2005

Error and attack tolerance of complex real networks

Vamsi Salla
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Salla, Vamsi, "Error and attack tolerance of complex real networks" (2005). *Graduate Theses, Dissertations, and Problem Reports*. 1680.
<https://researchrepository.wvu.edu/etd/1680>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Error and Attack Tolerance of Complex Real Networks

By

Vamsi Salla

**Thesis submitted to the College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Master of Science
in
Electrical Engineering**

Approved by

**Dr. Afzel Noore, Chair
Dr. Ronald Klein
Dr. Raymond Morehead**

Lane Department of Computer Science and Electrical Engineering

West Virginia University

**Morgantown, West Virginia
2005**

Keywords: Error and Attack tolerance, Hierarchical Networks, Cascade attacks

ABSTRACT

Real networks are often subjected to failures and attacks. Real networks are found to be robust to errors and vulnerable to attacks. This behavior of real networks is attributed to their non-homogeneous degree distribution. Non-homogenous networks are also known as scale free networks. Real networks along with the scale free property show high modularity.

Many network models have been proposed to model real networks. Erdos-Renyi Random network model is the first attempt, but fails to incorporate both properties - scale free as well as the high modularity of the real networks. Small world model shows high clustering but lacks non-homogeneous distribution. The scale free network model has non-homogeneous degree distribution but lacks the modularity. In 2002, Ravasz and Barabasi proposed a Hierarchical Network model that combines non-homogeneous degree distribution as well as high modularity showed by real networks.

The objective of this research is to study the error and attack tolerance of different network models. The static as well as dynamic tolerances of attacks are analyzed. The effect of an attack on the network model is quantified by considering the dynamic flows of quantities and using the impact factors. The results of the study show that though the scale free as well as the hierarchical models are vulnerable to attacks, the performance can be highly secured by protecting some key nodes.

ACKNOWLEDGEMENTS

The author would like to take this opportunity to thank the people who helped in the completion of this thesis. Thanks to all those whose supports made this project a reality. The author wishes to thank Dr. Afzel Noore, a graduate adviser and committee chairman, for his invaluable guidance, which was the corner stone for this problem report. Thanks also goes to Dr. Raymond Morehead and Dr. Ronald Klein for their participation in the committee and keen eye for detail.

The author also would like to thank family and friends for their moral support and understanding.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES.....	vii
LIST OF TABLES.....	ix
LIST OF ABBREVIATIONS.....	x
CHAPTER 1: INTRODUCTION	1
1.1 PROBLEM STATEMENT.....	1
1.2 RESEARCH OBJECTIVES	2
CHAPTER 2: NETWORK TOPOLOGIES	3
2.1 CHARACTERISTICS OF NETWORKS.....	3
2.1.1 DEGREE.....	3
2.1.2 SHORTEST PATH.....	3
2.1.3 AVERAGE PATH LENGTH.....	4
2.1.4 CLUSTERING COEFFICIENT.....	5
2.1.5 DEGREE DISTRIBUTION.....	6
2.1.6 EFFICIENCY.....	7
2.2.NETWORK TOPOLOGIES.....	8
2.2.1 RANDOM NETWORKS.....	9
2.2.2 SMALL WORLD MODELS.....	11
2.2.3 SCALE FREE NETWORKS.....	13
2.2.4 HIERARCHICAL NETWORKS.....	17
CHAPTER 3: LITERATURE REVIEW.....	22
3.1 INTRODUCTION.....	22

3.2 ERROR AND ATTACK MODELS.....	22
3.2.1 ERROR.....	22
3.2.2 ATTACK.....	23
3.2.3 CASCADE ATTACK.....	24
3.3 ERROR AND ATTACK ANALYSIS.....	26
3.3.1 STATIC ATTACK ANALYSIS.....	26
3.3.2 CASCADE ATTACK ANALYSIS.....	29
3.3.3 IMPACT ANALYSIS.....	32
CHAPTER 4: CHARACTERISTICS OF TEST NETWORKS.....	35
4.1 GENERATION/SYNTHESIS OF TEST NETWORKS	35
4.1.1 NETWORK REPRESENTATION.....	35
4.1.2 RANDOM NETWORK.....	36
4.1.3 SCALE FREE NETWORK.....	37
4.1.3 HIERARCHICAL NETWORK.....	38
4.2 CHARACTERISTICS OF TEST NETWORKS.....	41
4.2.1 DEGREE DISTRIBUTION.....	41
4.2.2 CLUSTERING COEFFICIENT.....	43
CHAPTER 5: PARTIAL ATTACK MODEL	45
5.1 K-BASED PARTIAL ATTACK MODEL.....	46
5.2 L-BASED PARTIAL ATTACK MODEL	47
5.3 CASCADE PARTIAL ATTACK MODEL.....	48
CHAPTER 6: EXPERIMENTAL RESULTS AND DISCUSSIONS....	49
6.1 STATIC (K-BASED & L-BASED) ATTACKS	49
6.2 CASCADE EFFECT	51
6.2.1 SIZE, EFFICIENCY & SIF.....	51

6.2.2 CIF.....	54
6.3 PARTIAL ATTACKS	56
6.3.1 K-BASED & L BASED ATTACKS.....	56
6.3.2 CASCADE ATTACK.....	62
6.3.3 EFFECT OF ALPHA.....	65
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS	69
BIBLIOGRAPHY.....	70
Appendix A: Results of Partial L-based attacks.....	73
Appendix B: Matlab code for ER random network.....	77
Appendix C: Matlab code for BA Scale free network.....	79

LIST OF FIGURES

Fig 2.1 Degree Distribution of Random Networks [1].....	10
Fig 2.2 Degree Distribution of BA Scale free Network [1].....	14
Fig 2.3 Average Path Length for BA Scale free Network [1].....	15
Fig 2.4 Clustering Coefficient of BA Scale free Network [1].....	16
Fig 2.5 Degree Distribution of Hierarchical Network [1].....	18
Fig 2.6 Clustering Coefficient of Hierarchical Network [1].....	19
Fig 2.7 Clustering Coefficient of Hierarchical Networks plotted on Size ' N ' [1].....	20
Fig 4.1 An example of a network and its adjacency matrix.....	36
Fig 4.2 Example of Random network.....	37
Fig 4.3 Example of BA Scale free network.....	37
Fig 4.4 Example of Hierarchical network.....	38
Fig 4.5 Degree distribution of Random Network.....	41
Fig 4.6 Degree distribution of BA Scale free Network.....	42
Fig 4.7 Degree distribution of Hierarchical Network.....	42
Fig 4.8 Clustering Coefficient of Random Network.....	43
Fig 4.9 Clustering Coefficient of BA Scale free Network.....	44
Fig 4.10 Clustering Coefficient of Hierarchical Network.....	44
Fig 6.1 Error and K-based attack tolerance in three network models.....	49
Fig 6.2 Error and L-based attack tolerance in three network models.....	50
Fig 6.3 Cascading effect on System Size and SIF.....	51
Fig 6.4 Cascading effect on Efficiency and SIF.....	52
Fig 6.5 CIF values for three nodes due to an attack on node with highest load.....	54
Fig 6.6 Partial K-based attack analysis on Hierarchical Network.....	56
Fig 6.7 Partial K-based attack analysis on Random Network.....	57
Fig 6.8 Partial K-based attack analysis on BA Scale free Network.....	57
Fig 6.9 Comparison of efficiency for full and partial attacks in Hierarchical and Random networks.....	58
Fig 6.10 Comparison of efficiency for full and partial attacks in BA Scale free and Random networks.....	59
Fig 6.11 Attack analysis for full degree K-based attack.....	59

Fig 6.12 Attacks analysis for partial attack of 10% K-based attack.....	60
Fig 6.13 Variation of Efficiency with variation of degree of attack.....	62
Fig 6.14 Variation of System Size with variation of degree of attack.....	63
Fig 6.15 Variation of System Impact Factor with variation of degree of attack.....	63
Fig 6.16 Variation of System Size with Alpha under Full degree attack.....	65
Fig 6.17 Variation of Efficiency with Alpha under Full degree attack.....	66
Fig 6.18 Variation of System Size with Alpha 10% degree attack.....	66
Fig 6.19 Variation of Efficiency with Alpha under 10% degree attack.....	67
Fig A.1 Partial L-based attack analysis on Hierarchical Network.....	74
Fig A.2 Partial L-based attack analysis on Random Network.....	74
Fig A.38 Partial L-based attack analysis on BA Scale free Network.....	75
Fig A.4 Comparison of efficiency for full and partial attacks in Hierarchical and Random networks.....	75
Fig A.5 Comparison of efficiency for full and partial attacks in BA Scale free and Random networks.....	76
Fig A.6 Attack analysis for full degree L-based attack.....	76
Fig A.7 Attacks analysis for partial attack of 10% L-based attack.....	77

LIST OF TABLES

Table 2.1 Overview of key characteristics of different network models.....	21
Table 4.1 Adjacency matrix of 25 node module.....	40
Table 6.1 Efficiency for 100% degree of attack.....	60
Table 6.2 Efficiency for 10% degree of attack.....	61

LIST OF ABBREVIATIONS

ER- Erdos Renyi Random network

BA – Barabasi Albert Scale free network

CHAPTER I: INTRODUCTION

1.1 PROBLEM STATEMENT

Many real networks are often subjected to failures and attacks. Most of the real networks can be represented as complex networks consisting of nodes connected by edges. Many models have been proposed for real complex networks. The availability of topological data on large real networks led to the study of generic aspects of network structure. The existing empirical and theoretical results indicate that complex networks can be divided into two major classes based on the degree distribution of the networks [6]. Homogeneous networks have a Poisson degree distribution and non-homogeneous networks have a power law degree distribution. The non-homogeneous networks are also called as scale free networks. Empirical results show that many complex real networks like WWW, Internet, etc, show a power degree distribution. Ravasz and Barabasi [18] showed that the real networks show high modularity along with the power degree distribution and proposed a Hierarchical Network model that combines both the key characteristics of real networks.

