

SUBHENDU MONDAL

20121047

EXTREME EVENTS ON COMPLEX NETWORKS AND NETWORK ROBUSTNESS

A Thesis

submitted to

Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Subhendu Mondal



Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2017

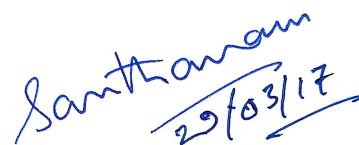
Supervisor: **Dr. M. S. Santhanam**

© **Subhendu Mondal** 2017

All rights reserved

Certificate

This is to certify that this dissertation entitled EXTREME EVENTS ON COMPLEX NETWORKS AND NETWORK ROBUSTNESS towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by **Subhendu Mondal** at Indian Institute of Science Education and Research under the supervision of **Dr. M. S. Santhanam**, Associate Professor, Department of Physics , during the academic year 2016-2017.

Handwritten signature of Dr. M. S. Santhanam in blue ink, with the date 29/03/17 written below it.

Dr. M. S. Santhanam

Committee:

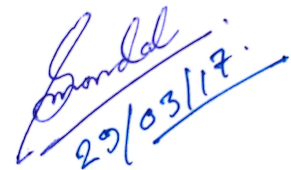
Dr. M. S. Santhanam

Dr. Apratim Chatterji

This thesis is dedicated to my parents, Subhas and Rekha Mondal, sister, Moumita Mondal.

Declaration

I hereby declare that the matter embodied in the report entitled EXTREME EVENTS ON COMPLEX NETWORKS AND NETWORK ROBUSTNESS are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, under the supervision of **Dr. M. S. Santhanam** and the same has not been submitted elsewhere for any other degree.

Handwritten signature of Subhendu Mondal and the date 29/03/17.

Subhendu Mondal

Acknowledgments

I would like to take this opportunity to express my profound gratitude and deep regard to my project supervisor **Dr. M. S. Santhanam**, for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout my project work. He gave me the opportunity to do this wonderful project on the topic extreme events on complex networks and network robustness, which will help me in doing a lot of research related to this topic, I am really thankful to him.

Abstract

We live in a modern world surrounded by networks ranging from transportation system to financial market. Network robustness is a matter of serious concern especially because a network can collapse completely due to overload failure. In this project my aim is to study overload failure of a network. Physical flow through a node is defined by load and capacity, capacity is the maximum load that a node can handle. I will model this situation using extreme events where population of walker on a node is the load. I use random walk simulation to prescribe a degree dependent capacity for each node. If a node encounters an extreme event, we will consider that situation as a node failure which causes redistribution of its load. I show that scale free networks are vulnerable against overload failures because of heterogeneous degree distribution but homogeneous networks (complete graph, Erdos-Renyi) are robust against overload failure. I will also show that an overloaded network undergoes a transition and define three different phases of network failure. Real life networks, internet, power grid has high heterogeneous distribution of loads. We will discuss a method to increase total capacity of the network.

Contents

Abstract	xi
1 Introduction	3
2 Definitions and concepts in graph theory	7
2.1 Graphs and sub graphs	7
2.2 Network structure	9
3 Random walk and Extreme events on complex networks	13
3.1 Random walk and extreme events on a graph	14
4 Extreme events and overload failure simulation	17
4.1 Network stability and Extreme events	18
4.2 Network stability and tolerance parameter	22
4.3 Network robustness and scale-free exponent	23
5 Results and Discussion	25
5.1 Results	25
5.2 Discussions	32

List of Figures

2.1	An example of a graph, G of size $N = 9$, $E = 12$ node 2,4,6,8 has degree 2, node 3,5,7,1 has degree 3, node 9 has maximum degree 4	8
2.2	Degree distribution of $G(N, P)$ random graph with $N = 5000$, $P = 0.05$ mean of best fit binomial curve is 247 and standard deviation 30.88 . The plot shows that degree distribution of a ER network indeed follows binomial distribution.	9
2.3	Scale free network degree distribution plot with $N = 1000$ (A) linear scale (B) log-log plot. I found $\gamma =$ slope of the straight line= 2.7 from (B) and y axis intersection is 1.03	11
2.4	An example of complete graph K_5 , consists of five nodes.	11
3.1	Binomial distribution for a random walkers probability of taking steps to right or left vs final position. Total step (N)=20, total 20000 iteration.	14
3.2	Extreme event probability as a function of degree for $N = 5000$, $E = 19815$, $W = 2E$, $m = 4$ and scale free network degree exponent $\gamma = 2.2$ averaged over 100 realization [Ref:5].	16
4.1	Number of surviving nodes vs $\eta =$ Capacity/load, $N = 1000$, $m = 3.5$, $W = 4000$	20
4.2	$\eta =$ Capacity/load vs Time for scale-free network, $N = 1000$, $m = 3.5$, $W = 4000$, transient time $T_t = 999$	21
4.3	$N = 100,000$ and $Q = 2.0 \times 10^6$, tolerance parameter m calculated from equation: 3.2 as a function of γ , Three lines shows the results for $\langle k \rangle = 0.5, 10$ and 15 from top to bottom.(Ref: [21] Fig.2)	24

5.1	Overload failure rate(left),Number of surviving nodes as a function of time(right), $N = 1000, W = 4000, m = 3.5$	26
5.2	Giant component size vs η , $N = 1000, W = 4000, m = 3.5$	27
5.3	Phase of Network failure, $N = 1000, W = 6000, m = 4.4$, Combined all data got from 10 realization	28
5.4	Phase of SF, Complete graph, Small world Network , $N = 1000, W = 6000, m =$ 4.4 , Combined all data got from 10 realization	29
5.5	N/N_0 vs m , $N_0=1000, W=4000$, as we increase m N/N_0 increases.	31
5.6	N/N_0 vs m , $N_0=1000, W=4000$, as we increase m N/N_0 increases.	31

Chapter 1

Introduction

Network is a collection of nodes and connections between them called edges. Most of the complex systems can be brought within the framework of network [1]. For example, infrastructure networks such as World Wide Web (www), power grid, transportation networks and social networks play an important role in our life. Many times we face disruption of services due to network collapse. For instance, power grid failures lead to large scale blackouts or too many webserver requests can substantially slow down the world wide web. At a basic level, such network failures begin as overload or extreme events in individual nodes of the network. If no remedial action is taken quickly, they can sometimes spread and affect network performance. In this thesis, we study how extreme events affect network as a whole.

A network that is robust to extreme events is not easily vulnerable to network failures. If random attack or some failure happens in a network but the network still maintain its functionality then we will say that the network is robust against a particular failure. Proper knowledge of complete network structure is required to understand and predict the functionality and vulnerability of a network, because different complex systems are completely different in their function. Network with a given degree distribution may be very robust against one type of attack but not to other. In this context, the robustness of complex network against random and targeted failure has been studied [2–7].

As far as natural breakdown of a network is concerned (not random or targeted attack) we can model many transport phenomena on complex network and vulnerability of a network using random walk and extreme events (EE), extreme event is defined as exceedence above

a prescribed threshold. Traffic jams in roads, power blackout, web server not responding due to heavy load on the server these are common example of EE. Not only in physics the concept of random walk and EE has lots of applications in finance too [8]. Each of the node in a network has a capacity similar like extreme event threshold if traffic flow or load at any instant of time exceeds that capacity the node will fail. We can describe this network failure phenomena using extreme events.

Transport phenomena and network robustness has been studied in great detail in the last several years [9, 10] but they consider only single node failure at a time due to random or intentional attack only. In this project report I will be discussing overload failure in terms of Extreme events. I will also discuss a simple model of cascading failure on complex networks.

Random walk and transport phenomena on complex network has been studied with great effort in last several years [11, 15]. But they were not concerned about extreme events which is an important phenomena to explain network robustness. Extreme event is related to the threshold of each node which is not necessarily related to real capacity of the node. It comes from the mean and fluctuation in the traffic passing through a node not through the constrained imposed by real capacity. Hence in this report I will consider nodes capacity which comes from random walk simulation (traffic flow) on the network and try to correlate network robustness with EE. Thus, it is important to estimate the probability of a node failure and EE to design a network that can sustain after node failure due to heavy load or random attack.

