Study of Dynamical Processes on Complex Networks

A THESIS

submitted for the award of Ph.D. degree of

MOHANLAL SUKHADIA UNIVERSITY

in the

Faculty of Science

by

Vimal Kishore

Under the Supervision of

M. S. Santhanam

Ex Reader, Physical Research Laboratory

Ahmedabad, India

and

Ravindra E. Amritkar

Professor, Physical Research Laboratory

Ahmedabad, India

DEPARTMENT OF PHYSICS

MOHANLAL SUKHADIA UNIVERSITY

UDAIPUR

Year: 2012

All our knowledge brings us nearer to our ignorance…

–T. S. Eliot

Dedicated to,

My Family

DECLARATION

*I **Mr. Vimal Kishore**, S/o Mr. Mannu Lal Saxena, resident of Room no. 204, Thaltej Hostel, Bodakdev, Ahmedabad, hereby declare that the work incorporated in the present thesis entitled, **"Study of Dynamical Processes on Complex Networks"** is my own and original. This work (in part or in full) has not been submitted to any University for the award of a Degree or a Diploma.*

Date :

(Vimal Kishore)

CERTIFICATE

I feel great pleasure in certifying that the thesis entitled, **"Study of Dynamical Processes on Complex Networks"** embodies a record of the results of investigations carried out by Mr. Vimal Kishore under my guidance.

He has completed the following requirements as per Ph.D. regulations of the University.

(a) Course work as per the university rules.

(b) Residential requirements of the university.

(c) Presented his work in the departmental committee.

(d) Published minimum of two research papers in a referred research journal.

I am satisfied with the analysis of data, interpretation of results and conclusions drawn.

I recommend the submission of thesis.

Date :

Madabushi Srinivasan Santhanam

Ex-Reader

(Thesis Supervisor)

Countersigned by

Head of the Department

Acknowledgements

I get by with a little help from my friends.

–John Lennon, Paul McCartney

This thesis, an extreme event in my life, is an outcome of my random walk on the path leading to PhD. It took almost six years for me to travel this road which has always been little less traveled. The journey of was full of many felicitous moments because I had the privilege to work with valuable colleagues, to enjoy the friendship of many and the love of the ones most close to me. I can not mention them all, neither I can acknowledge them enough, but all I can say surely is that my true acknowledgements are ampler than these few pages.

First and foremost, I thank my advisor and my mentor, Dr. M. S. Santhanam, for his guidance, patience, and instructions in the journey of this thesis. He not only provided me the freedom and enough confidence but also encouraged me to explore the world of new ideas. His technical writing skills continue to strive me to live up to his high standards. I am indebted to Santhanam for providing me countless opportunities and unending support in all these six years.

It has been an honor working with Prof. R. E. Amritkar, my co-advisor, whose inputs were remarkably effective in shaping up my thesis as well as me as a person. He taught me how to question thoughts and express ideas. He gave me courage to explore on my own and at the same time the guidance to recover when my steps faltered. I am deeply grateful to his insightful comments, constructive criticism and enforcing strict validation for each research result.

I owe thanks to Prof. V. K. B. Kota, Dr. Navinder Singh, Dr. P. Sharma (ex-Head academics) and all the academic committee members for their inputs during the progress of this work. Also, I am thankful to the entire theoretical physics division for creating a healthy environment, necessary for research. In particular, I would like to thank Dr. Dilip Angom, Dr. Namit Mahajan and Dr. R. P. Singh for offering good advice at critical points along the way.

I have spent a part of my PhD life in oscillating between PRL Ahmedabad and IISER Pune. IISER Pune has been very supportive during my entire PhD track. Specifically, I would like to thank IISER for its prompt hospitality.

It is my duty to acknowledge High Performance Computing (HPC) facilities at PRL Ahmedabad and IISER Pune. The contribution of both the machines is immense in my research work. I am also grateful to all the staff-members of PRL and IISER Pune for their assistance and cooperation. Specifically, I would like to acknowledge the support of Jigar Bhai, Parul Madam, Ghanshayam Bhai, Ranganathan Ji, Senthil Bhai and Prabhakar (IISER Pune) who have been highly efficient.

During the ups and downs of my PhD life, I always had a special source of enjoyment and support: My evergreen friends from PRL. Amzad, Anand, Arvind, Ashok, Ashwini, Bhaswar, Ketan, Naveen, Pankaj, Rabiul, Rohit, Sandeep, Soumya, Sreekanth, Suman, Tarun and Vineet, . . . : Cheers to all the time killing *chai* sessions and several parties which brought ultimate fun to us.

I would like to express my thank to Subimal da, Suman, Soumya and Sandeep for solving my Linux and other computation related problems. Sincerity of Ketan and Ashok, which made my life easier in many ways, is highly appreciable. I want to thank Harinder, Suman, Swarna (BHU), Rajeev (IMSc) and Rashmi (IISER) for the great time we had in every single school/conference.

The memories of time spent in Tulip with Anand, Rabiul, Pankaj are still fresh in my mind. My stay at PRL has been made more enjoyable due to the lively atmosphere at hostel. Company of Kushwaha ji, Sir Ritesh, Bhavik, Santosh, Akhilesh, Alok, Rahman, Shreyas, Gyan, Kirpa, Patra, Vaity, Saxena and Sumanta at PRL main Hostel was enjoyable. I also cherished the time spent with the younger friends at Thaltej hostel. In particular, I cannot forget the taste of delicious food cooked by Negi, Yashpal, and Sharma. Teasers from Raju, Tomar and Girish were always served for dessert. I also thank Arvind, Ashwini, Vineet, Julfi and Pankaj for all the happy times together at Thaltej Hostel. I extend my thank to Puja, Neeta, Vikas, Anwar, Siddhu for giving us some wonderful moments.

I must not forget the support from my colleagues with whom I spent relatively less but valuable time towards the end of my thesis. In particular, I would like to extend my thanks to Shivani, Mala, Basak, Siddhu, Shashi, Gulgul, Mano, Gulab, Khan, Suneel and Midhun. I would like to acknowledge Viral for sharing some really tasty food with me.

During my time at IISER Pune, I could experience the professionalism and the warmth of the PhD students. My heartfelt thanks go especially to Kanika, Mayur and Rashmi and other members of physics group. In particular, I would like to acknowledge Kanika for all the proof readings, cups of tea and the friendship extended to me in spite of our disagreements over every possible topic where two opinions can be formed. She proved to be a very good friend, a consistent dinner partner and a wonderful colleague. Along with her, I would like to thank Harsha, Payal, Rini, Mithila, Abhishek and JP for all the serious and less-serious conversations and especially for all the non-sense that made me wiser. I am sure, our dinner time collaborative projects will leave the dinner table and see the day-light someday.

I mainly met these people mentioned above during my PhD trajectory but my old friends have played an important role in this journey. Abhishek, Amit, Smriti, Kallol, Shyamal and Tapan; a heartfelt thank for sharing the fun, enthusiasm, frustration, curiosities and what not.

It is difficult to express the profound sense of gratitude that I feel towards my family who have been unfailingly supportive throughout my life. No matter how small, my personal treasure of triumphs has all been possible because of my parents' faith in me. The acknowledgment remains incomplete without mentioning the cute smile of my niece Anushka. She has added an extra charm to my life.

List of Publications

- Extreme Events on Complex Networks,
  Vimal Kishore, M. S. Santhanam and R. E. Amritkar,
  Phys. Rev. Lett **106**, 188701 (2011).
  arXiv:1102.1789 [cond-mat.stat-mech]

- Extreme Events and Event-size Fluctuations in Biased Random Walks on Networks,
  Vimal Kishore, M. S. Santhanam and R. E. Amritkar
  Phys. Rev. E **85**, 056120 (2012).
  arXiv:1112.2112 [cond-mat.stat-mech]

Abstract

Networks are everywhere and most of the the networks support a number of transport processes. Fluctuations in traffic flow constitute one of the main factor affecting the dynamics of these systems. Fluctuations above a certain threshold can be labeled as extreme events. This definition of extreme events as exceedances above a prescribed quantile is not necessarily related to the constraints imposed by the capacity of the node. It arises from the natural fluctuations in the traffic passing through a node. The transport model that we have adopted is the random walk on complex network. Thus, in this thesis we place our results in the context of both the random walks and extreme events in a network setting.

In the case of a simple random walk, we show that the small degree nodes of a network are more likely to encounter extreme events than the hubs. The result remains unchanged even with the use of shortest path strategy on networks. We also obtain the extreme event probability in the case of topologically biased random walk and show that biasing the traffic towards hubs can increase the risk of bottlenecks on networks.

Using the above notion of extreme events, we study the nature of failure of a network by removing nodes which experience an extreme event and redistributing the walkers on the remaining or active nodes. We find that in an all-to-all network, cascade failures cause the sudden collapse of the network.

This thesis, as a whole, is an attempt to understand extreme events occurring on the nodes due to flow on networks. It discusses the importance of lower degree nodes in scale free network and presents a different but natural mechanism for the complete failure of a network. The work done in the thesis can help in designing the networks which will be better prepared to meet the expected extreme events. However, extreme events discussed here being due to inherent fluctuations will nevertheless take place and can not be avoided.

Contents

Chapter 1

Introduction

1.1 Motivation

We live in an environment which is full of extremely complicated systems. Our life unfurls in a society, whose functioning depends on co-operation among billions of individuals. Every day we use a transport system for commuting or the transportation of various physical entities, a communication system consisting of cell phones, computers, internet etc. Our existence depends on the interaction of genes and metabolites, forming metabolic network. Our thinking ability is an outcome of the interactions among billions of neurons. Behind all these systems, there are complicated networks which keep the information about the interactions between different components. Smooth functioning of all these systems must result in a hassle-free life, this is what we desire. But we always find disruptions in these systems. Burst of unwanted activities in societies, traffic congestions on transport networks, slow down of internet, non-availability of particular websites, physical illness or epileptic seizure in brain, they all exist in real world. In an individual system, such incidents may not be very frequent but still they account for great losses. Because of the strong impact and low frequency of these special events, the tag 'extreme events' seems appropriate.

Every time while facing these extreme events we wonder about the origin of such extreme events. They may arise because of some external disturbances or the dynamics of the system itself can give birth to such events. Their origin may also be due to the combinations of both. Most of the times, we do not have any control over external shocks but the system can be secured in some cases against these external disturbances. On the other hand, extreme events arising due to the internal dynamics are unavoidable. Of course, if the dynamics stops, these events will not occur but that is a trivial solution. Hence we need to come up with a solution where at least the occurrences or consequences of extreme events can be reduced without stopping the dynamics. For this, at first hand we need to understand the network structures, dynamical processes and the coupling between them. Then, we need to device a criterion to separate the extreme events from the normal events. Only after doing all this, we can hope to understand the role of network structure in the origin and the dynamics of extreme events.

1.2 Aim of the thesis

The accurate prediction of extreme events has been a dream of humans down the ages but that is not the main aim of this thesis. The specific aim of this thesis is to prepare a framework for studying the extreme events taking place on a network and to understand the role of network structure, which is a step taken in the direction of realizing our dream.

As mentioned earlier, there can be different genesis for extreme events, here we concentrate on the extreme events whose origins lie in the dynamics itself. The dynamics we choose to study extreme events is the transport process, which is an important process not only in physics but also in many other disciplines like biology, social sciences etc. Hence, the first and important aim is the study of transport mechanism on networks.

The earlier work done on extreme events was not on networks, hence we lack the criterion for calling an event as extreme event in network settings. We should define a natural criterion based on the network properties for differentiating the extreme events and the normal events. To predict extreme events on network, we should take a step forward from defining the extreme events to studying their occurrences in terms of network properties. To correlate the extreme events with the parameters related to network structure is the central part of the thesis.

Extreme events are often disastrous and when they occur on networks, they may even destroy the network itself. With this scenario, the question arises if a network can survive against such extreme events. Studying robustness of the network against extreme events is an important part of the thesis. The thesis also includes the nature of failures on the network, if the network fails. This can be helpful in stopping the spread of extreme events. The understanding gained from these studies can be useful in designing the network which can handle the extreme events smoothly. It is a small but important step in making the life better.

Following the preface of the thesis, in the next part we provide a brief introduction of the essential ingredients used in studying extreme events on complex networks. An outline of the thesis is presented in the last section of this chapter.

1.3 Prelude

This section includes the overviews of two different fields; networks and extreme events. They together form the basis of this thesis. The first section is about the networks and the terminology used in the field of network science. Next, there is a small discussion regarding the dynamical processes taking place on complex networks. And then, the field of extreme events is presented in a nutshell.

1.3.1 Complex networks

Networks are everywhere. They can be as noticeable as highways or subway systems, power grids, the internet or more abstract ones such as friendship network, ecological network or co-occurrence networks. The list of networks can be very long, as many of the systems can be simplified using the network description without loosing the complexity. This section provides a brief history of networks followed by the terminology of network science, properties of real world networks and their mathematical models. There are very good reviews on the network science and we would like to suggest the reviews by Barabási [1], Boccaletti [2], Newman [3], Barthélemy [4] and a book by Dorogovtsev [5].

History

The very first example of network based approach to problems dates back to year 1736 when Swiss mathematician Leonhard Euler solved the so called *Königsberg bridge problem* by reducing the problem to a graph - the set of nodes and links (see Fig. 1.1). After Euler's work, for almost 200 years, mathematicians developed the theory of graphs known as *graph theory* and around year 1920 social scientists started using the network representation for quantifying the relationships among social entities and developed tools for social network analyses. In year 1959, Paul Erdős and Alfréd Rényi defined random graph, a graph with n links chosen randomly among N nodes, and studied structural properties of these graphs. Later empirical results drawn from small data sets gave the hint that real world networks are not completely random and they have some structure in them. In late 1990, increasing computing power and emergence of large databases of real world systems enabled scientists to analyze networks with thousands or millions of nodes. Empirical results revealed that there are unifying principles and statistical properties common to completely different real world systems ranging from social to biological sciences. In year 1998, Watts and Strogatz proposed a simple model of real world networks and a year after in 1999, Barabási and Albert proposed a mechanism behind evolution of the real world networks. After these two seminal papers, field of complex networks has developed at a brisk pace and has seen a lot of advances mainly due to its interdisciplinary nature but the spirit, basic terminology and the basic structure of networks remains similar.

Glossary and notations

Historically, networks have been studied by the mathematicians under the term called a *graph theory*. A structure consisting of a set of entities having pairwise connections, is called *graph* in mathematical literature while in other sciences, the word *network* is extensively used. The set of entities is called *nodes or vertices* denoted by N and the pairwise connections are labeled as *edges or links* denoted by E. The set of N nodes and E edges define a graph $G = (N, E)$. A node is drawn as a dot while an edge is drawn as a line connecting two nodes, called *endpoints*.

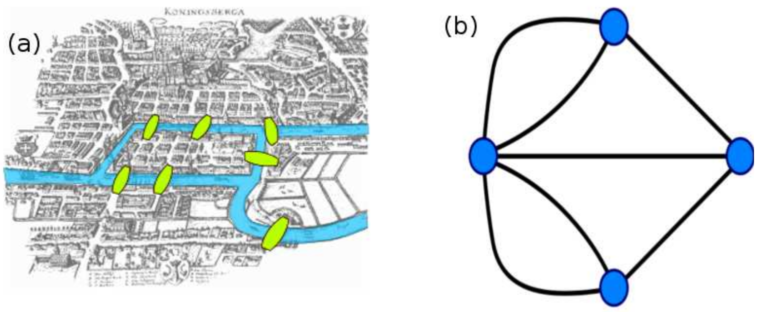

Figure 1.1: (a) Könisberg city had seven bridges connecting the four different land masses and the challenge was to take a walk through the town via all seven bridges but any bridge should not be crossed twice. (b) Mathematical representation of the city used by Euler. Each land mass is represented by a node (solid blue dots) and a bridge by a link (black lines). Euler proved that it is impossible to win this challenge. *Pictre Source: wiki.*

A *loop* is an edge having both the endpoints as a single node. If in between two endpoints, there are more than one links, the edge is called as the *multiple*. A graph without any multiple edges and loops is called *simple graph* otherwise it is a *multiple graph*. In general, graph is considered as simple graph, unless mentioned otherwise.

A central concept in graph theory is a *path or walk* which is a sequence of nodes $n_1, n_2, n_3 ...$ such that from each of its nodes there is a link to the next node in the sequence. A path may be infinite but if the path is finite, there is a start node and an end node. On a *simple path*, no two nodes appear twice. A graph is called *connected* if there exists a path in between every pair of nodes in the graph, otherwise the graph is *disjoint*. If the start node and the end node of a simple path are same, the path is called a *cycle*. The *length* of a path is the number of edges in it, counting multiple edges multiple times. For a single node, the path is of length 0. The smallest possible length in between two nodes (n_i, n_j), is called the distance between the nodes n_i and n_j and is denoted by $d(n_i, n_j)$. The Set of nodes at distance 1 from node n_i is called the neighborhood of n_i and is denoted by Γ_{n_i}. The maximum distance over the graph is called the *diameter* of the graph while minimum distance is the *radius* of the graph.

A cycle with an odd length is an *odd cycle* otherwise it is an *even cycle*. A graph without odd cycles is a *bipartite graph*. A graph without any cycle is an *acyclic graph.* A *tree* is a connected acyclic simple graph.

In some cases, edges can have directions (run only in one direction), these edges are called *arcs* and can be represented graphically by an arrow. A graph having arcs is a *directed graph. Undirected link* does not have any sense of direction and can be considered as a link with arrows drawn in both the directions. A graph with undirected links is called *undirected graph.* A graph having both directed and undirected links is called *mixed graph.*

It is also possible that edges or nodes can carry some weight, then the term used for such networks is *weighted networks.* If every pair of distinct nodes in a graph is connected by an unique edge, it is called a *complete graph.*

The graph $G(N, E)$ is represented in the form of a $N \times N$ matrix A called an *adjacency matrix.* It has the matrix entires A_{ij} such as,

$$A_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected} \\ 0 & \text{Otherwise.} \end{cases} \tag{1.1}$$

with

$$2E = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij}. \tag{1.2}$$

Number of non-zero elements in the i^{th} row, is called the *degree* of i^{th} node and is denoted by K_i. Degree is a local property of the node because it represents the number of links connected to a node.

The adjacency matrix of an undirected graph is the symmetric matrix $A_{ij} = A_{ji}$, and hence will have all real eigenvalues and an orthogonal eigenvector basis. The degree of a node i can be written as

$$K_i = \sum_{j=1}^{N} A_{ij}. \tag{1.3}$$

In the case of directed networks $A_{ij} \neq A_{ji}$ and there are two types of degrees, in-degree K^{in} (number of incoming links) and out-degree K^{out} (number of outgoing links), associated with a node.

The matrix entries in the $n - th$ power of an adjacency matrix, A^n, i.e. the

Networks	Size	Nodes	Links	References
Movie Actor	10^5	Actors	acted in same movie	[7]
Co-authorship	10^5	Researchers	co-authors	[8, 9, 10]
Citation	10^6	Research Papers	Citation	[11]
Phone call	10^7	Phone no.	Completed calls	[12]
Ecology	10^2	Animals	food web	[13]
Cellular	10^2	substrates	Chemical Reactions	[14]
Sexual	10^3	Individuals	Sexual relationships	[15]
Linguistic	10^6	words	used together	[16]
Protein folding	10^3	Distinct states	simple conformations	[17]
Power grid	10^3	Generator etc.	Transmission lines	[7]
Neural Network	10^2	Neurons	Synapse	[7]
WWW	10^8	Documents	hyper links	[18]
Internet	10^6	Routers	wires	[19]

Table 1.1: Different real world networks and their properties.

matrix product of n copies of A, reveals the number of paths of length n in between the nodes. The entry in $i-th$ row and $j-th$ column of the matrix A^3 represents number of paths of length 3 in between node i and node j and the diagonal entry of $i-th$ row represents the number of cycles of length 3 for node i. This implies, for example, that the number of triangles in an undirected graph is exactly the trace of A^3 divided by 6.

A complete graph is the one in which every node is connected to every other node. In this case, each node will be having degree $K = N - 1$ and total number of undirected edges $E = N(N-1)/2$.

There are many other terms from graph theory which are left untouched in this section but those terms are used at very specific places.

Real world networks

The aim of network approach is to understand various real systems ranging from financial to biological. There are many systems for which networks are relevant models. In this section, a brief description of such systems and their network models is provided. A small number of different networks (depending upon interactions) are listed (see table 1.1) to give an idea about the interdisciplinary nature of network based studies.

These networks have played a major role in the development of network theory. There have been many quantities which were defined to study the structural properties of such networks.

Properties of networks

After the empirical studies of the real world networks, it had been very clear that real networks stand somewhere in between random graphs and regular lattices. In this section, main properties of real world networks which appear to be shared by large number of systems are discussed.

Small world property In year 1967, Milgram's experiment suggested that the real world is very small [6] and later, this small world effect has been studied and verified in many real world networks having large number of nodes.

Consider an undirected connected network with N nodes and define the mean path length l between node pairs in a network

$$l = \frac{2}{N(N+1)} \sum_{i \geq j} d_{ij}, \tag{1.4}$$

where d_{ij} is the distance from node i to node j. As a network often is disconnected, the distance between disconnected nodes is ∞ and hence, it is useful to study l of the largest connected component or the sum of reciprocal distances [20].

A network is called a small world if the average distance l scales logarithmically or slower with the number of nodes N. Many real world networks exhibit the small world property.

Betweenness centrality A measure of the importance of the node related to shortest path is the betweenness. It is defined as the number of all possible shortest paths in a network passing through node i. It is defined as,

$$b(i) = \sum_{u \neq i \neq v} \frac{\sigma_{uv}(i)}{\sigma_{uv}}, \tag{1.5}$$

where σ_{uv} is the total number of shortest paths in between the nodes u and v and $\sigma_{uv}(i)$ denotes the number of those paths passing through node i. A node with the highest betweenness is the the most central node in the network.

Clustering Clustering is another feature of real world networks which makes them different from the random graphs. It is observed in many networks that if node A is connected to node B and node B is connected to node C, it is more likely that node A will also be connected to node C. The clustering coefficient of a node is the probability that two neighbors are sharing a common neighbor. Mathematically, if node i has the degree K_i and there are E_i connections in between these K_i neighbors, the local clustering coefficient is

$$C_i = \frac{2E_i}{K_i(K_i - 1)}. \tag{1.6}$$

The clustering of the whole network is given by its average $C = \sum_i C_i/N$. The value of clustering coefficient tends to be considerably higher for real world networks than the random graph with a similar number of nodes and edges. The clustering coefficient decreases as the degree of the node increases and it varies approximately as K_i^{-1}.

Degree Distribution Degree distribution is the probability distribution of the degrees of the nodes in the network. Mathematically, the degree distribution $P(K)$ is defined as the probability that a randomly chosen node in a network has the degree K.

In the case of random network, the links are random and hence the majority of the nodes have the number of connections close to the average degree $\langle K \rangle$. The degree distribution of a random graph is Poisson distribution with peak at $P(\langle K \rangle)$ and it decays as $P(K) \sim 1/K!$.

In contrast, real world networks have degree distribution which significantly deviates from Poisson and it is a slowly decaying degree distribution. An asymptotic power-law distribution is an example of a slow decaying distribution and for a large number of networks, the degree distribution has a power law tail,

$$P(K) \sim K^{-\gamma}, \tag{1.7}$$

with exponent varying in the range $2 < \gamma < 3$. Such networks are called *scale-free networks*. In scale free networks, small degree nodes are present in large numbers but higher degree nodes are very few. These higher degree nodes are called *hubs*. Presence of hubs has been noticed in many real world networks. Other common forms for the degree distribution are exponential and it has been noticed in many real world networks.

