AN APPROACH TO FIND INFLUENCERS ANALYZING COMPLEX SOCIAL NETWORK

DIPANKAR CHAKI

SUPERVISOR: DR. MOINUL ISLAM ZABER

Inspiring Excellence

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF ENGINEERING & COMPUTER SCIENCE
BRAC UNIVERSITY
MOHAKHALI, DHAKA-1212
BANGLADESH

AN APPROACH TO FIND INFLUENCERS ANALYZING COMPLEX SOCIAL NETWORK

A Thesis submitted in
partial fulfillment of the requirements for the degree of
Master of Science in Computer Science & Engineering of
BRAC University

By

Dipankar Chaki

Supervisor:
Dr. Moinul Islam Zaber

July 2017

© 2017
Dipankar Chaki

All Rights Reserved

DECLARATION

This is to certify that the research work titled "*An Approach to Find Influencers Analyzing Complex Social Network"* is submitted by Dipankar Chaki (ID: 15166004) to the Department of Computer Science & Engineering, BRAC University in partial fulfillment of the requirements for the degree of Master of Science in Computer Science and Engineering. The contents of this thesis have not been submitted elsewhere for the award of any degree. I hereby declare that this thesis is my original work based on the results I have found. The materials of work found by other researchers and sources are properly acknowledged and mentioned by reference. I have carried out my work under the supervision of Dr. Moinul Islam Zaber.

Dated: July 27, 2017

Signature of Author

____

Dipankar Chaki (15166004)

BRAC UNIVERSITY

FINAL READING APPROVAL

Thesis Title: An Approach to Find Influencers Analyzing Complex Social Network.

Date of Submission: 27th of July, 2017

The final form of the thesis report is read and approved by Dr. Moinul Islam Zaber. Its format, citations, and bibliographic style are consistent and acceptable. Its illustrative materials including figures, tables, and charts are in place. The final manuscript is satisfactory and is ready for submission to the Department of Computer Science & Engineering, School of Engineering and Computer Science, BRAC University.

Supervisor

____

Dr. Moinul Islam Zaber
Associate Professor
Department of Computer Science and Engineering
University of Dhaka, Bangladesh

COMMITTEE MEMBERS

Dr. Amitabha Chakrabarty, Committee Chair

Dr. Jia Uddin, Committee Member

Dr. Amin Ahsan Ali, Committee Member

Chairperson
Department of Computer Science and Engineering
BRAC University, Dhaka, Bangladesh

ACKNOWLEDGMENT

At first, I would like to thank the Almighty for providing me with positive mental spirit, wellbeing, courage to carry out this research successfully. My journey towards the Master degree would not have been possible without the help of many people. It is my great pleasure to take this opportunity to thank them for the support and advice that I received.

I am grateful to my thesis supervisor Dr. Moinul Islam Zaber, Associate Professor, Department of Computer Science and Engineering, University of Dhaka, for his inspiration, idea, guidance and overall suggestions to improve this work. He has offered me understanding and support at all stages of my work. His practical suggestions have kept me rooted to reality, and helped me to complete my research work on time.

I also acknowledge the contribution of Mr. Farhad Alam Bhuiyan, Senior Research Associate, Data and Design Lab, Department of Computer Science and Engineering, University of Dhaka. He has helped me in every step grasping the web data scrapping knowhow for my research work, and provided assistance required for the thesis work.

Nevertheless, I would like to thank Department of Computer Science and Engineering, BRAC University and my teachers for helping me with all the necessary support. Last but not the least, I would like to show my heartiest gratitude towards my parents for continuously supporting me throughout my journey and giving me strength to complete this degree.

ABSTRACT

Popularity of social media in Bangladesh is prodigious. 80 percent of internet users are on social networking websites like Facebook, Twitter. That is over 16 million people and counting. The rate of new Facebook users is outpacing the country's birth rate as one new Bangladeshi Facebook account is opened every 20 seconds. This makes social media a great platform for government to reach out to citizens and stay up-to-date with current events and trends in society. That is why, a Facebook group named "Public Service Innovation Bangladesh" has been created. In this group, discussions related to public service innovation, public service related problems and solutions, decision making in administrative works etc. are being prioritized. The focus of this study is to construct complex network from posts given by the members of this Facebook group, analyze features of the complex network including degree distribution, assortative mixing and betweenness centrality. It is important to detect influencers of that Facebook group. We have analyzed group data from January 1, 2016 to June 30, 2017 and generated a report which has given some interesting insights about that group. During this time frame, 5183 posts have been posted and most amazingly, majority of these posts have been posted from November, 2016 to till date. So, it can be said that, this group is growing now. In our constructed network, we have seen that the people who give more posts, get more likes and comments. That is how, they tend to be connected with other highly connected people. If a person who has many connections, gives a post, gets more attention meaning likes and comments than other. Our study helps to understand the structure of this group and finds the influencers of the group.

Index Terms: Complex Network Analysis, Social Network Analysis, Betweenness Centrality, Closeness Centrality, Degree Centrality, Characteristics Path Length

Table of Contents

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

BC:	Betweenness Centrality
CC:	Closeness Centrality
CNA:	Complex Network Analysis
DC:	Degree Centrality
EV:	Eigenvector Centrality
CPL:	Characteristics Path Length
IA:	Information Asymmetry
PAP:	Principal Agent Problem
NV:	Network Visualization
SNA:	Social Network Analysis

LIST OF PUBLICATIONS

Chaki, D., Zaber, M. I. (2017). Understanding Social Complex Network of Government Officials in Decision Making, accepted at CPRSouth Conference, Yangon, Mayanmar. Social Science Research Network (SSRN) Database.

Chaki, D., Zaber, M. I. (2017). An Approach to Find Influencers Analyzing Complex Social Network, submitted at 6th International Conference on Complex Networks and Their Applications, Lyon, France. IEEE Digital Express.

CHAPTER 1

Introduction and Overview

1.1 Introduction

Humans are inclined to form social network from the very beginning of their existence. In the era of technology, social media provides the instruments to expand this network even drastically further. Networks and their interacting nature have lain dormant innumerable information about the behavior of the different groups and the nature of the interaction of groups within the groups. Analyzing complex network getting more and more significant in the social media analysis to identify substantial insight among different groups.

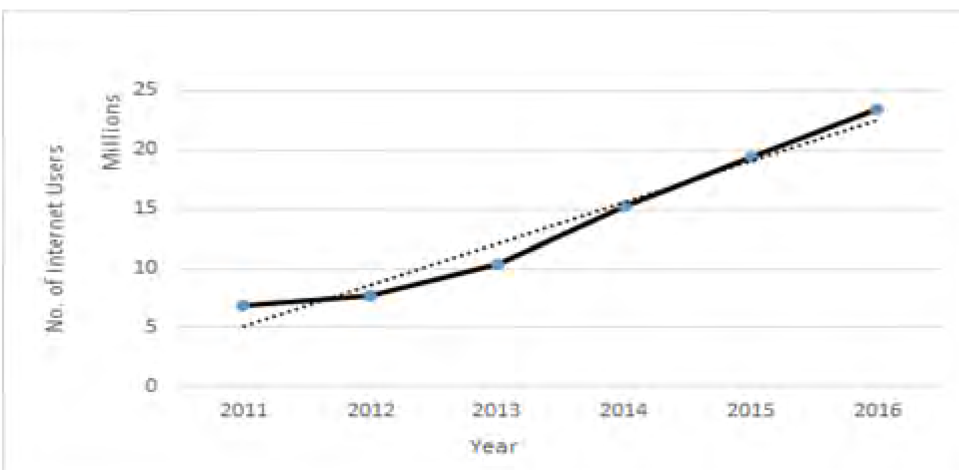


Figure 1.1: Internet users of Bangladesh over the year of 2011 to 2016.
See: http://www.internetlivestats.com/internet-users/bangladesh/

According to Figure 1.1, internet users are proliferating in Bangladesh. In the time period of 2011 to 2016, the number of internet users are increasing in such a way that it has become almost 4 times. The increasing numbers denoting the increasing engagement in internet and social media which leading towards more interactions with the peers contributing in forming complex network which is highly in need of research to fathom the network.

1.2 Problem Statement

Principal Agent Problem (PAP) is one of the crucial issues in any government. PAP (Pi-ji, C. H., 2000) means in an environment where the agents are obliged to act according to the requirements established by the principle and the environment is built by the principles. It supports the idea that the agent's motivations won't get along with the principles one. Suppose there is a restaurant and there is a social media like food bank. Here restaurant is "principles" and the food bank is "agents". If the food bank gives low rating on the restaurant, then it would be loss for the principles. So it has an incentive that the agency should give higher rating than what they may deserve on behalf of the principles. Another issue hinders the progress of the entire government is Asymmetric connectivity. Asymmetric means imbalance. Information asymmetric means that there is two group such as seller and buyer where the seller has more information/knowledge regarding the goods than the buyer. And that creates an imbalance knowledge between two sides. It indicates an imbalance power in any activity such as it could lead to a "market failure". It could be an advantage for a company in the market to lead its company in the right tracks. Suppose in a classroom the teacher definitely has more knowledge about the subject than the students.

Principle agent problem points out information asymmetry (Pi-ji, C. H., 2000) but it creates an uncertainty that indicates to connectivity problem. Due to less connectivity, principle agent problem gets occurred. Information asymmetry also creates a huge gap between two groups of people that leads to connectivity problem as well. There are usually one of two kinds of connectivity situated in the government; one of them is vertical and another one is horizontal. The connectivity exists in Bangladesh government is vertical as shown in Figure 1.2. The problem with the vertical connectivity is lower level of the government officials get restricted from decision making process. One of the solutions of the connectivity problem could be the platform of social media where all the officials can participate in directly or indirectly in decision making by expressing their knowledge and ideas which might significantly address the vertical connectivity problem result in enforce the efficiency in the government. There is an organization named as Organization for Economic Co-operation and Development (OECD). Member countries of this organization deliver better public services via ICT based platforms such as institutional websites, social media and public service related apps. There are 34 OECD member countries and among them, 26 countries have their own Facebook page and twitter account for various ministries (Mickoleit, 2014). Lansing is the capital of the US state of Michigan and they have their own verified twitter account for better public service.

Figure 1.2: Organogram of Ministry of Public Administration in Bangladesh (simplified*).
*Detailed organogram of Bangladesh government found at http://www.mopa.gov.bd/en/home/content/1/1/4

Bangladesh government has created a Facebook group of government officials and to make it more effective platform cabinet division of People's Republic of Bangladesh enforced a propaganda to join all the government officials in this Facebook group. Popularity of social media in Bangladesh is prodigious. 80 percent of internet users are on social networking websites like Facebook, Twitter. Bangladeshi Facebook account is opened in every 20 seconds. This makes social media a great platform for government to reach out to citizens and stay up-to-date with current events and trends in society. That is why, a Facebook group named "Public Service Innovation Bangladesh" has been created. This Facebook group network is not closely connected. There are only few people give posts and statuses and others just comment or like that person's post. So, those who publish more posts, we call them leaders or influential person of that group. Also, in this group, discussions related to public service innovation, public service related problems and solutions, decision making in administrative works etc. are being prioritized. Following image displays how this Facebook group is working as a common platform amongst government officials.

Figure 1.3: Screenshot of a post regarding a problem and get the immediate attention of an appropriate government official.

Figure 1.3 shows that one lower rank official post a problem statement on 26th may 2017 which indicate that one primary school was in under construction for 18 years and never been completed. The commissioner, highest ranked official of a district, of that district has seen the statement and commented on 27th May,2017, with all the dispatches, he has assigned an UNO (chief official of a sub-district) to look into that matter to solve the problem as soon as possible. The problem which was in the curse of red tape for 18 years solved in 1 day. It has only been possible to contact with a high rank official through this Facebook group platform. So, relation between governance and connectivity is created. Unlike vertical relation, it is important to have a horizontal relation between governance and connectivity. Vertical relationship causes the failure of planning commission. Hence, this Facebook group is creating the horizontal relation.

1.3 Aim of this Study

Our study focuses on the understanding of the complex network of this group of governmental officials based on their interactions in the group. To better understand the interaction network and to fathom its nature in the group we have analyzed different metrics of the complex network such as closeness centrality, degree distribution, average path length and betweenness centrality and other parameters of the network. Betweenness centrality is a measure of the node's influence in the network showing how often a node appears on shortest paths between any two randomly selected nodes in the network (Brandes U., 2001). Eigenvector centrality is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes (Ruhnau B., 2000). Eccentricity is the distance from a given starting node to the farthest node from it in the network (Hage, P., &Harary, F., 1995). Influence is a mental state which has notable impact on decision making. In any governmental network, influence can play even stronger role in decision making and, in some extent, in governing the country. In our study we try to analyze whether an influence network existed in the mentioned group.

