VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belgavi-590018, Karnataka, INDIA

PROJECT REPORT

On

"Monitoring Individuals in Drug Trafficking Organization using Social Network Analysis"

Submitted in partial fulfillment of the requirements for the VIII Semester

Bachelor of Engineering

In

INFORMATION SCIENCE AND ENGINEERING

For the Academic year
2020-2021

BY

Rahul Bijoor	**1PE17IS066**
Roshan R Achar	**1PE17IS072**
Sanjana Shetty	**1PE17IS077**
Sumukh Balu	**1PE17IS092**

Under The Guidance Of

Prof. Sreenath M.V
Assistant Professor
Department of Information Science & Engineering, PESIT-BSC

Department of Information Science and Engineering
PESIT Bangalore South Campus
Hosur Road, Bengaluru-560100

PESIT Bangalore South Campus

Hosur Road, Bengaluru-560100

Department of Information Science and Engineering

CERTIFICATE

This is to certify that the project work entitled **"Monitoring Individuals in Drug Trafficking Organization using Social Network Analysis**" carried out by **Rahul Bijoor, Roshan R Achar, Sanjana Shetty, Sumukh Balu** bearing USNs **1PE17IS066, 1PE17IS072, 1PE17IS077 and 1PE17IS092**, in partial fulfillment for the award of Degree of **Bachelors of Engineering in Information Science and Engineering of Visvesvaraya Technological University, Belagavi** during the year 2020-2021. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library.

The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for said degree.

Signature of Guide	Signature of HOD	Signature of Principal
Prof. Sreenath M.V	**Dr. Annapurna D**	**Dr. Subhash Kulkarni**
Asst. Professor, ISE	HOD, ISE, PESIT-BSC	Principal, PESIT-BSC

External Viva

Name of the Examiners	Signature with date
1.	
2.	

DECLARATION

We, Rahul Bijoor (1PE17IS066), Roshan R Achar (1PE17IS072), Sanjana Shetty (1PE17IS077), and Sumukh Balu (1PE17IS092) hereby declare that the dissertation entitles, 'Monitoring Individuals in Drug Trafficing Organization using Social Network Analysis', is an original work done by us under the guidance of Prof. Sreenath M.V, Assistant Professor, Department of Information Science and Engineering, PESIT-BSC, Bengaluru, is being submitted in partial fulfillment for the award of B.E. in Information Science and Engineering, VTU of the requirements for 8th semester.

Date: 16/07/2021

Signature of the candidate

Place: Bangalore

Rahul Bijoor

Roshan R Achar

Sanjana Shetty

Sumukh Balu

PLAGIARISM REPORT

Final year plag report

ORIGINALITY REPORT

11%	8%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	suryodaybasak.info Internet Source	6%
2	Kaustav Basu, Arunabha Sen. "Monitoring individuals in drug trafficking organizations", Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2019 Publication	2%
3	mafiadoc.com Internet Source	1%
4	www.scribd.com Internet Source	<1%
5	Submitted to National University of Singapore Student Paper	<1%
6	Submitted to Lovely Professional University Student Paper	<1%
7	klff.prolocofumone.it Internet Source	<1%

TABLE OF CONTENT

TABLE OF FIGURES

CHAPTER 1

INTRODUCTION

The United Nation Office on Drugs and Crime (UNODC) released a shocking report on the usage of drugs in 2018 where it was brought to the world's notice that the worldwide drug use had gone up to 56% and predominantly in the age group of 15-64 years.

More than 276 million people were reported to have abused some or the other form of drugs like marijuana, cannabis or cocaine at least twice in the year 2016 alone. Out of this, 15 million individuals in between the tender age groups of 14-16 were responsible for this alleged drug abuse. 4 and a half lakh people have died from drug related activity out of which 1 lakh 68 thousand of them have died as a direct result of drug overdose

These staggering numbers are only indicative of the persistent and increasing demand for illegal drugs all over the world. Alarmingly the various Drug Trade Organizations have seemed to hear the call for more drugs and they are only increasing the production rates of these illicit substances such as cocaine, pot, marijuana, opium and LSD.

As a unified resistance to this mass production of drugs worldwide, the DEA (Drug Enforcement Agencies) spend to the tune of $50 billion dollars every year to monitor and track all the people concerned with any form of drug related offences, manufacturing or consuming. Upon thoroughly investigating these drug offences the DEA is left with a list of suspects and their respective drug offences. However, year after year the number of alleged suspects keep growing and so do their drug offences. Thus, keeping track of every single suspect and their drug offences becomes an exhaustive and next to impossible task, not to mention that the manpower usage would become insanely high and ridiculously expensive.

Despite these challenges, it's pleasing to note that authorities around the world have achieved massive success by successfully dismantling and busting several Drug Trade Organizations (DTO's) all over the world, by analyzing transcripts obtained after questioning these individuals and also making use of phone tap and wiretap data extensively. They have also

questioned key actors and other important and influential people to build a Social Network that shows the relationships between different people in the Drug Trade Organization.

In this project we propose a unique drug network organization monitoring DNM approach based on the mathematical and statistical concept of identifying codes that not only completely reduces resource requirements of the authorities but also provides the functionality of uniquely identifying an alleged suspect when the alleged suspect becomes completely or partially active in drug related activities.

Our approach to this project is that we assume that when a potential drug offender becomes engaged in/ active in any illegal drug activity he/she may not be under direct surveillance by cops but his friends/colleagues or partner who may be under surveillance will definitely have an idea of their work.

Then this individual involved in any nefarious activities can be uniquely identified easily using their connections to their associates. Interestingly enough we also show that monitoring people using standard centrality metrics in SNA also does not guarantee any unique MO of the individual and it also leads to a tremendous wastage of human resources on the part of the authorities.

1.1 PURPOSE

The purpose of our project is to reduce the manpower and resources currently being used by policing and drug-control organizations across the globe to curb drug production rates Law enforcement agencies have been known to spend up to the tune of 40 billion every year and this amount will only increase as the number of drug offences increase after some time keeping track of all the alleged suspects in the database may become next to impossible. Hence our project aims to resolve the above problem. Our project also helps to find the "kingpin" of the drug trade organization (DTO) by using various social network analysis (SNA) centrality metrics.By identifying the "kingpin" of the Drug Trade Organization, Law Enforcement will be able to bring the kingpin down. As a result, the entire drug network gets completely destabilised. This leads to power fights within the Drug Trade Organization, which in turn leads to creation of multiple factions and divisions within the DTO, thus the entire DTO will be temporarily weakened. Law Enforcement can now use other methods to bring down the entire drug supply chain once and for all. An additional side benefit to our project is that law enforcement will also be able to track the drug users as well and once their current supplier is no longer supplying them with drugs, addicts are bound to turn elsewhere or try to find some other drug supplier for their fixes. Police will now be able to crack down on major drug syndicates in this way. Drug users will also be tracked and admitted to rehabilitation centers where they can recuperate and overcome their addictions. Thus, Our Project not only helps law enforcement bring down these anti-social drug networks but also helps the victims/drug users recover.

1.2 SCOPE

Our project is applicable to nearly all organizations that may need it. Our project is more like a plug-and-play feature to an already existing drug-monitoring software. All the user needs to do is plug in a database of call records of all the criminals/suspects local to their police regions the algorithm takes care of the remaining things and outputs out the kingpin of the DTO and even the suspects closest to the kingpin so that we can effectively bring down the drug trade organization. Our project also outputs an interactive visualization of the entire Drug Trade Organization which helps us to get an overall bigger picture of all the people involved and how they are connected to each other.

1.3 DEFINITIONS, ACRONYMS AND ABBREVIATIONS

1.3.1 Definitions

Drug Trafficking Organization: "complex organizations with highly defined command-and-control structures that produce, transport, and/or distribute large quantity "

Social Network Analysis: "investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes and the ties, edges, or links that connect them".

Graph Colouring Algorithm: A graph coloring is an assignment of labels, called colors, to the vertices of a graph such that no two adjacent vertices share the same color. Vertex coloring is the most common graph coloring problem. The problem is, given m colors, find a way of coloring the vertices of a graph such that no two adjacent vertices are colored using the same color. Chromatic Number: The smallest number of colors needed to color a graph G is called its chromatic number.

Chromatic Number: When the number required to colour the graph cannot be reduced anymore it is said to be the chromatic number of the graph

1.4 LITERATURE SURVEY

For a very long time, a lot of notable research was held on Drug Trade Organisations and this was conducted using a very unique concept called Social Network Analysis. has been conducted utilizing Social Network Analysis (SNA). One of the people who worked in this called Natarajan took a lot of efforts to analyse the wiretapped data and create the hierarchy structure and try to understand the individual roles. She had analyzed over 2000 wiretap conversations by performing Social Network Analysis on the call records, and was able to successfully reveal a huge and not so structured group of 294 individuals .

A drug trafficking network which is based in Australia, was analysed by Bright and Delaney who used the concept of Social Network Analysis to study the network.They inferred that the different roles of different individuals were associated with their respective centrality scores. Bright et. al. , to identify key actors in drug trafficking networks, tried to use individual attributes which are added with the centrality measures. Bright et. al. observed the judges decision and comments and with the help of that he created a network of individuals who were potentially involved in the distribution of methamphetamine in Australia in the 1990s.