The complexity of a network can be defined by the connectivity between nodes, called degree. In terms of degree distribution complex network can be divided into two categories, homogeneous and heterogeneous network. Homogeneous networks degree distribution tend to follow binomial or poisson degree distribution eg. random graph, small world model, where most of the nodes degree are concentrated around the mean degree. On other hand heterogeneous network have a heavy tail degree distribution eg. Scale-Free (SF) network. In this report I will mainly focus on random walk on scale-free network because most of the large network including the internet network follows a power law degree distribution $P(k) \sim k^{-\gamma}$ where k is the degree of a node and γ is the scale-free exponent. In a transport process each node in SF network does not contribute equally because of the distribution in degree (heterogeneity). Random walk is a fundamental dynamic process which helps us to understand how the structural heterogeneity affects the traffic flow through a node. Hence

random walk simulation on a unknown network can tell us more or less about the structure (homogeneity/ heterogeneity) of the unknown network.

Robustness of complex network is another important issue in complex network research. If initial shock (failure) trigger a large portion of the network failure, called cascading failure [12]. As an example August 10, 1996 cascading failure occurred in the western United State power transmission grid. This kind of failure are extremely difficult to predict even if each component of the network are well understood. As we know that all of the node in a network has some capacity (threshold). Hence the failure of nodes can causes redistribution of load of failure nodes to the rest of the network. This redistribution of load can make the load of some nodes exceed their capacities, which will lead to further failure and load redistribution. This process is very fast, at last the entire network will collapse. If a node has relatively small load its failure will not have major effect on the network. On the other hand if a node has high load its failure causes redistribution of huge load on the network, causes failure of significant portion of the network and possibly start a sequence of overload failure. I will discuss network failure in terms of extreme events, so if load on a node exceeds a prescribed threshold we will call it extreme event which is equivalent to node failure and I will deactivate this node from the system.

Chapter 2

Definitions and concepts in graph theory

Before starting further discussion it is important to discuss some definitions and concepts in graph theory which will be used throughout the rest of this project report.

2.1 Graphs and sub graphs

A undirected graph G (Fig. 2.1) is defined by a pair of sets $G = (n, e)$, where n is a set of non empty elements called nodes and e is set set of unordered pairs of different nodes, called edges.

Number of links connected to a node is called degree (k) and if two nodes i, j are connected together by an edge (i, j) we call those nodes are nearest neighbor or neighbor. Hence number of neighbor of a node is equal to the coordination number or degree of that particular node. In the context of complex network we identify the size of of a network by total number of nodes (N) in the network. So number of elements in set n is called the size of the network, and number of elements in set e is same as the number of edges in the network. In Fig. 2.1, the network consists of 9 nodes and 12 edges node 9 has maximum degree(=4). Nodes with large connections or links are called hub, in the mentioned network we can consider node 9 as a hub. I will consider only undirected graph throughout the rest of this report.

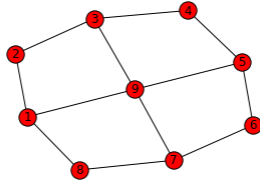


Figure 2.1: An example of a graph, G of size $N = 9$, $E = 12$ node 2,4,6,8 has degree 2, node 3,5,7,1 has degree 3, node 9 has maximum degree 4

In many cases we are also interested in the sub graph of a graph. If $G' = (n', e')$ is a sub graph of the graph $G = (n, e)$, then all the nodes in n' belong to n and all the edges in e' belong to e i.e. $e' \subset e$ and $n' \subset n$. Generally we define a graph as $G(N, E)$ to indicate the size N of the graph and number of edges E in the graph. In a mathematical point of view we can represent a graph by a matrix called adjacency matrix A . The elements of the matrix indicate whether pairs of nodes are connected or not in the graph. Element of the matrix is defined such that

$$A_{ij} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{otherwise} \end{cases} \quad (2.1)$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Adjacency matrix for the network in Figure:1

degree of i^{th} node is $\sum_j A_{ij}$. For undirected graphs the adjacency matrix is symmetric, $A_{ij} = A_{ji}$. Suppose two graphs G_1 and G_2 with adjacency matrices A_1 and A_2 are given then G_1 and G_2 are isomorphic iff there exists a permutation matrix P such that $PA_1P^{-1} = A_2$. That means A_1 and A_2 are similar graph with same eigenvalues of A_1 and A_2 .

2.2 Network structure

In this subsection I present a review of different types of network structure depending on their degree distribution.

2.2.1 Erdős Renyi model

Erdős Renyi (ER) model is used to generate random graphs with a given probability of the connection between two nodes. There are two closely related variant of Erdős Renyi model $G(N, E)$ and $G(N, P)$ [13].

In $G(N, E)$ model a graph is chosen randomly from a collection of graphs which consists of N nodes and E edges with equal probability. As an example $G(3, 2)$ implies this network is randomly chosen from a collection of graphs which consists of 3 nodes and 2 edges is included with probability $1/3$.

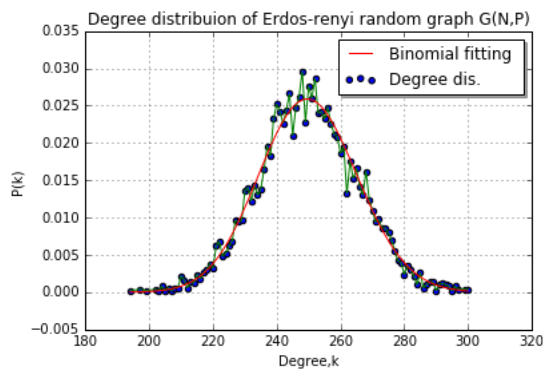


Figure 2.2: Degree distribution of $G(N, P)$ random graph with $N = 5000, P = 0.05$ mean of best fit binomial curve is 247 and standard deviation 30.88 . The plot shows that degree distribution of a ER network indeed follows binomial distribution.

$G(N, P)$ is a random graph consists of N nodes and each possible edge has probability P of existing. We have ${}^N C_2$ ways to connect any two nodes out of N nodes, so the number of edges in a $G(N, P)$ graph is a random variable with expected value ${}^N C_2 P$. There are ${}^N C_k$ ways to choose k nodes out of N nodes and P^k probability of having edge of all nodes. Then ${}^N C_k P^k$ is the probability of node has degree k . Hence there will be $(N - k)$ nodes with no edges with probability $(1 - P)^{(N-k)}$, so the degree distribution is

$$P(k) = {}^N C_k P^k (1 - P)^{(N-k)}$$

which is a binomial distribution. Fig. 2.2 shows the degree distribution of $G(N, P)$ graph (using *Networkx*, Python programming). It can be shown that the expectation value for the node degree of a $G(N, P)$ graph is $P(N - 1)$. In other word our best guess at what the node degree of a arbitrarily chosen node from an $G(N, P)$ graph is equal to $(N - 1)P$.

2.2.2 Scale-free Networks(SF)

Scale-free network is a graph whose degree distribution follows power law $P(k) \sim k^{-\gamma}$ where k is the degree of a node and γ is the scale-free exponent. Many real networks are (approximately) scale-free e.g. social networks, internet, airline networks etc are examples of scale-free network. Scale free network mainly characterized by a highly heterogeneous degree distribution. It has a long tail, it contains some nodes with very high degrees often those nodes are called hubs. Scale free property is strongly correlated with robustness and network failure. It is observed that major hubs are surrounded by smaller one and small degree nodes are surrounded by even smaller nodes. This kind of structure is highly robust against random attack but fragile under targeted attack on nodes with high degrees [2–7].

Barabasi and Albert (BA) were the first to develop an algorithm to construct a scale-free network, which works based on preferential attachment model [1]. The BA algorithm combines the attachment of a new node to existing nodes with certain preferences as follows—

Consider a relatively small ER network with N_0 nodes and E_0 edges. We want to make a SF graph of size N . At each step $S \geq 1$

- Add a new node, u_{new} to the existing network (u_{new} is the new node index).