Network models

The observations discussed in the previous section motivated the introduction of new mathematical models to explain the structural properties of the real world networks. Regular lattices are popular among physicists since a very long time and then on the other hand, random models are well studied by mathematicians. Structurally, these models display two extremes, ordered and random. The real world networks occupy a position in between them. Real world networks were modeled on the basis of their properties, small world and scale free. In this section, two very popular models of real world networks are discussed along with the random network models.

Random Networks In year 1959, Paul Erdős and Alfred́ Rényi defined random graph, a graph $G(N, E)$ with E links chosen randomly from the $N(N-1)/2$ possible connections [21]. The Erdős Rényi (ER) graph can be generated by connecting a pair of nodes with probability p.

The degree distribution is binomial and it is given by,

$$P(K) = \binom{N-1}{K} p^K (1-p)^{N-1-K}, \tag{1.8}$$

with average number of links $\langle K \rangle = pN(N-1)/2$. For large N, the distribution takes Poissonion form. The graph remains connected for $\langle K \rangle \gtrsim ln(N)$. The average path length $l \sim ln(N)/ln(K)$ shows small world property but for large networks the clustering coefficient $C \approx \langle K \rangle / N$ is very small, which is very different from real world networks. Hence, this model could not represent the real world networks. However, for random networks most of the quantities of interest can be calculated analytically.

Small world Networks Real world networks have small diameter and large clustering coefficient which could not be explained by random network model. In year 1998, Watts and Strogatz proposed a model, now known as W-S model, which could have both the properties [7]. In the model, every link on a regular lattice with K neighbors ($K > 2$) is randomly rewired with probability p such that self connections and multiple connections are avoided. For $p = 0$, the network is a regular network but for large p it has the properties of a random network. Even for smaller values of p, the average path length is of the order of that for random network but due to regular structure present in the network, clustering coefficient remains very high. Rewiring process keeps the average degree of the network unchanged but the local degree of the nodes changes and hence, the degree distribution also changes.

This was the first model that could explain the high clustering and small diameter present in real world networks. However this model could not explain the scale-free feature in the real world networks and a search for alternative models started.

Scale-free Networks One year after the WS model, Barabási and Albert suggested a mechanism responsible for the emergence of networks with power law distribution [22]. It is a network growth model with preferential attachments which favors attachment to the nodes with higher degree. This manifests the "rich get richer" principle.

Initially there are m_0 nodes with no connections in between them. At every time step, a new node with $m(\geq m_0)$ edges is added. It connects with m different nodes already present in the network. The probability π that a new node connects to node i depends upon the degree K_i such that,

$$\pi(K_i(t)) = \frac{K_i(t)}{\sum_j K_j(t)}. \tag{1.9}$$

After t time steps, the resulting network has $N = t + m_0$ nodes and mt links. Simulations show that network has the power law degree distribution $P(K) \sim K^{-\gamma}$

with exponent $\gamma = 3$. The model has shorter path length than the small world networks and is given by,

$$l \sim \frac{lnN}{ln(lnN)}. \tag{1.10}$$

The clustering coefficient also scales as power law, $C \sim N^{-.075}$. Power law degree distribution in the BA model arises only for linear attachment, hence the model is not a general model but it was the first model to discuss a mechanism behind power law degree distribution.

After these models, there have been many other models which could capture the properties of real world networks [23, 24, 25]. But the BA-model, ER model and the WS model belong to the first generation of simple mathematical models of networks and hence, these models have their own charm and importance in the theory of complex networks.

1.3.2 Dynamics on complex networks

Going beyond the structural properties of networks, the dynamical processes on the networks have generated considerable interests. Spreading of a disease, power black-outs, efficient navigation, emergence of social behavior etc. are some examples of dynamical processes taking place on networks. All These processes are strongly affected by the network topology and the understanding obtained from these studies is relevant in very different fields like biology, physics, computer sciences and social sciences.

Many of the dynamical processes have been studied on regular lattices but the outcomes of the dynamics on underlying heterogeneous network structure often differ from the standard results. Most of these results are documented in textbooks like [26, 27]. Therefore, it becomes necessary to understand the impact of network characteristics on the basic features of dynamical processes. The aim of studying dynamics on networks is to understand the emergence of co-operation and other properties of the dynamical processes from the very basic and simple microscopic interactions among the system elements.

There are two approaches to model dynamics on complex networks. In the first, each individual node of the network is like an element of the system. In the second approach, dynamic entities such as energy, people, packets flow through a network whose nodes identify the location of entities. In both the cases, each node i has the dynamical variable f_i characterizing the dynamical state of the node. The dynamical evolution of the system is given by the dynamics of $\{f_i\}$ and is described by the transition from one state to another state i.e. $f^a \to f^b$.

In general it is impossible to follow the microscopic dynamics of a network. The master equation approach becomes useful to study the dynamics of the system. The master equation consists of evolution of probability $P(f,t)$ of finding a system in state f and time t. It can be written as,

$$\partial_t P(f,t) = \sum_{f'} \left[P(f',t)\Pi_{f' \to f} - P(f,t)\Pi_{f \to f'} \right], \tag{1.11}$$

where the $\Pi_{f \to f'}$ is the transition probability of going to state f' from f and the sum is over all the possible states. In the case of networks, where the transition takes place among the neighbors only, the transition probability for the whole system can be decomposed into the products of single node transition probabilities. Now, the transition probabilities for a node depend on its local structure. Therefore, the network structure becomes evident and strongly influences the dynamics.

While it is not possible to solve the master equation in general, it is possible to obtain the stationary distribution which can be defined as,

$$\lim_{t \to \infty} P(f,t) = P_\infty(f). \tag{1.12}$$

The stationary distribution may not exist in all the dynamical systems but if the system is ergodic, the existence of a unique stationary distribution is guaranteed.

In the case of equilibrium dynamics, the stationary distribution can also be obtained using detailed balance condition,

$$P_{eq}(f')\Pi_{f' \to f} = P_{eq}(f)\Pi_{f \to f'}, \tag{1.13}$$

which means that the net flow in between the pairs of state is zero. The detailed balance is a strong condition but it is not true in the case of non-equilibrium dynamics. However, the detailed balance does not imply the lack of stationary distribution. As mentioned earlier, it is not possible to obtain the complete solution of the master equation representing the dynamical process. In such cases, approximate solution for the system can be obtained using various approximation schemes.

In a more complicated processes, detailed computer simulations are performed. In these simulations, each node is assumed to be in one of the several states and the state updates depend on the microscopic dynamics applied to each node. This approach mimics the whole system with detailed dynamical rules in a computer and provide us the understanding of the dynamical processes.

There are many dynamical processes which have been well studied earlier in statistical physics or nonlinear dynamics and later, the generality of the results have been tested on networks. A review article by S. N. Dorogovtsev et al. is a good reference for dynamics in complex networks [28]. Following is a small list of dynamical processes studied on networks with the main references.

- Ising model in networks [29, 30, 31].
- Synchronization phenomena in complex networks [32, 33, 34].
- Percolation in complex networks [35, 36, 37].
- Diffusion in networks [38, 39, 40, 41].
- Epidemic models in networks [42, 43, 44, 45].
- Opinion formation and the voter model [46, 47, 48].
- Transport processes on networks [49, 50, 51].

Out of these processes, transport processes are very well studied in complex network. One of the reasons behind it is the importance of transportation and technological infrastructure in our life. In most of the transport processes, traffic can be modeled in many different ways, from simple non-interacting random walkers [52, 53] to interacting random walkers with random traps [54, 55, 56]. In all the transport processes, congestion arises as an essential part of the flow on network. Consider, for an instance, a web server not responding to due the http requests due to heavy load. In most of the cases, one of the reason behind the congestion can be the limited handling capacity of the nodes or links through which the flux is passing. In this case, the overloading of node may lead to their failure and the redistribution of the traffic can trigger other failures in the network. This is called the cascading failure and can be treated as an example of extreme event on the transport network. The major power blackouts across the world [57], gridlocks on highways are the examples of cascading failures. Transport on networks continues to be widely studied in terms of congestions [58] and cascades [59] but much less attention has been focused on it from the point of view of extreme events arising due to the inherent fluctuations in the flux passing through the nodes.

1.3.3 Extreme events

Extreme events are just like any other events taking place in a system but they are very different in terms of their magnitude. The very large/small magnitude of these events creates a very high impact on the system and their rarity makes them special. Statistically, extreme events occur in the tail of probability distributions that define the occurrence of events of a given size, shown in Fig. (1.2). Floods, earthquakes, tsunami stock market crash, an epileptic seizure, a storm, magnetic storms or any other such events come under the category of extreme events. Extreme events are often associated with catastrophic consequences which lead to huge socio-economic losses. Therefore, it becomes very important to understand extreme events. The reasons behind the occurrences of extreme events are not precisely known. The questions related to their predictions can help in bringing down the losses but unfortunately they also remain unanswered in most of the cases.

The origin of extreme events is mainly related with the complex dynamics of the system. They occur usually where the system's variability and collective effects are dominant. Hence, the search of dynamic mechanism which allows system to visit the state away from its normal states has inspired us to look at the collective dynamics on the networks. But without knowing the dynamics behind the extreme events, it is possible to study the statistical properties of extreme events using a well developed theory known as extreme value theory [60, 61]. Main results of the

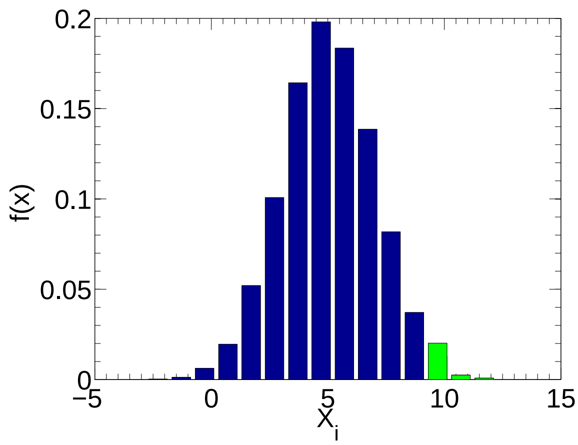

Figure 1.2: The complete Probability distribution function of a random variable (Blue bars). Extreme events lie in the tail of the distribution (shown in green bars)

extreme value theories are presented here in a very concise form.

Extreme value theory

The extreme value theory is an established area in mathematical statistics and is widely used in many disciplines ranging from structural engineering to economics. Extreme value theory was initiated by Fisher and Tippet and later developed by Von Mises, Gumbel, Gnedenko and many more. The aim of extreme value theory is to find the probability distribution functions of extreme events and calculating the probability of extreme events of a given size in a given time interval or the largest event that may occur in a given period of time. The limitations of the theory comes from the basic assumption made in it which says events are independent and identically distributed (iid), which is hardly a case in real life. This theory is important because it sets a basic benchmark and is useful in weakly correlated cases. It estimates for the magnitudes of extreme events which helps in the preparedness to meet the extreme events.

Extreme value theory deals with the events in the tail part of the distribution. In real life data, there are two ways of defining extreme events which result in two different types of distributions.

- Block maxima: It is the traditional approach where the data is separated into blocks (days, months etc) and from each block only the maximum values are

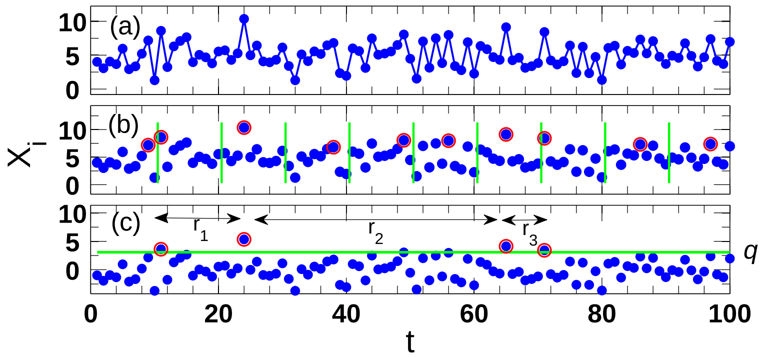

Figure 1.3: Two approaches for selecting extreme data. (a) Variable $\{X_i\}$ as a function of time. (b) Block maxima approach:$\{X_i\}$ is divided into blocks of equal intervals (green lines) and Red blue dots are the maximum values attained in each block which are called as extreme events here. (c) Peak over threshold: The green line here represents some threshold q and any value exceeding it is called as extreme event (Red blue dots). r_i is the time interval between two successive extreme events.

picked up (refer to Fig. 1.3 (b)). This lead to the classical extreme value distributions. Here, the width of blocks plays an important role.

- Threshold exceedances: A threshold is decided on the basis of the data and the data above the threshold is called an extreme event as shown in Fig. 1.3(c). The choice of threshold becomes very critical here in determining the extreme events. This lead to different probability distributions.

The probability distributions associated with these two ways of defining an extreme event are presented below.

Extreme value Distributions Suppose, we have a stationary sequence of iid random variables, $\{X_1, \ldots, X_n\}$ with a common cumulative distribution $F(x) = Pr(X_i > x)$. The maximum of these random variables is M_n. Then, the cumulative distribution of M_n is given by

$$Pr\{M_n \leq x\} = \{F(x)\}^n, \tag{1.14}$$

which converges to zero for $x < x_+$ and to 1 for $x \geq x_+$, where x_+ is the upper end point of F, defined as the smallest value of x for which $F(x) = 1$. However if there exists a sequence of constants $\{a_n > 0\}$ and $\{b_n\}$, such that the renormalized variable,

$$M_n^* = \frac{M_n - b_n}{a_n}, \tag{1.15}$$

has a non-degenerate limiting distribution, i.e.,

$$\lim_{n\to\infty} Pr\left\{\frac{M_n - b_n}{an} \leq x\right\} = F^n(a_n x + b_n) = G(x), \tag{1.16}$$

where G is a non-degenerate cumulative distribution function. Then, G belongs to one of the following extreme value cumulative distributions:

$$\text{Gumbel: } G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \qquad -\infty < x < \infty; \tag{1.17}$$

$$\text{Fréchet: } G(x) = \begin{cases} 0, & x \leq b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & x > b, \alpha > 0; \end{cases} \tag{1.18}$$

$$\text{Weibull: } G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)^{-\alpha}\right]\right\}, & x < b, \alpha > 0; \\ 1, & x \geq b. \end{cases} \tag{1.19}$$

These three classes of distributions are called the extreme value distributions. While a and b are the scale and the location parameters respectively for each distribution, Fréchet and Weibull have an additional parameter α, known as shape parameter. The remarkable result of extreme value theory is that these three distributions are the only possible limiting distributions of M_n^* regardless of the distribution F of random variables. This result is similar to the central limit theorem and can be considered as an extreme value analog of it.

The three types of extreme value distributions have their own domain of attractions and the limiting distribution is decided on the behavior of the tail of the distribution F of random variables X_i. Gumbel distribution is obtained as a limiting case of any distribution F with the tail falling faster than a power law. In the case of Fréchet, the tail of F must fall as a power law. For both the distributions, the upper end point x_+ of F is infinite while in the case of Weibull distribution, F

has the finite upper end point.

All the three extreme value cumulative distributions can be combined into a single distribution function known as generalized extreme value (GEV) cumulative distribution of the form,

$$G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{1.20}$$

for $(1+\xi(x-\mu)/\sigma) > 0$ with location parameter $-\infty < \mu < \infty$, scale parameter $\sigma > 0$ and shape parameter $-\infty < \xi < \infty$. Fréchet and Weibull distributions are obtained respectively for $\xi > 0$ and $\xi < 0$. The Gumbel distribution is obtained in the limit $\xi \to 0$.

The similar limiting distributions can be obtained for block minima by using the trick $Y_i = -X_i$.

Exceedances over thresholds Consider the cumulative distribution of X_i exceeding some high threshold q. The cumulative distribution will be given by,

$$Pr\{X > q+y|X > q\} = \frac{1-F(q+y)}{1-F(q)}, \quad y > 0. \tag{1.21}$$

For very high threshold q, the cumulative distribution function of $(X-q)$ conditional on $X > q$ is approximately

$$H(y) = 1-\left(1+\frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}, \tag{1.22}$$

with $y > 0$ and $(1+\xi y/\tilde{\sigma}) > 0$ where $\tilde{\sigma} = \sigma+\xi(q-\mu)$. The cumulative distribution is known as Generalized Pareto Distribution(GPD). ξ is the shape parameter but it no longer depends on the width of blocks. Like the GEVs, here also shape parameter ξ determines the qualitative behavior of GPD and the values of ξ are similar to the previous model. However, the scale parameter $\tilde{\sigma}$ depends on the threshold except in the case of $\xi = 0$.

Recurrence of extreme events

As discussed earlier, extreme events are rare but they are recurrent. Hence, one important issue related with extreme events is the predictability and the quantity of interest can be the time interval between events that exceed a threshold (say, q), known as return intervals (r_q) (refer to Fig. 1.3(c)). The mean return time $\langle r \rangle_q$ and the distribution $P_q(r)$ are the important quantities from the predictability point of view. When the threshold (q) is small, extreme events are frequent and the return intervals are short and when the threshold is large, return intervals are also long.

The basic assumption in the extreme value theory is that events are iid, i.e., they are uncorrelated. When events are uncorrelated, their return intervals are also uncorrelated. The correlation is measured by the two-point correlation function as $N \to \infty$ and is given by,

$$C(\tau) = \frac{\langle x(t)x(t+\tau)\rangle}{\langle x^2(t)\rangle}, \tag{1.23}$$

where $\langle . \rangle$ denotes the time average. In the case of uncorrelated events, $\sum_\tau C(\tau) < \infty$ and the return intervals follow the exponential distribution.

$$P_q(r) = \frac{1}{\langle r \rangle_q} \exp\left(\frac{-r}{\langle r \rangle_q}\right). \tag{1.24}$$

But, if the events are correlated i.e. $\sum_\tau C(\tau) = \infty$, the return interval distribution $P_q(r)$ differs from the exponential distribution. Most of the natural phenomena e.g. earthquakes, river run-off, daily atmosphere temperature etc. display long range correlations or in other words, long memory [62, 63]. Hence, in these cases time intervals in between extreme events should also be correlated and should differ from exponential distribution. In the case of long range correlated time series, Santhanam et al. have shown that, under some approximations, the return interval distribution gets modified and it turns out to be a product of power law and a stretched exponential when the average return intervals are large [64].

1.4 Thesis outline

The thesis is organized as follows.

Chapter 2 is about random walk and extreme events on complex networks. In that, we discuss the random walk dynamics on network with fluctuating number of walkers on the network. Based upon the random walk dynamics, extreme events on networks are defined as the exceedances over threshold and the probability of occurrence of extreme events on nodes is calculated. The correlation among extreme events using their return intervals on the nodes is discussed. Also, the scaling relation in extreme event probabilities is obtained for different thresholds. Finally, the results related to shortest path strategy are compared against the random walks.

In chapter 3, Extreme events on networks are investigated under the situation where dynamics has some preferences among nodes. They are discussed with the biased random walk dynamics. An appropriate parameter 'generalized strength' is used to characterize the extreme events. Event sizes are calculated under different biases and the role of network topology under different biasing parameters is discussed.

Chapter 4 is devoted to the study the resilience of network against extreme events. A model of network failure based on extreme events is proposed. The nature of failures and the conditions responsible for network failures are presented on a fully connected network.

A brief summary of our results described in Chapter 5 will conclude this thesis.

Chapter 2

Extreme Events on Complex Networks

2.1 Introduction

Extreme events taking place on the networks is a fairly commonplace experience. Traffic jams in roads and other transportation networks, web servers not responding due to heavy load of web requests, floods in the network of rivers and power blackouts due to tripping of power grids are some of the common examples of extreme events on networks. Such events can be thought of as an emergent phenomena due to the transport on networks. As extreme events lead to losses ranging from financial and productivity to even life and property [65], it is important to estimate probabilities for the occurrence of extreme events and, if possible, incorporate them in the design of networks so that it can handle such extreme events.

The transport phenomena on networks have been studied vigorously in the last several years [66, 67] though these studies were not focused on the analysis of extreme events. An extreme event can be defined as the exceedances above a prescribed quantile and is not necessarily related to the handling capacity of the node in question. It arises from the natural fluctuations in the traffic passing through a node and not due to constraints imposed by the capacity. Thus, in the rest of this chapter, we discuss the transport on networks and analyze the probabilities for the occurrence of extreme events arising in them without having to prescribe capacity at each of the nodes.

The transport model we adopt is the random walk on complex networks [68]. Random walk is of fundamental importance in statistical physics though in real network settings many variants of random walk could be at work [69, 70]. For instance, in the case of road traffic, the flow typically follows a fixed, often shortest, path from node A to B and can be loosely termed deterministic. Thus, given the operational principle of network dynamics, *i.e.*, deterministic or probabilistic or a combination of both, we obtain the probabilities for the occurrence of extreme events on the nodes. The study reveals a significant and unexpected result; namely that the extreme events are more prone to occur on a small degree node than on a hub. This feature is robust against fluctuating traffic and even upon the application of intelligent routing algorithms (e.g. shortest paths). This principal result implies that the design principles for networks should focus on small degree nodes which are prone to extreme events. Further, these probability estimates allow us to design nodes that can have sufficient capacity to smoothly handle extreme events of certain magnitude. Currently, for a univariate time series, there is a widespread interest on the extreme value statistics and their properties, in particular in systems that display long memory [64, 71]. We use multivariate time series and place our results in the context of both the random walks and extreme events in a network setting.

2.2 Random walk on network

We consider a connected, undirected, finite network with N nodes with E edges. The links are described by an adjacency matrix A whose elements A_{ij} are either 1 or 0 depending on whether i and j are connected by a link or not respectively. On this network, we have W noninteracting walkers performing the standard random walk. A random walker at time t sitting on the ith node with K_i links can choose to hop to any of the neighboring nodes with equal probability. Thus, transition probability for going from ith to jth node is A_{ij}/K_i. We can write down a master equation for the $n-$step transition probability of a walker starting from node i at time $n = 0$ to node j at time n as,

$$P_{ij}(n+1) = \sum_k \frac{A_{kj}}{K_k} P_{ik}(n). \tag{2.1}$$

Here, the $n-$step time-evolution operator corresponding to this transition, acting on an initial distribution, leads to a stationary distribution $P_i^\infty = p_i$ with eigenvalue unity [68]. The stationary distribution exists if and only if the network contains an odd loop, which is not true for a bipartite network. In a bipartite network, all the cycles are of even length [72] and hence, one of the eigenvalues is $\lambda = -1$ for which the distribution never converges to a stationary distribution.

The transition probability P_{ij} to go from node i to node j in $n+1$ steps can be written as,

$$P_{ij}(n+1) = \sum_{j_i \cdots j_n} \frac{A_{ij_1}}{K_i}.\frac{A_{j_1 j_2}}{K_{j_1}} \ldots \frac{A_{ij_n}}{K_n}. \tag{2.2}$$

A similar expression can be written for $P_{ji}(n+1)$,

$$P_{ji}(n+1) = \sum_{i_1 \cdots j_n} \frac{A_{ji_1}}{K_i}.\frac{A_{i_1 i_2}}{K_{i_1}} \ldots \frac{A_{ji_n}}{K_n}. \tag{2.3}$$

Now, if we compare these two equations for P_{ij} and P_{ji}, we see that

$$K_i P_{ij}(n) = K_j P_{ji}(n). \tag{2.4}$$

In the limit $n \to \infty$,

$$K_i p_j = K_j p_i, \tag{2.5}$$

which implies that,

$$p_i = C K_i \tag{2.6}$$

where C is a constant and using the normalization condition $\sum_i p_i = 1$ we get,

$$C = \frac{1}{\sum_i K_i} = \frac{1}{2E}. \tag{2.7}$$

So, the stationary probability turns out to be

$$\lim_{n\to\infty} P_{ij}(n) = p_j = \frac{K_j}{2E}. \tag{2.8}$$

It is easy to confirm if the obtained distribution is stationary. Using the definition of stationary probability,

$$\begin{aligned} p_j &= \sum_{i\to j} p_i P_{ij} \\ &= \sum_{i\to j} \frac{K_i}{2E}\frac{1}{K_i} \\ &= \sum_{i\to j} \frac{1}{2E} \\ p_j &= \frac{K_j}{2E}. \end{aligned}$$

The stationary distribution obtained here is unique and it is the consequence of connected network. Here, the stationary distribution depends only on the degree of the nodes which means that the network topology does not play any role as far as the stationary state is concerned but how fast the distribution converges to the stationary distribution depends on the network structure. The convergence time is bounded by $\mathcal{O}(N^3)$ time steps for any graph with N nodes because a walker can take maximum $\mathcal{O}(N^3)$ time steps to commute in between any two nodes [72].

