New Network Generation Models using concept of Friends of Friends

A *Thesis Submitted*

in Partial Fulfilment of the Requirements

for the Degree of

Master of Technology

by

Kumar gaurav

to the

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

March, 2018

CERTIFICATE

It is certified that the work contained in the thesis entitled **"New Network Generation Models using concept of Friends of Friends"** being submitted by **Mr. Kumar Gaurav** has been carried out under my supervision. In my opinion, the thesis has reached the standard fulfilling the requirement of regulation of the M.Tech. degree. The results embodied in this thesis have not been submitted elsewhere for the award of any degree or diploma.

Dr. Yatindra Nath Singh

Professor

Department of Electrical Engineering

Indian Institute of Technology Kanpur

Kanpur, INDIA

28th **March, 2018**

Abstract

Networks are everywhere around us. They can represent how people are friends with each other, how the papers cite other papers, how the routers are connected in actual network. The structure of the network characterizes the behaviour of all the processes happening over the network, e.g., diffusion of information. Study of network structures and its characterization is now know as network science. In last two decades, the network science has grown as an independent area of research and has found its application in multiple fields of science, technology, medicine and humanities.

To generate a network structure which can mimic real world networks in a more better manner has always been of interest. It is expected to give insights into how the real world networks function.
In this thesis, we have proposed a model of network generation which modifies the existing ones on the basis of these two common observations.

1. A person makes new connection via someone he/she already knows i.e., chances to make new connection among friends of friends is more.

2. A person does not make all of its friends at once, rather it is done in multiple steps in iterative fashion.

Including these two features results into a network having all three desirable qualities i.e., Small world property, Scale free degree distribution and high clustering coefficient. One more advantage is that average Clustering coefficient of the network does not drop to zero with increase in size of network.
A new parameter f i.e., tendency of a node to make new connection among friends of

friends circle has been introduced. It has been shown that by varying the parameter f, we can generate network having desired level of clustering coefficient. We have also proposed the strategy for estimation of this parameter f from a given network structure. It has been shown that estimation gives better result if we have images of the network structure at multiple instances during evolution.

Dedicated

To that Friend

Who Sweated for Hours,

Who Struggled Hard,

Who Took a Stand,

so that all of us can get an M.Tech. Degree...

Saumik,

You are a gem, mate!

Till now, Direct Ph.D. students used to get only Ph.D. degree. It is for the first time that we are getting M.Tech degree too. Saumik has taken this initiative forward knowing the fact that he will leave the institute before new rule will come up.

Acknowledgements

This M.Tech thesis is not a sudden event but only a step of the stair that I am climbing for a long time. A journey that started with the teaching of my parents and evolved with my schooling. Even if a small step was not there in time, it could never be possible to come this far. In every turn of my life there was someone to guide me for the path ahead and this is an opportunity to thank each of them for their presence in my life.

If I talk about this thesis particularly, then the key person will be none other than my supervisor who helped me technically at every steps with suggestions and supports. I am no one to judge a person but I feel he is good with technical knowledge but far better as a human being.

Friends were always an integral part of this journey- whether the boat was sailing or not. Sateesh is one such friend with whom I have discussed many perspectives of life and education. Ruchir sir has guided me in initial phase of my programme. I have learnt a lot about presenting a work from Rameshwar Ji. Anupam, Pallavi, Satish Awasthi Ji, Nitin, Varsha, Amit - all these people made this journey easy and tolerable.

Outside my lab, Saumik was always there with me from the very first day in IIT.

Manjeer, Amrita, Panchajanya, Nandini, Manoranjan and Aasif have always made this campus a home away from home for me. I associated with HSS in the last phase but they reduce my distance from Hindi literature and I am especially thankful to Shreya and Faizal for that.

Shyam Bhaiya was even more careful than me so that I can really complete my thesis in time. If he did not remind me regularly, I do not know whether I would ever be able to complete my thesis or not.

The way Sapna, my younger sister, took care of my family and house in this entire duration is unimaginable. My words will never be sufficient to thank her properly for that.

(Kumar Gaurav)

List of Symbols

A	Adjacency matrix of a network.
N	Total number of nodes in the network.
M	Total number of links in the network.
m_0	number of initial nodes in a network generated by Barabasi model.
k	Degree of the node.
$p(k)$	fraction of nodes having degree k.
$\langle k \rangle$	Average degree of the network.
$\langle k^2 \rangle$	Second moment of degree.
σ	Variance of the degree distribution.
d_{ij}	length of a geodesic path from node i to node j.
l	mean shortest distance between nodes of the network.
C_i	Local clustering coefficient of i^{th} node.
z_m	Mean value of number of neighbors at m hops away.
z_2	Mean value of number of neighbors at two hops away.
q_k	Probability that a node at the end of the random edge having excess degree k.
$\Pi(k)$	probability of a new node to connect to node having degree k.
f	probability of a node to make connection among his/her friends of friends circle .
α	Power law degree exponent.
v	A vertex.
$\boldsymbol{V}$	Set of all vertices of the network.
$\boldsymbol{N}(v)$	Set of neighbors of a node v.
$\mathbf{\Gamma_i}(\mathrm{v})$	Set of nodes at distance i from node v.
$\mathbf{\Gamma_2}(i)$	Set of nodes at distance 2 from node i i.e., set of friends of frends.
$\lvert\mathbf{\Gamma_2}(i)\rvert$	Cardinality of the set, number of friends of friends.
$\boldsymbol{S}(i)$	Set of nodes at distance 3 or more from node i.
p_{ij}	Probabilty for a node i to make connection with node j.
Q	Modularity Index.
fof	Acronym for friends of friends.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Nothing in this nature is isolated. Things are connected and interrelated. Number of components linked together in a certain way results into various networks. Example includes social network of people; technological network like Internet, telephone, power grid network; biological network of proteins, cells, molecules, neurons; and ecological networks like food web [1]-[5].

Interconnections in any system can be represented as a network. Network, in its simplest form, is a collection of points joined together in pairs by lines. Points are referred as nodes (or vertices) and the lines are referred as edges. While representing any system as a network, the components of the system are treated as nodes and the connections as edges. Network structure and the pattern of interactions can have a big effect on the behaviour of the systems. Pattern of connection between computers on the internet affects the routes that data take over the network and in turn affects the network data transportation efficiency. Connection in a social network affect how people learn, gather news and form opinions. It also affects phenomena such as spread of disease in a population. Information regarding structure of the networks is essential to understand how the corresponding systems work.

Real networks possess many characteristics. To investigate into the reasons behind a particular characteristic or to understand why a certain property is there in the network, different models of network generation have been proposed from time to time. These models try to generate a synthetic network which can mimic the real world network bahaviour.

Apart from models of network generation, network science deals with measurement of network properties, analysis of network data and study of dynamic processes on network like percolation, epidemic spread and network search [6]. In this work, we are mainly dealing with the models of network generation and measurement of some of the properties of the network.

1.1 Outline of the Thesis

In chapter 2 we have given background of network science necessary for this thesis work. We have discussed parameters which are used to characterize a network. We have also discussed few network generation models in brief which are the basis for our proposed model. We have also pointed out what extra features our model will introduce.

In chapter 3 we have discussed our proposed network generation model and its all possible variants. Finally we have discussed our more generalized model of network generation. Networks generation have been simulated with the proposed models and relevant parameters have been calculated.

In chapter 4 we have tried to estimate the fraction f which depicts the tendency of a node to make connection among friends of friends. We have also shown that by varying f, we can control average clustering coefficient of the network as well as the number of

communities in the network.

Chapter 5 concludes the thesis and provides possible future directions of investigations.

Chapter 2

Fundamentals of Network Theory

Network is most commonly represented as an adjacency matrix. The adjacency matrix **A** of a simple network graph is a matrix with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1 & \text{if link exists between i and j,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.0.1}$$

For a network with no self-loops, the diagonal elements of the matrix **A** are all zero. The Adjacency matrix for an undirected graph is always symmetric, since an edge between i and j implies an edge between j and i. In case of directed network, each edge has a direction, pointing from one vertex to another. Adjacency matrix of such a network is generally asymmetric. If edge of a network represents certain strength or frequency of certain event, number 1 is replaced by various rational numbers representing weights. Such a network is called weighted network.

2.1 Network Measures and Metrics

There are a large number of metrices and measures for a network. We have discussed a few of them which are relevant to our thesis work.

2.1.1 Degree Centrality

To find out the most important or central nodes in the network according to definition of importance, different centrality measures have been defined. Simplest centrality measure in a network is the degree of a node i.e. the number of links connected to it. Highest degree node is the most central node.

There are many other centrality measures like Eigenvector centrality, Katz centrality, Closeness centrality, Betweenness centrality, page rank, etc. These have not been used in our present work thus have not been elaborated any further.

2.1.2 Degree Distribution

In a network, fraction of vertices having degree k is denoted by p_k . It can also be thought of as a probability of a randomly chosen node to have degree k. The distribution of the probability p_k is called degree distribution in the network. The directed networks will have in-degree distribution and out-degree distribution corresponding to number of links ending at and beginning from a node.