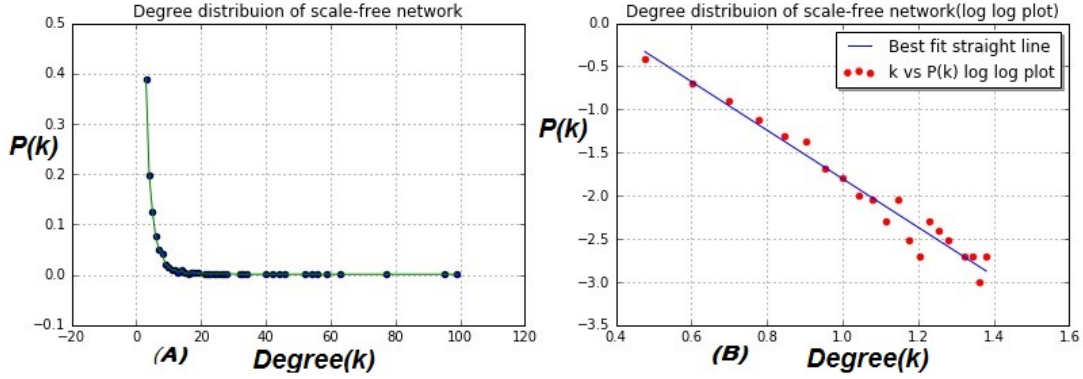


Figure 2.3: Scale free network degree distribution plot with $N=1000$ (A) linear scale (B) log-log plot. I found γ =slope of the straight line=2.7 from (B) and y axis intersection is 1.03

- Add $E_{new} < E_0$ edges to the network which incident from u_{new} node to existing node u with probability $P(E_{new} \rightarrow u)$

$$P(E_{new} \rightarrow u) = k_u / \sum_i k_i$$

i is the existing node index and k_i is the degree of node i .

- If the size of the network become N then stop the process, else repeat first two steps.

From 2nd point it is clear that the added nodes has higher probability to connect with an existing node which has higher degrees that implies BA algorithm works based on preferential attachment model. I generated a SF network using Barabasi Albert algorithm Fig. 2.3 shows its degree distribution indeed follows power law. I also calculated $\gamma = 2.7$ and proportionality constant = 1.01 from plot (B).

2.2.3 Complete graph

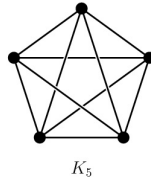


Figure 2.4: An example of complete graph K_5 , consists of five nodes.

Complete graph K_N is a simple undirected graph with N nodes and any pair of node is connected by one distinct edge. Each node is connected to $(N-1)$ number of neighbors so K_N has ${}^N C_2$ number of edges and its degree distribution is a delta function $P(k) = \delta(k - (N-1))$. K_N has the maximum number of edges among all possible graphs consist of N number of nodes.

Chapter 3

Random walk and Extreme events on complex networks

Random walk is a stochastic process which describe a path which consists of succession of random steps [14]. There are many real phenomena which can be modeled by random walk e.g. path trace by a molecule in gas, price of a fluctuating stock, transport phenomena on a network etc. Consider a 1D array of equally spaced nodes/ site centered at origin. If a drunk started walking from origin he can take either right step or left step with same probability. After each step the process renewed. The walker starts walking from the origin so if we plot the probability of taking left/right steps vs final position of the walker we will get a binomial distribution (Fig. 3.1) with mean zero. After total N steps, each of length l (our case $l = 1$) the walker located at $x = ml$ where m is any integer between $-N$ to N . So the probability of finding the particle at position $x = ml$ is a Binomial distribution given by

$$P_N(m) = \frac{N!}{[(N+m)/2]![(N-m)/2]!} (1/2)^N. \quad (3.1)$$

We can represent this process in terms of random variables, suppose $R = (R_1, R_2, \dots, R_N)$ is a sequence of random variables each of R_i can take value $+1$ or -1 with same probability $(1/2)$. We can construct another sequence $X = (X_1, X_2, \dots)$ such that $X_n = \sum_i^N u_i$, each u_i represents a single jump and X_n represents total displacement.

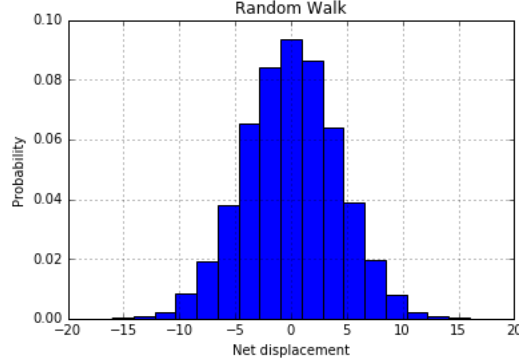


Figure 3.1: Binomial distribution for a random walkers probability of taking steps to right or left vs final position. Total step (N)=20, total 20000 iteration.

3.1 Random walk and extreme events on a graph

We have just discussed random walk in 1D, in this section we will discuss random walk on a graph which is more complicated than 1D case. Let $G(N, E)$ is a graph with N nodes and E edges, connection between two nodes are described by adjacency matrix A_{ij} . Random walk on a graph implies we start walking from a node and choose its neighbor randomly. After reaching its neighbor the process is renewed. Hence the sequence of the nodes are chosen in this process are random variables. In a mathematical way, if at time t_0 the walker on a node V_0 and after taking t steps it reaches at a node V_t , then the walker move to the neighboring node with probability $1/k_t$ where k_t is the degree of V_t node.

Our next aim is to find the transition probability of a walker from one node to another in a time interval. Assume a walker at node i at time $t = 0$ and at time t the walker selects one of its k_i neighboring node randomly with same probability as I mentioned in the first paragraph. At time $(t + 1)$ the walker reaches node j so the transition probability from node i to j is A_{ij}/k_i where k_i is the degree of node i . Now we can write a master equation which gives us the transition probability from node i to j in t time steps—

$$P_{ij}(t + 1) = \sum_n \frac{A_{nj}}{k_n} P_{in}(t) \quad (3.2)$$

In equation 3.2 n is the node index. It can be shown that the largest eigenvalue of the corresponding time evolution operator is 1 corresponding to a stationary distribution [15] and it turns out to be

$$\lim_{t \rightarrow \infty} P_{ij}(t) = P_j = \frac{k_j}{2E} \quad (3.3)$$

From equation 3.3 its clear that the probability of visiting at a given node j is proportional to the degree of the node k_j , nodes with high degrees will be visited by the random walker more often.

Now we need to get an expression for the distribution $f(w)$ of number of walker on a given node having degree k . Lets assume there are total W walkers on the network and probability of finding w walkers on a node with degree k is p . The walkers on the network are non interacting so finding w walkers on a node with degree k is p^w and other $(W - w)$ walkers are spread on the rest of the network. The probability distribution $f(w)$ turns out to be a binomial distribution

$$f(w) = \binom{W}{w} p^w (1 - p)^{(W-w)} \quad (3.4)$$

Mean number of walkers $\langle f \rangle$ and variance σ^2 can be written as [16]

$$\langle f \rangle = \frac{Wk}{2E} \quad \text{and} \quad \sigma^2 = \frac{Wk}{2E} \left(1 - \frac{k}{2E}\right) \quad (3.5)$$

The quantity $k/2E \ll 1$, from equation 3.5 it is clear that mean $\langle f \rangle$ and standard deviation satisfies the relation σ is $\sigma \sim \langle f \rangle^{1/2}$. Extreme events defined as exceedences of the number of walker at a node above a prescribel threshold or capacity. The threshold q_i for node i can not be a constant irrespective of the degree of a node because as we have seen before that flux through a node is an increasing function of its degree. If we set a constant threshold some node will be prone to EE but others would not encounter any EE. To avoid this problem we define a degree dependent threshold $q_i = \langle f \rangle_i + m\sigma_i$ of node i where m is any positive real number . In extreme value statistics probability of occurance of an extreme

event is very small. Once we have that threshold quantity for each node then we can define an extreme event as the exceedence of number of walker on a node above the threshold value. The probability distribution of extreme event can be obtained as [16]

$$F(k) = \sum_{k=\lfloor q \rfloor + 1}^W \binom{W}{k} p^k (1-p)^{(W-k)} = I_p(\lfloor q \rfloor + 1, W - \lfloor q \rfloor) \quad (3.6)$$

where $\lfloor q \rfloor$ is the floor function define as the largest integer not greater than q and $I_p(x, y)$ is the regularized incomplete beta function [18].

