.5);
                //! [downscale]
        }
        else
        {
                left_for_matcher = left.clone();
                right_for_matcher = right.clone();
        }
}
```

Fig.7.2 The disparity filter of c = 0.01 was used to identify the decade-by-decade growth of the network backbone from 1970 to 2010.



7.1.4 Identifying the terror hubs as influential nodes in the network

Diffusion is the process of information being transferred from one place to another by experiences. The acceptance rate is a statistic that can be used to calculate the rate of diffusion at a certain point in time. This figure reflects the number of new citizens who have followed the invention at any given time. Clearly, network layout has a huge impact on diffusion, such that a diffusion mechanism that starts from a node in a central area can cover a significant portion of the target population in a shorter time. The amount of exposure for a single person in a network at any given time is determined by the percentage of an individual's neighbors who have applied the invention before that time.

If that person has had the requisite amount of exposure, he or she will approve the invention and start infecting others. The strength of a person's social network is calculated using the following equation:

$$E_{i=} \frac{\sum W_{ij}\, y_j}{\sum W_i}$$

W denotes a network weight matrix describing a single person's direct connections, and y denotes adoption operation.

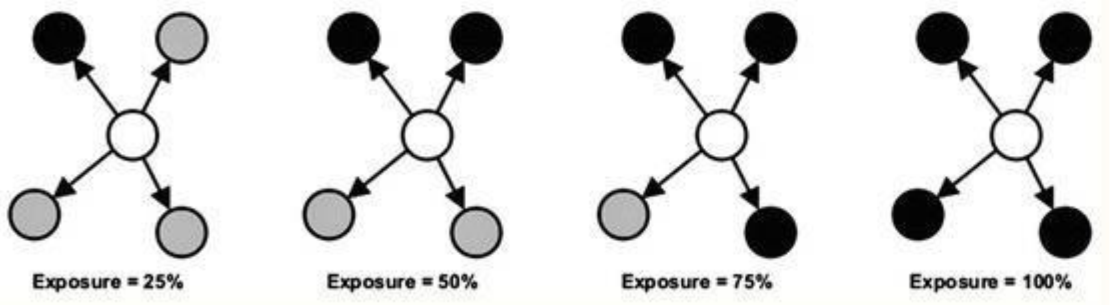


Fig.7.3 Amount of exposure at four-time steps

Some of the hand-picked algorithms for recognizing influential nodes are mentioned below.

Greedy Algorithm: It simply goes through k rounds, picking a new entry with the largest marginal gain in f in each round.

```
Greedy_InfluenceMaximization( )
 {
        Input: Graph, integer k, any set function f
        Output: Seed set S
        Initialize S with empty set
        While (|S|≤k)
        {       Select node which have maximum function weight;
                Update S with S ∪{u};
        }
        Return S;
 }
```

When it comes to massive social networks, the greedy algorithm, on the other hand, is much slower. This bug can be effectively resolved using the two algorithms mentioned below.

CELF (Cost-Effective Lazy Forward selection) is based on the idea that a node's marginal advantage in the current iteration cannot exceed its marginal value in previous iterations. The marginal value of u in comparison to S is u(S). u(S) is only re-evaluated for the top node at each stage, and the table is only resorted when it is needed. If a node stays at the end, the next seed will be selected. In practice, a heap Q is used to keep track of the sorted table information and to represent each node's priority. Except for the first, this optimization avoids re-computation of residual gains for all nodes in each iteration. The CELF optimization speeds up the greedy algorithm by 700 times as a result of the experimental results.

```
Greedy Algorithm optimized with CELF()
{
        Input: Graph, integer k, $\sigma_m$ // where $\sigma_m$ is marginal gain.
        Output: Seed set S Initialize S and Q with empty set For
        each u ε V
        do
        {
                Calculate marginal gain for each {u};
                Initialize round with zero;
                Add u to Heap Q;
        }
        While (|S|≤k)
        {
                u ← root element in Q;
                if (round == |S|)
                {
```

```
 Update S with S ∪{u};
                    Update Q with Q-{u};
              }
              Else
              {
                    Update Marginal gain= σm(S ∪{u}) -σm(S);
                    Round = |S|;
                    Reinsert u into Q and heapify.
              }
       }
       Return S;
}
```

Consider vertex v to be a neighbor of vertex u in terms of the general definition. We should not consider the edge v u when selecting v as a new seed based on its degree if u has already been selected as a seed. We should discount v's degree by one as a result of including u in the seed set, and we should do the same discount on v's degree on every neighbor of v who is still in the seed set. This is a basic degree discount heuristic that works for any cascade model. When using the Fibonacci heap, the running time of Algorithm is O(k log n + m). As a consequence, the Degree Discount Heuristic outperforms the original greedy algorithm greatly.