2.1.3 Excess Degree

Except a randomly chosen link to reach on to the node, all other links connected to the node are considered to be excess degree in the network [6]. Hence, if excess degree of a node is k, then the total degree is $(k+1)$. q_k is the probability of a node at the end of a randomly chosen link having excess degree k . Thus, the excess degree distribution distribution q_k is given by

$$q_k = \frac{(k+1)p_{(k+1)}}{\sum_{k=1}^{N} kp_k}. \tag{2.1.2}$$

where p_k is the degree distribution defined in section 2.1.2. Therefore, average number of outgoing edges of a neighbor vertex is

$$\begin{aligned} \sum_{k=0}^{\infty} kq_k &= \frac{\sum_{k=0}^{\infty} k(k+1)p_{k+1}}{\sum_j jp_j} \\ &= \frac{\sum_{k=0}^{\infty} (k-1)kp_k}{\sum_j jp_j} \\ &= \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \end{aligned} \tag{2.1.3}$$

Number of Next-Nearest Neighbors : The average number of neighbors m hops away is denoted by z_m. Thus z_1 gives average number of neighbors which are one hop away i.e., average degree of the node.

$$z_1 = \langle k \rangle \tag{2.1.4}$$

z_1 is commonly written as z. Equation (2.1.3) gives the average number of nodes two hops away from the starting node via a specific neighbor vertex. Multiplying the Equation (2.1.3) by average degree of starting node, $z_1 \equiv z$, the mean number of second hop neighbors, z_2 is

$$z_2 = \langle k^2 \rangle - \langle k \rangle. \tag{2.1.5}$$

The average number of edges emerging from a second neighbor (but not leading back), is also given by Equation (2.1.3). Hence, the average number of neighbors at m hops

away as mentioned in [6] is given by

$$z_m = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} z_{m-1} = \frac{z_2}{z_1} z_{m-1}. \tag{2.1.6}$$

Solving recursive relation gives

$$z_m = \left[\frac{z_2}{z_1}\right]^{m-1} z_1. \tag{2.1.7}$$

2.1.4 Clustering Coefficient

The clustering coefficient measures the average probability that two neighbors of a node are themselves neighbors. In effect it measures the density of triangles in the network. Local clustering coefficient for a single vertex i having degree k_i is defined as

$$C_i = \frac{\text{number of pairs of neighbours of i that are connected}}{\text{number of possible pairs of neighbours of i}} \tag{2.1.8}$$

where total possible number of pairs of neighbors is $\frac{k_i(k_i-1)}{2}$. Average local clustering coefficient of the network is the mean of local clustering coefficients of all the nodes in the network.

Another way to define clustering on the scale of whole network is in terms of Global clustering coefficients [7].

$$C = \frac{(\text{number of triangles}) \times 3}{(\text{number of connected triplets})} \tag{2.1.9}$$

Here a "connected triplet" means three vertices abc with edges (a, b) and (b, c). The edge (a, c) may or may not be present. The factor of three in the numerator arises because each triangle gets counted three times when we count the connected triplets in the network.

In our thesis work, we have used average local clustering coefficient as the clustering parameter.

2.1.5 Geodesic Distance

A geodesic path, also called simply a shortest path, is a path between two vertices such that no other shorter path exists. Number of edges traversed along this path is called geodesic distance or shortest distance.

The diameter of a network is the length of the longest geodesic path between any pair of nodes in the network for which a path actually exists.

Suppose d_{ij} is the length of a geodesic path from node i to node j, then the mean geodesic distance of node i from all other nodes in the network is calculated by average of all possible d_{ij} pair in the network (except $j = i$)

$$l_i = \frac{1}{(n-1)} \sum_j d_{ij} \tag{2.1.10}$$

For a connected network (having only one component), the mean distance between pairs of vertices is average of l_i's for all i's.

$$l = \frac{1}{n} \sum_i l_i \tag{2.1.11}$$

l is referred as mean shortest distance between nodes of the network.

2.2 Three Important Characteristics of real network

Different real world network possess different characteristics but most of them have few characteristics in common. Most important common characteristics of real world networks are: Small world effect, Power law degree distribution and high clustering coefficient. We are going to discuss each of these characteristics one by one.

2.2.1 Small-World Effect

The small-world effect is the hypothesis that most nodes can be reached from every other node by a small number of hops. In mathematical terms, mean shortest distance between nodes (as discussed in section 2.1.5) is small. Most of the real world network possess the characteristic of small-world [8].

The idea of six degrees of separation is a beautiful example of small-world effect. It conveys that anyone in this world is six or fewer steps away from anyone else so that a chain of friend of a friend can be made to connect any two people in a maximum of six steps. It was originally set out by Frigyes Karinthy in 1929 and popularized in an eponymous 1990 play written by John Guare [9].

Small-world effect has substantial implications for networked systems. Suppose a rumor or a disease is spread over a social network. Rumor will reach out to entire population much faster as people are only six steps far from each other. If average distance between them would have been in hundreds, the same process must have taken a very long time.

2.2.2 Scale Free Network

In statistics, a power law is a functional relationship between two quantities, where one quantity varies as a power of another. A scale-free network is a network whose degree distribution follows a power law, at least asymptotically.

$$p_k \sim k^{-\alpha} \tag{2.2.12}$$

The constant α is known as the exponent of the power law whose value is typically in the range $2 < \alpha < 3$, although occasionally it may lie outside these bounds [10].

One attribute of power laws is their scale invariance i.e. scaling by a constant c simply multiplies the original power-law relation by the constant $c^{-\alpha}$. All power laws with a

particular scaling exponent are equivalent up to constant factors, since each is simply a scaled version of the others. That's why networks having power law degree distributions are also called scale free networks.
Many real networks are scale free. Examples include in and out degrees of the World Wide Web, Co-authorship network of mathematicians, Protein-protein interaction network and many more [3],[11],[12].

2.2.3 High Clustering Coefficient

Most of the real networks have a high value of average local Clustering coefficient (as defined in Section 2.1.4). It signifies that nodes in a network tend to cluster together. Another interesting property is that even for very large network size, clustering coefficient does not vanishes and remains nonzero [7].

2.3 Models of Network Generation

Network science aims to build models that reproduce the properties of real networks. Here, we are going to describe three fundamental models of network generation which helped a lot to understand why a network possess a particular characteristics. These are the basic models which have been modified by us to achieve desired network characteristics.

2.3.1 Erdős-Rényi Random Graph Model

At first inspection, large network appears to be random. Paul Erdős and Alfréd Rényi [13] used this apparent randomness to generate networks that are truly random.

There are multiple ways to define and generate random network. Most common is $G(n, p)$ model where n is number of nodes and p is probability with which each pair of node is connected by a link.

To construct a random network, these steps are followed:

1. Start with n isolated nodes.
2. Select a node pair and generate a random number between 0 and 1. If the number is less than p, connect the selected node pair with a link, otherwise leave them disconnected.
3. Repeat step (2) for each of the $\frac{n(n-1)}{2}$ node pairs.

Consequently the degree distribution of a random network follows the binomial distribution

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{2.3.13}$$

Most real networks are sparse, meaning that for them $k << n$. In this limit the degree distribution 2.3.13 is well approximated by the Poisson distribution

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{2.3.14}$$

which is often called the degree distribution of a random network. Here, $\langle k \rangle$ is average degree of the network.

Average local clustering coefficient of a random network is equal to p which is independent of n. But, The average clustering coefficient of real networks is much higher than a random network of similar size (n) and equal number of links. Out of three desired properties mentioned in Section 2.2, random network possess only one i.e. small world phenomenon. Scale free property and high clustering coefficient are missing in the random networks.

2.3.2 Watts-Strogatz Model

Before Watts and Strogatz claim, network topology was assumed to be either completely regular or completely random. The Watts-Strogatz model [14] interpolates between a regular lattice, which has high clustering but lacks the small-world phenomenon, and a random network, which has low clustering, but displays the small-world property. By adding randomness on the top of regular graph structures like ring, the model was able to achieve high clustering coefficient along with small world properties.
To construct a network using this model, these steps are followed:

1. We start with a ring of nodes in which each node is connected to their immediate and next neighbors in both clock and anti-clock directions.

2. With probability p each link is rewired to a randomly chosen node.

For $p = 0$, no rewiring has been done and we get regular lattice. For $p = 1$, all links have been rewired, so the network turns into a random network. For small values of p, only few links are rewired therefore the network maintains high clustering while the random long-range links can drastically decrease the shortest possible distances between the nodes resulting into small world phenomenon.
This model possess small world effect and high clustering coefficient but lacks scale free property.

2.3.3 Barabási-Albert (BA) Preferential Attachment Model

Initially, large networks were assumed to be random and having Poisson degree distribution. With first map of WWW generated by Hawoong Jeong at University of Notre

Dame, it became clear that all networks do not have Poisson degree distribution, rather many of them follow power law degree distribution [15]. Using techniques to generate random graphs with any given degree distribution, it was possible to generate a network following power law. But, these models were not able to explain what is the reason for the network to have a power law degree distribution.

This investigation led to generative network models which focuses on the mechanism by which networks are created. If the structures produced by these models are similar to those of networks we observe in the real world, it suggests that similar generative mechanisms may be at work in the real networks. The Barabási-Albert model is the best known generative network model in use today [16].

According to this model, two essential features responsible for power law degree distribution are:

1. **Growth:** Real networks are the result of a growth process that continuously increases number of nodes (n). This is in contrast with the random network model which assumes that the number of nodes is fixed.

2. **Preferential attachment:** In real networks new nodes prefers to link to higher degree nodes i.e. nodes having more number of connections. In contrast, nodes in random network choose their interaction partners randomly.

To construct a network using this model, following steps are followed:

1. We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link.

2. At each timestep, a new node is added with $m(\leq m_0)$ links that connect the new node to m nodes which are already in the network.