Previous research emphasized the study of error and attack tolerances of real network models [2] [6]. The analysis was performed on two network models, Random networks (with homogeneous degree distribution) and Scale free networks (with non-homogeneous degree distribution). The results showed that the scale free networks are vulnerable to attacks. Initial analysis consisted of only static analysis, Motter and Lai [15] proposed a model considering the dynamic properties of networks. It was shown that a single attack can lead to a potential network breakdown by considering the dynamic

properties of networks. The results showed that the cascading effect of an attack is higher in scale free networks than the random networks making them vulnerable to potential single attacks that can cause a global network breakdown.

This study models error and attack tolerance of hierarchical networks and compares the performance to Random and Scale free network models. The hierarchical networks are unique in the sense that they show *high modularity along with the power degree distribution*, a feature shown by real networks. Previous error and attack analysis [2] [6] based their study on a full degree attack rendering a node completely useless. Partial attacks exist in real world and we aim to provide similar analysis for different degree of attacks and show how their vulnerability to attacks can be reduced.

1.2 RESEARCH OBJECTIVES

- To study static and dynamic tolerances of different network models.
- To study the performance of different network topologies in both normal and partial attacks.
- To quantify the effect of network topology on error and attack tolerance of network models based on model for partial attacks.
- To quantify the effect of an attack on a network using the impact factors - System Impact Factor (SIF) and Component Impact Factor (CIF).
- To investigate methods of improving the performance of the real networks.
- To provide a framework for an overall analysis of robustness of a given network.

CHAPTER II: NETWORK TOPOLOGIES

This chapter provides an introduction to different network topologies for modeling real networks. It provides the reader with necessary background on the network models and their key characteristics. The first section of the chapter provides an explanation of all the terms used throughout this report. Each characteristic term is explained along with its significance. The last section of the chapter provides an explanation of existing network topologies and their key characteristics.

2.1 CHARACTERISTICS OF NETWORKS

2.1.1 Degree

The degree k_i of a node denotes the number of edges it has or the number of neighboring nodes it is connected. The average degree $\langle k \rangle$ is the average of degrees of all nodes in a network and is given by

$$\langle k \rangle = \frac{\sum_i k_i}{N}$$

where k_i is the degree of node i and N is the size of the network.

Significance: The significance of average degree $\langle k \rangle$ is that it defines the average number of edges each node has in a network. It defines the average number of neighbors for each node in a network.

2.1.2 Shortest Path

The small-world concept describes that despite their large size, there exists a relatively short path between any two nodes [1]. The shortest path between a source node

and a destination node is the least number of hops or minimum number of edges traversed to get from source to destination. There may be a number of ways to travel from a source to a destination, but the shortest path would be the path where it has to pass through the least possible edges to get to the destination. In case of graphs where the edges have a weight associated with them, the shortest path would be the path with least weight (added weights of the edges through which it traverses). Therefore by definition the shortest path between node i and node j represented by $d_{ij} \geq 1$. And if there exists a direct edge connecting i and j , then the shortest path between these nodes is $d_{ij} = 1$.

Significance: The significance of the shortest path is that in a large complex network, the connectedness of two nodes can be represented by the path length. Path length is the least number of edges from source to destination. The larger the path length, the smaller is the connectedness. Presence of a shortest path decreases the path length and in a way increases connectedness. The presence of a large number of short paths between number of nodes makes a network more connected. It makes the communication in the network easier.

2.1.3 Average Path Length

The average path length is the average of all shortest paths between all possible pairs of source and destination in a network graph. It is also called the characteristic path length or diameter of a network and can be defined as

$$L = \frac{1}{N(N-1)} \sum_{i \neq j \in N} d_{ij}$$

where d_{ij} is the shortest path between node i and node j

and N is the total number of nodes in the network.

This equation is valid only if the network is totally connected, and only if all the nodes in the network are reachable. If a node, suppose j^* , is not reachable from i^* then the shortest distance $d_{i^* j^*} = +\infty$, which makes L an ill-defined quantity [7].

Significance: The significance of the average path length is that it measures the separation of two generic nodes in a network. It characterizes the connectedness of the whole network. In a way it characterizes the communication of the network. If the average path length is small, the network is closely connected and the communication is easy and vice versa.

2.1.4 Clustering Coefficient

Forming of cliques, a circle of acquaintances in which one member knows every other member, is a common property of social networks. This tendency to form clusters can be defined in a quantitative manner using the clustering coefficient. For a node i in a network, suppose there are k_i edges connected to k_i other nodes. There would be at most $k_i(k_i - 1)/2$ edges between these nearest neighbors, if they were part of a clique. The clustering coefficient of node i is defined as the ratio between the number of edges E_i existing between k_i nodes and the total number of $k_i(k_i - 1)/2$ edges.

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

where E_i is the actual number of edges connecting the neighbor nodes,

and k_i is the degree of node i .

The clustering coefficient of the whole network is obtained by averaging all individual C_i 's. Similarly, the average of C_i 's of all nodes (i 's) with degree k_i gives the average clustering coefficient at that degree $C(k)$.

Significance: The clustering coefficient measures average cliquishness of a node. A higher clustering coefficient value indicates the presence of a highly connected cluster with the maximum value for the clustering coefficient being '1' if all the neighbor nodes are connected to each other. Suppose for nodes with degree k_1 the clustering coefficient is higher than the nodes with degree k_2 . It means that higher density clusters are formed for nodes with degree k_1 .

2.1.5 Degree Distribution

The degree distribution of a network defines how the degree is distributed over the whole network. It defines how many numbers of nodes have a given degree. Suppose $P(k)$ is the probability that a given node has the degree k , then the plot of the probabilities as function of the degree k will give the degree distribution of a network. In

a network of degree k , the probability $P(k)$ is given by $P(k) = \frac{n_k}{N}$, where n_k is the number of nodes with degree k and N is the total number of nodes in the network.

Significance: The degree distribution of a network specifies how different degrees are distributed in the whole network. If a large number of nodes have the same degree then the network can be termed as a homogeneous network. For such networks, $P(k)$ is higher for a specific degree k and average degree $\langle k \rangle \cong k$. Similarly if $P(k)$ is distributed over k with some nodes having one degree k_1 and other nodes having a different degree k_2 , then the network can be termed as non-homogeneous network.

2.1.6 Efficiency

Efficiency ε_{ij} is described as $\varepsilon_{ij} = 1/d_{ij}$ where d_{ij} is the shortest distance between nodes i and j . The global efficiency is defined as

$$E_{glob}(N) = \frac{\sum_{i \neq j \in N} \varepsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in N} \frac{1}{d_{ij}}$$

where d_{ij} is the shortest path between nodes i and j

and N is the total size of the network.

Significance: The main advantage of efficiency ε_{ij} is that when there is no path in a graph between nodes i and j , as discussed above $d_{i^*j^*} = +\infty$ and as a result $\varepsilon_{ij} = 0$.

Therefore E_{glob} quantifies the communication in a network. If E_{glob} of a network is high, the communication in the network is good.

2.2 NETWORK TOPOLOGIES

Traditionally graph theory has been used to study complex networks. Since 1950's large scale graphs with no design principles have been described using random graphs. But in the past few years, there have been advances in this field and many new concepts have been proposed. Out of these advances, three concepts occupy a prominent place in contemporary thinking about complex networks: Small worlds, Clustering and Degree distribution, leading to three main network modeling paradigms [1].

- Variant models of original Random graphs. These are still widely used for modeling many networks.
- Small-world models which are motivated by clustering.
- Scale-free models which are based on the power-law degree distribution.

Additionally there is a fourth type of network model, the hierarchical network model proposed by Ravasz and Barabasi [18]. Like the real networks, the model shows both high clustering and high modularity. The four network topologies namely Random Graphs, Small-World Models, Scale-free Networks and Hierarchical Networks will be described using the key characteristics such as Degree Distribution, Average Path Length and the Clustering coefficient.