Carley , in her studies, found out that Social Network Analysis has a lot of potential when it comes to analysing a network. In Fact it can be used as an amazing tool to destabilize a network.

Hughes took his time to analyse the transcripts of court to identify the key actors. Then a social network is formed around these key actors. He wanted to explore how the products were diversified in the three major drug organisations in Australia..

In the last few years, Social Network Analysis gained importance. That's why a lot of research was done on it, mainly focusing on Identifying codes and how to apply this concept to networks.

The concept of identifying Codes was actually introduced by Karpovsky et. al. . It can provide almost accurate results for various types of graphs which differ in topologies, including binary trees and normal trees. Another Scientist Laifenfeld et. al. used Identifying codes to study Joint monitoring and routing in sensor networks which are wireless.

The issues that are related to computing all these minimum Identifying Codes for the trees or graphs was studied by Charon et. al. She concluded that in most of the graphs, the problem is NP-hard.

A minimal algorithm which can be used to compute the Identifying codes was proposed by Ray et. To identify the key actors in any network, the social network analysis usually utilises the centrality measures.

1.5 EXISTING SYSTEMS

Our project is the first of it's kind Drug Monitoring software using Social Network Analysis to bring out interactive visualizations and identifying the sole decision makers/kingpin of the drug trade organizations while at the same time showing a more efficient approach to do the same using far lesser resources thus saving costs for the law enforcement authorities. Our project uses a unique Graph Colouring Algorithm to save resources for the police.

The closest product in the market purely in terms of Drug Monitoring would be the Prescriptions Drug Monitoring Associations (PsDMA's). These PsDMA's are state level interventions i.e the state takes these preventive measures for making sure that opioid prescription is regulated and that it doesn't go out of control. That however, has no bearing or relevance to what we are aiming to accomplish in our project.

1.6 PROPOSED SYSTEM

In this project we used datasets from kaggle.com and from the references in the paper we used to implement the entire project. The Datasets themselves are the call records of all the suspected individuals that are allegedly involved somehow in the Drug Trade Organization either as a consumer or as a drug supplier, from this dataset we obtain adjacency matrices that we then use as input for our graph colouring algorithm.

The graph colouring algorithm then outputs a list of nodes i.e in a interactive manner it shows the way the Drug Trade

Organization looks after applying the GCA, it shows us a pictorial representation in the form of bar graphs and other data visualization tools. It also tells us which node is most likely to be the Kingpin of the Drug Trade Organization. It shows in percentage format; the amount of resources we are saving using our approach.

1.7 PROBLEM STATEMENT

Design a software that makes use of Social Network Analysis and Graph colouring Algorithm to accurately paint a wholesome picture of the Drug Trade Organization in a region . The model should be accurate, reliable and cost effective. Doing this using SNA centrality metrics should also help law enforcement authorities to reduce the amount of resources that they deploy to crack down on major Drug Syndicates.The tool should also be user friendly so that law enforcement agents can easily use it without having to refer to any technical manual or take any external help whatsoever .

1.8 SUMMARY

This chapter depicts the drug problem prevalent in many of the developed and the developing countries of the world. Related works and other existing systems have also been discussed in brief. The idea of how computer science and in particular Social Network Analysis can be used in a totally different field such as Drug Monitoring is also given as a hint. The chapter serves to shed light into the model which we developed this, will be discussed in detail in the future chapters.

CHAPTER 2

SOFTWARE REQUIREMENTS SPECIFICATIONS

Software requirements specifications are the detailed enlisting of all necessary requirements that arise in the project. The aim of having these requirements is to gain an idea of how the project is to be implemented and what is to be expected as a result of the project. The sections in this chapter deal with the various kinds of software, hardware and other functional and non-functional requirements of the project. A brief description of the various users of the system is also mentioned.

2.1 OPERATING ENVIRONMENT

This section gives a brief overview of the hardware and software prerequisites for the project.

2.1.1 Hardware Requirements

- 1.69GHz or faster processor
- 2 GB(32 bit) or 4GB(64 bit) RAM
- 200GB of available hard disk space

2.1.2 Software Requirements

- Operating system: Windows 10/11
- Programming language: Python
- IDE: Jupyter
  Documentation: MS Word

2.2 FUNCTIONAL REQUIREMENTS

Functional requirements are a way of representing the compulsory services of a project. We have identified the functional requirements for our project as follows:

- The input must be datasets that are call records, preferably in the form of adjacency
- matrices.
- The output must show the law enforcement the minimum number of nodes that must be monitored and also expose the node that is most likely to be the kingpin of the drug trade organization.
- Interactive Visualization have to be provided for the same, in the form of bar-graphs and other data visualization tools

2.3 NON-FUNCTIONAL REQUIREMENTS

These are the various capabilities offered by the system. These have nothing to do with the expected results, but focus on how well the results are achieved.

- **ACCURACY:** One of the major challenges to any project is to deliver exact results on every run, especially in the case of health care projects which can't afford erroneous results.
- **ROBUSTNESS:** The system should be accurate for a wide range of inputs for patient details. This helps build confidence among users.
- **MAINTAINABILITY:** The code should be split into smaller modules each targeting a different section of the project such as prediction, classification etc so as to easily maintain it even after the completion of the project.
- **USABILITY:** The graphical user interface is to be designed such that it is easy for a naïve user such as a patient as well. Text boxes and buttons should be unambiguous.

- **RESPONSE TIME:** The results should be provided within a very short interval of time. The response time should not exceed a few seconds to get results and recommendations.
- **RELIABILITY:** Dataset must be reliable and accurate in order to avoid errors and/or inaccuracies in the output

2.4 USER TYPES

There are of two types of users associated with the system:

- **Developers:** These are the people/designers of the system who are in charge of writing code and actually building the system from the ground up.
- **Law enforcement agents:** These are the police or the government agents in charge of tracking and monitoring the drug offenders. These are the primary users of our system.
-
- **Doctors/Rehabilitation Experts:** Doctors can also be potential users of our system, they can find drug consumers/addicts and attempt to rehabilitate them so they can get the adequate care and treatment that they need to overcome their debilitating addictions.

2.5 APPLICATIONS OF SYSTEM

- Understanding how to cut costs and reduce expenditures while tracking suspects.
- Understanding what drug networks look like and how to take them down effectively and permanently.
- Using Social Network Analysis and data visualization tools how we can find the mastermind/Kingpin of the Drug Trade Organization.

2.6 ADVANTAGES OF SYSTEM

- The system is reliable and accurate.
- The system provides a user-friendly GUI.
- The system has an enormous untapped potential to save costs for law enforcement agencies.
- Not only cracks down on illegal drug use but also helps rehabilitate and stabilize the lives of drug addicts .

2.7 SUMMARY

This chapter discussed both the software and hardware requirements needed for the smooth running of our project. We then dive into the functional and non functional requirements which give an idea of what to expect from the final project once it has been developed completely. The users, advantages and applications of the system are also pointed out quite succinctly in this chapter.

CHAPTER 3

HIGH LEVEL DESIGN

Starting with the design is the first step where we get to know the overview of the system. In this step we find out the different elements, the level of abstraction and how the data flows between the levels. Designs help the developers understand how the subsystems work together. The Design process is not completed in just one step, it is modified, corrected, with each step. Backtracking process is done to go through any defects in design.

3.1 DESIGN CONSIDERATIONS

Before starting with the design of the system there are various aspects to be considered. Considering these aspects at the earliest helps the developers in completing the project. Each design component should contribute in bringing the system to the mentioned requirements.

3.1.1 Assumptions

The first assumption is that the end user of the system that is the investigation department has a clear understanding of how our system works. The department should be able to use the system to get the desired result out of the input that they provide to our system. The department should be also able to understand the graph properly to identify the criminal. The success of the application is dependent on the dataset that we provide to the system. Keeping the performance factor of the system in mind we should make sure that system is usable by the end user.

3.1.2 Goals and Constraints

- The system should be simple to understand by the investigation team for them to obtain the result out of it.
- The system should fetch the details of suspects based on the dataset we provide and predict the outcomes accordingly.
- The result obtained should provide some value to the investigation department.

- The constraints are:

    - Time is one such constraint to be considered in training the data.
    - The Datasets must be clear and concise and without a lot of errors.

3.2 SYSTEM ARCHITECTURE

System architecture is a way of analyzing how the system should work and it gives a basic layout of the system. It distinguishes various components of the system and establishes the relationship between these components. The relationship is defined in terms of the data flow between the subsystem or logic of the system. Each component interacts differently to complete the overall requirements of the system.

All the data that is required by the system is provided as a dataset to the system. The system performs the data checks on the dataset to find the missing and NULL values in the dataset. These data are used by the system to apply the algorithm to find the output from the dataset. We look into the detailed design in this section.