Probability to make connection with any node is proportional to degree of that node. In this mechanism higher degree node always have more chances to attract new node towards them. Hence, higher nodes keep on accumulating more links and their degree will keep on increasing. In sociology, this is referred as Rich gets Richer phenomenon. Mathematically, the probability $\Pi(k)$ that a new node will make a connection with the node having degree k is

$$\Pi(k) = \frac{k}{\sum_j k_j} \tag{2.3.15}$$

Denominator is a normalizing factor which is summation of degree of all the nodes existing in the network. It was the first widely accepted model on evolving networks where nodes were added to the network over time and were more likely to link to nodes with higher degree. The model demonstrated small world effect with power law degree distribution but was not able to demonstrate the feature of high clustering coefficient which is also observed in real life networks.

2.3.4 Other Models

All of the above three models demonstrate maximum of two characteristics out of three desired characteristics of a real world networks. Later, the models were created to possess all the three characteristics simultaneously.

Jost and joy in [17] used distance preference to get high clustering coefficient with scale free degree distribution. Here, nodes at small distance were given high preference over nodes at larger distance or vice versa. To generate communities in the network, Xie *et.al.* in [18] started evolving the network with a few isolated communities. During network

growth, they first selected a community to which a new node will join and then a node from that community on the basis of preferential attachment. Recently, Meghanathan [19] has come up with a two hop neighbor preference based model which generates random graph with high clustering coefficient and large variation in nodal degree. All these models capture some of the properties which exist in our surroundings.

2.4 Our Proposed Model

We have proposed a network generation model which demonstrates all the three desired characteristics of a real network i.e. Small world effect, Scale free network and high clustering coefficient. It also uses growth as well as preferential attachment while generating the network. But there are few points of difference, which we will discuss in detail in next chapter. Along with that, few other features of our model are the following.

1. We can control the average clustering coefficient of the network by varying the value of a parameter f defined in our model. In this way we can generate network with desired level of clustering.

2. Without doing any thing specific for communities, our model generates communities which is characteristic of many real networks.

Chapter 3

Proposed Model of Network Generation

In society, a person makes new connections or gets introduced to a new person through someone he already knows. Someone acts as a middleman for growth of his neighborhood set. In Network Science, this phenomenon is referred as connecting with two hop neighbors or making connection among friends of friends. Most of the social sites like Facebook, Twitter, LinkedIn, etc., are using this principle for recommending new friends to its users [20]. This property has been widely used in the area related to link prediction over a network but very few models use this concept for generation of networks from initial stage itself [21],[22]. In our model, we have included this feature while generating the network. Rather than giving preference to higher degree nodes present any where in the network, in our model we prefer connection among friends of friends. In basic model, we have allowed a node to make friends only among friends of friends. But, later we have relaxed this assumption and allowed the node to make connection even to other nodes. We have chosen a fraction f with which a node can make connection among friends of friends and with fraction $(1 - f)$ it can make connection with others.

Other feature included in our model is regarding when to make connections. In Barabasi-Albert Model, a node is allowed to make connections only when it enters the network, not after that (Although, new incoming nodes can make a connection with it if they prefer). But, in real world, a person does not make all friends at once. It does so by adding friends one by one at intervals. To include this feature in our model, we have allowed nodes to make connections in iterations, one in each round. The network can move to next iteration, only after every node has created one link in the current iteration.

Network grows in terms of nodes as well as links but, increment in nodes is limited to first iteration only. Afterwards network grows only in terms of links.

To construct a network, following steps are followed:

1. In first iteration, at each timestep, a new node is added to the network with one link that connects new node to any one of the existing nodes with uniform probability.

   Two important points to note regarding first iteration are:

   (a) We have taken uniform probability considering the fact that a node can be brought in the network by any one. Only after entering the network, node can figure out whom to connect on the basis of parameter of interest like degree of nodes, common neighborhood or randomly.

   (b) This whole process generates a connected graph considering the fact that when a new node gets connected to a connected graph by a link, it generates a connected graph. At the end of the first iteration, there is a connected graph having N nodes and (N-1) links; basically a tree.

2. In subsequent iterations, each node is allowed to make one new connection in every iteration. A node needs to figure out the nodes to which it is allowed to make connection in current iteration and what is the parameter to assign probability of getting connected to each of them.

   Two important decisions to be made are:

   (a) New connections are allowed to be made only among friends of friends or even to other nodes.

   (b) Probability of making connections to a particular node among allowed set of nodes depends on degree of the nodes or number of common neighborhood or it is entirely random.

Depending upon what strategy we choose, five variants of network generation model have been proposed. Steps followed till first iteration is same in each of these models, difference exists only in the subsequent iterations. So, while describing different models, we have skipped the steps of first iteration and have discussed the strategy followed in subsequent iterations only. In next section, we have discussed first three models where connections are allowed only among friends of friends and remaining two models are discussed afterwards.

3.1 Models allowing connection among friends of friends only

As discussed in the last chapter, network is a collection of links and nodes. Total number of nodes (or vertex) in the network is denoted by N and number of edges by M. In this thesis, we have considered simple networks which does not have any self-loops

(edge connecting nodes to itself) or multi-edge (more than one link between same pair of nodes). The network is represented by an adjacency matrix A with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1 & \text{if link exists between nodes i and j} \\ 0 & \text{otherwise.} \end{cases} \tag{3.1.1}$$

Diagonal elements of the matrix are all zero, as there are no self-loops. Matrix is symmetric since there must be a link from node i to node j if there exists a link from j to i. A node (vertex) is represented by v where v= 1,2,3,. . .,N. Set of all nodes is represented by $\boldsymbol{V}$. Set of neighbors of a node v is represented by $\boldsymbol{N}$(v). Set of nodes at distance i from node v is denoted by $\mathbf{\Gamma_i}(v)$. For example, Set of nodes at distance 2 from node v is denoted by $\mathbf{\Gamma_2}(v)$.

All three models discussed in this section allow connection among friends of friends only. Difference lies in the parameter being used to assign probability of getting connected to friends of friends.

3.1.1 Random Selection among friends of friends

As clear from the name itself, a node is allowed to randomly make a new connection in each iteration to one of the friends of friends only. In this network generation model, when a node wants to add a link to some other node, a list of all the nodes at distance two (friends of friends) from the node is created and one node from this list is chosen randomly. Set of all such node is denoted by $\Gamma_2(i)$ as per the notation given in Section 3.1. Hence, the probability for a node i to connect to a node j is given by

$$p_{ij} = \begin{cases} \frac{1}{|\mathbf{\Gamma_2}(i)|} & \forall j \in \mathbf{\Gamma_2}(i) \\ 0 & \text{otherwise.} \end{cases} \tag{3.1.2}$$

Here, $|\mathbf{\Gamma_2}(i)|$ is cardinality of the set $\mathbf{\Gamma_2}(i)$, representing the number of elements in the set.

Same process is repeated for every node one by one. In the next iteration again every node makes one new connection. The whole process is iterated again and again until the targeted average nodal degree is achieved.

3.1.2 Preferential Attachment among friends of friends

In this variation, like in previous approach, connection is made to one of the nodes at distance two but probability to connect to a particular node is made proportional to the degree of that node as done in preferential attachment scheme of Barabasi-Albert model. Degree of node j is given by cardinality of neighborhood set of the node j i.e. $\boldsymbol{N}(j)$. In this case, probability for the node i to connect to node j is given by

$$p_{ij} = \begin{cases} \frac{|\boldsymbol{N}(j)|}{\sum_{k\in\boldsymbol{\Gamma_2}(i)}|\boldsymbol{N}(k)|} & \forall j \in \boldsymbol{\Gamma_2}(i) \\ 0 & \text{otherwise.} \end{cases} \tag{3.1.3}$$

Denominator is sum of degree of all two hop neighbors of node i which preserves the sum of the total probability of connecting to two hop neighbors by a node to unity.

3.1.3 Common Neighborhood Based Selection among friends of friends

In this model, a node from the friends of friends is chosen with probability proportional to number of common neighbors it shares with the node. Number of common nodes in neighborhood of node i and node j (number of common neighbors of node i and node j) is given by $|\boldsymbol{N}(i) \cap \boldsymbol{N}(j)|$. Hence, Probability for a node i to connect to a node j is given by

$$p_{ij} = \begin{cases} \frac{|\boldsymbol{N}(i) \cap \boldsymbol{N}(j)|}{\sum_{k \in \boldsymbol{\Gamma_2}(i)} |\boldsymbol{N}(i) \cap \boldsymbol{N}(k)|} & \forall j \in \boldsymbol{\Gamma_2}(i) \\ 0 & \text{otherwise.} \end{cases} \tag{3.1.4}$$

Denominator is again chosen such that the sum of probabilities to choose all allowed nodes is 1. Adding a link to friends of a friend creates as many new triangle in the network as the number of common neighbors between the two nodes. Clustering coefficient depends on number of these triangles. Hence, this model results into a network with high clustering coefficient as compared to the previous two models.

3.2 Simulation Results and other observations

We have simulated network generation with 128, 512 and 1024 nodes each having an average nodal degree of 8. Results of the simulations are tabulated in Table 3.1. We have compared the clustering coefficient of our models to Random Networks and Barabasi-Albert Preferential attachment model having similar average nodal degree. Our model provides a significant increment in the average clustering coefficient of the network. It is not surprising as we are making connections within friends of friends circle only. Out of three proposed models, Clustering coefficient is maximum in the network where probability is assigned proportional to number of common neighbors. Intuition tells that this is due to the fact that if a node with more number of common neighbors is given high probability, the increment in number of triangles centered at the node will be more.

