Analysis of Autonomous systems and its temporal dimensions using Social network analysis

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology
in
Computer Science and Engineering

By

Team Member	Registration Number
Vinay G	16BCE0497
Pavan Teja V	16BCE0512
Bhaskar G	16BCE0225

Under the guidance of

Prof. Meenakshi S P

Scope, School of Computer Science and Engineering
VIT, Vellore.

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
May, 2020

<u>DECLARATION</u>

I hereby declare that the thesis entitled "Analysis of Autonomous systems and its temporal dimensions using Social network analysis" submitted by Vinay, Venkata pavan teja, Bhaskar for the award of the degree of *Bachelor of Technology in Computer Science and Engineering* to VIT is a record of bonafide work carried out by me under the supervision of Meenakshi SP.

I further declare that the work reported in this thesis or project has not been submitted and will not be submitted, also in part or in full, for the award of any other degree or diploma in this institute or any other institute.

Place : Vellore

Date :27-05-2020

Venkata pavan teja

Signature of the Candidate

CERTIFICATE

This is to certify that the thesis entitled "Analysis of Autonomous systems and its temporal dimensions using Social network analysis" submitted by Vinay& 16BCE0497, Venkata pavan teja & 16BCE0512, Bhaskar & 16BCE0225 Scope, VIT, for the award of the degree of *Bachelor of Technology in Computer Science and Engineering*, is a record of bonafide work carried out by him / her under my supervision during the period, 01. 12. 2018 to 30.04.2019, as per the VIT code of academic and research ethics.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The thesis or project fulfills the requirements and guidelines of the University and in my belief meets the necessary standards for the submission.

Place : Vellore

Date :27-05-2020

Venkata pavan teja

Signature of the Guide

Internal Examiner **External Examiner**

Prof. Santhi .V

Computer Science and Engineering

ACKNOWLEDGEMENT

I wish to express our heartfelt gratitude to Dr. G.Viswanathan, Chancellor, VIT, Vellore, for providing facilities for the final year project. I am highly grateful to our Vice President, Dr.G. Sekar Viswanathan, Vice Chancellor Dr. Anand A. Samuel, and Pro-Vice Chancellor Dr. S. Narayanan, for providing the necessary resources.

We would like to thank Head of the Department of Computer Science and Engineering, VIT University, **prof. Santhi.V** and our Project Guide **prof. Meenakshi.SP** who guided us through doing these projects, provided with invaluable advice, helped us in difficult periods and provide information and suggestions for our projects. Their willingness to motivate us contributed tremendously to the success of these projects. Besides we would like to thank all CSE staff members who helped us by giving advice and providing information which we needed.

I thank the management of VIT, Vellore for permitting me to use the library resources. I also thank all the faculty members of VIT, Vellore for giving me the courage and strength I needed to complete my goals. This acknowledgement would be incomplete without expressing my whole hearted thanks to my family and friends who motivated me during the course of the work.

I have to appreciate the guidance given by other faculties as well as the panels especially in our project presentation that has improved our presentation skills thanks to their comments, advices and suggestions.Last but not in least, we would like to thank all who helped and motivated us to complete the project.

Place: Vellore

Date:27/5/2020

Executive Summary

The main goal of this project is to building a successful modeling tool able to generate realistic graphs for use in networking simulations. We took BGP routing data set, which consists of router numbers and connections among them. In BGP network, it have many routers, each router has one Autonomous number (AS). First, we converted the BGP dataset into a graph.

In this project we provide a detailed analysis of the inter-domain topology of the route view Internet. For analyzing and visualizing the network or graph we used the concepts of Social Network Analysis. The collected data and the resulting analysis will be based on one-year span routing dataset, we give results concerning major topology properties (nodes and edges number, average degree and distance, routing policy(AS), etc.) and main degree distributions , distance, etc. We provide different types of visualizations for each and every property in the network, to the users for their convenience. Data modeling is a technique used to model a data and to enhance the network for the future purpose. Eventually, we will analyze and research data modeling based on the properties and visualizations in the network.

CONTENTS

List of Figures

List of Abbreviations

BGP	Broad Gateway Protocol
SNA	Social Network Analysis
DD	Degree Distribution
CC	Closeness Centrality
DM	Data Modeling

1.Introduction

1.1 Objective

In this project, we are using BGP(Broad Gateway Protocol) dataset. We generate the graphs and analyze the trend among the different countries from the network by using social network analysis in different parameters of social network analysis, also with different visualization layouts. We are generating dashboard for the route view of the BGP dataset for the user convince. Also, we are developing the data modeling to the social network for the future purpose.

1.2 Motivation

Social network analysis is a field of study that attempt to understand relationships between entities in a network based on the assumption about the significance of relationship between entities. There are numerous meanings of social network analysis and the most used one is: Social network analysis has appeared as a set of methods for the analysis and examine of social structures, methods which are specifically geared towards an exploration of the relational aspects of these structures.

The use of these particular methods, depends on the availability of interpersonal, relational rather than attribute data. Social network analysis is mainly focused also on visualization techniques for exploring the networks structure with nodes and edges. Graph structure is often used to represent social networks and their size becomes increasingly huge as the progression of the means for data congregation and the storage gradually strengthens. This call for new methods in graph visualization and analysis for dealing with the problem of large data, graphs and networks.

