|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PROJECT REQUIREMENTS SPECIFICATION  Community detection in dynamic networks  UE18CS390A – Project Phase – 1  ***Submitted by:***   |  |  | | --- | --- | | Name | SRN | | Mahammad Thufail | PES2201800646 | | Manne Vasanth | PES2201800425 | | Purushotham S | PES2201800480 | | Pulle Manikya Sri Manasa | PES2201800468 |   Under the guidance of   |  | | --- | | **Prof. Sreenath MV**  Assistant 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 |

TABLE OF CONTENTS

|  |  |
| --- | --- |
| 1. Introduction | 3 |
| 1.1 Project Scope and Motivation | 3 |
| 1. Literature Survey or Existing System | 3 |
| 1. Product Perspective | 3 |
| 3.1 Product Features | 3 |
| 3.2 User Classes and Characteristics | 3 |
| 3.3 Operating Environment | 3 |
| 3.4 General Constraints, Assumptions and Dependencies | 3 |
| 3.5 Risks | 4 |
| 1. Functional Requirements | 4 |
| 1. External Interface Requirements | 4 |
| 5.1 User Interfaces | 4 |
| 5.2 Hardware Requirements | 4 |
| 5.3 Software Requirements | 4 |
| 5.4 Communication Interfaces | 5 |
| 1. Non-Functional Requirements | 5 |
| 6.1 Performance Requirements | 5 |
| 6.2 Safety Requirements | 5 |
| 6.3 Security Requirements | 5 |
| 1. Other Requirements | 5 |
| Appendix A: Definitions, Acronyms and Abbreviations | 5 |
| Appendix B: References | 5 |

# Introduction

Terrorism is a type of collective violence having direct impact on peace, normal routine of a country/community and security and it is also a way to generate fear in civilians using violence.Terrorism is an evolving phenomenon, thus it is vital to provide counter-terrorism operators with tools for the prevention of it.

The main objective of this work is to define an approach aiming at eliciting knowledge on terrorist attack perpetrators by analyzing terror events along the timeline. The idea is to construct a sociogram, i.e., a network of perpetrators, where the nodes represent terrorist groups and the edges represent generic relations occurring between two groups.

Considering terrorist group networks along the timeline allows to study the temporal evolution of relations between such groups. With respect to the adopted data source, the aforementioned conceptualization is realized by considering the historical data provided by the Global Terrorism Database (GTD).

Community detection in a social network is identifying sets of nodes in such a way that the connections of nodes within a set are more than their connection to other network nodes.

# Project Scope

Our project contributes to this development by providing comprehensive collection of terrorist events including both domestic and international incidents. While these data have some well-known limitations, they also provide a wide variety of analysis opportunities.

Improve their quality and to expand the types of analysis being conducted with the data. As event databases improve, new avenues for validating them become feasible. An important method for assessing the quality of event databases on terrorism will be to do systematic comparisons between different sources of event data and between terrorism data drawn from other sources

Using internet as platform to analyze terrorist group networks : Online discussions provide an opportunity to present opposing viewpoints or to engage in constructive debate, which may have the effect of discouraging potential supporters. Successful messages may also demonstrate empathy with the underlying issues that contribute to radicalization, such as political and social conditions, and highlight alternatives to violent means of achieving the desired outcomes.

# Literature Survey or Existing System

**Understanding the composition and evolution of terrorist group networks: A rough set approach.**

**Authors** : Vincenzo Loia, Francesco Orciuoli

**Results and advantages** : The paper provides an original approach to elicit terrorist group networks from a database of terrorist events by using a similarity function based on rough set theory. The approach has been described and illustrated by providing several illustrative examples. Considerations on the time dimension have been also provided. Furthermore, the proposed approach has been demonstrated and evaluated by using a Python implementation of rough set operators (realized by the authors ) and comparing the results to the expert knowledge obtained by the resources provided by the GTD database (START project).

