

Study of Extreme Events on Complex Networks

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

V. Sowmya



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

April, 2019

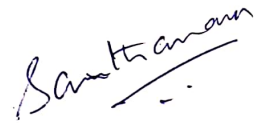
Supervisor: Dr. M.S. Santhanam

© V. Sowmya 2019

All rights reserved

Certificate

This is to certify that this dissertation entitled **Study of Extreme Events on Complex Networks** towards the partial fulfillment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by **V. Sowmya** at **Indian Institute of Science Education and Research, Pune** under the supervision of **Dr. M.S. Santhanam, Associate Professor, Department of Physics**, during the academic year **2018-2019**.



Dr. M.S. Santhanam

Committee:

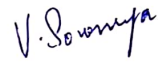
Dr. M.S. Santhanam

Prof. G. Ambika

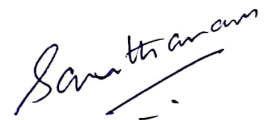
I would like to dedicate this thesis to my family . . .

Declaration

I hereby declare that the matter embodied in the report entitled **Study of Extreme Events on Complex Networks** are the results of the work carried out by me, **V. Sowmya**, at the **Department of Physics, Indian Institute of Science Education and Research, Pune**, under the supervision of **Dr. M.S. Santhanam, Associate Professor, Department of Physics**, during the academic year **2018-2019** and the same has not been submitted elsewhere for any other degree.



V. Sowmya



Dr. M.S. Santhanam

Acknowledgements

I would first like to thank my supervisor, Dr. M.S. Santhanam for his guidance and support, without which this thesis would not have been possible.

I am thankful to my TAC member, Prof. G. Ambika for her valuable suggestions.

I thank IISER-Pune's former Director, Prof. K.N. Ganesh and the current Director, Prof. Jayant B. Udgaonkar, the former Dean of Graduate Studies, Prof. G. Ambika and the current Dean of Graduate Studies, Dr. Bhas Bapat, for providing me with the opportunity and facilities to pursue my interests at the institute.

I would finally like to extend my heartfelt gratitude to my family and friends. Without their constant support and encouragement this project would not have been possible.

Abstract

The aim of the project was to study extreme events on complex networks. The study of complex networks, especially that of scale free and small world networks, has been an important topic of research in recent times. In this project, extreme events on complex networks were studied by using random walks to model flow on the network. By considering a set of independent random walkers to be moving on the network simultaneously, an extreme event on the network was defined as an event when the number of walkers exceeds a certain threshold. The existence of a stationary probability distribution, that is, the probability that a walker will be found on a node at a given time, allows for this definition of an extreme event. There is abundant literature on stationary occupancy probability distributions of random walks on networks. Review of literature on the dynamics and stationary probability distributions in the case of discrete and continuous time random walks constituted the first part of the project. In the next part of the project, some statistics of extreme events were studied. The probability of extreme events was obtained through simulations on a scale free network both in the discrete and continuous random walk cases. Correlation between magnitude differences of two consecutive extreme events and the time interval between the occurrence of the two was computed analytically and through simulations on scale-free, small-world and random networks. As a separate part of the project, the spectral properties of the adjacency matrices of scale-free networks were studied.

Table of contents

List of figures

1	Introduction	1
2	Theory	7
2.1	Random walks on networks	7
2.1.1	Discrete time random walks on networks	8
2.1.2	Continuous time random walks on networks	9
2.2	Statistics of extreme events on networks	13
2.2.1	Probability of occurrence of an extreme event	13
2.2.2	Correlation between difference in magnitudes of two consecutive extreme events and the time between their occurrences	14
3	Results: Higher Order Spacing Ratios of Adjacency Matrices of Networks	19
4	Results: Extreme Events on Complex Networks	25
4.1	Type of networks studied	25
4.2	Random walk simulations on scale-free networks	26
4.3	Statistics of extreme events	31
5	Outlook	43
	References	45

List of figures

3.1	Distributions of eigenvalue spacing ratios (of orders $k = 1, 2, 3, 4$) for scale-free networks with 1000 and 5000 nodes. The solid line is the analytic expression for the distribution obtained from Eq. 3.1 and 3.5.	22
3.2	Degree distribution of the co-authorship network (log-log plot). The distribution follows a power-law asymptotically	23
3.3	Distributions of eigenvalue spacing ratios (of orders $k = 1, 2, 3, 4$) for a real-world scale-free network (a co-authorship network) [1].The solid line is the analytic expression for the distribution obtained from Eq. 3.1 and 3.5. . . .	24
4.1	Time series of number of walkers on nodes of three different degrees for the power-law waiting time distribution case. The plots show the mean number of walkers μ and the threshold for an extreme event $\mu + 2\sigma$	27
4.2	Time series of number of walkers on nodes of three different degrees for the exponential law waiting time distribution case. The plots show the mean number of walkers μ and the threshold for an extreme event $\mu + 2\sigma$	28
4.3	Probability distribution of number of walkers on nodes of three different degrees for the power law waiting time distribution case. The histogram has been obtained from simulations of the random walk. The solid line is the analytic expression for the distribution which is given by the binomial distribution in Eq.4.3	29
4.4	Probability distribution of number of walkers on nodes of three different degrees for the exponential law waiting time distribution case.The histogram has been obtained from simulations of the random walk. The solid line is the analytic expression for the distribution which is given by the binomial distribution in Eq.4.3	30
4.5	Probability of the occurrence of an extreme event as a function of the degree of the node. The solid line is the analytic expression for the probability, which is given by Eqn. 4.4	32

4.6	Distribution of relaxation time (semi-log plot) for three nodes in the network. The solid line gives the linear fit for the plot, which shows that the distribution is exponential as given by Eq. 4.6	35
4.7	Distribution of the difference in magnitudes of consecutive extreme events on a node (Z). The solid line gives the analytic expression for the distribution, which is given by Eq. 4.7	36
4.8	Distribution of the absolute difference in magnitudes of consecutive extreme events on a node ($ Z $). The solid line gives the analytic expression for the distribution, which is given by Eq. 4.8	37
4.9	Pearson's correlation coefficient between the magnitude difference of two consecutive extreme events and the time interval between the occurrence of the two.	38
4.10	Pearson's correlation coefficient between the absolute magnitude difference between two consecutive extreme events and the time interval between the occurrence of the two.	39
4.11	Pearson's correlation coefficient between the absolute magnitude difference between two consecutive extreme events and the time interval between the occurrence of the two.	40
4.12	Pearson's correlation coefficient computed analytically and through simulations for a scale-free network. The top two plots show the correlation when the difference between magnitudes of consecutive extreme events Z is considered. The bottom two plots show the correlation when the absolute value of the difference ($ Z $) is considered.	41

Chapter 1

Introduction

In the past few decades, network science has emerged as an academic research field that involves the study of real-world systems by modeling them as complex networks. While providing newer insights into the structure and characteristics of various complex systems, it has also provided a formalism that enables the study of real-world systems which may be difficult to study otherwise. A network, as defined in graph theory, is a theoretical construct consisting of a set of vertices/nodes and edges/links, in which vertices are connected through edges in some manner that is characteristic of the network. Complex networks are networks in which the edges or links between different nodes is neither random nor regular, but such that they mimic or can be used to model real-world systems. Network science has applications in various fields including biology, physics, computer science, economics and sociology. Some complex systems that have been studied include biological protein or gene regulatory networks, food webs, the internet, the internet, the world wide web, etc. [2–6].

One important characteristic of a network is its topology, that is, the pattern in which various nodes are linked together in the network. While considering the topology of a network, scale-free and small-world properties are particularly of interest, as it has been observed these are properties of many real-world networks. To understand what a scale-free network means, one needs to define a quantity called the degree of a node, which is the number of links that the node has. A scale-free network is one in which the degree distribution of the nodes follows a power law distribution, that is, there are fewer nodes of higher degree than there of nodes of a smaller degree and the fall in the fraction of nodes with increasing degree is given by the power law distribution. Many real-world exhibit this property, few among which are the world wide web, citation networks, protein interaction networks and scientific co-authorship networks [4–7]. Another property of networks that is of interest is the small-world property. In small-world network, the degree of most nodes is very small, but

two nodes that are linked to the same node are very likely to be linked to each other as well. This property is better understood in terms of cliques and near-cliques. A subset of a network is called a clique if all the nodes in the subset are linked to each other. A near-clique, as indicated by its name, is a subset in which there exist links between most of the nodes of the subset. The small-world property of the network then refers to the tendency of nodes to form a number of cliques or near-cliques, which are poorly connected to other cliques or near-cliques, or in other words, the number of links between nodes of different cliques is small. A well-known example of such network, which has recently grown in popularity in network studies, is the human interaction or social relationship network. Electrical power grids, the internet, co-authorship and citation networks also exhibit this property.[4, 5] In this project, dynamics on both scale-free and small-world networks were studied. Dynamics on a random network (a network in which there is no apparent pattern in which the nodes are linked) was also considered. The Barabási-Albert model [8, 9], Watts-Strogatz model[10] and the Erdos-Renyi model[11] were used to construct scale-free, small-world and random networks respectively.

Diffusion or flow on networks, for example, the flow of current in an electrical circuit, spread of diseases in human interaction networks, flow of information on the internet, spread of dementia in the neural network of the brain[4, 12], have been widely studied for the past few decades. Flow on networks can be modeled using random walks on the networks, where a random walker moves from a node to one of the nodes that it is linked to, that is, a neighboring node. A random walk is a stochastic or random process in which the length of a jump or step and the time of the jump are random variables. In the case of networks, only the time of the jump is a random variable, the length of the jump is a constant. However, the neighboring node to which the walker jumps becomes a random variable and the transition probability is accordingly modified for random walks on networks. It is possible to consider two types of random walks, discrete time (DTRWs) and continuous time random walks (CTRWs) on networks. As the names suggest, the distinction between the two is in the distribution of the time interval after which the walker jumps from a node to one of its neighbors. In the discrete time case, the walker makes every jump after a constant period of time, that is, the time between two consecutive jumps of the walker is a constant. In the continuous time case, however, the time interval after which the walker moves is a random variable, and its distribution is given by a continuous probability density. In this project, both continuous and discrete time and continuous time random walks were considered.

One advantage of modeling flow using DTRWs is that it is possible to obtain analytic expressions for quantities of interest in most cases. However, since most physical stochastic processes are more accurately modeled by CTRWs, the study of CTRWs on networks has increased in the past few decades even though they might be more difficult to analyze. A formal analysis of continuous time random walks by considering the waiting time as a random variable began with the work of Montroll and Weiss[13], Montroll[14], Montroll and Scher [15]. In this project, two types of waiting time distributions have been considered, the Poisson or exponential distribution and the Pareto or power law distribution. While the exponential distribution is easier to analyze in the Montroll-Weiss formalism, the power law distribution provides a more accurate model for real-life processes. The power law distribution, being quite heavy tailed, can be used to model random processes like flow in human interaction networks such as mailing or voice calling, spread of diseases, etc., in which the distribution of waiting time between two consecutive jumps is quite heavy-tailed.

The study of extreme events on networks is a fairly new topic of research. A sharp spike in packet flow in the internet, traffic jams in transportation networks, a sudden increase in the expression of a gene in gene regulatory networks are some examples of extreme events on networks[16–19]. A fairly large change in flow on the network that exceeds the nodal capacity can be called an extreme event as it has a significant impact on the flow in the networks. By considering a fixed number of independent random walkers on the network, it is possible to define an extreme event as the event that occurs when the number of random walkers on a particular node exceeds a certain threshold, which can be defined based on the degree of the node. Even though the probability of occurrence of an extreme event might be small, the study of these events is important as they can disrupt normal flow or transport on the network. In this project, the probability of occurrence of an extreme event were studied in the continuous time random walk case by extending the existing analysis for the discrete time case.

The stationary probability distribution of a random walk on a network is of interest, as the existence of a stationary distribution allows for the above definition of an extreme event. The stationary probability of a given node in the network refers to the occupancy probability of the node after a steady state has been reached. A steady state may or may not be reached depending on the type of random walk or the network. In order to derive an expression for the stationary probability one would then consider the dynamics of the random walk on the network. A random walk on a strongly connected (a strongly connected network is one in which it is possible to go from any node on the network to any other node by walking on

existing edges in the network), undirected (that is, on any given edge, flow from both the end nodes is possible and flow is not restricted to one direction) network with N nodes has been considered. It is possible to define an $N \times N$ adjacency matrix for the network whose entries are given by either 1 or 0 depending on whether the two nodes corresponding to that entry are adjacent to each other or not. In other words, A_{ij} is 1 if nodes v_i and v_j are adjacent to each other and 0 if they are not. After each time interval, the walker makes a jump from one node, say v_i , to one of its neighbors, say v_j with probability that is inversely proportional to the total number of edges going from node v_i to other nodes in the network. An occupation probability for each node at a given time can be defined as the probability that the walker is occupying that node at that time. It is then possible to show that the occupation probability of a node goes to a stationary value after some time and that it is proportional to the degree of the node [20]. The stationary probability distribution has been widely studied in complex networks in the context of ranking in networks. Ranking refers to the attribution of a certain amount of importance to all the nodes in the network based on their connectivity to other nodes and the amount of transport or flow through the nodes. The PageRank algorithm, which is used to rank web pages in the world wide web and is a major component of the Google search algorithm, is one application of random walks on networks[2]. The PageRank vector that is computed for the web network and some modifications to it is equivalent to the stationary probability distribution of a discrete time random walk on the network[21]. A large amount of literature exists on the subject of stationary probability distributions, relevant portions of which have been elaborated in the next chapter.