Physically, the stationary probability in Eq. (2.8) implies that more walkers will visit a given node if it has more links. The existence of stationary distribution is crucial for defining extreme events.

Consider a situation, where more than one walkers are present and preforming random walks. These W walkers are independent and they do not interact with each other. A schematic diagram mimicking the scenario for time steps, $n = 0, 1, 2, 3, 4$, is shown in Fig. 2.1. There are $W = 5$ walkers moving randomly on the network with $N = 6$ nodes and $E = 8$ links. Their random movement leads to fluctuations in the number of walkers present on a node.

Now we can obtain the distribution of random walkers on a given node. We ask

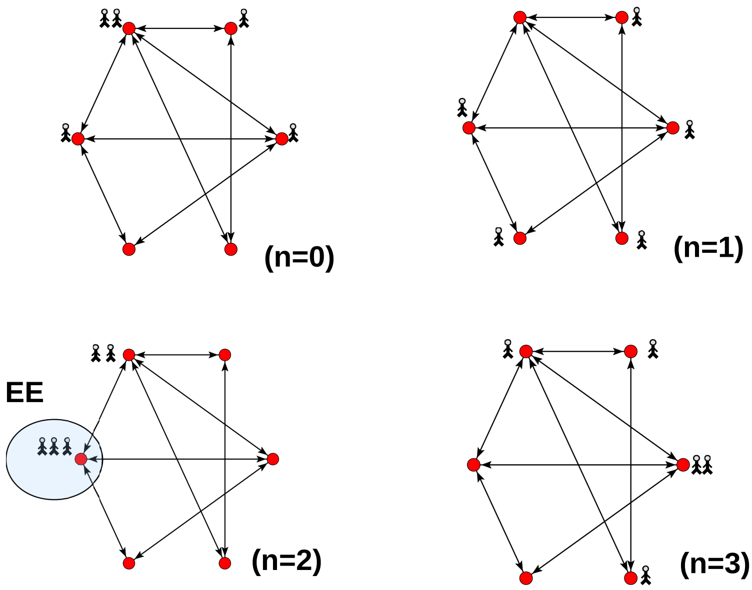

Figure 2.1: Schematic diagram to show random walk and extreme events. There are $W = 5$ independent and noninteracting walkers performing random walks on a small network($N = 6, E = 8$). Number of walkers on nodes are shown for $n = 4$ time steps. Extreme event is defined as the number of walkers exceeding a predefined threshold, say ($q = 2$) here, on a node at any time instant. At $n = 2$, there are 3 walkers accumulated on a particular node. This node is said to have an extreme event at $n = 2$.

for the probability $f(w)$ that there are w walkers on a given node having degree K. Since the random walkers are independent and non-interacting, the probability of encountering w walkers at a given node turns out to be binomial distribution, given by

$$f(w) = \binom{W}{w} p^w \ (1-p)^{W-w}. \tag{2.9}$$

Now, the mean and variance for a given node can be explicitly written down as

$$\langle f \rangle = \frac{WK}{2E}, \qquad \sigma^2 = W\frac{K}{2E}\left(1 - \frac{K}{2E}\right). \tag{2.10}$$

As expected, the mean and variance depends on the degree of the node for fixed W and E. The maximum convergence time is $\mathcal{O}(N^3)$ for an N node graph but in a crude sense, the minimum time required for the convergence can be equal to the diameter of the graph. To show the convergence in terms of number of walkers, the

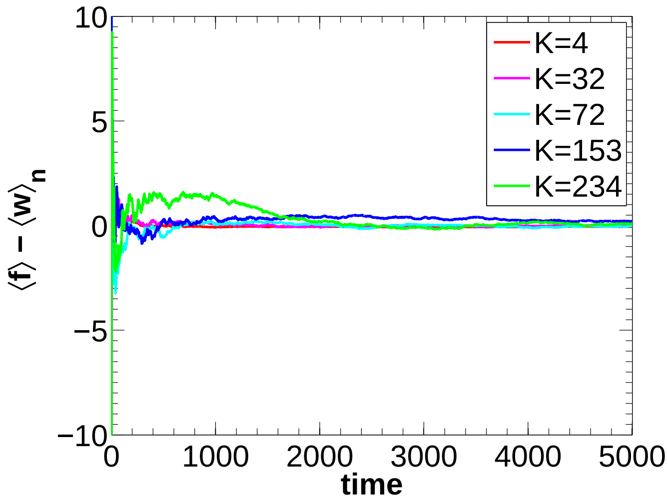

Figure 2.2: Convergence of average number of walkers $\langle w \rangle_n$ to the analytically calculated stationary distribution $\langle f \rangle$ for nodes with different degree is shown here. Numerically, the distribution is said to converge if $(\langle f \rangle - \langle w \rangle_n) < 0.05\sigma$.

difference between the stationary distribution $\langle f \rangle$ and the time average of walkers on a node, $\langle w \rangle_n$, is plotted as a function of time in Fig. 2.2. The nodes with smaller degree converge faster to their stationary distribution than the high degree nodes. Overall, the distribution converges to a stationary distribution in $\mathcal{O}(N)$ time steps. Note that $K/2E << 1$ and hence $\sigma \approx \langle f \rangle^{1/2}$. This reproduces the relation proposed in Ref. [73], later shown to have limited validity [74].

One natural extension of the result in Eq. (2.9) is to account for the fluctuations in the number of walkers. We assume that the total number of walkers is a random variable uniformly distributed in the interval $[W-\Delta, W+\Delta]$. Then the probability of finding w walkers becomes

$$f^{\Delta}(w) = \sum_{j=0}^{2\Delta} \frac{1}{2\Delta+1} \binom{\widetilde{W}+j}{w} \; p^{w} \; (1-p)^{\widetilde{W}+j-w}, \tag{2.11}$$

where $\widetilde{W} = W - \Delta$. The mean of the distribution should remain unchanged because of the mean field approach and is given by,

$$\langle f^{\Delta} \rangle = \langle f \rangle. \tag{2.12}$$

For estimating variance, we need to calculate the second moment of the distribution, $\langle f^{2,\Delta}\rangle = \langle w^2 f^{\Delta}(w)\rangle$, which is given by,

$$\langle w^2 f^{\Delta}(w)\rangle = \frac{1}{W}\sum_{w=0}^{W} w(w-1)f^{\Delta}(w) + \langle f^{\Delta}\rangle, \tag{2.13}$$

substituting the value of $f^{\Delta}(w)$ from Eq. (2.11),

$$\begin{aligned}
\langle w^2 f^{\Delta}(w)\rangle &= \frac{1}{2\Delta+1}\sum_{j}^{2\Delta}\left[(W+\Delta+j)(W+\Delta+j-1)\right]p^2 + \langle f^{\Delta}\rangle, \\
&= p^2\left[W^2 - W + \frac{\Delta^2}{3} + \frac{\Delta}{3}\right] + \langle f^{\Delta}\rangle, \\
\langle f^{2,\Delta}\rangle &= \langle f^{\Delta}\rangle^2 + (pW)^2\left[-\frac{1}{W} + \frac{\Delta^2}{3W^2} + \frac{\Delta}{3W^2}\right] + \langle f^{\Delta}\rangle.
\end{aligned}$$

and the variance of this distribution turns out to be,

$$\sigma_{\Delta}^2 = \langle f^{\Delta}\rangle\left[1 + \langle f^{\Delta}\rangle\left\{\frac{\Delta^2}{3W^2} + \frac{\Delta}{3W^2} - \frac{1}{W}\right\}\right]. \tag{2.14}$$

2.3 Extreme event probability

In the spirit of extreme value statistics, an extreme event is the one whose probability of occurrence is small, typically associated with the tail of the probability distribution function. In the network setting, we will apply the same principle to each of the nodes. Based on Eqns. (2.9-2.10), we will designate an event to be extreme if more than q walkers traverse a given node at any time instant (shown in Fig. 2.1,$n = 4$). The probability for the occurrence of extreme event can be obtained as

$$F(K) = \sum_{j=0}^{2\Delta}\frac{1}{2\Delta+1}\sum_{k=\lfloor q\rfloor+1}^{\widetilde{W}+j}\binom{\widetilde{W}+j}{k}p^k\,(1-p)^{\widetilde{W}+j-k}, \tag{2.15}$$

where $\lfloor u\rfloor$ is the floor function defined as the largest integer not greater than u.

Notice that necessarily the cutoff q will have to depend on the node (or rather, the traffic flowing through the node) in question. Applying uniform threshold independent of the node ($q =$ constant) will lead to some nodes always experiencing an extreme event while some others never encountering any extreme event at all. Hence we define the threshold for extreme event to be,

$$q = \langle f \rangle + m\sigma, \tag{2.16}$$

where m is any real number and it decides the rarity of the extreme events. The threshold, chosen in this way, automatically takes care of the variability in the average flux passing through the nodes with different degrees. It also incorporates the fluctuations in the number of walkers on a node.

It does not seem possible to write summation in Eq. (2.15) in closed form. However, for the special case when $\Delta = 0$, Eq. (2.15) simplifies to

$$F(K) = \sum_{k=\lfloor q \rfloor + 1}^{W} f(k) = I_p\left(\lfloor q \rfloor + 1, w - \lfloor q \rfloor\right), \tag{2.17}$$

where $I_p(.,.)$ is the regularized incomplete Beta function [75].

For a given choice of network parameter E and number of walkers W, the extreme event probability at any node depends only on its degree. In Fig. 2.3 we show $F(K)$ as a function of degree K superimposed on the results obtained from random walk simulations. The agreement between Eq. (2.15) and the simulated results is quite good. Further, each point in the figure represents an average over all the nodes with the same degree. We emphasize that the oscillations seen in Fig. 2.3 are inherent in the analytical and numerical results and not due to insufficient ensemble averaging. These oscillations are the consequences of the fact that the number of walkers on a node can take integer values only.

An important feature of this result is that the nodes with smaller degree ($K < 20$) reveal, on an average, higher probability for the occurrence of extreme events as compared to the nodes with higher degree, say, $K > 100$. By careful choice of parameters, the probability $F(K)$ can differ by as much as an order of magnitude. This runs contrary to a naive expectation that higher degree nodes garner more traffic and hence are more prone to extreme events. While the former contention is still true in the random walk model we employ, the results here indicate that the latter one is not generally correct. As shown in Fig. 2.3(b,c), this feature is

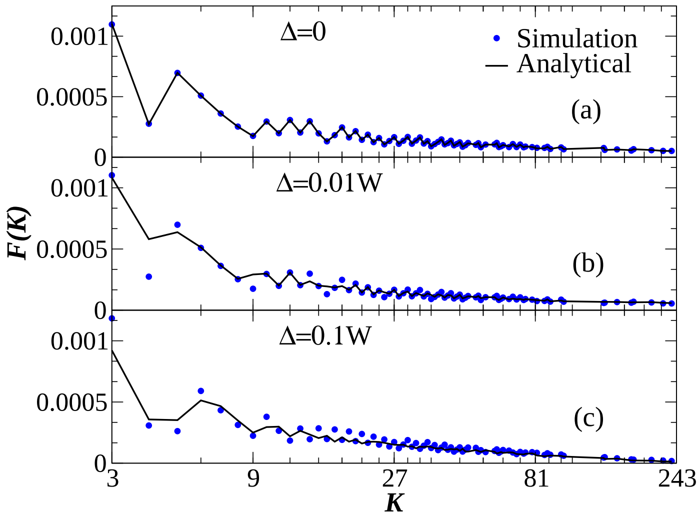

Figure 2.3: Probability for the occurrence of extreme events as a function of degree K with fluctuations Δ in the total number of walkers on semilog plot. The threshold for extreme events is $q = \langle f \rangle + 4\sigma$. The solid lines are from the analytical result in Eq. (2.15). All the simulations shown in this chapter are obtained with a scale-free network (degree exponent $\gamma = 2.2$) with $N = 5000$ nodes, $E = 19815$ vertices and $W = 2E$ walkers averaged over 100 realizations with randomly chosen initial conditions.

robust even when the number of walkers becomes a fluctuating quantity. Though, the analytical and simulation results shown in Fig. 2.3(b,c) are in good agreement with each other but they do not match exactly as in Fig. 2.3(a). This may be the consequence of mean field approach used in writing Eq. (2.15). The mean field approach does not account for the fluctuations in extreme events arising due to fluctuating number of walkers on the network.

We note that Eqns. (2.15-2.17) for the extreme event probability do not depend on the topology of the network. Even though the simulation results are shown for scale-free graphs, it holds good for other types of graphs with random and small world topologies shown in Fig. 2.4(a) and 2.4(b) respectively. However, the difference in probability for extreme events between hubs and smaller degree nodes is not pronounced in the case of random graphs.

The threshold q that defines an event to be extreme depends on the traffic flowing through a given node. The choice $q = \langle f \rangle + m\sigma$ is arbitrary. Now, we show that the extreme event probability in Eq. (2.17) scales with the choice of threshold q or, equivalently, m. In the Fig. 2.5(a) we show $F_m(K)$ for various choices of

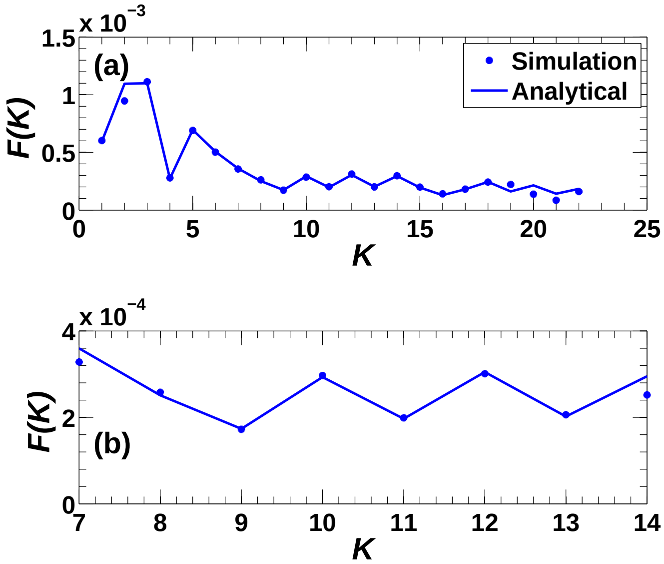

Figure 2.4: Extreme event probability F_K for $\Delta = 0$ for two different networks, (a) random network $(N = 1000, E = 4929)$ and (b) small world network $(N = 1000, E = 5000)$. The threshold for extreme events is defined according to Eq. (2.16) with $m = 4$. The extreme event probability depends on the degree of the node K irrespective of the network topology.

m in log-log scale. Clearly, as m decreases, ignoring the local fluctuations, the curves tend to become horizontal. Physically, this can be understood as follows; $q \to 0$ implies that the threshold for extreme events decreases and this leads to larger number of extreme events and hence higher probability of occurrence. In the limiting case of $q = 0$, $F(K) = 1$ for all nodes and all the events would be extreme.

The graph in Fig. 2.5(a) suggests that it might be scaling with respect to q or m. Starting from Eq. (2.17), we were not able to determine the scaling analytically. Hence, we empirically show that the following scaling relation holds for the probability of extreme event,

$$\frac{F_m(K)}{K^{1-S_m}} = \text{constant}, \tag{2.18}$$

where $F_m(K)$ represents extreme event probability for threshold value q with parameter m. In this, S_m is the slope of the curves $F_m(K)$ in the Fig. 2.5(a). Using Eq. (2.18) on the simulated data for $\Delta = 0$, we find that all the curves for the probability of extreme event, shown in Fig. 2.5(b), collapse into one curve to a

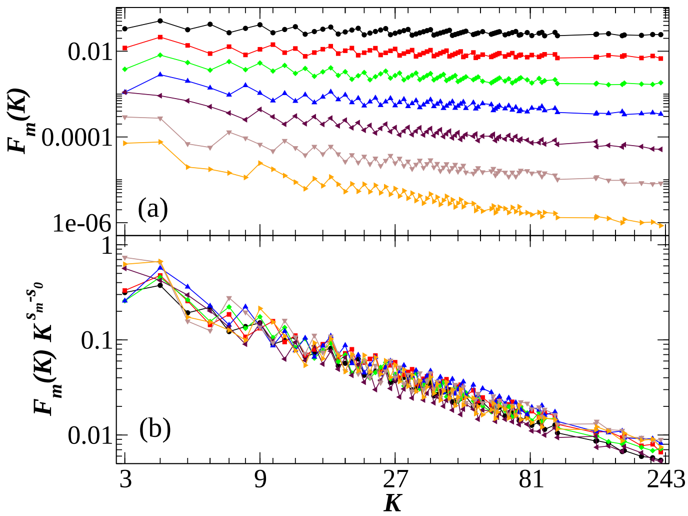

Figure 2.5: Probability for occurrence of extreme events for several values of threshold $q = \langle f \rangle + m\sigma$. (a) shows the extreme event probabilities in log-log plot obtained from simulations with $\Delta = 0$. (b) shows scaling extreme events probabilities. S_0 represents the reference slope with $m = 2$. The thresholds applied for the curves from top to bottom are $m = 2.0, 2.5, 3.0, 3.5, 4.0, 4.5$ and 5.0.

good approximation.

2.4 Return interval distribution

In the study of extreme events, distribution of their return intervals is an important quantity of interest. This carries the signature of the temporal correlations among the extreme events and is useful for hazard estimation in many areas. We focus on the return intervals for a given node of the network. Since the random walkers are noninteracting, the events on the individual nodes are uncorrelated. Then, the recurrence time distribution is given by $P(\tau) = e^{-\tau/\langle\tau\rangle}$, where the mean recurrence time is $\langle\tau\rangle = 1/F(K)$. In the inset of Fig. 2.6, we show $P(\tau)$ obtained from simulations for three nodes with different degrees. In semi-log plot, they reveal an excellent agreement with the analytical distribution $P(\tau)$ (shown as solid line). The main graph of Fig. 2.6 shows the mean recurrence time $\langle\tau\rangle$, the only parameter that characterizes the recurrence distribution, as a function of K and it agrees with the analytical result. The mean recurrence time can be obtained analytically by taking the inverse of extreme event probability calculated using Eq. (2.17).

However, there can be some spatiotemporal correlations among extreme events

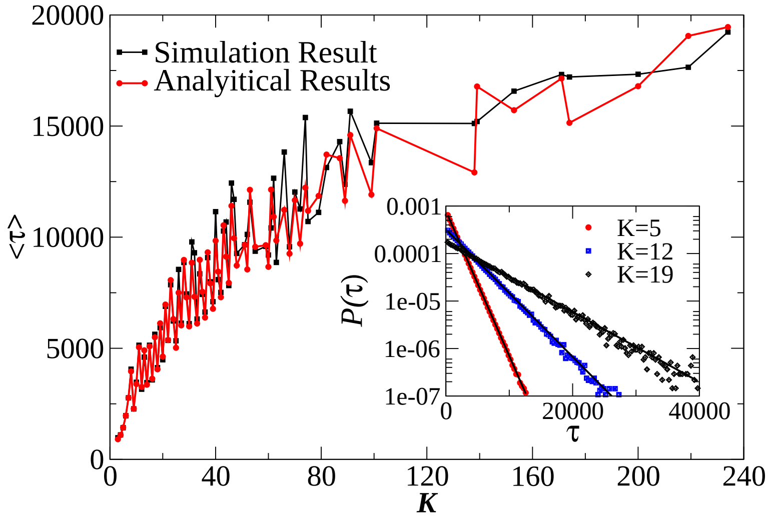

Figure 2.6: The inset shows the recurrence time distribution for extreme events from simulations (symbols) with $\Delta = 0$ for nodes with 5, 12 and 19 links. The solid line is the analytical distribution. The main figure shows the mean recurrence time as a function of degree K.

occurring on different nodes of the network. These correlations may arise due to the network structures and the fact that total number of walkers are constant. Think about the extreme situation, when all the walkers are on a single node i at time $n = 0$. The probability of such situation is infinitesimally small but it provides the arguments in the favor of correlations among extreme events on the network. At $n = 0$ there is only one extreme event occurring on node i. At $n = 1$, when these walkers perform the random walk from node i, the nearest neighbors of node i may experience extreme events. At $n = 2$, some walkers may choose to return to the node i while others get diffused on the rest of the network. So, at $n = 2$, it is more likely that node i may experience the extreme event again. So, if an extreme event of very large magnitude occurs on node i, it is more likely that extreme events may continue to occur on successive time steps on node i and/or in the neighborhood of node i. Once the walkers get diffused on the whole network, the extreme events occur on network almost at random. In the long time limit, statistical correlations among extreme events on networks becomes very small.

2.5 Walk through shortest paths

As pointed out before, many types of flow on the network, such as the information packets flowing through the network of routers and traffic on roads, use more intelligent routing algorithms [76] rather than a random walk. To check the robustness of results in Eqns. (2.15-2.17), we implemented the random walk simulation with the constraint that the traffic from node i to j takes the shortest path on the network. The source i and destination j for a walker are assigned randomly and the total number of walkers on the network are kept constant. In the case of shortest paths, walkers need to have information about the whole network and calculating shortest paths in between all pairs of nodes on a network is a computationally challenging task. There can be more than one shortest paths in between a pair of nodes. In this situation, walkers can choose any one of them with equal probability.

In this setting, for every random choice of source-destination pair the paths are laid out by the algorithm. Once the source and the destinations are chosen, a walker walks on a predetermined path. Hence, the dynamical process is mostly deterministic. The stochasticity arises only when multiplicity of shortest paths is available. Thus, this can be thought of as a walk with large deterministic component. The simulation results with shortest path algorithm [77] shown in Fig. 2.7 are qualitatively similar to the trend displayed in Fig. 2.3. In this scenario of predominantly deterministic dynamics, it is conceivable that the degree of a node does not determine the flux passing through it. This role is played by the centrality of the node with respect to the shortest paths in the network, quantified by the betweenness centrality b of a given node [78, 79, 80].

In general, the definition of betweenness centrality does not include the source and destination nodes. It results into zero betweenness for nodes with degree 1 i.e. nodes on the periphery of the network. Here, while calculating the betweenness, the end points are included.

