New Network Generation Models Using Concept of Friends of Friends

Md Arsad (12040880)
Aniket Mohanthy(12040200)

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 the 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 known as network science.

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 a new connection via someone he/she already knows i.e., chances to make new connections 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.

When these two traits are combined, the network gains the small world property, the scale-free degree distribution, and a high clustering coefficient. Another benefit is that when network size grows, the average Clustering Coefficient does not decrease to zero.

A new parameter f i.e., tendency of a node to make new connections among friends of friends circle has been introduced. It has been shown that by varying the parameter f, we can generate a network having desired level of clustering coefficient.

Introduction

Networks are a collection of nodes or vertices connected by edges, and their structure and pattern of interactions have a significant impact on the system they represent.Real networks possess many characteristics.
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 behavior.
Apart from models of network generation, network science deals with measurement of network properties, analysis of network data and study of dynamic processes on networks like percolation, epidemic spread and network search.

Three Important Characteristics of Real networks.

Different real world networks 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.**

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 is small.

2.Scale Free Network: A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. $\mathbf{pk \sim k^{-\alpha}}$

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.

3. High Clustering Coefficient: 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.

There are various Network Generation Models exist based on some properties of the network for example Erd˝os-R´enyi Random Graph Model,Watts-Strogatz Model, Barabasi-Albert (BA) Preferential Attachment Model etc.

If we analyze the Erd˝os-R´enyi Random Graph Model we found that 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 , random networks possess only one i.e. small world phenomenon. Scale free property and high clustering coefficient are missing in the random networks.
Similarly for Watts-Strogatz Model possesses small world effect and high clustering coefficient but lacks scale free property, and Barabasi-Albert 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.

Our Proposed Model

I have tried to design 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.

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 a network with desired level of clustering.
2. Without doing anything specific for communities, our model generates communities which is characteristic of many real networks.

Our proposed Model has two main features . The first feature is the concept of connecting with friends of friends, which is commonly used in social networking sites to recommend new friends to users. In this model, connections are made preferentially among friends of friends rather than higher degree nodes.

The second feature is the timing of making connections. In the Barabasi-Albert Model, nodes are only allowed to make connections when they enter the network. However, in reality, people do not make all their friends at once but rather add them gradually over time. To account for this, the model allows nodes to make connections in iterations, with each node creating one link per round. The network can only move to the next iteration after every node has made one connection in the current iteration.

To construct a network, following steps are followed:

1. In the first iteration, at each timestep, a new node is added to the network with one link that connects the new node to any one of the existing nodes with uniform probability. After the first iteration it will form a connected graph of n nodes and n-1 edges.

2. In subsequent iterations, each node is allowed to make one new connection in every iteration.

Two important decisions to be made are:

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

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

A. Models allowing connection among 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.
Hence, the probability for a node i to connect to a node j is given by

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

Here, |Γ2(i)| is the cardinality of the set Γ2(i), representing the number of elements in the set.

B. Preferential Attachment among friends of friends:

In this variation, like in the 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 the preferential attachment scheme of Barabasi-Albert model.
Hence, the probability for a node i to connect to a node j is given by

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

C. Common Neighborhood Based Selection among friends of friends:

In this model, a node from the friends of friends is chosen with probability proportional to the number of common neighbors it shares with the node**.**

Hence, Probability for a node i to connect to a node j is given by

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

Adding a link to friends of a friend creates as many new triangles in the network as the number of common neighbors between the two nodes. Clustering coefficient depends on the number of these triangles. Hence, this model results in a network with high clustering coefficient as compared to the previous two models.

D. Models allowing connection among friends of friends as well as others

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 a decision in favour of friends of friends circle and with remaining probability (1 − f) the decision is taken otherwise.

It can control the average clustering coefficient of the network. In this model, the probability to choose a node from friends of friends is proportional to the number of nodes in a 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.

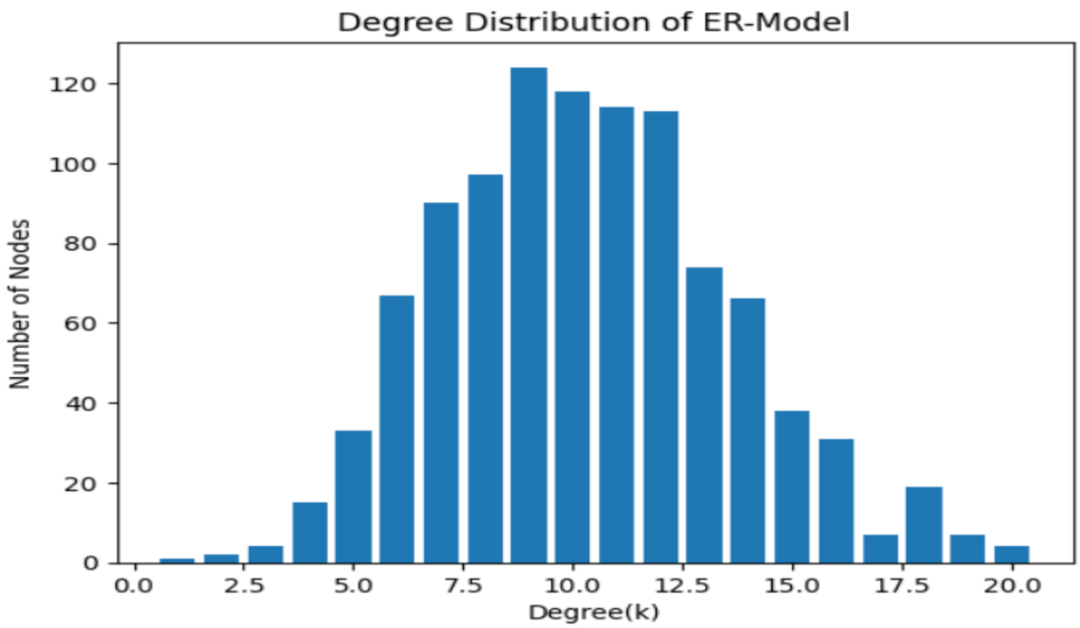
The probability for node i to connect to node j is given by