While studying the statistics of extreme events, one might be interested in studying the correlation between event size difference and recurrence time between consecutive events. Distribution of the magnitude of the extreme event on a particular node is given by the tail of the probability distribution of the number of walkers on that node. The recurrence time, that is, the time between two consecutive extreme events on the other hand has been found to be exponentially distributed [17]. A correlation between the difference in event sizes of two extreme events and the time interval between the two events can be interesting to study. The presence of a correlation or an anti-correlation between the two quantities would mean that given the occurrence of an event, something about the magnitude and the time of occurrence of subsequent extreme events can be predicted to a small extent.

Also, as an independent part of the project, the spectral properties of the adjacency matrices of scale-free networks were studied. In particular, we looked at the distributions of eigenvalue spacing ratios of these matrices. In Random Matrix Theory, the spectral

properties of a matrix are used for characterizing the matrix [22, 23, 1, 24–27]. Some of the spectral properties that have been studied include the distribution of eigenvalues, eigenvalue spacing and eigenvalue spacing ratios. One commonly studied property is the nearest-neighbour eigenvalue spacing. If the eigenvalues of the matrix that is studied are given by, E_1, E_2, E_3, \dots , then the nearest neighbour eigenvalue spacing is given as $s_i = E_{i+1} - E_i$. The eigenvalue spacing ratio, given by $r_i = \frac{s_{i+1}}{s_i}$ is another quantity that is widely studied. It is possible to define higher order spacing ratios by $r_i^{(k)} = \frac{E_{i+2k} - E_{i+k}}{E_{i+k} - E_i}$. In this project, the higher order spacing ratios of adjacency matrices of scale-free distributions have been studied.

Most real-world networks are time dependent. For example, in the world wide web, new web pages are created or removed every day. Extinction or the introduction of a new species in an ecosystem thus changing the food web of the ecosystem, or the formation or dissolution of social relationships, are some other examples of time dependency of real-world networks [20, 28–31]. However, since static networks are easier to analyze much of the research on diffusion in networks has focused on static networks. Time dependency of networks is a fairly recent topic of interest. In the final chapter of the thesis, some of the ways in which extreme event analysis can be extended to time dependent networks have been outlined.

Thesis overview

Chapter 1 provides an introduction to the thesis and outlines in brief the concepts that have been used in the project. Also, connections are drawn between different parts of the project. In Chapter 2, relevant portions of existing literature on the subject of discrete and continuous time random walks on networks have been elaborated. Theory on extreme events on networks and some statistics of the extreme events that were studied have been explained. Analysis, that exists in literature for the discrete time random walk case, has been extended to the case of continuous time random walks on networks.

As an independent part of the project, spectral analysis of the adjacency matrices of scale-free networks was conducted by studying the higher order eigenvalue spacing ratios of these matrices. Chapter 3 briefly describes the theory of eigenvalue spacing ratios studied in Random Matrix Theory (RMT) and provides the results that were obtained from the spectral study of adjacency matrices of scale-free networks.

Chapter 4 consists of all the results on extreme events on complex networks that were obtained as part of the project. The first section of this chapter explains the properties of different networks that were studied and the algorithms that were used to simulate these networks. Simulations of discrete time random walks, which already exist in literature, and our extension of the analysis to continuous time random walks forms the next part of the

chapter. In the next section, some of the results on the statistics of extreme events have been presented. Chapter 5 concludes the thesis with a brief summary of the results that were obtained and outlines some of the future work that can be carried out in this area.

Chapter 2

Theory

2.1 Random walks on networks

A network or a graph is a set of vertices/nodes and edges/links, in which the nodes are connected to each other in a specific pattern that is characteristic of the network. Complex networks are networks that are used to model real-life systems, in which the arrangement of links between different nodes is such that the network mimics or models a real-world system. The study of networks in the past few decades has focused on various properties of the network such as its topography, centrality measures for the nodes of the network, community detection, transport/ flow on the networks, etc. In this section, we focus on the study of stochastic processes on networks which can be theoretically modeled by random walks on the networks.

Random walks on complex networks can be used to model a diverse number of physical processes, such as, the spread of diseases in human interaction networks, the spread of information in social networks, particle diffusion in porous media, flow of electricity in a power grid etc. Random walks may be broadly classified into discrete and continuous time random walks, depending on the distribution of the waiting time after which the walker jumps to another node. The study of random walks has provided newer insights into the physical systems under consideration and transport or flow in these systems. A few examples of topics that have been studied using the random walk model include the stationary occupancy probability of the nodes of the network, ranking in networks, range of the random walk and search on networks using random walks. In our project, we are interested in studying the stationary probability distribution of both discrete and continuous time random walks on complex networks.

We consider a strongly connected, undirected, N node network. An $N \times N$ adjacency matrix, A , is defined for the network whose entries are given by either 1 or 0 depending on whether the two nodes corresponding to that entry are linked to each other or not. In other words, A_{ij} is 1 if nodes v_i and v_j are linked to each other and 0 if they are not.

2.1.1 Discrete time random walks on networks

As mentioned earlier, discrete time random walks are random walks in which the waiting time before a walker jumps is a constant. The dynamics of discrete time random walk are as follows[20]-

After every fixed time interval, the walker makes a jump from one node, say v_i , to another node, say v_j with probability given by

$$T_{ij} = \frac{A_{ij}}{k_i} \quad (2.1)$$

where T is called the transition probability matrix. Let $P_i(n)$ be the probability that the random walker is at node v_i after the n^{th} jump. The evolution of the probability that the walker is at a given node is then given by,

$$P_j(n+1) = \sum_{i=1}^{i=N} P_i(n) T_{ij} \quad (2.2)$$

where $j \in 1, 2, \dots, N$ and i denote the index for the nodes of the network.

This probability can be written in terms of a vector, $P = (P_1, P_2, \dots, P_N)$

It is possible to define a stationary probability distribution (that is, the probability distribution that remains unchanged as time progresses) $P^* = (P_1^*, P_2^*, \dots, P_N^*)$ by

$$P_i^* = \lim_{n \rightarrow \infty} P_i(n) \quad (2.3)$$

Thus,

$$P^* = P^* T \quad (2.4)$$

Thus, the stationary density is the left eigenvector of the transition probability matrix. In case of a strongly connected network, the existence and uniqueness of the stationary probability distribution is guaranteed by the Perron-Frobenius theorem [32] and is given by,

$$P_i^* = \frac{k_i}{\sum_{j=1}^N k_j} \quad (2.5)$$

where k_i is the degree of node v_i .

Thus, the stationary occupancy probability of a node is proportional to the degree of the node.

2.1.2 Continuous time random walks on networks

The dynamics of a random walk on the network is given by a generalized master equation, considering the waiting time as a random variable, initially proposed by Montroll and Weiss[13], which has been later adapted to several other continuous time random processes on networks. The derivation of the generalized master equation in this section follows the work of [33–35].

An N node undirected, strongly connected network is considered whose adjacency matrix is denoted as A . As in the discrete time random walk case, it is possible to define a transition probability density, $T(v_i, t|v_j, t')$, as follows,

$$T(v_i, t|v_j, t') = \phi(v_i|v_j, t)\psi(t - t'|v_j) \quad (2.6)$$

which is the probability that the walker at node v_j at time t' jumps to the node v_i at time t . $\phi(v_i|v_j, t)$ denotes the jump density, that is, the probability that the walker jumps from node v_j to v_i at time t and $\psi(t - t'|v_j)$ is the waiting or interval time distribution, that is, the probability that the walker does not move from node v_j in the time interval from t' to t . In our case, the jump density is independent of time and hence can be written as,

$$\phi(v_i|v_j, t) \equiv \phi(v_i|v_j) \quad (2.7)$$

$$\phi(v_i|v_j) = \frac{A_{ij}}{k_j} \quad (2.8)$$

where A_{ij} is the adjacency matrix element of the network and k_j is the degree of the node v_j .

Also, since the waiting time distribution is independent of the node in which the walker is present in our case, it can be written as,

$$\psi(t - t'|v_j) \equiv \psi(t - t') \quad (2.9)$$

The probability density that the walker arrives at node v_i at time t after n jumps starting from v_0 at time $t = 0$ is given by the recursive relation,

$$r_{n+1}(v_i, t|v_0) = \sum_{j=1}^N \int_0^t T(v_i, t|v_j, t') r_n(v_j, t'|v_0) dt' \quad (2.10)$$

with the following initial condition,

$$r_0(v_i, t|v_0) = \delta_{v_i, v_0} \delta(t) \quad (2.11)$$

It is now possible to obtain the probability density that a walker is at node v_i at time t ,

$$P(v_i, t|v_0) = \int_0^t \Psi(t-t') r(v_i, t'|v_0) dt' \quad (2.12)$$

where, $\Psi(t-t')$ is the probability that the walker does not jump off the node v_i in the time interval $t-t'$, which can be written as,

$$\Psi(t-t') = 1 - \int_0^{t-t'} \psi(\tau) d\tau \quad (2.13)$$

and,

$$r(v_i, t|v_0) = \sum_{n=0}^{\infty} r_n(v_i, t|v_0) \quad (2.14)$$

By taking a Laplace transform, one will be able to simplify the equations further as a convolution becomes a product in the Laplace space. Taking the Laplace transform of Eq. 2.12 gives,

$$\widehat{P}(v_i, s|v_0) = \widehat{\Psi}(s) \widehat{r}(v_i, s|v_0) \quad (2.15)$$

The hat is used to denote the Laplace transform of the variable. While taking the Laplace transform of Eq. 2.10 gives,

$$\widehat{r}_{n+1}(v_i, s|v_0) = \sum_{j=1}^N \widehat{T}(v_i, s|v_j) \widehat{r}_n(v_j, s|v_0) \quad (2.16)$$

Putting Eq. 2.16 back in the Laplace transform of Eq. 2.14 gives,

$$\widehat{r}(v_i, s|v_0) = \sum_{n=0}^{\infty} \widehat{r}_n(v_i, s|v_0) \quad (2.17)$$

$$\widehat{r}(v_i, s|v_0) = \sum_{n=0}^{\infty} \widehat{T}(v_i, s|v_j) \widehat{r}(v_j, s|v_0) + \widehat{r}_0(v_i, s|v_0) \quad (2.18)$$

Thus,

$$\widehat{r}(v_i, s|v_0) = (1 - \widehat{T}(v_i, s|v_j))^{-1} \widehat{r}_0(v_i, s|v_0) \quad (2.19)$$

Putting, Eq. 2.19 back in Eq. 2.15, gives,

$$\widehat{P}(v_i, s|v_0) = \widehat{\Psi}(s)(1 - \widehat{T}(v_i, s|v_j))^{-1}\widehat{r}_0(v_i, t|v_0) \quad (2.20)$$

The Laplace transform of $\Psi(s)$ can be obtained from Eq. 2.13,

$$\widehat{\Psi}(s) = \frac{1}{s}(1 - \widehat{\psi}(s)) \quad (2.21)$$

Putting, this back in Eq. 2.20 gives,

$$\widehat{P}(v_i, s|v_0) = \frac{1}{s}(1 - \widehat{\psi}(s))(1 - \widehat{T}(v_i, s|v_j))^{-1}\widehat{r}_0(v_i, t|v_0) \quad (2.22)$$

This, gives the generalized master equation of a continuous time random walk in Laplace space.

Finally, by taking the inverse Laplace transform of the above equation and some simplification, the generalized master equation for a continuous time random walk in real space can be obtained as follows,

$$\frac{\partial}{\partial x}P(v_i, t|v_0) = \sum_{j=1}^N \phi(v_i|v_j) \int_0^t K(t-t')P(v_j, t'|v_0)dt' - \int_0^t K(t-t')P(v_i, t'|v_0)dt' \quad (2.23)$$

which gives the time evolution of the occupancy probability of any given node on the network. $K(t-t')$ is the inverse Laplace transform of

$$\widehat{K}(s) = \frac{\widehat{\psi}(s)}{\widehat{\Psi}(s)} \quad (2.24)$$

In this project, one of the continuous time waiting densities that was considered is the exponential or Poisson probability density and the Pareto or power law waiting time density and random walk simulations were carried out for both cases. The stationary probability distributions for the random walk in these two cases have been considered in the next two sections.

Poisson waiting time density

In this case, the waiting time density is given by the exponential or Poisson distribution as,

$$\psi(t) = \lambda e^{-\lambda t} \quad (2.25)$$

For this waiting time density, it is possible to simplify the generalized master equation given by Eq.2.23 to give,

$$\frac{\partial}{\partial x} P(v_i, t|v_0) = \sum_{j=1}^N \lambda \phi(v_i|v_j) P(v_i, t|v_0) - \lambda P(v_i, t|v_0) \quad (2.26)$$

The steady state occupancy probability can then be obtained by setting the first derivative of the occupancy probability to 0 and it is found that it is the same as in the discrete time random walk case and is given by $P_i^* = \frac{k_i}{\sum_{j=1}^N k_j}$, where i and j are node indices, P_i^* is the stationary occupancy probability of the node v_i and k_i is the degree of the node v_i .