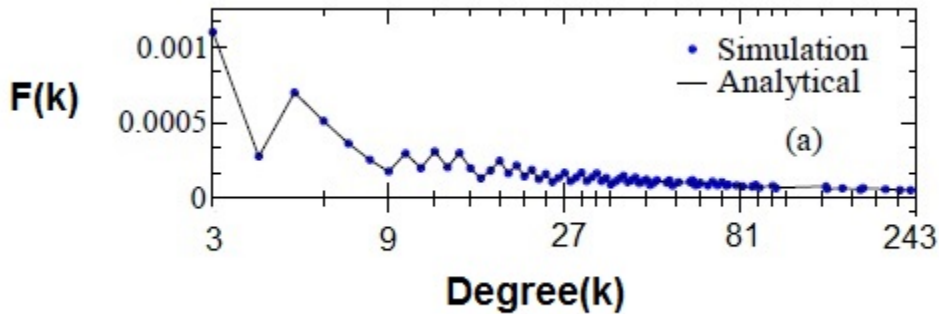


Figure 3.2: Extreme event probability as a function of degree for $N = 5000$, $E = 19815$, $W = 2E$, $m = 4$ and scale free network degree exponent $\gamma = 2.2$ averaged over 100 realization [Ref:5].

From the equatin 3.6 its clear that for a given graph $G(N, E)$ and number of walker W , the extreme event probability is depends only on the degree of a node as shown in Fig. 3.2. An important feature of this result is EE probability is higher for smaller degree nodes as compare to higher degree nodes. There is another model of EE where number of walkers are not fixed. After a overload failure total load W on a network increases as time progress. In this kind of situation when load on a network is greater than initial load on the network, the extreme event probability is an increasing function of degree k , which implies that high degree nodes are more likely to experience overload failure as compare to low degree nodes [16,17]. Although in our case load on the network will be fixed.

Chapter 4

Extreme events and overload failure simulation

In this chapter I am going to discuss about the simulation procedure and numerical results. I used a network generator module called *Networkx* which is available in Python programming language and for simulation and data analysis purpose I used Python programming language. We will mainly focus on scale free network because many real world networks have been reported to be scale-free.

To generate a SF network I used Barabasi-Albert algorithm which. In *Networkx* there is a function *barabasi-albert-graph*($n, m, seed = None$) which is used to generate a scale-free network where n is the number of node in the network, m is number of edges to attach from a new node to existing nodes. If we put $seed = none$ it will give us different graphs with different configuration in other word it will generate different graphs with different adjacency matrix in different trials. But if we put $seed = i$ where i is any integer then each time it will produce a graph with same adjacency matrix.

As I mentioned before that we will discuss network robustness in terms of extreme events. After performing a random walk simulation on a given network we will get a list of threshold value for different nodes which depends on the degree of a node. This threshold is equivalent to the capacity of a node. If walker/ traffic flow through a node exceeds that capacity the node will be damaged and redistribute its load to its neighbors. Redistribution of the load of a failed node can trigger another failure and that process can trigger a cascading failure.

Cascading failure is very difficult to control and which causes sudden collapse of a network.

4.1 Network stability and Extreme events

Complex networks play an important role in modern human society. Internet, electrical power grid, rail transport, spread of a viral disease these are some example of complex networks. There are several approach to understand the transport phenomena through a network but we used random walk approach to study network dynamics. Our main aim is to explain network robustness, we can define network load by the population of the walkers and redistribute them if a node gets overloaded which is similar to real life traffic flow.

A network collapse occur mainly because of three reason random attack, intentional attack, failure due to heavy load. First two phenomena have been studied by many research group [2–7] but the overload failure and cascading failure process still not well understood. During my project my primary goal was to understand network failure and cascading failure in terms of extreme events.

The network robustness has been studied in the context of percolation problem , mainly random failure and targeted attack studied [19,20]. There fore it is important to understand network robustness against overload failure in terms of extreme events. In this context, some over load nodes are simultaneously removed from a network and the network tend to collapse.

4.1.1 Random walk and Extreme events simulation

To do a random walk simulation I generated some networks (SF, ER, complete graph etc.) using *Networkx* package. Once a network is prepared then the degree distribution, number of edges (E), adjacency matrix (A), degree of each node (k) everything is known to us. As the name random walk suggest there has to be certain number of walker W on the network. Initially ($t = 0$) I choose $W = 2E$ (this choice is only for SF) number of walkers randomly distributed on the network. At each unit time interval each walker makes a jump to its neighboring node randomly. Suppose a walker is sitting on a node i at time t at time $t + 1$ it make a jump and reaches one of the neighbor of i^{th} node at random so probability of reaching

one of its neighbor is $1/k_i$ where k_i is the degree of i^{th} node. Each time step t each node are occupied with certain number of walkers (it could be zero also) the number of walker on a node at any instant of time t called the load on that node L_i and if we take the average load with respect to time for each node its called mean flux $\langle f \rangle_i$ through a node, similarly standard deviation of walker flow through a node is called flux fluctuation σ_i . If we monitor the flux through each node at each time interval at the end of the simulation we would be able to calculate the quantity $\langle f \rangle_i$ and σ_i . Once we know mean and standard deviation its easy to calculate can the capacity or threshold $q_i = \langle f \rangle_i + m\sigma_i$ of each node where m is called tolerance parameter which has significant role in network stability. This simulated results agrees with analytical results given in equation 3.5 . Mean flux and flux fluctuation these are function of degree k_i of a node i so first I did one random walk simulation to know $\langle f \rangle_i$, σ_i and q_i . Once we know capacity q_i for each node we are ready for second simulation for extreme event study because second random walk on same network does not change q_i because capacity depends only on degree of a node.

For extreme event investigation we have to do second random walk simulation. The random walkers are hopping from one node to nearest neighbor at random as we discussed before. An important thing about this simulation is I did not consider any EE in first few time interval, first few time steps are required for thermalisation of the walkers in other word first few time steps are needed to mix the walkers on the network, this time interval is called transient time T_t . Now we prescribed the threshold q_i for each node which I got from first random walk simulation. Now we have to keep track of the number of walker on a node at each instant of time, if the number of walker on a node exceeds the prescribed threshold we would consider that as extreme event. At the end of the simulation if we calculate average extreme events at a node and then add the mean number of EE of same degree nodes and divide by the number of nodes with that degree in the network we would get extreme events probability $F(k)$ as a function of degree k as shown in Fig. 3.2 and analytical expression of $F(k)$ is given in equation 3.6 .

4.1.2 Extreme events and network failure

From extreme event simulation on network its clear that we can assign a particular capacity (equivalent to threshold of a node in EE) for each node. If number of walkers or load exceed that capacity we call that extreme events which is similar to network failure because due

to overload failure load on a node exceeds its capacity. Therefore EE is a great concept to study network failure. We can control a network robustness against overload failure by changing two parameters first one is tolerance parameter m which is purely internal property of nodes that is not a network configuration property and another one is γ in scale free network ($P(k) \sim k^{-\gamma}$) which is a property of network configuration. That means we can make a scale free network robust against overload failure by changing network configuration or by changing tolerance parameter m . First we will discuss how tolerance parameter related to network robustness then we will discuss about γ parameter. Under random attack scale free network as well as ER network both are robust. In random failure each time step only one node gets removed from the network at random. We define a quantity $\eta = \sum_i q_i / W = Q / W$ where i is the surviving node index and $Q = \sum_i q_i$ called total capacity of the network. As we remove node from the network total tolerance will decrease because we are adding threshold of surviving nodes to get Q , but we are distributing the load of collapsed node on the network so total load W of the network is a constant quantity.

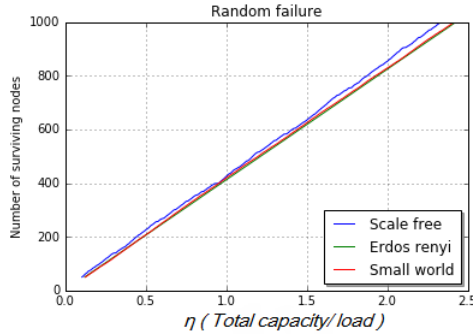


Figure 4.1: Number of surviving nodes vs $\eta = \text{Capacity}/\text{load}$, $N = 1000$, $m = 3.5$, $W = 4000$

Fig. 4.1 is number of deleted nodes vs η plot under random failure. It shows that number of surviving nodes vs η plot follows almost same pattern for SF, ER and small world. Hence SF and ER network both of them are equally robust under random attack that is why there is not much deviation between SF and ER network plot. Another kind of network is also there, small world network in that case network stability depends on the number of edges also in our case small world network consists of $N = 1000$ and $E = 1240$ and that plot is also linear in nature. I will mostly discuss about SF and ER networks although next few plots are also contains small world networks.