2.2.1 Random Networks

Random graphs are used to study Complex networks with unknown organizing principles as they often appear random. The theory of random graphs was introduced by Paul Erdos and Alfred Renyi (1959, 1960, and 1961). In their model, Erdos and Renyi defined a random graph as N labeled nodes connected by n edges, which are chosen randomly from $N(N-1)/2$ possible edges. As the probability that any node in a random graph is connected to its neighbor is equal to the probability that any two nodes in the random graph are connected, the clustering coefficient of a random graph is given by

$$C_{rand} = p = \frac{\langle k \rangle}{N}.$$

where $\langle k \rangle$ is the average degree and N is the size of the network.

a) Degree Distribution

Degree distribution of the graph is obtained by studying the number of nodes with degree k . The number of nodes with degree k is given by the probability of k , $P(k)$. The number of nodes with degree k follows a Poisson distribution for random networks. Figure 2.1[1] shows the degree distribution of a numerical simulation of a random graph with $N = 10000$ nodes and connection probability $p=0.0015$.

Fig 2.1 Degree Distribution for Random Network

This indicates that the network is homogeneous, with a large number of nodes having approximately same number of edges (i.e. degree 'k').

b) Average Path Length

The diameter of a network is the average of all shortest path lengths between all pairs of nodes in the network. This characterizes the connectedness of a network. The average path length of random networks is found to be close to that of real networks.

c) Clustering Coefficient

In a random graph if we consider a node and its nearest neighbors, the probability that two of these neighbors are connected is equal to the probability that two randomly selected nodes are connected. Hence the clustering coefficient of random networks is given by

$$C_{rand} = p = \frac{\langle k \rangle}{N}$$

This indicates that if we plot $C_{rand} / \langle k \rangle$ as a function of N , the graph will align along a straight line of slope N^{-1} . Thus the Random Graphs predict that the clustering coefficient of network depends on the size of the network.

2.2.2 Small World Models

The real world networks have small-world characteristics (i.e. existence of the shortest path) and usually have large clustering coefficient [1][21][22]. Earlier it is shown that the clustering coefficient in real networks is independent of the size N of the network. This is a characteristic of ordered lattices, whose clustering coefficient is independent of size N .

Watts and Strogatz (1998) proposed a one-parameter model that interpolates between an ordered finite-dimensional lattice and a random graph. It is known as the **WS** model. These small-world models have short average path length and high clustering coefficient.

a. Degree Distribution

In the Small-World model proposed by Watts-Strogatz, each node has the same degree k and hence the WS model has a Poisson degree distribution similar to random graph. The topology of this type of network is relatively homogeneous, with all nodes having approximately same number of edges.

b. Average Path Length

The WS model has a change in the scaling of the characteristic path length l . For small p , l scales linearly with system size, while for large p the scaling is logarithmic. The origin of rapid drop in l is the appearance of shortcuts between nodes. Every shortcut, created randomly, is likely to connect widely separated parts of the graph, and thus has a significant path length of the entire graph.

c. Clustering Coefficient

In addition to a short average path length, small-world networks have a relatively high clustering coefficient. This is in accordance with the properties of real networks which show a high degree of clustering coefficient.

2.2.3 Scale Free Networks

Empirical results on real networks show that many large networks are scale free, i.e. their degree distribution follows a power law for large k . Even for the networks that show an exponential tail, the degree distribution significantly deviates from a Poisson distribution. Both the random graph model as well as the **WS** model cannot produce this feature. Assuming that if we capture the process of assembling networks correctly, we will attain their topology correctly, scale free networks emphasize on the network dynamics.

The Barabasi-Albert {BA} Model [1] proposed for the scale-free models, is governed by two principles.

- Real networks grow by continuous addition of new nodes making it an open system. This was not included in the two previous models, which start with fixed number of N vertices and are then connected or re-wired without modifying N .
- Most of the Real Networks follow preferential attachment while connecting to other nodes, making it dependent on the degree of a node. In the two previous models the probability that two nodes would be connected was independent of the node's degree.

a) Degree Distribution

The main objective of the BA model was to obtain a scale-free network i.e. to obtain a network with a power law degree distribution. Figure 2.2[1] shows a degree distribution of a BA model with $N = 300000$ and $m_0 = 1$. 'm0' is the number of nodes at the start, from where other nodes are added to build the system.

Fig 2.2 Degree Distribution for BA Scale free Network

The graph in Figure 2.2[1] indicates that the number of nodes having large number of edges (high 'k') is less and decreases with 'k'. It means that the number of nodes having fewer edges is very high. In scale-free networks few nodes are highly connected to hold the whole network together, and a large number of nodes are loosely connected. The scale-free model creates significantly heterogeneous networks.

b) Average Path Length

Figure 2.3[1] shows the comparison of average path length ' l ', for the BA model and the random graph model with same average degree ' k ' and size ' N '.

Fig: 2.3 Average Path Length for BA Scale free network

Figure 2.3 indicates that the average path length is smaller in case of the BA model, indicating that the non-homogeneous scale-free network topology has densely connected nodes than the homogeneous topology of the random graphs.

c) Clustering Coefficient

There is no analytical prediction for the clustering coefficient of this model. The clustering coefficient of a Barabasi-Albert model with average degree $\langle k \rangle = 4$ is compared with that of a Random graph. The plot of the clustering coefficients of these two models is shown in Figure 2.4[1]. Here we find that the clustering coefficient of Scale free model is five times higher than the random graph model and it slowly increases with number of nodes.

Fig: 2.4 Clustering Coefficient of BA Scale free Network

The clustering coefficient of the Barabasi-Albert model decreases with increasing N (network size) and follows a power law $C \cong N^{-0.75}$. Though this is slower than the $C \sim N^{-1}$ decay for the random graph, it is still dependent on the network size and is in contrast with the small world models where C is independent of N .

2.2.4 Hierarchical Networks

As described above Real Networks in society and nature have two generic properties: they are scale-free and they display high degree of clustering [18]. Ravasz and Barabasi showed that these two features of the real networks is a consequence of a hierarchical organization - small groups of nodes hierarchically organize themselves to form increasingly large networks.

The scale-free property and clustering are found together in a number of real networks, metabolic networks [11] [22], protein interaction network [22], World Wide Web [3] and even in some social networks [16] [5]. But as we have seen, most of the models that were proposed to describe the topology of the complex network do not capture both features.

In this model Ravasz and Barabasi discuss that the hierarchical topology of the real networks is the feature that forms the root of the high degree of clustering and scale-free topology. Hence for their model they assumed that modules combine into each other in a hierarchical manner to form a hierarchical network. The presence of hierarchy and

scale-free property impose strict restrictions on the number different groups present in the network along with their degree of cohesiveness. The number of groups can be quantitatively captured using a scaling law that describes the dependence of the clustering coefficient on the node degree [18]. The presence of hierarchy in real networks is identified using this scaling law.

a) Degree Distribution:

The hierarchical model has the scale-free feature of the BA model. Hence it has a power-law degree distribution and produces a heterogeneous network model.

Fig 2.5 Degree Distribution of Hierarchical Network

Figure 2.5[1] shows the degree distribution of a hierarchical network with $N = 5^7$.

b) Clustering Coefficient

The clustering coefficient of this model is independent of the size of the network ‘N’.

In [10], Dorogovtsev, Goltsev and Mendes explained that in deterministic scale-free networks the clustering coefficient of a node with k links follows the scaling law

$$C(k) \sim k^{-1}$$

Ravasz and Barabasi show that the intrinsic hierarchy can be characterized in a quantitative manner using this scaling law. They showed that for the hierarchical model proposed, the clustering coefficient C follows the $1/k$ rule. The clustering coefficient of a hierarchical network with $N = 5^7$ is shown in the Figure 2.6[1].

Fig: 2.6 Clustering Coefficient of Hierarchical Network

Figure 2.6[1] shows that $C(k)$ is inversely dependent on the degree k . Hence this curve can be used to determine if there is a hierarchy present in a network.

Figure 2.7[1] shows a plot of the clustering coefficient with respect to the size of the network 'N'. Here we observe that for the hierarchical network model the clustering coefficient is independent of the size of the network.

Fig: 2.7 Clustering Coefficient of Hierarchical Networks plotted against Size 'N'

Table 2.1 summarizes the properties of the four network topologies discussed above and compares them with properties of the real networks.

Table 2.1 Overview of key characteristics of different network models

Network Topology	Degree Distribution $P(k)$	Clustering Coefficient
Real Networks	Power-law	Independent of 'N'
Random Networks	Poisson	Dependent on 'N'
Small World Models	Poisson	Dependent on 'N'
Scale-Free Model	Power-law	Dependent on 'N'
Hierarchical Model	Power-law	Independent of 'N'

CHAPTER III: LITERATURE REVIEW

3.1 INTRODUCTION

Many complex systems show a high degree of tolerance against errors. For example many organisms of the metabolic system grow, persist and reproduce despite drastic pharmaceutical and environmental conditions [19]. Even with regular malfunctions of key features and local features the communication networks that are complex show a surprising degree of robustness.

This chapter provides an introduction to different attack models. It summarizes the literature on current research in the error and attacks analysis of different network models. The first section explains all the error and attack models used for assessing the performance of different network topologies. The last section provides brief summary of important research and their main results in error and attack analysis.