Non-Poisson waiting time density

In general, it is not possible to obtain exact analytic expression for the generalized master equation, given by Eq. 2.23, for non-Poisson waiting time densities. It can be shown that the stationary probability distribution in the case on non-exponential waiting time densities is the same as in the discrete case and it is proportional to the degree of the node. The generalized master equation in Laplace space that was derived earlier (Eq. 2.22) can be written in matrix form as,

$$\widehat{P}(s) = \frac{1}{s} (I - \widehat{D}(s)) (I - \widehat{T}(s))^{-1} P(0) \quad (2.27)$$

where,

$$\widehat{D}_{ij}(s) = \psi(s) \delta_{ij} \quad (2.28)$$

The following arguments were used to obtain an expression for the stationary probability distribution of a general continuous time random walk on the network [34].

By the final value theorem, the stationary probability vector P^* is given by,

$$P^* = \lim_{t \rightarrow \infty} P(t) = \lim_{s \rightarrow 0} s \widehat{P}(s) = MP(0) \quad (2.29)$$

The leading eigenvector of M will give the stationary probability vector. In the $\lim_{s \rightarrow 0}$, an approximation to the Laplace transform can be used to get an expression for M and one can show that the leading eigenvector of M is the same as the stationary probability vector in the discrete time random walk case.

Thus, the steady state occupancy probability of a node in the case of continuous waiting time densities is proportional to the degree of the node and is given by $P_i^* = \frac{k_i}{\sum_{j=1}^N k_j}$.

2.2 Statistics of extreme events on networks

Extreme events on networks are sudden large changes in the amount of flow in the network, that could possibly significantly impact future flow on the network. A few examples of extreme events on networks include sharp spikes in the number of packets moving in the internet that temporarily stops the transport of information, traffic jams in transportation networks, a sudden increase in the number of organisms of a particular species in a food web, etc. While studying extreme events, the distribution of the magnitude of extreme events, the time between two consecutive extreme events, the probability of occurrence of an extreme events are some of the quantities of interest[36]. A lot of recent research in this topic has been focused on the prediction and control of such events on networks. In this project, the statistics that we have analyzed are the probability of occurrence of an event and the correlation between the magnitude difference between two consecutive extreme events and the time interval between their occurrences (recurrence time). In all derivations in this section (Section 2.2), a random variable is denoted by an upper case letter (T, X, Y, Z, \dots) and the value it assumes is denoted by a lower case letter (t, x, y, z, \dots), unless stated otherwise. $P(X = x)$ has been used to denote the probability that the variable X assumes the value x .

2.2.1 Probability of occurrence of an extreme event

Consider a total of N_W walkers moving independently on the network. At any given time after the steady state has been reached, the probability that a walker is found on a particular node v_i is given by $p_i^* = \frac{k_i}{\sum_{j=1}^N k_j}$. Given that the occupancy probability density for a random walk reaches a steady state value, it is possible to obtain the distribution of the number of random walkers on a particular node. Let W denote the random variable corresponding to the number of walkers on a given node after steady state has been reached and $f_W(w) = P(W = w)$ denote the probability that there are w walkers on the node that is being considered. In order to find the distribution of W , one would consider the probability of finding w walkers on a particular node, which is given by p^w since all the walkers move independently[17, 18]. To find the distribution one would then choose w walkers out of the total N_W number of walkers, consider that they occupy one particular node at a given time and that the rest of the walkers are present on other nodes in the network. Thus, the distribution of the number of walkers on

a particular node as a function of the stationary occupancy probability p of the node is given by,

$$f_W(w) = P(W = w) = \binom{N_W}{w} p^w (1-p)^{(N_W-w)} \quad (2.30)$$

Let μ and σ denote the mean and the standard deviation of the binomial distribution given above.

An extreme event is then defined as follows- when the number of random walkers on a given node exceeds a threshold, defined as $q = \mu + m\sigma$, where m can be any real number. The probability of occurrence of an extreme event on a node is thus dependent on the degree of that node. Given this definition of an extreme event, one can then find the probability of the occurrence of an extreme event on the network, which is a function of the degree of a node (k) and is given by,

$$P(W > q) = \sum_{j=\lfloor q \rfloor + 1}^{N_W} \binom{N_W}{j} p^j (1-p)^{(N_W-j)} \quad (2.31)$$

where $\lfloor q \rfloor$ denotes that floor integer of q .

2.2.2 Correlation between difference in magnitudes of two consecutive extreme events and the time between their occurrences

Another statistic that we are interested in is the correlation between the difference in magnitude of two consecutive extreme events and the time between the occurrence of the two (recurrence time) on a particular node in the network. In extreme value theory, this quantity is often computed to see if the occurrence of extreme events in the system has some degree of predictability. If the two quantities are either correlated or anti-correlated to some extent, then one can speculate to a certain extent the magnitude or time of occurrence of an extreme event from the knowledge of extreme events in the past. In order to compute this correlation quantity, one would first need to compute the distribution of the recurrence time and that of the difference in magnitudes of consecutive extreme events on a given node on the network. To compute this quantity, we have considered discrete time random walks in three types of networks- scale-free, small-world and random networks. We consider a generic node on the network and compute this statistic on the node. The same analysis extends to all the nodes in the network.

Distribution of recurrence time- Assume that a total of N_E extreme events occur on a particular node in the finite time for which the walkers are considered to move on the network. Let $T_1, T_2, T_3, \dots, T_{N_E}$ denote the times of occurrence of successive extreme events on a given node. The recurrence time then refers to the time interval between two consecutive extreme events, say, $T_{R_n} = T_{n+1} - T_n$, $n = 1, 2, 3, \dots, N_E - 1$. We are interested in the probability distribution of this variable T . The recurrence time T , in the case of discrete time random walks, has been found to be distributed exponentially from computer simulations of the same. The distribution is given by,

$$P(T_{R_n} = t) = \lambda e^{-\lambda t} \quad (2.32)$$

where, $n = 1, 2, 3, \dots, N_E - 1$ the constant λ depends on the the node that is being considered. The mean $E[T_R]$ and the standard deviation σ_{T_R} of the distribution are then given by,

$$E[T_R] = \sigma_{T_R} = \frac{1}{\lambda} \quad (2.33)$$

Distribution of difference in magnitudes of two consecutive extreme events- By the definition of a extreme event we have used, the magnitude or size of an extreme event on a given node is the number of walkers on the node. Let $M_1, M_2, M_3, \dots, M_{N_E}$ denote the time series of the magnitudes of the N_E extreme events on the node. We are interested in computing the distribution of the difference in magnitudes of consecutive extreme events, denoted as, $Z_n = M_{n+1} - M_n$, $n = 1, 2, 3, \dots, N_E - 1$ and the distribution of the absolute value of the difference in magnitudes of consecutive extreme events, denoted as, $|Z_n| = |M_{n+1} - M_n|$, $n = 1, 2, 3, \dots, N_E - 1$. In order to find the desired distributions, we first consider the distribution of the number of random walkers on the node, which is given by 4.3,

$$f_W(w) = \binom{N_W}{w} p^w (1-p)^{(N_W-w)} \quad (2.34)$$

We then find the distribution of the magnitude M_n of an extreme event on the node. This is given by,

$$P(M_n = w) = P(W = w | W > q) = \frac{1}{P(W > q)} P(W = w \cap W > q) \quad (2.35)$$

where, q is the threshold used for defining an extreme event.

Let

$$c = \frac{1}{P(W > q)} \quad (2.36)$$

$$P(M_n = w) = P(W = w | W > q) = cP(W = w)\mathbf{1}_{(q, N_W]}(w) \quad (2.37)$$

where $\mathbf{1}_{(q, N_W]}(w)$ is the indicator function which takes the value 1 for $w \in (q, N_W]$ and the value 0 for all other w and $P(W = w)$ is the binomial distribution $f_W(w)$ of the number of walkers on the node. Thus, the distribution of the magnitude of an extreme event on a network can be written as,

$$P(M_n = w) = \begin{cases} c \binom{N_W}{w} p^w (1-p)^{(N_W-w)} & w > q \\ 0 & \text{otherwise} \end{cases} \quad (2.38)$$

where c is given by Eq. 2.36.

One can then obtain the distribution of Z_n and $|Z|_n$ as follows. For simplicity, the Gaussian approximation to the binomial equation is considered. With this approximation, Eq. 2.38 can be written as,

$$P(M_n = x) = \begin{cases} \frac{c}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & x > q \\ 0 & \text{otherwise} \end{cases} \quad (2.39)$$

where the mean and standard deviation (μ and σ) of the Gaussian distribution are the same as that of the binomial distribution that it approximates.

For notational simplicity, let $Y (= M_n)$ and $X (= M_{n+1})$, $n \in 1, 2, 3, \dots, N_E - 1$, denote the magnitudes of two consecutive extreme events on a node and $Z (= Z_n)$, $n \in 1, 2, 3, \dots, N_E - 1$ denote their difference. The distribution of $Z_n = X - Y$ can then be obtained as,

For $z \geq 0$,

$$P(Z = z) = P(X - Y = z) = P((X = z + y) \cap (Y = y)) \quad (2.40)$$

$$P(Z = z) = \int_q^\infty P(X = z + y)P(Y = y)dy \quad (2.41)$$

$$P(Z = z) = \int_q^\infty \frac{c^2}{2\pi\sigma^2} e^{-\frac{(z+y-\mu)^2}{2\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \quad (2.42)$$

since X and Y are identically distributed as given by Eq. 2.39. On simplification, this yields for $z \geq 0$,

$$P(Z = z) = \frac{c^2}{2\sqrt{\pi}\sigma} e^{-\frac{z^2}{4\sigma^2}} \left[\frac{1}{\sqrt{2\pi}(\sigma/\sqrt{2})} \int_q^\infty e^{-\frac{(y-(\mu-\frac{z}{2}))^2}{2(\sigma/\sqrt{2})^2}} dy \right] \quad (2.43)$$

$$P(Z = z) = \frac{c^2}{2\sqrt{\pi}\sigma} e^{-\frac{z^2}{4\sigma^2}} \left[1 - \Phi_{((\mu-\frac{z}{2}), (\frac{\sigma}{\sqrt{2}})^2)}(q) \right] \quad (2.44)$$

where, $\Phi_{(\mu, \sigma^2)}(q)$ denotes the cumulative distribution function of the Gaussian distribution with mean μ and variance σ^2 at q .

Now, we consider $z < 0$. Note that, $P(Z = -z) = P(X - Y = -z) = P(Y - X = z)$. Since, X and Y are identically distributed, $P(Z = -z) = P(Z = z)$. Thus, the distribution is symmetric around 0. This allows us to write, for $z \leq 0$,

$$P(Z = z) = \frac{c^2}{2\sqrt{\pi}\sigma} e^{-\frac{z^2}{4\sigma^2}} \left[1 - \Phi_{(\mu + \frac{z}{2}, (\frac{\sigma}{\sqrt{2}})^2)}(q) \right] \quad (2.45)$$

Hence, the distribution of Z can be written as,

$$P(Z = z) = \frac{c^2}{2\sqrt{\pi}\sigma} e^{-\frac{z^2}{4\sigma^2}} \left[1 - \Phi_{((\mu - \frac{|z|}{2}), (\frac{\sigma}{\sqrt{2}})^2)}(q) \right], \forall z \quad (2.46)$$

To find the distribution of the absolute value of the difference, $|Z|$, we note that,

$$P(|Z| = z) = [P(Z = z) + P(Z = -z)] \mathbf{1}_{(0, \infty)}(z) \quad (2.47)$$

where $\mathbf{1}_{(0, \infty)}(z)$ is an indicator function, which takes the value 1 for $z \in (0, \infty)$ and the value 0 otherwise. Thus, from Eq. 2.46, we get,

$$P(|Z| = z) = \begin{cases} \frac{c^2}{\sqrt{\pi}\sigma} e^{-\frac{z^2}{4\sigma^2}} \left[1 - \Phi_{((\mu - \frac{|z|}{2}), (\frac{\sigma}{\sqrt{2}})^2)}(q) \right] & z > 0 \\ \frac{c^2}{2\sqrt{\pi}\sigma} \left[1 - \Phi_{(\mu, (\frac{\sigma}{\sqrt{2}})^2)}(q) \right] & z = 0 \end{cases} \quad (2.48)$$

The correlation that we are interested in can then be defined as,

$$C_P = \frac{E[T_R H] - E[T_R]E[H]}{\sigma_{T_R} \sigma_H} \quad (2.49)$$

where, $H \in (Z, |Z|)$, $E[T_R]$ and $E[H]$ denote the expectation values of the random variables T_R and H , σ_{T_R} and σ_H denote the standard deviations of the two variables.

Simulations of discrete time random walks were carried out on scale-free, small-world and random networks to compute this correlation. The above analysis was used to obtain a semi-analytic expression for the same.

