DYNAMICS OF JOINING SOCIAL NETWORKS

By

PRADEEP KODURU

Bachelor of Technology in Computer Science
Jawaharlal Nehru Techonological University
Hyderabad, Andhra Pradesh, India
2005

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2008

DYNAMICS OF JOINING SOCIAL NETWORKS

Thesis Approved:

Dr. Subhash Kak

Thesis Advisor

Dr. John Chandler

Dr. Nohpill Park

Dr. A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to dedicate my work to my younger brother Satish Koduru who tragically died in an accident.

I would like to express my deepest gratitude to my graduate advisor Dr. Subhash Kak for his guidance and help in completing my work. His sound advice and his help in writing academic papers was invaluable.In Addition, I would like to thank my committee for their advise.

I would like to thank my friends Ravikanth Sama, Prasad Banala, Sumanth Gangasani for supporting me during difficult times. I especially want to thank my loving family their loving support, affection and encouragement during my studies. I would also like to thank my brother Srinivas Koduru for his love and support in both good and bad times. Finally, many thanks to all of my friends for their companionship and support which i will always cherish and enjoy.

TABLE OF CONTENTS

Chapter Page

1 INTRODUCTION **1**

1.1 Background and motivation 1

1.2 Research Goals and Approach 2

2 SOCIAL NETWORK MODELS **6**

2.1 Random Networks 6

2.2 Scale Free Networks 8

2.3 Small World Networks 10

3 JOINT NETWORK AND GROWTH **15**

4 RESULTS **25**

4.1 Two Scale Free Networks Joined 25

4.2 Small World and Random Network 28

4.3 Small World Network and Scale Free Network 30

5 CONCLUSION **33**

BIBLIOGRAPHY **34**

LIST OF TABLES

Table Page

2.1 Differences between network models 13

LIST OF FIGURES

Figure Page

1.1 Clustering coefficient C_B 3

1.2 Joining 2 social networks 4

2.1 Random undirected graph generated and visualized using pajek and with 1000 nodes and linking probability '0.001' 7

2.2 Degree distribution graph of random network shown in Figure 2.1 . . 8

2.3 (a)simple random network. (b)scale free network showing highlighted nodes as hubs [1] 9

2.4 Scale free network with 1000 nodes generated and visualized using Network Workbench. 11

2.5 Degree distribution graph of Scale free network shown figure 2.4 . . . 12

2.6 Diagram of Watts-Strogatz model with various p values. 13

2.7 Diagram on path length and clustering coefficient with different p values [2]. 14

2.8 Small world network with 100 nodes generated and visualized using Network Workbench 14

3.1 Pajek file format 15

3.2 Joint Network in pajek format 17

3.3 Linear function for probability of association (x-axis is degree and y-axis is probability of association) 21

3.4 Step function for probability of association 22

4.1 Two sample scale free networks joined using preferential attachment growth algorithm 26
4.2 Two sample scale free networks joined using preferential attachment growth algorithm 26
4.3 Two sample scale free networks joined using Random attachment . . 27
4.4 Two sample scale free networks joined using proposed approach with linear probability of association. 27
4.5 Two sample scale free networks joined using proposed approach with step function as probability of association 28
4.6 Sample small world network and random network joined and grown using preferential attachment 28
4.7 Sample small world network and random network joined random attachment 29
4.8 Sample small world network and random network joined proposed approach with linear probability of association 29
4.9 Sample small world network and random network joined using proposed approach with step function as probability of association 30
4.10 Sample small world network and scale free network joined preferential attachment 30
4.11 Sample small world network and scale free network joined random attachment 31
4.12 Sample small world network and scale free network joined proposed approach with linear probability of association 31
4.13 Sample small world network and scale free network joined using proposed approach with step function as probability of association 32

CHAPTER 1

INTRODUCTION

As social beings, humans are bound to one another by various interdependencies that include common cause, values or kinship. They form social networks which are interconnections of nodes (people or organizations) that come together and communicate through various channels. Here we concentrate on online social networks formed by various means like email, newsletters and instant messaging. A social network graph maps the interconnections between individuals or 'actors'. The nodes in the graph represent the actors and the edges represent the communication, friendship, relation or 'ties' between actors. In recent years there has been immense growth of online social networks worldwide some of which are Facebook, MySpace and Orkut. A detailed study of social networks was limited to 100 individuals in 1970s [3]. With the advent of very large online social networks we have the opportunity to learn of classes of social networks and their dynamics with time.

1.1 Background and motivation

Along with popular online social networks such as Facebook, Friendster, MySpace there are others that are not well known which serve graphs with more limited intents. We now have websites which allow users to set up their own private social networking site for family or friends. One can expect that networks that are not so popular will join hands with each other or with large networks to expand their user database as well as increase their reach to many individuals. An example of this is the merger of Wink.com and Reunion.com [4]. By merger, their users could have a comprehensive search of user profiles which would span both wink.com and reunion.com databases. Other example of this kind is the integration of xing.com and zoominfo.com [5] to increase their database of number of user profiles accessible by integrated search engine. This move helped them to increase their network size to counter their common rival linkedin.com.

As social networks join with one another there is a need to study how structure of the joined network evolves over time and how the network characteristics would change with the growth of joint network and what does it mean to the companies which are willing to have their databases merge to have larger user base.

1.2 Research Goals and Approach

The primary goal of the thesis is to emulate real world network evolution of joining two social networks.

We would like to know how the metrics of joint social network change when two online social networks are made visible to each other and are grown using a network growth algorithm. We assume that the joining takes place between the nodes on either side of the two networks using methods such as preferential attachment as seen in generation of scale free network, random attachment and other kinds of attachment methods. We would also examine various metrics of social network and obtain a glimpse of how the network evolves over time. The two networks used in joining are the networks created using algorithms that generate specific models that will be introduced in next chapter. The metrics we would be considering for evaluation include:

- **Degree Distribution**

  Degree of a node is number of interconnections with other nodes in the network. We would look at probability distribution of degrees of nodes in entire network.

- **Diameter**

  It is maximum path length between the nodes in the network. We will consider how the network diameter varies for different models of social networks.

- **Average Clustering Coefficient**

  Clustering coefficient was introduced by Duncan J. Watts and Steven Strogatz [6] and it is a measure by which we determine how close the node and its adjoining nodes are being a complete graph. Let node A be connected node B and node B be connected to node C then clustering coefficient of node B is the probability that nodes A and C are connected.

  The clustering coefficient of a node *'i'* is given by

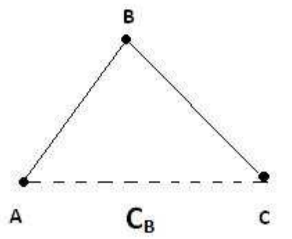

Figure 1.1: Clustering coefficient C_B

$$C_i = \frac{e_i}{k_i * (k_i - 1)} \tag{1.1}$$

In the above equation e_i is the number of edges between node i's neighbors or number of triangles through the node 'i' and k_i is number of neighbors of node *'i'*.

The average clustering coefficient of the entire graph given by the equation

$$\overline{C} = \frac{1}{n} \sum_{i=1}^{n} C_i \tag{1.2}$$

where n is the number of nodes in the entire social network graph.

We approach the problem by first generating social network models using algorithms which have been proposed by Erdos and Renyi for generation of random network model [7], the small world model by Watts-Strogatz algorithm [2] and the scale free network model generated using Barabasi's method [6, 8].

After the network is generated we make a joint intermediate network by renaming the nodes in one of the networks so there won't be any conflict between the nodes of one network with the other.

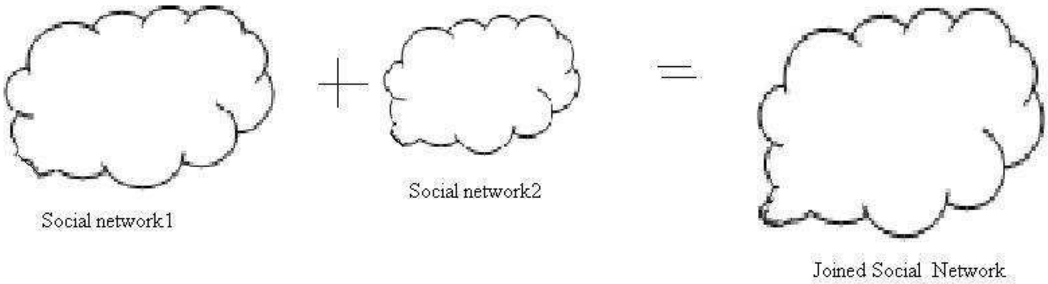

Figure 1.2: Joining 2 social networks

Later, the combined network obtained from renaming nodes is used as input to the social network growth algorithms that emulate the evolution of joint network over time. Algorithms that we would be using for evolution are:

- **Preferential Attachment**

  In this network growth algorithm at every time step a new node or actor gets added to the combined social network and ties are established between already existing actors and the new one. The actors with high connectivity have more chance of attracting the newly added node. This means higher the degree of the pre-existing node, the more are the chances of having edge to the newly added node.

- **Random Attachment**

  It is a network growth in which at every time step a new node gets added and edges of newly added node gets attached to the pre existing nodes in a random manner.