1.3 Back ground

In generally, social network analysis is a common word but there are many definitions of it. Today, people think that social networks are Facebook, Twitter, instagram, Google or any website with the keyword “social network”. Generally, social network is known as a social structure in which actors or individuals called nodes, and all nodes connected but the some sort of relationships called edges such as: friendship, common interest, relationship, beliefs, knowledge, neighbors, prestige etc. The main principles or characteristics of a social network are:

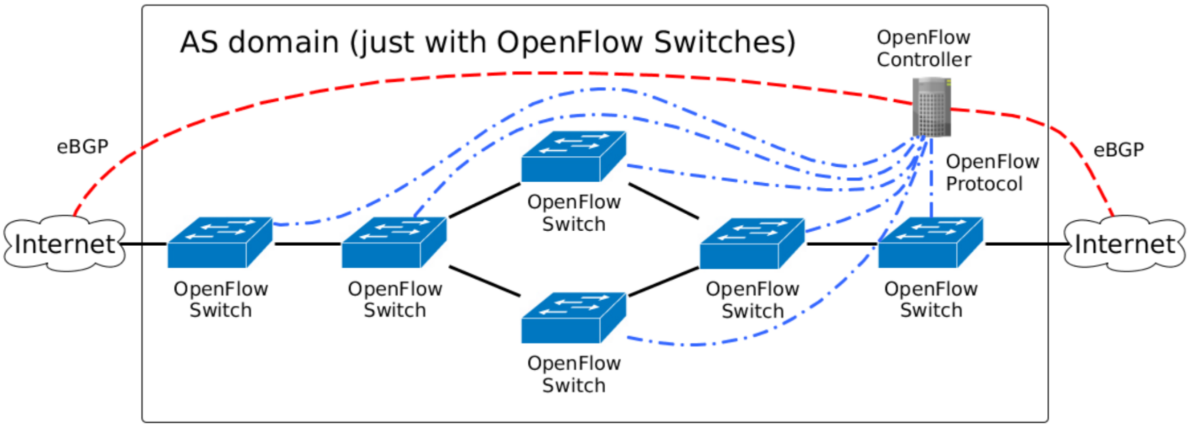
- There exists a collection of entities that present in the network. Typically, these entities are people or individuals in social network.

- There exist some sort of or at least one type of relationships between entities. The relationship can be a type of “all-or-nothing” or has a degree.

- There exist an assumption of non-randomness and irregularities. It means that if node A has relationships with both nodes B and C, then there is a high probability that B and C are related.

2. PROJECT DESCRIPTION AND GOALS

Our tool consisting a data is BGP dataset which it consists of a AS number (router number), edges (connection between them).In the existing network tools of BGP dataset, they didn't have a different visualization techniques, network properties. We proposed the different visualization techniques among the various properties of the network. Some of them are degree centrality, closeness centrality, betweeness centrality, community detection, cliques, hubs, degree distribution, clustering coefficient etc.



2. BGP Route-View

We designed the dashboard for the route view of data set based on the different countries, years and months. Also we proposed some of the visualization features, like highlighting the adjacent nodes and edges of a particular node, visualizing in a circular way. In this project, we are going to analyze the trend of a dataset (route view) of a particular country, particular month. Eventually, based on the properties of a network we are going to do a data modeling like preferential networks, small world networks.

BGP dataset consists of a AS number of every router and the connections between them. We are applying the social network analysis to the autonomous system (bgp dataset) to analyze the trend and social behavior of network on country wise analysis. We are designing the dash board with different visualization techniques and layouts. For developing this, we are using R Studio. It is an open source software and work under different platforms (Windows, linux, Mac).

3. TECHNICAL SPECIFICATION

In generally Border Gateway Protocol (BGP) is the postal service of the Internet. When someone drops or posts a letter into a mailbox, the postal service processes that piece of mail and it chooses a very efficient route to post that letter to its recipient or receiver. Similarly, when someone generates and submits data in the Internet, BGP is responsible for looking at all of the available routes that data could travel and choosing the best and optimal path, which hopping between the autonomous systems. We took the BGP dataset, generated the graph and also identified the many entities like closeness, betweeness, community detection, degree distribution, identifying and highlighting the adjacent routers(nodes). Some of the properties we are using to visualize the data are

Closeness Centrality

Closeness centrality is the way of finding nodes that are able to pass the information very efficiently through a graph which consists of nodes and edges. The closeness of a node measures its average distance (inverse distance) to all other nodes. Node consisting with high closeness value, have the shortest distances to all other neighbor nodes.

Betweenness centrality

Betweenness centrality is a measure of the influence of a particular node over the flow of information between every pair of nodes under the assumption that information largely flows over the shortest paths between the nodes in the graph.

Degree Centrality

Degree centrality is the simplest centrality, to compute. Node degree is simply a count of how many social connections or edges passing through it. The degree centrality for a node in network, is simply its degree. A node with 20 social connections or edges would have a degree centrality of 20.