**Limitations** : the lack of a method to automatically deal with time intervals and the lack of a method to automatically analyze the temporal evolution of terrorist groups networks

**Finding influential nodes in social networks based on neighborhood correlation coefficient.**

**Authors** : Ahmad Zareie, Amir Sheikhahmadi, Mahdi Jalili, Mohammad Sajjad Khaksar Fasaei

**Results and advantages** : Social networks analysis and mining have recently gained ever-increasing importance with many potential applications in diverse industries. Influence maximization is one of the topics that has attracted much attention in this field. An important challenge in influence maximization is to find the most influential nodes based on their structural location in the network

**Limitations** : Need to process entire data to find the most influential node.

**Community detection in large‑scale social networks: state‑of‑the‑art and future directions**

**Authors** : Mehdi Azaouzi, Delel Rhouma, Lotf Ben Romdhane

**Results and advantages** : The main goal of this paper is to give a comprehensive survey of community detection algorithms in social graphs. Taxonomy of existing models based on the computational nature and thus in static and dynamic social networks. Comprehensive overview of existing applications of community detection. The refinement step converge very fast . Very good balance between structural and attribute similarities.Good quality of partition.

**Limitations** : Eliminates many interesting communities. The number of communities should be known in advance. Noise sensitivity.

**Hidden community detection in social networks**

**Authors** : Kun Hea, Yingru Li a, Sucheta Soundarajanc, John E. Hopcroft

**Results and advantages** : This paper introduces a new graph-theoretical concept of hidden community for analysing complex networks, which contain both stronger or dominant communities and weak communities.. We propose a meta approach, namely HICODE (Hidden Community Detection), for identifying the hidden community structure as well as enhancing the detection of the dominant community structure.

Due to the difficulty of labeling all ground truth communities in real-world datasets, HICODE provides a promising technique to pinpoint the existing latent communities and uncover communities for which there is no ground truth. Our finding in this work is significant to detect hidden communities in complex social networks.

**Limitations** : Through experiments on a variety of real-world networks, we demonstrate that the higher the hiddenness value a community is, the harder for an algorithm to locate such community; HICODE outperforms several state-of-the-art community detection methods on uncovering the hidden communities.

**TI-SC: top-k infuential nodes selection based on community detection and scoring criteria in social networks**

**Authors** : Hamid Ahmadi Beni, Asgarali Bouyer

**Results and advantages** : The TI-SC algorithm selects the infuential nodes by examining the relationships between the core nodes and the scoring ability of other nodes. After selecting each seed node, the scores are updated to reduce the overlap in selecting the seed nodes. This algorithm has efcient performance in high Rich-Club networks.

In this paper, They have proposed an efcient community-based algorithm combined with a scoring measure for selecting top-K infuential nodes. In the TI-SC algorithm, the scoring criterion reduces the overlap of seed nodes, and this leads to the selection of K-node with optimal infuence spread.

**Limitations** : Infuence maximization is a classic optimization problem to fnd a subset of seed nodes in a social network that has a maximum infuence with respect to a propagation model. This problem sufers from the overlap of seed nodes and the lack of optimal selection of seed nodes.

# Product Perspective

We lead and integrate the national counterterrorism (CT) effort by fusing foreign and domestic CT information, providing terrorism analysis, sharing information with partners across the CT enterprise, and driving whole-of-government action to secure our national CT objectives.

# Product Features

Elicit terrorist group’s networks: Selecting relevant features, building rough conceptualizations of terrorist groups, building boundary regions for terrorist groups and designing terrorist group networks.

Finding most influential nodes: The prediction of terrorist network and identifying main actors is an important issue for intelligence and security informatics. We propose a method to analyze social network using machine learning techniques. The proposed technique uses k-core concept to remove unwanted and passive nodes from the whole network. It then extracts multiple features and uses hybrid classifier to identify main actors. The proposed technique is tested on a publicly available dataset and results show significance of proposed system.

Analysis of temporal evolution of terrorist groups supporting counter-terrorism : Temporal evolutions of relations are due to possible changes in groups’ behaviors along the timeline. In fact, groups maintaining their behaviors tend to maintain also their relational states, groups that change their behaviors tend to modify the relations they are involved in. Lastly, it is possible that groups that change their behaviors maintain the relations among them because changing their behavior accordingly.