```
DegreeDiscount(G; k)
{
       initialize S with empty set;
       for each vertex v
       do
```

```
    {
        Compute the degree of every vertex $d_v$;
        $D(d_v) = d_v$;
        initialize tv =0;
    }
    for I = 1 to k do
    {
        Select u = maximum
        {
            $D(d_{v)}\}$;
            Update S with $S \cup \{u\}$;
        }
            for each neighbor v of u and v ε V \S
            do {$t_v = t_v + 1$;
            update $D(d_v)$ with degree discount;
    }
Return S;
}
```

CHAPTER-8

CONCLUSION OF CAPSTONE PROJECT PHASE 1

In order to solve unique problems in this field, many algorithms have been suggested. We are using the difference filter procedure to separate the backbone of the terrorist network and identify areas of terror and weak foreign terrorism motives. We have chosen an $\alpha c=0.01$ which allows us to track the development of terrorist centers and weak motifs in the backbone systems and to extract the backbones of the networks over the various evolution cycles. The network size on which the programmes underlying the system must be used tends to increase over time. Reliable and productive mechanisms of social control are also highly sought after. This paper presents a number of algorithms that demonstrate their advance and disadvantage and how effectively they can be used to maximize influence in social networks.

CHAPTER-9

PLAN OF WORK FOR CAPSTONE PROJECT PHASE 2

The Phase 2 of the Capstone project involves bringing work under way to apply and simplify the selected algorithms. In order to achieve the desired performance, a variety of execution and hand notes must be made. The convergence of the reduction algorithm and the influential discovery of the node must be completed and the result must be eventually released.

REFERENCES/BIBLIOGRAPHY

[1]“Uncloaking Terrorist Networks” by Valdis Krebs, 2012. https://journals.uic.edu/ojs/index.php/fm/article/view/941

[2]“Measuring Link Importance in Terrorist Networks” by Uffe Kock Wiil, Jolanta Gniadek, Nasrullah Memon, 2010.
https://www.researchgate.net/profile/Nasrullah-Memon/publication/221273632_Measuring_Link_Importance_in_Terrorist_Networks/links/00b7d526005eab9991000000/Measuring-Link-Importance-in-Terrorist-Networks.pdf

[3] Identifying the global terror hubs and vulnerable motifs using complex network dynamics By Syed Shariq Husain, Kiran Sharma, Vishwas Kukreti , Anirban Chakraborti Link, 2018.
https://www.researchgate.net/publication/322950246_Identifying_the_global_terror_hubs_and_v ulnerable_motifs_using_complex_network_dynamics

[4] Modelling the re-emergence of information diffusion in social network by Dingda Yang a,d , Xiangwen Liao b,d, *, Huawei Shen c,e , Xueqi Cheng c,e , Guolong Chen a, 2016.
http://www.bigdatalab.ac.cn/~shenhuawei/publications/2018/pa-yang-2.pdf

[5] Information Diffusion Models and Algorithms for Influence Maximization in Social Networks-A Survey by Lal Vikram Singh and S.Siva Sathya, 2014.
https://www.researchgate.net/publication/328837460_Information_Diffusion_Models_and_Algorithms_for_Influence_Maximization_in_Social_Networks-A_Survey

[6] Information Diffusion in Social Networks by Mohammed Zuhair Al-Taiecorresponding author and Seifedine Kadry, 2017.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7123536/

[7] Practical Issues and Algorithms for Analyzing Terrorist Networks by Tami Carpenter, George Karakostas, and David Shallcross, 2002

http://www.cas.mcmaster.ca/~gk/papers/wmc2002.pdf

[8] Network Topography, Key Players and Terrorist Networks By Sean F Everton, PhD is an Assistant Professor at the Naval Postgraduate School in Monterey, California, in the Defense Analysis Department, 2017.

https://my.nps.edu/documents/104382430/104582412/Everton+2012+(Topography).pdf/c6c7a7c1-3178-4bfb-aeaf-aa6120bdca2e

[9] Tractable Models for Information Diffusion in Social Networks by Masahiro Kimura and Kazumi Saito, 2006.
https://link.springer.com/chapter/10.1007/11871637_27

[10] Predicting Information Diffusion in Social Networks Using Content and User's Profiles by Cedric Lagnier, Ludovic Denoyer, Eric Gaussier, and Patrick Gallinari, 2013.
https://link.springer.com/chapter/10.1007/978-3-642-36973-5_7

APPENDIX A DEFINITIONS, ACRONYMS AND ABBREVIATIONS

Terror Hubs: Groups that use terror as a weapon to achieve its goals.

Vulnerable motifs: Easily noticeable targets which has a potential treat from terror attacks. Covet Network: Network members try and keep their identities secret.

GTD: Global Terrorism Dataset

IDA: Information Diffusion Algorithm

DFA: Disparity Filter Algorithm