- **New Attachment Method**

  Barabasi's method of network evolution only includes evolution of network by adding new nodes and having edges between the added node and pre-existing nodes at every time step. In this proposed method at every time step we will not only have preferential attachment on the newly added node but also have internal evolution of the network itself. We would examine how the network would structure itself by internal evolution as well as addition of new nodes.

CHAPTER 2

SOCIAL NETWORK MODELS

Based on topological characteristics online social networks could be modeled into three major types. Random networks that we will be discussing are usually very rare in occurrence in online social networks as they do not effectively mirror the social structure but they are the most extensively studied networks in the field of complex networks. We will also examine scale free and small world networks.

2.1 Random Networks

Random networks were introduced by Erdos and Renyi in 1959 [7]. A random network graph could be easily constructed taking nodes or actors and putting edges or ties between each pair of nodes selected with a probability given by p. We construct random graph by using the Erdos-Renyi model in which the actors are connected with equal probability 'p' with $0 < p < 1$ [7]. We will have a graph or network which is given by G (n, p)where 'n' is the number of actors and 'p' being the probability that having an edge between any two actors.

Consider an actor in the graph and he/she is connected to all other N-1 actors with same probability as p and N being total number of actors in graph. The probability P(k) that an actor with a degree equivalent to k is given by a binomial distribution.

$$P(k) = C_k^{N-1} p^k (1-p)^{N-1-k} \tag{2.1}$$

The average degree of an actor in the network is x = (N-1)p. So we could also write the above equation as

$$P(k) = C_k^{N-1}[\frac{x}{N-1-x}]^k[1-\frac{x}{N-1}]^N - 1 \approx \frac{x^k}{k!}e^-x \quad (2.2)$$

From the above equation we could say that degree of connectivity of actors follows the Poisson distribution.

Below is a sample undirected Erdos-Renyi random graph with 1000 nodes and p = 0.001 generated using pajek [9].

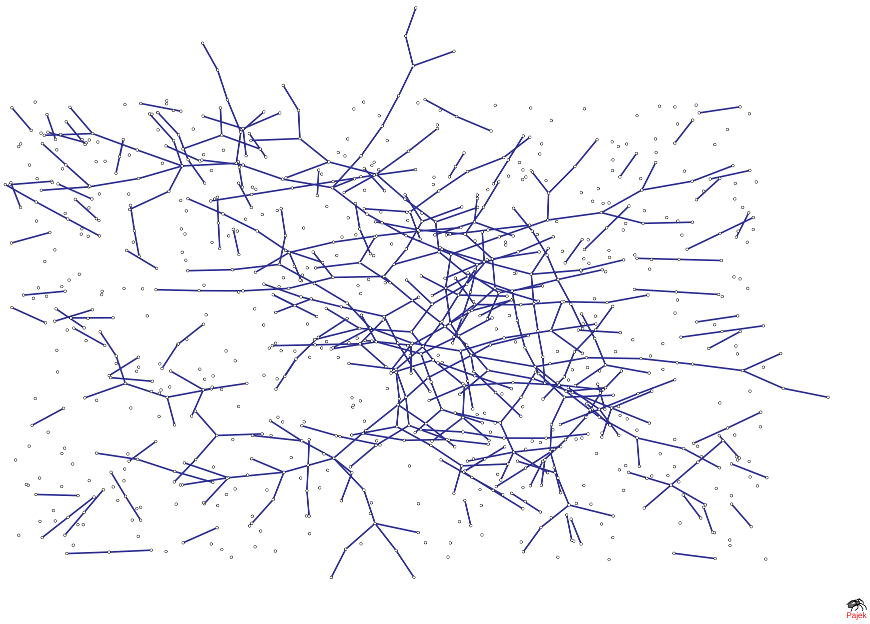

Figure 2.1: Random undirected graph generated and visualized using pajek and with 1000 nodes and linking probability '0.001'

The degree distribution of the above visualized social network is given in Figure 2.2. Here x-axis is the degree of a node - k and the y-axis is probability of node having degree k indicated by P(k). We could see that average degree is around 1 which could calculated using equation (N-1)*p. In this case its $999 * 000.1 = .999 \approx 1$.

Random networks are extensively studied and are usually used as reference in robustness tests and how a rumor or virus spreads around the network. Very few real world networks could be modeled around random graph but we do want to include

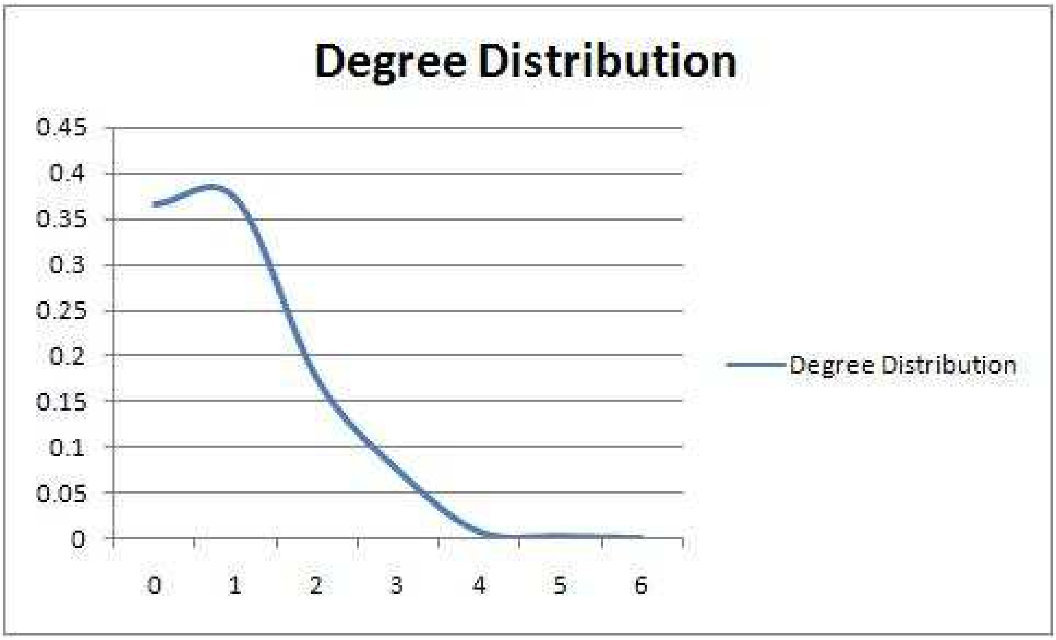

Figure 2.2: Degree distribution graph of random network shown in Figure 2.1

them in our study of joining of different network models.

2.2 Scale Free Networks

We have seen that in random network the degree distribution follows Poisson distribution, in scale free network the degree distribution obeys the power law distribution [10] and the probability of a node having a degree K is given by the equation:

$$P(K) = cK^{-\lambda} \tag{2.3}$$

where λ is a constant and has a range $2 < \lambda < 3$. The term 'scale free network' was coined by Barabasi and his team and it has the characteristics of continuous growth and preferential attachment [6, 8]. The clustering coefficient of nodes in the network also obeys the power law distribution. As degree distribution follows power law distribution the topology of network formed is different from other network models. The nodes having large degree and are called hubs in the network. In Figure 2.3 the filled circles are the nodes with high connectivity which are hubs in scale free network. Many real world networks could be modeled around this type as most of them have slow growth and their growth follows preferential attachment as seen in scale free networks.

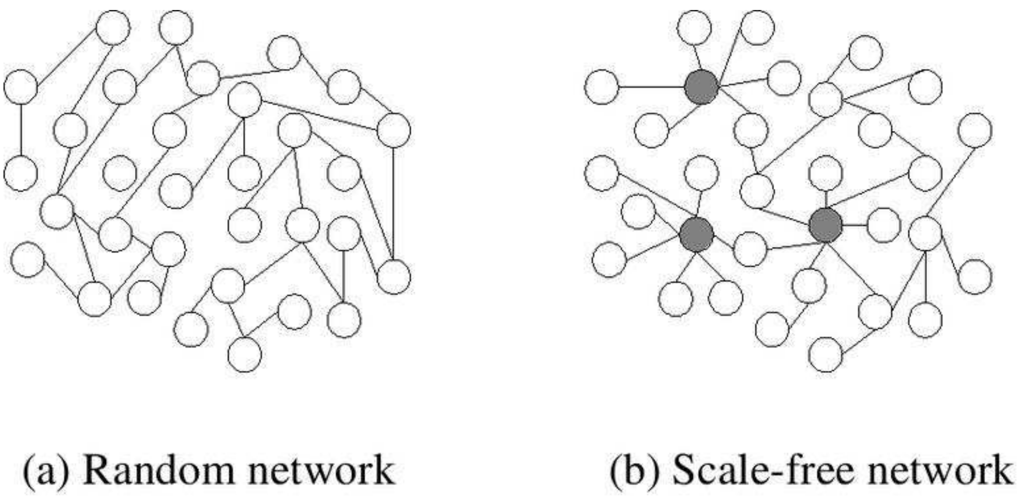

Figure 2.3: (a)simple random network. (b)scale free network showing highlighted nodes as hubs [1]

Growth using preferential attachment are common to many complex networks as many follow the phenomenon of "rich become richer". Examples include world wide web [11], frequency of words found in a text [10], links between java classes [12].