Communities are seen as groups, clusters, coherent subgroups, or modules in different fields; community detection in a social network is identifying sets of nodes in such a way that the connections of nodes within a set are more than their connection to other network nodes.

# User Classes and Characteristics

The proposed approach which elicit terrorist group’s networks, finds most influential nodes and temporal evolution of terrorist networks is not an open-source model instead these data are shared with national intelligent management for protecting nation from terrorist threats and implement counter terrorism activity . This has been implemented under the guidance of college professors.

Admin can modify the dataset and improvise the algorithms and users are data analysts, National Intelligent Management and Government

# Operating Environment

Since it is a Social Network Analysis project it can run on any modern CPUs

which has at least 4 cores, 4 threads and 3.5GHz Clock speed.

# General Constraints, Assumptions and Dependencies

* Regulatory policies

We use the Global Terrorism Database (GTD) provided by START project that can be analyzed in order to provide, for instance, prediction models. The main idea underlying this work is using the historical data provided by GTD, which offers information related to terrorist attacks perpetrated since 1970, in order to conceptualize the behaviors of terrorist groups in specific time intervals .

* Hardware limitations.

Analyzing the dataset and applying the algorithm we implement takes time depending upon the hardware we use. Task completes faster when we use any modern CPU.

Minimum storage required to store the dataset.

* Safety and security consideration

GTD analysis and implementation of algorithms should not be misused since it is issued for educational and Development (couter terrorism ) purposes only. This processed data is allowed only for national intelligent management for protecting nation from terrorist threats and implement counter terrorism activity .

* Assumption: Terrorism cannot be defeated

Terrorism has arguably been one of the defining factors of our age. It frequently makes headlines, threatening or attacking governments, private business and ordinary citizens. And in many parts of the world, it has been one of the most important threats to peace, security and stability.

# Risks

Data leakage which leads to situations in which sensitive or otherwise confidential data escapes organizational infrastructures, making that data vulnerable to potential unauthorized disclosure or malicious use. Mitigating the risks of handling such data and leakage can be an expensive undertaking .

# Functional Requirements

Data is collected from Global Terrorism Database (GTD) and necessary data Pre-processing steps are done. Approximate conceptualizations of terrorist group’s behaviors through Rough Sets. Later on selecting relevant features, building rough conceptualizations of terrorist groups, building boundary regions for terrorist groups and designing terrorist group networks.

Further finding influential nodes in social networks based on neighborhood correlation coefficient . The proposed method is based on local clustering coefficient and uses similarity of connections between neighboring nodes. The method is based on k-shell decomposition approach where influentiality of a node depends on how the node has shared connections with its neighbors.

To categorize and compare the theoretic community detection algorithms. To bring together techniques related to characterize, identify, and extract communities with two strategies: centralized and distributed. To analyze the strategy of detection according to the dynamic of networks: static or evolves over time. To provide the metrics for detecting partition: the structural-based partition or the semantic and structural based partition.

# External Interface Requirements

# User Interfaces

Project UI (web page) consist of the option such as checkout (gather the information provided in the text field) display (display the visual output of the network) etc. on the top of the web page. In the project we build web page that display the result of our project. Basically, the user interface is basic HTML page that enable the user to interact with result of our project.  Intermediate part of the webpage is the working area where user can enter the details and see the result obtained.  Project going to use the python and json to interact with the server.  If the user provides the invalid input or input format, error message will pop out in the webpage.

# Hardware Requirements

Any Intel(7 th gen or higher) or AMD(2 nd gen or higher) processors with at

least 3.5 GHz base lock.

All the result obtained from the project is displayed on the screen. In order to obtain the result from the server we use the TCP protocol. All the XML request and response between the client and the server is uses the TCP protocol.

# Software Requirements

Python Version : 3.7 or higher

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

Tools and Libraries (Open-source):

igraph – The network analysis software. igraph is a collection of network analysis tools with the emphasis on efficiency, portability and ease of use.