Chapter 3

Results: Higher Order Spacing Ratios of Adjacency Matrices of Networks

As an independent part of the project, the spectral properties of the adjacency matrices of scale-free networks were studied. In particular the eigenvalue spacing ratio distributions of the adjacency matrix was studied. In this chapter, a brief overview of the theory on eigenvalue spacing ratio distributions has been given, following which the results of studying these distributions in simulated and real-world scale-free networks have been presented.

In random matrix theory (RMT), eigenvalue distributions, eigenvalue spacing distributions and spacing ratios are some measures of spectral fluctuations that are used to characterize a system. RMT can be used to study the statistical properties of the spectra of certain complex systems [24–27]. One of the most commonly studied measure of spectral fluctuations is the nearest neighbour eigenvalue spacing, which is defined as, $s_i = E_{i+1} - E_i$ for $i = 1, 2, \dots$, where E_i denotes the i^{th} eigenvalue of the matrix. However, the study of eigenvalue spacing requires one to unfold the eigenvalue spectra system, which is a complicated process by which spectral features that are system dependent are removed, in order to study spectral fluctuations of the system [24, 25, 37]. The eigenvalue spacing ratio, defined as $r_i = \frac{s_{i+1}}{s_i}$ for $i = 1, 2, \dots$, is independent of the local density of states of the system and hence one does not have to unfold the eigenvalue spectra to study spacing ratios [22].

One of the widely studied models in RMT is the Gaussian Orthogonal Ensemble (GOE), which has been used to model systems which can be represented by symmetric matrices, whose entries are derived from a standard normal distribution. It has been found that the RMT average for spacing ratio distribution corresponding to the Gaussian orthogonal ensemble ($\beta = 1$) is given by [37, 22],

$$P(r, \beta) = C_\beta \frac{(r + r^2)^\beta}{(1 + r + r^2)^{1 + \frac{3}{2}\beta}} \quad (3.1)$$

Where C_β is a constant depending on β ,

and,

$$r_i = \frac{s_{i+1}}{s_i} \quad (3.2)$$

$$s_i = E_{i+1} - E_i \quad (3.3)$$

for $i=1,2,\dots$ where E_i denotes the i^{th} eigenvalue of the adjacency matrix, s denotes the i^{th} eigenvalue spacing and r_i denotes the i^{th} spacing ratio ,

Higher order spacing ratios are defined by,

$$r_i^{(k)} = \frac{E_{i+2k} - E_{i+k}}{E_{i+k} - E_i}, (i, k = 1, 2, \dots) \quad (3.4)$$

where $r_i^{(k)}$ is the k^{th} order spacing ratio.

It has been observed that the k^{th} order spacing ratio distribution is given by [22],

$$P^k(r, \beta) = P(r, \beta') \quad (3.5)$$

$$\beta' = \frac{k(k+1)}{2}\beta + (k-1) \quad (3.6)$$

for $k \geq 1$

In this project, we tried to find out if the eigenvalue spacing ratio distribution of the adjacency matrices of scale-free networks follows GOE statistics. In order to do so, the spacing ratio distributions were computed for the adjacency matrix of simulated scale-free networks with 500, 1000 and 5000 nodes. The spacing ratio distributions of a real-world scale-free network were also computed.

To simulate a scale-free network, the Barabási-Albert model proposed in [8, 9] was used. A scale-free network is a network in which the degree distribution of the nodes follows, at least asymptotically, a power-law distribution. The fraction $\omega(k)$ of nodes that have a given degree k (connect to k other nodes) is given by,

$$\omega(k) = ck^{-\gamma} \quad (3.7)$$

where, c and γ are positive constants.

In the Barabási-Albert model, new nodes are added sequentially one by one and the links that the new node makes after each new addition are made preferentially. Preferential attachment, in this case, means that the new node is more likely to form links with nodes that have higher degree, a rich get richer scenario. This model was initially developed based on network growth in the world wide web, which is a good example of a scale-free network. When new web pages are added to the network, they are more likely to form links with web pages that have a high degree, also known as hubs, rather than to pages that have fewer links. In this model, a network is built using the following algorithm. One first starts with an initial connected network. Newer nodes are added one by one and the probability that an already existing node forms a link with the new node is proportional to the degree of the existing node. In this way, the rest of the nodes are added to build a scale-free network.

The eigenvalue spacing ratios of a scale-free networks with 1000 and 5000 nodes, that were simulated using the Barabási-Albert algorithm, were studied and the results have been presented in Fig. 3.1. The figure (Fig.3.1) shows the distribution of the eigenvalue spacing ratios for $k=1,2,3$ and 4. It is observed that the spacing distributions follow the GOE statistics given by Eq. 3.1 and Eq. 3.5.

The analysis was also carried out for a real-world scale-free network. A co-authorship network in Condensed Matter Physics with 23,133 nodes and 93,497 edges, obtained from the article [1], was used for analysis. A co-authorship network is a network in which the authors of journal papers form the nodes of the network and an edge between two nodes indicates that the two authors corresponding to these nodes have co-authored a paper together. Co-authorship networks have generally been known to be scale-free networks. The degree distribution of this network was computed to verify that the network is scale-free. This has been shown in Fig.3.2. Next, higher order eigenvalue spacing ratios were computed for the adjacency matrix of this network and the plots are shown in Fig.3.3 .The results that were obtained again suggest that the eigenvalue spacing ratios of the adjacency matrix of scale-free networks follow the GOE statistics given by Eq. 3.1 and Eq. 3.5.

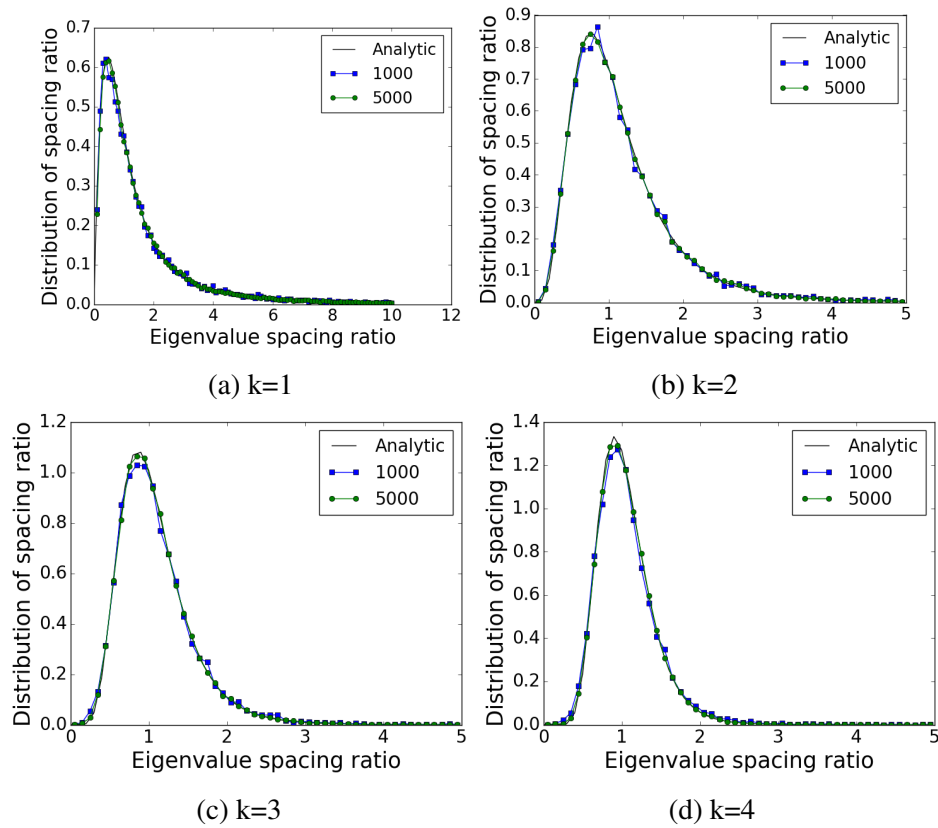


Fig. 3.1 Distributions of eigenvalue spacing ratios (of orders $k = 1, 2, 3, 4$) for scale-free networks with 1000 and 5000 nodes. The solid line is the analytic expression for the distribution obtained from Eq. 3.1 and 3.5.

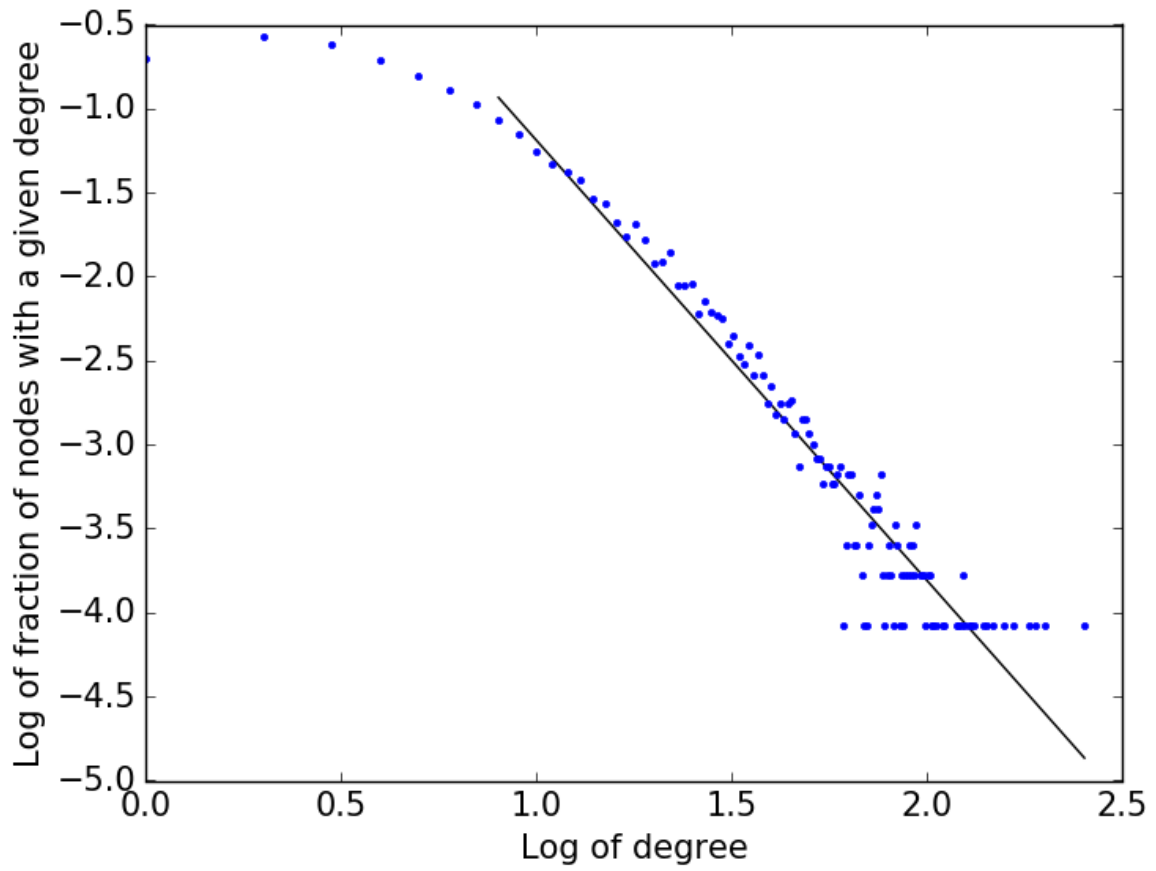


Fig. 3.2 Degree distribution of the co-authorship network (log-log plot). The distribution follows a power-law asymptotically

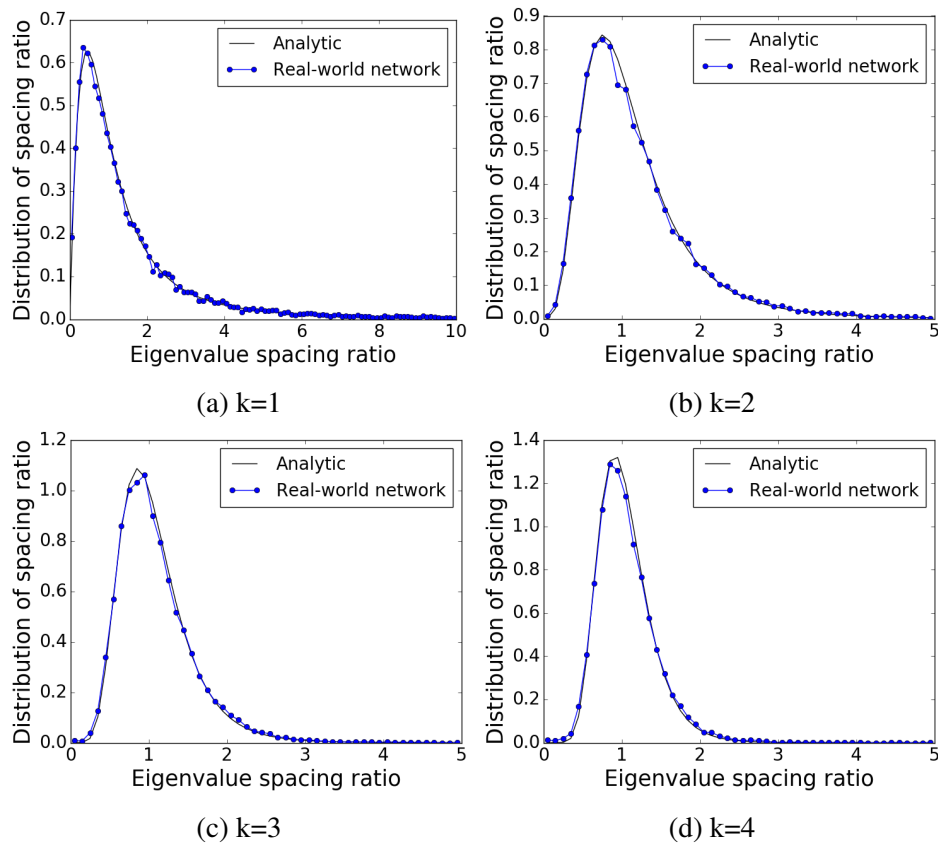


Fig. 3.3 Distributions of eigenvalue spacing ratios (of orders $k = 1, 2, 3, 4$) for a real-world scale-free network (a co-authorship network) [1]. The solid line is the analytic expression for the distribution obtained from Eq. 3.1 and 3.5.