In Fig. 4.1 the failure occurred due to random failure but we want to study failure due to overload. As time progress number of nodes decreases so η will also decrease for both random

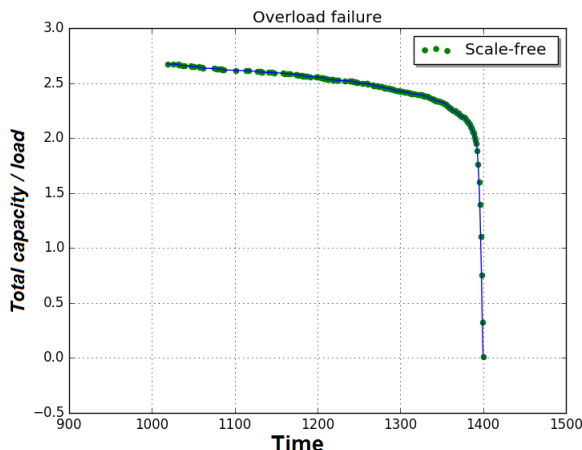


Figure 4.2: η =Capacity/load vs Time for scale-free network, $N = 1000$, $m = 3.5$, $W = 4000$, transient time $T_t = 999$

and overload failure. We want to see how total capacity of a scale-free network changes as a function of time against overload failure. Depending on the total load (Q) of the network, initial process of deletion is not very fast but as time progress enormous amount of load on the network causes failure of large number of node. Fig. 4.2 indicate the same behavior as we expected. Initially failure rate was very slow hence η was large as time progress deletion rate increases because load on the network is constant. Eventually failure rate speeds up and η decreases drastically.

4.1.3 Simulation for Network failure due to overload

As I mentioned before we can substitute the technical term failure by the mathematical term extreme events. So network failure due to overload simulation is the continuation of EE simulation. If there is any EE at a node we have to deactivate those nodes which is equivalent to deleting those nodes. In real network after deactivation of a node (power grid, internet network etc.) causes redistribution of the deactivated nodes load. After removing those nodes I redistributed the walkers which were on that node to its neighbors. During this process its quite possible to find some isolated nodes also I did not remove those isolated nodes unless there is an EE. If EE occur on a isolated nodes there is no neighbors to to take that extra load in this situation I redistributed load of that isolated nodes on the remaining network at random to make total load W of the network constant. We defined η in the last

section to study different phase of over load failure called random failure, overload failure and cascading failure I will explain these three phases later.

This simulation is a deletion process, so if we start with a large connected graph after some time the graph will be divided into several components and isolated nodes but we measure the robustness in terms of the size of the largest connected component. Common terminology for large component is giant connected component. As time progress size of the giant component will decrease, we consider the network to be robust till the point when the size of the giant component is comparable with the original network size, after that point network will collapse very fast. The module *Networkx* does not give the giant component size directly, to get the giant component size I calculated size of all subgraphs in the remaining network and take maximum size out of them.

4.2 Network stability and tolerance parameter

For a given network each time steps information or relevant quantity is exchanged between two nodes, we modeled that transport phenomena in terms of random walk. Number of walkers on a node at a particular time is the load on that node and each node has a prescribed capacity $q_i = \langle f \rangle_i + m\sigma_i$. If we want to make a network more robust against overload failure we have to increase the total capacity of the network. For a given network and total load (total number of walkers) mean flux and flux fluctuation through a node are a fixed quantity. There is only one way to increase the capacity q_i that is by changing m , m is called tolerance parameter which decides to what extent the network can sustain overload. Any failure leads to redistribution of load as a result subsequent failure can occur. This step by step process called *cascading failure*. In the case of node failure the damage causes by a failure is quantified in terms of the relative size of the giant component N/N_0 where N is the size of giant component and N_0 is the initial network size. In Results and discussions chapter I will explain the network robustness as a function of tolerance parameter. Changing m does not affect the network topology which we can change by hand. In physical network changing m is equivalent to just changing the capacity of each electrical transformer or internet server etc. But we need to change something in the network configuration itself which will increase total capacity of the network.

4.3 Network robustness and scale-free exponent

In previous section we found that by changing the network configuration we can make a network more robust. Now one question appear automatically that how can we relate network robustness with a configuration parameter.

Most of the networks demands to be scale-free network so I will consider only SF. In this section we will show that total network tolerance or total capacity Q is an increasing function of γ where γ is the scale free exponent which satisfies the relation $P(k) \sim k^{-\gamma}$ which is SF degree distribution. Our goal is to make a network more tolerant by changing γ but keeping Q fixed.

The total tolerance of a network with degree distribution $P(k)$ is given by

$$Q = N \sum_{k=1}^{k_{max}} q_k P(k) \quad (4.1)$$

If we wanto to change γ but keep total tolerance (Q) constant then we have to change m' according to the following equation [21]

$$m' = \frac{Q - W}{N \sum_k \sigma_k P(k)} \quad (4.2)$$

We defined m as tolerance parameter in equation 4.2. However, m' also has the same physical significance though m and m' are not identical quantities. m is a quantity which we can change by hand but m' is a quantity which is fixed for given values of N , Q , W , k_i and $P(k)$. We can see that m' is a function of γ through $P(k)$. If we keep N , Q , W , k_i and $P(k)$ fixed then Fig 4.3 suggests that m' is a decreasing function of γ . So if we increase γ but keep m' fixed that will increase total tolerance Q . By increasing the exponent γ we can make a SF more robust against overload failure.

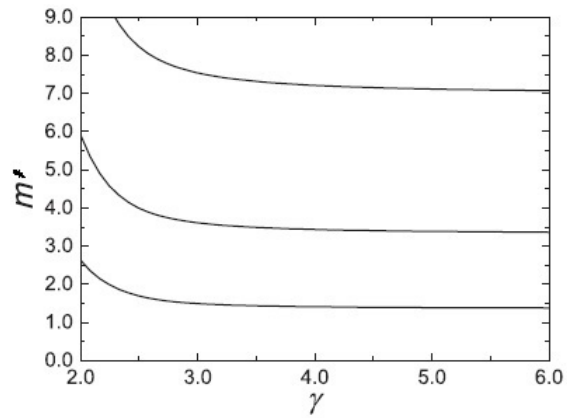


Figure 4.3: $N = 100,000$ and $Q = 2.0 \times 10^6$, tolerance parameter m calculated from equation: 3.2 as a function of γ , Three lines shows the results for $\langle k \rangle = 0.5, 10$ and 15 from top to bottom. (Ref: [21] Fig.2)

Chapter 5

Results and Discussion

Till this point we discussed simulation and theory behind the results. In this chapter I will explain the simulated results and try to correlate them with some examples.

5.1 Results

A network mainly collapse due to three reasons (1) Random attack (2) Intentional attack (3) Overload failure. Network stability against random attack and intentional attack have been studied in detail [19,20] although these network robustness against these two attacks are not completely known. On the other side we have overload failure which is not well understood. My aim is to investigate overload failure of a network. There are many models to understand transport process through network, I used random walk and EE model to study network robustness.

5.1.1 Rate of overload failure

To know about the rate of overload failure we need to have the instant of time when a failure occur and the number of node breakdown or failure happened at that time. I got those data from overload failure simulation on a network of $N = 1000, W = 4000, m = 3.5$. I wanted to understand how fast SF and ER network collapse against overload failure which is

indirect measure of network robustness. From Fig. 5.1(left) we can see that SF completely disintegrate much before ER, that implies ER is more robust than SF against overload attack. The reason behind that kind of behavior is that SF has a power law degree distribution which has long tail. Initial failure mainly occur at the node with lower degree because from Fig. 3.2 its clear that EE probability is higher for lower degree nodes. Once lower degree of nodes are deactivated then their loads are re distributed to its neighbors. One of the important property of SF is hubs are surrounded by comparatively lower degree nodes and those nodes are surrounded by even lower degree nodes. From equation 3.3 its clear that, when loads are redistributed to the neighbors then higher degree nodes has higher probability to collect extra loads from its neighbors because it has large number of links. And the known fact about SF is that it is very unstable against targeted attack. Once high degree nodes gets overloaded and fail then one failure trigger next failure and the network collapse very fast.