1.4 Thesis Outline

The rest of the thesis report is organized as below:

- In chapter 2, complex network definition, architecture and previous works related to complex networks have been discussed.
- In chapter 3, data description along with data acquisition, data preparation, preparing graph format files, etc. have been explained briefly.
- Chapter 4 represents research methodologies, statistical analysis, network analysis, results and findings.
- Chapter 5 concludes the paper with general remarks and future works.

CHAPTER 2

Background

2.1 Introduction

The study of networks pervades all of sciences, from neurobiology to statistical physics. The most basic issues are structural: how does one characterize the wiring diagram of a food web or the internet or the metabolic network of the bacterium Escherichia coli? Are there any unifying principles underlying their topology? Scientists have been conducting researches to find the answers of these questions by analyzing complex networks.

2.2 Complex Network Definition

In the context of network theory, a complex network is a graph (network) with non-trivial topological features—features that do not occur in simple networks such as lattices or random graphs but often occur in graphs modelling of real systems (Complex Network (2000)). The study of complex networks is a young and active area of scientific research (since 2000) inspired largely by the empirical study of real-world networks such as computer networks, technological networks, brain networks and social networks. Network anatomy is so important to characterize. Because structure always affects function. For instance, the topology of social networks affects the spread of information and disease, and the topology of the power grid affects the robustness and stability of power transmission (Strogatz, S. (2001)). From this perspective, the current interest in networks is part of a broader movement towards research on complex systems. In the words of E. O. Wilson, "The greatest challenge today, not just in cell biology and ecology but in all of science, is the accurate and complete description of complex systems. Scientists have broken down many kinds of systems. They think they know most of the elements and forces. The next task is to reassemble them, at least in mathematical models that capture the key properties of the entire ensembles."

2.3 Complex Network Architectures

There are many regular networks available such as chains, grids, lattices and fully-connected graphs (Figure 2.1). Those simple architectures allowed us to focus on the complexity caused by the nonlinear dynamics of the nodes.

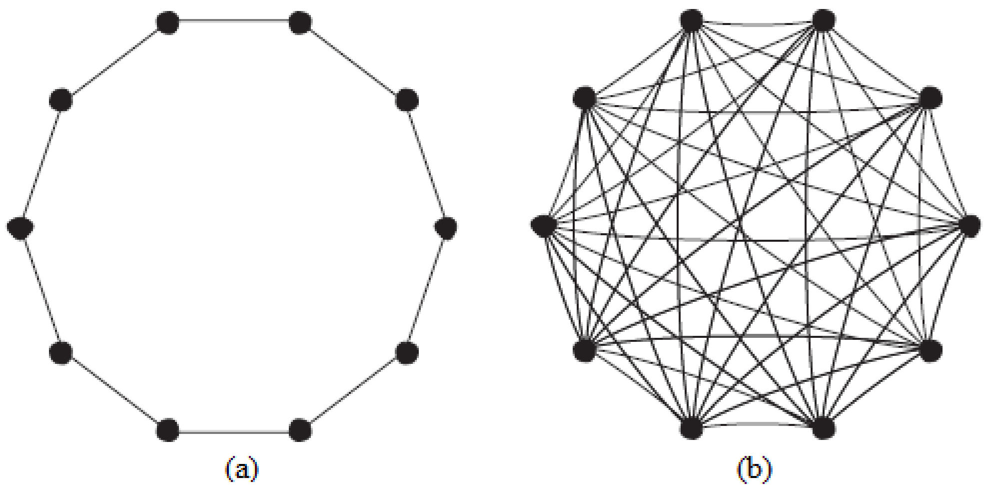


Figure 2.1: Schematic illustration of regular network architectures. (a) Ring of ten nodes connected to their nearest neighbors. (b) Fully connected network of ten nodes.

2.3.1 Random Graphs

Imagine $n>1$ buttons strewn across the floor (Kauffman, S. (1995)). Pick two buttons at random and tie them together with thread. Repeat this process m times, always choosing pairs of buttons at random. (If m is large, you might eventually select buttons that already have threads attached. That is certainly allowed; it merely creates clusters of connected buttons.) The result is a physical example of a random graph with n nodes and m links. Now slowly lift a random button off the floor. If it is tied to other buttons, either directly or indirectly, those are dragged up too. Erdös, P. and Rényi, A. (1960) studied how the expected topology of this random graph changes as a function of m. When m is small, the graph is likely to be fragmented into many small clusters of nodes, called components. As m increases, the components grow, at first by linking to isolated nodes and later by coalescing with other components. Furthermore, all nodes in the giant component are connected to each other by short paths.

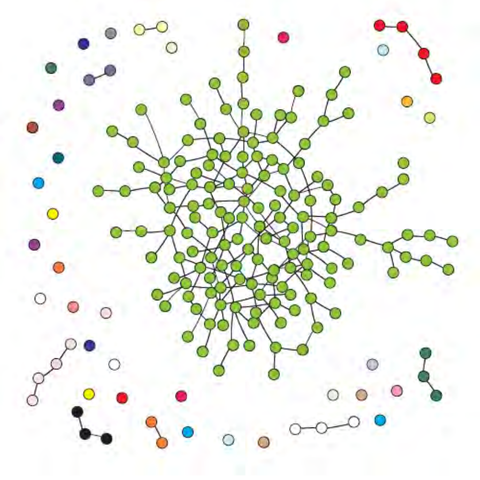


Figure 2.2: Sample of a random graph.

Figure 2.2 depicts a random graph, constructed by placing n nodes on a place. then joining pairs of them together at random until m links are used. Nodes may be chosen more than once, or not at all. To clarify it, I have segregated the different connected components, colored them. The main topological features are the presence of a single giant component.

2.3.2 Small World Networks

Although regular networks and random graphs are both useful idealizations, many real networks lie somewhere between the extremes of order and randomness (Figure 2.3). Watts, D. J. and Strogatz S. H. (1998) studied a simple model that can be tuned through this middle ground: a regular lattice where the original links are replaced by random ones with some probability 0 ~ 1. They found that the slightest bit of rewiring transforms the network into a 'small world', with short paths between any two nodes, just as in the giant component of a random graph. Yet the network is much more highly clustered than a random graph, in the sense that if A is linked to B and B is linked to C, there is a greatly increased probability that A will also be linked to C (a property that sociologists call 'transitivity') (Wasserman, S. and Faust. K. (1994)).

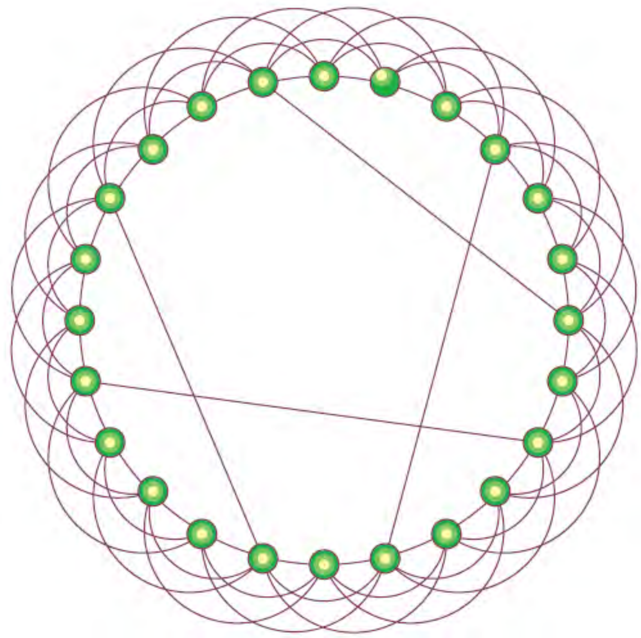


Figure 2.3: Solvable model of a small-world network. The model starts with a ring lattice of n nodes, each connected to its neighbors out to some range k (here n = 24 and k = 3).

2.3.3 Scale Free Networks

In any real network, some nodes are more highly connected than others are. To quantify this effect, let Pk denotes the fraction of nodes that have k links. Here k is called the degree and Pk is the degree distribution. For many real networks, Pk is highly skewed and decays much more slowly than a Poisson. For instance, the distribution decays as a power law $Pk \sim k^{-y}$ for the Internet backbone (Faloutsos, M. et. al (1999)), metabolic reaction networks (Jeong H. et. al (2000)), the telephone call graph (Abello, J. et. al (1998)) and the World-Wide Web (Abello, J. (1998)). Figure 2.4 demonstrates a scale free network, grown by attaching new nodes at random to previously existing nodes. The probability of attachment is proportional to the degree of the target node; thus richly connected nodes tend to get richer, leading to the formation of hubs and a skewed degree distribution with a heavy tail. Colors indicate the three nodes with the most links (red, k = 33 links; blue, k = 12; green, k = 11). Here n = 200 nodes, m = 199 links. Figure provided by D. Callaway. Network visualization was done using the Pajek program for large network analysis.

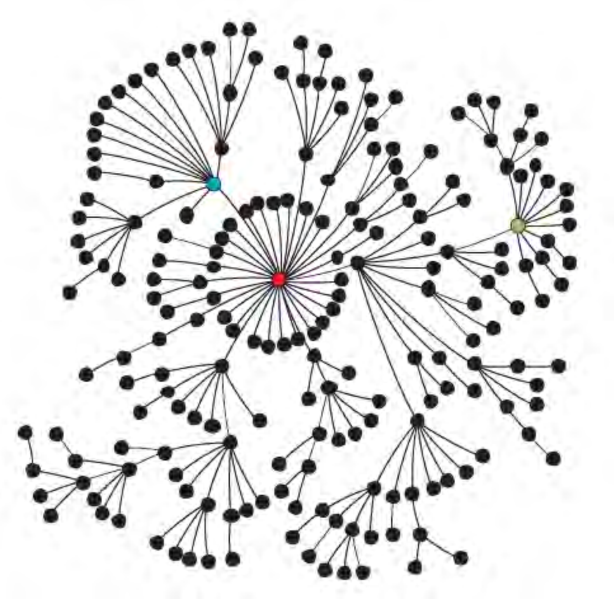


Figure 2.4: Example of a scale-free graph.

2.4 Previous Works Related to Complex Network

Past studies have documented the impact of complex network analysis to understand structure of several governance systems. Analysis of complex network is carried out in the study of Renfro and Dekro (2001) based on interaction and other official relation and chain of command. They have also located the influence network within the government and mention the most influential person in the government. Identifying the most influential nodes in social networks is a key problem in social network analysis. In another study, Ilyas and Radha (2011) have described a way to identify the influential nodes in social media using different complex network techniques such as measuring Eigen vector and principal component centrality. Besides these, other forms of networks do exist like terrorist network, traveler's online community network, scientists' collaboration network and so on.

Dynamic Network Analysis was used as to find the approaches to assessing destabilization tactics of terrorist network (Carley et. Al., 2004). Moon and Carley (2007) have analyzed the terrorist network and the geospatial dimensions by investigating the networks based on their interaction, task location and regional knowledge. They have measured relative similarity and relative expertise of the network along with probability of the interaction to identify terrorist network among other networks. Figure 2.5 portrays terrorist network (Moon & Carley, 2007), travelers network (Social Networks and Travel Behavior, 2015) and scientists' collaboration network (Newman M. E., 2001). Some sample pictures of several networks are given below. From the below figure, it can be said that terrorist network (a) is closely connected, travelers network (b) is clustered connected and scientists' collaboration network (c) is ring connected. It means that, terrorists are narrowly connected to themselves whereas travelers are connected community or group wise. On the other hand, in Figure 2.5 (c), the point in the center of the figure denotes the author of the paper, the centered ring represents his collaborators, and the second ring demonstrates their collaborators. Collaborative bonds between members of the same ring, of which there are numerous, have been omitted from the figure for clarity (Newman M. E., 2001).