Chapter 4

Results: Extreme Events on Complex Networks

4.1 Type of networks studied

In this project, three types of networks- scale-free, small-world and random networks were simulated and statistics of extreme events were analyzed for these network configurations. A brief overview of the algorithms that were used in the simulations is described in this section. In the previous chapter, scale-free networks were discussed in detail. In this section, the algorithms that were used for simulating small-world and random networks have been discussed.

Small-world networks

Small-world networks are networks that show a high degree of clustering and the tendency to form cliques or near-cliques. Another characteristic of the network is that the average path length between two nodes in the network is very small. The Watts-Strogatz model [10] was used to construct a small-world network. The algorithm that was used is as follows. In order to construct an N node network, with an average degree of k , first a ring lattice is constructed with N nodes with each of them connected to their k nearest neighbour, that is, $\frac{k}{2}$ on either side. A rewiring probability β is then considered. The $\frac{k}{2}$ edges on one side, say the right, of each node is then rewired with probability β to connect to any other node in the network. A value of β close to 0 creates an almost regular network and a value close to 1 creates an almost completely random network. By choosing, a suitable value of β one will then be able

to construct a small-world network.

Random networks

Random networks are graphs in which the vertices are connected randomly, with no apparent pattern to the arrangement of edges. In order to construct a random network, the Erdos-Renyi model [11] was used. In this model, an N node system is considered initially with no links between vertices. The edge between any two nodes is then added independently (not depending on the presence or absence of other edges in the network) based on some probability p . The number of edges in the network would then be determined by p . In this way, a random graph can be constructed.

4.2 Random walk simulations on scale-free networks

All discrete and continuous time random walk simulations in this section were carried out on a scale-free network with 1000 nodes, 3132 edges and with 5000 random walkers. For continuous time random walk simulations, the exponential waiting time density and the power law waiting time density were chosen. The exponential waiting time distribution that was used in the simulation is given by,

$$\psi(t) = \lambda e^{-\lambda t} \quad (4.1)$$

with $\lambda = \frac{1}{2}$.

The power-law distribution that was used in simulations is the Pareto distribution, which is given by,

$$\psi(t) = \begin{cases} \frac{\gamma}{t^{\gamma+1}} & t \geq 1 \\ 0 & t < 1 \end{cases} \quad (4.2)$$

In this section the time series of the number of random walkers on three different nodes of the network (Fig. 4.1, Fig. 4.2) and the probability distribution of the number of walkers on these nodes (Fig. 4.3 and Fig. 4.4) have been shown. The three nodes have been chosen based on their degree, one of high, moderate and small degrees each, to effectively show the dynamics of the random walk in three different cases. The simulations for the discrete time case already exist in literature and have not been shown here.

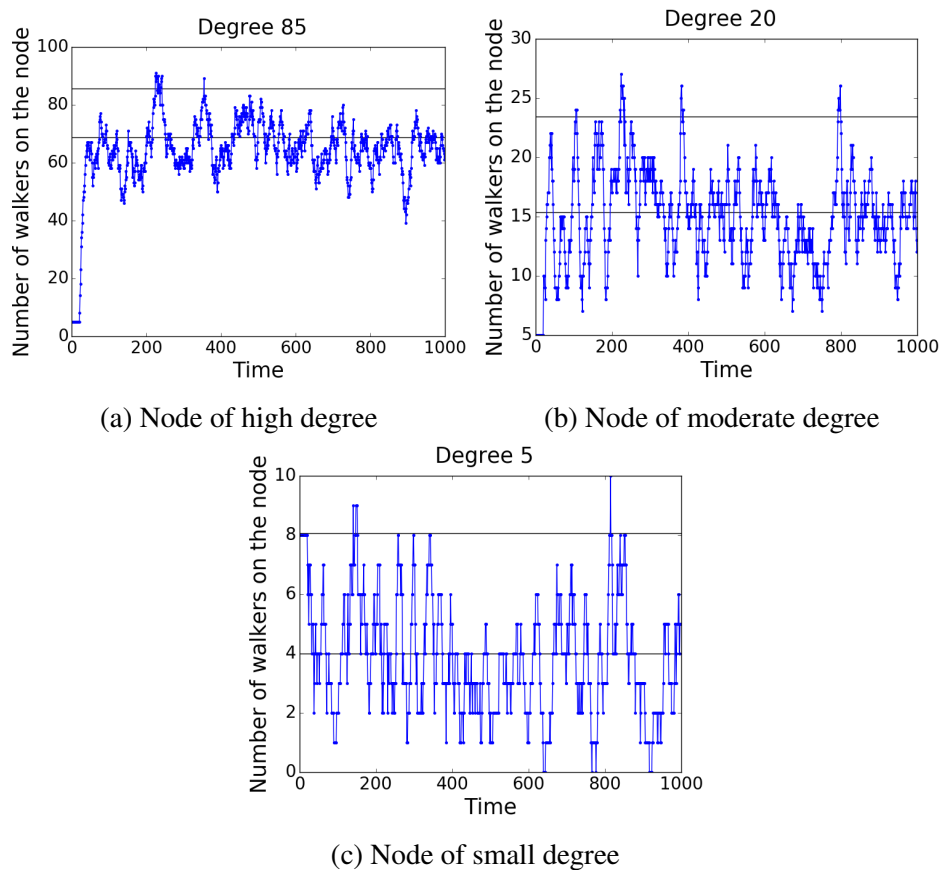


Fig. 4.1 Time series of number of walkers on nodes of three different degrees for the power-law waiting time distribution case. The plots show the mean number of walkers μ and the threshold for an extreme event $\mu + 2\sigma$

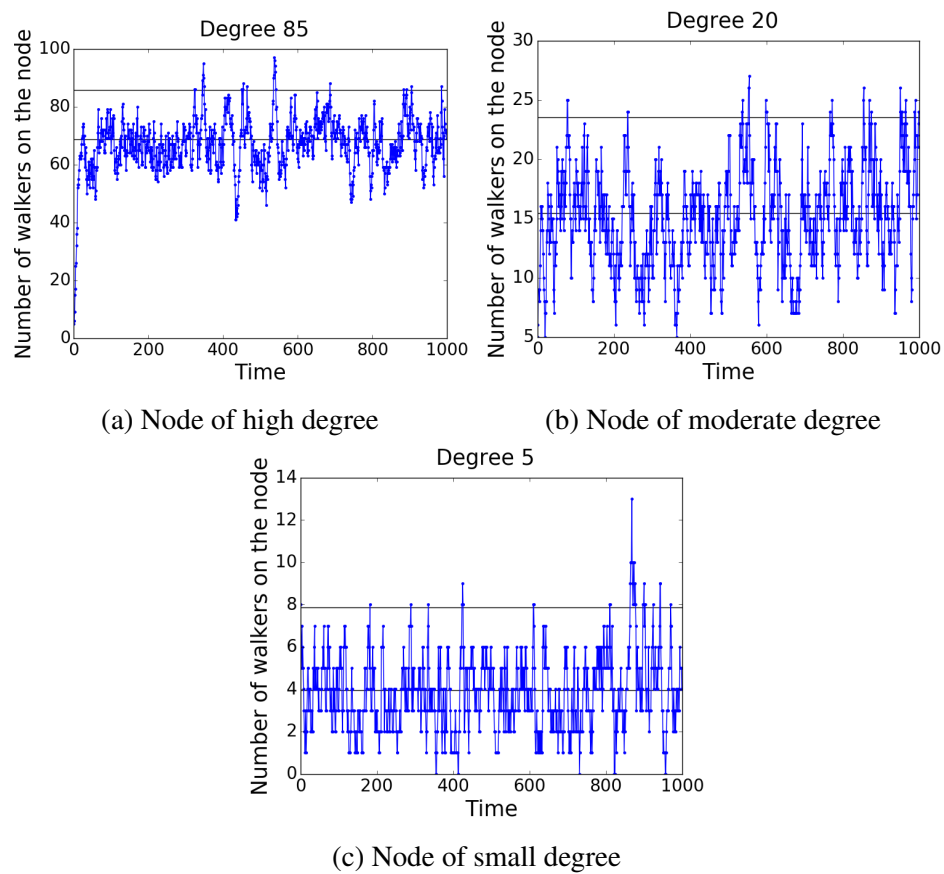


Fig. 4.2 Time series of number of walkers on nodes of three different degrees for the exponential law waiting time distribution case. The plots show the mean number of walkers μ and the threshold for an extreme event $\mu + 2\sigma$

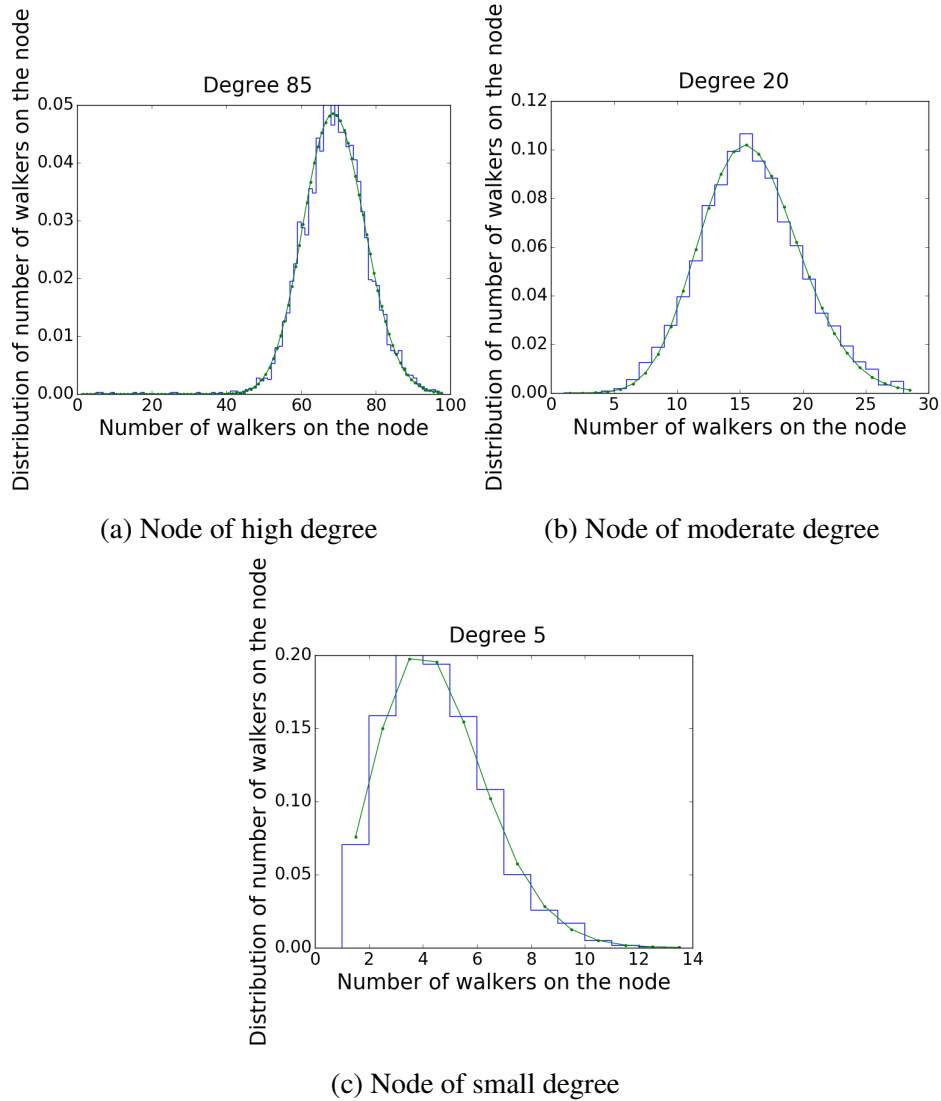


Fig. 4.3 Probability distribution of number of walkers on nodes of three different degrees for the power law waiting time distribution case. The histogram has been obtained from simulations of the random walk. The solid line is the analytic expression for the distribution which is given by the binomial distribution in Eq.4.3