Degree Distribution

The degree of a node in a network is the number of connections it has to other nodes and the degree distribution is the probability distribution of these degrees of nodes over the whole social network or graph.

Product Perspective

To visualize the BGP route view network data visually in a simplified manner. It allows users to understand at a glance without knowing any complex terminology. Usually the route view data sets so complex and it needs so much space and functional requirements just to open it our systems , and then it is very hard to understand and relate each other of whatever we want to know and study. So our main perspective is to visualize these complex networks datasets.

Product features

Product or tool it shows the routing information among the routers i.e., closeness, betweeness, degree, degree distribution etc. Our tool will display the data with different visualizations techniques. Also, we designed the dashboard to visualize the data clearly and it is very convenient and easy to use.

User characteristics

Now a day's internet is becoming more involved in our day to day lives. Every piece of work is relatable and depending on the internet and many internet service providers, social network analysts and other data scientists needs some simplified visualization to be in grasp of the work they are doing. The main characteristics regarding our project are to make these users understand the whole network. We derive so many characteristics like Closeness centrality, Betweenness centrality, Degree Centrality, Degree Distribution

Assumption & Dependencies

There are lot of assumptions that we need to focus on while doing these kind of work. The output we get will conclude and give different ways of answers to the users in their own way of getting their answers.

Domain Requirements

A normal pc of any operating system with R Studio software installed in it. And the data BGP route view data sets to visualize them. And good internet connection to download and install some libraries online for the R Studio.

R Studio is a IDE which is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to R Studio Server or R Studio Server Pro (Debian/Ubuntu, Red Hat)

Uses and Advantages

- Access R Studio tool locally
- Highlighting the syntax, code completion, and indentation
- Execute and implement R code directly from source editor
- Quickly jump to the function definitions
- Easily manage multiple working directories using projects
- Integrated R help and documentation about libraries
- Interactive debugger to recognize and fix errors quickly
- Extensive package development tools

4.DESIGN APPROACH AND DETAILS (as applicable)

4.1 Design Approach / Materials & Methods

Our project, is related to software. In our approach, we design three steps

1. Collecting the BGP dataset && converting the BGP data into graph or network:

We collected daily BGP routing data of different countries. BGP data set consists of router numbers and connections among them. Eventually, we converted the data into graph or network with nodes and edges.

2. Analyze and visualize the network

For analyzing the network, social network analysis is the best way by using the IDE as R studio. We used packages like visnetwork, igraph etc., for visualizing the network by using the R language. SNA concepts like closeness, degree, betweeness, highlighting the nodes and edges etc and for visualization we used various techniques like bar plot, scatter plot, round view etc. Eventually, we found the trend analysis of a network.

3. Dash board and data modeling

We developed the dashboard for the network for the user and operator convince. Also, we researched and analyzed the data modeling like small world networks, scale free networks.

4.2 Codes and Standards

The IDE we used R Studio and the language is R. Some of the codes and standards are

```
install.packages("networkD3")
install.packages("plotly")
library(visNetwork)
library(igraph)
library(networkD3)
library(plotly)


links <- read.csv("C:/Users/pavan/Desktop/Capstone Project/proj.csv", header=T, as.is=T)
nodes <- read.csv("C:/Users/pavan/Desktop/Capstone Project/proj1.csv", header=T, as.is=T)
edges <- data.frame(from = links$From,to = links$To)
head(nodes)
graph <- graph_from_data_frame(d=edges,vertices = nodes,directed = FALSE)
graph
plot(graph)
tkplot(graph,edge.arrow.size=.5)
data <- toVisNetworkData(graph)
nodes<-data.frame(data$nodes)
edges <- data.frame(data$edges)
visNetwork(nodes, edges) %>% visEdges(arrows = 'to')



#visulization
visNetwork(nodes, edges) %>%
  visIgraphLayout() %>%
  visNodes(
    shape = "dot",
    color = list(
      background = "#0085AF",
      border = "#013848",
      highlight = "#FF8000"
    ),
    shadow = list(enabled = TRUE, size = 10)
  ) %>%
  visEdges(
    shadow = TRUE,
    arrows = list(to = list(enabled = TRUE,scaleFactor=2)),
    color = list(color = "lightblue", highlight = "red")
  ) %>%
  visOptions(selectedBy = "ID",
             highlightNearest = TRUE,
             nodesIdSelection = TRUE) %>%
  visPhysics(stabilization = TRUE)%>%
  visLegend()%>%
  visLayout(randomSeed = 1100)



#added hover to edges
edges <- data.frame(data$edges,title = paste("Edges :","<br>",'From',links$From,"<br>","To",links$To, sep = " " ))
nodes <-data.frame(data$nodes,title = paste("Node:",nodes$id))
visNetwork(nodes, edges) %>%
  visIgraphLayout() %>%
  visNodes(
```