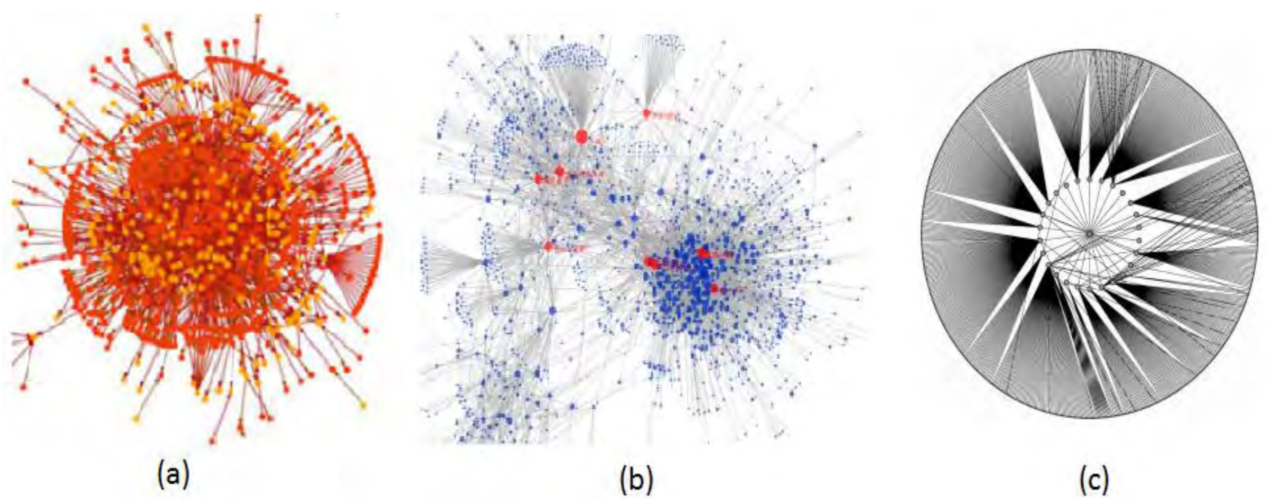


Figure 2.5: (a) Terrorists Network. (b) Travelers Network. (c) Scientists Collaboration Network.

Christakis and Fowler (2008) analyzed the smoking network. They have analyzed the graphs using the Kamada-Kawai algorithm, in addition with, different centrality metric such as eigenvector and clustering techniques to compare the networks for different smoking behaviors. Costa et.at (2008) discussed about the techniques to analyze real world phenomenon using complex network such as music, sports, sexual relation, collaboration, religion and many more.

CHAPTER 3

Data Description and Network Construction

3.1 Introduction

The first step in data analysis is to collect data and improve data quality. Data scientists correct spelling mistakes, handle missing data and weed out nonsense information. This is the most critical step in the data value chain - even with the best analysis, junk data will generate wrong results and mislead the research. This chapter demonstrates briefly regarding data acquisition and data preparation process for network analysis. The last section in this chapter represents data description.

3.2 Data Description

At first, we have collected data from a Facebook group named as "Public Service Innovation Bangladesh". Currently, there are 14500+ members in that group and counting. All members in that group are government officers from different ministries and departments. A picture of that Facebook group is displayed below (Figure 3.1). Top bureaucrats like the Cabinet Secretary and Principal Secretary are not only writing posts on topics like innovation but even encouraging their colleagues, particularly those at the field level who are closer to citizens, to challenge their propositions and have online debates through comments on Facebook to decide on the best course of action. Acquiring data from Facebook is very challenging due it's security issues. However, we have managed to collect data by web scraping. Web Scraping is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in the computer or to a database in table (spreadsheet) format (Web Harvy (2017)).

Figure 3.1: Cover picture of Public Service Innovation Bangladesh Facebook group.

At first, we have collected all the data available in that group between October 13th, 2013 and June 30th, 2017. Later, we have calculated year wise mean number of posts per month and found out that the number of posts are increasing rapidly from the year 2016 (Figure 3.2). This is why, this research focuses on the data from January 1st, 2016 to June 30th, 2017.

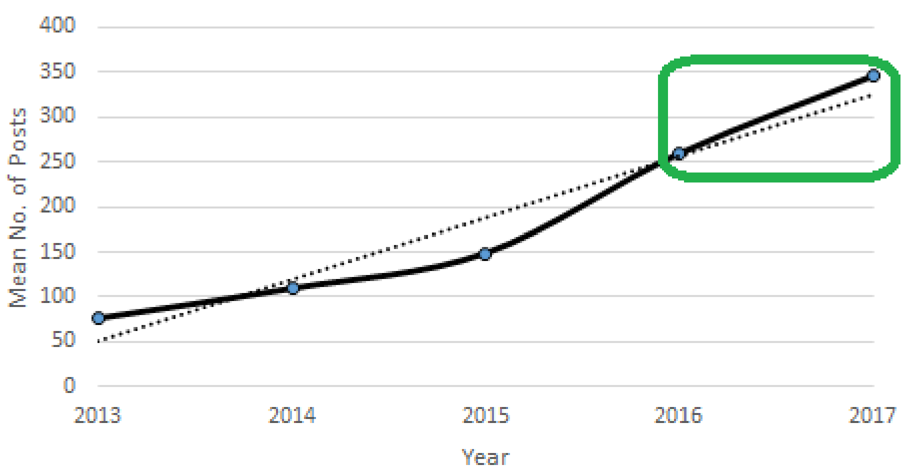


Figure 3.2: Year wise mean number of posts per month. This research focuses on the data between January 1st, 2016 and June 30th, 2017.

In order to conduct this research, we have gathered data of 5183 posts published in that group between January 1, 2016 and June 30, 2017. Among 14500+ members, 1571 members posted at least 1 post in that group. Figure 3.3 depicts the screenshot of the dataset. Obviously, names of post givers are anonymized for security reason. Table 3.1 (appendix A) gives the list of attributes of our dataset. From this dataset, statistical analysis has been conducted. Doing some exploratory analysis, we have found few interesting insights of this group which are being discussed in chapter 4 statistical analysis section.

	us_message	status_author	link_name	status_type	status_link	status_published	num_r	num_cor	num_sh	num_	num_l	num_wo	num_ha	num_sac	num_angrys
	5_736794066522483	[redacted]		status		2017-07-23 8:11:27	1	0	0	1	0	0	0	0	0
1721772696	https://m.facebook.com/story.p	[redacted]	ইউএনও জনাব গাজী তা	link	https://dmpn	2017-07-23 8:11:15	6	0	1	6	0	0	0	0	0
1721772696	Innovation Workshop	[redacted]	Photos from Say	photo	https://www.f	2017-07-23 6:47:11	10	0	0	10	0	0	0	0	0
1721772696	নবীনগর এই কর্মতৎপর uno মহোদয় এর	[redacted]		status		2017-07-23 6:47:00	4	0	0	4	0	0	0	0	0
1721772696	প্রবল বর্ষনে উপ-স্বাস্থ্য কেন্দ্রটি টই টুম্বর৷ উ	[redacted]	Photos from Mith	photo	https://www.f	2017-07-23 6:46:26	5	3	0	5	0	0	0	0	0
1721772696	কাপ্তাই হ্রদে কার্প প্রজাতির মাছের প্রজনন	[redacted]	কাপ্তাই হ্রদে ১৪ বছর প	link	http://www.pr	2017-07-23 6:45:59	11	0	1	11	0	0	0	0	0
1721772696	ডিজিটাল পানি উন্নয়ন বোর্ড গঠনে করনীয়	[redacted]		status		2017-07-23 6:45:27	1	0	0	1	0	0	0	0	0
1721772696	আজ রবিবার ২৩শে জুলাই,২০১৭খ্রিঃ 'জাতী	[redacted]	Photos from Abja	photo	https://www.f	2017-07-23 6:44:57	13	1	1	13	0	0	0	0	0
1721772696	১৮ বছর বা তদুর্ধ বয়সের যাদের জাতীয় প	[redacted]		status		2017-07-23 6:44:40	5	0	0	5	0	0	0	0	0
1721772696	সরকারি দপ্তরসমূহে সেবা সহজিকরণে নানা	[redacted]		status		2017-07-23 6:44:01	7	1	0	7	0	0	0	0	0
17217726950835_736718659863357		[redacted]	ইউএনও তারিক সালমন	link	http://www.ka	2017-07-22 10:46:12	183	10	10	174	3	0	6	0	0
1721772696	https://m.facebook.com/story.p	[redacted]	DC Office, Kusht	photo	https://www.f	2017-07-22 7:49:47	37	2	1	37	0	0	0	0	0
17217726950835_736577476544142		[redacted]		status		2017-07-22 6:53:07	5	0	0	5	0	0	0	0	0
1721772696	https://m.facebook.com/story.p	[redacted]	Bangladesh Polic	photo	https://www.f	2017-07-22 6:52:35	14	0	1	14	0	0	0	0	0
1721772696	এই জোড় পূর্বক টাকা আদায়কে আমি কি ব	[redacted]		photo	https://www.f	2017-07-22 6:52:16	99	23	0	79	0	1	1	9	9
17217726950835_736589963209560		[redacted]		status		2017-07-22 6:52:03	7	0	0	7	0	0	0	0	0
1721772696	কৃষি শ্রমিক কার্ড-কৃষি শ্রমিক সমস্যা সমাধা	[redacted]		status		2017-07-21 23:35:47	11	0	1	11	0	0	0	0	0
1721772696	আনসার ভিডিপি ডাটাব্যাজ ইনোভেশন প্রক	[redacted]	projects.gov	link	http://projects	2017-07-21 23:35:41	6	0	0	6	0	0	0	0	0
17217726950835_736325159902707		[redacted]		status		2017-07-21 23:35:24	17	0	0	16	1	0	0	0	0
1721772696	https://m.facebook.com/story.p	[redacted]		status		2017-07-21 23:35:18	2	0	0	2	0	0	0	0	0

Figure 3.3: Screenshot sample of our dataset (names are being blurred).

3.3 Network Construction

From our dataset, we have prepared 2 graph formatted files for network analysis. One network file is for bipartite graph where nodes are either posts or users and another network file is for mono-partite graph which mainly represents interaction between users. In this research, only interaction network is used.

The network data set is composed of an n x n matrix M, where n is the number of nodes in the analysis. For our research, we have used such a network where all the nodes are users and edges represent the linkage between users. We have 3155 nodes and 11744 edges in our network. If one user publishes a post and another user likes or comments or gives reaction to that post, then, we consider them as connected users. Edge weight encodes the number of reactions or comments a user makes on another user's post(s). For instance, M_{ij} could be the connection between a node i (Mr. X) and j (Mr. Y). The connections can be directed or undirected. The directed edge from one node (i) to another node (j) depicts that node i gives likes or comments or reactions to node j's post. In cases when the origin and the destination of the likes/comments/reactions are differentiated, M is assumed to be asymmetrical ($M_{ij} \neq M_{ji}$). An undirected edge discards the information of who gives likes/comments/reactions to which post. Thus, M is assumed to be symmetrical ($M_{ij} = M_{ji}$). For this reason, in our research, we consider M to be asymmetrical. In order to understand the structure of the network appropriately, various statistical methods are applied to the network M. In this paper we have used standard network analysis, cluster analysis (Aldenderfer 1984) to describe the structure of the Facebook group network.

Figure 3.4 represents our interaction network. Here, in this picture, Node A publishes Post A. Node B, C and D either like or comment on that post A. Thus, Node B, C and D is connected to Node A. If Node C only likes post A, then, weight of edge between C and A is 1. If Node D likes post A and gives 2 comments, then, weight of edge between D and A is 3 (1 like + 2 comments). On the other hand, Node B publishes Post B and Node A likes that post B. In that case, edge weight of Node A and B is 2 (Node A gives 1 like to Node B's post + Node B gives 1 like to Node A's post). Again, Node D is not directly connected with Node B. But, it is connected with Node B via Node A. Since, Node C gives likes or comments to Post B, Node C and D is directly connected. Each node consists of several attributes.

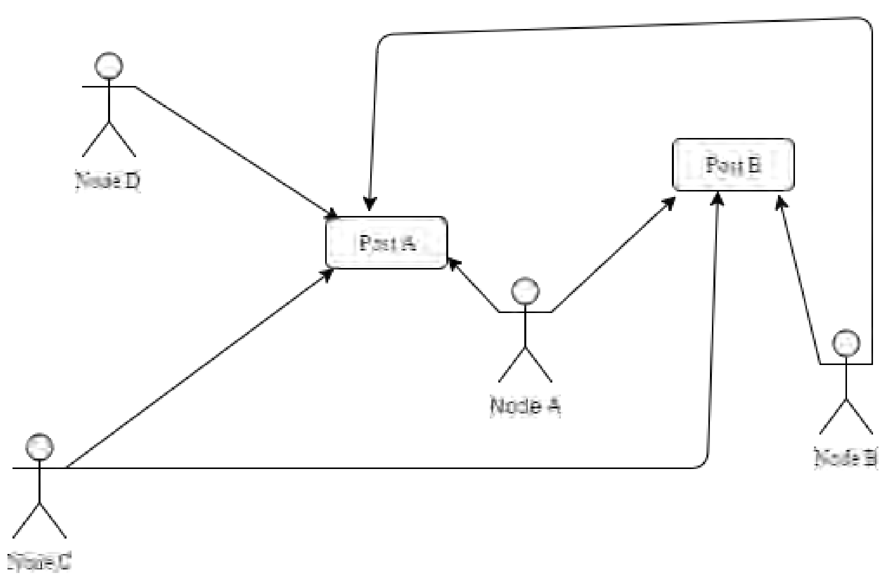


Figure 3.4: Sample picture of Facebook group's interaction network.