Based on this qualitative argument, the results in Fig. 2.7 can be understood if we replace Eq. (2.8) with $p = \beta b/B$ where B is a normalization factor that depends on the sum of betweenness centrality of all the nodes on the network. From the numerical simulations, we obtain $\beta \approx 0.94$. After obtaining the stationary

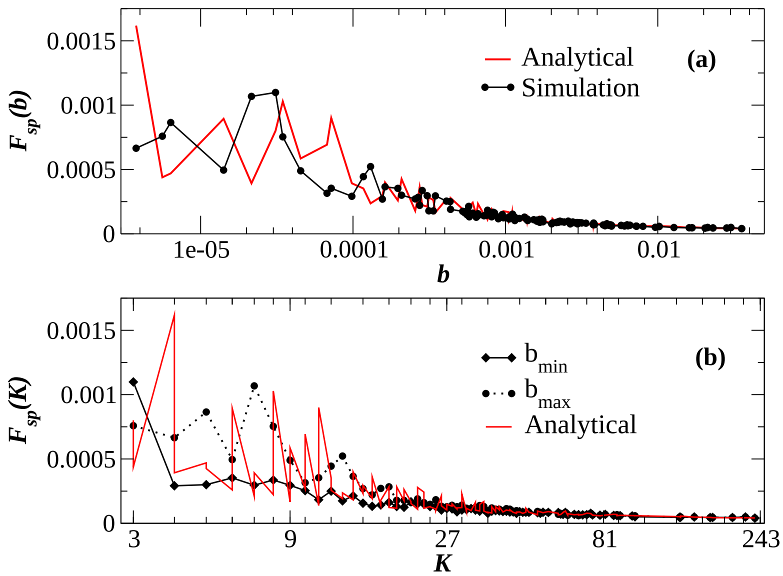

Figure 2.7: Extreme event probability F_{sp} for $\Delta = 0$ with shortest path algorithm implemented for random walkers. The data are plotted in two different ways. (a) $F_{sp}(b)$ as a function of betweenness centrality, (b) $F_{sp}(K)$ as a function of degree K of the node. Nodes with same value of K can have different betweenness centrality. In (b), in order to reduce the clutter, for every value of K, the extreme event probability for the node with largest (b_{max}, solid circles) and least value (b_{min}, solid square) of b is plotted.

probability, we can calculate the probability distribution of the walkers on a node. Again, the distribution is given by the binomial distribution due to the fact that walkers are independent. We can go through the same arguments as before and analytically obtain $\langle f \rangle, \sigma^2, q$ and the probability $F_{sp}(b)$ for occurrence of extreme events. In Fig. 2.7(a), $F_{sp}(b)$ is shown as solid curve. Here, the nodes with smaller betweenness centrality have higher chances of receiving extreme events than the nodes with higher betweenness. In general, on average betweenness centrality is directly proportional to the degree of the node in scale free graph. Hence, the result that hubs are less likely to experience extreme events remains valid. In Fig. 2.7(b), the same data for $F_{sp}(b)$ is shown as a function of K for an easier comparison with Fig. 2.3. Since nodes with the same degree can have different betweenness centralities, we have plotted $F_{sp}(K)$ only for the nodes with maximum betweenness b_{max} and minimum betweenness b_{min} for a given degree. Thus, even with the SP algorithm thrown in, the extreme event probabilities are higher for the nodes with smaller degree ($K < 20$) than for the ones with larger degree ($K > 100$).

2.6 Discussion and summary

This work is an attempt to understand the extreme events occurring on the nodes due to the random flow on the network. In this work, we study random walk model with fluctuating number of walkers and analytically obtain the probabilities for the occurrence of extreme events on the nodes. In this framework, extreme events occur due to the fluctuations in the flux passing through any node and are defined as the exceedances above a chosen threshold q. The threshold is chosen to be proportional to the natural variability of the node. Here, we have shown that extreme events are more likely to occur on the low degree nodes in comparison with the high degree nodes and it is true irrespective of the network topology. The choice of threshold q is crucial in defining the extreme events and it is characterized by the parameter m, which decides the rarity of the extreme events. Here, we obtain the scaling relation for the extreme event probabilities with respect to m. Recurrence time distribution shows that the events on a node are independent from each other and hence, follow exponential distribution. Here, we also looked upon the cases where walkers perform the walk on network via shortest paths in between a pair of nodes and show that the main results associated with the extreme event probabilities remain unchanged and hubs experience less extreme events than the low degree nodes.

Finally we comment on how these results can be applied as a basis to design nodes of a network. The central result in this chapter in Eq. (2.15) allows us to *a priori* estimate the extreme event probabilities. These depend on whether operating principle of dynamics is deterministic or probabilistic. If the idea is to avoid congestion or other problems arising due to extreme events of certain magnitude, then these estimates can be used as an input to the design principles for the nodes. For instance, for the road traffic that operates broadly on the shortest path principle the probabilities can be used as a basis to provision for higher capacity to nodes that will avoid bottlenecks arising from extreme events of a given magnitude.

In scale-free networks, small degree nodes form the bulk, are more prone to encounter extreme events. But network design principles and practice generally focus on the hubs. Such evolved practices might work best for average conditions. Our work suggests that they might fail in the context of extreme events and hence a revised approach is necessary. A careful design for the capacity of small degree nodes is important as well. It must be emphasized that incorporating such extreme event estimates in design principles will only help in better preparedness to meet the expected extreme events. The extreme events discussed here being due to inherent fluctuations will nevertheless take place and cannot be avoided.

Chapter 3

Biased Random Walk and Extreme Events

3.1 Introduction

Random walk on complex networks is a useful fundamental model against which to compare other transport processes. Most realistic transport phenomena on networks, such as the flux of information packets passing through the network of routers or road traffic, do not proceed by performing random walk. In order to model the flux in a more realistic way, it is useful to generalize the standard random walk to a situation in which the flux is either biased toward hubs or small degree nodes. For example, consider the case of two remote airports which are not directly connected by flights. Typically, they would be connected through a major hub on the airline network. This is one practical scenario in which the traffic is biased toward the hubs. This happens in many a network settings; railways tend to connect the hinterland with the hubs, phones connect to nearest hubs on the network. Motivated by these physical examples, in this work, we model the transport process as random walks biased by the topology of the network and study the process from extreme event point of view.

In the world of internet, consider the most common experience of web surfers; a web server not responding due to the heavy load of http requests. This is an extreme event taking place on the network of world wide web. For example, the popular social networking site Twitter handled about 600 tweets per second in early 2010 [81]. According to an industry estimate, the Google search engine received approximately 34000 search requests per second by the end of 2009 [82]. For most websites on the world wide web that are unprepared for such a large number of http requests, these numbers would represent extreme events and could potentially disrupt the service. Grid locks in highways is an example of extreme event on transportation network. From the point of view of physics, all these events could be thought of as an emergent phenomena arising due to flux on the networks and could be regarded as extreme events arising primarily due to the limited handling capacity of the node.

However, extreme events happen not only because of the limited handling capacity of the node on a network but also because of inherent fluctuations in the flux passing through the node. These fluctuations in the flux passing through a node could be so large as they breach a prescribed threshold, in which case, we label the event as an extreme event for the node. Then, a relevant question is how the connectivity of the network affects the probability for extreme event occurrence. By modeling the transport as standard random walks on networks, it was shown in Ref. [83] that the probability for the occurrence of extreme events $P(k_i)$, arising due to inherent fluctuations, depends only on the degree k_i of the i-th node in question. In this work, the threshold q_i was chosen to be proportional to typical fluctuation size on i-th node. Thus, the extreme events are identified after taking care of the natural variability of the flux passing through the given node. Further, it was shown that, on the average $P(k)$ is higher for small degree nodes than for hubs.

Here in this chapter, we study biased random walk on network and calculate extreme event probabilities and event-size distributions. We show that the biased random walk leads to extreme fluctuations in the event sizes on the network. In the subsequent sections, we discuss the topologically biased random walk model on a network and obtain analytical results for the probability of occurrence of extreme events on any node. We show that the analytical and simulation results are in good agreement.

3.2 Biased random walk on networks

Stationary distribution

We consider a connected, undirected, finite network with N nodes and E edges. The network is characterized by a symmetric adjacency matrix A with elements $A_{ij} = 1$ if nodes i and j are connected by an edge and $A_{ij} = 0$ otherwise. There are W independent walkers performing biased random walk on this network in the sense explained below. We denote by b_{ij} the transition probability for a walker to hop from a node i to a neighboring node j. Let P_{ij} be the probability that a walker starting at the node i at time $n = 0$ is at node j at time n. Then, the master equation can be written as

$$P_{ij}(n+1) = \sum_{l} A_{lj}\ b_{lj}\ P_{il}(n). \tag{3.1}$$

The random walkers are biased by taking the time-independent transition probability for hopping from l-th to j-th node to be [84, 85, 86]

$$b_{lj} \propto k_j^{\alpha}, \tag{3.2}$$

where α is a parameter that defines the degree of bias imparted to the walkers. Clearly, $\alpha = 0$ corresponds to the standard random walk and the transition probability is unbiased, where the walker can hop to any of the neighboring nodes with equal probability.

For $\alpha > 0$, the random walkers are biased toward the nodes with larger degree or hubs. The schematic diagram of the walk is shown in Fig. 3.1. In the scale-free network, there are large number of lower degree nodes but most of them are connected to hubs. So, even though the walk is biased towards the lower degree nodes, hubs remain occupied by the walkers. The walk, biased towards the lower degree nodes, helps in spreading the walkers on the network. In contrast, if $\alpha < 0$, walkers preferentially hop to the small degree nodes (refer to Fig. 3.2). There are very few large degree nodes in the scale-free network and it is more likely that they are connected among themselves. Hence, most of the walkers remain confined

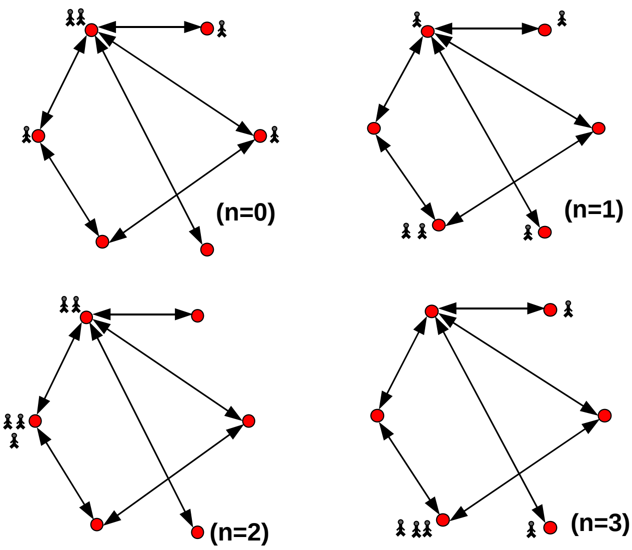

Figure 3.1: Random walk biased towards the lower degree nodes ($\alpha < 0$) is shown here using a schematic diagram. In the scale-free network, walkers visit the large degree nodes regularly though they prefer lower degree nodes.

to the hubs and small degree nodes receive the walkers occasionally. The larger (smaller) the α, stronger the bias toward the hubs (small degree nodes) is. Then, the normalized transition probability becomes

$$b_{lj} = \frac{k_j^\alpha}{\sum_{m=1}^{k_l} k_m^\alpha}. \tag{3.3}$$

The summation in the denominator runs over the nearest neighbors of node l. Using the transition probability in Eq. (3.3), the master equation becomes

$$P_{ij}(n+1) = \sum_l A_{lj} \frac{k_j^\alpha}{\sum_{m=1}^{k_l} k_m^\alpha} P_{il}(n). \tag{3.4}$$

By repeated iteration of Eq. (3.4), it can be shown that $P_{ij}(n)$, as $n \to \infty$ leads to the stationary distribution

$$\lim_{n\to\infty} P_{ij}(n) = p_j = \frac{k_j^\alpha \sum_{l=1}^{k_j} k_l^\alpha}{\sum_{m=1}^{N} \left(k_m^\alpha \sum_{l=1}^{k_m} k_l^\alpha \right)}. \tag{3.5}$$

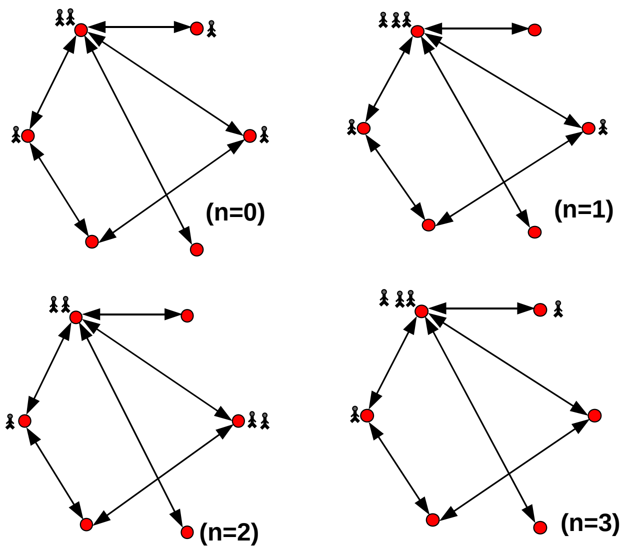

Figure 3.2: Random walk biased towards the hub ($\alpha > 0$) is shown here using a schematic diagram. Walkers remain confined to the large degree nodes.

We can define the generalized strength of j th node to be

$$\phi_j = k_j^\alpha \sum_{i=1}^{k_j} k_i^\alpha, \tag{3.6}$$

which is a measure of the ability of a node to attract the walkers. Note that ϕ_j depends on the bias parameter α and the degree of the nearest neighbors to which it is connected by an edge. Hence, it is possible for the nodes with the same degree to have different generalized strengths. Thus, the generalized strength of the node is independent of the global network structure but is dependent on the local connectivity structure around the node. This is in contrast to the case of standard random walk (on networks) in which the large-scale structure of the network topology plays no significant role. The local network structure is important for biased random walks on networks. In Fig. 3.3, we show how the generalized strength ϕ depends on the degree of a node, for several values of α, in a scale-free network with degree exponent $\gamma = 2.2$. For $\alpha = 1$ (crosses in Fig. 3.3), the generalized strength of a node is higher for large degree nodes (hubs) and an approximate linear relation is seen between ϕ_i and k_i . For $\alpha = 0$, which is the standard random walk case, the generalized strength of the node is the same as the degree of the node (solid circles in Fig. 3.3). However, for $\alpha = -1.0$, ϕ is independent of k especially for the large degree nodes (triangles in Fig. 3.3). In this case, the bias in the random walk represented by its generalized strength ϕ is balanced by the degree of the node. In a scale-free network, a large number of small degree nodes are present and they do not have identical values for the generalized strength ϕ. This explains the spread in ϕ for all values for $k < 50$. Upon further decrease in the bias parameter α below -1.0 (open squares in Fig. 3.3), nodes with a smaller degree or neighbors with smaller degree become important and the generalized strength decreases with increasing degree.

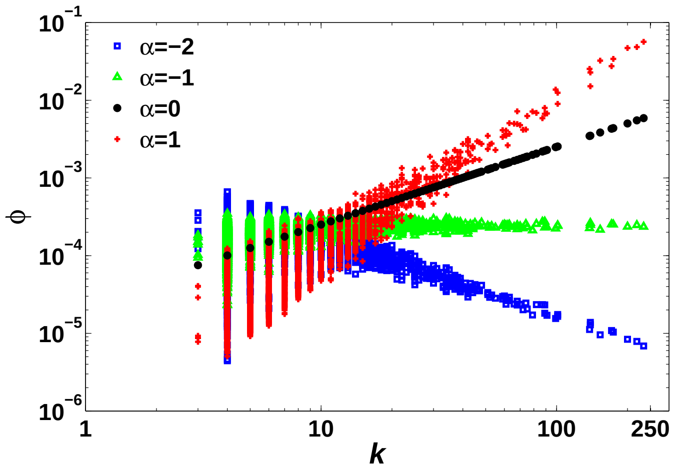

Figure 3.3: Normalized strength ϕ as a function of degree k for different values of α in log-log plot.

Extreme event probability

The stationary distribution for the number of walkers in j-th node can be rewritten in terms of the generalized strength ϕ as

$$p_j = \frac{\phi_j}{\sum_{l=1}^{N} \phi_l}. \tag{3.7}$$

Thus, every node can be uniquely characterized by its generalized strength ϕ. It is expected that two nodes with the same value of ϕ show similar behavior as far as biased walks on networks based on Eq. (3.2) are concerned. In the case of $\alpha = 0$, we get $\phi_i = k_i$ and the stationary distribution simplifies to $p_j = \frac{k_j}{2E}$, the result obtained for the case of standard random walk in Ref. [68]. Thus, in the case of a standard random walk, the degree k characterizes the node. In the case of uncorrelated random networks, the stationary occupation probability can be further simplified by using the mean field approximation and can be written as [84, 85]

$$p_j = \frac{k_j^{\alpha+1}}{N\langle k^{\alpha+1}\rangle}. \tag{3.8}$$

This approximate result suggests that the nodes with the same degree should have identical transition probabilities [84]. This is not necessarily as good for the nodes of correlated networks such as the scale-free networks. This is because in a scale-free network, the neighborhood of nodes with identical degrees are not identical. Hence, to study extreme events we use Eq. (3.7) instead of Eq. (3.8).

Given that Eq. (3.7) gives the probability to find one walker on i-th node with generalized strength ϕ_i, we can now obtain the distribution of random walkers on i-th node. The formulation is applicable to any node on the network and hence, in our further discussions, we suppress the index i of the node. Random walkers are independent and non-interacting and hence the probability $f(w)$ of finding w walkers on a node is p^w while the rest of the walkers, $W - w$ are distributed on the rest of the nodes of the network. When properly normalized, this leads to a binomial distribution given by

$$f(w) = \binom{W}{w} \, p^w \, (1-p)^{W-w}. \tag{3.9}$$

The mean and variance of the flux passing through a given node is

$$\begin{aligned} \langle f\rangle &= W\frac{\phi}{\sum_{l=1}^{N}\phi_l}, \\ \sigma^2 &= W\frac{\phi}{\sum_{l=1}^{N}\phi_l}\left(1-\frac{\phi}{\sum_{l=1}^{N}\phi_l}\right). \end{aligned} \tag{3.10}$$

Note that the results in Eqs. (3.9) and (3.10) depend only on the generalized strength ϕ that characterizes a node including its neighborhood. It does not depend on the large scale connectivity pattern. Hence, these results will hold good for any network, such as scale-free, random or small world, irrespective of its degree distribution. Further, in the cases for which $\sum_{l=1}^{N} \phi_l >> \phi$, we obtain the approximate relation $\sigma \approx \langle f \rangle^{1/2}$. This relation can be thought of as a generalization of a similar relation for the unbiased random walks reported in Ref. [83]. However, the exponent $1/2$ is not universal and instead depends on details such as the fluctuation in number of walkers and the sampling resolution of the flux [74]. The distribution of random walkers on two nodes with different degrees, $k = 4$ and $k = 234$, is plotted in Fig. 3.4. The biased random walk simulations were performed on a scale-free network with 5000 nodes with 19915 links and 39830 walkers. Initially, at time $n = 0$, the walkers are randomly distributed on N nodes. The simulation results presented in Fig. 3.4 have been obtained after averaging over 100 realizations with different initial conditions of random walkers. The simulation results, the solid lines in Fig. 3.4, show a good agreement with the analytical distribution given by Eq. (3.9).

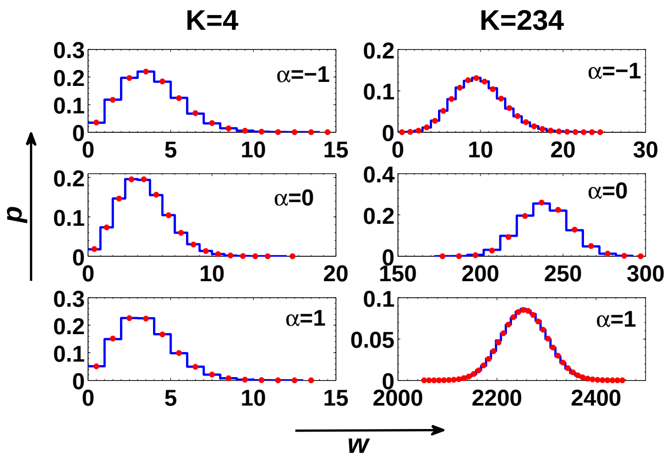

Figure 3.4: The distribution of walkers on two nodes with $k = 4$ and $k = 234$ for $\alpha = -1.0, 0.0$ and 1.0. The solid lines show the distribution of walkers obtained from simulation while the solid circles belong to the binomial distribution obtained analytically using the stationary probability in Eq. (3.7).

3.3 Probability for extreme events

We take an extreme event to be the one for which the probability of occurrence is small and is typically associated with the tail of the probability distribution function for the events. We extend this principle to the events on the nodes of a network [83]. Given that the number of walkers w passing through a node with generalized strength ϕ follow the Binomial distribution, if more than q walkers pass through the node, then it is an extreme event for the node. Then, the probability for the occurrence of extreme event is

$$F_i = \sum_{w=q_i}^{W} \binom{W}{w} p_i^w \, (1-p_i)^{W-w}, \tag{3.11}$$

$$= I_{p_i}(\lfloor q_i \rfloor + 1, W - \lfloor q_i \rfloor), \tag{3.12}$$

where $\lfloor u \rfloor$ is the floor function defined as the largest integer not greater than u and $I_z(a,b)$ is the standard incomplete Beta function [87]. In this form, the extreme event probability will depend on the choice of threshold q_i. First, we consider the case of constant threshold. If $q_i = 0$, using Eq. (3.11) we obtain $F_i = 1$ for all the nodes on the network. Thus, all the nodes will experience extreme events all the time. On the other hand, if we set $q_i = W$, then we obtain

$$F_i = p_i^W. \tag{3.13}$$

Since $p_i << 1$, we get $F_i \approx 0$ for all the nodes implying that there are no extreme events anywhere in the network. Hence, these choices of threshold values are not physically interesting cases. Any other arbitrary choice such as $q_i = q_0$, where q_0 is a constant, will predominantly lead to some nodes encountering extreme events nearly all the time and others having no events at all. This too is not an interesting case. The foregoing arguments imply that an interesting scenario would arise if the threshold is chosen to be proportional to the natural variability of the flux passing through a node. Thus, we choose the threshold for extreme events to be [83]

$$q_i = \langle f_i \rangle + m\sigma_i, \tag{3.14}$$

where $m \geq 0$. The mean flux $\langle f_i \rangle$ and standard deviation σ_i are given by Eq. (3.10). Substituting q_i in Eq. (3.12), it is clear that the probability for the occurrence of extreme events is dependent only on the generalized strength ϕ of the node. In

Fig. 3.5, we show the simulation and analytical results for the probability of extreme events as a function of ϕ for several choices of α. The numerical results are based on simulations with $W = 39380$ walkers on a scale-free network with $N = 5000$ nodes evolved for 10^7 time steps. An unusual feature is that F_i predicts higher probability of occurrence of extreme events, on average, for nodes with small values of generalized strength ϕ than for the nodes with higher values of generalized strength ϕ. For instance, in Fig. 3.5(a), the probability of extreme event occurrence is generally higher for nodes with $\phi < 10^{-5}$ than for nodes with $\phi > 10^{-3}$. A similar effect is seen in Figs. 3.5(b)- 3.5 (e). Even though the nodes with higher generalized strength ϕ attract more walkers as given by Eq. (3.5), this does not imply that they also have higher probability for extreme events. This is a generalization of the result obtained in Ref. [83] for the standard random walk on networks which shows that the extreme events are more probable for nodes with small degree than for the ones with high degree. The local fluctuations seen in Fig. 3.5 are inherent in the system and not due to insufficient ensemble averaging. Further, notice that Eq. (3.12) does not depend on the large scale structure of the topology and hence it is valid for biased random walks on any topology, random or small-world or scale-free.

However, the local connectivity patterns in the vicinity of any node plays a crucial role in the diffusion of an extreme event. Suppose an extreme event takes place at node A at time n, then one interesting question is how probable it is for an extreme event to take place in its immediate neighborhood at time $n+1$, i.e, after the first jump. We call it the first-jump probability and it is similar to the one reported in [52]. In the case of a standard random walk ($\alpha = 0$), our simulations (not shown here) indicate that in general if the node A is a hub, then the probability to encounter an extreme event in its neighborhood is higher (at least by a factor of 3-4) compared to the case when the node A is a small degree node. For biased random walks, the results suggest a higher likelihood for an extreme event to be transferred to its neighborhood in the case when $\alpha < 0$ compared to the case when $\alpha > 0$.