In [19], Meghanathan has already shown that if nodes are connected only to two hop neighbors, resulting network demonstrate power law degree distribution. In our models, links are limited to two hop neighbours only as in [19], but the nodes are

Table 3.1: Average Local Clustering Coefficient

Type of Network	**N=128**	**N=512**	**N=1024**
Random among fof	0.555	0.467	0.466
Prefrential among fof	0.465	0.468	0.470
Common Neighbour fof	0.563	0.561	0.558
Random Network	0.062	0.015	0.008
Preferential Attachment	0.143	0.054	0.038

chosen according to a different preference. It means our model will also show power law degree distribution. To validate the statement, we have plotted degree distribution of networks having 1024 nodes and average degree 8 generated by all three proposed models. As evident from the figures 3.1 - 3.4, degree distribution is following power law in each of the three cases.

In case of Preferential attachment among fof, network has few nodes having a very large degree which is not the case in other two models. It is due to rich gets richer phenomenon observed in Barabasi-Albert model. Once the degree of a node is high it will always have higher probability to be chosen for new connections by the other nodes and this phenomenon keeps on increasing its selection probability with time. That's why we can observe a node having degree 661 in a network of size 1024. We have also plotted the degree distribution in logarithmic scale which is the best representation of power law distribution. In logarithmic scale, power law changes to a linear relation. To get the average statistical value, we have carried out the simulation 1000 times and averaged the results. As these very large degree nodes are very less and a node with 661 degree might have appeared only once in all these simulations. Next time number might have changed to 662 and so on. That's why we are getting fractional value of number of nodes in our simulation results and same is the reason why plot is linear till the average value of number of nodes is greater than 1 and afterwards it is distorted. We have not been able to show logarithmic plot in other two models because logarithmic scale is not

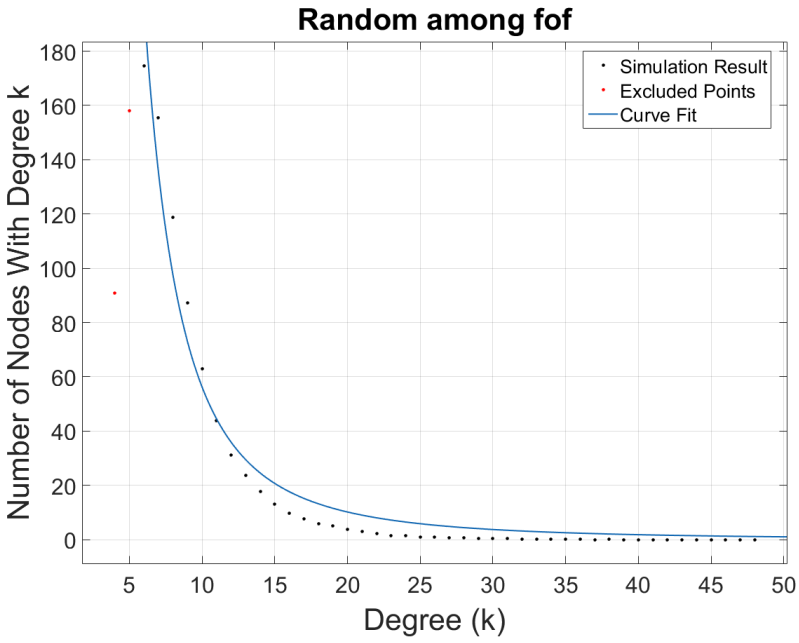

Figure 3.1: Degree distribution of a network generated by Random selection among friends of friends and having 1024 nodes and average degree 8.

very apt for smaller value and in those cases maximum nodal degree in the network is quite less. We may be able to see logarithmic scale plot even in these models if size of the network or average nodal degree of the network is large.

As far as small world property is concerned, it was present even in fundamental models (Random and Preferential Attachment) and is preserved in our network generation model too. From second iteration we are allowing connection only among second hop neighbors. In first iteration only, the connections are made randomly. That randomness results into an irregular network which is the basic feature responsible for small world property.

In this way, we can say that our models generate networks with all the three desired characteristics: Power law degree distribution, Small world property and high clustering coefficient.

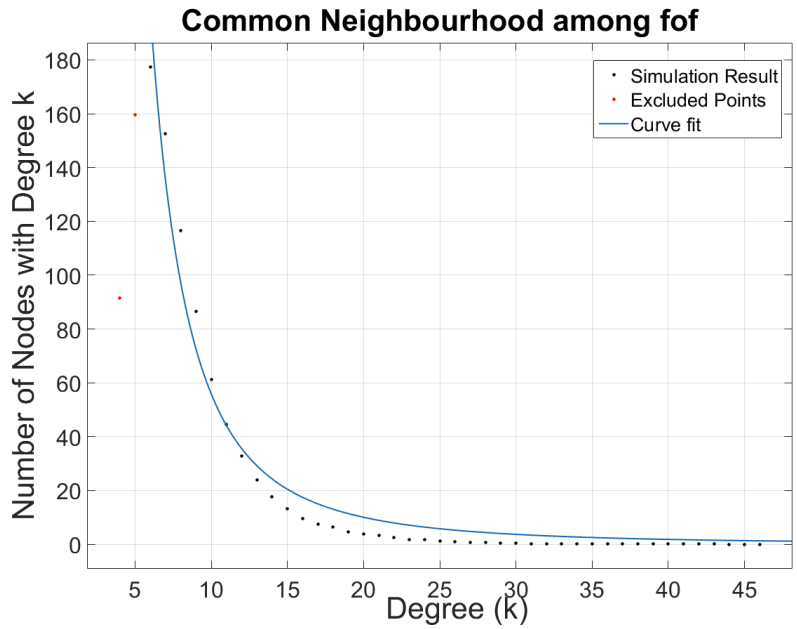

Figure 3.2: Degree distribution of a network generated by Common Neighborhood based selection among friends of friends and having 1024 nodes and average degree 8.

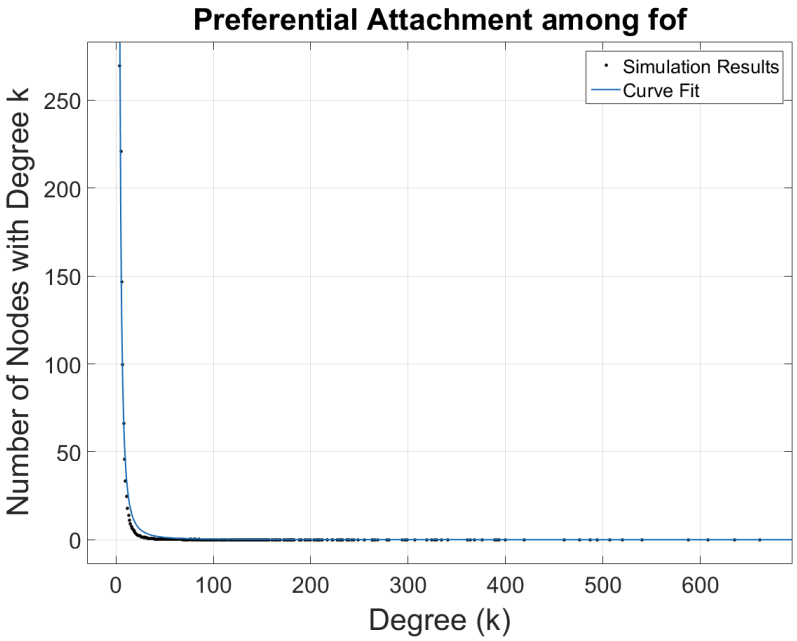

Figure 3.3: Degree distribution of a network generated by Preferential Attachment scheme among friends of friends and having 1024 nodes and average degree 8.

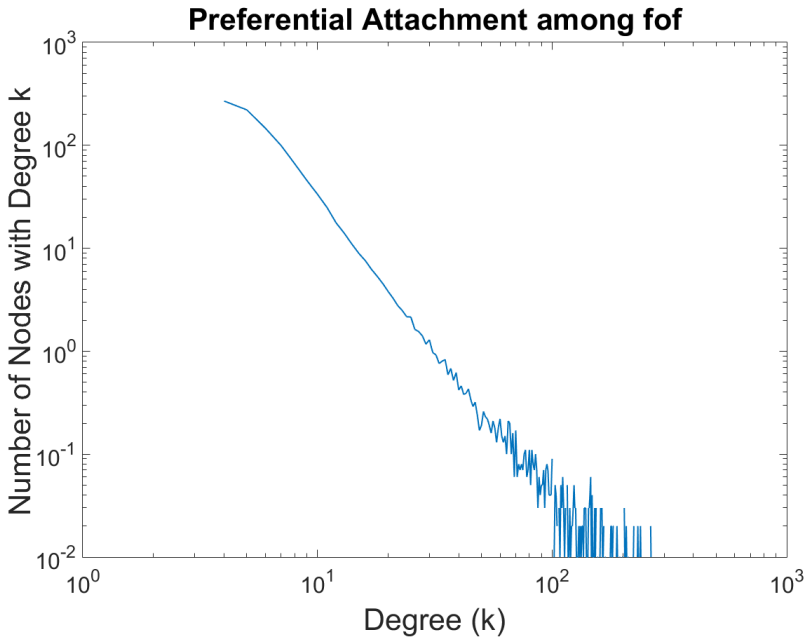

Figure 3.4: Degree distribution of a network (on logarithmic scale) generated by Preferential Attachment scheme among friends of friends and having 1024 nodes and average degree 8.