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

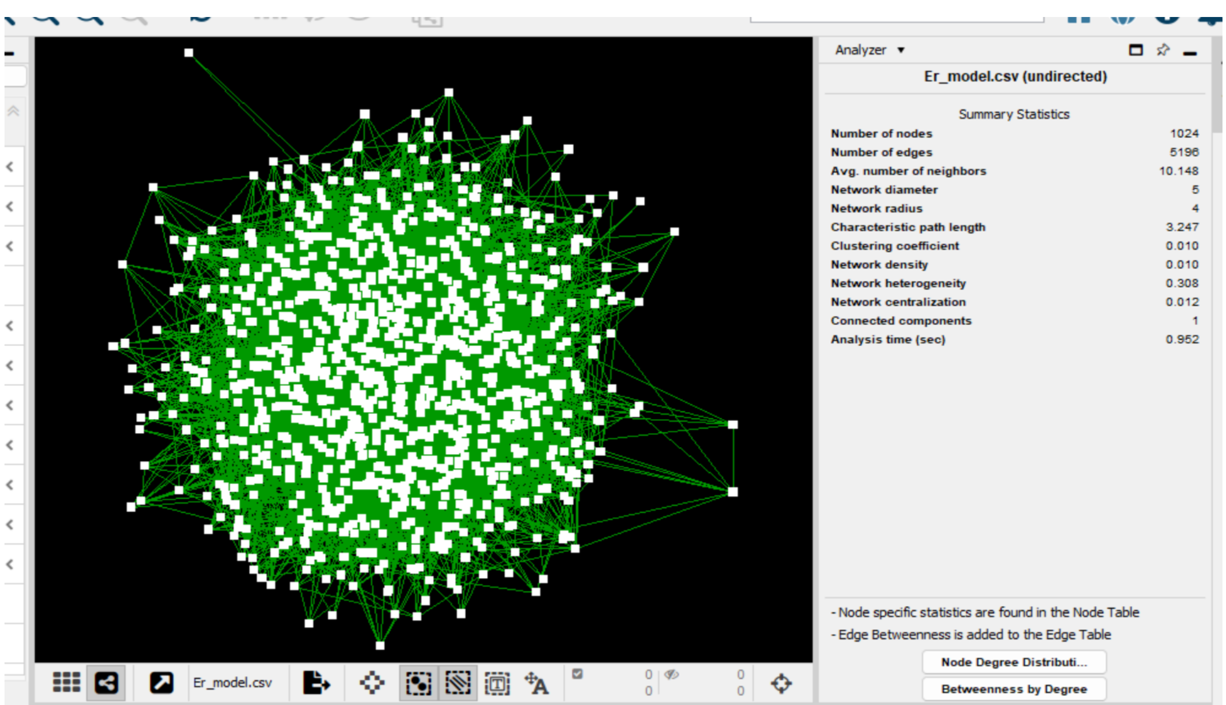
Where set $S(i) = V \setminus \{\{i\} \cup N(i) \cup \Gamma_2(i)\}$ i.e set of nodes which is at a distance more then two.

Analysis of existing Network Generator Model
ER-Model:

```
number of nodes : 1024
number of edges : 5291
Average degree: 10.333984375
Average clustering coefficient: 0.011224258360708035
```