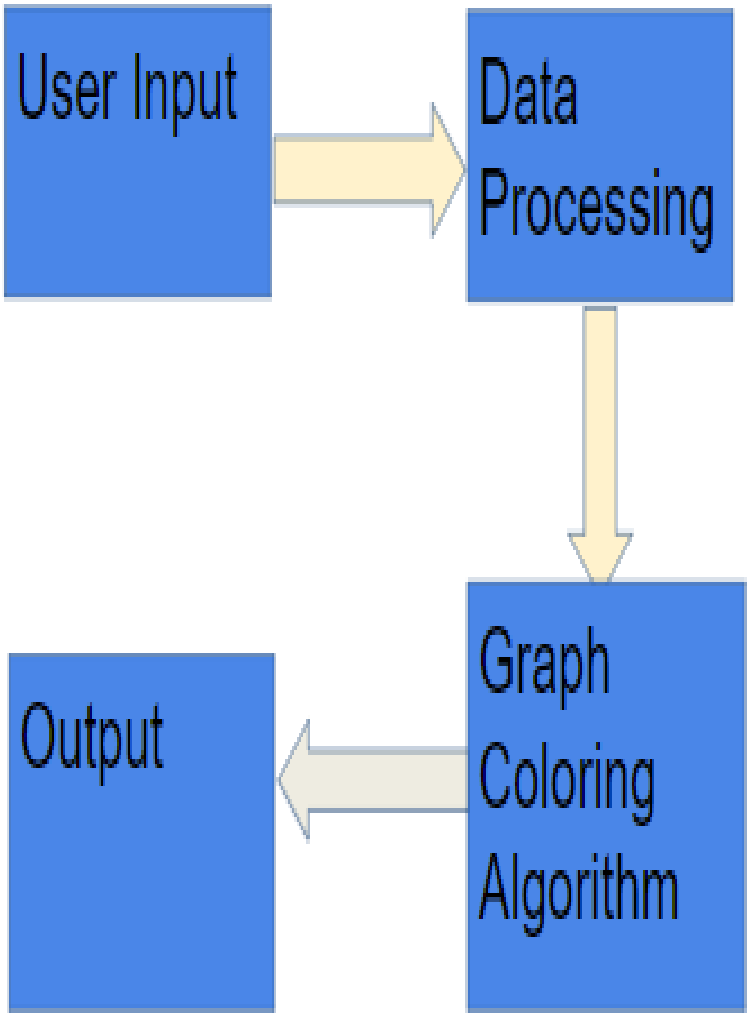


Figure 3.2.1 System Architecture Diagram

3.3 SEQUENCE DIAGRAM

A sequence diagram shows the interactions of the objects and how their processes operate in time. The various processes are depicted in the order of execution. The objects exchange messages and data and parameters to carry out their specific functionalities. The thing that our system gets trained is with the dataset that we provided.

The dataset gets trained to show the result. The results of the algorithm are to see who the suspect is.

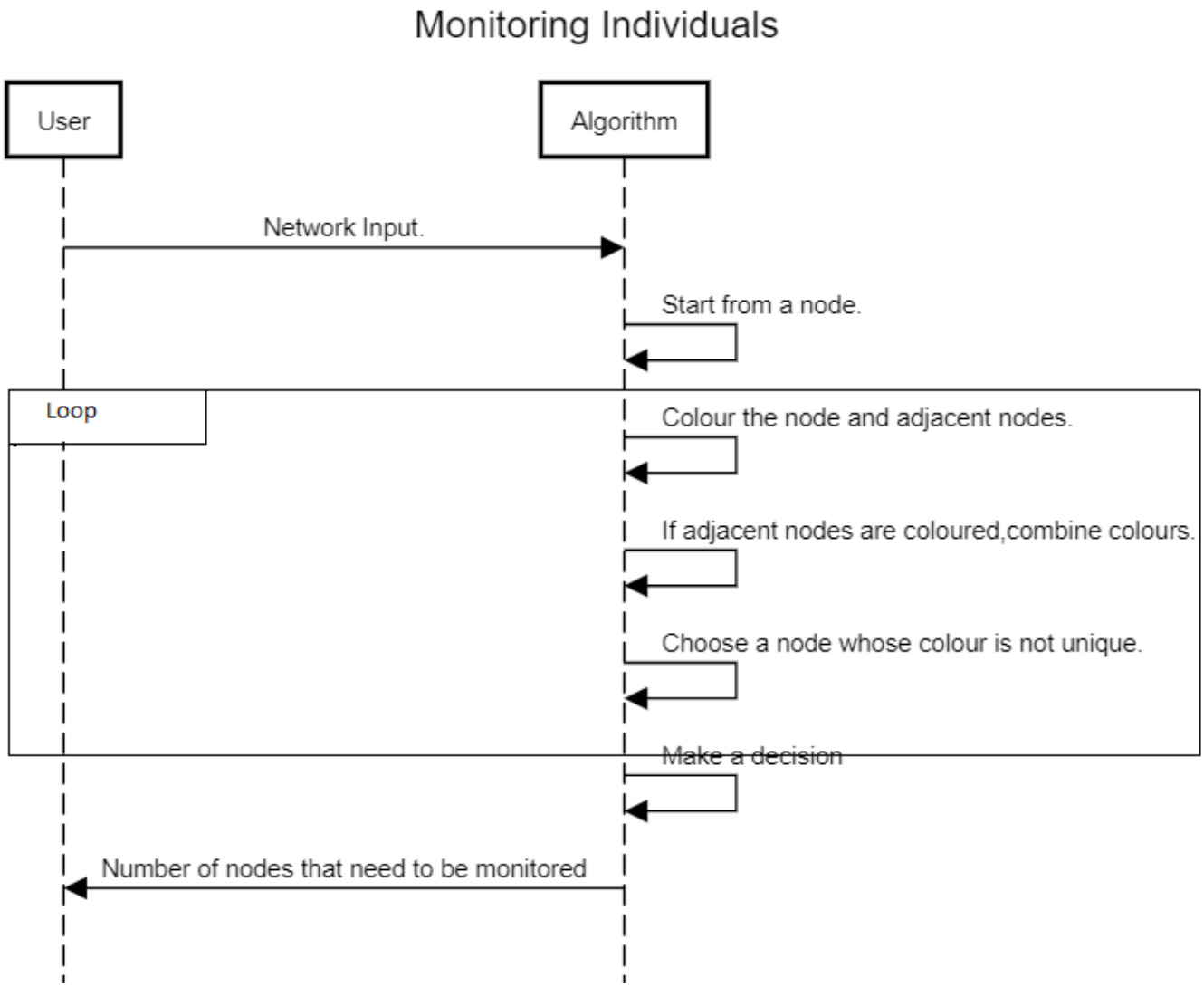


Figure 3.5.1 Sequence Diagram

3.4 USE CASE DIAGRAM

Use case diagram shows the interaction of the users with the system. The Use case diagram for this system is followed by its explanation. The steps involved in the use case diagram is explained as follows:

- At first we get the names and contacts of the individuals working in the DTO. We make a dataset out of the name and contacts we find from the DTO
- Next we monitor the call records of the individuals of the DTO. We use this dataset for our system for the algorithm to work.
- We see if any of the call records is related to the individuals in the DTO.
- We use the graph coloring Algorithm on the dataset that we obtained from the previous steps to get the output for it.

If we find the suspect we report it to the respective department. If not we report that the suspect doesn't belong to the current dataset.

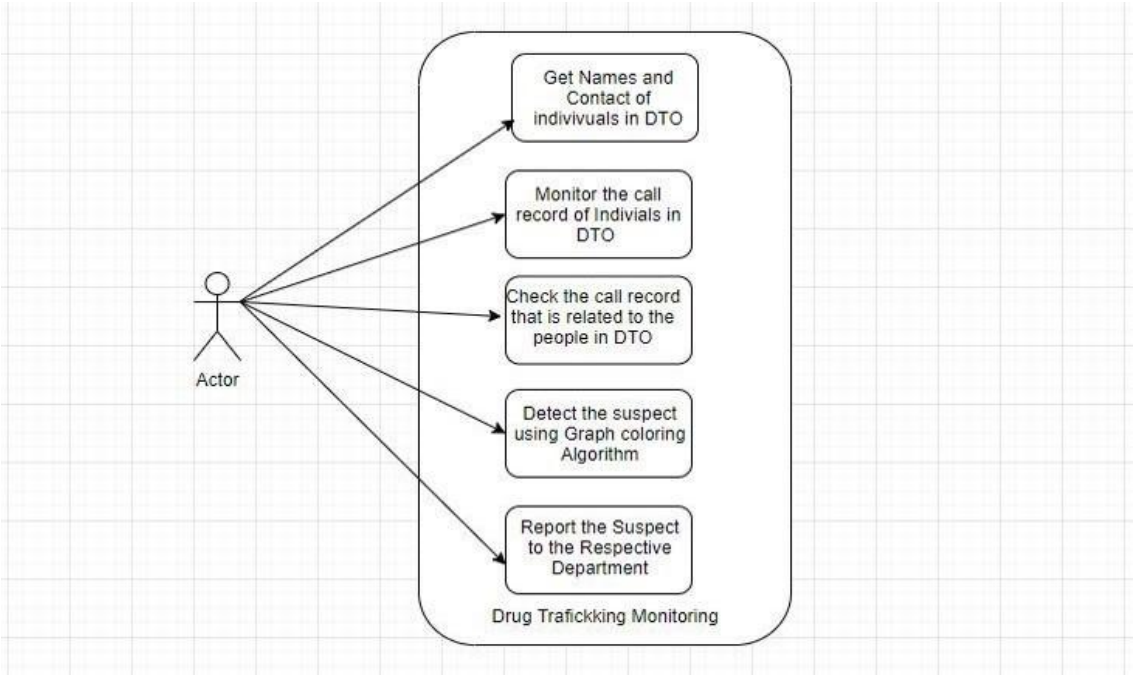


Figure 3.6.1 Use Case Diagram

3.5 SUMMARY

This chapter explains the design consideration, limitations and constraints. We presented the whole system as system architecture. The use case diagram depicts the interaction between the user and the system. The sequence diagram shows the sequence in which the steps take place to meet the requirements of the project. The data flow diagram shows how the data is processed in the system.