CHAPTER 4

Statistical Analysis, Network Analysis and Experimental Result

4.1 Introduction

This chapter comprises of statistical analysis, network analysis procedure, network evaluation and experimental results. The Public Service Innovation Bangladesh Facebook group's network has been analyzed to better understand its characteristics. For this, we measure several complex network features including average shortest path, degree distribution, assortative mixing, betweenness centrality, closeness centrality along with basic statistical analysis.

4.2 Statistical Analysis

At the beginning of our network analysis, we have done some statistical analysis to get the insight of this group. We have analyzed how this group is being interacted through its members, what types of posts have been published, posts' and comments' engagement, how posts and comments are distributed, etc. Figure 4.1 represents the group interactions of our Facebook group. This group is interacted mainly via liking (79%) and commenting (14%). This means that, in our 5183 posts, there are 122236 likes (79%) and 20855 comments (14%).

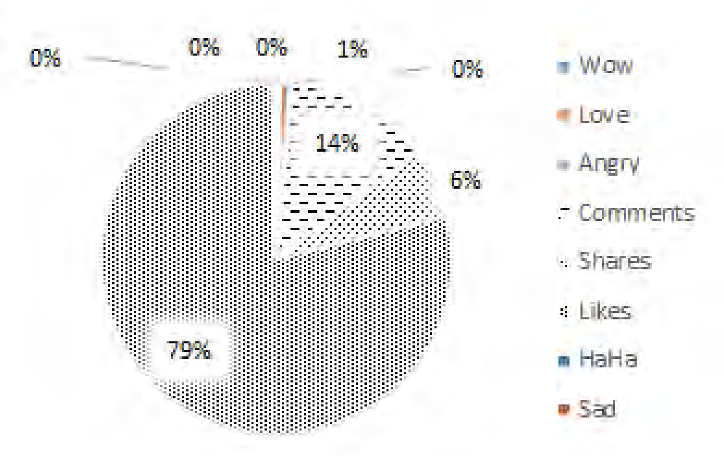


Figure 4.1: Pie chart of group interactions based on posts.

Members of this group who are actually government officers, mostly post photos (46%), statuses (36%), links (12%) and videos (5%). Figure 4.2 represents various types of posts published in the group.

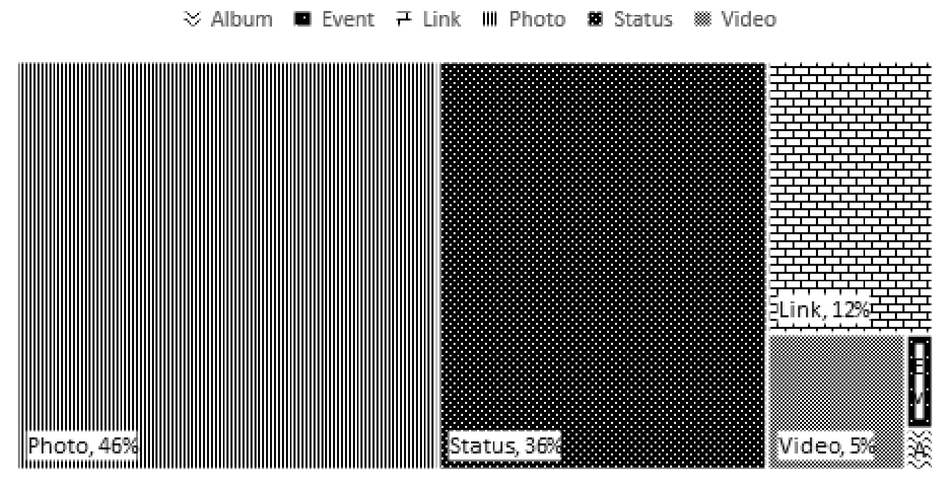


Figure 4.2: Various types of posts in Public Service Innovation Facebook group.

This group is getting popular day by day. Number of posts have been increased than earlier. Trend-line shows that number of posts are growing. Figure 4.3 denotes month wise mean number of posts per day of this group. Figure denotes that during November 2016 to January 2017, highest number of posts have been published. From January 2016 to June 2017, number of posts have been increased almost double. However, in recent months, from January 2017 to June 2017, number of

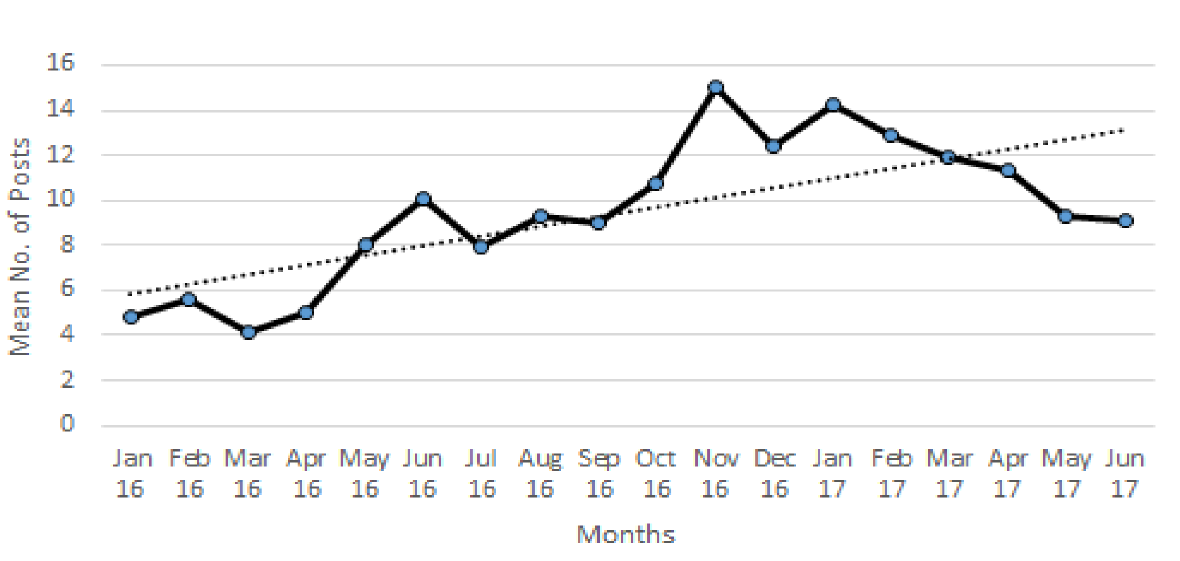


Figure 4.3: Month wise mean number of posts per day.

posts are decreasing which might indicate that members of this group are becoming less active. Post engagement is moderately good in this group. Post engagement occurs when a post is being liked or commented or shared or reacted. Figure 4.4 suggests that, 99.65% posts (5165 out of 5183 posts) have been reacted or commented or shared. If we consider individual category, then, we can see 65.55% posts have been commented. However, only 35.92% posts have been shared.

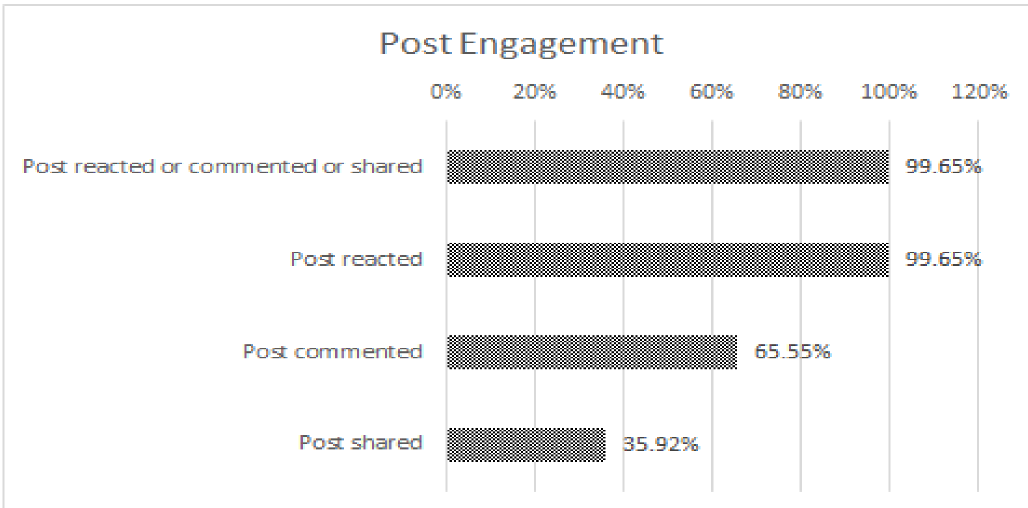


Figure 4.4: Engagement of Posts in Public Service Innovation Facebook group.

Another interesting fact is, this group is mainly active in terms of posts between Mondays and Wednesdays. Figure 4.5 symbolizes post distribution by day. Majority of posts have been submitted in the group on Mondays and minority of posts have been submitted on Sundays.

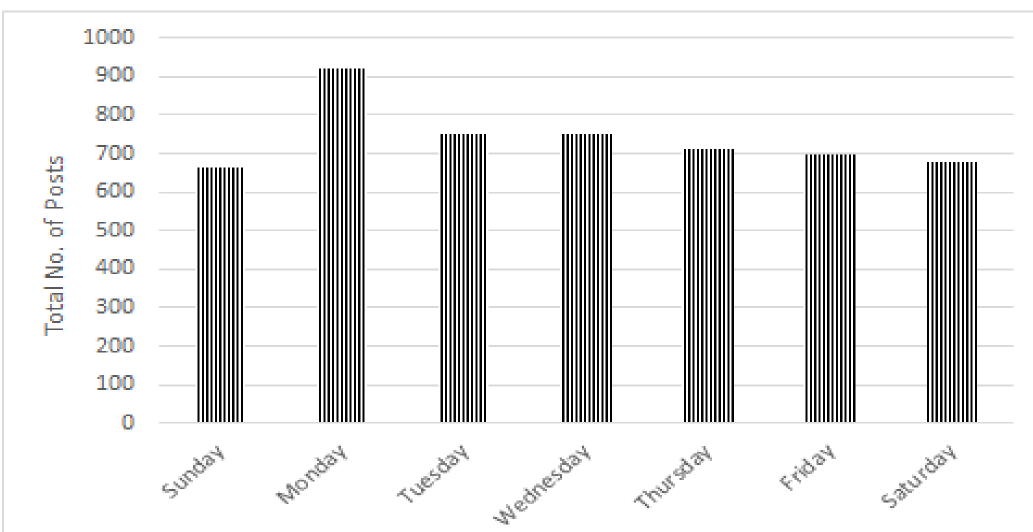


Figure 4.5: Day wise post distribution in Public Service Innovation Facebook group.

Government officials are more active in this group between 8 PM and 11 PM. Among last 5183 posts, 1518 posts have been submitted during this hour. Figure 4.6 shows hour wise post distribution in the group.

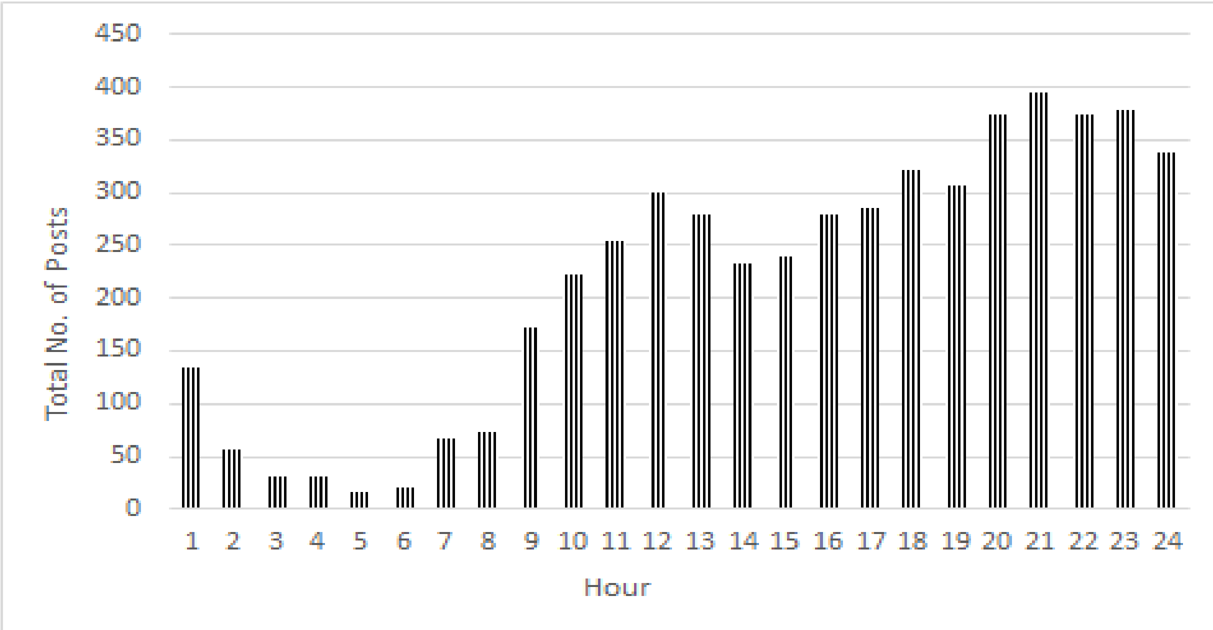


Figure 4.6: Hour wise post distribution in Public Service Innovation Facebook group.