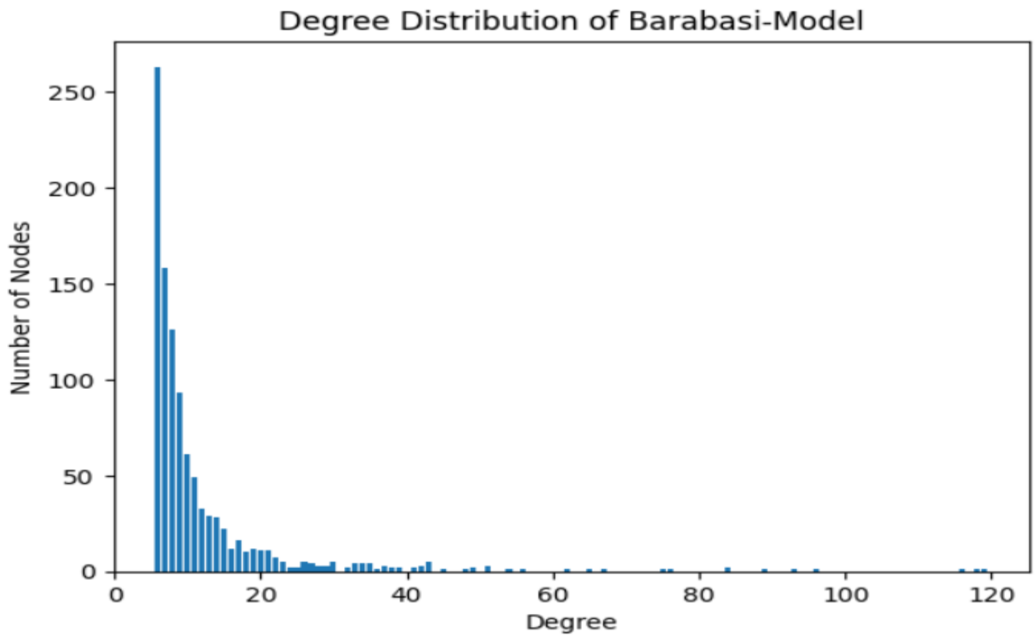


Visualization in Cytoscape

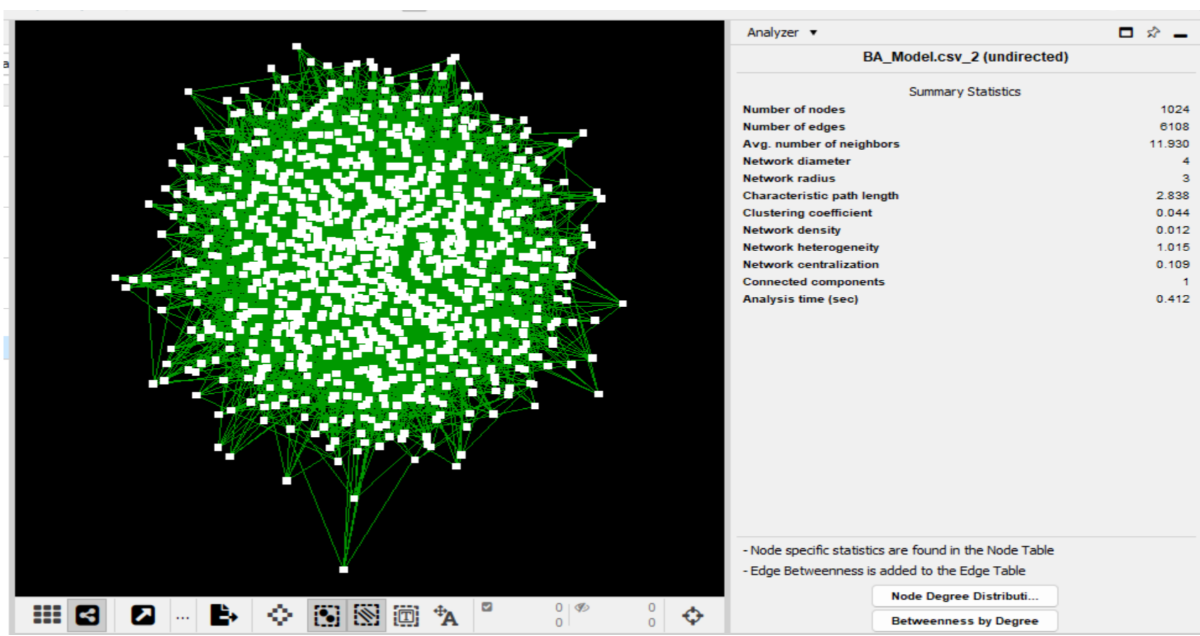


Barabasi-Albert Model

```
number of nodes : 1024
number of edges : 6108
Average degree: 11.9296875
Average clustering coefficient: 0.04372936342936496
```