4.2.1Codes and standards

```
transitivity(graph, type="local")
#diameter
diam <- get_diameter(graph, directed=T)
diam
vcol <- rep("gray40", vcount(graph))
vcol[diam] <- "gold"
ecol <- rep("gray80", ecount(graph))
ecol[E(graph, path=diam)] <- "orange"
# E(net, path=diam) finds edges along a path, here 'diam'
plot(graph, vertex.color=vcol, edge.color=ecol, edge.arrow.mode=0)


# node degree
deg <- degree(graph, mode="all")
degree(graph, mode="all")
d<-degree(graph, mode="all")
table(d)
d
tkplot(graph, vertex.size=deg*3)

fig <- plot_ly()
fig <- fig %>%
  add_trace(
    type = "funnel",y=c("1740", "5459", "5413", "1849",  "702",  "701", "3333",  "286", "1239", "2548",  "568", "3847", "234","266"," 237", "3561", "2497'
x=c( 8 ,   5 ,   5   , 3   , 2 ,  19  ,  2  ,  9 ,  11  ,  5 ,   9 ,   3 ,   1 ,   2,    2,  18 ,   6  ,  8 ,   2 ,   8 ,   2 ,   3   , 2 ,   3  ,  1, 2
    )
fig
barplot(d)
hist(deg, breaks=1:vcount(graph)-1, main="Histogram of node degree")



#degree frequency
deg <- degree(graph, mode="all")

tkplot(graph, vertex.size=deg*3)
deg.dist <- degree_distribution(graph, cumulative=T, mode="all")

plot( x=0:max(deg), y=1-deg.dist, pch=19, cex=1.2, col="orange",

      xlab="Degree", ylab="Cumulative Frequency")



###hubs and authority
hs <- hub_score(graph, weights=NA)$vector
as <- authority_score(graph, weights=NA)$vector
par(mfrow=c(1,2))
tkplot(graph, vertex.size=hs*50, main="Hubs")
tkplot(graph, vertex.size=as*30, main="Authorities")
```

4.2.2 codes and standards

~/finalyear - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

server.r × ui.r ×

Run App

```
#install.packages("shiny")
install.packages("visnetwork")
library(igraph)
library(shiny)
library(visNetwork)
library(shinydashboard)
links <- read.csv("C:/Users/pavan/Desktop/Capstone Project/ind_1jan_6_00.csv", header=T, as.is=T)
nodes <- read.csv("C:/Users/pavan/Desktop/Capstone Project/ind1_1jan_6_00.csv", header=T, as.is=T)
edges <- data.frame(from = links$from,to = links$to)
graph <- graph_from_data_frame(d=edges,vertices = nodes,directed = TRUE)
data <- toVisNetworkData(graph)
nodes<-data.frame(data$nodes)
edges <- data.frame(data$edges)


server<-function(input,output)
{
  output$network<-renderVisNetwork({
    visNetwork(nodes, edges) %>%
      visIgraphLayout() %>%
      visNodes(
        shape = "dot",
        color = list(
          background = "#0085AF",
          border = "#013848",
          highlight = "#FF8000"
        ),
        shadow = list(enabled = TRUE, size = 10)
      ) %>%
      visEdges(
        shadow = TRUE,
        arrows = list(to = list(enabled = TRUE,scaleFactor=2)),
        color = list(color = "lightblue", highlight = "red")
      ) %>%
      visOptions(selectedBy = "ID",
                 highlightNearest = TRUE,
                 nodesIdSelection = TRUE) %>%
```

51:2 (Top Level)

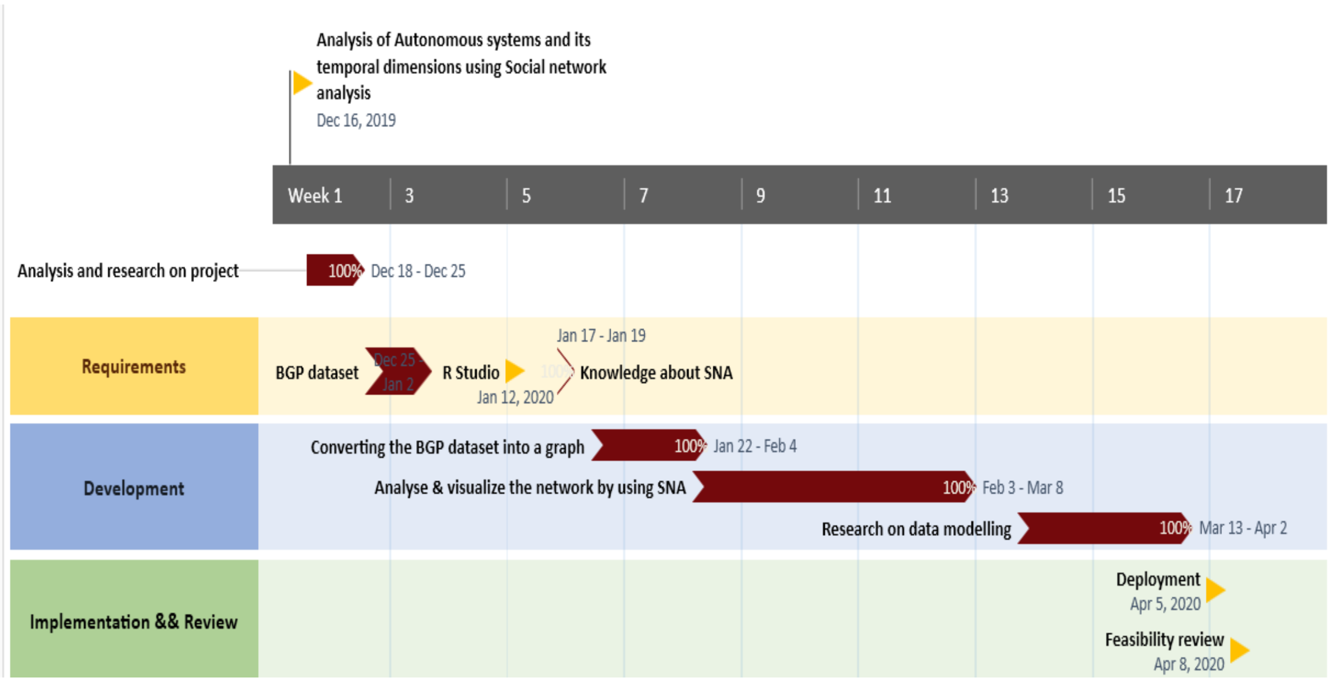
4.2.3 Codes and standards(Dash board)