3.3 Models allowing connection among friends of friends as well as others

We have discussed in the beginning of this chapter that a person makes new friends via someone he already knows. In real world, this is not the only possible mechanism of making friends. Many a times we make friends on our own. We have some information about a particular person, we meet him and become friends. Sometimes we meet some people during morning walk, while having food in a restaurant, while traveling and during many other activities and end up making new friends. Some source of information has definitely played its role in this process but no individual person is involved. In all these interactions, mostly, a common friend is not needed. Considering such possibilities, we proposed new variants of our models, wherein a node is allowed to make connection not only among friends of friends but also among others.

3.3.1 Generalised random network with common neighborhood preference

In this model, a node has an option to choose nodes either from friends of friends circle or from other remaining nodes. We have assigned a probability f with which a node takes decision in favour of friends of friends circle and with remaining probability $(1 - f)$ the decision is taken otherwise. We can say that f reflects the tendency of a node to make connections among friends of friends. This tendency may differ from society to society. In fact the tendency will be different even for different person but we have assigned equal value of f for every node considering f to be the average tendency of nodes in the network.

After deciding whether a node will be selected from friends of friends or from outside, next question is how to choose a node from that particular group. In this model, the probability to choose a node from friends of friends is proportional to number of nodes in common neighborhood and to choose one from remaining nodes is random. The scheme is apt for the situation when we meet a person accidentally and become friends. $\boldsymbol{N}(i)$ is the set of neighborhood of node i i.e. nodes to which node i is already connected. $\boldsymbol{\Gamma_2}(i)$ is the set of nodes at distance 2 from node i. The set corresponding to $(1-f)$ fraction is $\boldsymbol{V} \setminus \{\{i\} \cup \boldsymbol{N}(i) \cup \boldsymbol{\Gamma_2}(i)\}$. For simplicity, this set is denoted by $\boldsymbol{S}(i)$. Cardinality of this set is $(\mathrm{N} - 1 - |\boldsymbol{N}(i) \cup \boldsymbol{\Gamma_2}(i)|)$. Hence, the probability for node i to connect to node j is given by

$$p_{ij} = \begin{cases} f \times \frac{|\boldsymbol{N}(i) \cap \boldsymbol{N}(j)|}{\sum_{k \in \boldsymbol{\Gamma_2}(i)} |\boldsymbol{N}(i) \cap \boldsymbol{N}(k)|} & \forall j \in \boldsymbol{\Gamma_2}(i) \\ (1-f) \times \frac{1}{\mathrm{N} - 1 - |\boldsymbol{N}(i) \cup \boldsymbol{\Gamma_2}(i)|} & \forall j \in \boldsymbol{S}(i) \\ 0 & \text{otherwise.} \end{cases} \tag{3.3.5}$$

Apart from $\boldsymbol{\Gamma_2}(i)$ and $\boldsymbol{S}(i)$, remaining nodes are i itself and its direct neighbors which are already connected. Zero probability assigned in third part of the expression 3.3.5 corresponds to these remaining nodes.

3.3.2 Generalised PA network with common neighborhood preference

In this model, connections are made on the same basis as done in the previous variant except the probability to connect to rest of the network (nodes at distance three or more) depends on degree of the nodes as in Preferential attachment model of Barabasi and

Albert. Hence, probability for node i to connect to a node j in this model is given by

$$p_{ij} = \begin{cases} f \times \frac{|\boldsymbol{N}(i) \cap \boldsymbol{N}(j)|}{\sum_{k \in \boldsymbol{\Gamma_2}(i)} |\boldsymbol{N}(i) \cap \boldsymbol{N}(k)|} & \forall j \in \boldsymbol{\Gamma_2}(i) \\ (1-f) \times \frac{|\boldsymbol{N}(j)|}{\sum_{k \in \mathbf{S}(i)} |\boldsymbol{N}(k)|} & \forall j \in \boldsymbol{S}(i) \\ 0 & \text{otherwise.} \end{cases} \tag{3.3.6}$$

3.4 Simulation Results and other observations

We have again simulated network with 128, 512 and 1024 nodes and each having an average nodal degree 8. Fraction f is varied from 0.1 to 1 with an equal intervals of 0.1. Table 3.2 gives clustering coefficient in case of Generalized Random Network with common neighborhood preference among friends of friends. Here, connection is allowed among whole of the network with definite fraction f within friends of friend. Fraction f=1 means probability to join among friends of friends is 1 and among others is 0. It is equivalent to our third model discussed in section 3.1.3 where we allow connections among friends of friends only and probability to connect was assigned on the basis of size of common neighborhood. The fact is supported by clustering coefficient found in our simulations and listed in third row of lines of Table 3.1 and last row of Table 3.2.

As shown in Table 3.2, clustering coefficient increases monotonically with increase in value of f. This trend persists irrespective of size of the network. For a fixed value of f average clustering coefficient reduces consistently with increase in N, though by very small value. Reduction is more when we are considering networks of size 124 and 512 as compared to the case between networks of size 512 and 1024. It is also evident from overlapping curves of Fig. 3.5, where clustering coefficient as a function of f has been plotted for Generalised Preferential Attachment Model having sizes 128, 512 and 1024.

Table 3.2: Average Clustering coefficient in Generalized Random Network

fraction f	**N=128**	**N=512**	**N=1024**
0.1	0.043	0.034	0.034
0.2	0.085	0.070	0.067
0.3	0.125	0.106	0.103
0.4	0.167	0.145	0.141
0.5	0.205	0.186	0.183
0.6	0.255	0.234	0.230
0.7	0.311	0.288	0.285
0.8	0.374	0.356	0.352
0.9	0.452	0.440	0.439
1	0.563	0.558	0.558

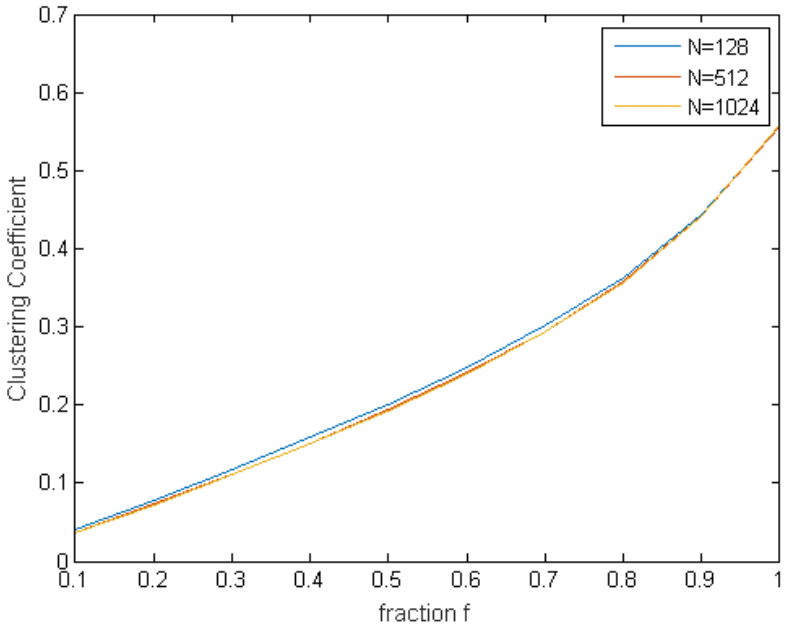

Figure 3.5: Variation in clustering coefficient with increase in the value of the fraction f for three networks of different sizes.

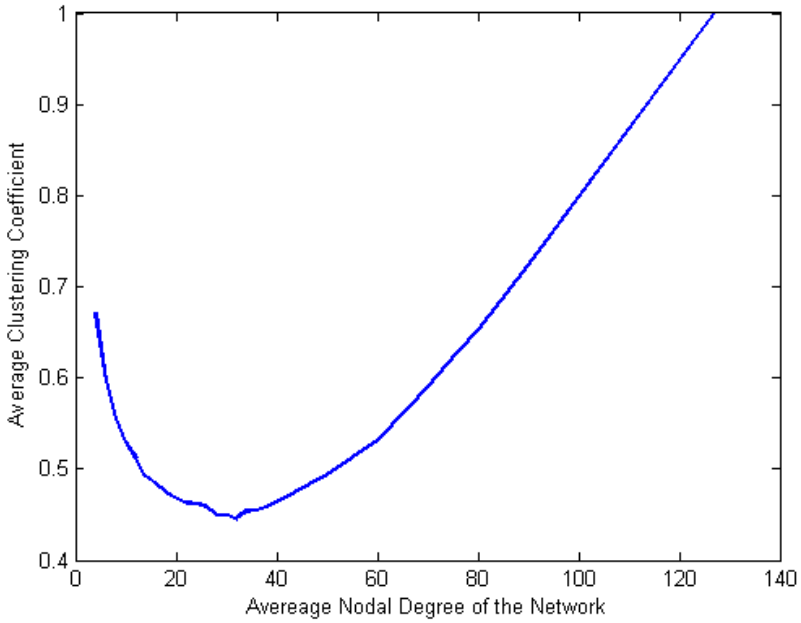

Figure 3.6: Variation in clustering coefficient with average degree of the network having 1024 nodes.

Although we have performed our simulation upto 1024 nodes only (due to memory limitation), but seeing the pattern we can predict that average clustering coefficient of the network will remain almost same even if number of nodes is increased further keeping average nodal degree fixed. This feature indicates that at local level, network is similar, even though size of the network is increasing. Mathematically, clustering coefficient converges to a finite nonzero value asymptotically for large number of nodes which is a common characteristic of almost all the real networks.