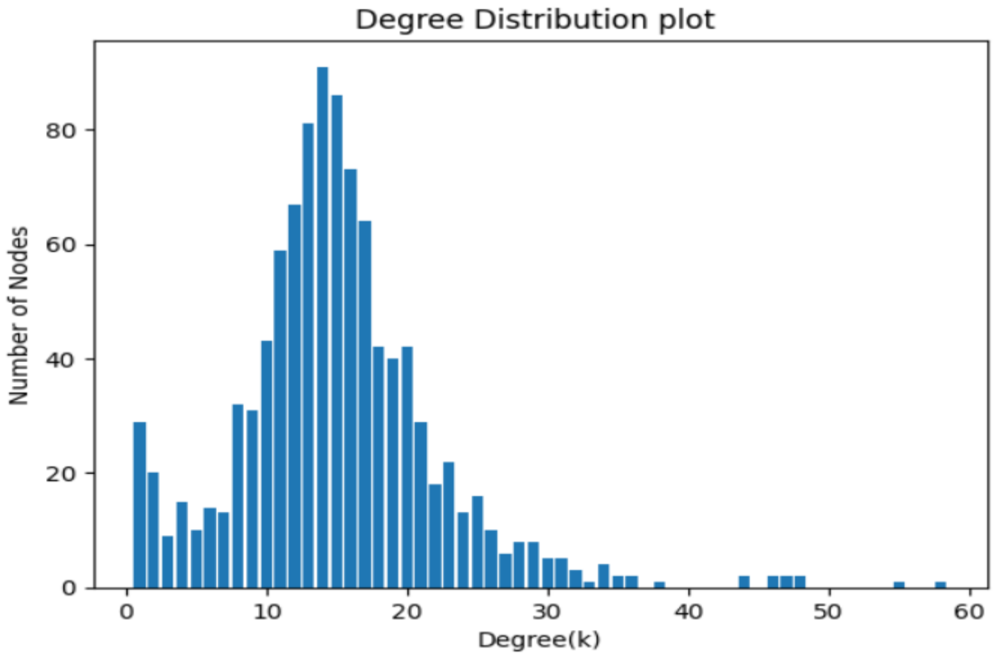
Visualization in Cytoscape



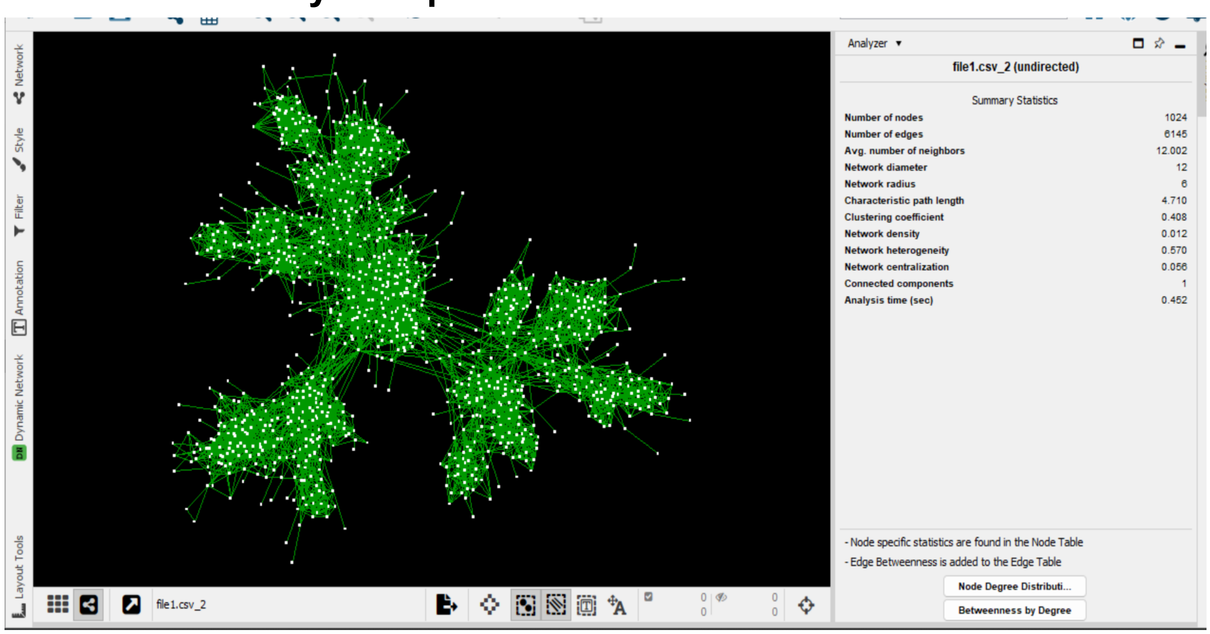
Result of our proposed model

A. Models allowing connection among friends of friends only:

```
avrage degree : 15.001953125
Number of nodes: 1024
Number of edges: 7681
Average clustering coefficient: 0.38542651998854827
Number of triangles: 13833.0
```