5. SCHEDULE, TASKS AND MILESTONES

The complete schedule, tasks and mile stones of our project

Title	T/M	Start	End		%
Analysis and research on project	T	18-12-2019	25-12-2019	7 days	100
Analysis of Autonomous systems and its temporal dimensions using Social network analysis	M	16-12-2019	16-12-2019	–	100
Requirements					
BGP dataset	T	25-12-2019	02-01-2020	8 days	100
Knowledge about SNA	T	17-01-2020	19-01-2020	2 days	100
R Studio	M	12-01-2020	12-01-2020	–	100
Development					
Converting the BGP dataset into a graph	T	22-01-2020	04-02-2020	14 days	100
Analyse & visualize the network by using SNA	T	03-02-2020	08-03-2020	34 days	100
Research on data modelling	T	13-03-2020	02-04-2020	21 days	100
Implementation && Review					
Deployment	M	05-04-2020	05-04-2020	–	90
Feasibility review	M	08-04-2020	08-04-2020	–	100

5.1Tasks and Milestones(Dataline)



5.2Tasks and Milestone(Timeline)

6. PROJECT DEMONSTRATION

In our project, the first step is collect, analyze the BGP routing data and convert into the graph or network which consists of nodes and edges, nodes are routers, edges are the relationships or connections between them. In BGP data set each router consists of Autonomous number. Secondly, we analyze and visualize the network with various social network analysis concepts and different visualization techniques. Some of the concepts or parameters we consider in social network analysis are closeness centrality, betweeness centrality, degree distribution, community detection, hubs, clustering coefficient etc.

We developed various visualization techniques like bar plot, scatter plot, round view etc., for the user convince. Also, we focused an developed the dashboard to visualize the country wise routing data with various other parameters. Eventually, we found and develop the trend analysis of a network. Also, we researched on the data modeling network like scale free, small world networks etc. This data modeling will help to analyze, visualize the network data easily and also helps to the future enhancements to the network

Operational Requirements

Economic

Network based companies will be beneficial by going through our tool and they will use this data for future enhancements.

Social

Social network users and data scientists will use our tool to get to know the routing information with different visualization techniques and dashboard.

Political

A formal social network analysis with various visualizations techniques outlook can be employed to understand political, economic and social organizations and individuals. In network, actors within the set of their interactions or relationships allows for intuitions on the distribution of power and the effective influence of social and route view based on the actors.

Ethical

This type of social network mapping, which consists of actors and relations between them, has more obvious ethical inferences because participants of the map may never know that they are actually going to be mapped. In survey-based, electronic mapping methodologies and route view based networks wants to keep the identities of individuals confidential, for protecting their privacy.

Legality
Social network analysis used to identify high-priority needs for law enforcement's use of social media and social network analysis.

Inspect ability
SNA is the process of exploring the social structures through the use of networks and graph concept. It exemplifies the networked structures in terms of nodes (individual actors, people) and the edges, or links considered as relationships that connect between the nodes.