CHAPTER 4

DETAILED DESIGN

4.1 PURPOSE

The purpose of this chapter is to discuss the detailed design of the system about how the interaction of the system takes place.

4.2 COMPLETE SYSTEM DESIGN

We have split the detailed design flowchart for a clear understanding of the flow of data across the subsystems. We explain the design in detail followed by the flowchart of the design.

4.3 MODULE 1: BUILDING THE DATASET FROM THE DATA

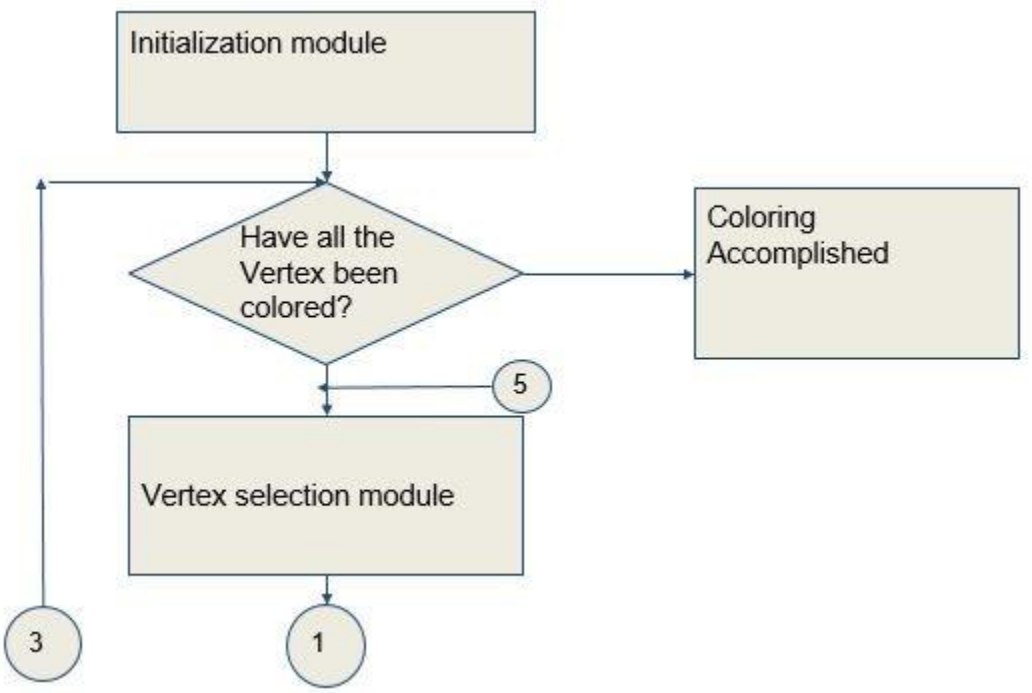
Each and every component of the project is equally important in building the projects that meet the end user requirement. Building the dataset for the project is the first step we consider in our project. We take the information of the user in the DTO and their call records.

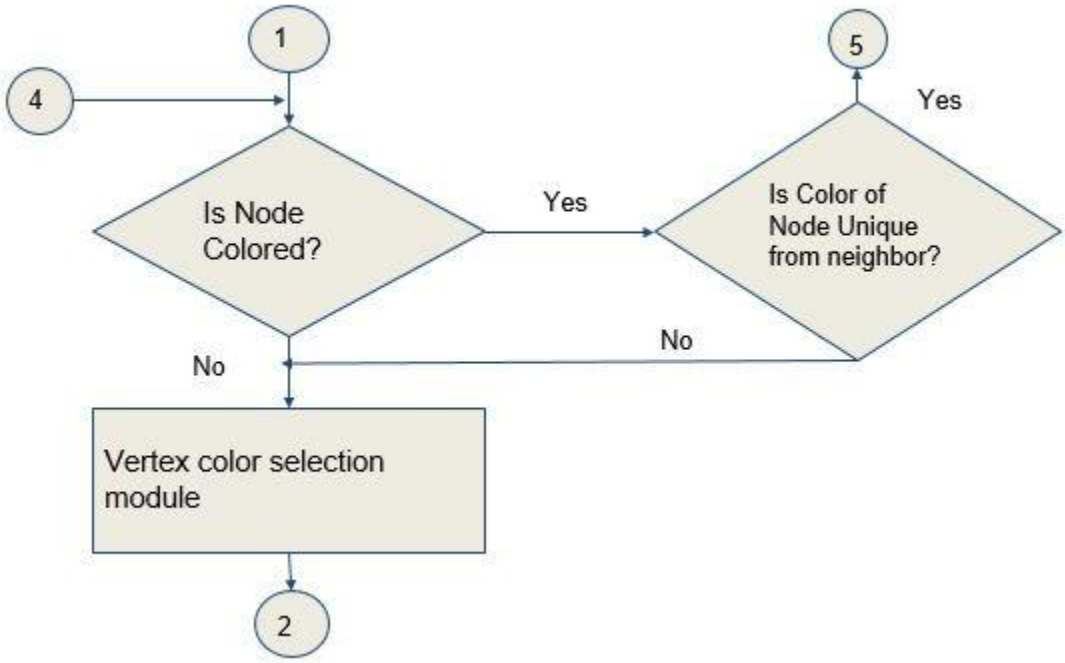
The data we used was structured and typed into CSV format. Initially the dataset contains missing data and error data. We go through the dataset, add those missing values and correct the error data. After this step we are having the dataset of the user and their call record.

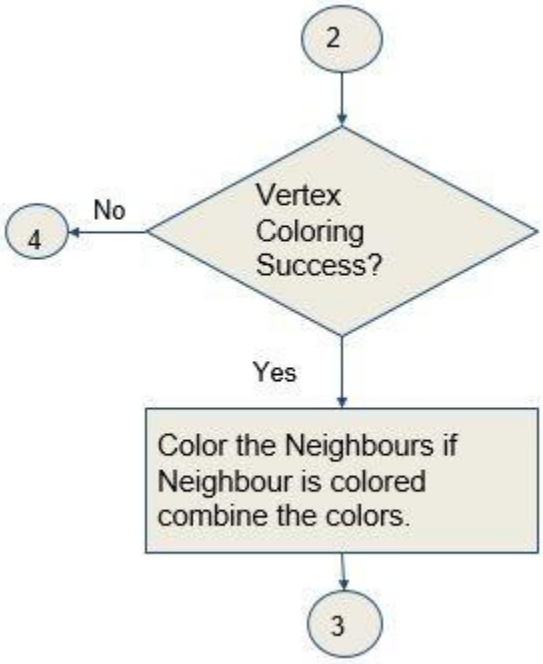
4.4 MODULE 2: USING THE DATASET AND APPLYING IT TO OUR ALGORITHM

- Now that we have our dataset we are using the dataset to apply it into our system. We apply the graph coloring algorithm on our dataset.

- Graph coloring algorithm is one of the machine learning algorithms. Graph coloring is the procedure of assignments of colors to the vertices. The main objective is to minimize the number of colors while coloring the graph. Here we apply the algorithm to find the suspect who is linked with the DTO
- The detailed design is as follows:




1
4
Is Node Colored?
Yes
5
Yes
Is Color of Node Unique from neighbor?
No
No
Vertex color selection module
2


2
Vertex Coloring Success?
No
4
Yes
Color the Neighbours if Neighbour is colored combine the colors.
3

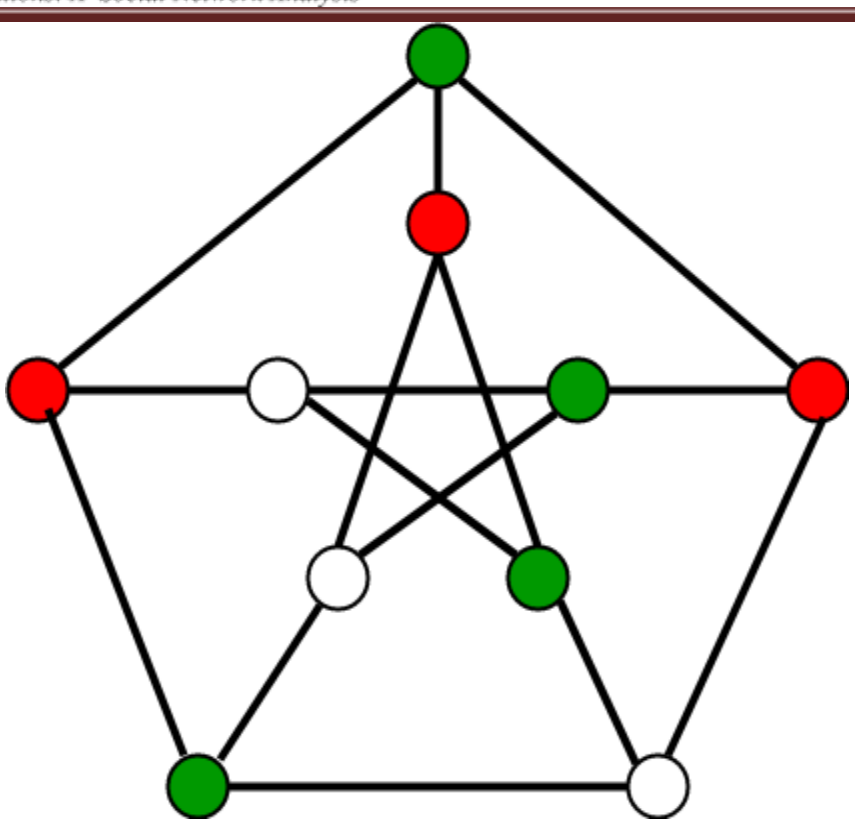


Figure 4.4 Graph Coloring

4.5 SUMMARY

This chapter was based on the entire module and their execution in the project. In this chapter we saw the detailed design of the entire project. Diagrams and plots corresponding to the modules have also been shown.