Earlier models of networks did not have addition of new actors or nodes over time and were static. Scale free network has the characteristic of adding new nodes over time which mirrors continuous growth. Preferential Attachment is done by addition of a new node at every time step whose edges are attached to already existing well connected nodes. Hence older nodes with high degree have more probability of establishing an edge with the new incoming node that gets added at each time step for the creation of a scale free network. The probability for a node attaching itself to a new coming node is given by

$$P(K_i) = \frac{K_i}{\sum_j K_j} \tag{2.4}$$

where degree of node i.

Barabasi-Albert algorithm of generating a scale free network starts out with small number of nodes (m_0) and at each time step we add a new node with small number ($m \leq m_0$) of edges that would establish a link between the new node and already existing nodes preferentially. After t time steps the resultant graph would have $t+m_0$ vertices and mt more edges. The network thus obtained follows the power law of degree distribution. We would be using Barabasi-Albert algorithm for generating a scale free network to simulate a scale free online social network as many real world online Virtual communities are scale free networks [6, 8]. The algorithm requires that we at least have more than two nodes initially with least one connection between them. After running the algorithm the resultant structure does not depend on the initial structure and the degree distribution follows the power law. we should not use this algorithm to generate graphs with small number of nodes as they are not necessarily become scale free. The algorithm used in generation of a scale free network has preferential attachment step repeated at every time step and the nodes are attached to pre-existing nodes with a probability given by P(K).

Figure 2.4 shows visualized sample scale free network generated using Network Workbench and figure 2.5 shows the degree distribution of sample network visualized in figure 2.4.

2.3 Small World Networks

Stanley Milgram in his classic paper of 1967 paper described his experiment of passing document from person to his acquaintance and then to immediate acquaintance could linkup two strangers in different parts of the country [13]. It was found that to link up any two strangers it usually took an average of six degrees of separation, which is defined to be the least number of ties between the two nodes or actors.

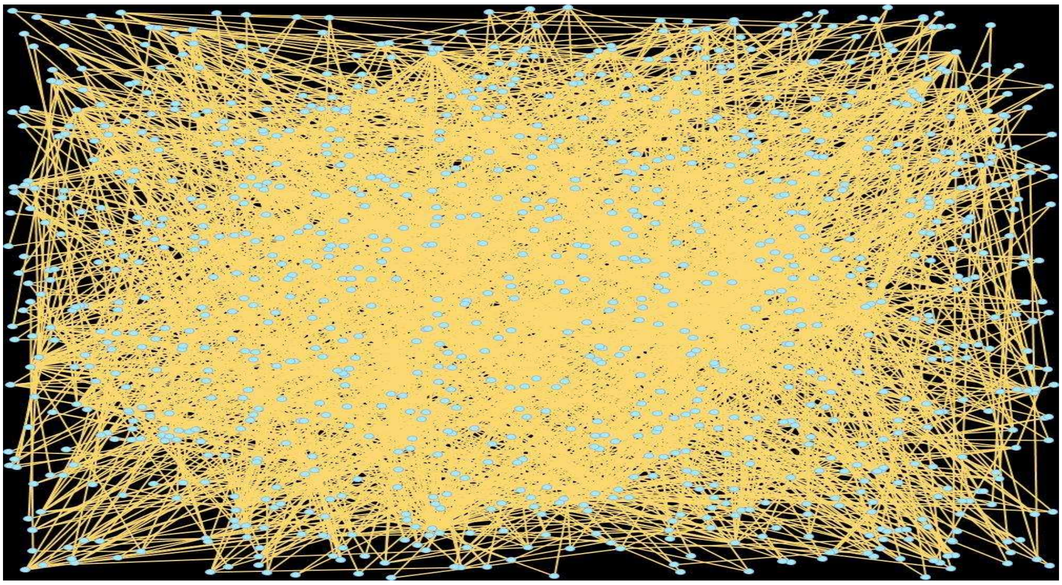

Figure 2.4: Scale free network with 1000 nodes generated and visualized using Network Workbench.

The Primary characteristics of small world networks are a high clustering coefficient and small average distance between the nodes of the graph and having a Poisson degree distribution. As an aside, the concept of small world was the basis of the play 'Six Degree of Separation' [14] which attests that this concept has entered popular consciousness. The small world property usually means network exhibiting short linking path between individuals. This property has been found to be in many real world examples. For example, protein -protein interactions have small world social networks [15]. Considering web page as node and hyper-link between one page to other as edge, Barabasi and his team also found that to jump from one web site to any other web site usually required average of 19 jumps [16].

The algorithm for generation of small world is given by Watt and Strogatz [2, 17].

Small World Algorithm

Input: Regular graph with N nodes. Average degree be K and rewiring probability be p with $(0 < p < 1)$

Output:small world network with N nodes.

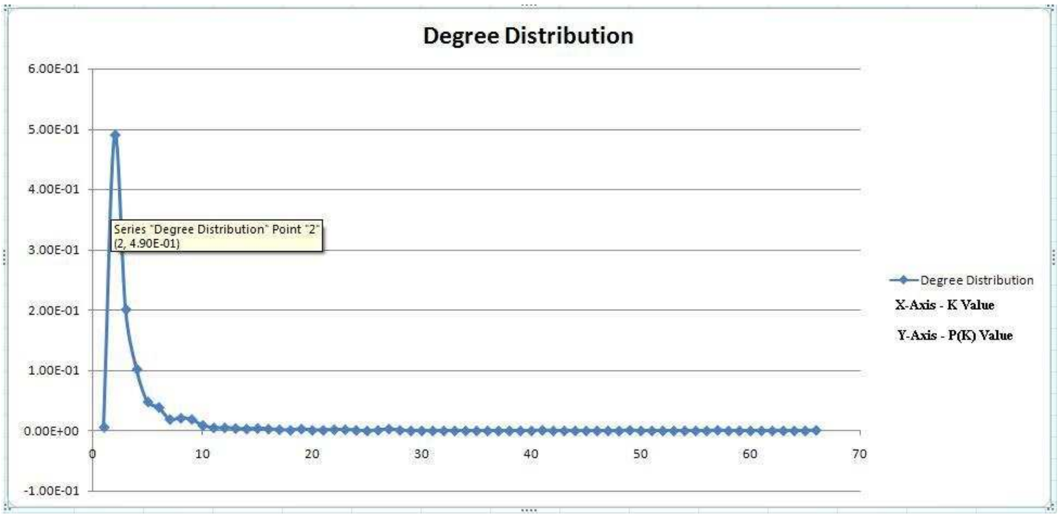

Figure 2.5: Degree distribution graph of Scale free network shown figure 2.4

Steps:

1. Start with a node in the graph and break its connections to other nodes with a probability p. If the connection is broken, establish an edge to any other node which is randomly chosen and which does not have a connection to it.
2. Return to step 1, until we have visited all nodes in network.

The algorithm discussed, repeatedly and randomly reassigns edges of input regular network and does not increase the number of nodes or edges in the graph. It opens up the regular graph in a way that there will be edges between different neighborhoods of the networks. The output is a network which has small average distance and large clustering coefficient. Small world network could be used to describe many real world networks. Value of 'p' determines the randomness in the network. So p=0 represents fully ordered graph and p=1 gives a random graph. Figure shows the transition of regular graph.

Figure 2.7 shows how the path length(L) and clustering coefficient(C) varies with various values of p. It shows that having small increment in p from 0 will decrease the network path length drastically while clustering coefficient is not varied and hence

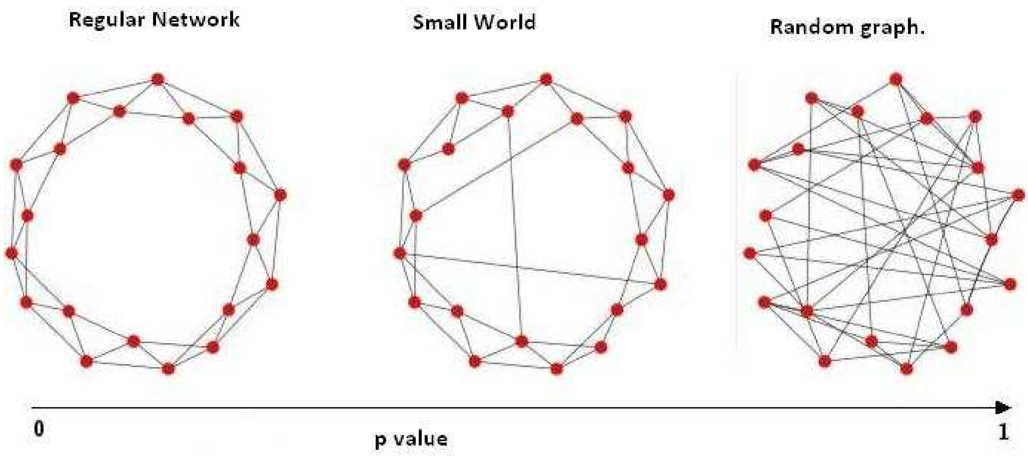

Figure 2.6: Diagram of Watts-Strogatz model with various p values.

denotes the characteristic of a small world.