3.2 ERROR AND ATTACK MODELS

3.2.1 Error

Errors frequently occur in real networks. Many systems are subjected to failures that are unexpected, which cannot be accounted for in network models. Behavior of different network models under such failures is described. The errors or failures in a network are modeled by random removal of nodes. As failures occur randomly in real world, this model is appropriate. A removal of a node due to an error removes all the

edges of the node. The performance of the network is studied by determining the global efficiency of the network.

3.2.2 Attack

Attack on a network is a deliberately planned and designed action to damage the integrity of a network. An attack on a network is usually planned in a way to damage a network to a maximum amount. Hence an attack is modeled by removing the most important node in the network. The importance of a node is determined in two different ways in a network.

- **K-Based:** In the first approach, the importance of a node is based on the degree of the node. As explained earlier the degree of a node is the number of edges that pass through the node. It can also be quantified as the number of nodes to which it is directly connected. The node with highest degree is considered the node with highest importance. The node with the next highest degree is the node with the next importance and is chosen to attack, and so on. The performance of the network is studied by determining the global efficiency of the network.
- **L-Based:** The second approach to determine the importance of a node is based on the load of the node. The load is also called as the betweenness of the node. The load of the node is the total number of paths that pass through the node. The loads of the node are calculated in the free-state i.e. when no nodes are removed. The node with highest load is selected as the node with highest importance. The node with next highest load is the second important, and so on. The performance of network is studied using the global efficiency of the network.

3.2.3 Cascade attack

The previous two attack models (K-based and L-based) focus on the static properties of the network. But if the dynamic redistribution of flows is taken into account, a single attack on a node is sufficient to collapse the entire network. The effect of an attack can be more devastating if the intrinsic dynamics of the flows of quantity in a network is taken into account.

The model followed is that in any network, initially when the network is fully functional i.e. in free-state each node carries some initial load. The load on a node is the total number of shortest paths passing through the node. We assume that the capacity of the node is proportional to the initial load of the node. It is given by

$$C_j = \alpha L_j(0), \quad j = 1, 2, \dots, N,$$

where $\alpha \geq 0$ is the tolerance parameter and N is the initial number of nodes.

When a node in a network is made non-functional due to either a failure or an attack, changes in the flow of the quantities occur and eventually the load on each node also changes. This leads to global redistribution of load all over the network. It can cause overload failures for some nodes in the network which again leads to redistribution of loads. This phenomenon can continue until a network breakdown occurs. This kind of attack is known as cascade attacks. The method followed is that if the load after redistribution exceeds the capacity of a node, then the node is considered not functional.

The node is removed from the network, thereby removing all the edges of the node. This is described as

$$e_{ij}(t+1) = \begin{cases} 0 & \text{if } L_i(t) > C_i \\ e_{ij}(0) & \text{if } L_i(t) \leq C_i \end{cases}$$

where e_{ij} is the weight of the edge connecting nodes i and j .

$L_i(t)$ is the load on the node at time t .

C_i is the initial capacity on node i .

Cascading failures are common in most of the complex networks. The cascade attacks are present in many infrastructure networks.

The parameters studied under this kind of attack model are:

- The size of the main cluster of the network
- The global efficiency of the network

3.3 ERROR AND ATTACK ANALYSIS

3.3.1 Static Attack analysis

For the first time [2], the authors studied the error and attack tolerances of two types of networks – ER random model and the BA model for scale-free networks. They compared the robustness of these two types of networks with same number of nodes and edges, differing only in degree distribution. Two types of node removal were considered.

- **Errors:** These are the errors that may occur naturally. Vertices are *randomly* broken and disappear along with their connections (edges).
- **Attacks:** These are damages that are caused by planned and designed attacks on the networks. The vertices are *deliberately* broken and the most important nodes are chosen based on some criteria.

By removing a node, all the edges that connect to it are also removed, disrupting some of the paths between the remaining nodes. The attack on the networks is based on the degree of the node (k – based attack). The parameters that are studied on the two types of the networks are the average path length (also called as the diameter) that describes the interconnectedness of a network and the size of the largest cluster as a fraction of the total system size.

- The average path length characterizes the closeness or connectedness of a network and is defined as the average length l of the shortest paths between any

two nodes in the network. The smaller the diameter, the shorter is the expected path between the two nodes and hence has a better ability to communicate with each other.

- The size of the largest cluster as a fraction of the total size characterizes the dependency of the network integrity on the degree distribution. After the removal of the nodes due to an attack, the relative size of the largest cluster S that remains connected quantifies the dependency of the network on the effected nodes. The relative size of the largest cluster is given by the number of nodes in the cluster as a fraction of the total size of the network N . Hence $0 \leq S \leq 1$, with $S = 1$ when there are no nodes removed.

The main results and conclusions from their study are:

- a) The error and attack tolerances of the network are dependent on the topology of the network.
- b) There are two types of network topology – Poisson and Power law.
- c) Poisson distribution networks are homogeneous (Erdos-Random network) and their error and attack tolerance is almost similar.
- d) Power law distribution networks are non-homogeneous (BA network model) and are robust to errors and vulnerable to attacks.

A similar study for the www as well as the Internet networks is also provided [2].

In 2003, [7] the authors studied the effect of error and attack tolerance on the random graph model and the BA scale-free model. The attack on the networks is based on the load of the node (L-based attack). They studied the performance of these networks with the average path length l , clustering coefficient C , and they introduced two more parameters – Global efficiency (E_{glob}), and Local efficiency (E_{loc}). These two parameters are described as follows:

➤ Global Efficiency (E_{glob}): A characteristic efficiency ε_{ij} is described as

$\varepsilon_{ij} = 1/d_{ij}$ where d_{ij} is the shortest distance between nodes i and j . The global efficiency is defined as

$$E_{glob}(N) = \frac{\sum_{i \neq j \in N} \varepsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in N} \frac{1}{d_{ij}}$$

where d_{ij} is the shortest path between nodes i and j

and N is the total size of the network.

The main advantage of efficiency ε_{ij} is that when there is no path in a graph between nodes i and j , $d_{ij} = +\infty$ and $\varepsilon_{ij} = 0$. Therefore the E_{glob} in a way quantifies the communication of a network. When E_{glob} of a network is high, it specifies that the communication in the network is good.

➤ Local Efficiency (E_{loc}): Local efficiency E_{loc} is analogous to the clustering coefficient C , and is defined as:

$$E_{loc}(N) = \frac{1}{N} \sum_{i \in N} E_{glob}(n_i)$$

where n_i is the number of neighbors of the node i

and N is the total size of the network.

The E_{loc} quantifies the local communication, i.e. the communication between the neighbors of each node.

The main results and conclusions from their study are:

- a) The error and attack tolerances of a network is dependent on the topology of the network.
- b) Poisson distribution networks are homogeneous (Erdos-Random network) and their error and attack tolerance is almost similar.
- c) Power law distribution networks which are non-homogeneous (BA network model) are robust to errors and vulnerable to attacks.
- d) The usage of the efficiency is better than that of the average path length (L).

3.3.2 Cascade Attack Analysis

The previous K-based and L-based attacks mainly study the static properties. But the effect can be more devastating if the intrinsic dynamics of the flows of quantity in a network is taken into account [15]. In any network, each node has an initial node. When a node in a network is damaged, due to either a failure or an attack, the flow of the quantities and hence the load of each node changes. This leads to global redistribution of loads all over the network and can cause overload failures for some nodes in the network

that again leads to redistribution of loads. If this phenomenon continues, then it causes a network to breakdown. This kind of attack is known as cascade attacks.

In 2002 [15], A. E. Motter and Y. C. Lai first proposed a dynamical model to study the cascade-based attacks on complex networks. They studied the changes in the size of the main cluster due to the effect of an attack to show the effect. The model proposed is as follows:

Suppose in a given network, at each time step of one unit a relevant quantity is exchanged between every pair of nodes along the shortest path [15]. The load of any node is the total number of shortest paths passing through that node. Every node has a capacity which is the maximum load the node can handle at any instance. The capacity is limited by cost. In their model Motter and Lai assumed that the capacity of node is proportional to the initial load of the node.

$$C_j = (1 + \alpha)L_j, \quad j = 1, 2, \dots, N,$$

where $\alpha \geq 0$ is the tolerance parameter and N is the initial number of nodes.

Under normal conditions the system works in free-flow state. But removal of a node results in redistribution of the shortest paths and the load at a node can exceed the capacity resulting in failure of the node. In this manner, the cascades triggered by single node removal are modeled. The main results are:

- a) The real networks are robust to failures but are vulnerable to attacks due to the presence of few nodes with large exponential load.

- b) By taking the dynamic flows of a network into consideration, the attacks with cascading effect can be more devastating than the normal K-based and L-based attacks.

In 2004, Crucitti *et al* [6] proposed a dynamical model for the cascading effects of an attack. They studied the efficiency to show the effect of an attack. The method proposed is as follows:

Similar to the model proposed by Motter and Lai, here the authors assume that the capacity of a node is proportional to the initial load of the node, where load is the total number of shortest paths passing through the node. Hence capacity can be given by

$$C_j = \alpha L_j(0), \quad j = 1, 2, \dots, N,$$

where $\alpha \geq 1$ is the tolerance parameter of the network

and $L_j(0)$ is the load at time 0, initial load.