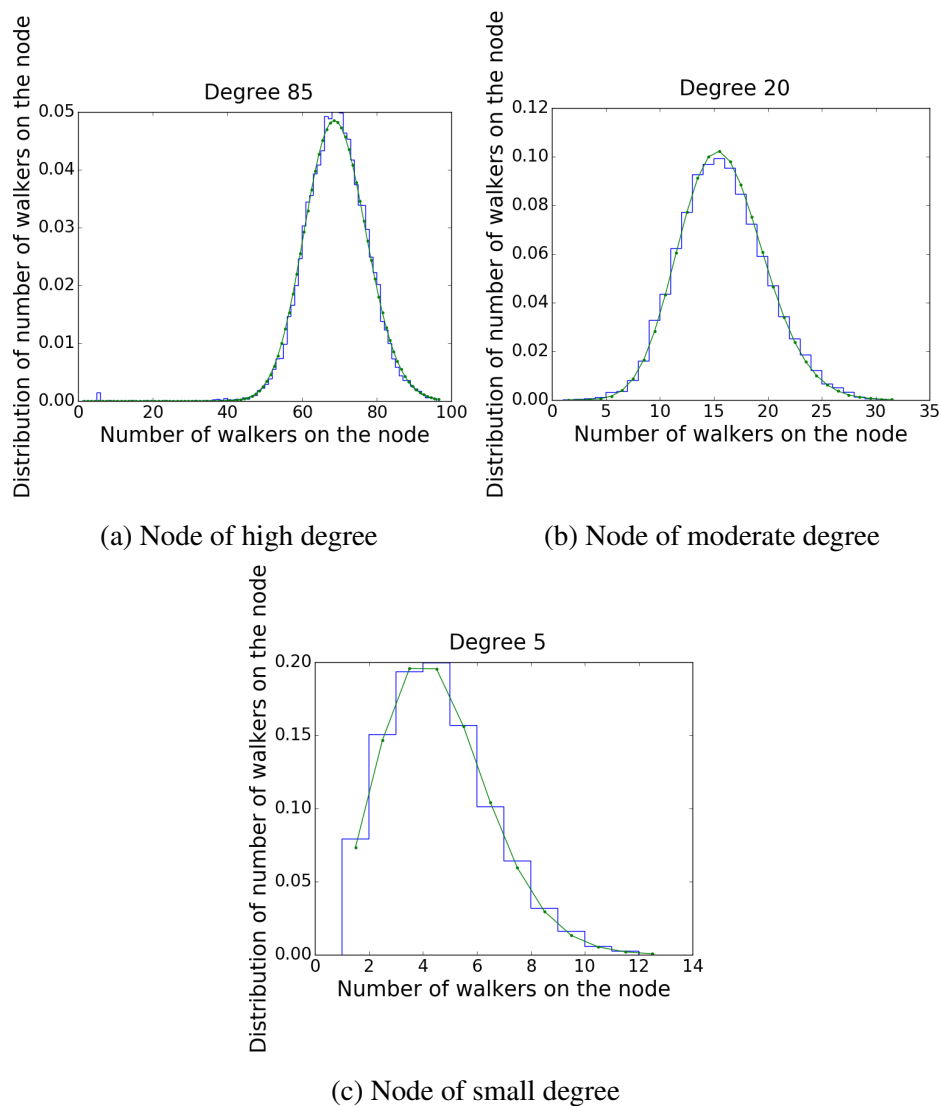


Fig. 4.4 Probability distribution of number of walkers on nodes of three different degrees for the exponential law waiting time distribution case. The histogram has been obtained from simulations of the random walk. The solid line is the analytic expression for the distribution which is given by the binomial distribution in Eq.4.3

The time series of number of walkers on a given node gives a sample of the random walk on that node, that is the dynamics of the random walk on the node (see Fig. 4.1 and Fig.4.2). From the plots of the time series of the random walks on the networks, one can see that the number of random walkers fluctuates around a certain value, which is the steady state value of the number of walkers on the node (μ). The threshold used for the definition on an extreme event, $q = \mu + 2\sigma$ has been shown in the figures. Only the time series plots for the continuous time random walks have been shown here. Plots for the discrete time case are similar to those of the continuous time cases.

Fig. 4.3 and Fig.4.4 show the probability distribution of the number of walkers on three different nodes for Pareto and exponential waiting time cases. These plots give a picture of the probability of finding a certain number of walkers on a given node at a given time. Since this has already been studied for discrete time random walks, only the plots for continuous time random walks have been shown here. The probability distribution of the number of random walkers on the nodes follow the expected binomial expression (Eqn. 4.3) for both discrete and continuous time random walks.

$$f_W(w) = \binom{N_W}{w} p^w (1-p)^{(N_W-w)} \quad (4.3)$$

where p is the stationary occupancy probability of the node, which is dependent on the degree of the node. The mean of this distribution (μ) for a node gives the number of walkers on that node at steady state.

4.3 Statistics of extreme events

Probability of an extreme event-

The probability of an extreme event as a function of the node was then computed for both discrete and continuous time random walk cases. These simulations were carried out on a scale free network with 1000 nodes, 3132 edges and with 5000 random walkers. Fig. 4.5 shows the probability of occurrence of an extreme event as a function of the degree of the nodes. From the plots, it is observed that the probability of occurrence of an extreme event is greater for nodes of smaller degrees than for nodes of higher degree. Even though flow through higher degree nodes is greater, the probability of the occurrence of an extreme event is smaller on these nodes than nodes with smaller degrees. (See Fig. 4.5).

In both discrete and continuous time random walk cases, the probability was computed by computing the number of times the number of walkers exceeded a threshold of $q = \mu + 2\sigma$,

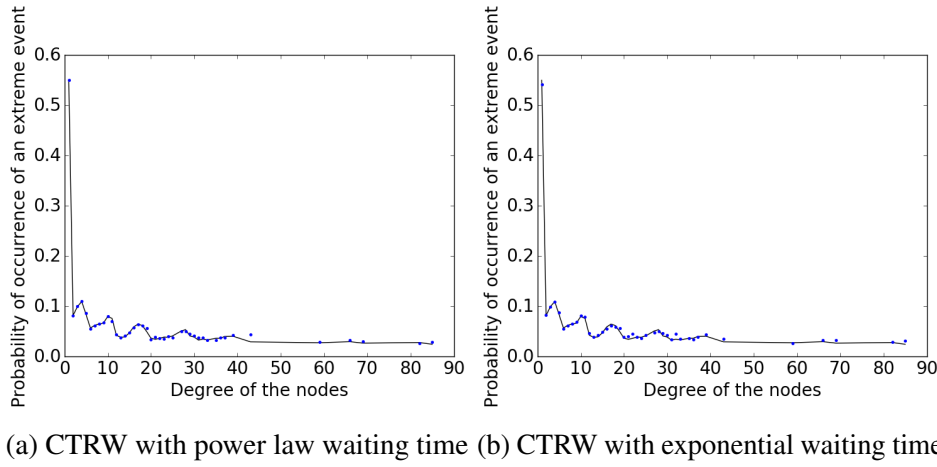


Fig. 4.5 Probability of the occurrence of an extreme event as a function of the degree of the node. The solid line is the analytic expression for the probability, which is given by Eqn. 4.4

where μ and σ are the mean and standard deviation of the binomial distribution given above. An analytic expression for the probability of the occurrence of an extreme event is then given by Eqn. 2.31,

$$P(W > q) = \sum_{j=\lfloor q \rfloor + 1}^{N_W} \binom{N_W}{j} p^j (1-p)^{(N_W-j)} \quad (4.4)$$

Correlation between difference in magnitudes of consecutive extreme events and the time between their occurrences-

The Pearson's correlation coefficient was computed between the difference in magnitudes of two consecutive extreme events and the time between the occurrence of the two. Simulations were carried out on three different networks with 1000 nodes and 5000 random walkers. A scale-free network with 3132 edges, a small-world network with 3000 edges and a random network with 3012 edges were considered. This was discussed in Section 2.2.2. The quantity that was computed is given by,

$$C_P = \frac{E[T_R H] - E[T_R]E[H]}{\sigma_{T_R} \sigma_H} \quad (4.5)$$

where, $H \in (Z, |Z|)$, Z denotes the difference in magnitudes of consecutive extreme events, $|Z|$ denotes the absolute difference in magnitudes of consecutive extreme events, T_R denotes the recurrence time, $E[T_R]$ and $E[H]$ denote the expectation values of the random variables T_R and H , while σ_{T_R} and σ_H denote the standard deviations of the variables. The Pearson's

correlation computed between two quantities takes a value between -1 and 1 . A significantly large negative value (closer to -1) indicates that the two quantities are anti-correlated, that is, a larger value of one quantity is associated with a smaller value of the other quantity. A significantly large positive value (close to 1) indicates that the two quantities are correlated, that is, a larger(smaller) value of one quantity is associated with a larger(smaller) value of the other quantity. A value close to 0 indicates that the two quantities are neither correlated nor anti-correlated.

The recurrence time $T_{R_n}, n = 1, 2, \dots, N_E - 1$ was found to be exponentially distributed from simulations of discrete time random walks on scale-free, small-world and random networks. The distribution of T_{R_n} is given by,

$$P(T_{R_n} = t) = \lambda e^{-\lambda t} \quad (4.6)$$

where, $n = 1, 2, 3, \dots, N_E - 1$ the constant λ depends on the node that is being considered. From existing literature, we know that λ is given by $\lambda = \frac{1}{P(W > q)}$ [17]. The distributions of recurrence time for discrete time random walks on three nodes of a scale-free network with 1000 nodes have been shown in Fig. 4.6.

The distribution of the difference in magnitudes of consecutive extreme events, Z , was calculated in Section 2.2.2. The distribution of Z , by considering the Gaussian approximation of the binomial was obtained as (Eq. 2.46),

$$P(Z = z) = \frac{c^2}{2\sqrt{\pi}\sigma} e^{-\frac{z^2}{4\sigma^2}} \left[1 - \Phi_{((\mu - \frac{|z|}{2}), (\frac{\sigma}{\sqrt{2}})^2)}(q) \right], \forall z \quad (4.7)$$

The distribution of the absolute value of the difference was obtained as (Eq. 2.48),

$$P(|Z| = z) = \begin{cases} \frac{c^2}{\sqrt{\pi}\sigma} e^{-\frac{z^2}{4\sigma^2}} \left[1 - \phi_{((\mu - \frac{|z|}{2}), (\frac{\sigma}{\sqrt{2}})^2)}(q) \right] & z > 0 \\ \frac{c^2}{2\sqrt{\pi}\sigma} \left[1 - \phi_{(\mu, (\frac{\sigma}{\sqrt{2}})^2)}(q) \right] & z = 0 \end{cases} \quad (4.8)$$

The correlation coefficient was computed for the two cases, one in which the the absolute value of the difference in magnitudes of consecutive extreme events was taken and the other case in which the magnitude of the previous event was subtracted from the magnitude of the later event. Simulations were carried out on $N = 1000$ node scale-free, small-world and random networks (Fig.4.9, Fig. 4.10 and Fig. 4.11) and we are currently working on

obtaining a semi-analytic expression for the same. When the difference between magnitudes (Z) of consecutive events is considered, the correlation is close to 0 for all three types of networks. This indicates that Z and T_R are uncorrelated. However, when the absolute difference between magnitudes ($|Z|$) is considered, a strong negative correlation is observed between the two quantities. This indicates that the two variables ($|Z|$ and T_R) are strongly anti-correlated, which allows some degree of predictability for the system. A large (small) difference in magnitudes of consecutive extreme events is associated with a small (large) recurrence time. One might be able to comment on the magnitude or time of occurrence of a future extreme event based on the magnitudes of extreme events in the past and the times at which they occurred.

Fig. 4.12 compares the Pearson's correlation coefficient that was computed analytically and through simulations. When Z is considered while computing the coefficient, simulations show that the value of the coefficient is close to 0. Semi-analytic values for the coefficients, however, show a moderate negative correlation, for some nodes of small degrees. However, when $|Z|$ is considered while computing the coefficient, the analytic and simulation results match closely.