CHAPTER 5

IMPLEMENTATION

If we consider any development life cycle, implementation is an important phase. It is a phase where the system requirements and specifications are identified and converted into a fully fledged model to satisfy the real world needs. Key functionalities identified from the design stage are converted into functions that are executable using appropriate programming languages.

Therefore we can say that the implementation phase holds a lot of importance where essential decisions on how to select a language or how to select a platform etc is taken. There are various factors such as security concerns,response time required, and data management which can influence a decision. The steps that are finally taken as decisions affects the quality and functionality of the final product.

5.1 PROGRAMMING LANGUAGE SELECTION

Selecting the right programming language to code the entire project is very much essential in the Implementation phase. One has to take the requirements into consideration while selecting the right programming language. It was evident that we would choose python for our project because it is simple in terms of syntax and it can be used easily to build any application. Python programming is very much preferred for ML and AI projects because you can import user defined and built in libraries easily and hence helps you save time and energy.

Overview of Python

In the late 1980s, a person who has worked with amazing companies like Google and is now working at DropBox, invented Python. His name was Guido van Rossum. As we can guess, Python was not preferred or known at that time. It gained popularity exponentially due to its easy adoption with ML projects and for easy accessibility of many libraries.

The good thing is that Python is not difficult to learn. In fact, Python is known to be one of the easiest languages to use and learn, provided that one is familiar with basic building blocks of programming languages.
Python is an interpreted language. It is object oriented and very dynamic and used for various purposes. It has a huge and inclusive standard library. It helps its user to convey a concept in limited lines of code which is impossible for languages like Java, C, C++ etc. It is known as one of the best programming languages especially when you need to do data visualization. It has a lot of open source resources which can be imported for the project. In addition to that, the Python Programming community is friendly,welcoming and spread across the globe, hence you can get assistance easily. Python has a wide range of applications. It can be used for automation . It is extended to create WEB applications as well. Since the project we have worked on to a large extent deals with social network analysis and data analysis, python is like the only preferred choice out of all the choices out there. Because Python is dynamic.

FEATURES OF PYTHON:

- **1. Effortless Learning**

  If we want to know which language is the most novice-friendly language, the answer is Python. The syntax of Python is the unornamented of all. This will help one to code in python without any complications.

  2. Highly Desired Language

  People are recognizing the importance of Python and its scope. Hence more and more people started using Python.

3. **Used extensively in latest Technologies.**

We all know how Machine Learning, AI and other latest technology is in demand. Python is the most widely used language for all these latest technologies.

4. **Corporate are adapting to Python**

It would be a lie if we say that Python has not gained any importance. It is used in big companies like Yahoo, IBM etc.These companies have adapted Python at the core of many of its applications.

5.2 PLATFORM SELECTION

Our project was developed in Windows 10. This OS was our choice given that our data needed a lot of fine tuning and visualizations which could be easily provided in the MS Excel software. Windows provide GUI for its computers. Windows minimizes the need for commands to operate OS by simply implementing mouse which navigates through menus, tabs, dialog boxes and icons. Windows was given this name due to its dynamic nature to allow multiple tasks to run at the same time. We have chosen this platform since its user friendly, since our project required more of analytics rather than focus on a very powerful development environment.

We also use a jupyter notebook as our IDE. Jupyter notebook is an open source software which can be used by a programmer to code seamlessly in python. It makes data visualization easier. It also makes sure that the packages that are needed to be imported for the project are done successfully. Jupyter notebook can either be downloaded directly or can be used via anaconda. The reason why we used jupyter is because it can be operated in Chrome/Firefox. Therefore collaboration can be done in a very efficient and easy way.

The Jupyter notebook consists of something called a “Dashboard” . It is actually a control panel where all the files are displayed, including the local files. When you open a Jupyter notebook, the very first thing you see is the Dashboard. The reason why we chose Jupyter notebook is because we can get to see the code and the results at the same place, hence becoming an ideal environment to showcase our work.

5.3 PYTHON LIBRARIES

- **Scikit-learn:** scikit learn is a wide python library which practices machine learning, cross validation of data, and preprocessing of data. It is built on NumPy and SciPy.

- **Pandas:** Pandas is a user defined python package.It basically offers a data structure that will help one in manipulation of number tables. It can be used for absolutely free.It makes the data work piece of cake. Whether it can be ordered, or unordered data, pandas can be used. But it prefers to deal with tabular data.

- **NumPy:** NumPy's is an multi dimensional array, it holds an array data structure. NumPy is a core python library which contains a collection of tools and techniques. One of these tools is multi dimensional object.This is the library created by several other developers and Jim Hugunin.

- **Maplotlib:** It is a library extensively used for visualizing the data. It was conceived by
- several other developers and John Hunter in the year 2002. It has an interface similar to MatLab.

- **SciPy:** It is a library which contains functions to perform more complex calculations, particularly scientific computations.

- **Networkx:** Networkx came into existence in May 2002. The original version was conceived , designed and created by Aric Hagberg, Pieter Swart and Dan Schult. Even though it was created in the year 2002, the first official public release was done in April 2005. This library is used to deal with the studies that are related to graphs and networks.

- **Gzip:** Just like the name suggests, gzip is the python module which deals with file compression and decompression. Using gzip module, you can perform basic operations like open(), close() and other kinds of operations.

- **Random:** Random is a module in Python which is built in. Its very purpose is to define random numbers. You can use this module to the way you want by giving a fixed range . It uses Mersenne Twister as the code generator.

- **Seaborn:** Just like Mayplotlib, Seaborn is used as a tool to visualise data. As a matter of fact, Seasborn is a library that is based on matplotlib. The reason why one can prefer seaborn over matplotlib is because seabron provides an amazing graphic interface. It is also useful for informative statistical graphics.

5.4 CODING CONVENTIONS

The developed product may not be maintained by the developer throughout its lifetime. Thus it is highly important to make the code comprehensible to any new reader as well. Making the code more readable helps in changing and maintaining the code even in the future. Besides writing the code in python that is relatively easy to understand, two other practices are also employed.

- **Naming variables and constants:** The naming is done such that the name indicates the variable's/ constant's obvious meaning without too much ambiguity. However some of the features used in the dataset are medical terms which may require a little bit of background study. Example height indicates the height of the pregnant woman.

- **Comments:** Appropriate comments are provided in the code to make the confusing parts of code easy to understand. Comments are provided as complete sentences to avoid any sort of ambiguity.

5.5 IMPLEMENTATION STEPS

1. Store the data collected from the call records in the form of a table in MS Excel, and save it as a .csv file.
2. The data has to be stored in the form of an adjacency matrix. To be more specific we have to consider the rows/columns individually, if there is a connection between two individuals, the value becomes 1, or else its 0.
3. Define a module for loading the .csv file. Check for rows with a few missing values and fill them with the mode value of each feature.
4. Delete the extra rows and columns that appear in the table.
5. Analyze the data statistically, which involves plotting the histograms and chi-square test for normality.
6. Once we make sure the adjacency matrix is complete and ready to use, read the adjacency matrix in the code.
7. The adjacency matrix serves as an input to the algorithm that is implemented. Use slicing operation to remove column header and row header and then feed it to the algorithm.
8. The output from the algorithm is as a matter of fact the answer that is required and the result that we obtain.
9. The result is further analysed by using various tools and formulas. Then we come into a conclusion. The conclusion will talk about how many minimum number of police officials that are required to monitor the suspected individuals.
10. The conclusion also talks about how much reduction of resources is done and how much man-power can be saved by carefully selecting the individuals that are to be monitored by the police organization.

5.6 SUMMARY

This chapter summaries about different techniques that could be utilised in the development phase of the project. It starts with the process undergone to select the proper language for the project and then ends with the entire overview of implementation steps.

CHAPTER 6

TESTING

6.1 SOFTWARE TESTING

The process of evaluating the system as well as its individual components with the aim of finding any bugs or errors is known as Software Testing.It is very important to contend the quality of the software and the possible bugs in the software which might arise due to various reasons before it is put into actual use. The main purpose of software testing is to find out whether the software product satisfies its requirements and specifications as it was intended to early on in the design phase.Any gaps or missing requirements in comparison to the expected requirements can be found using testing.

We can reduce the cost sustained by testing the code in parallel to the implementation.In each phase of the software development life cycle ,Testing is done with the intent of improving the phase. Besides debugging, testing also involves verification and validation of the product to ensure both functional and nonfunctional requirements are met.