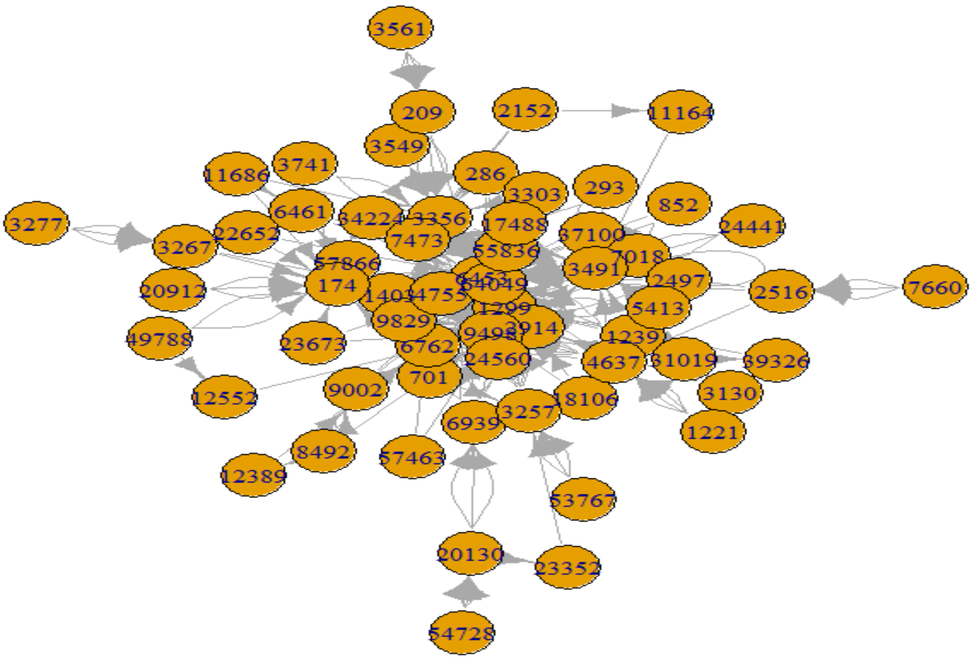
Efficiency
By using social network analysis in BGP, we can easily find out which routers routing the more information or data and links between them within short period of time. It is very efficient because at a time it has a different type of visualizations of the network.

Reliability
A measure is said to be a high reliability, if it produces a similar results under consistent conditions and standards. It is highly reliable under consistent datasets which consists nodes and edges.

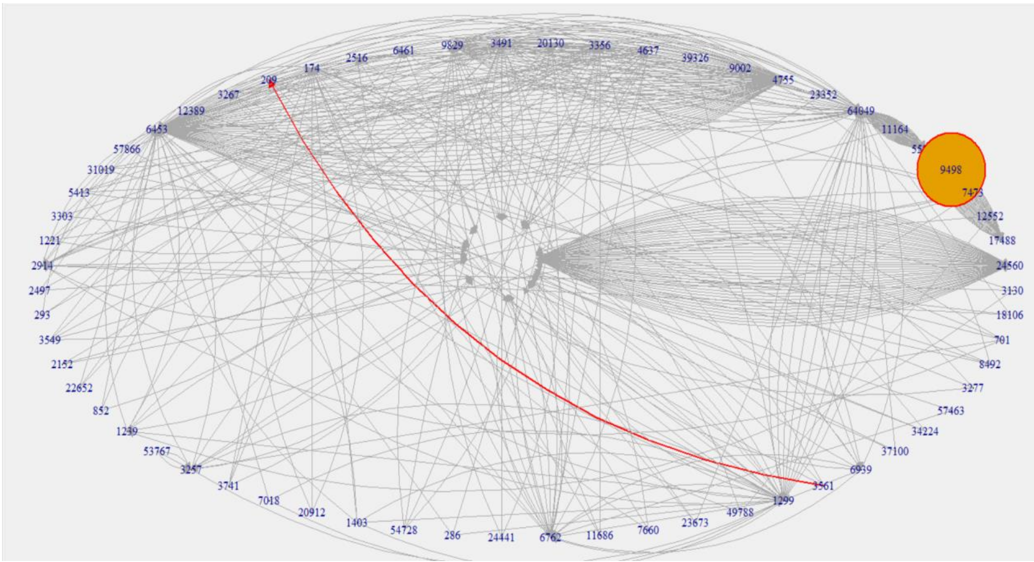
Portability
Portability, in software, is a measure of how an application can be transferred or transmitted from one computer environment to another environment. Our project or tool is portable under any platforms(mac/linux/windows)

Usability
Usability is a degree to measure, in which how customers or users using a particular application and also measure how the users efficiency, effectiveness and satisfaction in terms of usage of software. It is easily and highly usable to the consumers, we designed the dashboard and different visualizations.