We have analyzed the comment engagement in this Facebook group to understand how much comments each post gets, how much likes each comment gets and so on. In order to do that, we have got Table 4.1. Collected 5183 posts got 20855 comments. We observe that each post has on average 4.02 comments and each comment has on average 1.54 likes.

Table 4.1: Comment Engagement in Public Service Innovation Facebook Group

Total Number of Comments	**Average Comments Per Post**	**Total Number of Likes on Comments**	**Average Likes Per Comment**
20855	4.02	25543	1.94

We have also examined comment distribution by day and hour. Figure 4.7 shows comment distribution by day which suggests that, the group is mainly active in terms of comments on Saturdays, Sundays, Mondays. Majority of the comments have been given on Saturdays and minority of the comments have been given on Fridays. In Bangladesh, Friday is the government holiday. That might be the reason behind less social networking site engagement on Fridays.

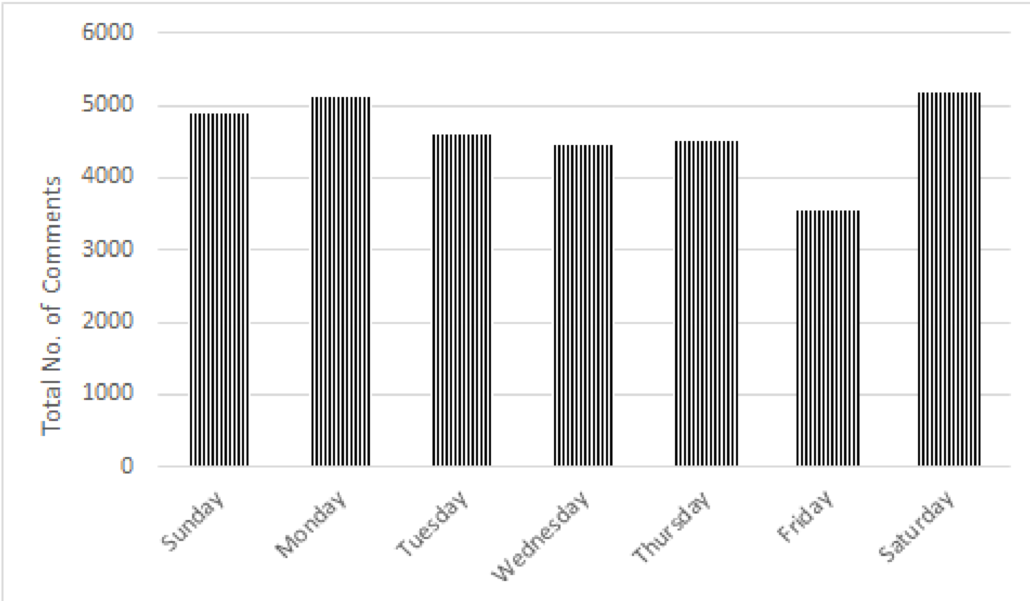


Figure 4.7: Day wise comment distribution in Public Service Innovation Facebook group.

Government officials are more active in this group between 10 PM and 12 AM. Among last 20855 comments, 8354 comments have been submitted during this hour. Figure 4.8 shows hour wise comment distribution in the group.

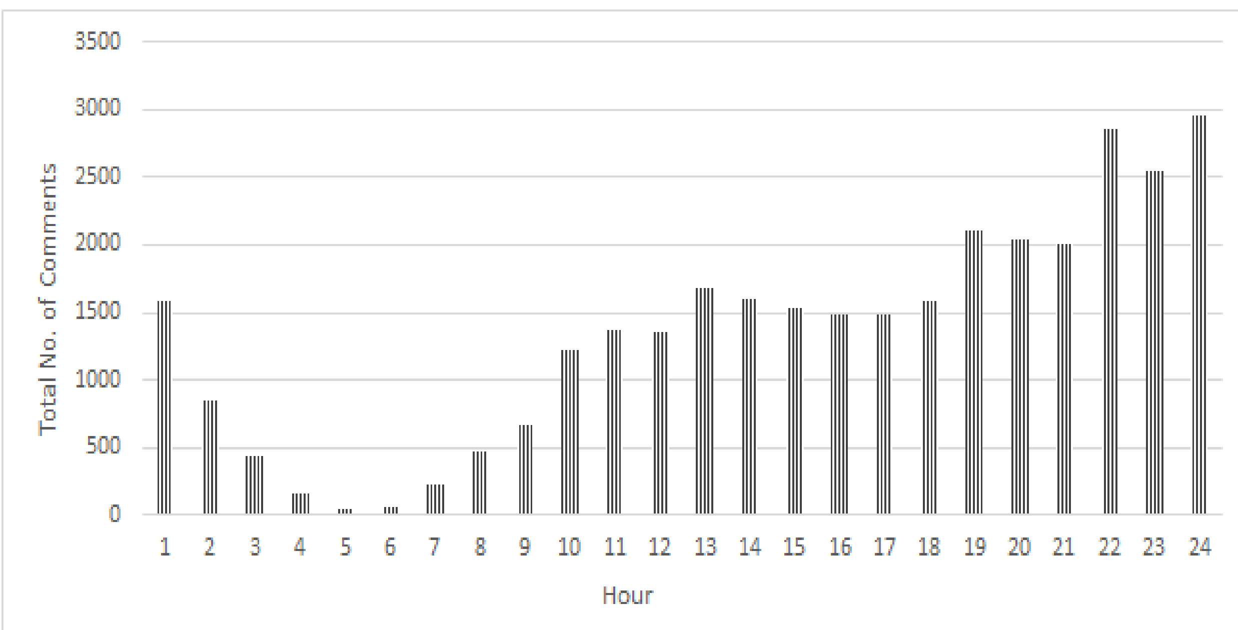


Figure 4.8: Hour wise comment distribution in Public Service Innovation Facebook group.

Finally, we have analyzed members' engagement in that group. Surprisingly, we have found out that, there are 14500+ members in the group, but, only 1571 members (10.83%) have put at least 1 post and 5891 members (40.62%) never posted or commented or reacted in a single post. We call them inactive members or loners in the group. Table 4.2 gives more illustration of members' engagement.

Table 4.2: Members' Engagement in Public Service Innovation Facebook Group

Activity	Numbers	Percentage (%)
Total Members	14502	100%
Unique Members Who Gave At Least 1 Post	1571	10.83%
Unique Members Who Gave At Least 1 Comment	3141	21.66%
Unique Members Who Gave At Least 1 Reaction	8609	59.36%
Members Who Never Posted or Commented or Reacted	5891	40.62%

4.3 Network Analysis

The objective of this research is to identify the structure and key influences of this government official's Facebook group. We have started our research from the two key questions. One is, how does the structure of this group affect the capability to fruitfully accomplish the objectives of this group. And, the second one is, what kind of people play the most vital role in this group. Visualizing the group as a network and ascertaining the utmost leading communities and the most influential members can answer both of these questions. Additionally, it can specify the main points of entrance: the people and clusters within the communities that should be addressed in order to communicate a message to the group in a fast and efficient manner.

4.3.1 Analysis Procedure

We have used ORA (ORA-NetScene software, CASOS (2017)), a complex network analysis and visualization tool for our analysis. Then, we have used a layout algorithm which helps to find connected nodes and clusters. When running this algorithm, it pushes non-attracted nodes away and it brings attracted nodes closer. The principal is easy, linked nodes attract each other and non-linked nodes are pushed apart (Jacomy M., et. Al 2014). After that, we have calculated average degree which represents number of connections each node has to other nodes.

Then, we have run some statistics functions to uncover our network properties. At the beginning, we have calculated the average path length for the network. It computes the path length for all possible pairs of nodes and give information about how nodes are close from each other. Later, we have calculated betweenness centrality, closeness centrality, eccentricity distribution. Finally, we calculate graph density and average clustering co-efficient. The average clustering coefficient distribution gives the average of the clustering coefficients for all nodes n with k neighbors for k = 2.

4.3.2 Experimental Result

Figure 4.9 depicts the structure of our Facebook group network. The structure of this network seems similar to both the core periphery network (Csermely, P. et. Al, 2013) and the scale free network (Caldarelli, G., 2007). The structure has a dense core and many periphery nodes (which are connected to the core and have no connections among themselves). On the other hand, the structure also seems kin to the scale free network with most nodes are connected to few other nodes and few nodes are connected to many other nodes.

Figure 4.10 represents the zoom in version of our interaction network. This picture clearly suggests that out network has dense core and many periphery nodes. Everyone is not connected in this network. Few are connected with most of the nodes. Only few members publish posts, likes, comments, shares and so on. It means some are more eager to connect than others. Some are more influential in terms of posts/likes/comments than others in that group. The characteristic path length (the number of people to reach from one randomly selected member to another) is 3.940 and the clustering co-efficient is 0.446. So, the clusters in the group are pretty well connected. It means to have a pattern in the group which is, there is a certain group of people who are regularly give posts and there is a certain group of people who are regularly give comments and likes.

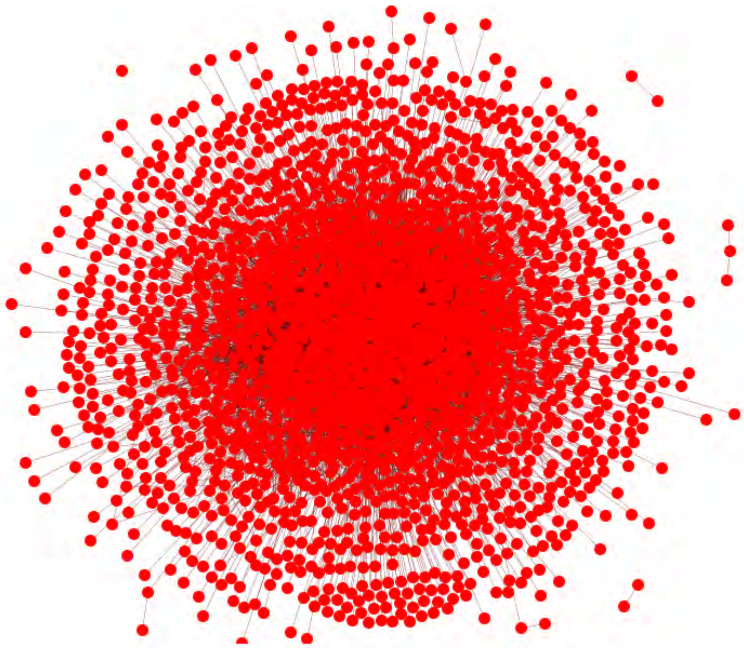


Figure 4.9: Interaction network of Public Service Innovation Bangladesh Facebook group. The structure of this network seems similar to both core periphery network and scale free network.