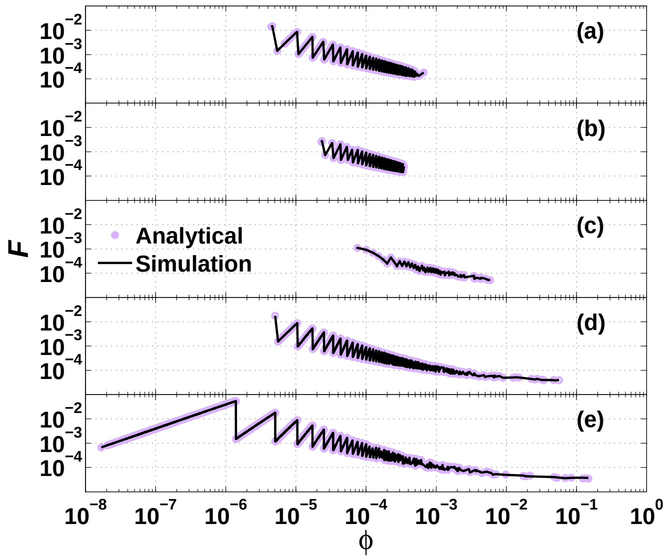

Figure 3.5: The probability for the occurrence of extreme events plotted as a function of node generalized strength ϕ (normalized) for different values of bias parameters (a) $\alpha = -2.0$, (b) $\alpha = -1.0$, (c) $\alpha = 0.0$, (d) $\alpha = 1.0$ and (e) $\alpha = 2.0$. The threshold for extreme event is $q = \langle f \rangle + 4\sigma$. The circles are from analytical results in Eq. (3.12) while solid lines are the simulation results performed on a scale-free network ($N = 5000$, $E = 19915$) with $W = 2E$ walkers averaged over 100 realizations with randomly chosen initial positions of walkers.

3.4 Fluctuations in event size

The size of an event is measured in units of the standard deviation σ of the flux passing through a node. In this section, we show that the extreme fluctuations in the flux of walkers are realized in the case of $\alpha = 2$ which implies that the walkers are biased toward the nodes with larger generalized strength ϕ (hubs). An event on a given node is of size m if $m\sigma \leq w - \langle w \rangle < (m+1)\sigma$, where w is the number of walkers on the given node.

Then, the probability for the occurrence of an event of size m can be written down as,

$$\mathcal{P}_m = I_p(\lfloor q_m \rfloor + 1, W - \lfloor q_m \rfloor) - I_p(\lfloor q_{m+1} \rfloor + 1, W - \lfloor q_{m+1} \rfloor). \tag{3.15}$$

To illustrate the result, we show the distribution of event sizes in Fig. 3.6 for $\alpha = -2, -1, 0, 1, 2$ in a scale-free network obtained from simulations evolved for

10^7 steps and averaged over 100 ensembles. Here, the events with probability of occurrence of less than 10^{-8} have been discarded to maintain the numerical accuracy. In the case of $\alpha = 0$ (standard random walk), the distribution of events is shown in Fig. 3.6(c). The events of size $m = 0$ are highly probable with $\mathcal{P}_0 \sim 0.1$. In contrast, the probability for the events of size $|m| > 0$ decrease and thus the extreme events of size $m = -2, 8$ occur with probability $\mathcal{P}_{-2} \sim \mathcal{P}_8 \sim 10^{-8}$. The limitation on the lower limit of event sizes is restricted by the minimum possible number of walkers on a node, i.e., 0. For lower degree nodes, events of sizes -2σ to 8σ are observed but in the case of higher degree nodes $k > 100$, events sizes range from -5σ to 6σ only. In the case of a standard random walk, for the whole network, event size m varies from -5σ to 8σ.

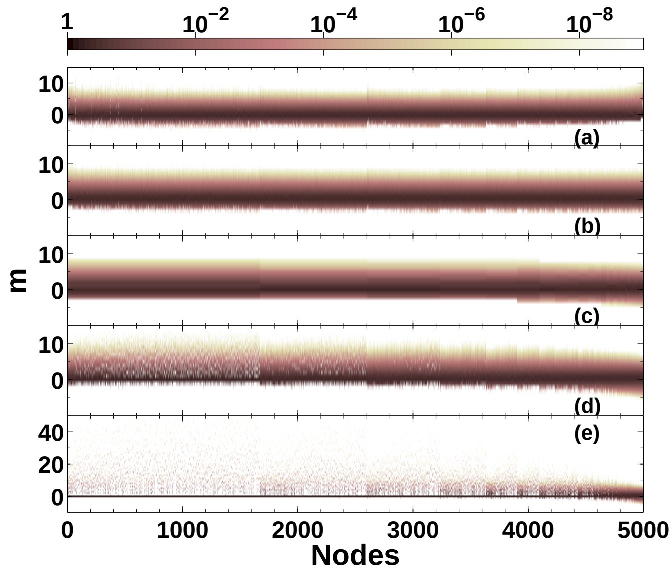

Figure 3.6: The distribution of event sizes for biased random walks as a function of node number on x-axis obtained from simulations performed on a scale-free network for different values of bias parameters (a) $\alpha = -2.0$, (b) $\alpha = -1.0$, (c) $\alpha = 0.0$, (d) $\alpha = 1.0$ and (e) $\alpha = 2.0$. The nodes are arranged in the order of increasing degree. The probability values $\mathcal{P}_m$ are color coded. This should be compared with analytical results in Fig. 3.7.

In comparison, for the case of $\alpha = 1$ shown in Fig. 3.6(d) the events of size 8 have higher probability of occurrence ($\mathcal{P}_8 \sim 10^{-7}$) and the events of even higher sizes are also possible. For $\alpha = 2$, even higher size events, as large as 40, become highly probable for small degree nodes as seen in Fig. 3.6(e). Thus, in general, for larger α, larger size events become probable when compared with the case of $\alpha = 0$. Physically, this can be understood as follows. With $\alpha = 0$, the random walkers perform unbiased random walk. However, for $\alpha = 2$, the walkers preferentially choose to hop to the nodes with larger degree (hubs). Since the large degree nodes are mostly well connected among themselves, very few walkers reach small degree nodes. Hence the average flux through the small degree nodes becomes so small that even occasional visits by a few walkers lead to extremely large size events. These occasional visits lead to probabilities of order 10^{-6} even for events of size 40. Hence, in the case of biased random walks, extremely large fluctuations in event sizes can be observed in small degree nodes. This effect is also seen in the analytical results obtained using Eq. (3.15) shown in Fig. 3.7.

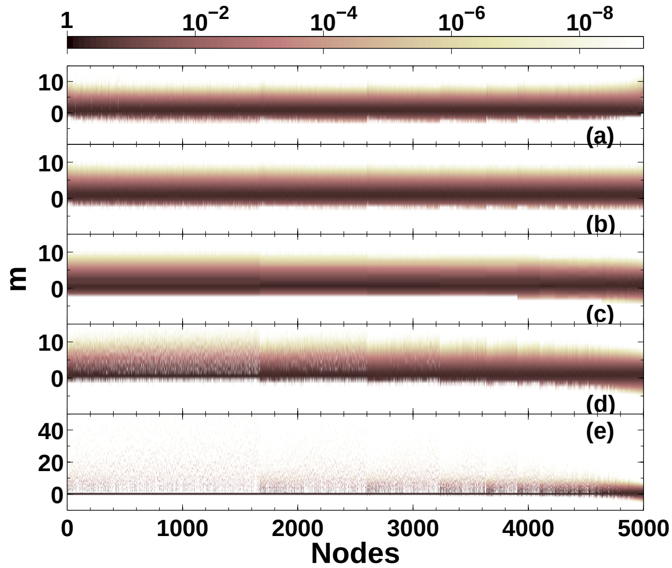

Figure 3.7: The distribution of event sizes for biased random walks as a function of node number on x-axis obtained analytically using Eq. (3.15) for different values of bias parameter (a) $\alpha = -2.0$, (b) $\alpha = -1.0$, (c) $\alpha = 0.0$, (d) $\alpha = 1.0$ and (e) $\alpha = 2.0$. The nodes are arranged in the order of increasing degree. The probability values $\mathcal{P}_m$ are color coded.

On the other hand, for the cases $\alpha = -2, -1$ such large fluctuations are not visible in the event sizes in Fig. 3.6(a) and 3.6(b). For $\alpha = -1$ in Fig. 3.6(b), there is a small increase in the event sizes (when compared to $\alpha = 0$) for the small degree nodes but it is not as large as in $\alpha = 1$ case. Further, with $\alpha = -1$, it must also be noted that the probability profile remains similar for most of the nodes irrespective of the large differences in their degree. This is because ϕ is an approximate constant for most of the nodes since, in this case, the effect of the bias is balanced by the degree of these nodes. For $\alpha = -2$, the flux is strongly biased towards small degree nodes and again events of sizes $m = 10$ can be seen in Fig. 3.6(a) though only on the higher degree nodes. The event sizes for hubs are not as large as observed in the case of $\alpha = 2$ for lower degree nodes. It can be explained as follows; when $\alpha = -2$, the flux preferentially flows through the small degree nodes which form the bulk in a scale-free network. Most small degree nodes do not have a direct link with other small degree nodes but are connected through a hub. Hence, despite the biased walk favoring the small degree nodes, sufficiently large flux flows through the hubs as well. Hence, abnormally large event size fluctuations are not seen in hubs for $\alpha = -1, -2$. All these features show a good agreement with the analytical result obtained in Eq. (3.15) and shown in Fig. 3.7.

3.5 Discussion and summary

This work is an attempt to understand the extreme events occurring on the nodes due to flow on networks which typically is directed toward or away from the hubs. In this work, we study a biased random walk model in which the traffic preferentially moves either toward or away from the hubs and we analytically obtain the probabilities for the occurrence of extreme events. In this framework, extreme events are due to inherent fluctuations in the flux passing through any node and is defined as the exceedances above a chosen threshold q. The threshold is chosen to be proportional to the natural variability of the node. Each node on the network is characterized by generalized strength ϕ which depends on its degree and that of its immediate neighborhood. It is a measure of how much traffic is attracted to the particular node. The larger the generalized strength of a node is, larger is its ability to attract walkers. In this chapter, we have shown that the nodes with a smaller generalized strength, on an average, have a higher probability for the occurrence of extreme events when compared to nodes with higher generalized strength. Further, we have also shown that when the flux is biased toward the hubs, abnormally large fluctuations in event sizes become highly probable. This is one possible signature of the topologically biased flow in a scale-free network.

In general, it is possible to conceive of many ways by which bias can be imparted to the independent random walkers on networks. These biasing strategies are motivated by real observations and the quest for efficient search strategies on networks. Various kind of biases based on the local environment, shortest paths, the entropy of random walk and various adaptive strategies are some examples of biased random walk on networks [52, 58, 88, 89, 90, 91]. It will be interesting to study the extreme event probabilities under such biasing strategies. However, we emphasize that if the stationary probability distribution equivalent to Eq. (3.5) exists for all the above strategies, then it would be possible to define extreme events and analyze them following the methods presented in this work.

In the context of scale-free network, it has been argued that hubs are important for better functioning of the network. Apart from being responsible for providing better connectivity, existence of hubs makes the scale-free network robust against the random node removal but fragile if the node removal is targeted [92, 93]. The results in this chapter show that extreme events due to natural fluctuations are more probable on the small degree nodes (when compared to the hubs). Hence special attention must be paid to designing the capacity of the small degree nodes so that extreme events can be smoothly handled without leading to disruption of the node. The results in this chapter can be used to estimate the capacity a node should possess if it should handle extreme events of size, say, m. If we want the node to handle 4σ events smoothly, then the required capacity can be obtained by inverting Eq. (3.12). Thus, the numbers so obtained can be useful as an input for arriving at a capacity to be built for the nodes on a network.

Chapter 4

Network vulnerability to Extreme Events

4.1 Introduction

Complex networks are everywhere and they support important dynamical processes on them. Hence, smooth functioning of the networks is not only desirable but also essential for everyday life. Congestion on roads, slowdown of Internet, power blackouts etc are a few examples of unwanted situations which appear on the networks. Sometimes, these undesirable scenarios are not confined to a small part of the network but tend to cover the whole network and lead to serious economic and social consequences. It is therefore important to understand the mechanism behind such large scale failures on the networks.

In the past, the structural robustness of different networks have been studied against the external attacks on the node/links i.e. the removal of a finite numbers of nodes and/or links. It has been concluded that networks with large heterogeneity in their node-degrees, such as scale-free networks, are more robust against the random removal of the nodes than the homogeneous networks (random networks, small-world networks). In contrast, the heterogeneous networks prove to be more vulnerable against the deliberate attacks on the high degree nodes [94]. A fully connected network always remains fully connected the against node removals.

Later, the dynamics on the network was introduced and if the flow passing through the node exceeds its capacity, the node is removed from the network. Under such rules, the effect of a single or multiple node failures on the network topology was examined [95]. In these studies, it has been observed that such large scale collapses can happen due to the breakdown of a single node or link. Most of the times, the networks remain stable against small failures but sometimes major damages can be triggered by the small shocks arising due to external factor.

In this work, we explore the network failure mechanism based on the extreme events on the nodes in the network. After encountering an extreme event, a node stops working and is deleted from the network. The failure of a single node due to an extreme event on it, causes the redistribution of the load on the neighboring nodes. It enhances the load on rest of the nodes in network and the extreme events on network become more probable. The process continues until the whole network collapses or an steady state is reached. Since, these extreme events occur due to the internal fluctuations in the dynamics, the large scale collapses on networks can occur due to the dynamical process itself. A real life example of such failure is the power-grid failure where a transmission line gets overloaded [96] and causes the shutdown of a whole power grid. To avoid such failures, we need to understand the route to the state of complete network failure. Here, we discuss the nature of the network failure and show that even a fully connected network is not robust against the extreme events occurring in the dynamics taking place on network.

In this chapter, before we discuss the model of catastrophe based on extreme event, a brief review of the earlier approaches to study the robustness of the networks is presented. The earlier models relating to the robustness of the networks can be categorized in two main classes. These two classes are distinct in terms of physical quantities affecting the network.

Structural robustness

The simple model of the network failure is based upon the removal of the nodes/links from a network and studies the impact of it on the topological structure of the network.

In the year 2000, a comparative study of topological effects of node removal (random and targeted) was made on homogeneous and heterogeneous networks [94]. Each time, when a node is removed, all the links of the node are also deleted and the connectivity properties of the network is measured. The quantitative measure of damage is given by the relative size of the largest connected component to the original size of the network. As long as the largest connected component remains comparable to the original size of the network, the network has the ability to perform its tasks but when the largest connected cluster breaks into smaller clusters, network is considered to be non-functional. Results show that the heterogeneity (in terms of degree) in the network provides more protection to the network against the random removal of the nodes but the same heterogeneity in the network can be held responsible for the failure of the networks against the targeted attacks (removal of high degree nodes). There have been many analytical and numerical studies which discussed the tolerance of the network to the error and attacks [35, 92, 93, 97, 98].

This approach to study the structural robustness of the system does not include the detailed technical effects such as flow, capacity etc on the network but provides initial insight into the large scale failures on the network. Later, the concept of loads and capacities was introduced and models with these features could capture the cascading feature of the network failures. Cascades are observed during the failure of a power-grid network. The models incorporate the fact that networks can have dynamics and due to this dynamics, the load of a failed node should be redistributed on the network.

Dynamical robustness

The structural robustness was studied using the static properties of the networks however the main aim to understand the network robustness remained incomplete because the dynamical processes taking place on the networks were not taken into account. In a network, the components, nodes/links can support different type of flows but the quantity of the flux passing through them is often limited due to their transmission capacities. There can be financial or physical reasons behind these finite capacities. If the flow property and capacity of a network, are taken into account, the failure of a single component due to external factors can have serious consequences.

These models argue that failure of a component causes the redistribution of the flow over rest of the network. It results into some extra load on the components. Not all the components can tolerate the excessive load due to their limited capacities and hence, their failure may trigger a cascade of overload failures.

In 2002, fiber bundle model [63] on scale-free network was studied to model cascading failures[99]. In this model, the capacities of the nodes are randomly distributed using weibull distribution and a constant external pressure (load) is imposed on each node of the system. All the nodes with capacities less than the load fail in the first instance. The individual load carried by each of the failed node is equally transferred to their surviving neighbors which may result in secondary failures and then, tertiary failures and so on. The equilibrium is finally attained when there are no more failures on the network. The results indicate that at the critical load, network loses its functionality and the final breakdown of the network arises more abruptly and catastrophically. In 2003, Moreno et al discussed the updated version of a similar model [100] but there are several important differences. In this model, the rules are performed on the links and, in one case, the redistribution of the load among the neighbors of the failed node is done randomly. The results point out that above the critical average load, any small failure in scale free network leads to the whole network collapse.

In a network, the average load and the capacity of the components play a major role in cascade failures. In a simple model, suggested by Motter and Lai, the capacity of a component is assumed to be proportional to the betweenness centrality [95]. The model shows that global cascades in a network with the capacities defined according to the flux passing through the component can either be triggered by high degree nodes or the nodes with high load or both. Subsequent studies introduced more realistic redistribution mechanism and also used different measures for the capacities of the components of a network [101, 102, 103, 104, 35, 105, 106]. Watts modeled cascades in a system where an agent on the network can choose one of the binary states and also, influence the neighbors [107]. The model finds its usefulness in the social and economic systems. In 2010, the cascading behavior was studied in interdependent networks [108]. These networks can not be considered as a single large network because they have their own dynamics. Authors studied the vulnerabilities of the interconnected network, the electrical network spatially coupled with the Internet network.

All these studies were performed under the assumption that a node/link of the network fails due to some external reason. Then due to the redistribution mechanism, failure spreads and affects the whole network. After this brief review, we study the robustness of the networks against extreme events taking place on it due to internal fluctuations.

Network robustness against extreme events

The mechanisms behind the growth of local failure vis-a-vis the global failure have been extensively studied in the past but the reason behind the first node failure is always assumed to be external. Here, we argue that the first trigger responsible for the failure of a network need not be external, it can be an outcome of the dynamics taking place on the network. Extreme event on a node can cause the node failure and also the network failure. The model based upon the extreme events is not only a different model from the earlier ones but also it is a realistic model of network failures.

We consider a connected, undirected finite network consisting of N nodes and E links. The dynamics on the network has been modeled using the standard random walk model with multiple random walkers. There are W walkers performing random walk on the network and a walker on node i can hop to any of the neighboring nodes with equal probability. The random walk model is discussed in details in chapter 2. As discussed earlier, the distribution of the walkers on a node is a binomial distribution with probability $p_i = k_i/2E$. The mean and the variance of the number of walkers passing through the node i is given by,

$$\langle f_i \rangle = W\frac{k_i}{2E} \tag{4.1}$$

$$\sigma_i{}^2 = W\frac{k_i}{2E}\left(1 - \frac{k_i}{2E}\right). \tag{4.2}$$

An event is called an extreme event on node i at time t if the flux passing through the node exceeds a given threshold q_i. The threshold can be defined in many ways but in the network context, it is natural to define it as

$$q_i \;=\; \langle f_i \rangle + m\sigma_i. \tag{4.3}$$

where m can be any real number and the numerical value of m decides the rarity of an extreme event. The threshold, defined in Eq. (4.3), is function of the average flux passing through a node and the variance in it. Hence, it not only takes care of the load(here, W walkers) on the network but also incorporates the fluctuations in the flux passing through a node.

The number of walkers on a node fluctuates with time and when it exceeds the capacity of the node defined as Eq. (4.3), it experiences an extreme event, and the probability of such an event is,

$$F(K) = \sum_{k=\lfloor q \rfloor+1}^{W} f(K) = I_p\left(\lfloor q \rfloor + 1, W - \lfloor q \rfloor\right), \tag{4.4}$$

where $I_p(.,.)$ is the incomplete beta function.

4.2 Model of network failure

There are W walkers on the network and they walk on the nodes following the rules of random walk on a network and extreme events occur on different nodes due to the fluctuation in number of random walkers on the nodes.

In our model, a node after experiencing an extreme event stops functioning and it sheds all its incoming links with the neighboring nodes. So that, at the next step, walkers present on the node could walk out to its neighboring nodes as per the rules of the dynamical process, however, due to the dropped incoming links, the node does not receive any walker. Hence, the node stops supporting the dynamics taking place on the network and gets deleted from the network.

The random walk continues on the modified network but the capacities of the nodes on the modified network are not altered. A cartoon of the model is shown

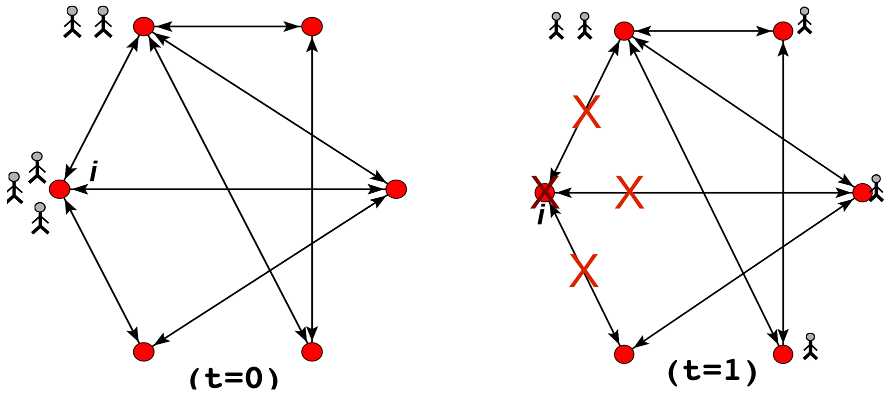

Figure 4.1: A network failure model is proposed based on extreme events. At $t = 0$, extreme event occurs at node i and it gets deleted from the network at the next time step i.e. $t = 1$. Before the node is deleted, walkers present on the node move to neighboring nodes as per the dynamical rules.

in Fig. 4.1. At $t = 0$, an extreme event occurs of node i. At $t = 1$, the node i and its links are deleted from the network and walkers continue to walk randomly on the modified network.

In this process, a node may get isolated on the network due to its non-functioning neighbors. In such situation, walkers present on the isolated node, stay there for infinite time and do not participate in the dynamics. Walkers on the isolated nodes can be considered as dead.

The dynamical process stops with the collapse of the network which means that there does not remain a single node which could support the dynamical process. The dynamics can also stop when there does not remain a single walker alive to perform the random walk which is highly improbable. In either case, the network stops functioning and fails.

The model has many qualitative differences with other models used to study the network failures. It incorporates the internal dynamics of the system and does not require any external factor unlike other models of dynamic failures. The failure mechanism is based upon the fluctuations in walkers on the node and these large fluctuations may arise due to the regular dynamics, following the extreme event probability given by (refer to chapter 2), or due the redistribution dynamics after the failure of a node, or due to both. In previous models, the redistribution of the load among neighbors was either uniform or random. In this model walkers always follow the rules of the dynamics and therefore, the redistribution of walkers is also governed by rules of the dynamical process. Hence, the model provides the generalized framework to check the robustness of the network against the dynamical process taking place on it.

In previous models, an external trigger is provided by the removal of a single node or a group of nodes and based on the properties of this nucleation node, the network may fail or survive. But here, due to the internal dynamics of the network, failure does not depend on the properties of the initial node failure. Hence, it lacks the concept of nucleation points.

Also, in this model, capacities of the nodes take care of the total load presented on the network and also, incorporates the fluctuations present in the network dynamics. Hence, the network failure does not depend on the initial load on the network. The network failure model based on extreme events proposed in this work captures most of the features of the network failures discussed in earlier studies.