Below is sample small world generated using Network workbench:

The differences between different network models summarized by table 2.1.

Table 2.1: Differences between network models

Property	ER Graph	Small world model	Scale free network
Degree distribution	Poisson	Poisson	Power law
Clustering Coefficient(C)	small	large	large
Average path length(L)	small	small	small

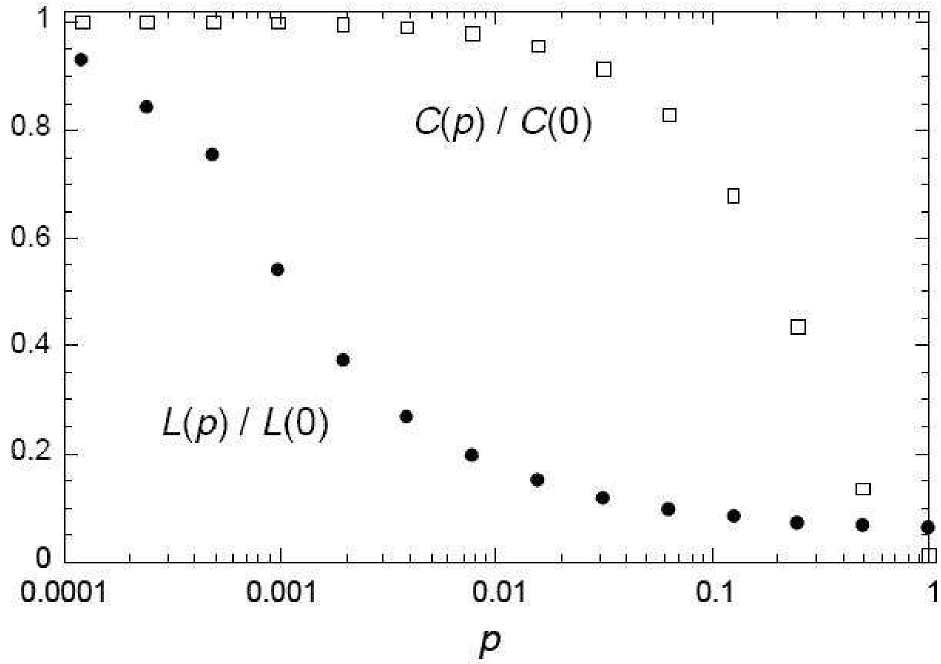

Figure 2.7: Diagram on path length and clustering coefficient with different p values [2].

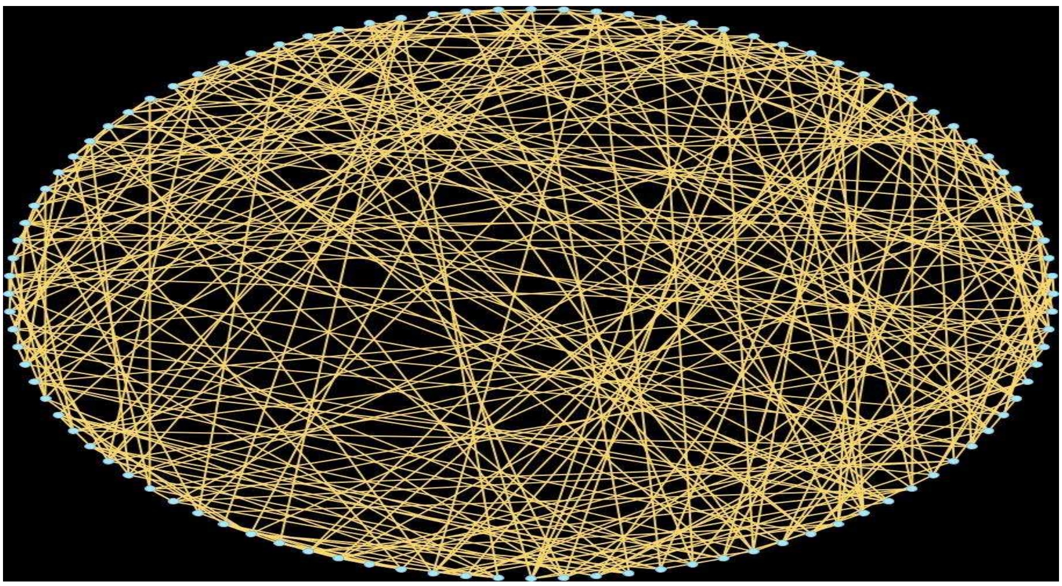

Figure 2.8: Small world network with 100 nodes generated and visualized using Network Workbench

CHAPTER 3

JOINT NETWORK AND GROWTH

We will be using computer simulation to generate joint network and emulate growth using one of the growth algorithms. Though computer simulations are a practical way of testing theories, it will not predict the future of a real online community. Nevertheless, it is a way of testing our assumptions by generative computer models.

We will be using pajek data file format to represent the network. It is plain text file which has an extension '.net' and the nodes and edges of a complex network are represented in the following format.

```
*Vertices n
1 "label1"
2 "label2"
...
*Edges
1 2
3 4
...
```

Figure 3.1: Pajek file format

In the above format 'n' represents the number of nodes in the network. This should be same as the number of lines that appear between '*Vertices' and '*Edges' lines. Each line after '*Vertices' line has vertex number and vertex name. The lines after '*Edges' represent the links between the vertices given their numbers.

Our first step in joining any two network models is to generate sample networks by using algorithms mentioned in chapter two. We generate sample data sets of different network models in pajek data file format (.net file) of size 1000 nodes according to respective parameters. Second step is to rename one of data sets in generated network models that come together for joining. Let's say

Small world sample: Node numbers 1-1000.

Scale free network: Node numbers 1-1000.

Joint network: 1-2000 will be the node numbers of the joint network with 1-1000 from scale free network and 1001-2000 node numbers are the nodes from small world which are renamed from 1-1000.

Figure 3.2 shows joint network that will be constructed in pajek's '.net' file format. Now this intermediate joint network is used on which growth algorithms are run. The idea is that newly added actors can examine public profiles and interests of pre existing actors of joint social network. The newly added actors thus see the joint network as single network and not as two separate networks. This will be possible if the search engine's search spans not only old user database but also the database of the new network which is bought to form the joint social network.

The growth functions that we will be considering to run our simulation on the joint network try to emulate real world social network. The simulation of growth algorithm on resultant joint network tries to achieve a balance between randomness and order. Hence the resulting structure after simulation will have both randomness and order. Interpersonal networks were locally dense and globally sparse. Watts in his paper [18] discussed how networks that followed small world phenomenon inspired from Milgram [13] were neither completely random nor completely ordered. In the

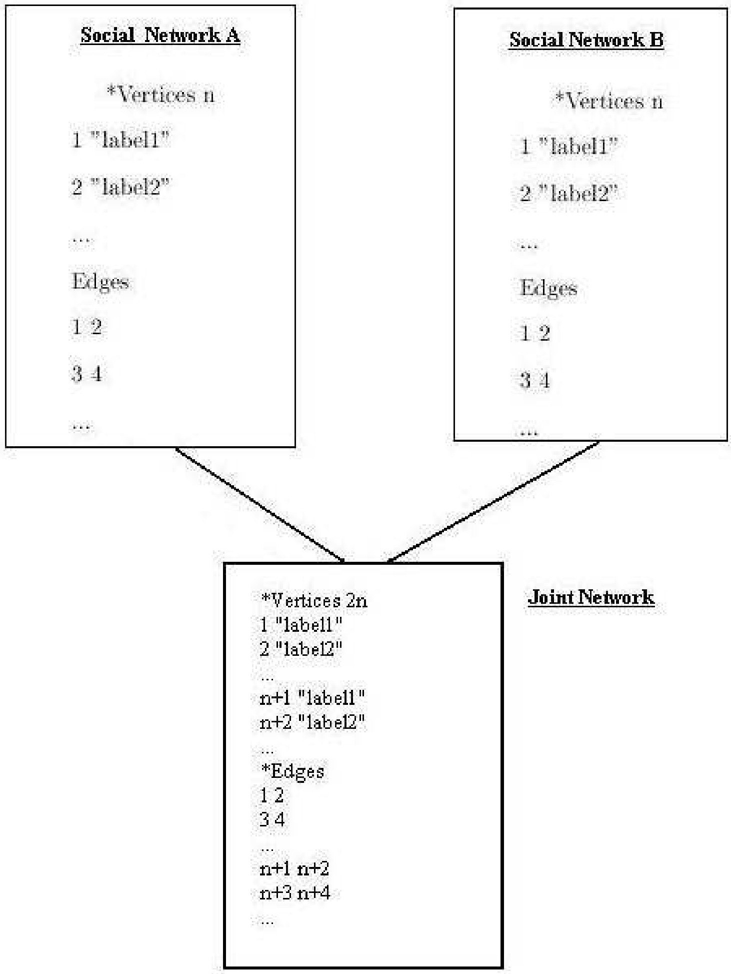

Figure 3.2: Joint Network in pajek format

same spirit we will have the methods or algorithms which we will be discussing will induce both randomness and structure.