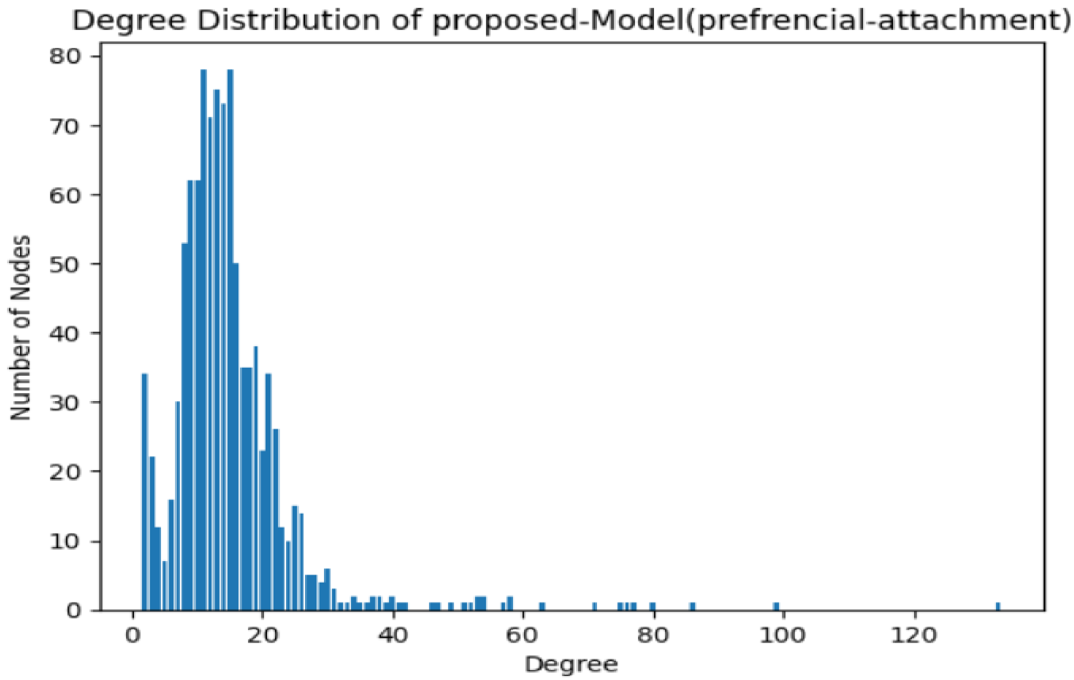


Visualization in Cytoscape

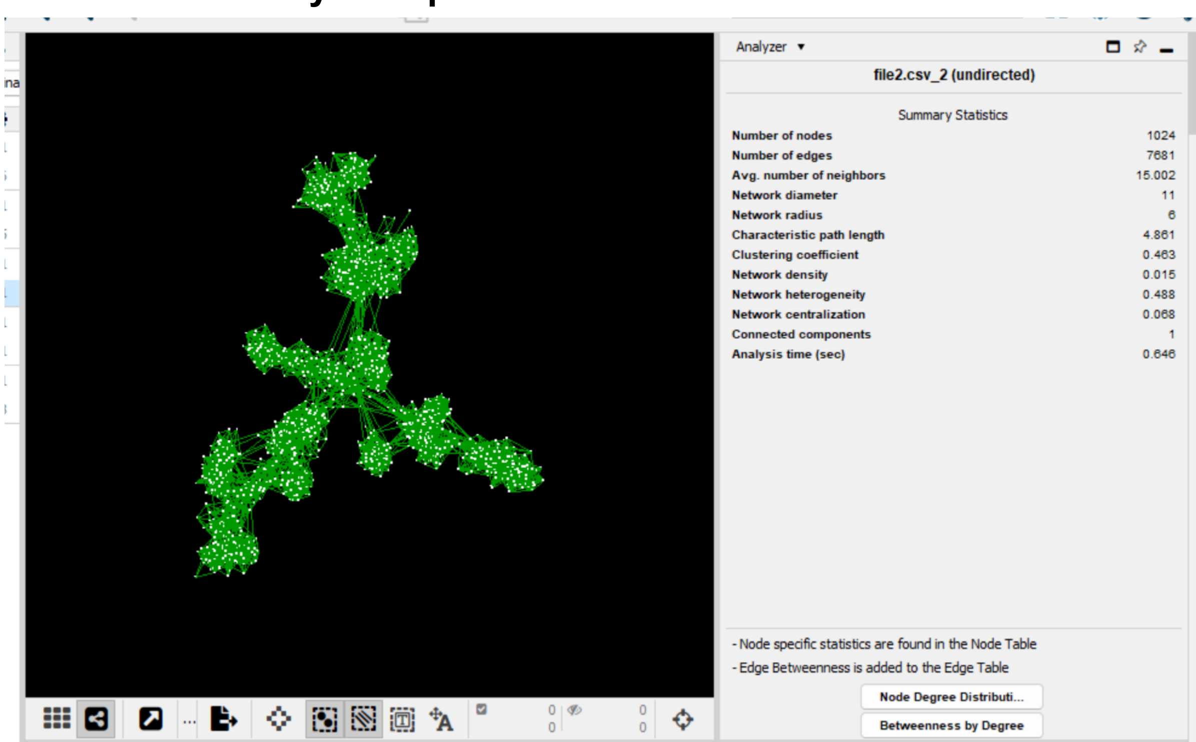


B. Preferential Attachment among friends of friends:

```
avrage degree: 15.001953125
Number of nodes:  1024
Number of edges:  7681
Average clustering coefficient: 0.45387837709706524
Number of triangles: 13833.0
```