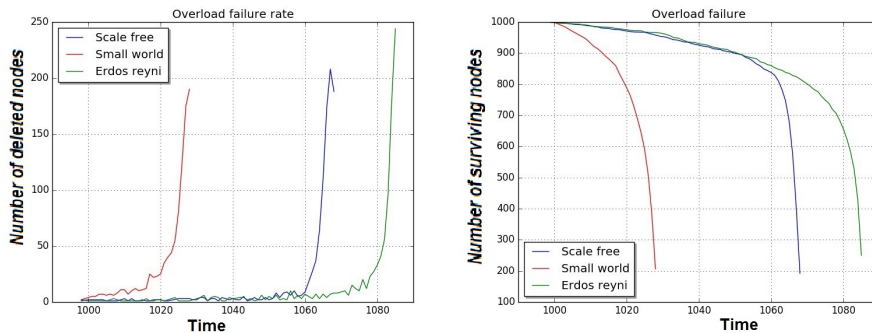


Figure 5.1: Overload failure rate(left),Number of surviving nodes as a function of time(right), $N = 1000, W = 4000, m = 3.5$

On other hand we have ER which degree distribution follows binomial distribution, that means less heterogeneous as compare to SF. Once network failure starts, the redistribution of load happen among the nodes which are not very different in terms of degree. In this network EE frequency is much less than SF.

Fig. 5.1 (right) is just the reverse plot of left side plot. It also shows that Number of surviving nodes in SF fails very fast. which implies SF can not sustain against overload failure for a long time. But ER networks surviving nodes removal happen in a slower rate which implies ER network more robust than SF against overload failure. Therefore, homogeneous networks appear to be more robust against overload failure as compare to heterogeneous networks.

5.1.2 Giant component size as a function of η

We defined a quantity $\eta = \Sigma_i q_i / W = Q / W$ and our goal is to study the giant component size as a function of η . We could have used other parameter instead of η but η is a quantity which depends on the network and other known parameters. As an example if we monitor the giant component size as a function of time then each realization we will get completely different graph because overload network failure is not a function of time, its a random process. But for a given size and configuration of network and given number of walkers we can have some idea about the transition point when network completely collapse. Hence η is a suitable parameter compare to others to study network robustness and giant network size. Another important point to choose η is that it gives the present capacity/load ratio which is a measure of how strong the overload is compare to network capacity, and giant component size vs η gives an idea to what extent of overload a network can sustain.

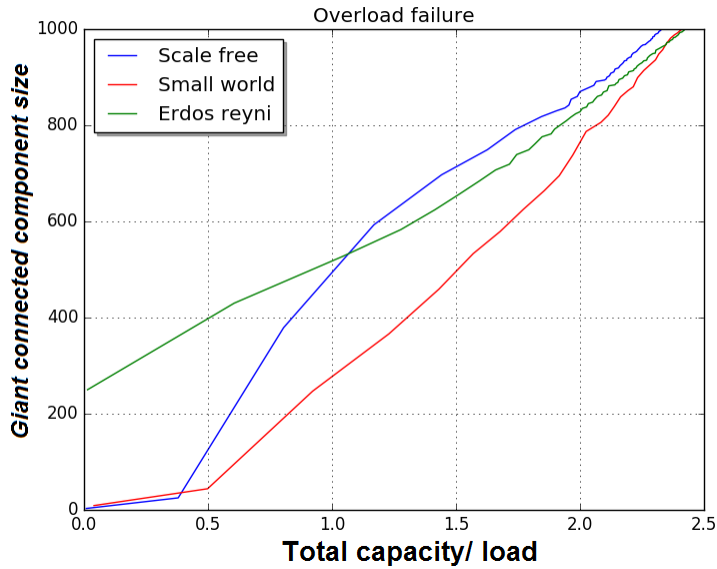


Figure 5.2: Giant component size vs η , $N = 1000, W = 4000, m = 3.5$

We expect giant network size is an increasing function of η . From Fig. 5.2 its clear that giant component size indeed an increasing function of η . If there is any overload failure in the network we are removing those nodes but the number of the walkers in the network remain constant which implies as time progress total capacity of the network will decrease. In this plot SF falls much faster than ER which means when load on a network is much higher than its total capacity (η is very small, in this case ~ 0.5) at this condition SF

network connectedness tend to zero that means there is no connection between nodes. On the other hand for ER when load on the network is very high then also the network maintain its connectedness ($N_{giant} \sim 240$) which implies even for overload condition ER is able to pass information within a big subnetwork.

5.1.3 Different phase of overload failure

We were discussing about overload failure and how network collapse occur due to overload failure. Initial overload failure in a network is not very effective because as we know initial failures are more likely to occur at a node with low degree and SF, complete graph, ER all of them are robust under random failure [2–7]. After deletion of certain number of nodes the load on the network increases rapidly. Then the network undergo a rapid failure. If the deletion process continues then enormous load on the network pushes the network towards cascading failure. Fig. 5.3 shows three different phases of overload failure.

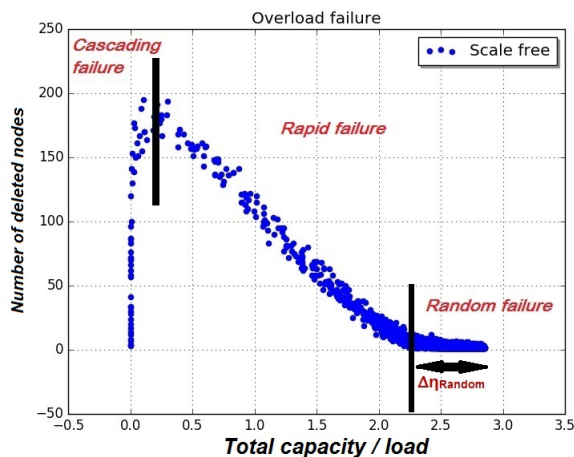


Figure 5.3: Phase of Network failure, $N = 1000$, $W = 6000$, $m = 4.4$, Combined all data got from 10 realization

Rapid failure and cascading failure are very fast process if a network enter into any of these two phase the entire network will experience a catastrophic failure. To control this sudden and total failure we need to stop the network failure process with in the random failure phase, because as we can see in this region network failure rate is smaller as compare to other two phase, so its easier to repair the damage nodes with in that phase. The time required to bring back those nodes into the network is called repair time τ_R .

We observe same pattern for different types of networks also. A network robustness depends on the population of deleted nodes and length of η belong to the random failure zone ($\Delta\eta_{Random}$). As time progress load on the network will increase so η will tends to zero. If the $\Delta\eta_{Random}$ for random failure zone greater than other network that means that network has the capability to work normally even under high load, network still belong to random failure zone without entering into overload failure zone. Another important point is the population of scattering points inside any of these zone. If population of points inside random failure zone is very high that means in each failure number of deleted nodes is very less which is good for the network. Any failure in the rapid failure region are the indication of deletion of large number of nodes so if population in rapid failure zone is less that means number of events when large number of nodes deleted from the network is rare.

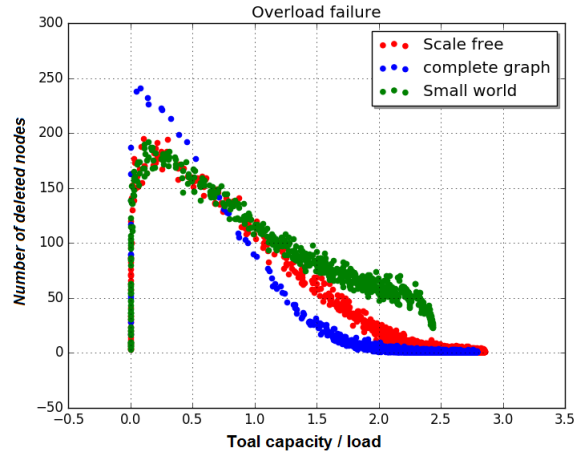


Figure 5.4: Phase of SF, Complete graph, Small world Network , $N = 1000, W = 6000, m = 4.4$, Combined all data got from 10 realization

In Fig. 5.4 I plotted three different phase for SF, Small world and complete graph among these three types of network complete graph has most homogeneous degree distribution, all nodes have $(N - 1)$ degree. As we discussed before that homogeneous networks are robust under overload failure, we can see that effect in this plot. $\Delta\eta_{Random}$ and population of points in random failure zone for SF is less than complete graph that means first several failure occurred in complete graph were not very effective, due to each failure number of deleted nodes was not sufficient to collapse the network. Similarly in SF first few failure were in random failure zone. But when rapid failure occur even for small η as compare to complete graph then at each failure step causes deletion of huge number of nodes.