This is realistic as the capacity of a network cannot be infinitely large because it is limited by cost. The network initially operates in a satisfactory state. But removing a node can cause redistribution of paths and loads, creating overloads at some nodes. The rule followed at each time t is

$$e_{ij}(t+1) = \begin{cases} e_{ij}(0) \frac{C_i}{L_i(t)} & \text{if } L_i(t) > C_i \\ e_{ij}(0) & \text{if } L_i(t) \leq C_i \end{cases}$$

where e_{ij} is the weight of the edge connecting nodes i and j .

$L_i(t)$ is the load on the node at time t .

C_i is the initial capacity on node i .

and j extends to all the first neighbors of i . If at any time t a node is congested, the load exceeds the capacity of a node, the efficiency of all the edges passing through the node is reduced, so that eventually the information takes alternate paths. The important results are:

- a) The cascading effect can lead to system breakdown if the node that is attacked is one with higher loads in the network.
- b) This is particularly important for real-world networks with highly heterogeneous distribution like the Internet.

3.3.3 Impact Analysis

In 2003, Salim *et al* [19] proposed an online monitoring and analysis framework, formulating a theoretical basis for constructing global metrics for impact analysis. An agent-based architecture is presented that continuously monitors network vulnerability metrics. It provides new ways to measure the impact of faults and attacks, and for monitoring and quantifying component behavior [19]. Two metrics are proposed to quantify the impact of a fault or an attack at the component or at the system level.

Component Impact Factor (CIF): The impact on individual network components can be characterized and quantified using the CIF. It can be defined as the ratio between the changes occurred due to the fault to the minimum change required to change the network to abnormal state.

$$CIF(FS_k) = \frac{|Norm - Fault|}{|Norm - Min|}$$

where FS_k is the fault scenario.

Norm is the value of the parameter observed under the normal working condition

Fault is the value of the parameter observed under the faulty working condition

Min is the least changed value of the parameter to make the system work in abnormal state

We can select the parameter and also set the threshold for obtaining the quantitative values to represent the normal, fault and minimum operating conditions. For example Salim and Ramkishore [19] calculated Component Impact Factor for an example network of the Internet level (synthesized) and used three different parameters – data transfer rate, buffer utilization, and connection queue length. Different thresholds were selected for different parameters.

System Impact Factor: The impact of a fault on the whole network can be characterized and quantified using SIF. It is obtained by evaluating the weighted impact factors of all network components [19]. That is, SIF gives the percentage of components operating in

vulnerable state (components whose CIF's exceed normal operating thresholds), compared to total number of components. It can be defined by the equation below

$$SIF(FS_k) = \frac{\sum_{\forall j, CIF_j > d} COS_j}{total\ number\ of\ clients}$$

where 'd' is the threshold and COS denotes the components operating state. COS is '1' if the component is not operating in normal state (i.e. $CIF > d$) and it is equal to '0' if it is operating in normal condition (i.e. $CIF < d$).

The important results of their work are:

- a) The parameters CIF (component impact factor) and SIF (system impact factor) were defined.
- b) They proposed a model to monitor the performance of a network system by keeping track of these parameters online.

CHAPTER IV: CHARACTERISTICS OF TEST NETWORKS

This chapter explains the networks used in this research for the study of error and attack tolerance. It explains the different networks used and how they are generated along with their important properties – degree distribution and clustering coefficient. The next section of the chapter explains the representation and generation of networks used. The last section explains the network properties along with their significance.

4.1 GENERATION/SYNTHESIS OF TEST NETWORKS

4.1.1 Network Representation:

A network can be represented in several ways – adjacency list, adjacency matrix, etc. Here the networks are represented in adjacency matrix form. A brief explanation about adjacency and adjacency matrix is given in the section below.

Adjacency: Two nodes in a graph are called adjacent if they are the endpoints associated with same arc/branch/edge.

The adjacency matrix is the matrix that defines the connectivity between the nodes of a network. If there are ' N ' numbers of nodes in a network, then the size of the adjacency matrix would be $N \times N$.

Each element of the matrix would be of value '**0**' or '**1**'. A value of **zero** denotes that there is no connection between nodes represented by the row and column number of the element. If there is a connection then a value of **one** is used.

Figure 4.1 below shows an example of a network and the adjacency matrix of the network.

Fig 4.1 An example of a network and its adjacency matrix

In the adjacency matrix, the highlighted '0' shows that there is no direct connection between node 1 (row) and node 4 (column) and hence the value zero. Similarly the highlighted '1' shows that there is a direct connection between node 4 (row) and node 2 (column). In this manner any network can be represented by a 2 dimensional $N \times N$ matrix where N is the number of nodes in the network.

4.1.2 Random Network

The Random network model proposed by Erdos and Renyi is generated using an existing algorithm in MATLAB. The code for the algorithm is shown in Appendix B [17]. An example of random network is shown Figure 4.2 [23].

Fig 4.2 Example of a random network

4.1.3 Scale Free Network

The Scale free network model proposed by Barabasi and Albert is generated using an existing algorithm in MATLAB. The code of the algorithm is added in Appendix C [14]. An example of BA scale free network is shown in figure 4.3 [23].

Fig 4.3 Example of BA Scale free network

4.1.4 Hierarchical Network

The hierarchical network model as proposed by Ravasz and Barabasi [18] is constructed using an algorithm implemented in JAVA. The basic idea behind the algorithm is that the hierarchical networks are made of small identical networks. Hence any hierarchical network can be considered as a network with repetitive small structures. These small repetitive structures will have similar connections within themselves. That is, the adjacency matrices of these small structures will be identical. This is possible only because we are considering non weighted networks. So, if we can identify the basic structure of a hierarchical network, the whole network can be constructed based on the specific network structure.

For example let us consider the network shown in Figure 4.4 [18]. The basic module is comprised of 5 nodes. Hence, we can generate the 125 node network, based on the 5 node structure.

Fig 4.4 Example of Hierarchical network

The adjacency matrices for the 5-node as well as the 25-node networks are shown below.

In the 25-node network's adjacency matrix, we show how the 5-node structure's adjacency matrices repeat itself.

Adjacency matrix of the 5-node module

nodes	1	2	3	4	5
1	0	1	1	1	1
2	1	0	1	1	1
3	1	1	0	1	1
4	1	1	1	0	1
5	1	1	1	1	0

Adjacency matrix of the 25-node module

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1
2	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
13	1	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
14	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
15	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0

17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0
18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0
19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0
20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1
23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
24	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1
25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0

Table 4.1 Adjacency Matrix of the 25 node module

The adjacency matrix of a 5-node basic structure repeats 5 times in the 25 node network as shown in Table 4.1. The darkened 1's show the connections between these five 5-node structures. Similarly the 125-node network can be constructed using the basic structure.

The algorithm used for generating the network model is as shown below:

Algorithm:

Inputs: Adjacency Matrix of Basic Module or Modules – a [] (stored in the array).
Number of nodes in the complex network – N.

For each $n \in N$ the sequence number of the node relates to the sequence number in the array of the adjacency matrix 'a'.

The adjacency matrix of node $n1 \in N$ is stored in a [1] and $n1 \in N$ in a [2] etc.

For k =1 to N (total number of nodes in the complex network)

- For size of a [k-1] < **row** > size of a [k]
 - For size of a [k-1] < **column** > size of a [k]
 - B [row] [column] = a [k][row] [column]
 - End
- End

End

Then the code for connections between N nodes is specific to the design of complex network.

4.2 CHARACTERISTICS OF TEST NETWORKS

4.2.1 Degree Distribution

The degree distribution for the Random, BA and Hierarchical network models generated are shown in the plots below. Figure 4.5 shows the degree distribution of an ER random graph. The random graph model generates a homogeneous network (a Poisson distribution), with most of the nodes having a degree equal to average degree $\langle k \rangle = 7.34$.

Fig 4.5 Degree distribution of Random Network
Number of Nodes = 125
Number of Edges = 394

Similarly Figure 4.6 and Figure 4.7 show the degree distributions of BA Scale free and hierarchical network models respectively. Here the degree distribution follows a power-law showing that the BA as well as the hierarchical networks are highly non-homogeneous. For both these networks, few nodes have high degree, whereas a large number of nodes have lower degree.

Fig 4.6 Degree distribution of BA Scale free Network
 Number of Nodes = 125
 Number of Edges = 394

Fig 4.7 Degree distribution of Hierarchical Network
 Number of Nodes = 125
 Number of Edges = 394

4.2.2 Clustering Coefficient

Ravasz and Barabasi [18] showed that a network can be quantified to have a hierarchical organization if the clustering coefficient follows a scaling law.

$$C(k) \sim k^{-1}$$

Hence the $C(k)$ vs. k curves are plotted for all the three network models under consideration. Figures 4.8 and 4.9 show the organization of clustering in a random network and BA scale free network respectively. The graphs clearly show that there is no hierarchical organization present in both the ER random model as well as the BA scale free model networks, since the clustering coefficient does not follow the scaling law.