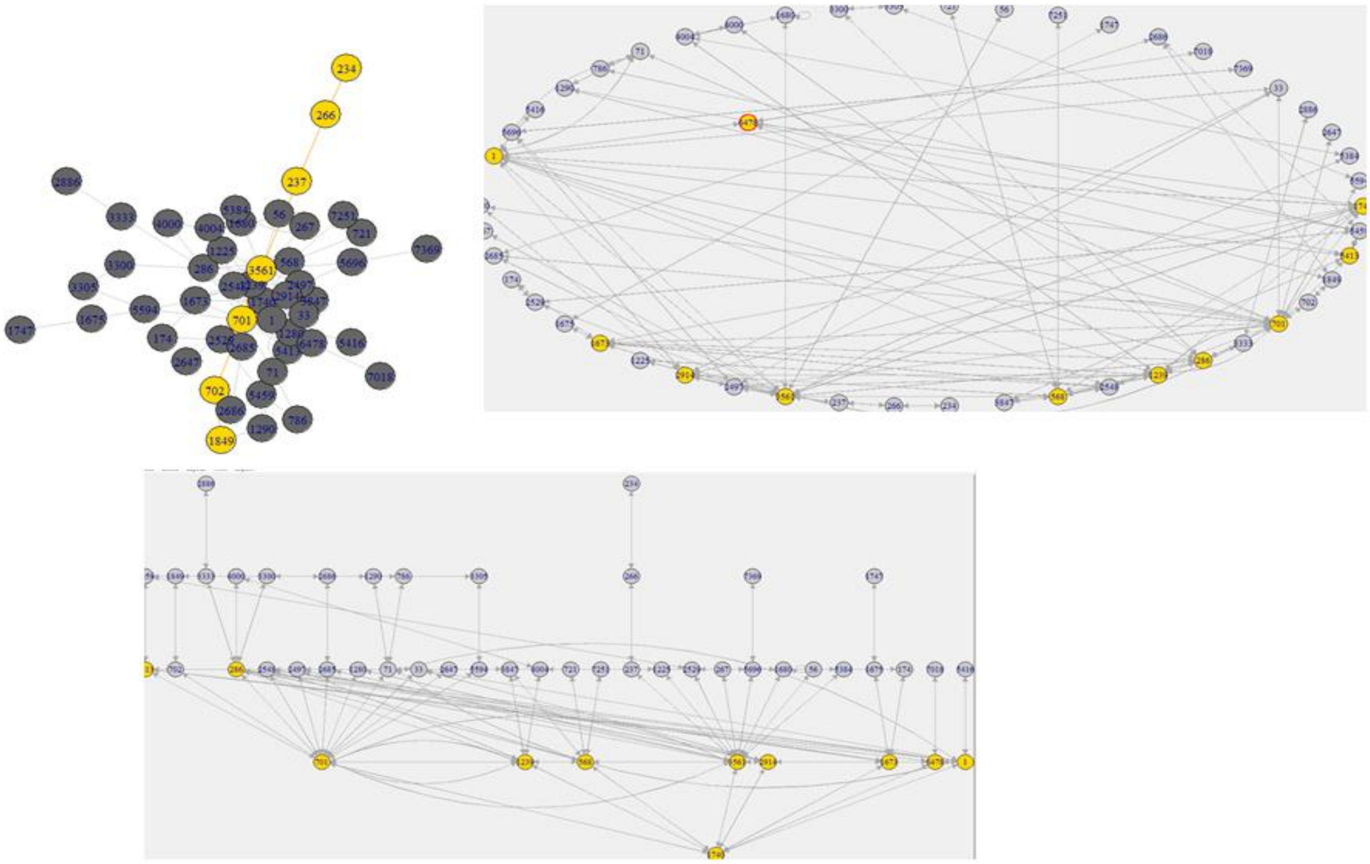
7. COST ANALYSIS / RESULT & DISCUSSION (as applicable)