The reason for this kind of behavior of SF is that SF has long tail power law degree distribution and we know SF is fragile against intentional attack on nodes with high degree [2–7]. In the random failure phase most of the nodes with lower degree are deleted because these nodes are more prone to experience EE. But after redistribution of load it is highly probable that nodes with higher degree will experience EE. Several failure at higher degree nodes catalyze the failure process to enter into rapid failure phase. On the other hand complete graph has homogeneous degree distribution and it is robust against targeted attack that is why failure of higher degree node in the 1st phase does not trigger next failure very fast, that is why we can see more scatter population in random failure region as compare to rapid and cascading failure zone.

5.1.4 Giant component and tolerance parameter

In this section we will discuss how network stability can be changes by changing tolerance parameter m . Network stability caused by a failure measured in terms of the relative size of the giant component N/N_0 where N is the size of the largest connected component and N_0 is the initial network size. Tolerance parameter is a constant which we put by hand. m is a mathematical parameter in this model but in real network m defines the robustness of a given network. I will give an example where we can relate m to a physical network. If we consider electrical power grid network, each transformer has certain capacity if current flow through that transformer exceeds that capacity it will be damaged. But some how if we manage to increase the capacity it will sustain against more electrical flow which is similar to increasing m .

To check how network stability changes by changing m common thing to plot relative giant component size vs m Fig. 5.5, Fig. 5.6. This simulation is just the continuation of overload failure simulation. In this case I collected the relative size of the giant component at the end of first 10 time steps. From Fig. 5.5, 5.6 till $m = 2$ the network disintegrate completely ($N/N_0 \sim 0$) but as we increase m probability of failure has gone down. The relative giant component size tends to 1. This results show us that failure frequency decreases as we increase m . Same argument is true for ER and complete graphs also Fig. 5.5 shows how N/N_0 changes as a function of m .

But there are two major difference between SF and ER (or complete graph). When m

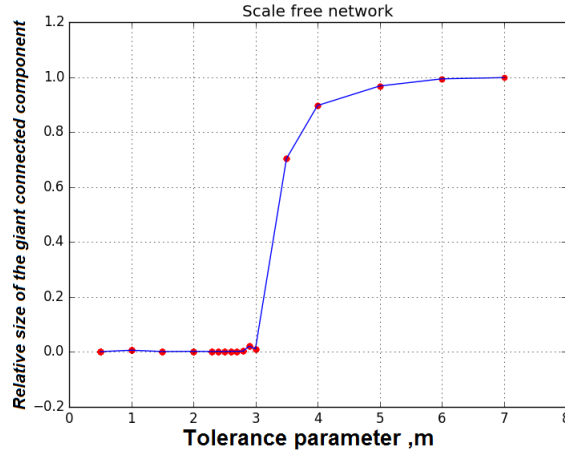


Figure 5.5: N/N_0 vs m , $N_0=1000$, $W=4000$, as we increase m N/N_0 increases.

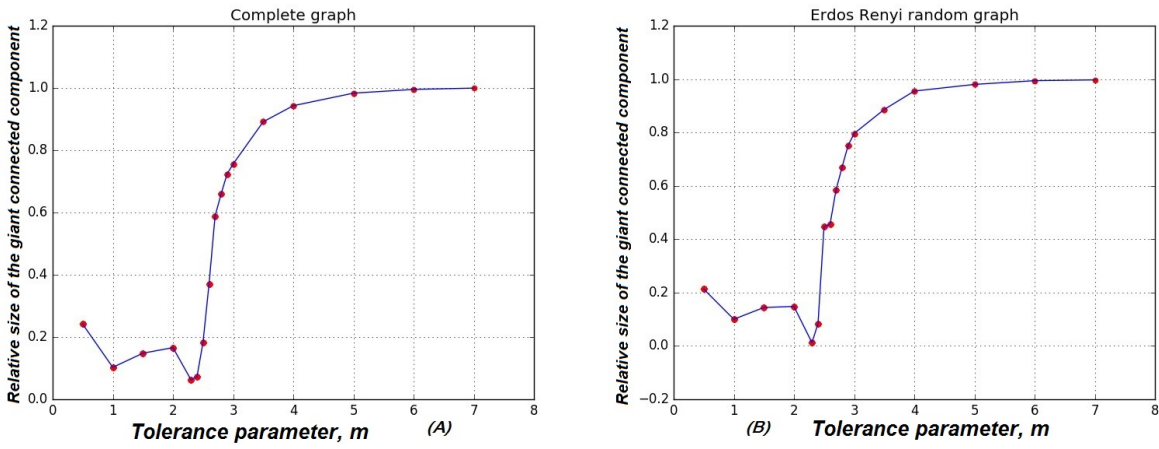


Figure 5.6: N/N_0 vs m , $N_0=1000$, $W=4000$, as we increase m N/N_0 increases.

is very small ER has a nonzero giant component size and another difference is that ER is more robust than SF against overload failure even for small m as compared to SF.

I plotted above three graphs at the end of first 10 time steps. The main reason behind that plot is to stop overload failure. Previously we defined three phases of a network failure. If we can decrease the failure frequency by changing m within the random failure zone then the network will not encounter an overload failure.

5.2 Discussions

We have seen many important results in previous section. In this section I will mention some important results and try to give possible explanation behind that results.

Our everyday life is surrounded by networks from internet to social relations. It is important to study the network robustness and find out some way to make a network more robust. My aim is to model overload failure on network using extreme events as I mentioned in the introduction part. Occurrence of EE depends on the degree of a node, an expression for EE is given in equation 3.6 . From the Fig. 3.2 we can see that nodes with lower degree are more prone to EE. I substituted the overload failure by EE because EE is the condition when number of walker on a node exceeds its capacity. In the next section we will discuss some possible way to change network stability.

5.2.1 Network stability

We can increase the total network tolerance factor Q for a given number of size network by increasing scale-free exponent γ for a scale free network. In equation 4.2 we see that m is an indirect function of γ through $P(k)$. From Fig. 4.3 we can see that as we increase γ m decreases slowly to keep total load Q constant. We know that $\langle f \rangle$ and σ both are increasing function of degree k that means capacity of each load also depends on their degree. keeping $\langle K \rangle$, Q and N fixed and increasing γ is equivalent to increasing individual capacity of load which implies total capacity of the network has to be increase but we kept Q as a constant. To keep Q as a constant we need to decrease the tolerance parameter m that is why in Fig. 4.3 m is a decreasing function of γ . That is a way to increase the total tolerance of the network.

5.2.2 Comparison between a heterogeneous and homogeneous degree distribution network stability

The definition of heterogeneous and homogeneous network is already mentioned before so we will start with two different networks belong to two different category. An example of

heterogeneous network is scale-free network(SF) and homogeneous network is Erdos-Renyi or complete graph(CG).

Robustness of network against random and intentional attack has been studied extensively in the context of percolation problem [19,20]. Study suggest that SF is robust against random failure but fragile against intentional attack, on the other hand ER, CG both are robust against intentional failure. Fig. 5.1 shows that SF network disintegrate faster than ER. As we know that there are three phase of network failure , initial failure happen with in the random failure zone. The probability of occurrence of EE at a node with lower degree is more than a higher degree node. Thus, initial failure occurs at lower degree nodes which does not have much effect on the connectedness of the network because SF is robust under random failure. But when redistribution of load happens then some nodes with high degree can face an EE but SF is fragile under intentional attack then a rapid failure can occur and the finally the network completely disintegrate. That is not the case for ER or CG , their degree distribution does not have tail like SF they have a uniform degree distribution that is why ER and CG are robust against intentional attack. If an intentional attack happens with in random failure zone that can be suppressed by other nodes that is why in Fig. 5.1 (left) SF collapse faster than ER.

5.2.3 Three phases of network failure and η

In the last section we have seen three different phase of network failure Fig. 5.3 . $\eta = Q/W$ is a quantity which indicates the transition from one phase to other. Instead of η if we choose time it will also produce same plot but network failure is a random process change of number of nodes in the network does not have any impact on time but number of surviving nodes has direct effect on η .