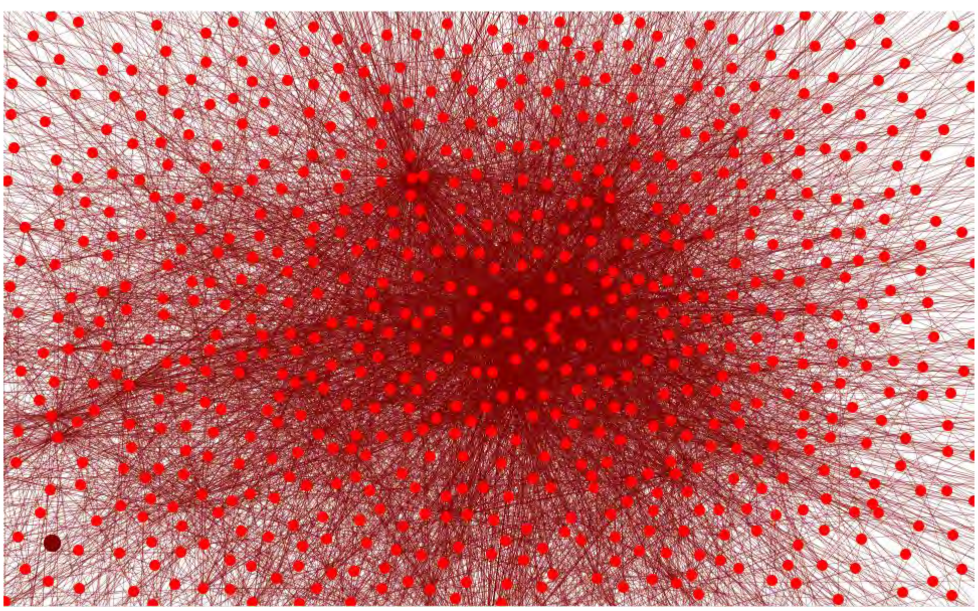


Figure 4.10: Dense core and many periphery nodes (zoom in version of interaction network).

Then, we have calculated network and node level measures for our network. We have found out that, in terms of eigenvector centralization, the Facebook group network is also tending to be very high. This shows that government officer's Facebook group network is a highly-clustered network with nodes of high eigenvector centrality. In terms of betweenness centrality this network is found to have the low betweenness. Table 4.3 represents various network level and node level measures of the network generated from that Facebook group.

Table 4.3: Network and Node Level Measures of Public Service Innovation Facebook Group

Network and Node Level Measures	**Values**
Characteristic Path Length	3.940
Clustering Coefficient	0.446
Network Assortativity	-1.031
Network Reciprocity	0.019
Betweenness Centralization	0.104
Eigenvector Centralization	0.909
Closeness Centralization	0.307
Total Degree Centralization	0.002

Closeness centrality is a measure that indicates how close a node is to all the other nodes in a network, whether or not the node lays on a shortest path between other nodes (Borgatti, S. P., 2005). A high closeness centrality means that there is a large average distance to other nodes in

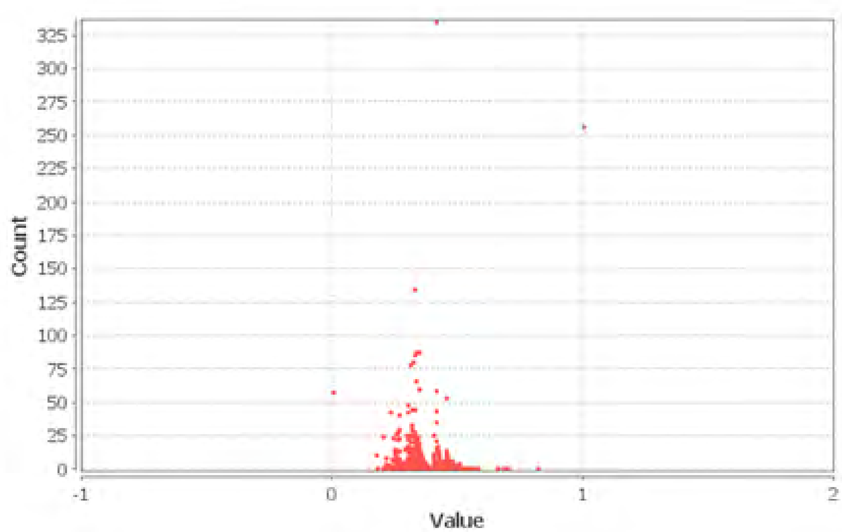


Figure 4.11: Closeness centrality distribution of Public Service Innovation Facebook group.

the network. So, a small closeness centrality means there is a short average distance to all other nodes in the network. Average closeness centrality for our network is 0.307 which is relatively small. Figure 4.11 displays closeness centrality distribution of our network. It can be clearly visible that most the nodes have closeness centrality between 0.25 and 0.45. It indicates that members of this group is very closely connected with each other since out network is constructed based on posts, comments, likes, shares.

Betweenness centrality is a measure based on the number of shortest paths between any two nodes that pass through a particular node (Brandes U., 2001). Nodes around the edge of the network would typically have a low betweenness centrality. A high betweenness centrality might suggest that the individual is connecting various different parts of the network together. From Figure 4.9, we see that our network looks like a scale-free network. There are only few nodes who are connected with most of the other nodes. From this observation, it can be said that, only few members of that Facebook group publish posts or links and majority members just give likes/comments/reactions in various posts. We figure out such influential nodes in our network and visualized below (Figure 4.12). Influential nodes are those who have high betweenness centrality. For privacy reasons, instead of names, nodes are assigned as P1, P2, P3 and so on. Here, betweenness centrality of top 20 influential nodes are displayed.

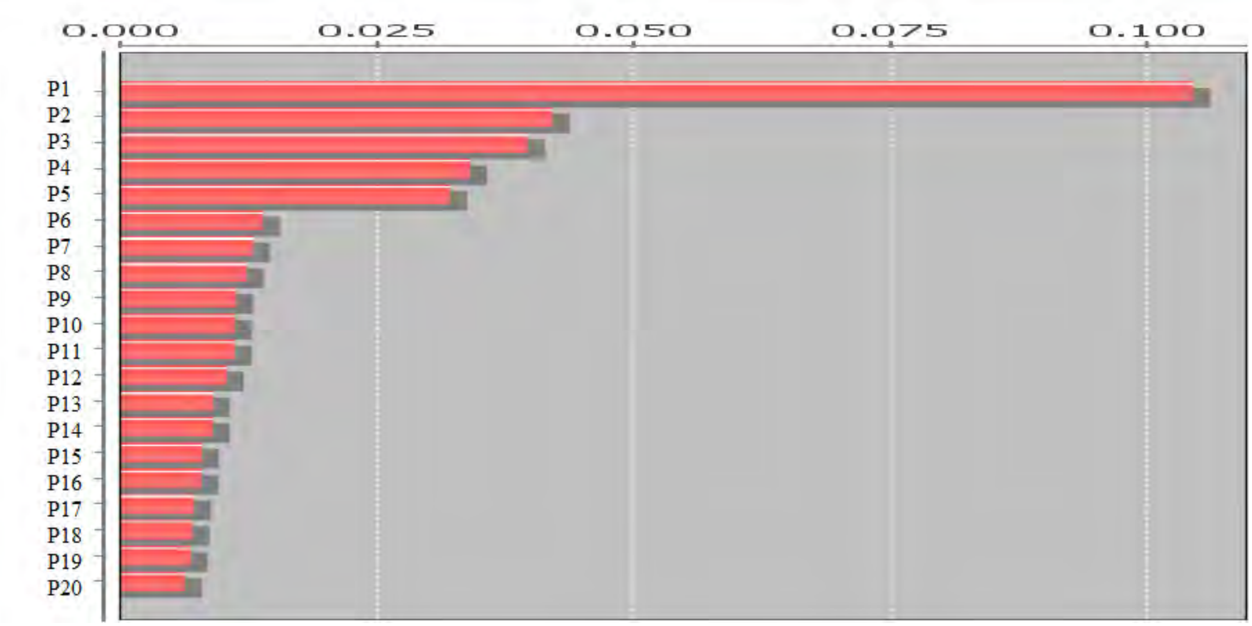


Figure 4.12: Betweenness centrality distribution of top 20 nodes of Facebook group.

In figure 4.13, influence network of top 3 nodes based on betweenness centrality is displayed. Green circled nodes have high betweenness score. Around 1500 nodes are connected only with these 3 nodes.

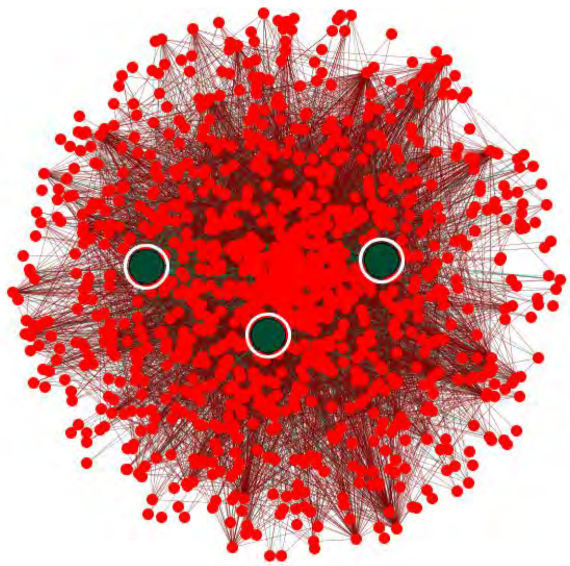


Figure 4.13: Influence network (top 3 nodes are green colored circle) of Public Service Innovation Facebook group.

Since our network is scale-free, the distribution of node linkages follows a power law distribution, in that most nodes have just a few connections and some have a tremendous number of links (Adamic, L. A., et. Al 2000). From figure 4.14, we observe that only 4-5 nodes contain the most of the links and other contain less number of links. Influential network justifies the power law distribution of node linkage.

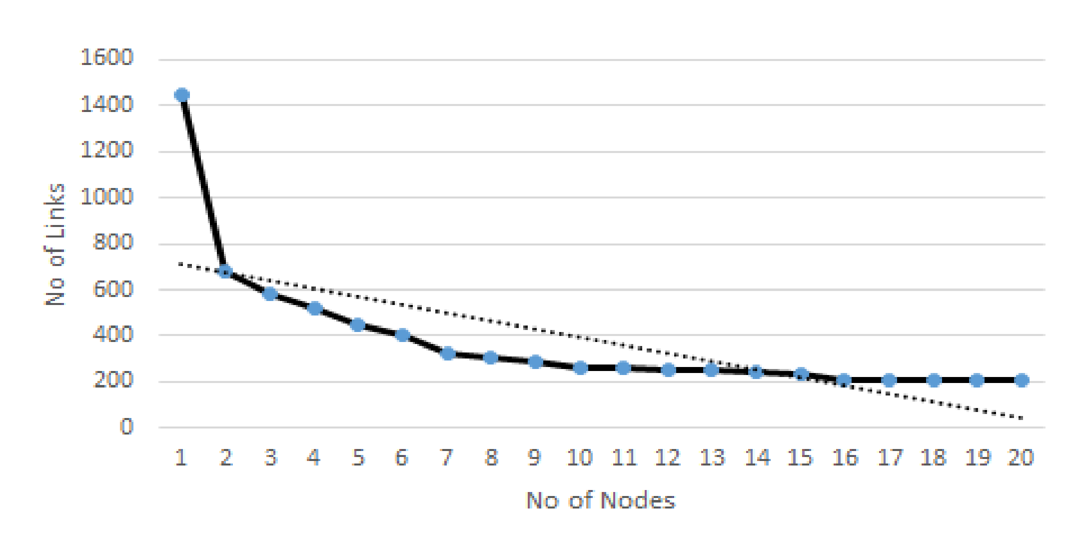


Figure 4.14: Power law degree distribution of node linkage.

In graph theory, eigenvector centrality is a measure of the influence of a node in a network; a measure of related influence, who is closest to the most important people in the graph? Kind of like "power behind the scenes" or "influence beyond popularity". It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. Figure 4.15 shows the top 10 nodes who are working as supporting influential nodes in our network based on eigenvector centrality distribution. It can be observed that P2, P3, P4 and P5 are supporting influential nodes as well as main influential nodes.