In Fig. 3.6, we have shown the variation in clustering coefficient with average degree of the network keeping N and f fixed. Results show that clustering coefficient of network decreases with increase in average degree of the network, reaches a minimum and then increases again. So we infer that apart from fraction f, the clustering coefficient is also a function of average nodal degree of the network.

3.5 Conclusions

1. Clustering coefficient is largest when connections are made among neighbors on the basis of number of nodes in common neighborhood.

2. In proposed model of network generation, clustering coefficient does not decreases to zero even if the network size grows infinitely.

3. We can create a network with desired level of clustering coefficient by changing the value of fraction f.

4. Clustering coefficient is a combined function of average degree of the network and fraction f.

Chapter 4

Community Generation and estimation of the fraction f

People having common interest tend to form their own groups. It leads to community structure in a network. Communities have been observed in many networks including social network, computer network, biological network and animal network [23]-[27].

When we try to generate network using algorithms, the community structure similar to real life network are desirable. To create communities, basic procedure is to start with a few communities, assign an incoming node to a particular community and create desired number of links to nodes within the same community and rest with nodes of other communities. The Choice of assigning a community to a node can be random or any other mechanism. Xu *et.al.*[24] pointed out that like nodal degrees, the preferential mechanism also exists in the evolution of communities. For example, while adding a new node to an existing community, communities with larger sizes are selected with higher probabilities. Xie *et.al.* [18] used this phenomenon of rich gets richer in terms of community scale to propose their model of evolving network. In all these models, network evolves with time but the number of communities in the network is always kept

constant. Later [25] proposed the model where number of communities can also change with time.

There are many existing models which generate networks having communities but most of these models grow from an inbuilt community structure whereas the community structure in real networks is formed along with the process of network evolution. It has been observed that Random network model or preferential attachment model of network generation are not able to generate communities as they exist in real world networks.

4.1 Community Structure and Modularity Index

There are various characterizing parameters for community structure in a network. One of them is modularity index. Extent to which a community is present in a network is depicted by the Modularity Index [28]. Modularity is the fraction of edges that fall within the given community minus the expected fraction if edges were to be distributed randomly across the whole network. It is zero in case of a random network; positive if the number of edges within groups is more than the number expected randomly and vice versa.

We have carried out community detection in the network by Newman fast algorithm [29]. It gives modularity index corresponding to all possible number of communities in the network. We have considered the number of communities which maximizes the modularity index. In Fig 4.1 we have shown the result of Newman fast algorithm carried over a network of 128 nodes generated by Generalised random scheme with common neighborhood preference among friends of friends with preference f=0.6. We get modularity index corresponding to number of communities in which we intend to

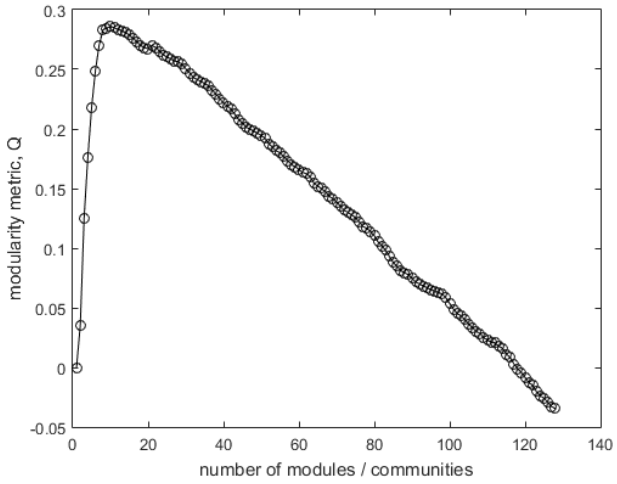

Figure 4.1: Horizontal axis is number of communities in which network has been divided and on vertical axis is the corresponding modularity index. Network considered is Generalised random with common neighborhood preference among friends of friends with preference f=0.6 and total number of nodes N=128.

Table 4.1: Comparision of Modularity Index

Type of Network	**N=128**	**N=512**	**N=1024**
Random Network	0	0	0
Random Selection among fof	0.610	0.799	0.808
Prefrential Attachment among fof	0.366	0.488	0.458
Common Neighborhood based selection among fof	0.606	0.823	0.827

divide the whole network. In our example, modularity index is maximum when 10 communities are assumed to exist in the network.

We have calculated the modularity index of networks created by all three basic network generation models proposed in Chapter 3 i.e., Random selection among fof, Preferential attachment among fof and Common neighborhood based selection among fof. We have again considered networks having 128, 512 and 1024 nodes. Results have been summarised in Table 4.1. Modularity index in case of Random Network is zero. In all other three networks where connection is allowed only among friends of friends, modularity index is quite high (in the range 0.35-0.85). Among these three networks,

Table 4.2: Number of communities (C) and Modularity Index (Q) in Generalized Network Models

f	**C in GR**	**Q in GR**	**C in GPA**	**Q in GPA**
0.1	21	-0.015	17	-0.111
0.2	17	-0.131	17	-0.100
0.3	16	0.201	15	0.170
0.4	15	0.194	15	0.176
0.5	12	0.208	13	0.179
0.6	10	0.287	11	0.260
0.7	9	0.257	10	0.275
0.8	8	0.418	9	0.408
0.9	6	0.481	8	0.507
1	5	0.578	6	0.660

modularity is least when preferential attachment is used as the selection strategy. It is because of the hubs present in the network, which are generated by preferential attachment. Hubs have very large number of neighbors. Every node of the network is generally connected to one or more hubs. It is obvious that a node will not have connection to many of the friends of these hubs. Other important point which can be deduced from the Table 4.1 is very high modularity index in many cases. For example, for a network of 1024 nodes with average nodal degree of 8, modularity index comes out to be around 0.8 which is far more than real world networks. This result indicates that even in real world, node is not making friends with friend of friend alone rather with others also.

In generalized model when we allow connection among outer world also, modularity index decreases as shown in Table 4.2. In this table, number of communities and modularity index have been denoted by C and Q respectively. Generalized random network with common neighborhood preference among friends of friends is written as GR and its preferential attachment variant has been written as GPA. For very small values of f i.e., probability of a node to make new connection among fof, like 0.1 and

0.2 modularity index is coming out to be negative implying the absence of community structure. Modularity index is monotonically increasing with fraction f and we have many instances when modularity index is in the range of 0.25 to 0.5, which is observed in many real world networks. In this way we have been able to generate communities in the network. Hence our claim that community can be generated by making connection among friends of friends in iteration is justified. Metric f in that sense does not take care about clustering coefficient alone rather it also takes care about communities in the network.

4.2 Estimation of fraction f

Till now, we have seen the effect of the fraction f on networks generation. We have seen how variation in f can result into network of desired clustering coefficient along with all other necessary features like power law degree distribution and small diameter in Section 3.2 and 3.4. We have also seen how incorporating f in our model results in communities without doing any thing specific for it. Now, we are trying to estimate the value of this fraction from a given network structure.

In a growing network, whenever a link is created by a node, it can easily be observed whether the link was created among friends of friends or outside that circle. For any particular node, it can be easily found that out of total link created by the node how many are among fof and how many are outside this domain. Hence, fraction of links created among fof can be calculated for each node.

Average of f's for all nodes present in the network will give average value of the fraction f for whole network. It will give average tendency of nodes in the network to

make new friends among friends of friends.

If we know initial structure of the network as well as the sequence in which links were created then it is very easy to calculate the average fraction f. But, if we do not have the temporal data i.e. we do not know the sequence in which links were created then it is not possible to decide which links were created among fof and which were not. Hence, value of f can not be found directly. In next section, we have tried to estimate this value where sequence of link formation is not known.

4.2.1 Random Sequencing

Let us assume that network has n nodes and m links. A random sequence of number from 1 to m is assigned to m links of the network. We assume that link with assignment 1 has been created first, then 2, then 3 and so on. In this way we have a time-stamp for each link in the network and we have generated a dynamic network from given static network. It is now easy to determine whether a link was connected among fof or outside. Hence, the value of f can be estimated for a node and can be further averaged to find network level value.

We have considered both Random as well as Scale free network for simulation. After generating a network, a random time stamp is assigned to each of the links and average value of the fraction f is calculated. We have repeated the experiment for a particular network structure with different random sequence and observed that estimated value of f is almost same in each of them, varying only after second decimal places. Average value of f for network of different size and nodal degree are summarized in Table 4.3.

From the simulation result, we estimated the average value of f but, it is not clear

Table 4.3: Range of estimated values of f for different networks

Type of Network	**Random Network**		**Scale Free Network**	
Avg. Degree	N=200	N=1000	N=200	N=1000
6	0.039-0.084	0.009-0.016	0.090-0.221	0.040-.051
10	0.130-0.170	0.029-0.037	0.209-0.291	0.091-0.110
20	0.390-0.445	0.112-0.134	0.470-0.520	0.224-0.256

whether the value obtained by random sequencing strategy is equal to actual f or not. To check the validity of our result we generated a dynamic network and noted down the sequence in which links were created. From this sequence we calculated the average f. We then applied random sequencing methodology on the final generated network and estimated the value of f. Similar result in these two schemes will confirm that our estimation is right.

As discussed in previous chapter, Generalised random network strategy with common neighborhood preference among friends of friends has been used to create a network of 1000 nodes and 10000 links making average degree 20. Probability to connect among friends of friends has been considered to be 0.6. Although probability is 0.6 for every node but in the resulting network, fraction f for individual nodes vary from 0.2 to 0.9. Histogram of the fraction f, plotted in the Fig.4.2, suggests that fraction f follows Gaussian distribution. We have also plotted the Gaussian fit of the obtained histogram in the same figure. Mean of the Gaussian distribution is 0.5731 and standard deviation is 0.1264. Mean of the distribution is closer to the value assigned to fraction f, which is 0.6 in this case.