Fig 4.8 Clustering Coefficient of Random Network
Number of Nodes = 125
Number of Edges = 394

Fig 4.9 Clustering Coefficient of BA Scale free Network
Number of Nodes = 125
Number of Edges = 394

Figure 4.10 shows the clustering organization in the hierarchical networks. The graph shows that the clustering follows a scaling law in the hierarchical network models quantifying the presence of high modularity in these network models like most of the real networks.

Fig 4.10 Clustering Coefficient of Hierarchical Network
Number of Nodes = 125
Number of Edges = 394

CHAPTER V: PARTIAL ATTACK MODEL

An attack on a node affects all the edges of the node. The attacks modeled so far are complete or full degree attacks, where the node is completely removed by making it unconnected to the network. This is obtained by removing the edges of the node that is attacked. In other words the efficiency of the edges of the node under attack is made zero. Hence we can say that the attacks modeled are based on the effective performance of the edges of the attacked node.

In all the previous studies [2] [6], the authors focused their study on the full degree attacks (K-based attack, L-based attack and the cascade attack). Normally in real world, partial attacks can occur. Instead of damaging the whole node, the performance of the node is reduced to some extent. The reduction in the performance of the node can be modeled by reducing the efficiency of the edges of the node. In this case the efficiency of the edges are not directly made to zero making the connection between the nodes useless, but is reduced by a degree. In other words, the distance between the nodes is increased or the weight of all the edges of the node is increased. The weight of the node characterizes the edge that connects the nodes. The lesser the weight of the edge the easier it is for information to pass through the nodes using the edge. Hence we can say that for an edge initially carrying certain amount of information, to carry the same amount of information, the cost will be higher as it has to travel a longer distance than in normal condition.

For example, suppose an attack of degree 10% is modeled on a node. It reduces the performance of the edges of the node by 10%. It will reduce the functionality of the edge by 10%. So each time the attack is modeled on a node, the functionality of its edges is reduced by 10%. The reduction of the edge functionality can be considered as reduction in the ability to communicate using the edge. This can be represented by reducing the efficiency of the edge i.e. increasing the weight of the edge. Here we modeled the attacks of varying degree from 10 to 100%. The 100% reduction being that all the edges of the attacked node are removed making the node unconnected to the network. These varying degree attacks are based on degree (K-based), load (L-based) and for the cascading attack models as described below.

5.1 K-based partial attack model

In the K-based attack model, we calculate the degree of each node. Then the nodes are attacked according to the ascending order of their degree. Since for varying degrees of attack, the edges of a node are not removed but instead the weights are increased by a percentage, we consider a parameter '*effective-k*', which is the sum of all the efficiencies of all the edges of each node. Now initially under normal conditions the efficiencies of the edges are either '1' or '0'. Therefore the degree 'k' of a node is equal to the sum of the efficiencies of the edges of the node. Hence the effective-k is proper assumption.

In this attack model, after an attack is modeled and the efficiencies of the edges are reduced, the effective-k is recalculated for all the nodes. Now the node with the highest effective-k is attacked. So at each step after an attack, effective-k is calculated. In this way at every step of the attack model, we attack the node that has edges with highest efficiencies or in other words the node that has the highest efficiency of sending information.

5.2 L-based partial attack model

Similar to that of the K-based varying degree of attacks, in the L-based model we recalculate the loads on the nodes after each attack. After an attack, the efficiency of the edges of the node that was attacked changes based on the degree of the attack. The shortest paths of the network get redistributed resulting in redistribution of the loads. Hence after each attack the loads on each node are recalculated and the node with the highest load at the time is selected and attacked. In this way at every step of the attack model, we attack the node that carries the highest load at that time to make sure the damage is highest.

5.3 Cascade partial attack model

The cascading effect of the varying degree attacks on the node with highest load is studied here. That is, instead of removing the edges of the node, the efficiency of the edges are reduced by the percentage degree of attack. As explained earlier, even by

changing the efficiency of the edges of a node, redistribution of loads can occur, leading to a cascading effect. So the cascading effect is studied for varying degrees of attacks from 10% to 100%. The parameters studied are:

- The size of the main cluster of the network.
- The global efficiency of the network.
- The System Impact Factor (SIF) quantifying the total percentage of nodes in the network working under abnormal state.
- The Component Impact Factor (CIF) quantifying the effect of an attack on individual nodes.

CHAPTER VI: EXPERIMENTAL RESULTS AND DISCUSSIONS

This chapter outlines all the experimental results along with important conclusions. The first two sections of the chapter show the results for the error and attack tolerances of static (k-based and L-based) and cascade attacks respectively. The results are in agreement with earlier studies, as will be explained in sections to follow. The final chapter of this section explains the results obtained by the proposed partial attack model with varying degree attacks.

6.1 STATIC (K-BASED & L-BASED) ATTACK

The error and attack tolerance of three types of networks - Random, BA scale-free and Hierarchical networks are tested under the error, K-based and L-based attack models. The efficiency of the network is the parameter considered to study the effect of error, K-based and L-based attacks. Figure 6.1 below shows the behavior of networks under random failures and K-based attack. Similarly Figure 6.2 shows the behavior of the networks under random failures and L-based attack.

Fig 6.1 Error and K-based attack tolerance in three network models

Fig 6.2 Error and L-based attack tolerance in three network models

From the plots above, we see that for the Random network, there is not much difference in behavior both in the error as well as attack mode. Whereas in the case of the hierarchical as well as the BA Scale free networks, there is a considerable drop in the efficiency with increasing 'p', in case of attack (both k as well as L). This confirms that the BA scale free and hierarchical networks are very vulnerable to attacks. It can be explained by the degree distribution of the network models, which show a power law distribution. This means that few nodes in the network have high degree values and when these nodes are attacked deliberately, the efficiency of the network drops considerably.

In conclusion, we can infer that the hierarchical as well as the BA Scale free networks are robust to errors but are very vulnerable to attacks due their non-homogeneous degree distribution. The random networks, since they are more homogeneous show a similar behavior to both errors and attacks.

6.2 CASCADE EFFECT

6.2.1 Size, Efficiency & SIF

When a node is attacked, the attack leads to a network breakdown due to the cascading effect. This can be explained by studying the effect of the attack on the network with respect to time on three different factors – the size of the main cluster, efficiency and the System Impact Factor. The capacity of a node is proportional to the initial load of the node and is given by equation.

$$C_j = \alpha L_j(0), \quad j = 1, 2, \dots, N,$$

where $\alpha \geq 1$ is the tolerance parameter of the network

and $L_j(0)$ is the load at time 0, initial load.

The value of alpha is taken to be equal to 1.05. The node to be attacked is based on load (L-based). The node that initially carries the highest load is attacked. Figures 6.3 and 6.4 show the results of cascading attack.

Fig 6.3 Cascading effect on System Size and SIF

Fig 6.4 Cascading effect on Efficiency and SIF

From Figures 6.3 and 6.4, we see that the size and the efficiency of the BA scale free and hierarchical networks drop significantly with time when compared to random networks. It shows that the cascading effect is significantly high in these networks. This can be attributed to the non-homogenous degree distribution of these networks. In these networks, system integrity is dependent on the key nodes and an attack on the key node can lead to network breakdown as shown.

In Figure 6.3, for the hierarchical network, the size drops down to 20% of the original size, just by attacking the key node, whereas the SIF is zero meaning that all the other nodes are operating on a load below their capacity. This is because the system has broken down to small clusters of connected nodes. In case of the BA scale free network model and the random network, SIF is almost same with time after the attack. This means that the percentage of nodes not functioning at any time is almost equal for the two networks. But at the same time the size of the main cluster is far less in case of the BA scale free

network, explaining that the size of the main cluster depends on few key nodes for these networks.

In conclusion, we can infer that the cascading effect of an attack is high in case of hierarchical and BA scale free networks than the random networks. This is attributed to their non-homogeneous degree distribution. These results emphasize that for real networks, since they have a non-homogeneous degree distribution, a single potential attack on a key node can lead to major breakdown.

6.2.2 CIF

The Component Impact Factor (CIF) of any node specifies the effect of the attack on that specific node. If there is an attack on node 'x', then the effect of this attack at any time on node 'y' can be obtained by the CIF of the node 'y' at that specific time. As an example, Figure 6.5 shows the CIF of nodes '69', '92', '75' respectively when the node with the highest load is attacked in ER network.

Fig 6.5 CIF values for three nodes due to an attack on node with highest load

From Figure 6.5, the CIF takes both positive as well as negative values. The negative value means that the recalculated load is in fact less than the initial node and vice versa. For example node '92', at time $t = 1$, the CIF is -5 which means that the recalculated load is less than the initial load at time $t = 1$. But for the same node, at time $t = 2$ the CIF is almost equal to 5, which means that the recalculated load at time $t = 2$ is higher than the

initial load. This is an effect of some additional number of nodes malfunctioning from time $t = 1$ to time $t = 2$.