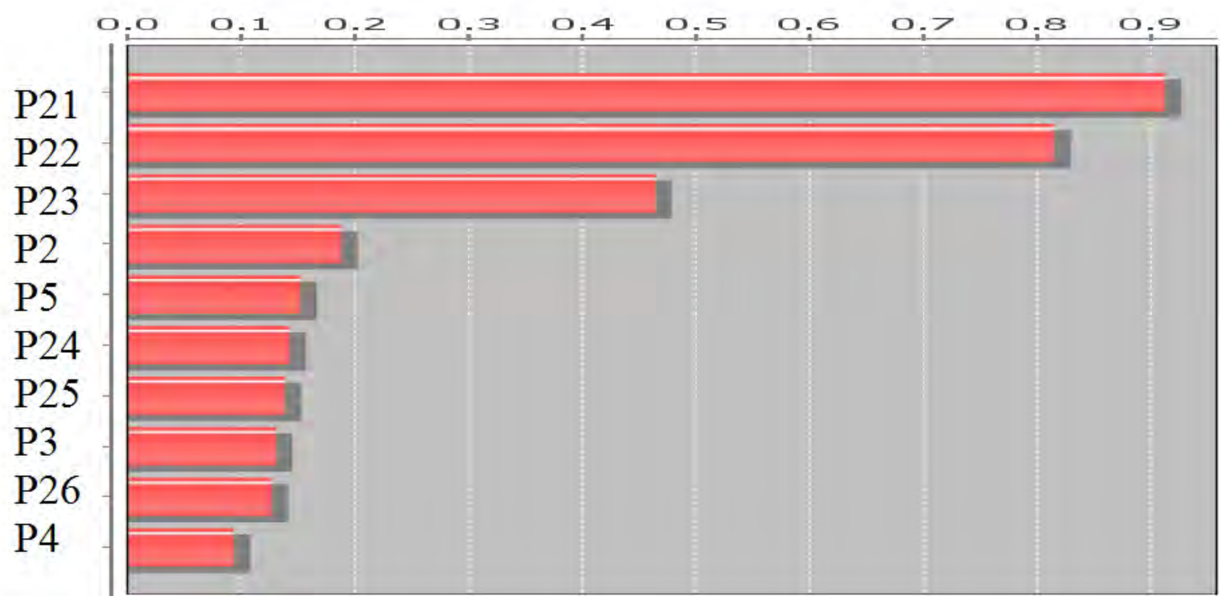
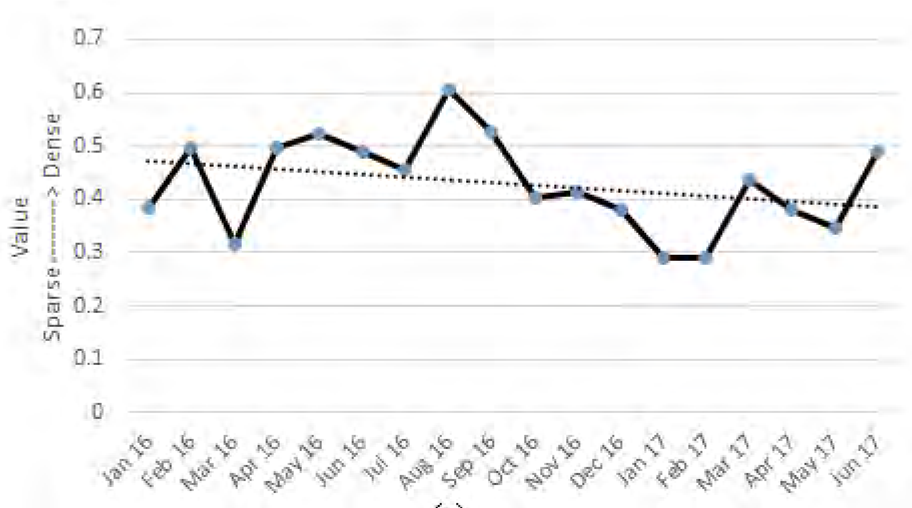
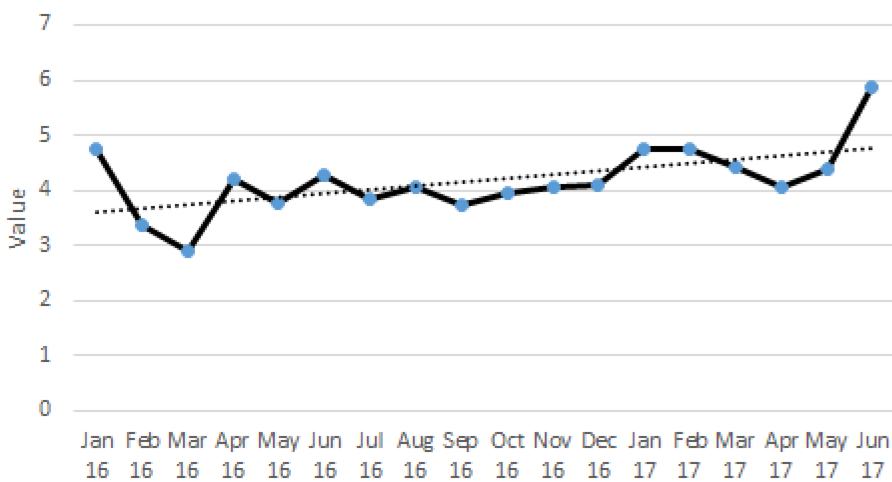


Figure 4.15: Eigenvector centrality distribution of top 10 supporting nodes of Facebook group.

After that, we have analyzed various network and node level measures of our network month wise. We have formed network based on each month's data and have calculated each month's characteristic path length, clustering coefficient, network assortativity, betweenness centrality, eigenvector centrality and so on. Detail values are given on Table 4.4 (Appendix A). Figure 4.16 presents month wise clustering coefficient (a) and characteristic path length (b). We observe that day by day, number of posts are increasing and clustering co-efficient is decreasing. It means that, our network is becoming sparser. Various types of posts are increasing and people are liking or commenting on various types of posts. Therefore, it becomes less dense in recent days. On the other hand, characteristics path length is increasing a bit with the advent of increasing number of posts.



(a)



(b)

Figure 4.16: (a) Month wise clustering coefficient. (b) Characteristic path length of Public Service Innovation Facebook group.

Another important network property is assortativity which measures the extent to which nodes of similar degree are inter-connected. Assortativity is a preference for a network's nodes to attach to others that are similar in some way (Assortativity. (2017)). Though the specific measure of similarity may vary, network theorists often examine assortativity in terms of a node's degree. The addition of this characteristic to network models more closely approximates the behaviors of many real world networks. Correlations between nodes of similar degree are often found in the mixing patterns of many observable networks. For instance, in social networks, nodes tend to be connected with other nodes with similar degree values. This tendency is referred to as assortativity.

High values (+1) mean that the nodes with high degree are clustered together and nodes of low degree are clustered together. Low values (-1) mean that high degree nodes connect to low degree nodes. Figure 4.17 represents month wise assortativity of our network. For our network, each month has negative assortativity. It means that, here, in this group, there is no cluster among members. Low degree nodes are connected with high degree nodes and vice versa. Only few members publish posts who are known as high degree nodes and there are numerous members of people who only give like and basically, they are known as low degree nodes. But, trend-line suggests that assortativity of this network is increasing which means that in future there will be clusters (high degree nodes' cluster and low degree nodes' cluster) in this group.

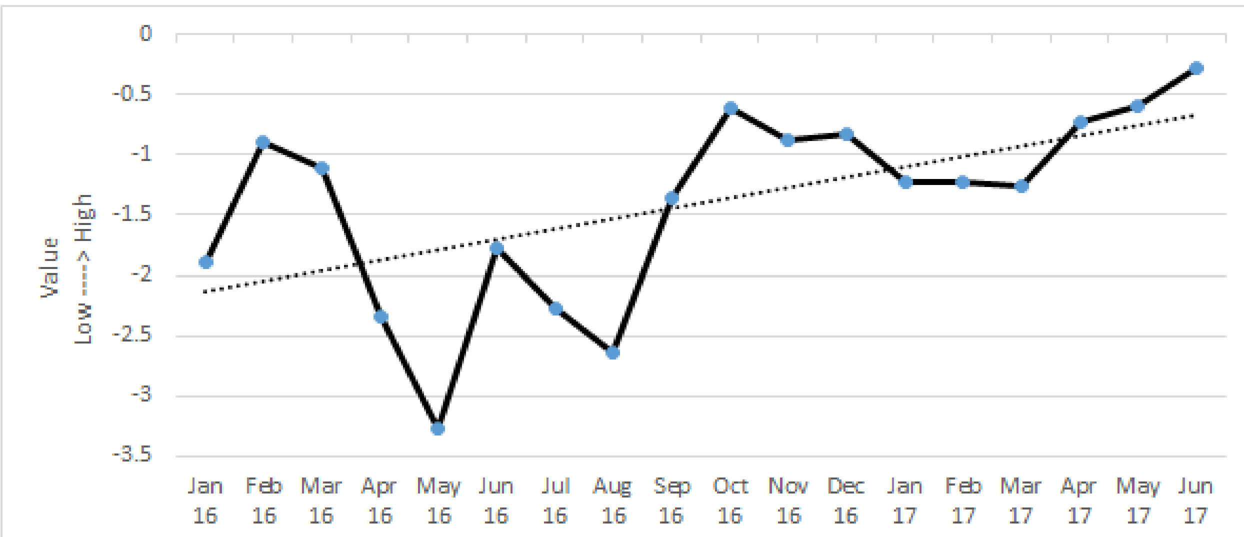


Figure 4.17: Month wise network assortativity of Public Service Innovation Facebook group.

Finally, we have analyzed overall (May 2016 – April 2017) influence based on betweenness centralization and total degree centralization by each month to observe whether one person holds the power of the group or others are also coming forward. We have found out top 10 influential persons in that group based on betweenness centralization and total degree centralization. List of top 10 influential persons are specified on Table 4.5 (Appendix A). Among them, top 3 nodes are considered for month wise analysis. Figure 4.18 depicts month wise betweenness centralization of top 3 influential nodes. It is observable from the image that node P1 is losing its betweenness centrality day by day. Whereas, node P2 is gaining its betweenness centrality. It means that node B is becoming more and more active and working as a medium of connecting so many isolating nodes.

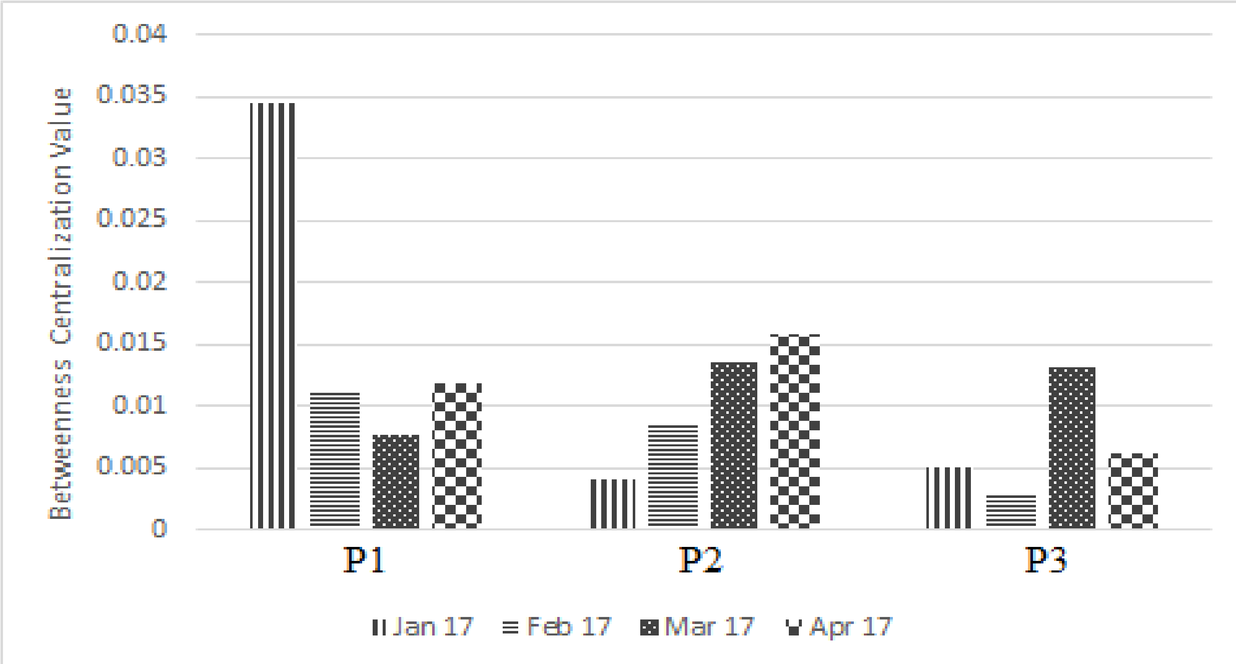


Figure 4.18: Month wise betweenness centralization of top 3 influential nodes

Later, we have found out each month's top 5 influential nodes based on betweenness centralization (Table 4.6) and observed that, Node A, B and C are quite consistent in leading the role of influential nodes. However, it is noticeable that, in each month, new nodes are becoming influential nodes too. Therefore, it can be said that, influential persons (in terms of group activity) in that group are increasing gradually which is a good sign for good governance and eradicating principal agent problem.