1. **Preferential Attachment:**

   This algorithm of growth could be justified by conjecture that people with large number of friends or acquaintances are more likely to attract new friends. It means that new nodes have a tendency to attach to already existing nodes that are well connected. Preferential attachment is used in generation of intermediate joint network. The joint network obtained from the above process is subjected to this growth algorithm.

Preferential Attachment Algorithm

Input: Social network1, Social Network2, N- number of nodes expected by end of simulation

Output: Grown joint network.

(a) Create a Joint network by renumbering the nodes of any one of the social networks to create a intermediate joint network. For the sake of simplicity we will assume that there are no common nodes which represent the same individual.

(b) A new node is added with small number of edges (≤ 2). The probability of attaching to an already existing node is given by

$$P(K_i) = \frac{K_i}{\sum_j K_j} \tag{3.1}$$

If above probability is greater than a random number from random number generator and is also greater than 0 then an edge is attached between the two nodes. We visit every node in the joint network.

(c) Return to step (a) until number of nodes in the network is equal to N given by input.

2. **Random Attachment:**

Random attachments usually happen in group of people with no previous contacts at all. It is a growth algorithm in which at each time step a new node with less than or equal to 2 edges (we are modeling slow growth as observed in real world) gets added every time step.

Random Attachment Algorithm

Input: Social network1, Social Network2, N- number of nodes expected by end of simulation

Output: Grown joint network.

(a) Create a Joint network by renumbering the nodes of any one of the social networks to create a intermediate joint network.

(b) A new node is added with small number of edges (≤ 2). We randomly choose nodes (≤ 2) in the joint network and have edges between the newly added node and the chosen nodes.

(c) Return to step (a) until number of nodes in the network is equal to N given by input.

3. **New Method of Attachment:**

This method of attachment is a mixture of preferential attachment and internal evolution. In this method at each time step we will add a new node which will attach preferentially to the already existing nodes in the network and there is also evolution in terms of growth of edges which represent development of interpersonal relationships between the individuals inside the network at that particular time step.

Anthropologist Robin Dunbar [19] has made an important claim regarding the extent of connectivity in effective social organizations. Arguing that the size of the brain is correlated with the complexity of function, he has developed an equation, which works for most primates, that relates the neocortex ratio of a particular species - the size of the neocortex relative to the size of the brain to the largest size of the social group. For humans, the max group size is 147.8, or about 150. This represents the maximum number of people who can be part of a close social relationship.

Dunbar considered 21 different hunter-gatherer societies around the world and found that the average number of people in each village was 148.4 . Another significant validating point is the community of Hutterites, followers of the sixteenth century Jakob Hutter of Austria, who are pacifists and believe in community property and live in shared communities called colony. Several thousand Hutterites relocated to North America in the late 19th century and their colonies are mostly rural. A colony consists of about 10 to 20 families, with a population of around 60 to 150. When the colony's population approaches the upper figure, a daughter colony is established.

Dunbar number is used to restrict the maximum degree of the nodes in our method of attachment. we will restrict the maximum number of interpersonal relationships a actor can have in the social network to 150. At every time step internal growth takes place. We will be adding small number of edges (≤ 2) as we are emulating small growth, between pair of randomly chosen vertices. We will be using the probability of association function to make a decision of whether to have an edge between the randomly chosen pair of vertices. The functions used are shown below.

In Figures 3.3 and 3.4, the x-axis denotes the degree of the node in the social network and y-axis denotes the probability of association or the probability of an actor willing to associate with other actor inside the joint network at that particular time step.

If we are using the linear function shown in figure 3.3 then for randomly picked node of degree k, the probability of association is given by equation.

$$P_A(k) = 1 - \frac{k}{150} \tag{3.2}$$

We will find $P_A(k_1)$ for the first randomly chosen node and $P_A(k_2)$ for the second randomly chosen node and we will use above function to find P_{total} which is given by equation

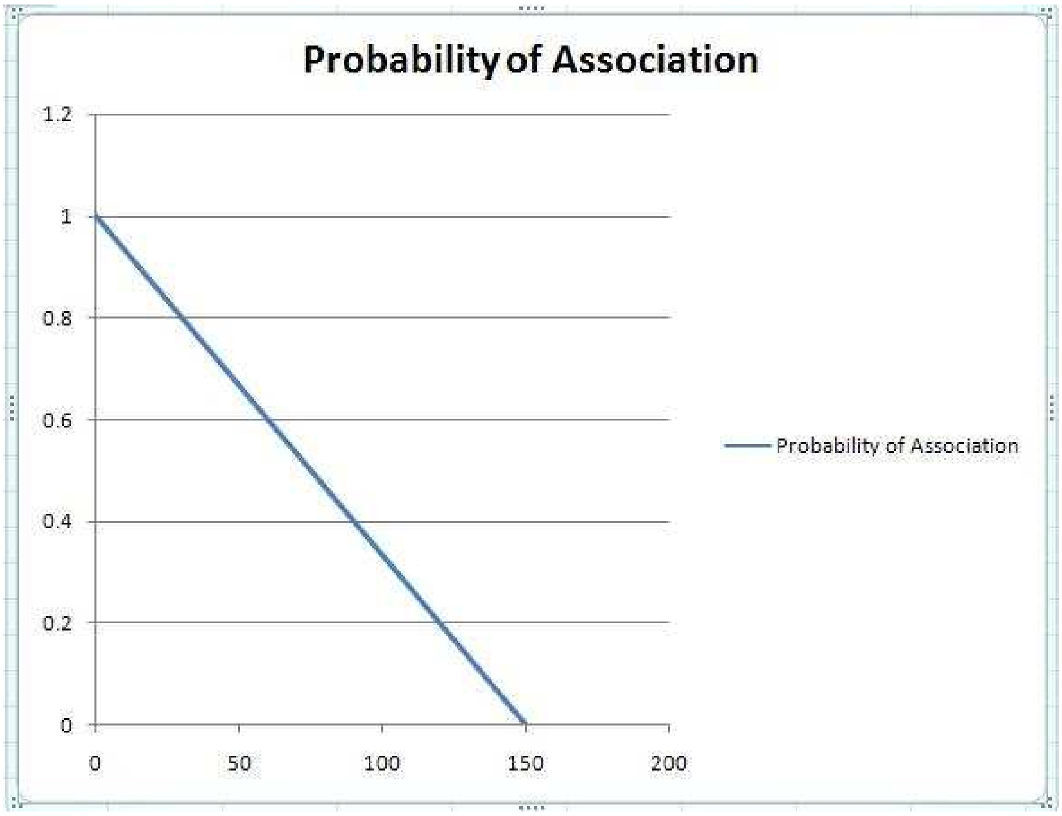

Figure 3.3: Linear function for probability of association (x-axis is degree and y-axis is probability of association)

$$P_{total} = P_A(k_1) + P_A(k_2) \tag{3.3}$$

If $P_{total} \geq 1$ then we will be having an edge between the chosen pair of nodes.

If we are using the step function to decide on having an edge between randomly chosen pair of nodes then probability of association of a chosen node with a degree k is given by the equation

$$P_A(k) = \begin{cases} 1.0 \text{ if } k = 0 \text{ and } k \leq 50 \\ 0.5 \text{ if } k > 50 \text{ and } k \leq 100 \\ 0.25 \text{ if } k > 100 \text{ and } k < 150 \\ 0 \; otherwise \end{cases} \tag{3.4}$$