6.2 PURPOSE

The purpose of this chapter is to check whether the system satisfies the various requirements both at the unit level as well as at the system level.

6.3 LEVELS OF TESTING

We used functional testing which is a form of black box testing that is based on the requirements that the system should fulfill. The system is provided with inputs at unit testing and integration testing levels and the results are checked to ensure they satisfy the functionality that they are intended for.

1. **Unit testing:** All the major components of the system are tested individually. We need to make sure that all the components are isolated from each other before testing. Testing is done on the smallest parts of the program for correctness. This process is done during the development of the system, in order to prevent the failure of the system as a whole when integrated. In our project unit testing was done for all the major modules that involved fetching of data, preprocessing, classification, evaluation of classifiers and the screening tool mechanism.
2. **Integration testing:**components might work individually but it is also important to determine how they can work when combined together. This testing is done once all the individual modules are combined together to ensure the system as a whole works without any errors, while satisfying its promises.

The testing procedure is described below.

6.4 UNIT TESTING

In this test, major modules are tested for their individual performance and correctness.

6.4.1 Unit testing for Data acquisition and cleaning

Test case ID	Unit test case 1
Purpose	Check if the data from CSV files is loaded as a structured dataset and fill in missing values.
Procedure	● Store the initial data set in the same directory as the executable program. ● Run the executable program, load the CSV file, and check for missing values.
Input	● Filename of the CSV file for loading. ● Mode values of each column to fill in the missing values in the structured dataset. ● Replace any inconsistent data.

Expected Results	• Required dataset is loaded. • Missing values are filled in.
Actual Results	Dataset loaded smoothly as expected.
Remarks	PASS

Table 6.4.1 Test case for data acquisition and cleaning module.

6.4.2 Unit testing for visualization of data

Test case ID	Unit test case 2
Purpose	Check if data can be displayed as a Graph
Procedure	● Preprocess the initial dataset. ● Find the list of nodes connected for each node. ● Draw the graph using the networkx function.
Input	Dataset stored as a matrix. [[1 0 1 ..., 0 1 1] [1 0 1 ..., 1 1 1] [1 1 0 ..., 0 1 1] ..., [1 0 0 ..., 1 1 1] [0 1 1 ..., 0 1 1] [0 0 1 ..., 1 0 1]]
Expected Results	The graphical representation of Dataset
Actual Results	The graphical representation of Dataset
Remarks	PASS

Table 6.4.2 Test case for data preprocessing module

6.4.3 Units testing for graph coloring algorithm

Test case ID	Unit test case 3
Purpose	Check if all the nodes are colored such that no two neighbouring nodes are same color.
Procedure	● Import the preprocessed data. ● Run the graph coloring algorithms.
Input	● The output from the graph coloring algorithm
Expected Results	All the nodes are colored such that no two neighbouring nodes are the same color.
Actual Results	All the nodes are colored such that no two neighbouring nodes are the same color.
Remarks	PASS

Table 6.4.3 Test case for classification module

6.4.4 Unit testing for finding the centrality of the nodes

Test case ID	Unit test case 4
Purpose	Find out the centrality of each individual nodes.
Procedure	● Run the graph coloring algorithms ● Find the number of edges of each node and store the data in a dictionary. ● Arrange the items in the dictionary in descending order.
Input	Data which is received as output from graph coloring.
Expected Results	● Display all the nodes with centrality in descending order. ● Result {'SAR': 11, 'AAR': 10, 'CEPO': 9, 'MMM': 5, 'H': 4, 'JMMM': 3, 'JLAR': 3, 'JCR': 3, 'EEG': 3, 'VA': 2, 'T': 2, 'RRS': 2, 'R': 2, 'JA': 2, 'PA': 1, 'NA': 1, 'JOR': 1, 'JGG': 1, 'JFS': 1, 'JCT': 1, 'HO': 1, 'DA': 1, 'ASB': 1, 'AN': 1, 'AL': 1}

Actual Results	● Display all the nodes with centrality in descending order. ● Result AR': 11, 'AAR': 10, 'CEPO': 9, 'MMM': 5, 'H': 4, 'JMMM': 3, 'JLAR': 3, 'JCR': 3, 'EEG': 3, 'VA': 2, 'T': 2, 'RRS': 2, 'R': 2, 'JA': 2, 'PA': 1, 'NA': 1, 'JOR': 1, 'JGG': 1, 'JFS': 1, 'JCT': 1, 'HO': 1, 'DA': 1, 'ASB': 1, 'AN': 1, 'AL': 1}
Remarks	PASS

Table 6.4.4 Test case for centrality of nodes

6.5 INTEGRATION TESTING

Once we complete the unit testing,To check for the coordination of the various modules in the project we performed the integration testing. This testing helps identify any errors or blunders when the system is working as a whole. This ensured us of the successful working of all the modules even as a group. The integration testing was done in phases.

The modules were combined sequentially to check if the input and outputs across the combined modules were running smoothly. The unstructured data with the missing values was sent to the preprocessing unit, the outputs were checked for missing values and then the data was sent to the graph coloring algorithm.The graph coloring algorithm returns the graph colored and the list of colors used to color the graph.The output is then used to find the centrality of each node.The nodes are arranged based on centrality to display the hierarchy of the drug trade organization.

6.6 SUMMARY

This chapter shows the testing done at all the levels of the project.The project is seen to fulfill all the requirements quite effectively. The levels of testing performed have also been explained in brief.

CHAPTER 7

RESULTS

We use Identifying Codes to color each node in the graph and make sure the no two connected nodes have the same color. The result is further analysed by using various tools and formulas. We use the centrality of nodes to determine how active the node is in the graph.Then we find the kingpin of the organization. The result will talk about how many minimum number of resources that are required to monitor the suspected individuals.The centrality metrics is used to determine the hierarchy of the members in the organization.The conclusion also talks about how much reduction of resources is done and how much man-power can be saved by carefully selecting the individuals that are to be monitored by the police organization.

7.1 EVALUATION METRICS

We count the number of nodes in each color and store it in a list.The most commonly occurring color is used to find the minimum number of nodes which needs to be monitored to cover all the transactions taking place in the organizations.The most uniquely colored node is the most important node in the graph and can be called the kingpin of the organization.
The reduction of resources is
calculated as follows:

Rr = ((Nt-Nc)/Nt)*100

Rr=Reduction in Resources in %

Nt=Total Number of Nodes

Nc=Number of Nodes Commonly occurring color

We use four datasets which has data about different drug trade organizations around the world .The data has been collected by investigative authorities based on the interaction between different members in the drug trade organization.

Datasets	Total No of Nodes	No of Nodes to be monitored	Reduction in resources	Type of Graph
ACERO	**25**	**13**	**48.00%**	**directed**
MAMBO	**31**	**19**	**31.71%**	**undirected**
JUANES	**51**	**30**	**41.18%**	**directed**
JAKE	**38**	**26**	**31.58%**	**undirected**

Table 7.1 Observations made during the experiment

The above table gives information about the reduction in resources by using this approach.The structure of the graph also determines the amount of resources that can be reduced.