To begin with, we check the robustness of a completely connected network against the random walk dynamics and in the next section, we show that even all-to-all network fails to support the random walk dynamics for infinite time.

4.3 Failure of an all-to-all network

We consider a fully connected network made of $N = 100$ nodes and $E = 4950$ links. There are $W = 9900$ walkers on the network performing the random walk. In a fully connected network, all the nodes have the same degree $K = 99$ and hence, have equal capacities, $q = 148$. The capacity of node is defined as in Eq. (4.3) with $m = 5$. The total capacity of the network is defined as

$$Q = \sum_{i=1}^{N} q_i. \tag{4.5}$$

Then, for our choice of W and N, we have

$$Q = Nq \sim \frac{3W}{2}. \tag{4.6}$$

The total capacity of the network is approximately 1.5 times the total load on the network. The first extreme event strikes a node on the network with probability $F = 1.49 \times 10^{-6}$, calculated using Eq. (4.4). The probability that a extreme event can occur anywhere on a fully connected network is,

$$F_N = 1 - (1 - F)^N \approx FN, \tag{4.7}$$

which means that first node from the network gets deleted with probability $F_N = 1.49 \times 10^{-4}$. Here in the case of fully connected network, the probability of the occurrence of an extreme event on any node always remains same. Another advantage of a fully connected network is that it always remains an all-to-all network even after the nodes are removed. However, the degree of the nodes decreases with every node removal.

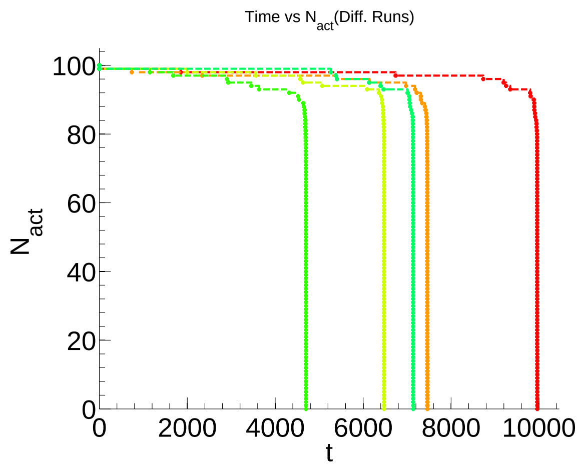

Figure 4.2: Number of active nodes on a network as a function of time. The capacity of each node is $q = \langle f \rangle + 5\sigma$. Different colors represents different initial conditions of walkers on the network.The simulation is performed on a fully connected network with $N = 100$ nodes, $E = 4950$ links and $W = 2E$ walkers.

In Fig. 4.2, the number of active nodes N_{act} is plotted against time and different realizations are shown using different colors. Active nodes are defined as the nodes which participate in the network dynamics at a given time. In each realization, the clock starts as soon as the first extreme event strikes the network, so at $t = 1$, the first node fails and $N_{act} = N - 1$. Solid dots indicate that a node has been removed at the time. Concentrating on any realization (say, red line) it is found that initially, the nodes fail at a slow rate but with every failure, the failure rate increases smoothly until a critical number of nodes get deleted. After that, there is a sharp decrease in the number of active nodes and the entire remaining network collapses in a very few time steps. It shows that even fully connected networks are not very robust against the extreme events which are inherent in the dynamics supported by the network. A similar behavior can be observed in other realizations(different colors). A sudden collapse of the network is a very robust feature against different initial settings. The collapse occurs at different times in different realizations due to the stochasticity involved in the dynamics.

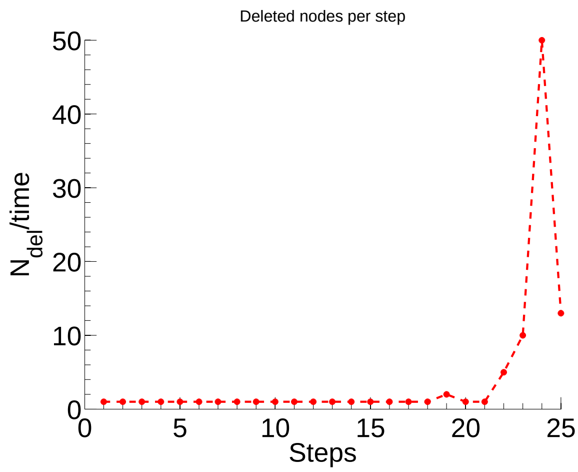

Figure 4.3: Number of deleted nodes on a network at each step. The step is a time at which at least one extreme event has occurred on the network.

To understand the dynamics of node failures, the number of deleted nodes is plotted against the steps in Fig. 4.3. A step is defined as the time at which at least one node gets removed from the network. It shows that initially there was one node failure at each step and the behavior continues until 20 nodes get deleted. After these 20 steps, number of deleted nodes at each step increases and at around 23^{rd} steps, almost 50 nodes fail at a single time which results in a network collapse. The remaining small number of nodes also get deleted after this big event. In Fig. 4.4, the average number of failures per step (red dots) is plotted. The average is over 200 different initial conditions of random walkers. Maximum and minimum number of failures per step is represented using error bars (Blue solid lines). The average does not depict the correct picture but it clearly shows that a large number of nodes fail

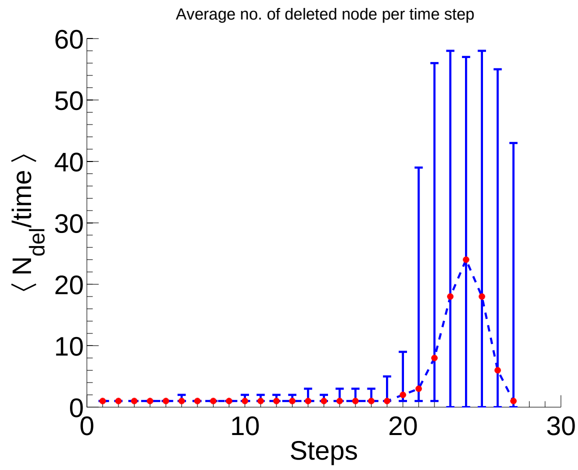

Figure 4.4: Average number of deleted nodes (red dots) on a network at each step. The step is a time at which at least one extreme event has occurred on the network. The average is taken over 200 realizations. Error bars are the minimum and maximum number of deleted nodes at a step in these 200 realizations.

in a single step and the major failure happens roughly after 20 steps. The error bar show that in one step even more than 50 nodes may fail and the remaining nodes fail at the very next step. It is clear that it takes a small number (< 30) of steps for an all-to-all connected network to fail completely. It is evident that with each node failure, the probability of another node failure increases and when, it crosses a certain value, the network as a whole fails. The time of failure remains uncertain due to the cumulative stochasticity present in the system. With the fact that an all-to-all connected network is almost certain to fail, it becomes very important to understand the underlying mechanism behind the network failure.

4.4 Nature of failures

We note that the total number of walkers on the network remains constant in the case of all-to-all network, but the total capacity of the network keeps decreasing with every failure. At the time of complete failure, capacity of the network becomes very small compared to the number of walkers present on the node. Hence, we can define a parameter,

$$\eta = \frac{Q}{W}. \tag{4.8}$$

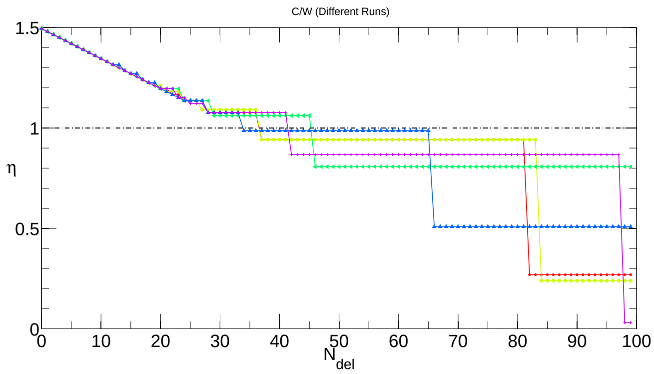

Figure 4.5: The ratio $\eta = C/W$ is plotted against the number of deleted nodes in the network. Different colors represent different realizations. On the black dash-dot line the capacity of the network equals the total load(walkers) on the network.

for the network and it can be an important parameter in studying network failures. The capacity to load ratio η is plotted in Fig. 4.5 against the deleted number of nodes. Initially, η decreases linearly, which means that the capacity of the network is decreasing linearly. Then, η shows small deviations from the linear behavior and towards the end, these deviations become larger and larger. The similar behavior is noticed in all the realizations of network failures, shown in different colors in Fig. 4.5. For an all-to-all network, nodes do not get isolated and therefor, W always remains constant. So η depends only on the capacity Q of the network which is a linear function of number of active nodes in the network. Hence, the deviations from the linear behavior represent the total number of nodes failed in one time step.

For an all-to-all network, the model is analytically tractable. As we know, the probability $F(K)$ that an extreme event occurs on a node depends on the degree (K_i), cutoff (q) and the total number of walkers on the network (W) (refer to Eq. (4.4)). With every node failure, degree K_i of the nodes changes while other parameters remain fixed for a fully connected network. The diameter of the network is 1, hence it can be assumed that random walkers spread all over the network immediately after the node fails. This assumption allows us to calculate the extreme event probability on a node using Eq. (4.4). The steady state probability of finding a walker on node i after the failures of N_{del} nodes can be written as,

$$P_i^{N_{del}} = \frac{K_i^{N_{del}}}{\sum_{j=1}^{N-N_{del}} K_j^{N_{del}}}. \tag{4.9}$$

For an all-to-all network.

$$K_i^n = N - 1 - N_{del}; \tag{4.10}$$

So,

$$\begin{aligned} P_i^{N_{del}} &= \frac{N - N_{del} - 1}{\sum_{j=1}^{N-N_{del}} (N - N_{del} - 1)} &(4.11)\\ &= \frac{N - N_{del} - 1}{(N - N_{del})(N - N_{del} - 1)} &(4.12)\\ &= \frac{1}{(N - N_{del})}. &(4.13) \end{aligned}$$

Using $P_i^{N_{del}}$, the probability of occurrence of an extreme events on a node can be obtained as,

$$F(N_{del}) = I_{P(N_{del})}\left(\lfloor q \rfloor + 1, W - \lfloor q \rfloor\right). \tag{4.14}$$

Where $I_x(.,.)$ is the incomplete beta function. Since, all the nodes have the same degree, one can calculate the expected number of extreme events on the network per time. Expectation value of extreme events on the network is equal to the expected number of deleted nodes per time. The formal equation can be written in the following way,

$$\langle N_{del} \rangle = N_{del} * F(N_{del}), \tag{4.15}$$

and it is plotted in Fig. 4.6 as the blue dashed line against the η. Initially, $\langle N_{del} \rangle = 0$ for $\eta = 1.49$. As the N_{del} increases, the network capacity decreases. In the beginning, $\langle N_{del} \rangle$ increase slowly until it becomes greater than 1. The green line in Fig. 4.6 indicates $\langle N_{del} \rangle = 1$ and it cuts the curve at $\eta = \eta_{c1} = 1.23$. After that, $\langle N_{del} \rangle$ increase rapidly and attains the maximum value at $\eta = .84$. The $\langle N_{del} \rangle$ starts decreasing until $\langle N_{del} \rangle = 0$ with $\eta = 0$. Soon after, reaching its peak value, $\langle N_{del} \rangle$ becomes comparable to the network size, N_{act} at $\eta = \eta_{c2} = 0.79$. The

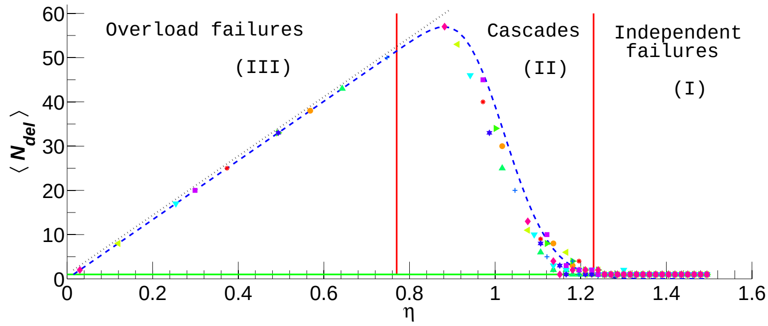

Figure 4.6: The number of deleted nodes are plotted against the capacity-load ratio. Blue dashed line is obtained from Eq. (4.15). Symbols represent the data obtained from simulations. Different symbols are used for different realizations. Along the black dot-dash line, $\langle N_{del} \rangle = N_{act}$. Red solid lines are used for separating the different natures of failure. Along the green solid line $\langle N_{del} \rangle = 1$.

behavior of $\langle N_{del} \rangle$ changes drastically at two points η_{c1} and η_{c2}, represented by solid red lines in Fig. 4.6. Based on the behavior of $\langle N_{del} \rangle$ in the regions demarcated by the critical points η_{c1} and η_{c2}, we can divide the dynamics of network failure in three different regions.

- Independent failures: $\langle N_{del} \rangle < 1$ and $\eta > \eta_{c1}$.
- Cascade failures: $N_{act} > \langle N_{del} \rangle \geq 1$ and $\eta_{c1} > \eta > \eta_{c2}$.
- overload failures: $\langle N_{del} \rangle \approx N_{act}$ and $\eta < \eta_{c2}$

The detailed descriptions of these three regions are as follows.

4.4.1 Independent failures

The region of independent failures is marked as region (I) in Fig. 4.6 for the range $\eta > 1.23$. In this region $\langle N_{del} \rangle < 1$ which means that it may take more than 1 time step for the extreme event to occur on any of node on the network. It happens because the network in the region (I) has excess capacity than the total load and hence, the individual nodes also have the capacities more than the average number of walkers. Hence, the redistributed random walkers get absorbed in the network without causing any other extreme event on the network at the next time step. As the number of deleted node, increases the probability of the node failure on the network also increases.The probability that a node gets deleted from the network in the region ranges from $\mathcal{O}(10^{-3})$ to $\mathcal{O}(10^{-1})$. Therefore, the time intervals in between the successive node failures also range from $\mathcal{O}(10^{3})$ to $\mathcal{O}(10^{1})$.

As we know, that the diameter of an all-to-all network is 1 and hence, with in a single time step, the modified network should attain a steady state. Therefore, the node failures occurring in the region (I) can be considered as independent from each other. The small dependencies in between the node failures arises due to the fact that each failed node decreases the total capacity of the network.

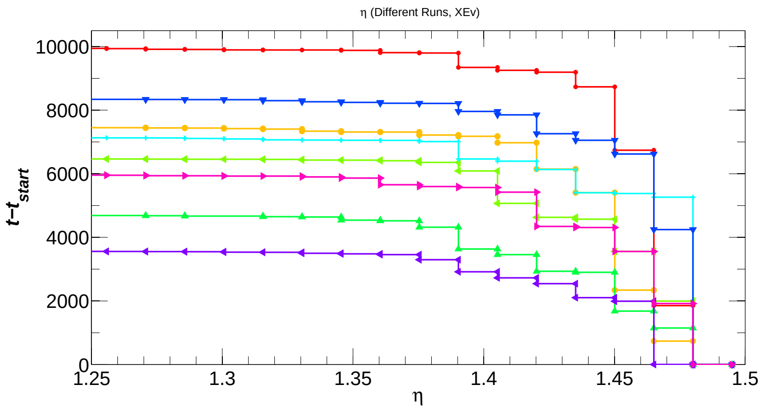

Figure 4.7: The time at which a node fails is plotted as a function of the capacity-load ratio η in the region of independent node failures($1.25 < \eta$). The time starts as soon as the first node fails. Again, different colors are used for different realizations.

Region (I) corresponds to the linear region ($\eta > 1.25$) in Fig. 4.5. This region is plotted against time in Fig. 4.7. The starting time t_{start} is set to 0 as the first extreme event knocks out a node from the network. Solid dots represent the node failures. Initially, the time differences in between the successive node failures are large $\mathcal{O}(10^{3})$ in the range $\eta > 1.42$ and they become $\mathcal{O}(10^{1})$ for $1.35 > \eta > 1.25$. In the Fig. 4.7, it seems that there are multiple failures at a single time step but it is just an artifact of the time scale.

For an easy comparison, the data of node failures obtained from simulations is also shown in Fig. 4.6 and is represented by the symbols. Different types of symbols correspond to different realizations of the network failure. In simulations, the number of failed nodes can take integer values only hence in region (I) of Fig. 4.6, differences between the analytically obtained $\langle N_{del}\rangle$, (blue dashed line) and

N_{del} obtained from simulations (symbols) can be observed. In this region, most of the symbols lie on top of each other, which represents that only one node may fail at a given time. However, there remains a very small but finite probability of multiple independent failures on the network and it is confirmed by the presence of one lower triangle, deviating from line of $\langle N_{del} \rangle = 1$. At this step, there were two nodes which failed together.

Though the numerical simulations give an approximate value, one can get an exact analytical estimation. The value of $\eta_{c1} = 1.25$ obtained from simulations is approximate and it varies, to some extent, from realization to realization. As the $\langle N_{del} \rangle$ approaches the green line, node failures become more frequent and another mechanism starts playing role in node failures.

4.4.2 Cascade failures

As η decreases, $\langle N_{del} \rangle$ increases and at $\eta = \eta_{c1} = 1.23$, the value of $\langle N_{del} \rangle$ exceeds one. It means that at every time step, at least one node gets deleted from the network. Here, the node failures are not independent and they tend to follow one another. Though, in some part of the region (II) the total capacity of the network $\eta > 1$ is still more than the total load present on the network but some of the nodes can no longer handle the extra walkers which arises due to the fluctuations in the number of walkers on a node and they get deleted due to excess load received from the redistribution of walkers. In an all-to-all network, all the nodes are the neighbors of each other and hence, these failures are correlated to one another. The failure of a small number of nodes is followed by the failure of a larger number of nodes at the next time step. Failure of nodes in such a manner is called the cascade failures. The cascade failures are shown as region (II) in Fig. 4.6. In this region also, simulation results deviate from the blue curve obtained analytically but still, they are in good harmony with each other. The deviation from the analytical results seems to be an effect of the region (I) in which the number of deleted nodes in simulations can take integer values only.

Cascades can be clearly seen in time domain in Fig. 4.8 where the region $\eta < 1.25$ of Fig. 4.5 is plotted against the time. Here, the time at which the entire

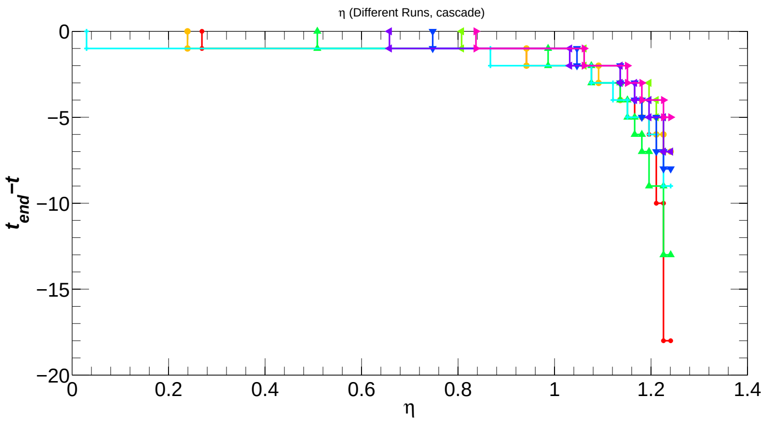

Figure 4.8: The time at which a node fails is plotted as a function of the capacity-load ratio in the region of cascade failures and over load failures ($0 < \eta < 1.25$). Here, the time of the total failure is considered as 0 and $'-'$ sign represents the time steps just before the network failure. Again, different colors are used for different realizations.

network fails is set to 0 and the history of node failures is shown. Again, the solid dots represent the node failures and it is evident from Fig. 4.8 that in the range $1.2 > \eta > 0$ nodes get deleted at each time step until the entire network fails. In the earlier studies, cascade failures have not been reported on a fully connected network.

The region of cascading failures, region (II) extends up to $\eta > \eta_{c2} = 0.79$ which means that the capacity of the network is less than the total load present on the network. In this case, the entire network should fail in single time step but stochasticity present in the dynamics prevents it from one step failure in the region of $C < W$. It also becomes clear from the Fig. 4.6, where $\langle N_{del} \rangle < N_{act}$ though, the network is overloaded.

4.4.3 Overload failures

The last stage of cascade failure is the overload failure because the network fails with deletion of the remaining nodes in one time step. In the region of overload failures marked as region (III) in Fig. 4.6 ,$\langle N_{del} \rangle$ becomes equal to N_{act} and the total load on the network is much greater than the capacity ($\eta \leq 0.79$) of the network. One can safely say that in this region, each node in the network is already overloaded and hence, it does not take more than one time step for a complete failure of the network. The range of region (III), also matches with the simulation results plotted in Fig. 4.5. In this region of Fig. 4.6, every symbol appears only once, it implies that the all the remaining node in the network fail together.

Therefore, the network collapse is triggered by the failure of a single load followed by the independent failures of few other nodes. These independent failures brings the network at the state of cascade failures which leads to the successive failures of the nodes and in the end, the network collapses in a single time step due to overload failure.

4.5 Discussion and summary

Here, in this chapter, a scenario is presented in which extreme events can cause node failures which lead to the complete failure of a network. We argue that our model is conceptually different from the models earlier studied. Based on the studies performed on a fully connected network, the nature of failures have been discussed. The network failure starts with independent node failures which gives rise to cascading failures of the nodes and in the end, the remaining nodes fail in one time step due to over load. The process of network failure is also studied analytically and the analytically obtained regions matches well with the simulation results. In contrast with the earlier studies we show that cascade failures can also occur in all-to-all networks. The model captures all the features of the network failures in an all-to-all network.

Chapter 5

Conclusion

A wide spectrum of extreme events ranging from the traffic jams to floods take place on networks. Motivated by these, this thesis was aimed for understanding the disruptions in the flow occurring on networks due to the internal fluctuations arising in the flow. This thesis is an attempt to perceive the effect of heterogeneity, inbuilt in complex networks, on the occurrences of extreme events and may be regarded as the first document to provide a frame work for studying extreme events on networks arising due to the internal fluctuations in the dynamical process. The extreme events studied here can occur in any dynamical process but here we have restricted ourselves to the transport process.

To begin with, we employ the random walk model for transport on networks and extreme events were defined on network using a carefully chosen criterion. We obtained various analytical and numerical results for the extreme events on networks. They revealed an unforeseen yet a robust feature: small degree nodes of a network are more likely to encounter extreme events than the hubs. Further, we also studied the recurrence time distribution and scaling of the probabilities for extreme events. Recurrence time distribution suggests that extreme events occurring on individual nodes are independent but it does not rule out the possibility of correlations among extreme events occurring on different nodes in the network. This cross-correlations among extreme events on different nodes can shed light on the dynamics of extreme events. It can be a useful tool to predict the extreme events on network provided that it has already occurred on some node. In the case of random walk, network topology does not play any role in the occurrence of extreme events. However the magnitude of the extreme events can depend on network structure but without receiving the support from the dynamics, effect of topology remains invisible.

In order to validate our expectations, the dynamics was modeled using generalized random walk. Here, the walk is biased by the network topology and tuning of a parameter allows the walk to be biased towards the hubs or small degree nodes. In this setting, we show that extremely large fluctuations in event sizes are possible on small degree nodes when the walkers are biased toward the hubs. Further, the probability for the occurrence of extreme events on any node in the network depends on its generalized strength, a measure of the ability of a node to attract walkers and nodes with smaller strengths display larger probability for the occurrence of extreme events compared to the nodes of higher strengths. The role of the network topology in the occurrence of extreme events could also be captured.