As we can see from Fig. 5.3 that depending on the rate of node failure there are three phases of a network failure. First phase is called random failure , second phase is called rapid failure and the third one is cascading failure. From the plot its clear that as far as failure due to over load is concerned a network failure can not start from rapid failure or cascading failure zone. Initial some failure depending on the types of the network has to belong to the random failure zone. If several nodes with high degree fails then the deactivation process gets elevated and a transition happens. First phase of the failure we call it random failure

although the node failure in that region is not random. Failure rate in this region is not very fast and most of the nodes with lower degree are likely to fail which has same effect as random failure on the network, that is why we define first region as random failure.

In the rapid failure region η is small because total capacity of the network decreases due to deletion process but total load on the network remain same this phase is very risky for a network because of high load. The population of deleted nodes in that region is a signature of network robustness. If scatter point density is very less that means most of the population are there in the random failure zone which is a safe and controllable phase. For this reason a ER is more stable than SF against overload failure as we can see from Fig. 5.4 .

5.2.4 Tolerance parameter m and network robustness

The quantity m called tolerance parameter because capacity of each node can be regulated by this quantity. During simulation I put it as a known parameter. If we increase m then the rate of network failure will be reduced, because each nodes should have higher capacity as compare to capacity with smaller m . Fig. 5.6 shows how relative giant component size changes as a function of m at the end of first 10 time steps. When m is very small then probability of node failure is very high that is why SF has relative giant component size $N/N_0 \sim 0$ which is not the case for ER or CG because they are robust under intentional failure. As m increase $N/N_0 \rightarrow 1$ which implies frequency of failure has been reduced. Hence, it is important to have large m to make a network robust, another advantage is that if the nodes have large m the frequency of failure will be less, that means repair time would be large. We should have sufficient time to bring a deactivated node back into the network.

5.2.5 Some new procedure I used in my project work and some new results:

I have already described all procedure that I used. Here I want to discuss about some procedure which are different from conventional procedure and I wold also like to discuss some new results which I found from my project work.

- People study network failure as a percolation model [19, 20], very few people used

random walk and extreme events model to study overload failure [21, 22]. In their model load on the network W was not a fixed quantity which changes as a function of number of edges or other parameters. In my case I did not changed total load I kept it as a constant.

- I defined a new quantity η which indicates how a network behaves as a function of capacity of total network. I found three phases of overload failure in each phase the rate of node failure is different, it increases as we go from random to cascading failure phase.
- I found that SF network is more fragile against overload failure as compare to CG and ER network.

5.2.6 Significance of the results

The results I have found some of these are comparative results and some of them has its own implication. As an example ER is more robust than SF this is a comparative result on the other hand there are three region in the network failure profile due to different rate of overload failure which has lots of information about repair time, frequency of failure, size of the network etc. The transition point when network shift one phase to another is very significant, although getting the exact transition point numerically is not possible till now but for a given network failure information we can study and compare the repair time distribution for three different phases.

We also came to know that we can increase a scale free network robustness by increasing the scale free exponent γ . SF is fragile against overload attack we can model a SF with same size and same $\langle k \rangle$ which is more robust than a SF with smaller γ .

Some examples of overload failure

- Blackout in northern India in 2012
- Blackout in northeast America in 2003
- Blackout in London in 2003

Conclusion

Overload failure can do great harms to complex network, especially to infrastructure network. It is important to understand failure process to protect the network against overload failure. During this project I have studied the network robustness against overload failure based on random walk and extreme event model. Random failure and intentional failure has been extensively studied [2–7], study suggest that SF is robust against random attack but fragile against targeted attack unlike ER and CG. Some of the research group have been studied overload failure based on extreme events [21, 22] but they did not consider total load on a network as a constant quantity, in my project I also used EE model but in my case I consider total load as a constant quantity.

Extreme event is a great model to study overload failure because we define EE if the flow through a node exceeds certain threshold, similar way we can define capacity of each node and consider a node failed when flow through that node exceeds the prescribed capacity. Extreme event probability for a smaller degree node is greater than a node with higher degree that implies first several failure will occur at some nodes with lower degree. Any failure leads to a new redistribution of load and, as a result, subsequent failure occurs. This failure will propagate and shutdown a large fraction of the network. A high degree node failure can causes a devastating overload failure for SF unlike ER, because SF is vulnerable against targeted attack. I found the signature of this comparison between SF and EE against overload failure in Fig. 5.1 which shows that certainly SF collapses much faster than ER.

I also wanted to study how network failure rate changes with respect to some quantity which is related to the present situation of the network. We defined a quantity η , from the construction of η we know that as time progress number of nodes in the system and η both decreases. I found two transition point in surviving nodes vs η plot so we can divide the network failure into three phases (1)*Random failure* (2)*Rapid failure* (3)*Cascading failure*. It is easier to repair a deactivated or failed node in the first phase because in that region number of deleted nodes is quiet low but if the network failure enters into the second phase its very difficult to bring a deactivated node back to the network because repair time is very short in this region. There is no numerical result which can predict the transition point in terms of η . Hence it is difficult to control a network transition from one to another phase.

I also found that we can increase the robustness of a SF by changing its scale-free ex-

ponent γ . If we increase gamma for a fixed m , fixed network size and fixed $\langle k \rangle$ then total tolerance of the network increases. I also explain how network robustness is related to network functionality, I showed that even for small m ER and CG is more robust than SF against overload failure. Heterogeneous degree distribution of SF make it versatile against overload attack, but ER has a binomial degree distribution which is homogeneous that is why ER is more robust under targeted attack or overload failure.

Implication of the presented results in the context of advancing the current research topic

The most important result is three phases of overload failure. One can try to derive a numerical result which can give the transition point. Once we know the transition point we can study the repair time distribution for three different phase and we can take effective safety measures according to the repair time distribution.

I mentioned two methods to make a network more robust one is changing m and another one is by increasing γ for SF. One can think of other ways to make a network more robust against overload failure.

In this report I only talked about the relative size of the giant component of a network, it is possible to study the complimentary network growth in terms of a percolating cluster which is equivalent to study how initial network failure spread over the network.

My results clarify that Erdos-Renyi and complete graph are more robust than scale-free network but it is difficult to replace all SF by ER or CG because in physical network links between two nodes are constrained by cost. I also found two transition point of overload failure and depending on these two point we can define three different phase of overload failure. These results could be help full to advancing the current research topic.

Bibliography

- [1] Reka Albert and Albert-Lszl Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002).
- [2] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
- [3] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
- [4] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [5] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, *Phys. Rev. E* **65**, 056109 (2002).
- [6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006).
- [7] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [8] *Extreme Events in Finance* by Francois Longin, M. F. M. Longin (wiley, 2001)
- [9] *Extreme Events in Nature and Society*, edited by S. Albeverio, V. Jentsch, and Holger Kantz (Springer, New York, 2005).
- [10] S. Boccaletti et al., *Phys. Rep.* **424**, 175 (2006); C. Nicolaides, L. Cueto-Felgueroso, and R. Juanes, *Phys. Rev. E* **82**, 055101(R) (2010); V. Tejedor, O. Benichou, and R. Voituriez, *Phys. Rev. E* **80**, 065104(R) (2009)
- [11] S. Boccaletti et al., *Phys. Rep.* 424, 175 (2006); C. Nicolaides, L. Cueto-Felgueroso, and R. Juanes, *Phys. Rev. E* **82**, 055101(R) (2010).
- [12] D.J. Watts, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5766 (2002).
- [13] P. Erdős, A. Renyi, "On Random Graphs-1", *Mathematicae*, **6**, 290–297 (1959).
- [14] F. Reif, *Statistical Physics*, (McGraw-Hill, 1965); C. Song et. al., *Nature Physics* **6**, 818b (2010).
- [15] J.D. Noh and H. Rieger, *Phys. Rev. Lett.* **92**, 118701 (2004).

- [16] Vimal kishore, M. S. Santhanam, and R. E. Amritkar. Extreme events on complex networks. *Phys. Rev. Lett.* **106**, 188701, (2011).
- [17] Mizutaka, S. & Yakubo, K. Structural robustness of scale-free networks against overload failures. *Phys. Rev. E*, **88**, 12803–12809 (2013).
- [18] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
- [19] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: structure and dynamics. *Phys. Rep.*, **424**, 175–308 (2006) .
- [20] Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. Critical phenomena in complex networks. *Rev. Mod. Phys.*, **80**, 1275 (2008).
- [21] S.Mizutaka, K.Yakubo, *journal of complex networks* **2**, 413–418 (2014)
- [22] Mizutaka and K. Yakubo, *Phys. Rev. E* **92**, 012814 (2015).