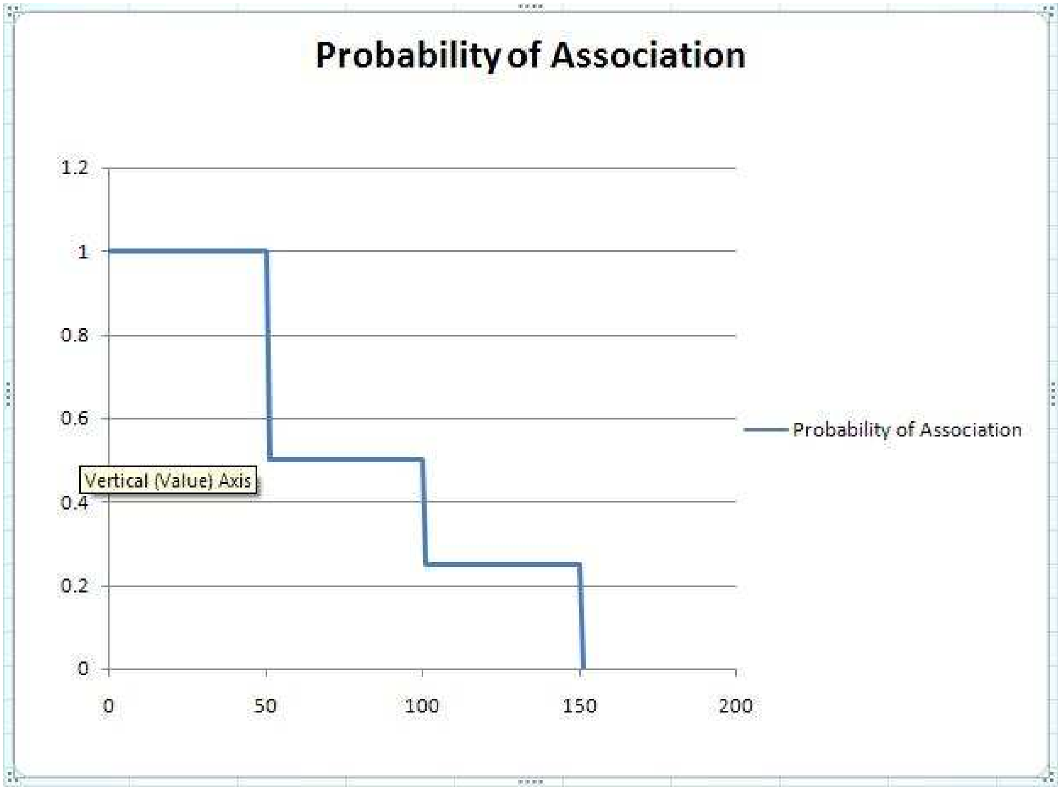

Figure 3.4: Step function for probability of association

We will find $P_A(k_1)$ for the first randomly chosen node and $P_A(k_2)$ for the second randomly chosen node, where k_1 and k_2 are the degrees of the nodes. As we did with linear association function previously we will find P_{total} using the equation 3.3. If $P_{total} >= 1$ then an edge or a relationship is established between the chosen pair of actors. The algorithm of growth using proposed method is given below.

New Attachment Algorithm

Input: Social network1, Social Network2, N- number of nodes expected by end of simulation, Probability Association function

Output: Grown joint network.

(a) Create a Joint network by renumbering the nodes. We will assume that there are no common nodes which represent the same individual in the joint network.

(b) A new node is added with small number of edges(≤ 2). The probability of attaching to an already existing node in joint network is given by

$$P(K_i) = \frac{K_i}{\sum_j K_j} \tag{3.5}$$

If above probability is greater than a random number from random number generator and is also greater than 0 then an edge is attached between the two nodes. We visit every node in the joint network.

(c) Randomly select any two pairs of nodes inside the joint network at this time step.

(d) If we are using linear probability association function for the first pair of nodes we find $P_A(k_1)$ and $P_A(k_2)$ using the equation 3.2 where k_1 and k_2 are the degrees of nodes. We find equation 3.3 and if $P_{total} \geq 1$ then an edge is established between the two nodes. Similar procedure is followed for the other pair.

(e) If we are using step function then for the first pair of nodes we find $P_A(k_1)$ and $P_A(k_2)$ using the equation 3.4 where k_1 and k_2 are the degrees of nodes in the pair. We find equation 3.3 and if $P_{total} \geq 1$ then an edge is established between the two nodes. Similar procedure is followed for the other pair.

(f) Return to step (a) until number of nodes in the network is equal to N given by input.

No attachment methods fully emulate real world and will not predict the future of the joined social networks. The proposed attachment method addresses internal growth of the network which has not been addressed in preferential attachment developed by Barabasi and his team. In the next chapter we present our results regarding joining various social network models using the algorithms mentioned above.

CHAPTER 4

RESULTS

We were able to generate various network models and join them by different growth algorithms that were described in earlier chapter. The growth algorithm used and the network models that come together for merger will determine structure of the resultant network. We run the simulation on different network models to obtain specific network sizes and we draw graphs of degree distribution which help in characterizing the grown network structure obtained by running a growth function on the joint network. The algorithms we have discussed for emulating continuous network growth do not accurately mirror real world growth of a social network as there might be situations in evolution where ties between nodes die. Growth algorithms described try to emulate real world that have both continuous and slow growth. In degree distribution graphs shown in Figures 4.1 to 4.13, the x-axis represents the degree of nodes (k) and the y-axis represents p(k), which is probability of having degree k. The p(k) used to plot the graph is mean of 4 simulations of growth algorithm on the joint network formed.

4.1 Two Scale Free Networks Joined

Two sample scale free networks of node sizes 1000 each are generated in pajek format using Barabasi's algorithm. As generated networks both have same node numbers (1-1000); once the network is renamed to form a combined network with node numbers 1-2000. A growth function is run on this combined network which is used to simulate the network evolution over time.

1. **Preferential Attachment:**

   Figure 4.1 shows degree distributions of the joint network after running preferential attachment algorithm on the joint network.

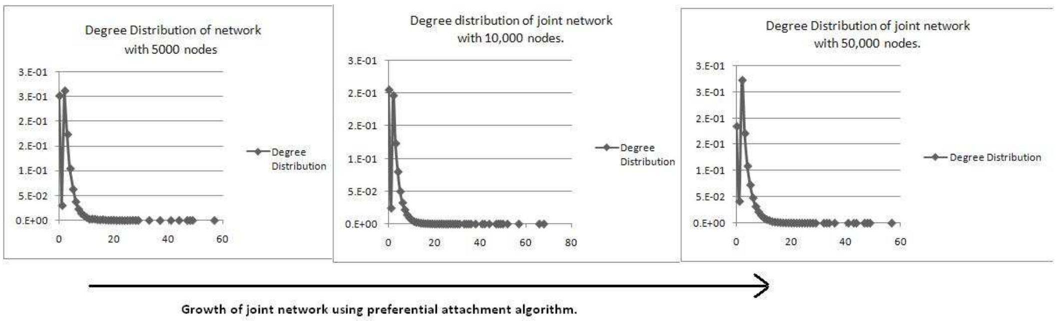

Figure 4.1: Two sample scale free networks joined using preferential attachment growth algorithm

   The logarithmic scale of Figure 4.1 is given in Figure 4.2.

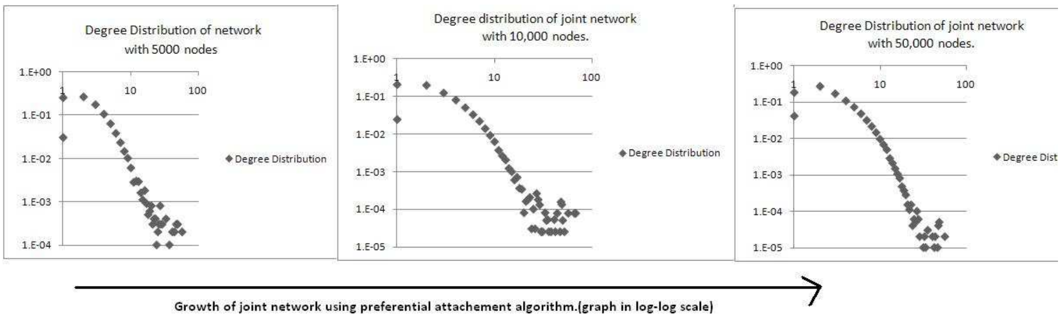

Figure 4.2: Two sample scale free networks joined using preferential attachment growth algorithm

   Figure 4.2 shows that the resulting network after running preferential attachment on joint network has power law distribution so the resulting network is also a scale free network.

2. **Random Attachment:**

   In this kind of growth the newly added node gets attached to already existing nodes in a random manner. The edges are established at a slow rate at every time step as seen in preferential attachment algorithm but in random manner.

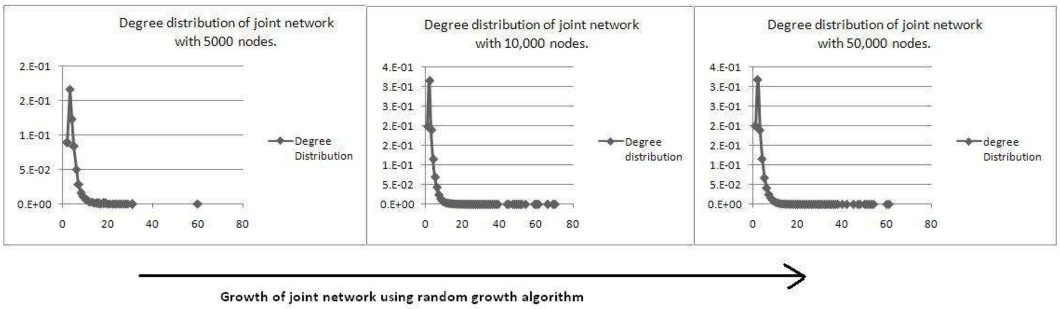

Figure 4.3: Two sample scale free networks joined using Random attachment

3. **New method of Attachment:**

   In this method at every time step the newly added node gets attached preferentially and there will also be random addition of edges at a slow rate which signify the internal evolution usually observed in social networks.