Result of Random sequencing estimation has been plotted in Fig.4.3. Average value of estimated f is around 0.39 which is quite far from the actual value 0.6. As we are assigning the sequence randomly, we have carried out the simulation 20 times and taken the average of the result. Distribution is Gaussian with mean 0.3890 and standard

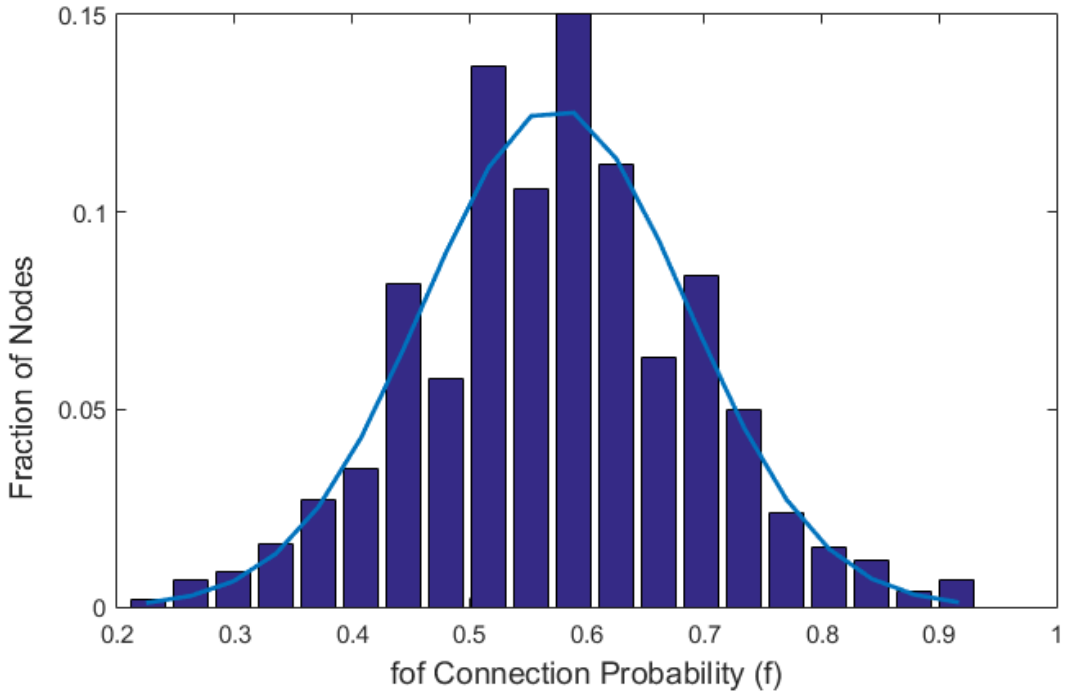

Figure 4.2: Histogram and the Gaussian fit of the fraction f of the network considering actual sequence in which links were created. Probability to connect among fof was considered to be 0.6 while generating the network.

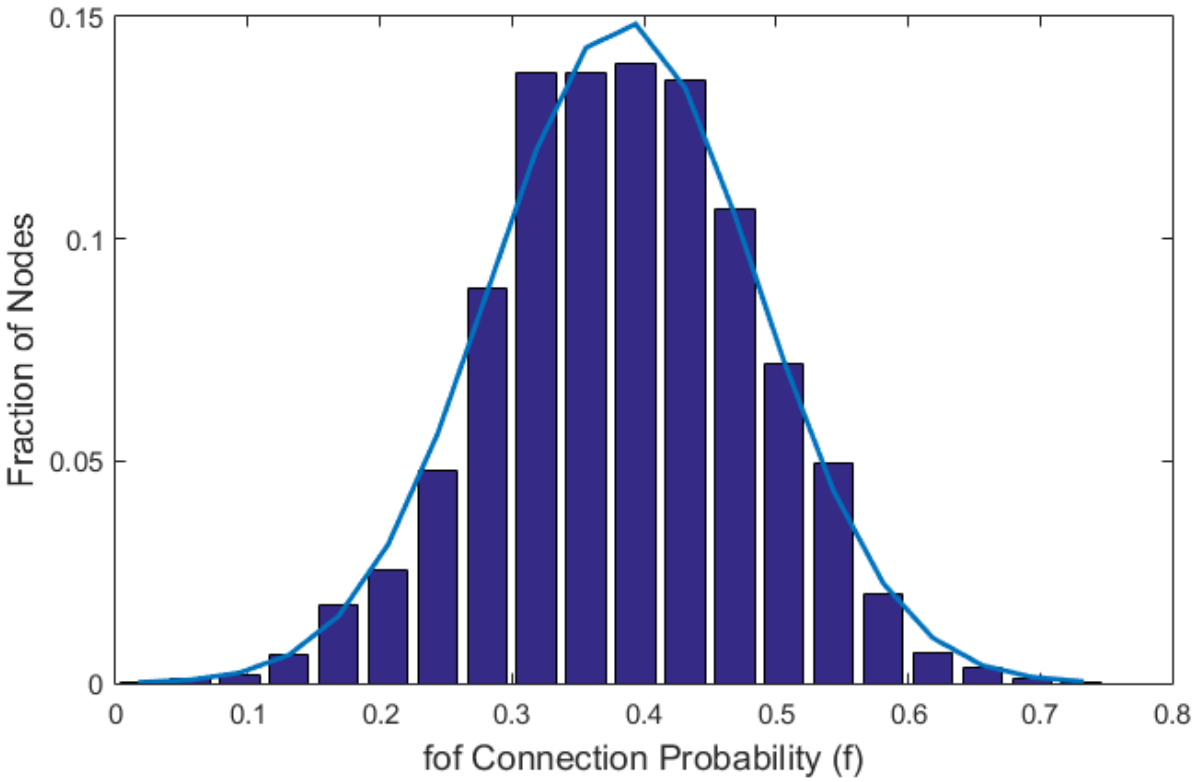

Figure 4.3: Histogram and the Gaussian fit of the fraction f of the network considering random sequencing. Probability to connect among fof was considered to be 0.6 while generating the network.

deviation 0.1407. We have carried out the similar simulation for different values of f and compared the results in Table 4.4. All these results confirm that there is a large error in estimation of f by random sequencing strategy.

Hence, prediction using random sequencing is not accurate. We need the information about sequence of link creation in order to predict the true value of f.

Table 4.4: Comparison of actual value of f and its estimated value

Actual f	**f calculated by actual sequence**	**f estimated by random sequence**
0.8	0.7662	0.5241
0.6	0.5739	0.3890
0.4	0.3866	0.3144
0.2	0.1892	0.2409

4.2.2 Random sequencing with multiple image of network

Let us suppose that we do not have complete information about evolution of the network, but we have the network structure at few time instants. For example let us say a network is one month old and we know the structure of network at regular interval of one week. It means we have four images of network structure at regular time interval. We do not have the complete information regarding sequence in which links were created but we have partial information about this sequence. We atleast know which links were created in first week, which in second week and so on.

At first, we have considered the examples when we have structures at two different instances. Even now, we do not know the complete sequence of the link creation but we know which links were created before a particular time and which were created after that. Now in our generated network which is being used for verifying our results, we randomize the sequence such that all links created before the particular time is kept before and remaining links are kept afterwards. we see that our result move closer to the actual f. With structures at three time instances results move more closer and so on. We have used a network of 1000 nodes, average degree 20 and probability of making connection among fof i.e., f=0.6. Table summarizes the mean and standard deviation of the distribution of f when network structure at only one, two, three and four instances are available with us. Same has been plotted in Fig.4.4. We can see

Table 4.5: Estimated Gaussian fit of distribution of f

Availability	**Mean**	**Standard Deviation**
Network Structure at one instance	0.3890	0.1407
Network structure at two instances	0.4734	0.1533
Network structure at three instances	0.5033	0.1574
Network structure at four instances	0.5184	0.1592

that as we are increasing the number of network images, distribution curve moves right, closer to the actual distribution.

It means in a scenario where it is possible to know the structure of network at different time instances we will be able to predict the value of f better.