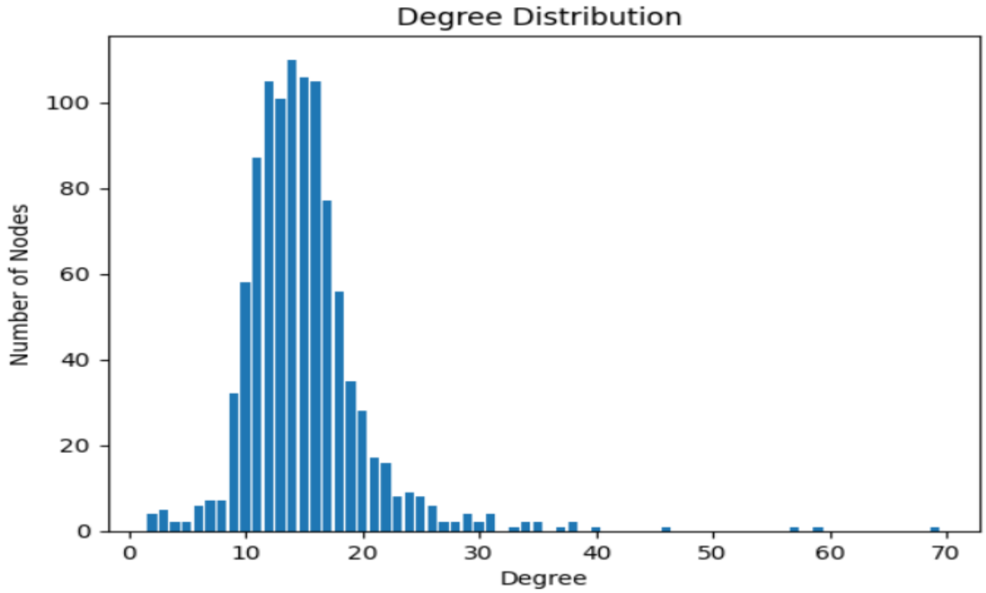


Visualization in Cytoscape

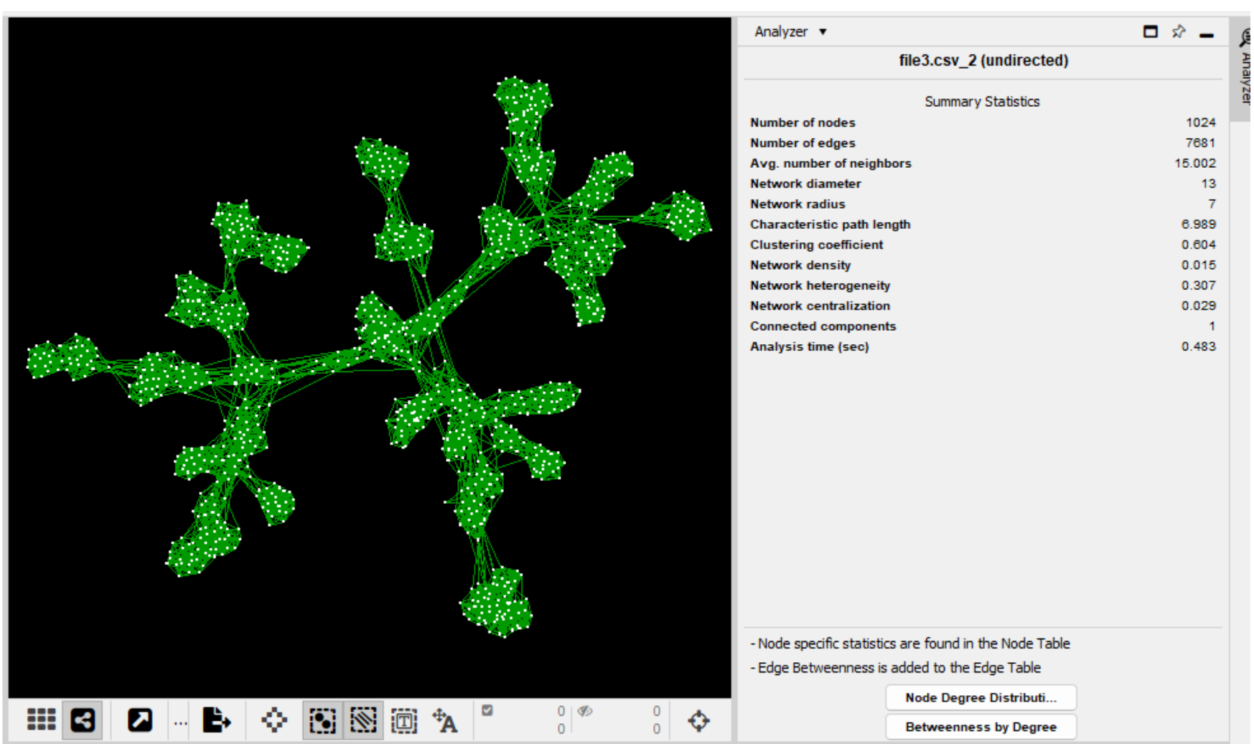


C. Common Neighborhood Based Selection among friends of friends:

```
avrage degree : 15.001953125
number of nodes: 1024
number of edges : 7681
Average clustering coefficient: 0.6043811016099492
Number of triangles: 21916.0
```