Table 4.6: Each Month's Top 5 Influential Nodes Based on Betweenness Centralization

Sl No.	January 2017	February 2017	March 2017	April 2017
1	P1 (0.034)	P22 (0.013)	P2 (0.013)	P2 (0.015)
2	P2 (0.028)	P1 (0.011)	P3 (0.013)	P1 (0.011)
3	P26 (0.020)	P23 (0.009)	P1 (0.007)	P11 (0.006)
4	P27 (0.018)	P24 (0.009)	P17 (0.006)	P3 (0.006)
5	P28 (0.017)	P2 (0.008)	P14 (0.005)	P5 (0.005)

.

CHAPTER 5

Conclusion and Future Works

In this chapter we have summarized the main findings with regards to the research questions and we have also provided general conclusions based on the findings of the studies presented in this thesis. Besides, we have discussed about the limitations of this thesis and our future plan of work on this thesis.

5.1 Conclusion

A social medial platform like Facebook group provide an indispensable instrument to connect a node with every other node which is the essence of efficient organization. Any platform builds upon a hierarchical chain of command sometimes hinder the progress of the entire system nevertheless. When the nodes are directly connected through each and every node regardless of hierarchical position then each can explore themselves and take part in decision making through sharing their opinion and views. We have seen the evidence of efficiency in the in solving problems and to disseminate innovative ideas by analyzing "Public Service Innovation Bangladesh" group, which was only been viable by participating members regardless of their hierarchical position.

This Facebook group has a remarkable opportunity to acknowledge how government can work if given a platform where everybody can be connected. We have attempted to see if it is true. So we scrutinize this network and complex network analysis makes way for this possibility. This Facebook group network is not closely connected. There are only a few people give posts and statuses and others just comment or like that person's post. So, those who publish more posts, we call them leaders or influential person of that group. One example is, one lower rank official post a problem statement on 26th may 2017 which indicate that one primary school was in under construction for 18 years and never been completed. The commissioner, highest ranked official of a district, of that district has seen the statement and commented on 27th May,2017, with all the dispatches, he has assigned an UNO (chief official of a sub-district) to look into that matter to solve the problem as soon as possible. The problem which was in the curse of red tape for 18 years solved in 1 day. It has only been possible to contact with a high rank official through this Facebook group platform.

Government suffers from principal agent problem and it seems that online communities such as Facebook groups may make it flatter. Consequently, this should reduce the information asymmetry problem given that everyone participates. Acknowledging the nature of these platforms does give a new opening for the policymakers and thinkers to come forward with new methods to understand how government works. In this paper, we have conducted complex network analysis to understand the structure of the community and have found out the key members or influential members of this group.

5.2 Future Works

For further analysis, we hope to use various other computational tools such as sentiment analysis, natural language processing techniques to understand the nature of the communication among the users. Also, in this research, we have not yet analyzed the contents of this group posts. Like, what type of posts are being posted; is it related to social problems, or is it related to public service, or is it related to social development, or is it related to innovative idea sharing and so on. In coming days, we plan to categorize and classify post contents or post materials along with sentiment analysis. Sentiment analysis might give some insights of government officials' thoughts and emotions regarding posts of this group. For that, we require knowledge related to Bengali language processing.

REFERENCES

Abello, J., Buchsbaum, A. & Westbrook, J. A functional approach to external graph algorithms. Lect. Notes Comput. Sci. 1461, 332–343 (1998).

Assortativity. (2017). En.wikipedia.org. Retrieved 31 July 2017, from https://en.wikipedia.org/wiki/Assortativity

Adamic, L. A., Huberman, B. A., Barabási, A., Albert, R., Jeong, H., &Bianconi, G. (2000). Power-Law Distribution of the World Wide Web. Science, 287(5461), 2115. doi:10.1126/science.287.5461.2115a

Aldenderfer, M.S. &Blashfield, R.K (1984). *Cluster analysis*. Sage Publication Inc.

Banavar, J. R., Colaiori, F., Flammini, A., Maritan, A. &Rinaldo, A. Topology of the fittest transportation network. Phys. Rev. Lett. 84, 4745–4748 (2000).

Borgatti, S. P. (2005). Centrality and network flow. Social Networks, 27(1), 55-71. doi:10.1016/j.socnet.2004.11.008

Brandes, U. (2001). A faster algorithm for betweenness centrality*. The Journal of Mathematical Sociology, 25(2), 163-177. doi:10.1080/0022250x.2001.9990249

Caldarelli, G. (2007). Scale-Free Networks. doi:10.1093/acprof:oso/9780199211517.001.0001

Carley, K. M., Reminga, J., & Kamneva, N. (2004). Destabilizing terrorist networks. *Institute for Software Research*, 45.

Christakis, N. A., & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine*, *358*(21), 2249-2258.

Complex network. (2017). En.wikipedia.org. Retrieved 25 July 2017, from https://en.wikipedia.org/wiki/Complex_network

Costa, L. D. F., Oliveira Jr, O. N., Travieso, G., Rodrigues, F. A., Villas Boas, P. R., Antiqueira, L., ... & Correa Rocha, L. E. (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, *60*(3), 329-412.

Csermely, P., London, A., Wu, L., &Uzzi, B. (2013). Structure and dynamics of core/periphery networks. Journal of Complex Networks, 1(2), 93-123. doi:10.1093/comnet/cnt016

Erdös, P. & Rényi, A. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 5, 17–61 (1960).

Faloutsos, M., Faloutsos, P. & Faloutsos, C. On power-law relationships of the internet topology. Comp. Comm. Rev. 29, 251–262 (1999).

Hage, P., &Harary, F. (1995). Eccentricity and centrality in networks. Social Networks, 17(1), 57-63. doi:10.1016/0378-8733(94)00248-9

Ilyas, M. U., &Radha, H. (2011, June). Identifying influential nodes in online social networks using principal component centrality. In *Communications (ICC), 2011 IEEE International Conference on* (pp. 1-5). IEEE.

Jacomy M, Venturini T,Heymann S, Bastian M (2014). A Continuous Graph Layout Algorithm for Handy Network Visualization. PLoS ONE 9(6): e98679. https://doi.org/10.1371/journal.pone.0098679

Jeong H., Tombor, B., Albert, R., Oltavi, Z, N. & Barabási, A.-L. The large-scale organization of metabolic networks. Nature 407, 651–654 (2000).

Kauffman, S. At Home in the Universe (Oxford, New York, 1995).

Krapivsky, P. L. ,Redner, S. &Leyvraz, F. Connectivity of growing random networks. Phys. Rev. Lett. 85, 4629–4632 (2000).

Liggett, R. S. (1980). The quadratic assignment problem: an analysis of applications and solution strategies. Environment and Planning B: Planning and Design, 7(2), 141-162. doi:10.1068/b070141

Mickoleit, A. (2014), “Social Media Use by Governments: A Policy Primer to Discuss Trends, Identify Policy Opportunities and Guide Decision Makers”, OECD Working Papers on Public Governance, No. 26, OECD Publishing, Paris.

Moon, I. C., & Carley, K. M. (2007). Modeling and simulating terrorist networks in social and geospatial dimensions. *IEEE Intelligent Systems*, *22*(5).

Newman, M. E. (2001). From the Cover: The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2), 404-409. doi:10.1073/pnas.021544898

Newman, M. E. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E, 64(1). doi:10.1103/physreve.64.016132

ORA-LITE: Software | CASOS. (2017). Casos.cs.cmu.edu. Retrieved 30 June 2017, from http://www.casos.cs.cmu.edu/projects/ora/software.php

Pi-ji, C. H. E. N. (2000). On the Information Asymmetry of the Tourism Market and the Government Administrative Behaviors [J]. Tourism Tribune, 2, 011.

Renfreo,II, C. R., &Deckro, R. F. (2001). *A social network analysis of iranian government*. Retrieved from https://fas.org/irp/eprint/socnet.pdf

Ruhnau, B. (2000). Eigenvector-centrality — a node-centrality? Social Networks, 22(4), 357-365. doi:10.1016/s0378-8733(00)00031-9

Social Networks and Travel Behaviour. (2015). doi:10.4324/9781315609560

Strogatz, S. (2001). Exploring complex networks. Nature, 410(6825), 268-276. http://dx.doi.org/10.1038/35065725

Wasserman, S. & Faust. K. Social Network Analysis: Methods and Applications (Cambridge Univ. Press, New York, 1994).

Watts, D. J. & Strogatz S. H. Collective dynamics of ‘small-world’ networks. Nature 393, 440–442 (1998).

Web Harvy (2017). Retrieved 25 July 2017, from https://www.webharvy.com/articles/what-is-web-scraping.html

Wilson, E. O. Consilience p.85 (Knopf, New York, 1998).

Appendix A

Table 3.1: Description of Dataset Attributes

Sl No.	Attribute Name	Attribute Description
1.	Actions	Sum of the number of posts, comments, and likes a user makes.
2.	Made Posts	Number of posts a user submits.
3.	Made Comments	Number of comments a user writes.
4.	Made Reactions	Number of reactions a user makes.
5.	Receive Comments	Number of comments a user's post(s) receives.
6.	Receive Comments	Number of reactions a user's post(s) receives.
7.	Receive Comment Likes	Number of likes a user's comment(s) receives.
8.	Type	Facebook's post classification (e.g. photo, status, etc.).
9.	Post by	Author of the post.
10.	Post Link	Direct link to the post.
11.	Post Message	Text of the post.
12.	Picture	The picture scraped from any link included with the post.
13.	Full Picture	The picture scraped from any link included with post (full size).
14.	Link	Link URL (if the post points to external content).
15.	Link Domain	Domain name of link.
16.	Post Published	Post publishing date.
17.	Post Published Unix	Publishing date as Unix timestamp (for easy conversion and ranking).

18.	Post Published SQL	Publishing date in SQL format (some analysis tools prefer this).
19.	Likes	Number of actually retrieved likes a post received or a user made.
20.	Likes Count	Facebook provided like count for posts.
21.	Comments Count	Facebook provided comment count for posts.
22.	Reactions Count	Facebook provided reactions count for posts (includes likes).
23.	Shares Count	Facebook provided share count for posts.
24.	Engagement	Sum of comment, reaction, and share counts.
25.	Comments Retrieved (not in “stats only” mode)	Actually retrieved comments (may be lower than Comments Count due to privacy or deletion).
26.	Comments Base (not in "stats only" mode)	Number of base level comments (in threaded conversations).
27.	Comments Replies (not in "stats only" mode)	Number of reply level comments (in threaded conversations).
28.	Comment Likes Count (not in "stats only" mode)	Number of likes on comments for this post.
29.	Rea_X (not in "stats only" mode)	These columns give counts for the retrieved reactions (NONE, LIKE, LOVE, WOW, HAHA, SAD, ANGRY, THANKFUL).
30.	Comment ID	ID of the comment.
31.	Comment by	Author of the comment.
32.	Is Reply	Whether the comment is a reply to another comment (in threaded conversations).
33.	Comment Message	Text of the comment.
34.	Comment Published	Publishing date of the comment.

Table 4.4: Month Wise Various Network and Node Level Measures of Facebook Group Network

	Characteristic Path Length	**Clustering Coefficient**	**Network Assotativity**	**Betweenness Centralization**	**Eigenvector Centralization**	**Eccentricity Centralization**	**Total Degree Centralization**
May 16 – April 17	3.123	0.368	-1.121	0.036	0.693	5.625	0.003
Apr 17	4.050	0.526	-0.573	0.016	0.967	5.316	0.005
Mar 17	4.023	0.475	-0.712	0.013	0.603	5.467	0.005
Feb 17	4.360	0.548	-0.949	0.013	0.393	6.050	0.006
Jan 17	5.743	0.470	-0.713	0.034	0.721	6.501	0.002
Dec 16	4.002	0.561	-0.797	0.018	0.721	6.874	0.004
Nov 16	4.104	0.583	-0.676	0.017	0.933	3.871	0.007
Oct 16	4.978	0.573	-1.110	0.015	0.622	5.124	0.003
Sep 16	3.709	0.603	-0.773	0.014	0.762	4.964	0.007
Aug 16	4.805	0.406	-1.231	0.027	0.848	7.252	0.003
Jul 16	5.192	0.445	-1.024	0.019	0.615	7.896	0.005
Jun 16	3.845	0.575	-0.731	0.026	0.986	5.266	0.004
May 16	4.109	0.580	-0.937	0.014	0.543	6.053	0.008

Table 4.5: Overall Influence (May 2016 – April 2017) based on Betweenness Centralization and Total Degree Centralization

Node	Betweenness Centralization	Total Degree Centralization
P1	0.036437	0.002937
P2	0.031945	0.003467
P3	0.030391	0.002434
P4	0.02021	0.001341
P5	0.019056	0.00107
P6	0.012462	0.000752
P7	0.007194	0.000469
P8	0.006604	0.000588
P9	0.005714	0.002465
P10	0.005407	0.000439