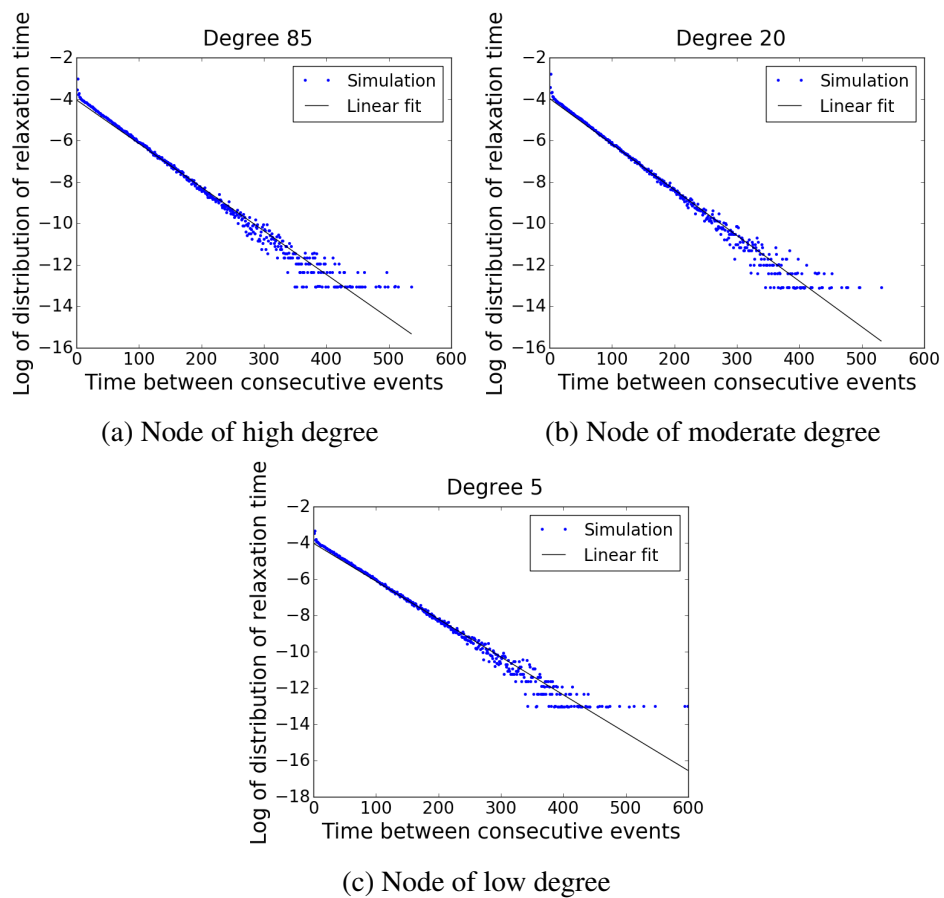


Fig. 4.6 Distribution of relaxation time (semi-log plot) for three nodes in the network. The solid line gives the linear fit for the plot, which shows that the distribution is exponential as given by Eq. 4.6

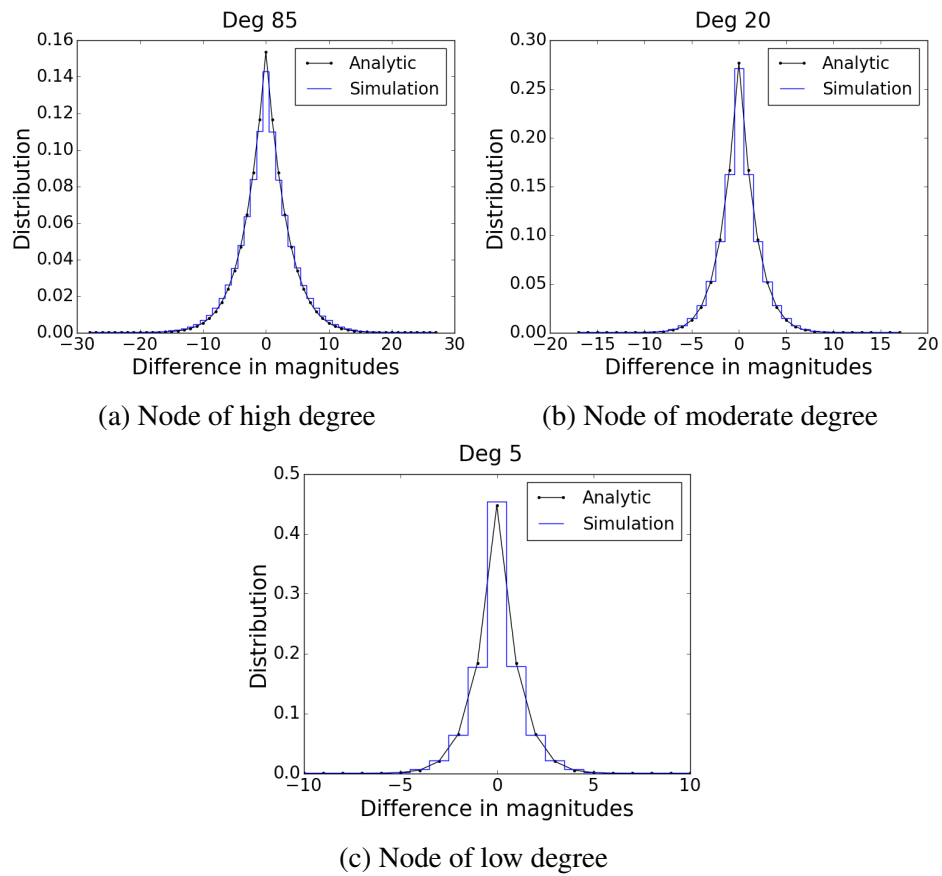


Fig. 4.7 Distribution of the difference in magnitudes of consecutive extreme events on a node (Z). The solid line gives the analytic expression for the distribution, which is given by Eq. 4.7

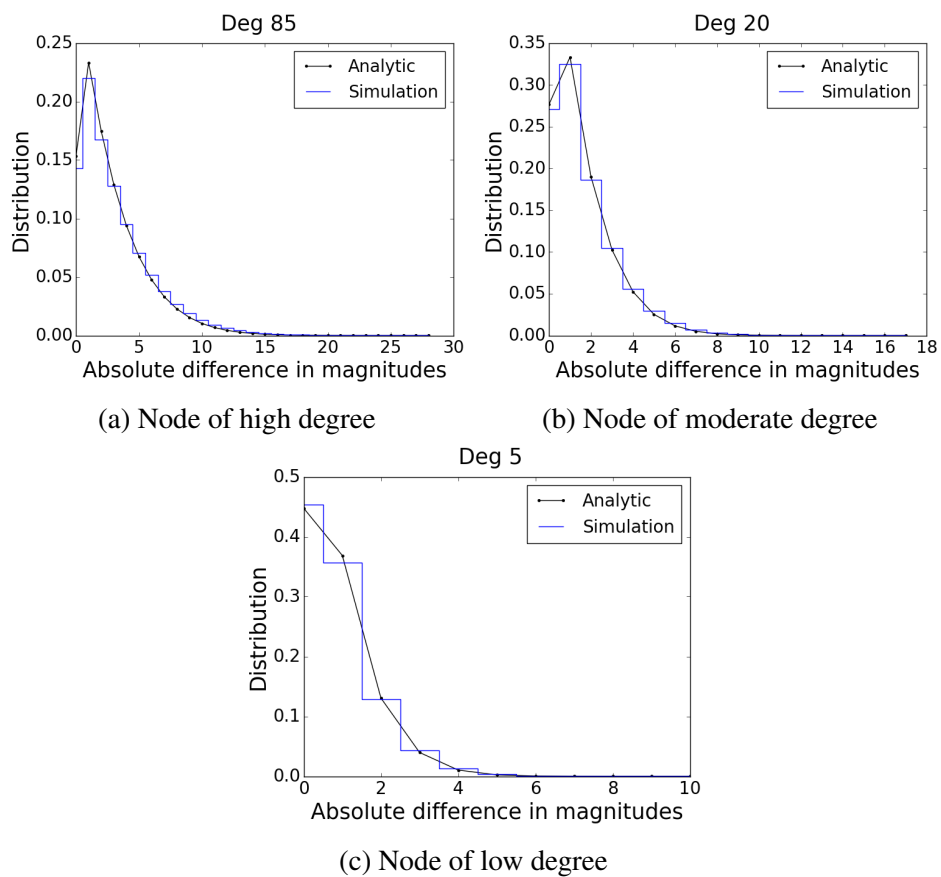


Fig. 4.8 Distribution of the absolute difference in magnitudes of consecutive extreme events on a node ($|Z|$). The solid line gives the analytic expression for the distribution, which is given by Eq. 4.8

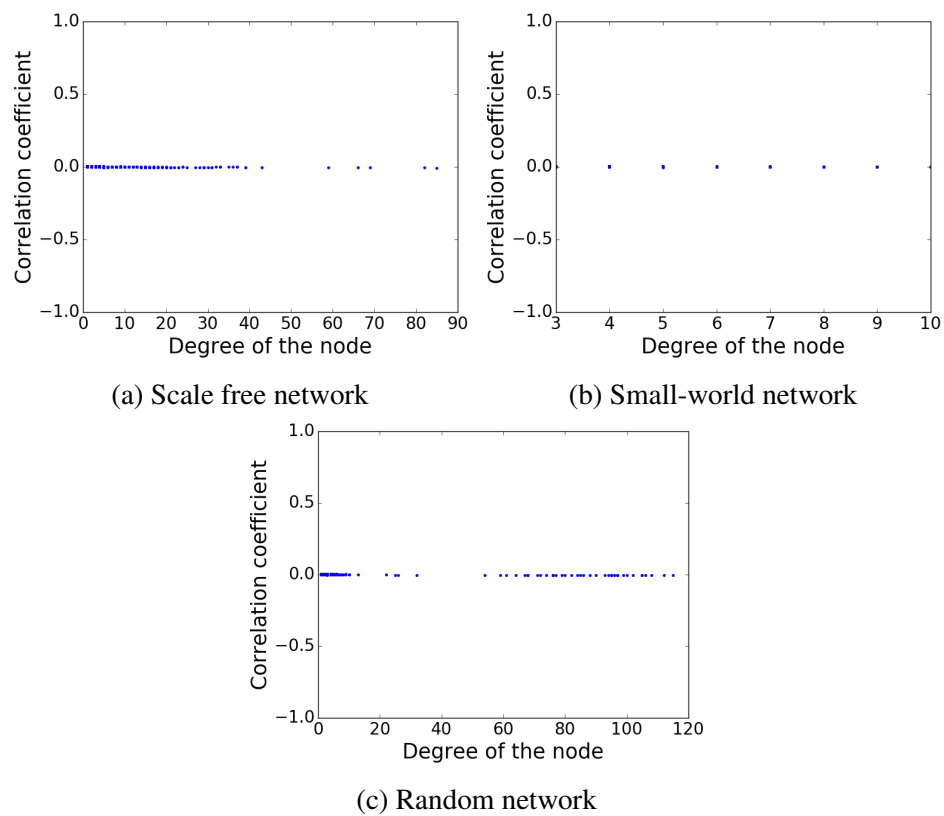


Fig. 4.9 Pearson's correlation coefficient between the magnitude difference of two consecutive extreme events and the time interval between the occurrence of the two.

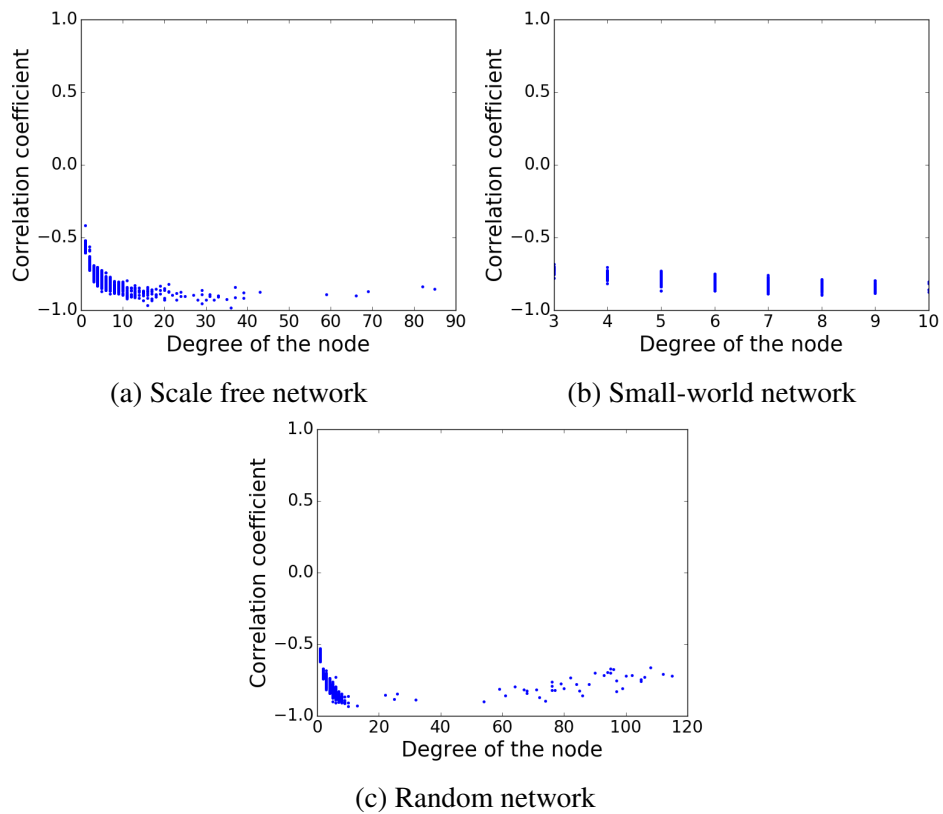


Fig. 4.10 Pearson's correlation coefficient between the absolute magnitude difference between two consecutive extreme events and the time interval between the occurrence of the two.