After studying the occurrences of extreme events on networks, we studied the effect of extreme events on networks under the assumption that extreme events destroy the node on which they occur. We realized that our model of network failure based on extreme events is conceptually different from the earlier models of network failures. To understand the dynamics arising due to this model, it was tested on a full connected network. At least in this case, we expected our results to follow the earlier known results of network failure. But to our surprise, results revealed more than what was expected based on the results reported by others. The network not only collapses against extreme events but also the collapse occurs just after the failure of only $10\% - 15\%$ nodes. Moreover, cascades of node failures were observed in all-to-all networks other than the usual over-load failures. On the all-to-all network, the model was analytically tractable and conditions for different kinds of failures were obtained.

All the results documented in this thesis suggest that the nodes with smaller degree are very important from extreme events point of view. On a scale-free network they are large in numbers and carry most of the load. At the same time, they are the most vulnerable to face the extreme events. These results can be used to design the network which can handle the extreme events smoothly.

The work reported in this thesis provides a general framework to study the extreme events in the case when dynamics is taking place on the networks. Though, here we have studied extreme events on networks using the random walk dynamics but it is possible to study extreme events for other dynamical processes on networks. Hence, this thesis opens a new window to look at the various phenomena occurring on networks and understand them in terms of extreme events.

I would like to conclude this thesis with the following quote,

Now this is not the end.
It is not even the beginning of the end.
But it is, perhaps, the end of the beginning.
– Sir Winston Churchill (1942)

Bibliography

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys., **74**, 47 (2002).

[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U.ng. Hwa, Phys. Rep., **424**, 175 (2006).

[3] M.E.J. Newman, SIAM Rev., **45**, 167 (2003).

[4] Marc Barthélemy, Phys. Rep., **499**, 1 (2010).

[5] S. N. Dorogovtsev, *Lectures on Complex Networks* (Oxford University Press, Oxford, 2010).

[6] S. Milgram, Psychol. Today, **2**, 60 (1967).

[7] Watts, D. J., and S. H. Strogatz, Nature, **393**, 440 (1998).

[8] Newman, M. E. J.,Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).

[9] Newman, M. E. J., Phys. Rev. E **64**, 016131 (2001).

[10] Newman, M. E. J., Phys. Rev. E **64**, 016132 (2001).

[11] Redner, S., Eur. Phys. J. B **4**, 131 (1998).

[12] Aiello, W., F. Chung, and L. Lu, Proceedings of the 32nd ACM Symposium on the Theory of Computing (ACM, New York), 171 (2000).

[13] Montoya, J. M., and R. V. Solé, preprint cond-mat/0011195. (2000).

[14] Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).

[15] Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).

[16] Ferrer i Cancho, R., and R. V. Solé, Santa Fe Institute working paper 01-03-016 (2001).

[17] Jeong, H., S. P. Mason, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **411**, 41 (2001).

[18] Broder, A., R. Kumar, F. Maghoul, P. Raghavan, S. Rajalopagan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).

[19] Govindan, R., and H. Tangmunarunkit, in Proceedings of IEEE INFOCOM 2000, Tel Aviv, Israel (IEEE, Piscataway, N.J.), **3**, 1371 (2000).

[20] V. Latora, M. Marchiori, Phys. Rev. E, **71**, 015103 (2005)(R).

[21] Erdős, P., and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[22] A. -L. Barabási, and R. Albert, Science **286**, 509 (1999).

[23] A.F. Rozenfeld, R. Cohen, D. ben Avraham, S. Havlin, Phys. Rev. Lett., **89** , 218701 (2002).

[24] J.S. Andrade Jr., H.J. Herrmann, R.F.S. Andrade, L.R. da Silva, Phys. Rev. Lett., **94**, 018702 (2005).

[25] S.S. Manna, P. Sen, Phys. Rev. E, **66** , 066114 (2002).

[26] K. Huang, *Statistical Mechanics*, (Wiley, 1987).

[27] S. K. Ma, *Modern Theory of Critical Phenomena*, (The Perseus Books Group, 2000).

[28] S. N. Dorogovtsev, A. V. Goltsev, and J. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008)

[29] S.N. Dorogovtsev, A.V. Goltsev, and J.F. Mendes, Phys. Rev. E **66**, 016104 (2002).

[30] Takehisa Hasegawa and Koji Nemoto , Phys. Rev. E **80**, 026126 (2009).

[31] Carlos P. Herrero, Phys. Rev. E **77**, 041102 (2008)

[32] S. Boccaletti, J. Kurths, G. Osipov, D. L. Valladares, C. S. Zhou, Physics Reports, **366**, 1 (2002).

[33] S.H. Strogatz, Nature (London) **410**, 268 (2001).

[34] S. Jalan and R. E. Amritkar, Phys. Rev. Lett. **90**, 014101 (2003).

[35] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).

[36] Dimitris Achlioptas, Raissa M. D'Souza and Joel Spencer, Science **323** 1453 (2009).

[37] Amir Bashan, Roni Parshani, and Shlomo Havlin, Phys. Rev. E **83**, 051127 (2011)

[38] B. Kozma, M.B. Hastings, G. Korniss, Phys. Rev. Lett., **95** 018701 (2005).

[39] S. Jespersen, I.M. Sokolov, A. Blumen, Phys. Rev. E, **62** 4405 (2000).

[40] C. Song, T. Koren, P. Wang, A.-L. Barabasi, Nat. Phys. **6**, 818 (2010).

[41] M. Szell, R. Sinatra, G. Petri, S. Thurner, Vito Latora, Scientific Reports **2**, 457, (2012).

[42] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[43] P. Wang, M. Gonzalez, C.A. Hidalgo, A.-L. Barabási, Science,**324**, 1071 (2009).

[44] S. Meloni, A. Arenas, Y. Moreno, Proc. Natl. Acad. Sci. U.S.A. **106**, 16897 (2009)

[45] M.J. Keeling, M.E.J. Woolhouse, R.M. May, G. Davies, B.T. Grenfell, Nature, **421**, 136 (2003).

[46] V. Sood and S. Redner, Phys. Rev. Lett. **94**, 178701 (2005).

[47] Naoki Masuda, N. Gibert, and S. Redner, Phys. Rev. E **82**, 010103 (2010).

[48] Claudio Castellano, Santo Fortunato, and Vittorio Loreto , Rev. Mod. Phys. **81**, 591 (2009).

[49] T. Ohira and R. Sawatari, Phys. Rev. E **58**, 193 (1998).

[50] R. Pastor-Satorras, A. Vzquez, and A. Vespignani, Phys. Rev. Lett., **87** 258701 (2001).

[51] M. E. Crovella and A. Bestavros, IEEE/ACM Transactions on Networking, **5**, 835 (1997).

[52] B. Tadić, Europ. Phys. J. B **23**, 221 (2001).

[53] R. Guimera, A. Diaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).

[54] Lazaros K. Gallos, Phys. Rev. E **70**, 046116 (2004).

[55] A. Kittas, S. Carmi, S. Havlin, and P. Argyrakis, Europhys. Lett. **84**, 40008 (2008).

[56] Zhongzhi Zhang, Yihang Yang, and Yuan Lin, Phys. Rev. E **85**, 011106 (2012).

[57] `http://en.wikipedia.org/wiki/List_of_power_outages`.

[58] S. Chen, W. Huang, C. Cattani, and G. Altieri, Math. Probl. Eng. **2012**, 732698 (2012).

[59] C.D. Brummitt, R.M. DSouza, and E.A. Leicht, Proc. Natl. Acad. Sci. U.S.A. **109**, E680 (2012).

[60] A. Gumbel, *Statistics of Extremes* (Dover Publications, New York, 2004)

[61] S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, London, 2001).

[62] D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics* (Cambridge University Press, Cambridge, 1997).

[63] D. Sornette, *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-organization and Disorder: Concept and tool*, (Springer, Berlin, 2004)

[64] M. S. Santhanam and H. Kantz, Phys. Rev. E **78**, 051113 (2008).

[65] S. Albeverio, V. Jentsch and Holger Kantz (Ed.) , *Extreme events in nature and society*, (Springer, 2005).

[66] C. Nicolaides, L. Cueto-Felgueroso, and R. Juanes, Phys. Rev. E **82**, 055101(R) (2010).

[67] V. Tejedor, O. Bnichou, and R. Voituriez, Phys. Rev. E **80**, 065104(R) (2009).

[68] J. D. Noh and H. Reiger, Phys. Rev. Lett. **92**, 118701 (2004).

[69] F. Reif, *Statistical Physics*, (McGraw-Hill, 1965).

[70] D. Brockmann, L. Hufnagel, and T. Geisel, Nature **439**, 462 (2006).

[71] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, Phys. Rev. E **75**, 011128 (2007).

[72] L. Lovasz, Random walks on graphs: A survey, Combinatorics **99**, 1-46 (1993).

[73] M. Argollo de Menezes and A.L. Barabasi, Phys. Rev. Lett. **92**, 028701 (2004).

[74] S. Meloni, J. Gomez-Gardenes, Vito Latora and Y. Moreno, Phys. Rev. Lett. **100**, 208701 (2008).

[75] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).

[76] D. Methi and K. Ramasamy, *Network Routing: Algorithms, Protocols and Architectures*, (Morgan Kaufmann, 2007).

[77] Cormen T. H., Leiserson C. E., Rivest R. L. and S. Clifford, *Introduction to Algorithms*, (MIT Press, 2001).

[78] M. Barthelemy, Eur. Phys. J. B **38**, 163 (2004).

[79] H. Wang, J.M. Hernandez, and P.V. Mieghem, Phys. Rev. E **77**, 046105 (2008).

[80] S. Dolev, Y. Elovici, and R. Puzis, J. of ACM **57**, 25 (2010).

[81] `http://blog.twitter.com/2010/02/measuring-tweets.html`

[82] `http://www.comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009`

[83] Vimal Kishore, M. S. Santhanam and R. E. Amritkar, Phys. Rev. Lett. **106**, 188701 (2011).

[84] Agata Fronczak and Piotr Fronczak, Phys. Rev. E **80**, 016107 (2009);

[85] Wen-Xu Wang, B.H. Wang, C. Y. Yin, Y.B. Xie and T. Zhou, Phys. Rev. E **73**, 026111 (2006).

[86] S.-J. Yang, Phys. Rev. E **71**, 016107 (2005).

[87] J. L. Lopez and J. Sesma, *Integral Transforms and Special Functions* **8**, 233 (1999).

[88] V. Sood and P. Grassberger, Phys. Rev. Lett. **99**, 098701 (2007).

[89] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, Phys. Rev. Lett. **102**, 160602 (2009)

[90] R. Sinatra, J. Gmez-Gardees, R. Lambiotte, V. Nicosia, and V. Latora, Phys. Rev. E **83**, 030103(R) (2011).

[91] A.N. Mian, R. Baldoni, R. Beraldi, *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops, Montreal, QC* (IEEE, Piscataway, NJ,2009). pp. 153-157.

[92] R. Cohen, K. Erez, ben-Avraham D. and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).

[93] R. Cohen, K. Erez, ben-Avraham D. and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).

[94] Albert, R., Jeong, H. and Barabási, A.-L., Nature **406**, 378 (2000).

[95] Motter, A.E., lai, Y., Phys. Rev. E **66** 065102(R) (2002).

[96] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, Chaos **12**, 985 (2002).

[97] Cohen, R., Avraham, D. Ben , Havlin S. , Phys. Rev. E **66**, 36113 (2001).

[98] Vázquez, A. , Moreno, Y., Phys. Rev. E **67**, 015101(R) (2003).

[99] Moreno, Y. , Gómez, J. B., Pacheco, A.F., Europhys. Lett. **58**, 630 (2002).

[100] Moreno,Y., Pastor-Satorras, R. , Vázquez, A., Vespignani, A., Europhys. Lett. **62**, 292 (2003).

[101] Crucitti, P., Latora, V., Marchiori, M., Phys. Rev. E **69**, 045104(R) (2004).

[102] Huang, L., Yang, L., Yang,K. , Phys. Rev. E **73**, 036102 (2006).

[103] Bakke, J., Hansen, A., Kertész, Europhys. Lett. **76**, 717 (2006).

[104] Dall'Asta, L., Barrat, L. , Barthélemy, M. , Vespignani, A. J. Stat. mech. P04006 (2006).

[105] Lehman, J., Bernasconi J., Phys. Rev. E, **81** 031129 (2010).

[106] Simonsen, I., Buzna, Lubos, Karsten, P., Bornholdt, S., Helbing D., Phys. Rev. Lett. **100**, 218701 (2008).

[107] Watts, D.J., Proc. Natl. Acad. Sci. USA, **99**, 5766 (2002)

[108] Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., Havlin, S., Nature **464** 1025 (2010).

Extreme Events on Complex Networks

Vimal Kishore,[1] M. S. Santhanam,[2] and R. E. Amritkar[1]
[1]*Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India*
[2]*Indian Institute of Science Education and Research, Pashan Road, Pune 411 021, India*
(Received 12 February 2011; published 3 May 2011)

A wide spectrum of extreme events ranging from traffic jams to floods take place on networks. Motivated by these, we employ a random walk model for transport and obtain analytical and numerical results for the extreme events on networks. They reveal an unforeseen, and yet a robust, feature: small degree nodes of a network are more likely to encounter extreme events than the hubs. Further, we also study the recurrence time distribution and scaling of the probabilities for extreme events. These results suggest a revision of design principles and can be used as an input for designing the nodes of a network so as to smoothly handle extreme events.

DOI: 10.1103/PhysRevLett.106.188701 PACS numbers: 89.75.Hc, 05.40.−a, 05.40.Fb

Extreme events (EE) taking place on networks is a fairly commonplace experience. Traffic jams in roads and other transportation networks, web servers not responding due to heavy load of web requests, floods in the network of rivers, and power blackouts due to tripping of power grids are some of the common examples of extreme events on networks. Such events can be thought of as emergent phenomena due to transport on the networks. As EE lead to losses ranging from financial and productivity to even life and property [1], it is important to estimate probabilities for the occurrence of extreme events and, if possible, incorporate them to design networks that can handle such extreme events.

Transport phenomena on the networks have been studied vigorously in the last several years [2,3] though they were not focused on the analysis of EE. However, one kind of extreme event in the form of congestion has been widely investigated [4]. For instance, a typical approach is to define rules for (a) generation and transport of traffic on the network and (b) capacity of the nodes to service them. Thus, a node will experience congestion when its capacity to service the incoming "packets" has been exceeded [5]. In this framework, several results on the stability of networks, cascading failures to congestion transition have been obtained. An extreme event, on the other hand, is defined as exceedances above a prescribed quantile and is not necessarily related to the handling capacity of the node in question. It arises from natural fluctuations in the traffic passing through a node and not due to constraints imposed by capacity. Thus, in the rest of this Letter, we discuss transport on the networks and analyze the probabilities for the occurrence of EE arising in them without having to model the dynamical processes or prescribe the capacity at each of the nodes.

The transport model we adopt is the random walk on complex networks [3]. Random walk is of fundamental importance in statistical physics though in real network settings many variants of random walk could be at work [6]. For instance, in the case of road traffic, the flow typically follows a fixed, often shortest, path from node A to B and can be loosely termed deterministic. Thus, given the operational principle of network dynamics, i.e., deterministic or probabilistic or a combination of both, we obtain the probabilities for the occurrence of EE on the nodes. This reveals a significant and unexpected result: namely, that the EE are more prone to occur in a small degree node than in a hub. This feature is robust against fluctuating traffic and even upon the application of intelligent routing algorithms (e.g., shortest paths). This principal result implies that the design principles for networks should focus on small degree nodes which are prone to EE. Further, these probability estimates allow us to design nodes that can have sufficient capacity to smoothly handle EE of a certain magnitude. Currently, for univariate time series, there is a widespread interest on the extreme value statistics and their properties, in particular, in systems that display long memory [7]. Thus, we place our results in the context of both the random walks and EE in a network setting.

We consider a connected, undirected, finite network with N nodes with E edges. The links are described by an adjacency matrix $\mathbf{A}$ whose elements A_{ij} are either 1 or 0 depending on whether i and j are connected by a link or not, respectively. On this network, we have W noninteracting walkers performing the standard random walk. A random walker at time t sitting on the ith node with K_i links can choose to hop to any of the neighboring nodes with equal probability. Thus, transition probability for going from the ith to the jth node is A_{ij}/K_i. We can write down a master equation for the n-step transition probability of a walker starting from node i at time $n = 0$ to node j at time n as,

$$P_{ij}(n+1) = \sum_k \frac{A_{kj}}{K_k} P_{ik}(n). \tag{1}$$

It can be shown that the n-step time-evolution operator corresponding to this transition, acting on an initial distribution, leads to stationary distribution with eigenvalue unity [3] and it turns out to be

$$\lim_{n\to\infty} P_{ij}(n) = p_j = \frac{K_j}{2E}. \tag{2}$$

The existence of stationary distribution is crucial for defining EE. Physically, the stationary probability in Eq. (2) implies that more walkers will visit a given node if it has more links.

Now we can obtain the distribution of random walkers on a given node. We ask for the probability $f(w)$ that there are w walkers on a given node having degree K. Since the random walkers are independent and noninteracting, the probability of encountering w walkers at a given node is p^w while the rest of the $W - w$ walkers are distributed on all the other nodes. This turns out to be binomial distribution given by

$$f(w) = \binom{W}{w} p^w (1-p)^{W-w}. \tag{3}$$

Now, the mean and variance for a given node can be explicitly written down as

$$\langle f\rangle = \frac{WK}{2E}, \qquad \sigma^2 = W\frac{K}{2E}\left(1-\frac{K}{2E}\right). \tag{4}$$

As expected, the mean and variance depends on the degree of the node for fixed W and E. Note that $K/2E \ll 1$ and hence $\sigma \approx \langle f\rangle^{1/2}$. This reproduces the relation proposed in Ref. [8], later shown to have limited validity [9].

One natural extension of the result in Eq. (3) is to account for fluctuations in the number of walkers. We assume that the total number of walkers is a random variable uniformly distributed in the interval $[W-\Delta, W+\Delta]$. Then the probability of finding w walkers becomes

$$f^\Delta(w) = \sum_{j=0}^{2\Delta} \frac{1}{2\Delta+1}\binom{\tilde{W}+j}{w} p^w (1-p)^{\tilde{W}+j-w}, \tag{5}$$

where $\tilde{W} = W - \Delta$. The mean and variance of this distribution can be obtained as

$$\begin{aligned} \langle f^\Delta\rangle &= \langle f\rangle, \\ \sigma_\Delta^2 &= \langle f^\Delta\rangle\left[1 + \langle f^\Delta\rangle\left\{\frac{\Delta^2}{3W^2} + \frac{\Delta}{3W^2} - \frac{1}{W}\right\}\right]. \end{aligned} \tag{6}$$

In the spirit of extreme value statistics, an extreme event is one whose probability of occurrence is small, typically associated with the tail of the probability distribution function. In the network setting, we will apply the same principle to each of the nodes. Based on Eqs. (3) and (4), we will designate an event to be extreme if more than q walkers traverse a given node at any time instant. The probability for the occurrence of an extreme event can be obtained as

$$F(K) = \sum_{j=0}^{2\Delta} \frac{1}{2\Delta+1} \sum_{k=\lfloor q\rfloor+1}^{\tilde{W}+j} \binom{\tilde{W}+j}{k} p^k (1-p)^{\tilde{W}+j-k}, \tag{8}$$

where $\lfloor u\rfloor$ is the floor function defined as the largest integer not greater than u. Notice that necessarily the cutoff q will have to depend on the node (or rather, the traffic flowing through the node) in question. Applying uniform threshold independent of the node ($q = \text{const}$) will lead to some nodes always experiencing an extreme event while some others never encountering any extreme event at all. Hence we define the threshold for extreme event to be $q = \langle f\rangle + m\sigma$, where m is any real number.

It does not seem possible to write the summation in Eq. (8) in closed form. However, for the special case when $\Delta = 0$, Eq. (8) simplifies to

$$F(K) = \sum_{k=\lfloor q\rfloor+1}^{W} f(k) = I_p(\lfloor q\rfloor + 1, w - \lfloor q\rfloor), \tag{9}$$

where $I_p(.,.)$ is the regularized incomplete beta function [10]. For a given choice of network parameter E and number of walkers W, the extreme event probability at any node depends only on its degree. In Fig. 1 we show $F(K)$ as a function of degree K superimposed on the results obtained from random walk simulations. The agreement between Eq. (8) and the simulated results is quite good. Further, each point in the figure represents an average over all the nodes with the same degree. We emphasize that the oscillations seen in Fig. 1 are inherent in the analytical and numerical results and not due to insufficient ensemble averaging.

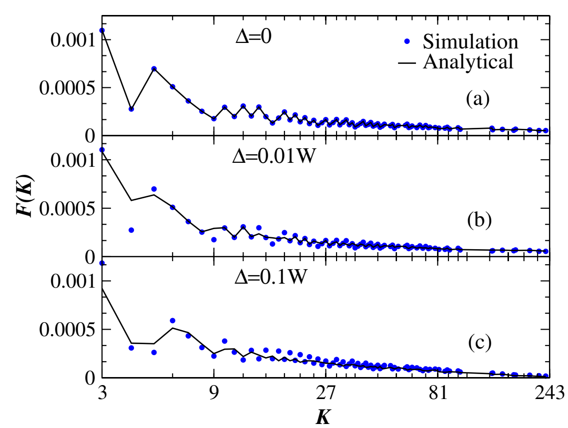

FIG. 1 (color online). Probability for the occurrence of extreme events as a function of degree K with fluctuations Δ in the total number of walkers on semilog plot. The threshold for EE is $q = \langle f\rangle + 4\sigma$. The solid lines are from the analytical result in Eq. (8). All the simulations shown in this Letter are obtained with a scale-free network (degree exponent $\gamma = 2.2$) with $N = 5000$ nodes, $E = 19\,815$ vertices, and $W = 2E$ walkers averaged over 100 realizations with randomly chosen initial conditions.

An important feature of this result is that the nodes with smaller degree ($K < 20$) reveal, on an average, a higher probability for the occurrence of EE as compared to the nodes with higher degree, say, $K > 100$. By careful choice of parameters, the probability $F(K)$ can differ by as much as an order of magnitude. This runs contrary to a naive expectation that higher degree nodes garner more traffic and hence are more prone to EE. While the former contention is still true in the random walk model we employ, the results here indicate that the latter one is not generally correct. As shown in Figs. 1(b) and 1(c), this feature is robust even when the number of walkers becomes a fluctuating quantity. We note that Eqs. (8) and (9) for the extreme event probability do not depend on the topology of the network. Even though the simulation results are shown for scale-free graphs, it holds good for other types of graphs (not shown here) with random and small world topologies. However, the difference in probability for EE between hubs and smaller degree nodes is not pronounced in the case of random graphs.

The threshold q that defines an event to be extreme depends on the traffic flowing through a given node. The choice $q = \langle f \rangle + m\sigma$ is arbitrary. Now, we show that the extreme event probability in Eq. (9) scales with the choice of threshold q or, equivalently, m. In the Fig. 2(a) we show $F_m(K)$ for various choices of m in log-log scale. Clearly, as m decreases, ignoring the local fluctuations, the curves tend to become horizontal. Physically, this can be understood as follows: $q \to 0$ implies that the threshold for EE decreases and this leads to larger number of EE and hence a higher probability of occurrence. In the limiting case of $q = 0$, $F(K) = 1$ for all nodes and all the events would be extreme. The graph in Fig. 2(a) suggests that it might be scaling with respect to q or m. Starting from Eq. (9), we were not able to determine the scaling analytically.