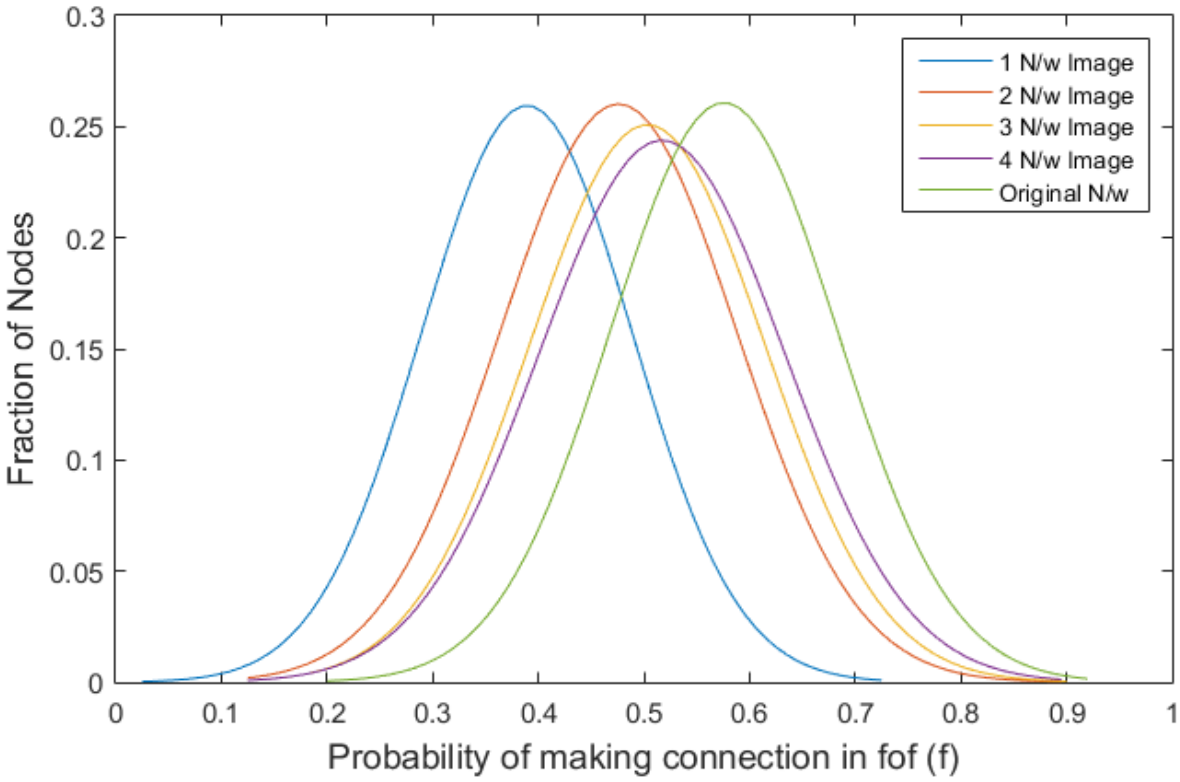

Figure 4.4: Comparison of distribution of the fraction f when network structure is available at one, two, three and four instances with original distribution. Distribution is moving closer to original as number of available images increases.

Chapter 5

Conclusions and Future works

Network structure is the backbone of all the processes going on the network. Structure of the network impacts the dynamics of these processes. To understand these processes synthetic network is generated and used for the simulation. Network is generated such that it is similar to real world network in its characteristics.

In this work, a new method of network generation has been proposed which generates a network having all three desirable characteristics i.e., small world property, power law degree distribution and high clustering coefficient. It has been achieved by accounting for two real life observations:

1. A person does not make all its friends in one go rather he/she does so at intervals.
2. It is more probable that a person will make a new friend among his/her friends of friends circle.

First observation has been added in the process of generation by allowing node to make new connections in iteration rather than once. Second observation has been included by introducing a probability (f) of making connection among friends of friends. While

creating a new connection, node picks up a node from his friends of friends circle with probability f and a node from outside this circle with probability 1-f. We have shown that including these two features in generation scheme, we are able to generate a network having all three desired quality simultaneously. The parameter f can be varied to generate network with desired level of clustering. One additional benefit of our model is that without doing anything specific for communities, we are able to generate a network having communities which is a common characteristic of most of the real world networks.

In second part of our work, we have tried to estimate the affinity with which nodes of the network are interested to make connection among friends of friends circle i.e., We have tried to estimate the average probability f from given network structure. We have shown that it is not possible to estimate the value of f using random sequencing scheme in networks where we do not have the information about the sequence in which links were created. But, if network is dynamically changing and we have network structure at multiple time instances, we can estimate the value of f more accurately.

Many more problems worth further investigations have been found during the course of this work. Some of the problem which can be further pursued are given below:

1. One obvious extension of the work is to investigate the value of f for real networks where temporal data is available with us.

2. While generating network, we have allowed a node to make one connection in every iteration. Although the proposed strategy relaxes the constraint of making all friends at once but node is not free to make his/her independent choice. In extension of our work, we can allow node to choose whether he/she is interested in making a new connection or not. We can even assign a probability with which

node is willing to make a friend or not.

3. At present, we have chosen a fix value of probability f for every node. We can use any apt probability distribution scheme instead. For choice of the distribution, we again need to investigate some real network data sets with temporal information.

Bibliography

[1] Mitra, K., Carvunis, A.R., Ramesh, S.K. and Ideker, T., 2013. Integrative approaches for finding modular structure in biological networks. Nature Reviews Genetics, 14(10), p.719.

[2] Pagani, G.A. and Aiello, M., 2013. The power grid as a complex network: a survey. Physica A: Statistical Mechanics and its Applications, 392(11), pp.2688-2700.

[3] Rao, V.S., Srinivas, K., Sujini, G.N. and Kumar, G.N., 2014. Protein-protein interaction detection: methods and analysis. International journal of proteomics, 2014.

[4] Thompson, R.M., Brose, U., Dunne, J.A., Hall Jr, R.O., Hladyz, S., Kitching, R.L., Martinez, N.D., Rantala, H., Romanuk, T.N., Stouffer, D.B. and Tylianakis, J.M., 2012. Food webs: reconciling the structure and function of biodiversity. Trends in ecology and evolution, 27(12), pp.689-697.

[5] Ganguly, N., Krueger, T., Mukherjee, A. and Saha, S., 2014. Epidemic spreading through direct and indirect interactions. Physical Review E, 90(3), p.032808.

[6] Newman, M., 2010. Networks: an introduction. Oxford university press.

[7] Barabsi, A.L., 2016. Network science. Cambridge university press.

[8] Zaidi, F., 2013. Small world networks and clustered small world networks with random connectivity. Social Network Analysis and Mining, 3(1), pp.51-63.

[9] Guare, J., 1990. Six degrees of separation: A play. Vintage.

[10] Onnela, J.P., Saramki, J., Hyvnen, J., Szab, G., Lazer, D., Kaski, K., Kertsz, J. and Barabsi, A.L., 2007. Structure and tie strengths in mobile communication networks. Proceedings of the national academy of sciences, 104(18), pp.7332-7336.

[11] Barabsi, A.L. and Albert, R., 1999. Emergence of scaling in random networks. science, 286(5439), pp.509-512.

[12] Barabsi, A.L., 2009. Scale-free networks: a decade and beyond. science, 325(5939), pp.412-413.

[13] Erds, P. and Rnyi, A., 1960. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci, 5, pp.17-61.

[14] Watts, D.J. and Strogatz, S.H., 1998. Collective dynamics of small-worldnetworks. nature, 393(6684), p.440.

[15] Barabsi, A.L., Albert, R. and Jeong, H., 2000. Scale-free characteristics of random networks: the topology of the world-wide web. Physica A: statistical mechanics and its applications, 281(1-4), pp.69-77.

[16] Barabsi, A.L., Albert, R. and Jeong, H., 1999. Mean-field theory for scale-free random networks. Physica A: Statistical Mechanics and its Applications, 272(1-2), pp.173-187.

[17] Jost, J. and Joy, M.P., 2002. Evolving networks with distance preferences. Physical Review E, 66(3), p.036126.

[18] Xie, Z., Li, X. and Wang, X., 2007. A new community-based evolving network model. Physica A: Statistical Mechanics and its Applications, 384(2), pp.725-732.

[19] Meghanathan, N., 2015, November. A Random Network Model with High Clustering Coefficient and Variation in Node Degree. In Control and Automation (CA), 2015 8th International Conference on (pp. 54-57). IEEE.

[20] Chen, J., Geyer, W., Dugan, C., Muller, M. and Guy, I., 2009, April. Make new friends, but keep the old: recommending people on social networking sites. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 201-210). ACM.

[21] Backstrom, L. and Leskovec, J., 2011, February. Supervised random walks: predicting and recommending links in social networks. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 635-644). ACM.

[22] Liben-Nowell, D. and Kleinberg, J., 2007. The link-prediction problem for social networks. journal of the Association for Information Science and Technology, 58(7), pp.1019-1031.

[23] Arenas, A., Danon, L., Diaz-Guilera, A., Gleiser, P.M. and Guimera, R., 2004. Community analysis in social networks. The European Physical Journal B, 38(2), pp.373-380.

[24] Xu, X.J., Zhang, X. and Mendes, J.F.F., 2009. Growing community networks with local events. Physica A: Statistical Mechanics and its Applications, 388(7), pp.1273-1278.

[25] Sallaberry, A., Zaidi, F. and Melanon, G., 2013. Model for generating artificial social networks having community structures with small-world and scale-free properties. Social Network Analysis and Mining, 3(3), pp.597-609.

[26] Girvan, M. and Newman, M.E., 2001. Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA, 99(cond-mat/0112110), pp.8271-8276.

[27] Andrade, R.F., Rocha-Neto, I.C., Santos, L.B., de Santana, C.N., Diniz, M.V., Lobao, T.P., Gos-Neto, A., Pinho, S.T. and El-Hani, C.N., 2011. Detecting network communities: An application to phylogenetic analysis. PLoS Computational Biology, 7(5), p.e1001131.

[28] Newman, M.E., 2006. Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), pp.8577-8582.

[29] Newman, M.E., 2004. Fast algorithm for detecting community structure in networks. Physical review E, 69(6), p.066133.

[30] Saramki, J. and Kaski, K., 2004. Scale-free networks generated by random walkers. Physica A: Statistical Mechanics and its Applications, 341, pp.80-86.

[31] Herrera, C. and Zufiria, P.J., 2011, June. Generating scale-free networks with adjustable clustering coefficient via random walks. In Network Science Workshop (NSW), 2011 IEEE (pp. 167-172). IEEE.