|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| HIGH LEVEL DESIGN DOCUMENT  Community detection in dynamic networks.  UE18CS390A – Capstone Project Phase – 1  ***Submitted by:***   |  |  | | --- | --- | | Name | SRN | | Mahammad Thufail | PES2201800646 | | Purushotham S | PES2201800480 | | Manne Vasanth | PES2201800425 | | Pulle Manikya Sri Manasa | PES2201800468 |   Under the guidance of   |  | | --- | | **Prof. Sreenath MV**  Assistant Professor  PES University |   **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 | 4 |
| 1. Current System | 4 |
| 1. Design Details | 7 |
| 3.1 Novelty | 7 |
| 3.2 Innovativeness | 7 |
| 3.3 Performance | 7 |
| 3.4 Reliability | 7 |
| 3.5 Maintainability | 7 |
| 3.6 Legacy to Modernization | 7 |
| 3.7 Application Compatibility | 7 |
| 3.8 Resource Utilization | 7 |
|  |  |
|  |  |
|  |  |
|  |  |
| Appendix A: Definitions, Acronyms and Abbreviations | 8 |
| Appendix B: References | 8 |
| Appendix C: Record of Change History | 8 |
| Appendix D: Traceability Matrix | 8 |

# 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.

We used an approach that will allow us to find clusters of similar terror groups using information on their operational characteristics. Specifically, using open access data of terrorist attacks occurred worldwide since 1970, we build network that includes terrorist groups and related information on tactics, weapons, targets, active regions.

We model this data with each partition joined to the terrorist groups. Later on we will find the most influential group with maximum number of relations between other networks and try to prevent the attacks. Community detection to identify sets of nodes in such a way that the connections of nodes within a set are more than their connection to other network nodes.

# Current System

The current basic system, extracts the input data from the Global Terrorism Dataset (GTD) according to requirements provided by the user. Later on preprocessing of data has to be done before analysis. After the preprocessing, visualization of the data using various python tools such as NetworkX ,Igraph etc. Finally it implements the algorithm on the data to extract the composition and evolution of terrorist group networks, influential nodes and detecting communities.

The existing system contains drawbacks in their respective algorithms , to overcome this we implement most efficient algorithms

* Understanding the composition and evolution of terrorist group networks: A rough set approach.
* Finding influential nodes in social networks based on neighborhood correlation coefficient.
* Community Detection : Dynamic Louvain using local modularity

1. **Design Details**
   1. **Novelty**

The prediction of terrorist network and identifying main actors is an important issue for intelligence and security informatics, and few of the researches lack in this aspect. So we propose a method to analyze social network using some machine learning techniques (ie.,K-Core Concepts). Once the networks are clustred accordingly then we then use methodology based on community detection to determine the relationships of the terrorist node within the same cluster as well as the terrorist nodes from different clusters.

* 1. **Innovativeness**

The approach that we proposed elicits terrorist group’s networks, finds most influential nodes and temporal evolution of terrorist networks, but it 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.

* 1. **Performance**

Performance is also the main aspect of this approach, because the approach that we proposed has a better efficiency since we measure it by considering the number of nodes that can instantly access a large number of different nodes through a relatively small number of ties. The nodes are treated as nonredundant contacts.

In some of the approaches the effectiveness targets the cluster of nodes that can be reached through non-redundant contacts. But in our case , 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.

* 1. **Reliability:**

The algorithm we use is capable of performing operations using the data got from the terrorist database and is producing results in a time which we could think of other tasks to perform. It is reliable for the kind of data we use and the efficient results produced at the end.

* 1. **Maintainability:**

We need to use a good ranking methods and algorithms to make sure it is showing right results to the users. We would also assign weights to the nodes so that there will be change in importance of identifying the community accordingly as the time goes on. These ranking methods have to be tested and updated time to time to ensure good working of the tool. The retrieval of results from the underlying search engines is achieved using their respective API’s which are free of cost. Maintenance would be required if their respective policies changes.

* 1. **Legacy to modernization:**

To gain operational efficiencies as per the legacy modernization we are updating and optimizing business systems by enabling graphical access to government agencies who could actually view each community growth and the influential node in a single display graphical view. This makes user achieve their expectations in their experience and getting easily adopted to other newer technology platforms.

* 1. **Application compatibility:**

As our project is capable of operating in different operating systems and its environment is also user-friendly, it allow us to automate the process of testing where all the application is tested for their compatibility at once. Enabling auto removing of the features which is not supported by the operating system is achieved to a certain extent.

* 1. **Resource Utilization:**

There is huge data to be processed and most of it are not needed for the kind of results we expect. So we kind of use only the results which has huge impact on terrorist growth and which also needed for the future link prediction among the groups. The community detection almost enables us to know the relation between different communities and their growth over time. So the data is utilized as per our intermediate needs and predictions.

# Appendix A: Definitions, Acronyms and Abbreviations

GTD: Global Terrorism Dataset

|  |
| --- |
|  |

# Appendix B: References

# Nesrine Hafiene,Wafa Karoui1, and Lotfi Ben Romdhane “Influential Nodes Detection in Dynamic Social Networks”. Paper  2019

# <https://link.springer.com/chapter/10.1007%2F978-3-030-20482-2_6>

# Mohamed EL-Moussaoui, Tarik Agouti, Abdessadek Tikniouine, Mohamed Eladnani “Community detection: Approaches and applications”. Paper 2019

# <https://www.sciencedirect.com/science/article/pii/S1877050919305046>

# Punam Bedi and Chhavi Sharma “ Community detection in social networks ". Paper 2016

# <https://www.researchgate.net/publication/295395520_Community_detection_in_social_networks>

# Vinicius da Fonseca Vieira, Carolina Ribeiro Xavier, Nelson Francisco Favilla Ebecken and Alexandre Goncalves Evsukoff “Performance Evaluation of Modularity Based Community Detection Algorithms in Large Scale Networks”. Paper 2014

# <https://www.hindawi.com/journals/mpe/2014/502809/>

# Appendix C: Record of Change History

[This section describes the details of changes that have resulted in the current High-Level Design document.]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Date** | **Document Version No.** | **Change Description** | **Reason for Change** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# Appendix D: Traceability Matrix

[Demonstrate the forward and backward traceability of the system to the functional and non-functional requirements documented in the Requirements Document.]

|  |  |
| --- | --- |
| **Project Requirement Specification Reference Section No. and Name.** | **DESIGN / HLD Reference Section No. and Name.** |
|  |  |
|  |  |
|  |  |