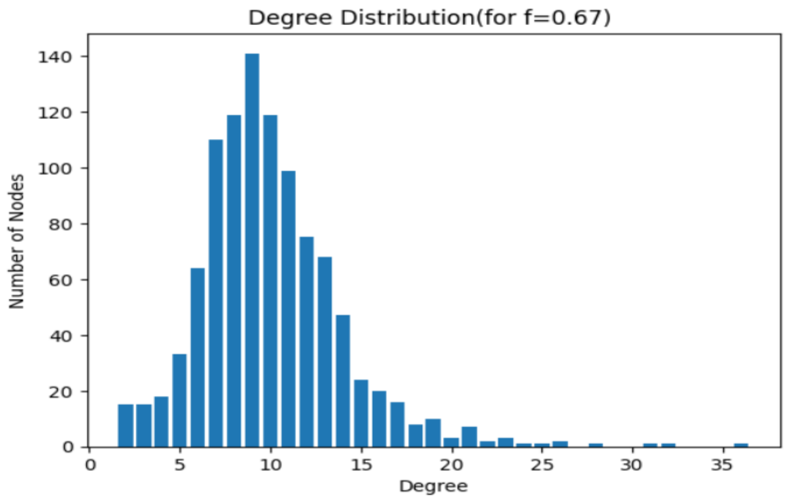
Visualization in Cytoscape



D. Models allowing connection among friends of friends as well as others.

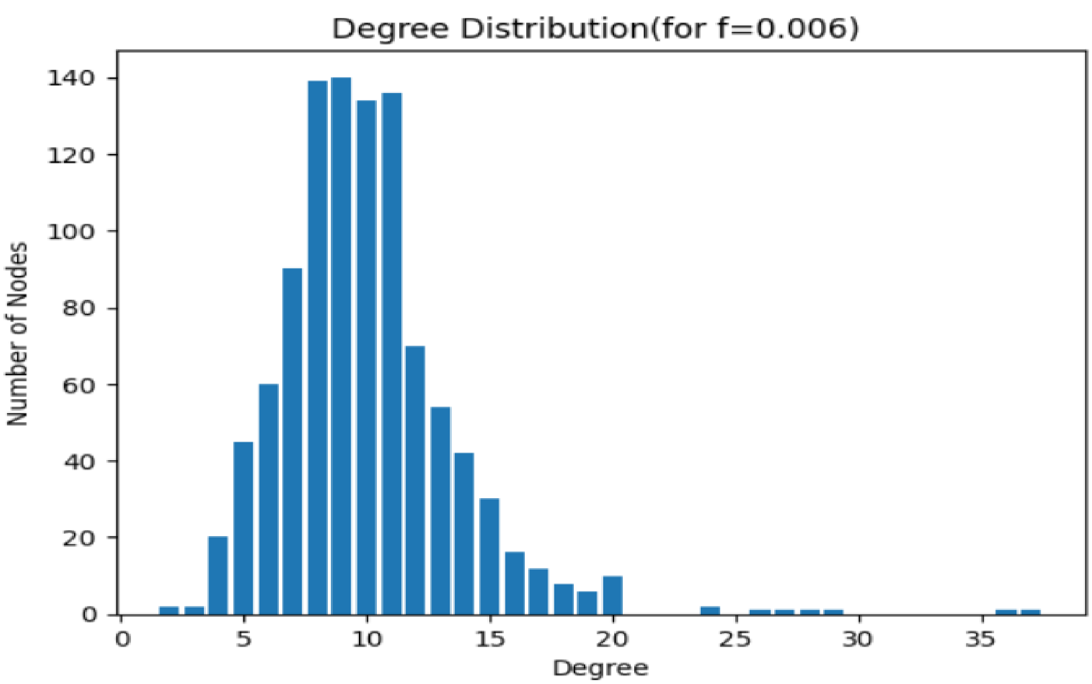
Degree distribution for large value of fraction ‘f’. It’s follow friends of friends model for large value of f and have high clustering coefficient.

```
avrage degree 10.001953125
number of nodes: 1024
number of edges : 5121
Average clustering coefficient: 0.6428543297456807
Number of triangles: 9723.0
```