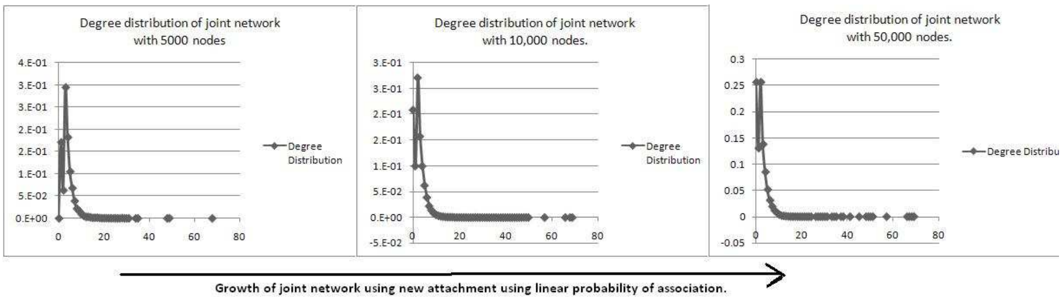

Figure 4.4: Two sample scale free networks joined using proposed approach with linear probability of association.

Figure 4.5 shows degree distributions of network grown to various sizes using step function as probability of association function. The degree distribution graphs of growth algorithms show that joint network grown using any of the network growth algorithms forms a scale free network as degree distribution exhibited by grown network is power law distribution.

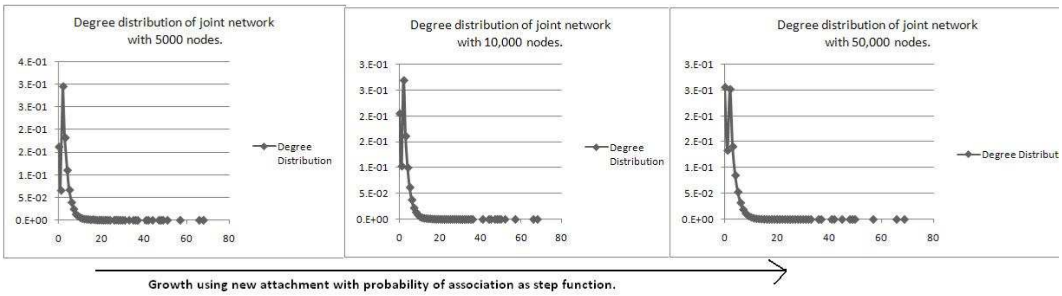

Figure 4.5: Two sample scale free networks joined using proposed approach with step function as probability of association

4.2 Small World and Random Network

A random network with 1000 nodes with average degree 1 is joined with a small world network of size 1000 nodes constructed giving input of mean degree k=10 and rewriting probability 0.1 generated using Watts and Strogatz method. We make a joint network and run the growth algorithms on them. The resultant degree distributions are shown below.

1. **Preferential Attachment:**

   Below figure shows that when preferential attachment is used to join small world and random networks, the degree distribution indicates that resultant network is totally scale free.

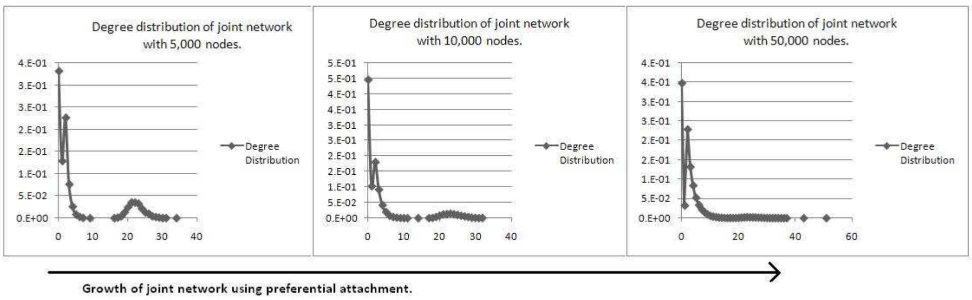

Figure 4.6: Sample small world network and random network joined and grown using preferential attachment

2. **Random Attachment:**

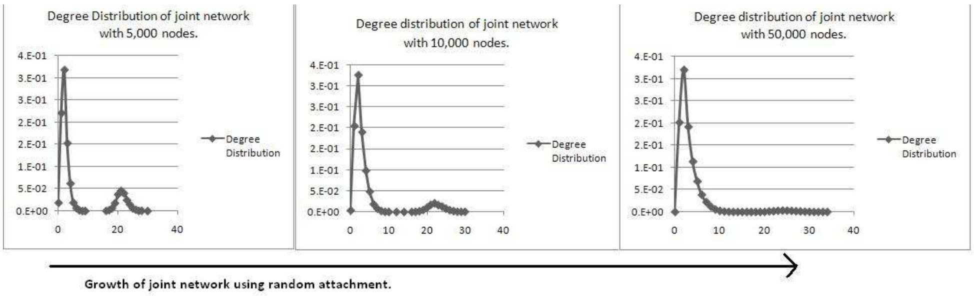

Figure 4.7: Sample small world network and random network joined random attachment

3. **New method of Attachment:**

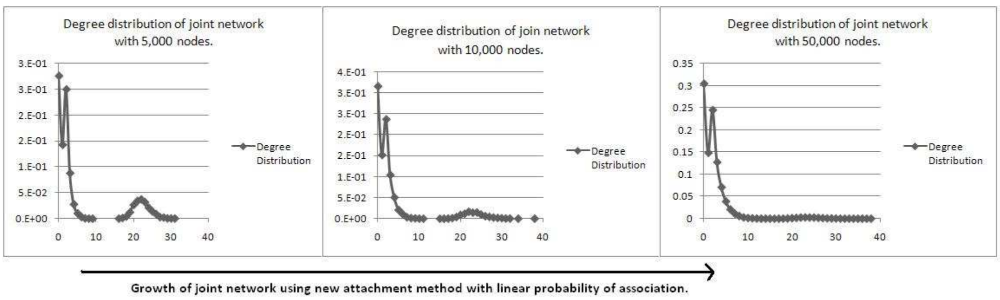

Figure 4.8: Sample small world network and random network joined proposed approach with linear probability of association

The above graphs indicate that networks obtained by joining random and small world networks have properties of both random and small world. When the new attachment method is used as growth function we see that as network size grew they spread out the Poisson distribution of degrees to have more of power degree distribution and so the resulting networks tend to gravitate more to scale free properties as network size grows.

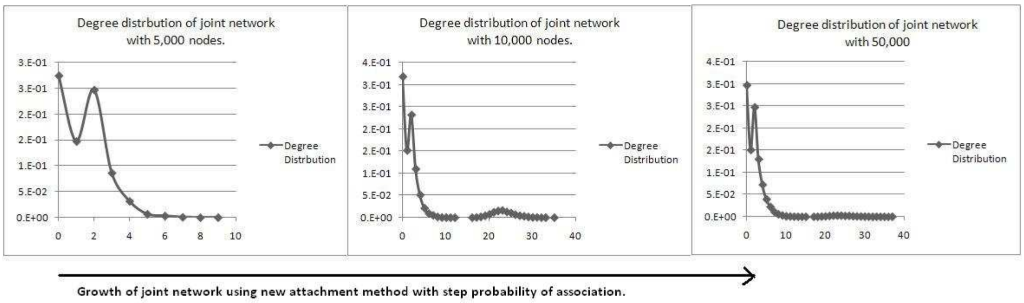

Figure 4.9: Sample small world network and random network joined using proposed approach with step function as probability of association

4.3 Small World Network and Scale Free Network

A small world network of size 1000 nodes with parameters k=10 rewriting probability 0.1 is generated using Watts and Strogatz method. A scale free network of size 1000 nodes is generated using Barabasi method of generation and they are combined to form joint network which is then subjected to the following growth algorithms.

1. **Preferential Attachment:**

   Preferential attachment used for growth decreases the average degree as more and more small degree nodes get added at each time step.

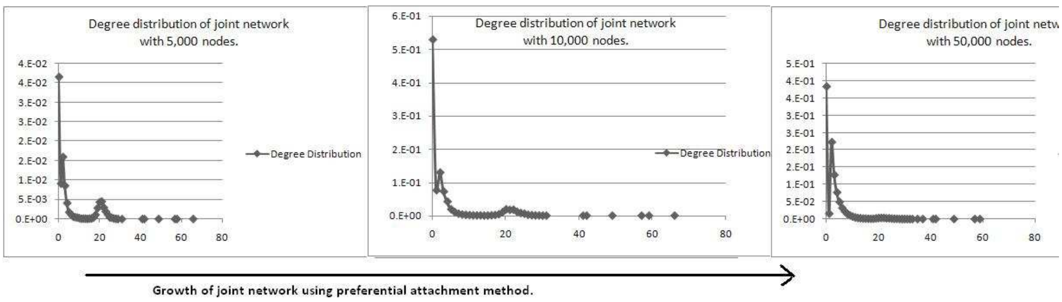

Figure 4.10: Sample small world network and scale free network joined preferential attachment

2. **Random Attachment:**

   In our experiment the random attachment growth when performed over combined network the resulting network ends up having less number of nodes having higher degree and more number of nodes having lower degree.