The significance of the CIF is that in combination with SIF, system size, and efficiency it can perform real time monitoring and track the effects of an attack. This can form the basis for studying existing critical infrastructure performance under different attack scenarios and provide critical information for planning better security.

6.3 PARTIAL ATTACKS

The partial attacks are modeled by not removing the node completely but by reducing the performance of the node. For example an attack of degree 10% is modeled by reducing the functionality (efficiency) by 10%. In a similar manner the attacks are carried for a varying degrees ranging from 10 to 100%, where 100% denotes a full attack, requiring removal of the node. The attacks are modeled on the three network topologies – Random, BA scale-free and Hierarchical network models.

6.3.1 K-Based & L-Based Attacks

The partial k-based and L-based attacks are carried on the three test networks and the results are shown in Figure 6.6

Fig 6.6 Partial K-based Attack Analysis on Hierarchical Network

Fig 6.7 Partial K-based Attack Analysis on Random Network

Fig 6.8: Partial K-based Attack Analysis on BA Scale free Network

In Figures 6.6, 6.7 and 6.8, as the degree of the attack is increased from 10% to 100%, the drop in the efficiency is increased. This result is common in all three networks. The interesting observation is the difference between the plots for 10% and 100% in all the three networks. We see that the difference between 10% and 100% for random networks is not as high as in case of BA scale free or hierarchical networks.

For explaining the above results in a better manner consider the Figures 6.9 and 6.10 that show the plots from 10% and 100% attacks for the three network models and Figures 6.11 and 6.12 that show the efficiency graphs for the initial attacks in both the 10% and 100% attacks.

Fig 6.9 Comparison of efficiency for full and partial attacks in Hierarchical and Random networks

Fig 6.10 Comparison of efficiency for full and partial attacks in BA Scale free and Random networks

Fig 6.11 Attack analysis for full degree K-based attack

Fig 6.12 Attack analysis for partial attack of 10% K-based attack

From Figures 6.9 and 6.10 we see that the performance of the BA Scale free network and the Hierarchical networks is considerably improved, especially for the first few attacks in case of the reduced degree of attack (10%). This result can be clearly seen from Figures 6.11 and 6.12, which show that the performance of three networks is almost similar for reduced degree 10% initial attacks.

The results are also shown in Tables 6.1 and 6.2, showing the change in efficiency values for the initial attacks.

Table 6.1 Efficiency for 100% degree of attack

Time	Hierarchical	Random	BA-Scale free
0	0.42	0.38	0.39
1	0.09	0.37	0.38
2	0.08	0.36	0.36
3	0.06	0.35	0.33
4	0.05	0.34	0.31
5	0.03	0.34	0.29

Table 6.2 Efficiency for 10% degree of attack

Time	Hierarchical	Random	BA-Scale free
0	0.42	0.38	0.39
1	0.40	0.38	0.39
2	0.37	0.38	0.39
3	0.35	0.38	0.38
4	0.32	0.38	0.38
5	0.30	0.38	0.38

The values show the improvement of efficiency for the BA scale free and hierarchical networks for reduced attack when compared to the random networks. This behavior can be attributed to the topology (i.e. the degree distribution) of the network models. As the random networks are homogeneous and has a Poisson degree distribution with all the nodes having almost equal degree, the drop in the efficiency is not very prominent from 10% to 100% attacks. But whereas for the BA scale-free network and the hierarchical network the drop is significant because they have a power-law degree distribution.

The conclusion we draw from these observations is that in case of the scale-free and the hierarchical networks, if we can identify key nodes and protect them properly, thereby reducing the effect of an attack on these nodes, the vulnerability of these networks to attacks can be reduced considerably.

Note: Similar types of results are obtained for the L-based attacks and the plots are shown in Appendix A.

6.3.2 Cascade Attack

The cascading failures are common in most of the complex networks. When a node is attacked, the attack leads to a network breakdown due to the cascading effect. The capacity of node is proportional to the initial load of the node and is given by equation.

$$C_j = \alpha L_j(0), \quad j = 1, 2, \dots, N,$$

where $\alpha \geq 1$ is the tolerance parameter of the network

and $L_j(0)$ is the load at time 0, initial load.

The node with the highest initial load is attacked. The cascading effect of different degree of attacks is studied for the three network models – Random, BA scale-free and Hierarchical network models. The results presented below are for a low value of $\alpha = 1.05$. The capacity is dependent on the value of α . The higher the value of α , the higher is the capacity. The plots are final values of efficiency, size and SIF obtained after either the network stabilizes or is broken down to small parts.

Fig 6.13 Variation of Efficiency with variation of degree of attack

Fig 6.14 Variation of System Size with variation of degree of attack

Fig 6.15 Variation of System Impact Factor with variation of degree of attack

From Figures 6.13 we see that for very low values of $\alpha = 1.05$, the efficiency drops rapidly for the BA Scale free and the Random networks, whereas it drops slowly in case of the Hierarchical networks. The cascade effect is low for hierarchical networks compared to the scale free or random networks in case of reduced degree attack (10%-80%) even for a small value of $\alpha = 1.05$. The capacity of nodes and thus the effect of a cascading attack is dependent on the value of α . An extremely large α can reduce the cascade effect to zero, but is not practically possible due to the limitations of cost.

Similarly from Figure 6.14 we see that the size of the main cluster also drops rapidly in case of the BA and Scale free networks. Whereas in case of the Hierarchical networks the size of the main cluster is maintained from 10-90% of degree of attacks and rapidly drops for 100% attack. This explains that the network is totally dependent on a single node. If this node is well protected, the system can be kept largely connected.

The conclusion we draw from the above results is that if we can identify the key nodes in real networks and protect them, reducing the effect of an attack, the cascading effect can be considerably reduced. The major factor is that it can be obtained at a very small α value.

6.3.3 Effect of Alpha

The cascade effect of an attack directly depends on the value of alpha, since the capacity of nodes is determined by the value of alpha. Though an extremely large value of alpha is not possible to select, due to the limitations imposed by cost, we have tried to summarize the effect of alpha on the cascading effect in the three network topologies.

So far all the results obtained and discussed are for a value of alpha of 1.05. The higher the value alpha the smaller is the cascading effect of an attack. The variation of the system size and the efficiency of the three networks with increasing alpha is plotted. The values plotted are the final values, the values at which the system stabilizes due to the cascading effect of the attack. Figures 6.16 and 6.17 for the normal or full degree attacks (all edges of the node removed) and Figures 6.18 and 6.19 are for reduced 10% attacks.

Fig 6.16 Variation of System Size with Alpha under Full degree attack

Fig 6.17 Variation of Efficiency with Alpha under Full degree attack

Fig 6.18 Variation of System Size with Alpha under 10% degree attack

Fig 6.19 Variation of Efficiency with Alpha under 10% degree attack

From Figures 6.16 and 6.17, the system size and the efficiency improve with the increasing value of alpha in case of the Random and BA scale free networks. For the random networks the cascading effect is negligible for a value of alpha 1.2, whereas for the BA scale free network models it is higher at a value of almost 1.5. This can be attributed to different network topologies of the network models.

In case of the hierarchical networks, from Figures 6.16 and 6.17, the size as well as the efficiency drop to a very low value and are constant with increasing value of alpha. This is because the network is dependent on the key nodes and once the attack damages these key nodes the network is broken down. Whereas from Figures 6.18 and 6.19, we see that for reduced attacks, the size and the efficiency is improved significantly and is a constant as the alpha value increases.

We conclude that if we identify and protect the key node in the hierarchical network model, the cascading effect can be reduced considerably. Also the network can be well protected even for small values of α .

CHAPTER VII: CONCLUSIONS AND FUTURE WORK

From the results obtained and discussions above, it is clearly shown that the error and attack tolerance of a network can be attributed to the topology of the network. Further it is shown that the scale free networks are vulnerable to attacks – both under the static attacks as well as the cascading effect attacks. The hierarchical networks that have both the power law degree distribution as well as the high modularity characteristics of real networks are highly vulnerable to attacks, since they depend on few key nodes. These results clearly show the need to carefully design the protection measures for real networks. Both the hierarchical as well as the BA scale free networks show much improved performance under reduced degree of attacks and are comparable to that of the random networks under such conditions. Hence if we are able to protect the key nodes in these networks the performance of the networks can be highly improved. In particular the cascading effect of an attack can be considerably reduced by protecting the key nodes in the hierarchical networks. The effect of an attack on the system as well as other nodes (components) is quantified using the impact factors.

The impact factors and different attack models can be used to test and identify the effect of different attacks on existing real scenario networks such as the western power grid, transport systems, Internet, WWW etc. An analysis of this type can help plan for critical malfunctions that may occur in real networks. Since the advent of the 9-11 attack there is considerable emphasis on protecting the national critical infrastructures. A network analysis of this type can be the basis for assessing and designing robust network systems.