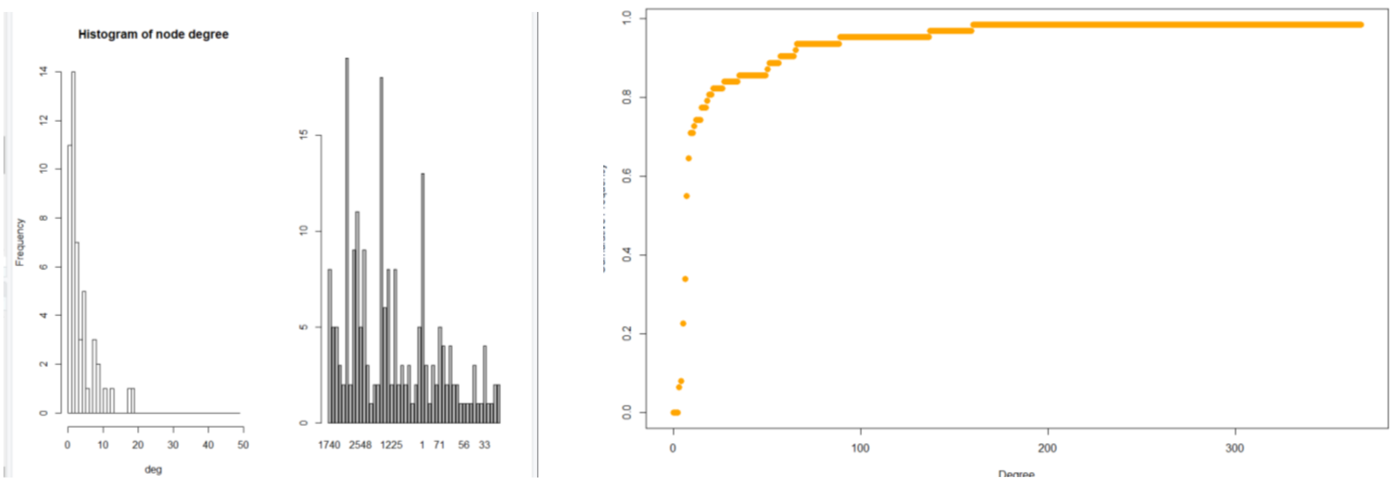
7.1 *BGP Graph or Network*



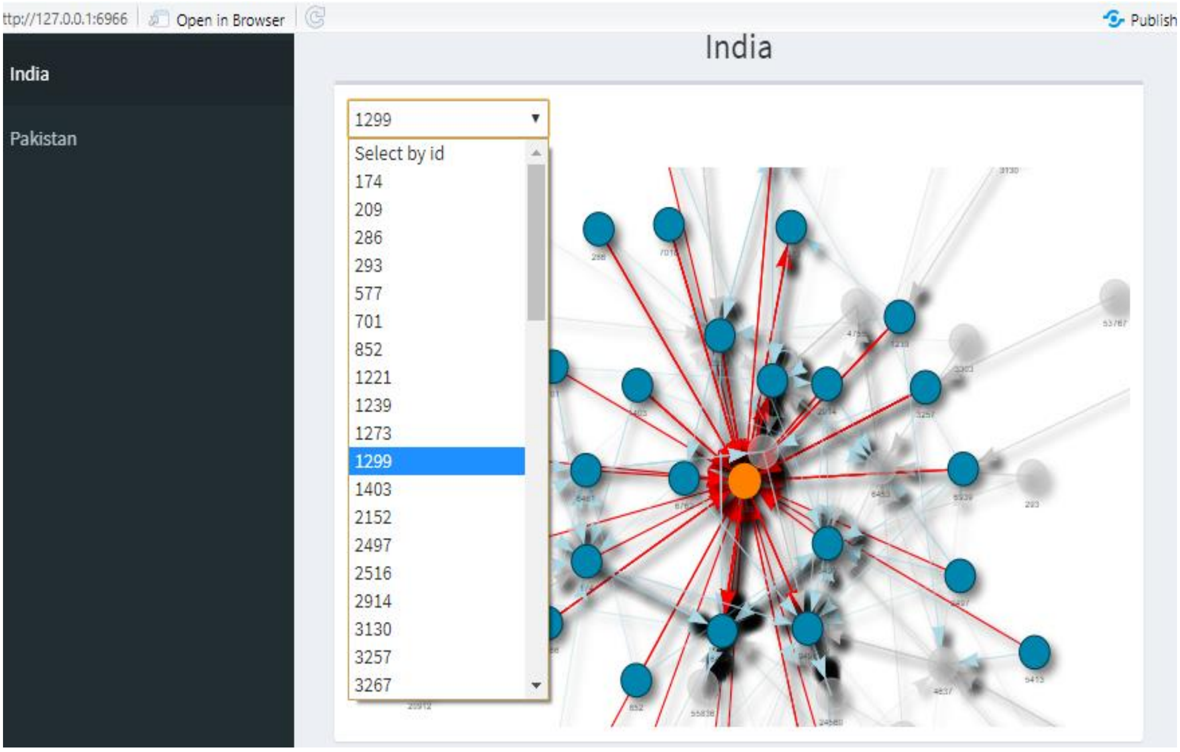
7.2 Round view visualization



7.3Some of the graphs(SNA) and visualizations



7.4Trend analysis of BGP network



7.5 Dash Board

Here the some of the results, of BGP data converting into graph and visualizing the network with various visualization techniques by using Social network analysis concepts. Also, we found out the trend analysis of a BGP dataset of a country wise and day wise. Eventually, we researched on the data modeling. Some of the data modeling techniques are small world networks, large world networks and scale free networks. Mainly, data modeling used for to model the data and used to enhance the network or graph in future.

A small-world network is a network or mathematical graph in which most nodes are not neighbors, but the neighbors of a node in network are mostly to be neighbors of each other and most nodes can be reached from every other node by a small number of hops. Scale free networks are the mathematical graphs or network which consists of large hubs. Scale free networks mainly consist of power law degree distributions, which power law consists same functions on different scales. As we researched on various research articles and various sizes of data, scale free networks model is suitable to the BGP routing dataset. Because, every day route view BGP data is inconsistent, it varies. So, on every scales, power law consists of same functional forms in scale free networks. Eventually, for our project scale free networks are best for data modeling.

References

[1] Scholz-Reiter B, Rekersbrink H, Go¨ rges M (2010) Dynamic flexible flow shop problems—scheduling heuristics vs. autonomous control. CIRP Annals – Manufacturing Technology 59(1):465–468.

[2] Norstedt JP, Aguren S (1973) The Saab-Scania Report, Swedish Employers' Confederation, Stockholm.

[3] Jovane F, Koren Y, Boer C (2003) Present and future of flexible automation: towards new paradigms. CIRP Annals – Manufacturing Technology 52(2):543–560.

[4] Van Brussel H, Wyns J, Valckenaers P, Bongaerts L, Peeters P (1998) Reference architecture for holonic manufacturing systems: PROSA. Computers in Industry 37(3):255–274.

[5] Butala P, Sluga A (2006) Autonomous work systems in manufacturing networks. CIRP Annals – Manufacturing Technology 55(1):521–524.

[6] Ueda K, Kito T, Fujii N (2006) Modeling biological manufacturing systems with bounded-rational agents. CIRP Annals – Manufacturing Technology 55(1):469–472.

[7] Palla G, Dere´ nyi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818.

[8] Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46(5):323–351.

[9] Baraba´ si AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512.

[10] Fortunato S (2010) Community detection in graphs. Physics Reports 486(3–5):