7.3 OUTPUT

OPERATION ACERO

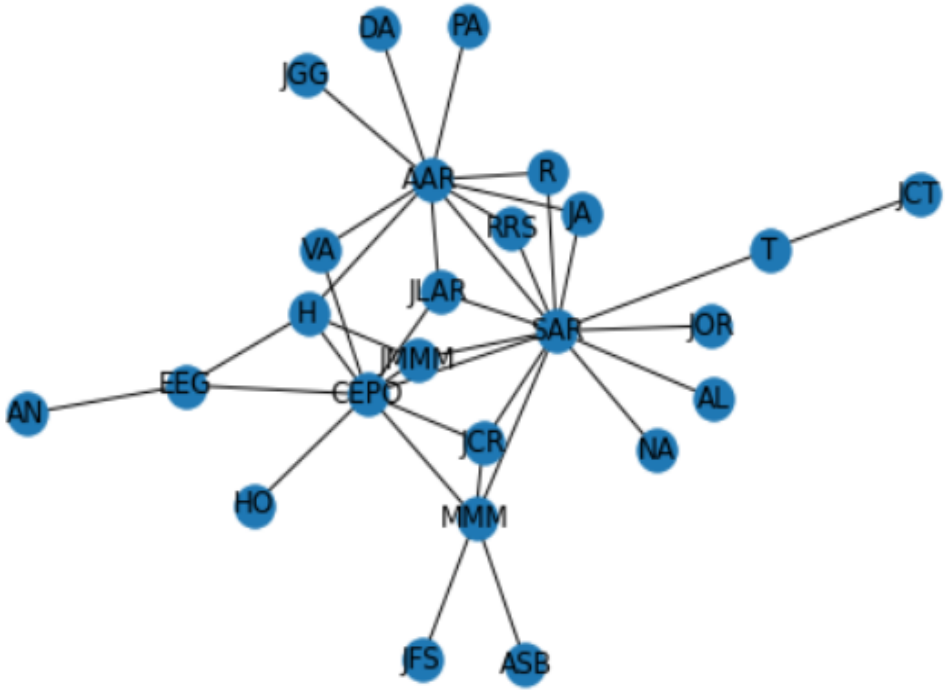


Figure 7.3.1:Operation Acero uncolored graph

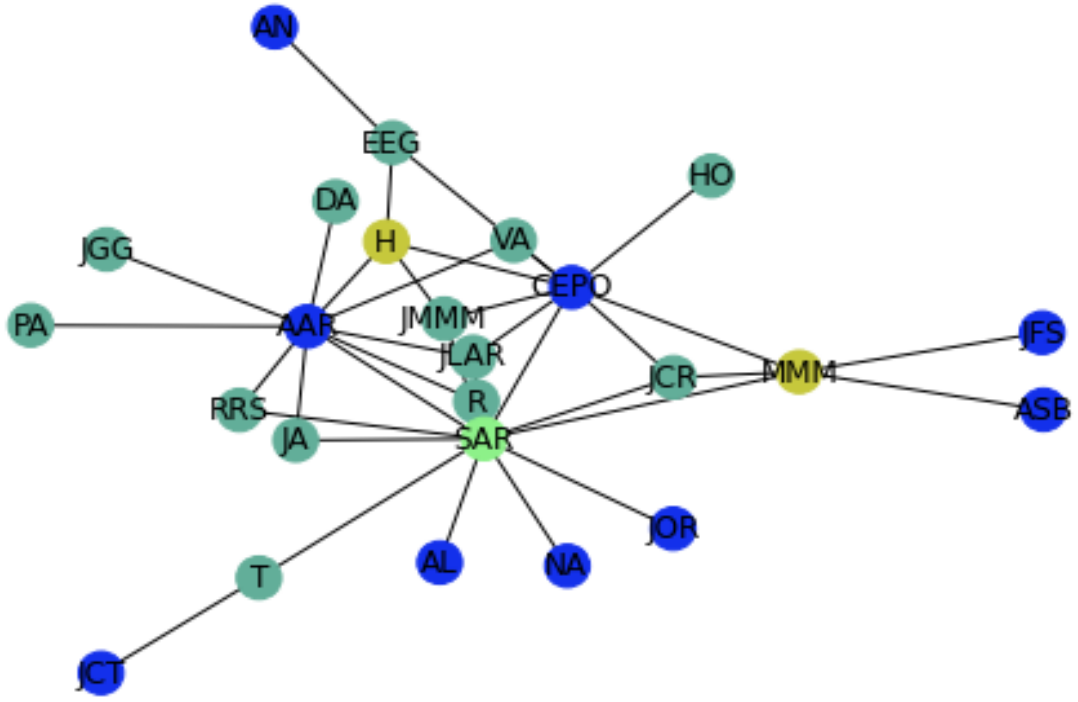


Figure 7.3.2:Operation Acero colored graph

OPERATION MAMBO

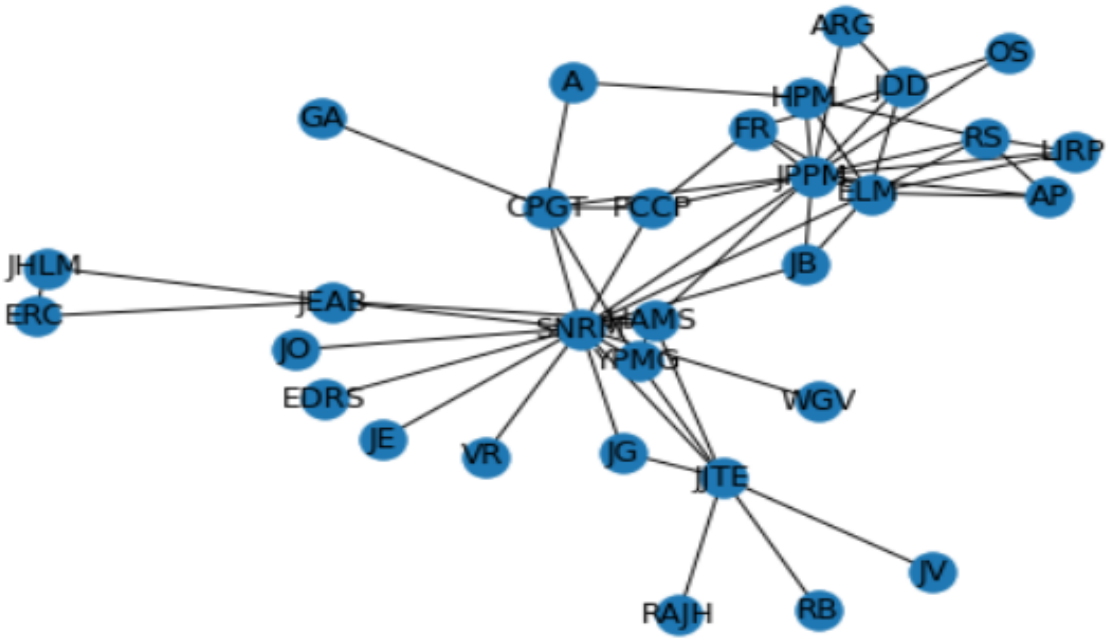


Figure 7.3.3:Operation Mambo uncolored graph

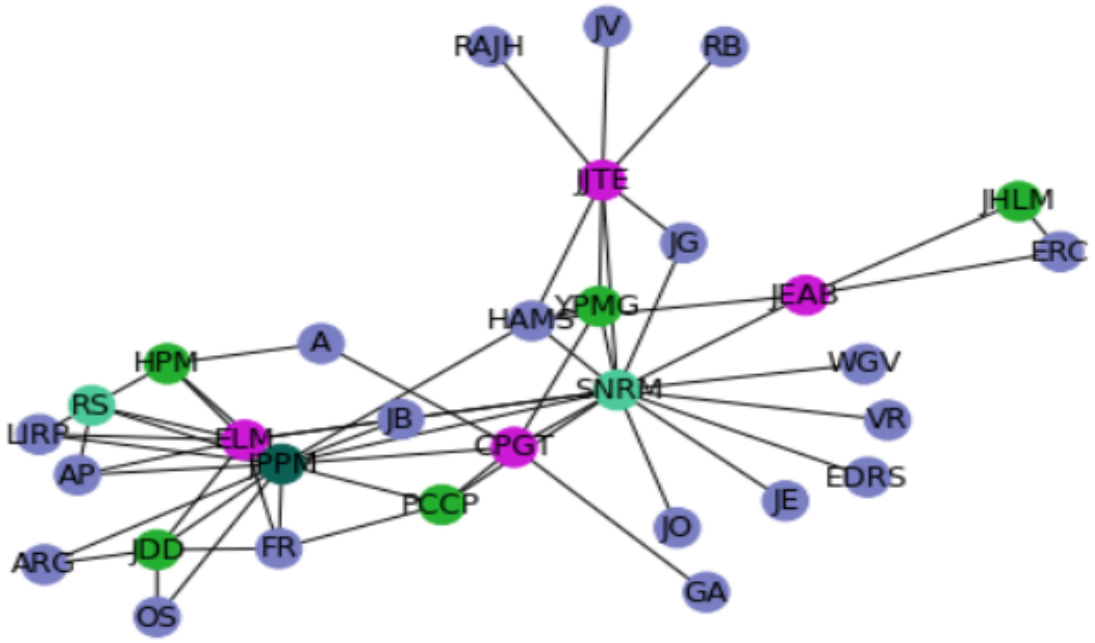


Figure 7.3.4:Operation Mambo colored graph

OPERATION JAKE

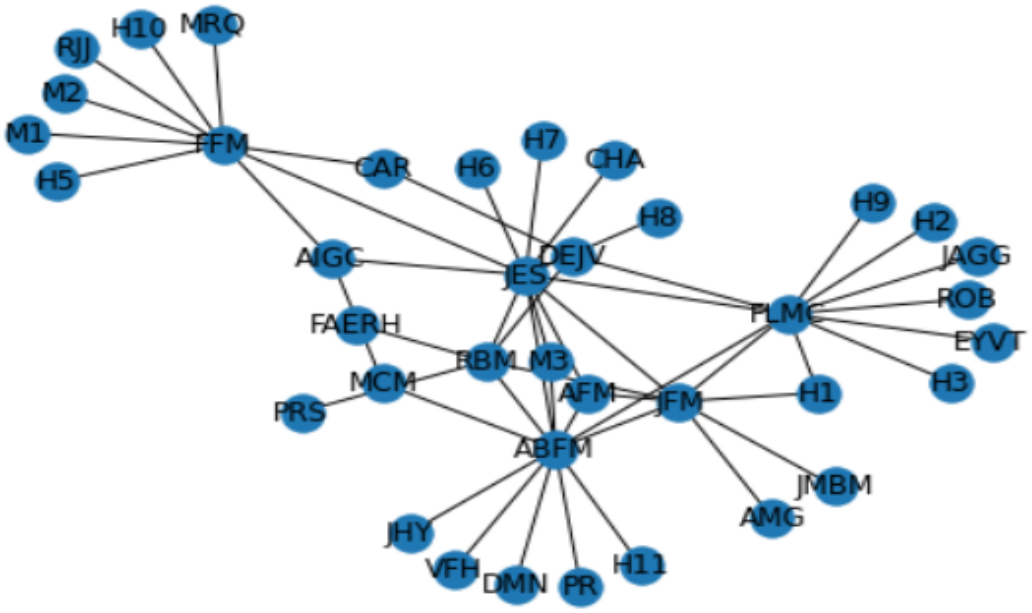


Figure 7.3.5:Operation Jake uncolored graph

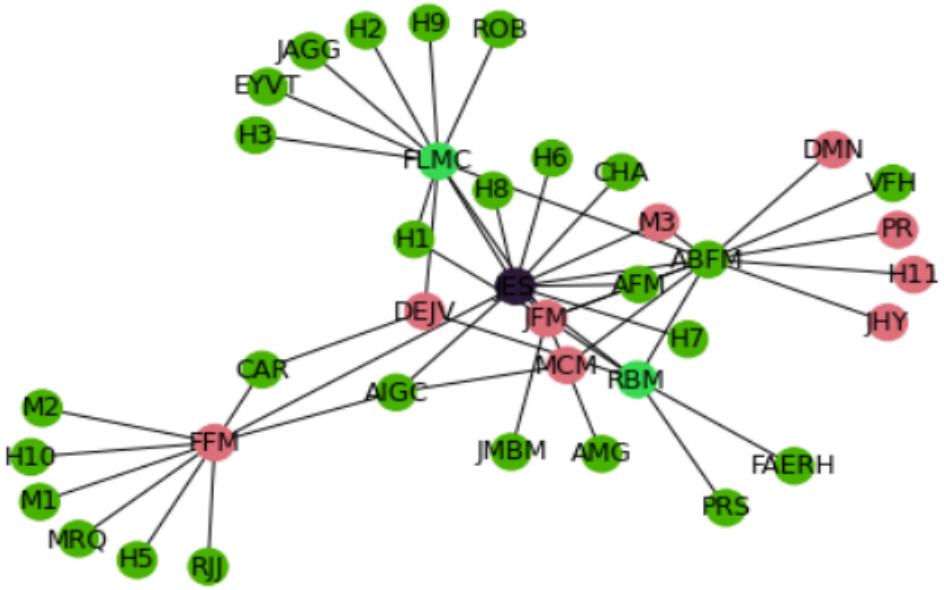


Figure 7.3.6:Operation Jake colored graph

OPERATION JUANES

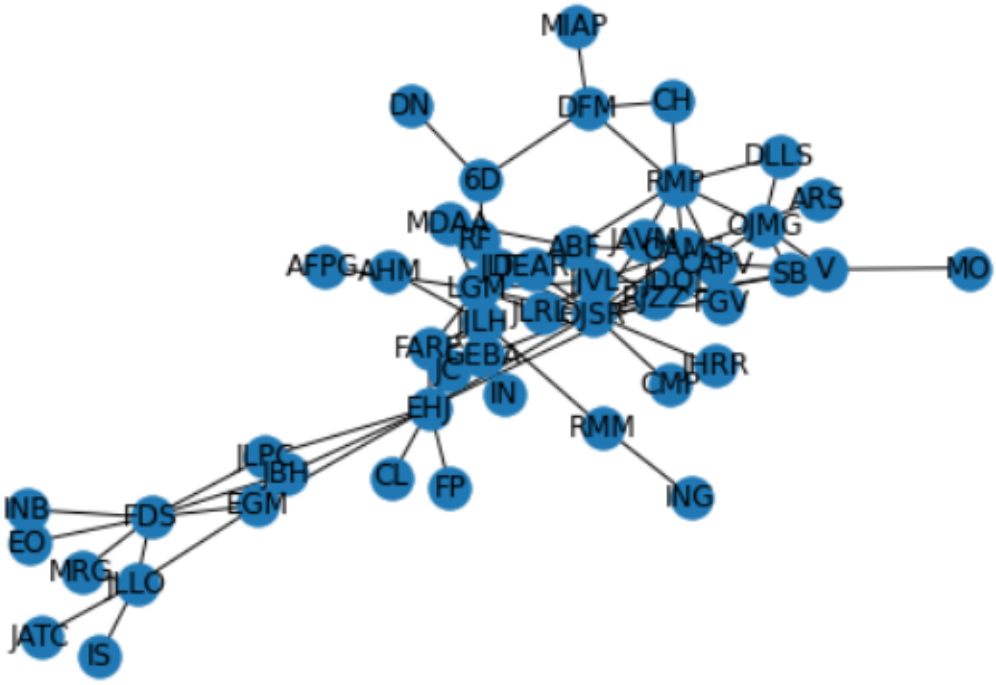


Figure 7.3.7:Operation Juanes uncolored graph

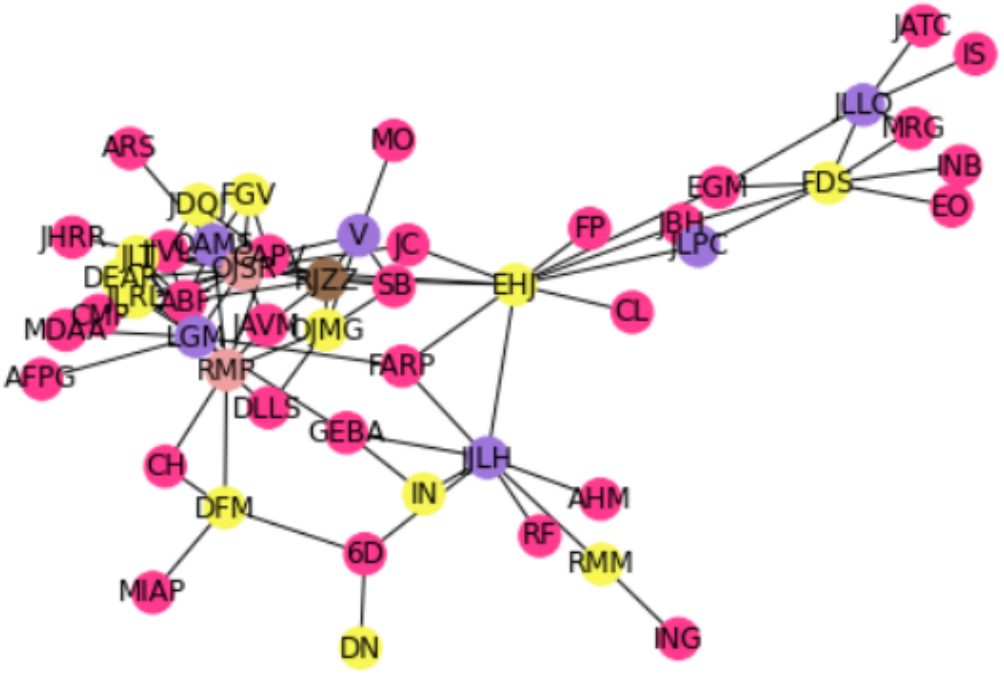


Figure 7.3.8:Operation Juanes uncolored graph

CHAPTER 8

CONCLUSION

We presented a solution technique which results in a significant reduction in resource requirements for monitoring individuals involved in drug trafficking organizations on the part of the authorities. We have used standard SNA centrality metrics to conduct extensive experimentation to show that for monitoring Organization , leads to wastage in the resources, on the part of the authorities and how Identifying Code can be used to get the minimum number of members who need to be monitored in Drug Trafficking Organization .

8.1 LIMITATIONS OF THE SYSTEM

- The number of datasets used is small and might lead to errors if large data is used.
- There is no information about the roles of each member of the organization.
- There might be other members in the organization who are not yet accounted for and if they are taken into consideration can change the result.
- There is no way to change the data real time if there is any change in the structure of the organization.

8.2 FUTURE ENHANCEMENTS

- We can build a more complex system by obtaining more data like monetary details ,roles of individual members.