BIBLIOGRAPHY

1. Albert, R. and Barabasi, A. L. (2002), 'Statistical mechanics of complex networks,' *Review of Modern Physics*, Vol. 74, pp. 47-97.
2. Albert, R., Joeng H., and Barabasi, A. L. (2000), 'Error and Attack Tolerance of Complex Networks,' *Nature*, Vol. 406, pp. 378-382.
3. Albert, R., Jeong, H. and Barabasi, A. L. (1999), 'Diameter of the World Wide Web,' *Nature*, Vol. 401, pp. 130-131.
4. Barabasi, A.-L., Albert, R., and Jeong, H. (2000), 'Scale-free characteristics of random networks: the topology of world-wide web,' *Physica A*, Vol. 281, pp. 69-77.
5. Barabasi, A. L., Jeong, H., Ravasz, R., Neda, Z., Vicsek, T. and Schubert, A. (2002), 'Evolution of the social network of scientific collaborations,' *Physica A*, Vol. 311, pp. 590-614.
6. Crucitti, P., Latora, V., Marichiori, M. and Rapisarda, A., (2004), 'Error and attack tolerance of complex networks,' *Physica A*, Vol. 340, pp. 388-394.
7. Crucitti, P., Latora, V., Marichiori M. and Rapisarda, A., (2003), 'Efficiency of scale-free networks: error and attack tolerance,' *Physica A*, Vol. 320, pp. 622-642.
8. Dong, Y., Hou, Y. and Zhang, Z.-L. (1999), 'A Server-Based Performance Measurement Tool within Enterprise Networks,' *Journal of Performance Evaluation*, Vol. 36-37, pp. 233-247.

9. Dorogovtsev, S.N. and Mendes, J.F.F. (2002), 'Evolution of Networks,' *Advanced Physics*, Vol. 51, pp. 1079-1187.
10. Dorogovtsev, S. N., Goltsev, A.V. and Mendes, J.F.F (2002), 'Pseudofractal scale-free web,' *Physical Review*, Vol. 65, 066122.
11. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. and Barabasi, A. L. (2000), 'The large-scale organization of metabolic networks,' *Nature*, Vol. 407, pp. 651-654.
12. Jeong, H., Mason, S. P., Barabasi, A. L. and Oltvai, Z. N. (2001), 'Lethality and centrality in protein networks,' *Nature* Vol. 411, pp. 41-42.
13. Jha S., Wing, J., Richard, C. L. and Thomas, A. L. (2000), 'Survivability Analysis of Network Specifications,' *Proc. Int'l Conf. on Dependable Systems and Networks (DSN2000)*, *IEEE Computer Security*, pp. 613–622.
14. Liebold, B. (2004), 'Exploration of effects of different network topologies on the ESN signal cross correlation matrix spectrum,' *School of Engineering & Science at International University Bremen*.
15. Motter, A.E. and Lai, Y.C. (2002) 'Cascade-based attacks on complex networks,' *Physical Review*, Vol. 66, 065102(R).
16. Newman, M. E. J. (2001) 'Scientific collaboration networks. I. Network construction and fundamental results,' *Physical Review*, Vol. 64, 016131.
17. Pablo, B. (2003), 'Erdos-Renyi Random Graph,' *MATLAB Central*,
<http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=4206&objectType=FILE>

18. Ravasz, E. and Barabasi, A. L (2003), 'Hierarchical Organization in Complex Networks,' *Physical Review*, Vol. 67, 026112.
19. Salim, H., Qu, G., Tushneem, Dharmagadda, Ramkishore M. and Raghavendra, C. S. (2003), 'Impact Analysis of Faults and Attacks in Large-Scale Networks,' Security and Privacy magazine, *IEEE Security & Privacy*, Vol. 1, no. 5, pp. 49-54.
20. Swiler, L.P., Phillips, C. and Gaylor, T. (1998), 'A Graph-Based Network-Vulnerability Analysis System,' *Proc of 1998 workshop on New security paradigms*, pp. 71-79.
21. Watts, D.J. and Strogatz, S.H. (1998), 'Collective dynamics of small-world networks,' *Nature*, Vol. 393, pp. 440-442.
22. Wagner D. and Fell A. (2001), 'The small world inside large metabolic networks,' *Proceedings of Biological Sciences*. Vol. 268, pp. 1803–1810.
23. <http://www.computerworld.com/networkingtopics/networking/story/0,10801,75539,00.html>

APPENDIX A

Results of Partial L-based attacks

Fig A.1 Partial L-based Attack Analysis on Hierarchical Network

Fig A.2 Partial L-based Attack Analysis on Random Network

Fig A.3 Patial L-based Attack Analysis on BA scale free Network

Fig A.4 Comparison of efficiency for full and partial attacks in Hierarchical and Random networks

Fig A.5 Comparison of efficiency for full and partial attacks in BA scale free and Random networks

Fig A.6 Attack analysis for full degree L-based attack

Fig A.7 Attack analysis for partial attack of 10% L-based attack

APPENDIX B

MATLAB code for Erdos Renyi Random Network Model [Pablo Blinder, 2003]

```
function [G]=erdosRenyi(nv,p,Kreg)
%Funciton [G]=edosRenyi(nv,p,Kreg) generates a random graph based on
%the Erdos and Renyi algoritm where all possible pairs of 'nv' nodes are
%connected with probability 'p'.
%
% Inputs:
%  nv - number of nodes
%  p  - rewiring probability
%  Kreg - initial node degree of for regular graph (use 1 or even numbers)
%
% Output:
%  G is a structure implemented as data structure in this as well as other
%  graph theory algorithms.
%  G.Adj  - is the adjacency matrix (1 for connected nodes, 0 otherwise).
%  G.x and G.y  - are row vectors of size nv wiht the (x,y) coordinates of
%                each node of G.
%  G.nv  - number of vertices in G
%  G.ne  - number of edges in G
%
%Created by Pablo Blinder. blinderp@bgu.ac.il
%
%Last update 25/01/2005

%build regular lattice

nv = input('Enter the number of the nodes:');
p = input('Enter the probability of rewiring:');
Kreg = input('Enter the initial node degree:');

A=sparse(nv,nv);
Kreg=fix(abs(Kreg)/2);Kreg=(Kreg<1)+Kreg;

for k=1:Kreg
    A=sparse(A+diag(ones(1,length(diag(A,k))),k)+diag(ones(1,length(diag(A,nv-k))),nv-
k));
end
ne0=nnz(A);
%find connected pairs
[v1,v2]=find(A);
% P=permPairs(nv);%my version is faster
Dis=(rand(length(v1),1)<=p);%pairs to disconnect
A(v1(Dis),v2(Dis))=0;
```



```

vDis=unique([v1(Dis),v2(Dis)]);%disconnected vertices
nDis=ne0-nnz(A);sum(Dis);

%cycle trough disconnected pairs
disconPairs=[v1(Dis),v2(Dis)];
for n=1:nDis
    %choose one of the vertices from the disconnected pair
    i=ceil(rand*size(disconPairs,1));
    j=logical(1+rand>0.5);
    vDisToRec=disconPairs(i,j);
    %find non adjacent vertices and reconnect
    adj=[find(A(:,vDisToRec)) ; find(A(vDisToRec,:))'];
    nonAdj=setdiff(1:nv,adj);
    vToRec=nonAdj(ceil(rand*length(nonAdj)));
    S=sort([vDisToRec vToRec]);
    A(S(1),S(2))=1;
end
[x,y]=getNodeCoordinates(nv);
%make adjacency matrix symmetric
A=A+fliplr((flipud(triu(A))));
G=struct('Adj',A,'x',x,'y',y,'nv',nv,'ne',nnz(A));

csvwrite('random-test',A);

```

APPENDIX C

MATLAB code for BA Scale Free Network [Markus Kaiser, 2002].

```
%clear all
%clc

%function matrix = scalefreegraph(n,fc);
% matrix = scalefreegraph(n,fc)
% yields matrix of scale-free graph
% with n nodes
% fc is a factor determining final density
% Author : Marcus Kaiser Date : 8 . 1 2 . 0 2

n=input('Enter the number of nodes in the network: ');
fc = input('Enter the factor for final density: ');

NODES = n ;
INITIALNODES = 3;

% generate initial matrix(two nodes; undirected link)
matrix = zeros (NODES,NODES) ;
matrix ( 1 , 2 ) = 1 ;
matrix ( 2 , 1 ) = 1 ;
matrix ( 1 , 3 ) = 1 ;
matrix ( 3 , 1 ) = 1 ;
matrix ( 2 , 3 ) = 1 ;
matrix ( 3 , 2 ) = 1 ;
nodes_incl= [1;2;3] ;

% aggregation of nodes to initial matrix
m = INITIALNODES ;
k = zeros (NODES, 1 ) ;

for i = 1 :m
k(i) = (sum(matrix(i,:))+sum(matrix(:,i)));
end;

while m < NODES
m = m + 1 ;
for i = 1 :m-1
P = k(i)/sum( k );
if (rand(1,1) <= P*fc)
k(i) = k(i) + 2 ;
k(m) = k(m) + 2 ;
matrix ( i ,m) = 1 ;
```

```
matrix (m, i ) = 1 ;  
end ; % if  
end ; % for  
end ; % while m  
  
csvwrite('BA-test',matrix);  
edges = sum(sum(matrix))  
return ;
```