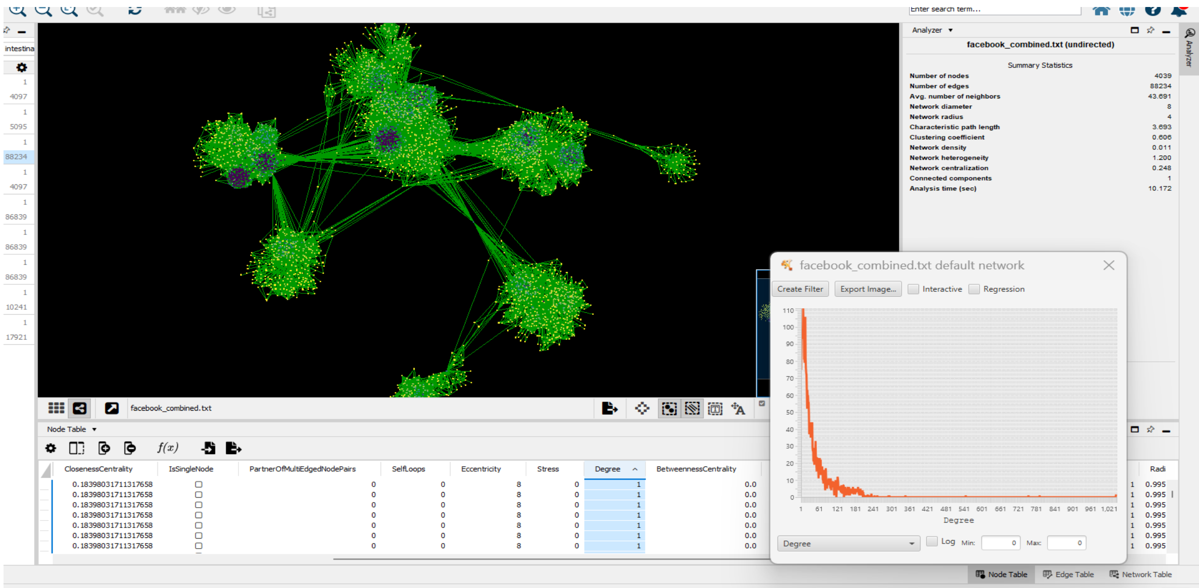
Degree distribution for small value of fraction ‘f’.have low clustering coefficient

```
avrage degree 10.001953125
number of nodes: 1024
number of edges : 5121
Average clustering coefficient: 0.266871014453624
Number of triangles: 3885.0
```

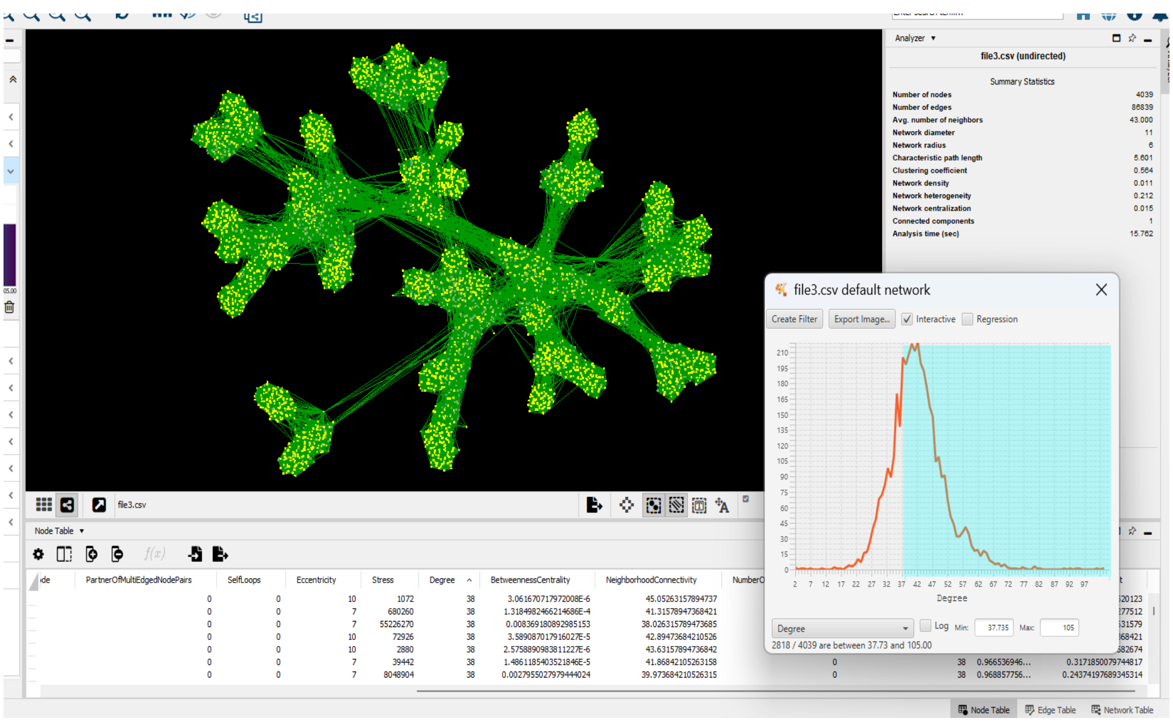


Now compare our third approach to a facebook dataset which i have taken from [SNAP: Network datasets: Social circles (stanford.edu)](https://snap.stanford.edu) .

visualization of facebook dataset in the Cytoscape .



Visualization our proposed model for same number of nodes and average degree



Since our proposed mode makes connections randomly in the first iteration, initially the degree distribution curve has a different shape but after some time it follows power law distribution.

Conclusion:

This work proposes a new method of network generation that produces networks with three desirable characteristics: **small-world property**, **power law degree distribution**, and **high clustering coefficien**t.

The method accounts for two real-life observations: people make friends at intervals and tend to make new friends among their friends' friends. By allowing nodes to make new connections in iterations and introducing a probability of making connections among friends of friends, the proposed method generates networks with all three desired qualities simultaneously. The parameter f can be varied to generate networks with different levels of clustering. Another benefit of this model is that it generates communities without any specific design.