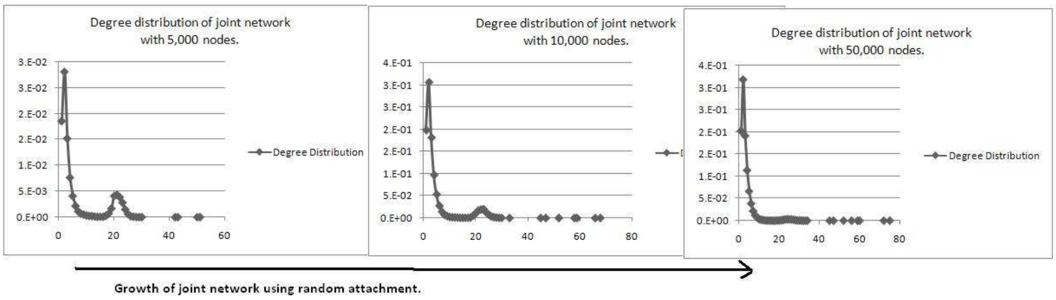

Figure 4.11: Sample small world network and scale free network joined random attachment

3. **New method of Attachment:**

   When a new attachment method with linear or square probability of association used for growth on the combined network of small world and scale free networks is used, it shows that resultant network has lower number of nodes with more degree and more number of nodes with low degree as seen in scale free network but it is not a smooth curve.

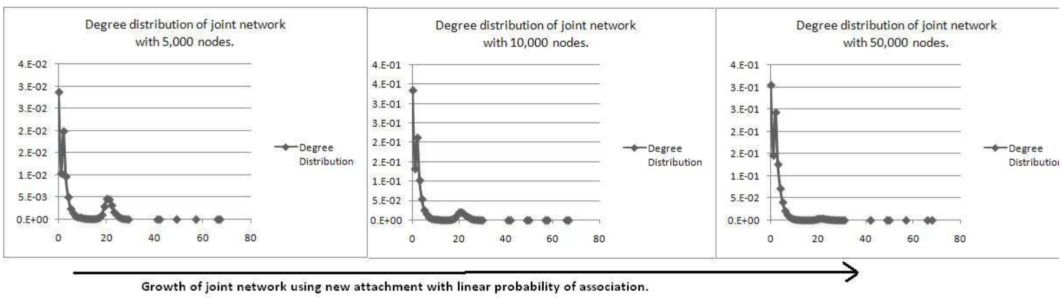

Figure 4.12: Sample small world network and scale free network joined proposed approach with linear probability of association

Figure 4.13 shows the growth of network with new method of attachment with the step function as probability of association used for internal evolution of network and

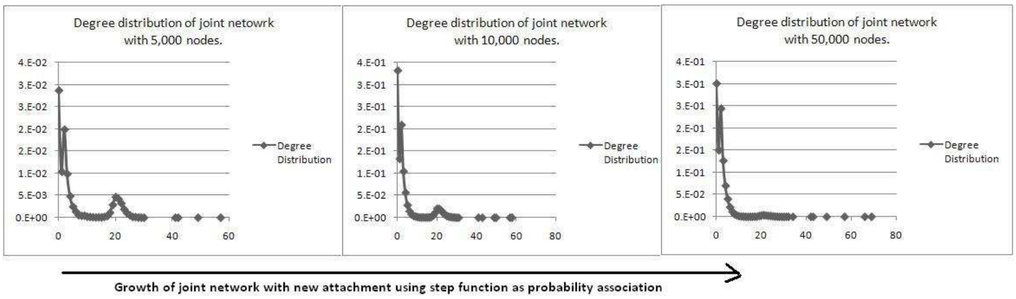

Figure 4.13: Sample small world network and scale free network joined using proposed approach with step function as probability of association

preferential attachment in addition of new nodes. The step function used is a way of classifying the nodes in a network by their degree and a function by which randomly selected pair of nodes have ties set up between them.

We have the degree distribution graphs of joined networks that were obtained by different growth algorithms on the joint network. From the degree distributions obtained we could say that the new attachment algorithm gives scale free properties to the joined network as it increases in size.

CHAPTER 5

CONCLUSION

We have simulated the growth of joint networks that are formed by merger of network models using growth algorithms. We have proposed a new growth algorithm which allows us to emulate the internal evolution within the network while new nodes are added, which was missing in Barabasi's method of evolution. Though no method could completely emulate the growth of a real world social network, our proposed approach does not only include external or newly added actors but also actors that are preexistent in the network. We have seen from the degree distributions of the grown network that our new method of attachment has accentuated the scale free properties of the joined network. We also found limits to degree of highly connected nodes in accordance with Dunbar's number.

The new method of attachment has both randomness and structure as seen in case of small world network model. Structure is achieved by emulation of internal evolution and limits on highest degree of a node and randomness by means of randomizing the selection process in internal growth. Though we put limits on maximum degree by Dunbar's number the model is random enough to allow nodes to have a degree more than this number. Degree distribution graphs suggest that joint network would eventually grow into a scale free network with gradual growth both internal and external to the joint network.

BIBLIOGRAPHY

[1] "Scale free network," http://en.wikipedia.org/wiki/Scale_free_network[Last accessed November 4, 2008].

[2] D. Watts and S. Strogatz, "Collective dynamics of'small-world' networks," *Nature(London)*, vol. 393, no. 6684, pp. 440–442, 1998.

[3] W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[4] "Wink to Merge with Reunion.com," http://www.socialnetworkingwatch.com/2008/11/wink-to-merge-w.html[Last accessed November 4, 2008].

[5] "Xing integrates with ZoomInfo, in move to overtake LinkedIn," http://venturebeat.com/2007/06/14/xing-integrates-with-zoominfo-in-move-to-overtake-linkedin[Last accessed November 4, 2008].

[6] A. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, p. 509, 1999.

[7] P. Erdos and A. Renyi, "Random graphs," *Maths Review*, pp. 290–297, 1954.

[8] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.

[9] "Program for Large Network Analysis," http://vlado.fmf.uni-lj.si/pub/networks/pajek/ [Last accessed November 4, 2008].

[10] M. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, p. 323, 2005.

[11] L. Adamic *et al.*, "Power-Law Distribution of the World Wide Web," *Science*, vol. 287, no. 5461, p. 2115, 2000.

[12] D. Puppin and F. Silvestri, "The social network of java classes," in *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing*, (New York, NY, USA), pp. 1409–1413, ACM, 2006.

[13] S. Milgram, "The small world problem," *Psychology Today*, vol. 2, no. 1, pp. 60–67, 1967.

[14] J. Guare, *Six Degrees of Separation: A Play.* Random House, 1990.

[15] P. Bork, L. Jensen, C. von Mering, A. Ramani, I. Lee, and E. Marcotte, "Protein interaction networks from yeast to human," *Current Opinion in Structural Biology*, vol. 14, no. 3, pp. 292–299, 2004.

[16] R. Albert, H. Jeong, and A.-L. Barabasi, "The diameter of the world wide web," *Nature*, vol. 401, p. 130, 1999.

[17] "Watts and Strogatz model," http://en.wikipedia.org/wiki/Watts_and_Strogatz_model [Last accessed September 9, 2008].

[18] D. Watts, "Networks, Dynamics, and the Small-World Phenomenon 1," *American Journal of Sociology*, vol. 105, no. 2, pp. 493–527, 1999.

[19] R. Dunbar, "Co-evolution of neocortex size, group size and language in humans," *Behavioral and Brain Sciences*, vol. 16, no. 4, pp. 681–735, 1993.

[20] E. N. Gilbert, "Random graphs," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.

[21] V. Batagelj and U. Brandes, "Efficient generation of large random networks," *Physical Review E*, vol. 71, no. 3, p. 36113, 2005.

VITA

Candidate for the Degree of

Master of Science

Thesis: DYNAMICS OF JOINING SOCIAL NETWORKS

Major Field: Computer Science

Biographical:

Personal Data:
Born on November, 28, 1983.

Education:

Completed the requirements for the degree of Master of Science with a major in your Computer Science Oklahoma State University in December, 2008.

Bachlor of Technology degree from B.V.Raju Institute Of Technology, Affiliated to J.N.T.U, Hyderabad, Andhra Pradesh, India, 2005, in Computer Science and Information Technology.

Experience:
Student Worker, Student Union, Oklahoma State University, Stillwater OK. Aug, 2005 - Sept, 2005.

FrontDesk and IT Migration Team Member, IT Services, Oklahoma State University, Stillwater. Sept, 2005-Dec,2005.

Software Developer, CEAT Labs, Oklahoma State University, Stillwater OK. Dec, 2006 - August, 2007.

Software Developer, RSRIT, Detroit MI. Mar, 2008 - Dec, 2008.

Name: Pradeep Koduru Date of Degree: December, 2008

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: DYNAMICS OF JOINING SOCIAL NETWORKS

Pages in Study: 35 Candidate for the Degree of Master of Science

Major Field: Computer Science

Online social networks constitute digital ecosystems. As social networks join with one another there is a need to study how the structure of joined network evolves over time. We have simulated the growth of joint networks that are formed by merger of network models using growth algorithms. We have proposed a new growth algorithm which allows us to emulate the internal evolution within the network while new nodes are added, which was missing in Barabasi's method of evolution. Though no method could completely emulate the growth of a real world social network, our proposed approach does not only include external or newly added actors but also actors that are preexistent in the network. We have seen from the degree distributions of the grown network that our new method of attachment has accentuated the scale free properties of the joined network. We also found limits to degree of highly connected nodes in accordance with Dunbar's number.

ADVISOR'S APPROVAL: ______________________________