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

# Communication Interfaces

We use the TCP protocol to obtain the result from the server and all the XML request

and response between the client and the server. The line speed should be at least 10 kbps to load few of image format outputs. Predefined functions in the application will handle the entire buffer size of the network .

# Non-Functional Requirements

# Performance Requirement

Efficiency:

Network efficiency can be measured by considering the number of nodes that can instantly access a large number of different nodes – sources of knowledge, status, etc., through a relatively small number of ties. These nodes are treated as nonredundant contacts.

Effectiveness:

Effectiveness targets the cluster of nodes that can be reached through non-redundant contacts. In contrast, efficiency aims at the reduction of the time and energy spent on redundant contacts. Each cluster of contacts is an independent source of information. One cluster around this non-redundant node, no matter how numerous its members are, is only one source of information, because people connected to one another tend to know about the same things at about the same time.

# Safety Requirements

We in the safety profession have had to re-think our positions relative to how to provide a secure and safe working environment. We've had to look at engineering safeguards, procedural methodologies and structural barriers, as well as ways to minimize the threats and even weapons of mass destruction.

If there is extensive damage to a wide portion of the database due to catastrophic

failure, such as a disk crash, the recovery method restores a past copy of the

database that was backed up to archival storage (typically tape) and reconstructs a

more current state by reapplying or redoing the operations of committed

transactions from the backed-up log, up to the time of failure.

# Security Requirements

Since the model depends on the data for learning purposes, hence it should

not be corrupted or poisoned, otherwise it could bring the system down to its knees. Security systems need database storage too.

# Other Requirements

Scalability : We believe that the extent of this study is not restricted to Global Terrorism Database (GTD), but generalizable to other social media platforms

Maintainability : The system should be developed in such a way that it is extensible. It should be easy to incorporate new feature requirements or accommodate a change in the existing requirements.

# Appendix A: Definitions, Acronyms and Abbreviations

GTD – Global Terrorism Database

# Appendix B: References

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

Authors : Vincenzo Loia, Francesco Orciuoli

<https://www.sciencedirect.com/science/article/pii/>[S0167739X19307757](https://www.sciencedirect.com/science/article/pii/S0167739X19307757)

1. Finding influential nodes in social networks based on neighborhood correlation coefficient.

Authors : Ahmad Zareie, Amir Sheikhahmadi, Mahdi Jalili, Mohammad Sajjad Khaksar Fasaei

<https://www.sciencedirect.com/science/article/pii/>[S0950705120300630](https://www.sciencedirect.com/science/article/pii/S0950705120300630)

1. Community detection in large‑scale social networks: state‑of‑the‑art and future directions

Authors : Mehdi Azaouzi, Delel Rhouma, Lotf Ben Romdhane

<https://www.sciencedirect.com/science/article/pii/>[S0020025517310101](https://www.sciencedirect.com/science/article/pii/S0020025517310101)

1. Hidden community detection in social networks

Authors : Kun Hea, Yingru Li a, Sucheta Soundarajanc, John E. Hopcroft

<https://link.springer.com/article/10.1007/>[s12652-020-01760-2](https://link.springer.com/article/10.1007/s12652-020-01760-2)

1. TI-SC: top-k infuential nodes selection based on community detection and scoring criteria in social networks

Authors : Hamid Ahmadi Beni, Asgarali Bouyer

<https://www.researchgate.net/publication/333197917>

1. Detection of Influential Nodes Using Social Networks Analysis Based On Network Metrics

Authors : Aftab Farooq,Muhammad Uzair,Gulraiz Javaid Joyia,Usman Akram

<https://ieeexplore.ieee.org/abstract/document/8346372>

1. Analysis of the Dynamic Influence of Social Network Nodes

Authors : Hong-Jian Yin, Hai Yu, Yu-Li Zhao, Zhi-Liang Zhu, Wei Zhang

https://www.hindawi.com/journals/sp/2017/5046905/