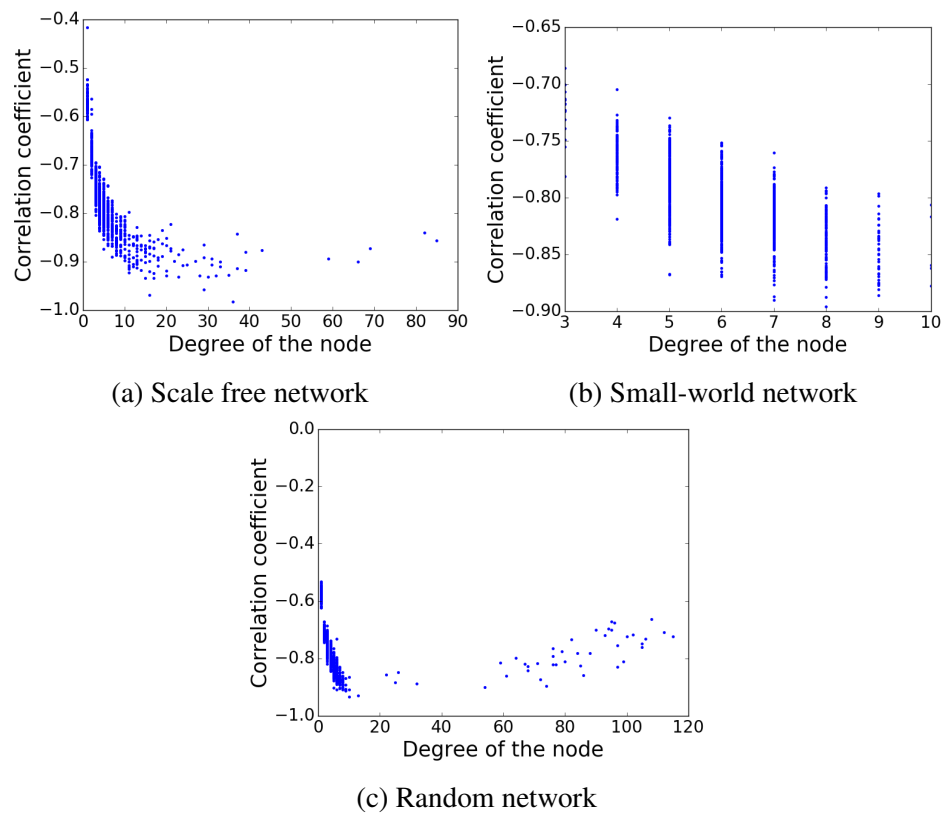


Fig. 4.11 Pearson's correlation coefficient between the absolute magnitude difference between two consecutive extreme events and the time interval between the occurrence of the two.

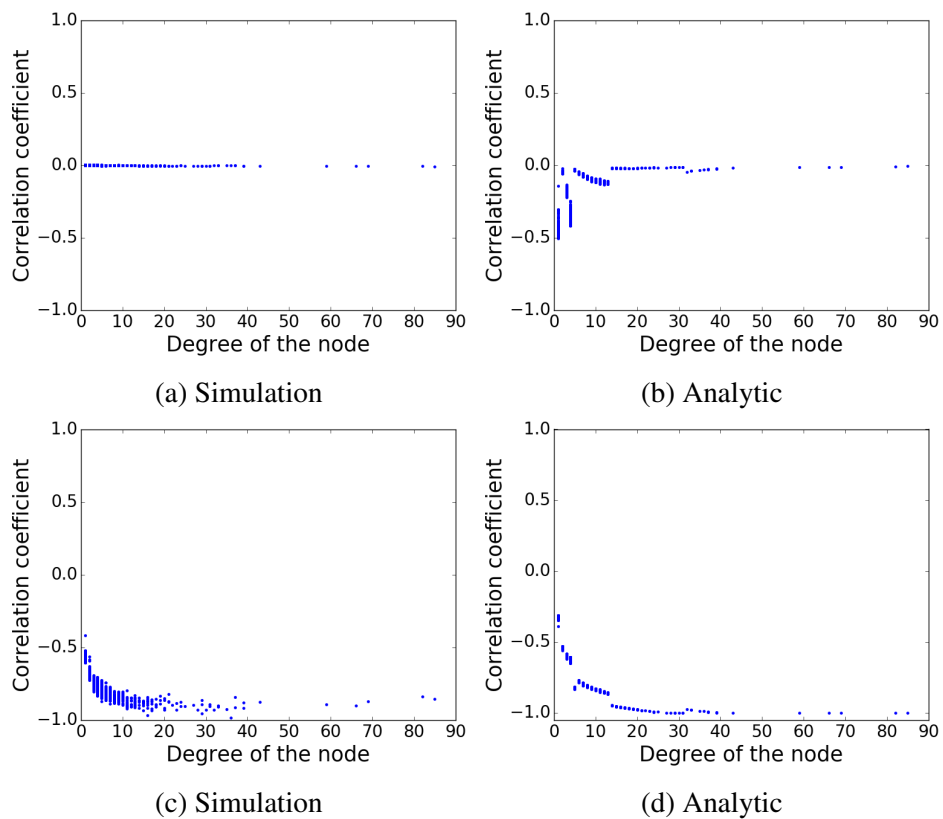


Fig. 4.12 Pearson's correlation coefficient computed analytically and through simulations for a scale-free network. The top two plots show the correlation when the difference between magnitudes of consecutive extreme events Z is considered. The bottom two plots show the correlation when the absolute value of the difference ($|Z|$) is considered.

Chapter 5

Outlook

In this chapter, the work that was done as a part of the thesis is briefly summarized and further work that can be done in this area has been outlined. In this project, discrete and continuous time random walks on networks were studied and the stationary occupancy probability densities were obtained. By considering multiple independent random walkers on the network, a definition for an extreme event was given as the event that occurs when the number of walkers on a particular node exceeds a certain threshold. Some statistics of the extreme events were studied. The study of extreme events with this definition of an extreme event already exists in literature for discrete time random walks on networks [17, 18]. The work was extended to continuous time random walks by first obtaining an expression for the stationary probability density and then studying extreme events defined based on these random walks. A correlation statistic was defined and was computed analytically and through simulations for the discrete time random walk case. In the future, it can be computed for the continuous time random walk case as well.

This analysis can be extended to other types of random walks on networks. It might be interesting to study biased continuous time random walks, in which the transition jump density of a random walker depends to some extent on the degree of the node to which the walker jumps [23]. In this project, we looked at uncoupled continuous time random walks in which the jump density and the waiting time density are independent. Some physical processes are better modeled by coupled continuous time random walks, in which the jump density and waiting time density are coupled and cannot be treated as independent random variables. Another type of random walk that one could consider is a random walk in which the waiting time density is dependent on some property of the node from which it makes the jump.

In this project, only static networks have been considered. However, real-world complex networks are hardly static. In the future, one could extend the study of extreme events to time dependent networks as well. One of the major problems that one would encounter is that the occupancy probability of a node may never reach a stationary value. It might be interesting to look at how different properties of the network, such as its degree distribution, centrality measures, etc. change with time. In this case, the above formalism would have to be modified to take into account the time dependency of the network. Some of the recent research in this field has taken into account the time dependency of networks by modeling the time dependency through the waiting time distribution of continuous time random walks[?]. Thus, by appropriately modeling the waiting time distribution, one can study time dependent networks in which the edges keep appearing or disappearing in time.

References

- [1] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2, 2007.
- [2] Larry Page, S Brin, R Motwani, and T Winograd. Pagerank: Bringing order to the web. Technical report, Stanford Digital Libraries Working Paper, 1997.
- [3] Bin Jiang. Ranking spaces for predicting human movement in an urban environment. *International Journal of Geographical Information Science*, 23(7):823–837, 2009.
- [4] Mark Newman. *Networks*. Oxford university press, 2018.
- [5] Mark Newman, Albert-Laszlo Barabasi, and Duncan J Watts. *The structure and dynamics of networks*, volume 12. Princeton University Press, 2011.
- [6] Eric Davidson and Michael Levin. Gene regulatory networks. *Proceedings of the National Academy of Sciences*, 102(14):4935–4935, 2005.
- [7] John Mingers and Loet Leydesdorff. A review of theory and practice in scientometrics. *European journal of operational research*, 246(1):1–19, 2015.
- [8] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [9] Soon-Hyung Yook, Hawoong Jeong, and Albert-László Barabási. Modeling the internet’s large-scale topology. *Proceedings of the National Academy of Sciences*, 99(21):13382–13386, 2002.
- [10] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440, 1998.
- [11] P ERDős. Rényi, a.: " on random graphs. *I*". *Publicationes Mathematicae (Debre)*, 1959.
- [12] James Gratwicke, Marjan Jahanshahi, and Thomas Foltynie. Parkinson’s disease dementia: a neural networks perspective. *Brain*, 138(6):1454–1476, 2015.
- [13] Elliott W Montroll and George H Weiss. Random walks on lattices. ii. *Journal of Mathematical Physics*, 6(2):167–181, 1965.
- [14] Elliott W Montroll. Random walks on lattices. iii. calculation of first-passage times with application to exciton trapping on photosynthetic units. *Journal of Mathematical Physics*, 10(4):753–765, 1969.

- [15] Elliott W Montroll and Harvey Scher. Random walks on lattices. iv. continuous-time walks and influence of absorbing boundaries. *Journal of Statistical Physics*, 9(2):101–135, 1973.
- [16] Yu-Zhong Chen, Zi-Gang Huang, and Ying-Cheng Lai. Controlling extreme events on complex networks. *Scientific reports*, 4:6121, 2014.
- [17] Vimal Kishore, MS Santhanam, and RE Amritkar. Extreme events on complex networks. *Physical review letters*, 106(18):188701, 2011.
- [18] Vimal Kishore, Abhijeet R Sonawane, and MS Santhanam. Manipulation of extreme events on scale-free networks. *Physical Review E*, 88(1):014801, 2013.
- [19] Niklas Boers, Bodo Bookhagen, Henrique MJ Barbosa, Norbert Marwan, Jürgen Kurths, and JA Marengo. Prediction of extreme floods in the eastern central andes based on a complex networks approach. *Nature communications*, 5:5199, 2014.
- [20] Naoki Masuda, Mason A Porter, and Renaud Lambiotte. Random walks and diffusion on networks. *Physics reports*, 716:1–58, 2017.
- [21] Amy N Langville and Carl D Meyer. *Google’s PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.
- [22] S Harshini Tekur, Udaysinh T Bhosale, and MS Santhanam. Higher-order spacing ratios in random matrix theory and complex quantum systems. *Physical Review B*, 98(10):104305, 2018.
- [23] Vimal Kishore, MS Santhanam, and RE Amritkar. Extreme events and event size fluctuations in biased random walks on networks. *Physical Review E*, 85(5):056120, 2012.
- [24] Madan Lal Mehta. *Random matrices*, volume 142. Elsevier, 2004.
- [25] Ashraf A Abul-Magd and Adel Y Abul-Magd. Unfolding of the spectrum for chaotic and mixed systems. *Physica A: Statistical Mechanics and its Applications*, 396:185–194, 2014.
- [26] Camellia Sarkar and Sarika Jalan. Spectral properties of complex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(10):102101, 2018.
- [27] Ginestra Bianconi. Spectral properties of complex networks. *arXiv preprint arXiv:0804.1744*, 2008.
- [28] Luis EC Rocha and Naoki Masuda. Random walk centrality for temporal networks. *New Journal of Physics*, 16(6):063023, 2014.
- [29] Hao Liao, Manuel Sebastian Mariani, Matúš Medo, Yi-Cheng Zhang, and Ming-Yang Zhou. Ranking in evolving complex networks. *Physics Reports*, 689:1–54, 2017.
- [30] Philip S Yu, Xin Li, and Bing Liu. Adding the temporal dimension to search—a case study in publication search. In *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI’05)*, pages 543–549. IEEE, 2005.

-
- [31] Hyounghick Kim and Ross Anderson. Temporal node centrality in complex networks. *Physical Review E*, 85(2):026107, 2012.
- [32] James P Keener. The perron–frobenius theorem and the ranking of football teams. *SIAM review*, 35(1):80–93, 1993.
- [33] CN Angstmann, IC Donnelly, BI Henry, and TAM Langlands. Continuous-time random walks on networks with vertex-and time-dependent forcing. *Physical Review E*, 88(2):022811, 2013.
- [34] Till Hoffmann, Mason A Porter, and Renaud Lambiotte. Generalized master equations for non-poisson dynamics on networks. *Physical Review E*, 86(4):046102, 2012.
- [35] Christopher N Angstmann, Isaac C Donnelly, and Bruce I Henry. Continuous time random walks with reactions forcing and trapping. *Mathematical Modelling of Natural Phenomena*, 8(2):17–27, 2013.
- [36] MS Santhanam and Holger Kantz. Return interval distribution of extreme events and long-term memory. *Physical Review E*, 78(5):051113, 2008.
- [37] YY Atas, E Bogomolny, O Giraud, and G Roux. Distribution of the ratio of consecutive level spacings in random matrix ensembles. *Physical review letters*, 110(8):084101, 2013.