- Create more in detail visualization to make it easier for users to understand.
- We can predict the motives of the suspects.
- Make observations to the system and understand the changes in the structure of the drug trade organization.

REFERENCES

[1] "Drug Trafficking"[Online]. Available:https://www.unodc.org/unodc/en/drug-trafficking/'

[2] "UNODC World Report" [Online]. Available: https://www.unodc.org/wdr2018/en/exsum.html

[3] C. E. Hughes, D. A. Bright, and J. Chalmers. "Social network analysis of Australian poly-drug trafficking networks: How do drug traffickers manage multiple illicit drugs?" Social Networks 51 (2017): 135-147.

[4] "How Drug Dealers Get Caught" [Online]. Available: https://www.robertkinglawfirm.com/blog/2013/november/how-drug-dealers-get-caught/

[5] M. Natarajan, "Understanding the structure of a drug trafficking organization: a conversational analysis." Crime Prev. Studies 11 (2000): 273-298.

[6] M. Natarajan, "Understanding the structure of a large heroin distribution network: A quantitative analysis of qualitative data." Journal of Quantitative Criminology 22.2 (2006): 171-192.

[7] D. A. Bright, and J. J. Delaney. "Evolution of a drug trafficking network: Mapping changes in network structure and function across time." Global Crime 14.2-3 (2013): 238-260.

[8] D. A. Bright, et al. "The use of actor-level attributes and centrality measures to identify key actors: A case study of an Australian drug traffic network." J. of Contemp. Criminal Justice 31.3 (2015): 262-278.

[9] D. A. Bright, C. E. Hughes, and J. Chalmers. "Illuminating dark networks: A social network analysis of an Australian drug trafficking syndicate." Crime, law and social change 57.2 (2012): 151-176.

[10] K. M. Carley, et al. "Destabilizing dynamic covert networks." Proceedings of the 8th international Command and Control Research and Technology Symposium. 2003