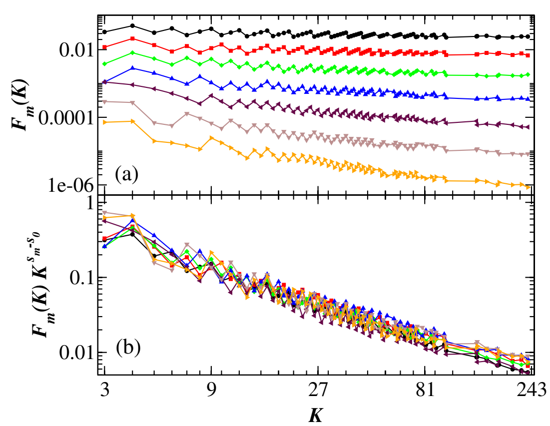

FIG. 2 (color online). Probability for occurrence of extreme events for several values of threshold $q = \langle f \rangle + m\sigma$. (a) shows the extreme event probabilities in log-log plot obtained from simulations with $\Delta = 0$. (b) shows scaling EE probabilities. S_0 represents the reference slope with $m = 2$. The threshold applied for curves from top to bottom are $m = 2.0, 2.5, 3.0, 3.5, 4.0, 4.5,$ and 5.0.

Hence, we empirically show that the following scaling relation holds for the probability of EE,

$$\frac{F_m(K)}{K^{1-S_m}} = \text{constant}, \qquad (10)$$

where $F_m(K)$ represents extreme event probability for threshold value q with parameter m. In this, S_m is the slope of the curves $F_m(K)$ in the Fig. 2(a). Using Eq. (10) on the simulated data for $\Delta = 0$, we find that all the curves for the probability of EE, shown in Fig. 2(b), collapse into one curve to a good approximation.

In the study of EE, distribution of their return intervals is an important quantity of interest. This carries the signature of the temporal correlations among the EE and is useful for hazard estimation in many areas. We focus on the return intervals for a given node of the network. Since the random walkers are noninteracting, the events on the nodes are uncorrelated. Then, the recurrence time distribution is given by $P(\tau) = e^{-\tau/\langle\tau\rangle}$, where the mean recurrence time is $\langle\tau\rangle = 1/F(K)$. In the inset of Fig. 3, we show $P(\tau)$ obtained from simulations for three nodes with different degrees. In semilog plot, they reveal an excellent agreement with the analytical distribution $P(\tau)$ (shown as a solid line). The main graph of Fig. 3 shows the mean recurrence time $\langle\tau\rangle$, the only parameter that characterizes the recurrence distribution, as a function of K and it agrees with the analytical result.

As pointed out before, many types of flow on the network, such as the information packets flowing through the network of routers and traffic on roads, use more intelligent routing algorithms [11] rather than a random walk. To check the robustness of results in Eqs. (8) and (9), we implemented the random walk simulation with the constraint that the traffic from node i to j takes the shortest path (SP) on the network. If multiple shortest paths are available to go from node i to j, the algorithm chooses any one of them with equal probability. Thus, in this setting,

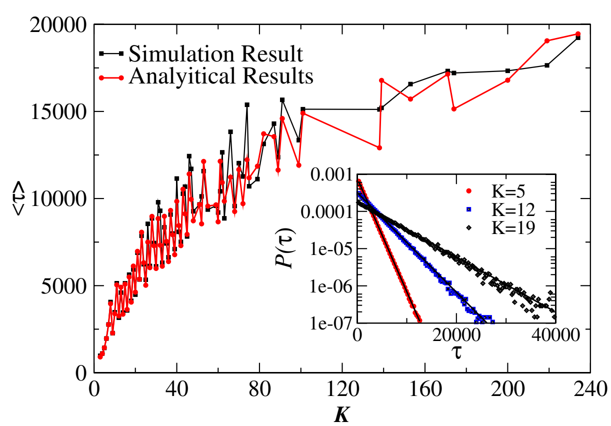

FIG. 3 (color online). The inset shows the recurrence time distribution for extreme events from simulations (symbols) with $\Delta = 0$ for nodes with 5, 12, and 19 links. The solid line is the analytical distribution. The main figure shows the mean recurrence time as a function of degree K.

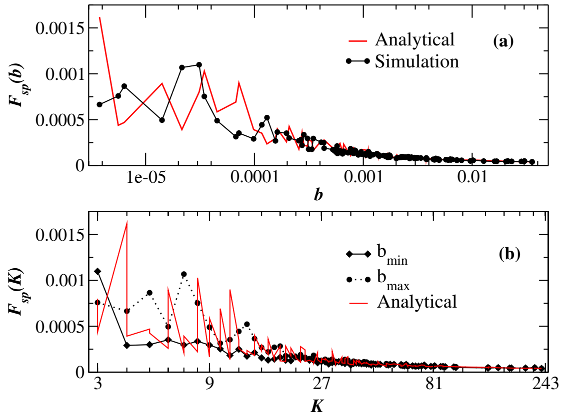

FIG. 4 (color online). Extreme event probability F_{sp} for $\Delta = 0$ with shortest path algorithm implemented for random walkers. The data are plotted in two different ways. (a) $F_{sp}(b)$ as a function of betweenness centrality, (b) $F_{sp}(K)$ as a function of degree K of the node. Nodes with same value of K can have different betweenness centrality. In (b), in order to reduce the clutter, for every value of K, the extreme event probability for the node with largest ($b_{\max}$, solid circles) and least value ($b_{\min}$, solid square) of b is plotted.

for every random choice of source-destination pair the paths are laid out by the algorithm and randomness arises only when multiplicity of SPs are available. Thus, this can be thought of as a walk with a large deterministic component. The simulation results with the SP algorithm [12] shown in Fig. 4 are qualitatively similar to the trend displayed in Fig. 1. In this scenario of predominantly deterministic dynamics, it is conceivable that the degree of a node does not determine the flux passing through it. This role is played by the centrality of the node with respect to the SPs in the network, quantified by the betweenness centrality b of a given node [13]. Based on this qualitative argument, the results in Fig. 4 can be understood if we replace Eq. (2) with $p = \beta b/B$ where B is the normalization factor that depends on the sum of betweenness centrality of all the nodes on the network. From the numerical simulations, we obtain $\beta \approx 0.94$. Using this p in Eq. (2), we can go through the same arguments as before and analytically obtain $\langle f \rangle$, σ^2, q, and the probability $F_{sp}(b)$ for occurrence of EE. In Fig. 4(a), $F_{sp}(b)$ is shown as solid curve. In Fig. 4(b), the same data for $F_{sp}(b)$ are shown as a function of K for easier comparison with Fig. 1. Thus, even with the SP algorithm thrown in, the EE probabilities are higher for the nodes with smaller degree ($K < 20$) than for the ones with larger degree ($K > 100$).

Finally, we comment on how these results can be applied as a basis to design nodes of a network. The central result in this Letter in Eq. (8) allows us to *a priori* estimate the EE probabilities. These depend on whether operating principle of dynamics is deterministic or probabilistic. If the idea is to avoid congestion or other problems arising due to EE of certain magnitude, then these estimates can be used as an input to the design principles for the nodes. For instance, for the road traffic that operates broadly on the shortest path principle the probabilities can be used as a basis to provision for higher capacity to nodes that will avoid bottlenecks arising from EE of a given magnitude.

In scale-free networks, small degree nodes form the bulk and are more prone to encounter EE. But network design principles and practice generally focus on the hubs. Such evolved practices might work best most of the time. Our work suggests that they might fail in the context of extreme events and hence a revised approach is necessary. A careful design for the capacity of small degree nodes is important as well. It must be emphasized that incorporating such EE estimates in design principles will only help in better preparedness to meet the expected EE. The EE discussed here being due to inherent fluctuations will nevertheless take place and cannot be avoided.

The simulations were carried out on computer clusters at PRL, Ahmedabad and IISER, Pune.

[1] *Extreme Events in Nature and Society*, edited by S. Albeverio, V. Jentsch, and Holger Kantz (Springer, New York, 2005).

[2] 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. Bénichou, and R. Voituriez, Phys. Rev. E **80**, 065104(R) (2009).

[3] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).

[4] D. De Martino *et al.*, Phys. Rev. E **79**, 015101(R) (2009); P. Echenique *et al.*, Europhys. Lett. **71**, 325 (2005); K. Kim *et al.*, Europhys. Lett. **86**, 58002 (2009); B. Tadic, G. J. Rogers, and S. Thurner, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 2363 (2007).D. J. Ashton, T. C. Jarrett, and N. F. Johnson, Phys. Rev. Lett. **94**, 058701 (2005); L. Zhao *et al.*, Phys. Rev. E **71**, 026125 (2005); R. Germano and A. P. S. de Moura, Phys. Rev. E **74**, 036117 (2006).

[5] Wen-Xu Wang *et al.*, Chaos **19**, 033106 (2009).

[6] F. Reif, *Statistical Physics* (McGraw-Hill, New York, 1965); C. Song *et al.*, Nature Phys. **6**, 818 (2010); D. Brockmann *et al.*, Nature (London) **439**, 462 (2006).

[7] J. F. Eichner *et al.*, Phys. Rev. E **75**, 011128 (2007); M. S. Santhanam and Holger Kantz, Phys. Rev. E **78**, 051113 (2008).

[8] M. Argollo de Menezes and A.-L. Barabási, Phys. Rev. Lett. **92**, 028701 (2004).

[9] S. Meloni *et al.*, Phys. Rev. Lett. **100**, 208701 (2008).

[10] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).

[11] D. Methi and K. Ramasamy, *Network Routing: Algorithms, Protocols and Architectures* (Morgan Kaufmann, San Francisco, 2007).

[12] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and S. Clifford, *Introduction to Algorithms* (MIT Press, Cambridge, 2001).

[13] M. Barthelemy, Eur. Phys. J. B **38**, 163 (2004); H. Wang, J. M. Hernandez, and P. V. Mieghem, Phys. Rev. E **77**, 046105 (2008); S. Dolev *et al.*, J. ACM **57**, 1 (2010).

Extreme events and event size fluctuations in biased random walks on networks

Vimal Kishore*
Physical Research Laboratory, Ahmedabad-380009, India

M. S. Santhanam†
Indian Institute of Science Education and Research, Pune-411021, India

R. E. Amritkar‡
Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India
(Received 12 December 2011; published 29 May 2012)

Random walk on discrete lattice models is important to understand various types of transport processes. The extreme events, defined as exceedences of the flux of walkers above a prescribed threshold, have been studied recently in the context of complex networks. This was motivated by the occurrence of rare events such as traffic jams, floods, and power blackouts which take place on networks. In this work, we study extreme events in a generalized random walk model in which the walk is preferentially biased by the network topology. The walkers preferentially choose to hop toward the hubs or small degree nodes. In this setting, we show that extremely large fluctuations in event sizes are possible on small degree nodes when the walkers are biased toward the hubs. In particular, we obtain the distribution of event sizes on the network. Further, the probability for the occurrence of extreme events on any node in the network depends on its "generalized strength," a measure of the ability of a node to attract walkers. The generalized strength is a function of the degree of the node and that of its nearest neighbors. We obtain analytical and simulation results for the probability of occurrence of extreme events on the nodes of a network using a generalized random walk model. The result reveals that the nodes with a larger value of generalized strength, on average, display lower probability for the occurrence of extreme events compared to the nodes with lower values of generalized strength.

DOI: 10.1103/PhysRevE.85.056120 PACS number(s): 89.75.Hc, 05.40.Fb

I. INTRODUCTION

Extreme events are typically associated with disasters of some kind or other, e.g., droughts, cold wave, cyclones, earthquakes, wind gusts, and economic recession. When a relevant variable, such as the wind speed $w(t)$ recorded at time t in the case of wind gusts, exceeds a certain prescribed threshold q due to its inherent fluctuations, i.e., $w(t) > q$, then it is taken to be an extreme event. In particular, it is important to note that the magnitude of the tremor, wind speed, temperature, economic growth, etc., are scalar variables. A large number of results, both theoretical and empirical, are known about the statistics and dynamics of extreme events for such univariate, scalar variables [1]. One significant result due to classical extreme value theory is that, depending on the probability distribution function of the variable, the distribution of block maxima for the uncorrelated sequence of random variables converges to only one of three possible forms, namely, Fréchet, Gumbel, and Weibull distributions [2].

In contrast to this scenario, extreme events can also take place on complex networks. Consider, for instance, the most common experience of web surfers, a web server not responding due to the heavy load of http requests. This is an extreme event taking place on the network of the World Wide Web. For example, the popular social networking site Twitter handled about 600 tweets per second in early 2010 [3]. According to an industry estimate, the Google search engine received approximately 34 000 search requests per second by the end of 2009 [4]. For most web sites on the World Wide Web that are unprepared for such a large number of http requests, these numbers would represent extreme events and could potentially disrupt the service. The power blackout in the northeastern United States in 2003 is also an example of extreme event on the power transmission grid network. The cascading failures shut down more than 508 power generating units at 265 power plants during the peak of this blackout [5]. Gridlock on highways is an example of an extreme event on the transportation network. From the point of view of physics, all these events could be thought of as an emergent phenomena arising due to flux on the networks and could be regarded as extreme events arising primarily due to the limited handling capacity of the node. Transport on networks continues to be widely studied, but much less attention has been focused on it from the point of view of extreme events. Generally, when the flux (packets of information or power or highway traffic in the case of the examples given above) exceeds the handling capacity, it turns out to be an extreme event for the particular node on the network. In earlier works related to congestion and cascade on networks [6–13], handling capacity is a key ingredient that needs to be prescribed upfront.

However, extreme events happen not only because of the limited handling capacity of the node on a network but also because of inherent fluctuations in the flux passing through the node. These fluctuations in the flux passing through a node could be so large that they breach a prescribed threshold, in which case we label the event as an extreme event for the node. This definition of extreme event for a node on any

*phy.vimal@gmail.com
†santh@iiserpune.ac.in
‡amritkar@prl.res.in

network is similar in spirit to that of the classical extreme value theory. Then, a relevant question is how the connectivity of the network affects the probability for extreme event occurrence. By modeling the transport as standard random walks on networks, it was shown in Ref. [14] that the probability for the occurrence of extreme events $P(k_i)$, arising due to inherent fluctuations, depends only on the degree k_i of the ith node in question. In this work, the threshold q_i was chosen to be proportional to typical fluctuation size on ith node. Thus, the extreme events are identified after taking care of the natural variability of the flux passing through the given node. Further, it was shown that, on average, $P(k)$ is higher for small degree nodes than for hubs. This is a surprising result because it implies that, within the framework of random walk on networks, even though hubs attract large flux (compared to small degree nodes) they are less prone to extreme events. Thus, in the context of a node on a connected network, larger flux does not necessarily translate into higher probabilities for extreme events. This feature is one possible signature of connectivity, i.e., the network setting on which the system operates. In contrast, for a scalar time series $w(t)$ larger flux would imply higher extreme-event probabilities.

Random walk on complex networks is a useful fundamental model against which to compare other transport processes. Most realistic transport phenomena on networks, such as the flux of information packets passing through the network of routers or road traffic, do not proceed by performing a random walk. In order to model the flux in a more realistic way, it is useful to generalize the standard random walk to a situation in which the flux is either biased toward hubs or small degree nodes. For example, consider the case of two remote airports which are not directly connected by flights. Typically, they would be connected through a major hub on the airline network. This is one practical scenario in which the traffic is biased toward the hubs. This happens in many network settings; railways tend to connect the hinterland with the hubs, phone connect to nearest hubs on the network. Motivated by these physical examples, in this work, we model the transport process as random walks biased by the topology of the network and study the extreme-event probabilities and event-size distributions. We show that a biased random walk leads to extreme fluctuations in the event sizes on the network. In the subsequent sections, we discuss the topologically biased random walk model on a network and obtain analytical results for the probability of occurrence of extreme events on any node. We show that the analytical and simulation results are in good agreement.

II. BIASED RANDOM WALK ON NETWORKS

A. Stationary distribution

We consider a connected, undirected, finite network with N nodes and E edges. The network is characterized by a symmetric adjacency matrix $\mathbf{A}$ with elements $A_{ij} = 1$ if nodes i and j are connected by an edge and $A_{ij} = 0$ otherwise. There are W independent walkers performing biased random walks on this network in the sense explained below. We denote by b_{ij} the transition probability for a walker to hop from node i to a neighboring node j. Let P_{ij} be the probability that a walker starting at node i at time $n = 0$ is at node j at time n. Then, the master equation can be written as

$$P_{ij}(n+1) = \sum_l A_{lj} b_{lj} P_{il}(n). \tag{1}$$

The random walkers are biased by taking the time-independent transition probability for hopping from the lth node to the jth node to be [15–17]

$$b_{lj} \propto k_j^\alpha, \tag{2}$$

where α is a parameter that defines the degree of bias imparted to the walkers. Clearly, $\alpha = 0$ corresponds to the standard random walk, where the transition probability is unbiased and a walker can hop to any neighboring node with equal probability. For $\alpha > 0$, the random walkers are biased toward nodes with larger degree or hubs. In contrast, if $\alpha < 0$, walkers preferentially hop to small degree nodes. The larger (smaller) α is, the stronger the bias toward hubs (small degree nodes) is. Then, the normalized transition probability becomes

$$b_{lj} = \frac{k_j^\alpha}{\sum_{m=1}^{k_l} k_m^\alpha}. \tag{3}$$

The summation in the denominator runs over the nearest neighbors of node l. Using the transition probability in Eq. (3), the master equation becomes

$$P_{ij}(n+1) = \sum_l A_{lj} \frac{k_j^\alpha}{\sum_{m=1}^{k_l} k_m^\alpha} P_{il}(n). \tag{4}$$

By repeated iteration of Eq. (4), it can be shown that $P_{ij}(n)$ as $n \to \infty$ leads to the stationary distribution

$$\lim_{n\to\infty} P_{ij}(n) = p_j = \frac{k_j^\alpha \sum_{l=1}^{k_j} k_l^\alpha}{\sum_{m=1}^{N} \left(k_m^\alpha \sum_{l=1}^{k_m} k_l^\alpha\right)}. \tag{5}$$

We can define the generalized strength of the jth node to be

$$\phi_j = k_j^\alpha \sum_{i=1}^{k_j} k_i^\alpha, \tag{6}$$

which is a measure of the ability of a node to attract walkers. Note that ϕ_j depends on the bias parameter α and the degree of the nearest neighbors to which it is connected by an edge. Hence, it is possible for nodes with the same degree to have different generalized strengths. Thus, the generalized strength of the node is independent of the global network structure but is dependent on the local connectivity structure around the node. This is in contrast to the case of a standard random walk (on networks) in which the large-scale structure of the network topology plays no significant role. The local network structure is important for biased random walks on networks. In Fig. 1, we show how the generalized strength ϕ depends on the degree of a node for several values of α in a scale-free network with degree exponent $\gamma = 2.2$. For $\alpha = 1$ (crosses in Fig. 1), the generalized strength of a node is higher for large degree nodes (hubs), and an approximate linear relation is seen between ϕ_i and k_i of the ith node. For $\alpha = 0$, which is the standard random walk case, the generalized strength of the node is the same as the degree of the node (solid circles in Fig. 1). However, for $\alpha = -1.0$, ϕ is independent of k, especially for large degree nodes (triangles in Fig. 1). In this case, the bias

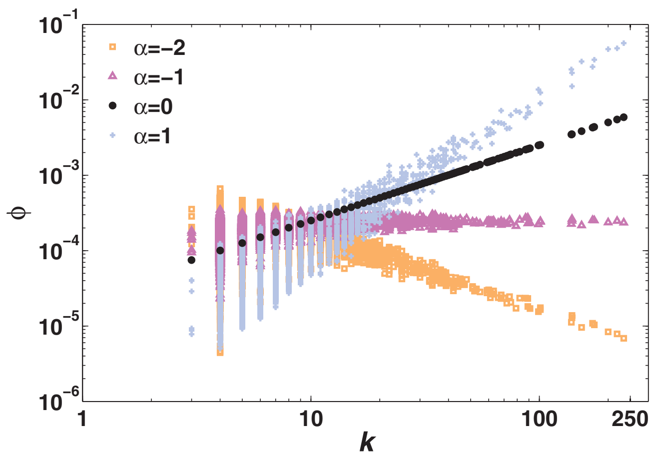

FIG. 1. (Color online) Strength ϕ as a function of degree k for different values of α in log-log plot.

in the random walk represented by its generalized strength ϕ is balanced by the degree of the node. In a scale-free network, a large number of small degree nodes are present, and they do not have identical values for the generalized strength ϕ. This explains the spread in ϕ for all values of $k < 50$. Upon a further decrease in the bias parameter α below -1.0 (open squares in Fig. 1), nodes with a smaller degree or neighbors with a smaller degree become important, and the generalized strength decreases with increasing degree.

B. Extreme-event probability

The stationary distribution for the number of walkers in the jth node can be rewritten in terms of the generalized strength ϕ as

$$p_j = \frac{\phi_j}{\sum_{l=1}^{N} \phi_l}. \tag{7}$$

Thus, every node can be uniquely characterized by its generalized strength ϕ. It is expected that two nodes with the same value of ϕ show similar behavior as far as biased walks on networks based on Eq. (2) are concerned. In the case of $\alpha = 0$, we get $\phi_i = k_i$, and the stationary distribution simplifies to $p_j = \frac{k_j}{2E}$, the result obtained for the case of a standard random walk in Ref. [18]. Thus, in the case of a standard random walk, the degree k characterizes the node. In the case of uncorrelated random networks, the stationary occupation probability can be further simplified by using the mean field approximation and can be written as [15,16]

$$p_j = \frac{k_j^{\alpha+1}}{N\langle k^{\alpha+1}\rangle}. \tag{8}$$

This approximate result suggests that the nodes with the same degree should have identical transition probabilities [15]. This does not necessarily hold well for the nodes of correlated networks, such as scale-free networks. This is because in a scale-free network, the neighborhoods of nodes with an identical degree are not identical. Hence, to study extreme events we use Eq. (7) instead of Eq. (8).

Given that Eq. (7) gives the probability to find one walker on the ith node with generalized strength ϕ_i, we can now obtain the distribution of random walkers on the ith node. The formulation is applicable to any node on the network, and hence, in our further discussions, we suppress the index i of the node. The random walkers (W) are independent and noninteracting, and hence the probability $f(w)$ of finding w walkers on a node is p^w, while the rest of the walkers, $W - w$, are distributed on the rest of the nodes of the network. When properly normalized, this leads to a binomial distribution given by

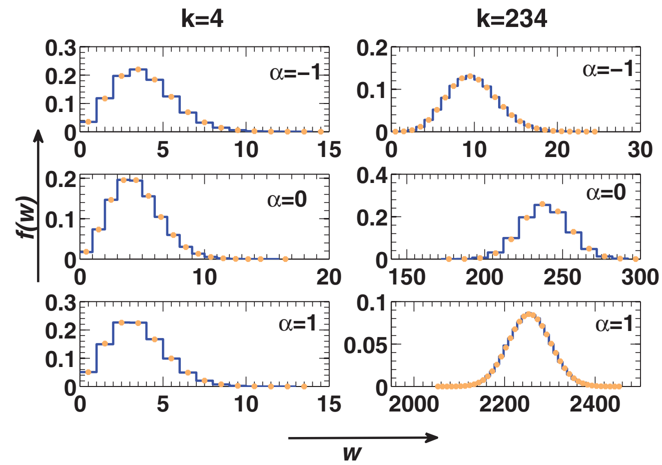

FIG. 2. (Color online) The distribution of walkers on two nodes with $k = 4$ and $k = 234$ for $\alpha = -1.0$,0.0, and 1.0. The solid lines show the distribution of walkers obtained from simulation, while the solid circles belong to the binomial distribution obtained analytically using the stationary probability in Eq. (7).

$$f(w) = \binom{W}{w} p^w (1-p)^{W-w}. \tag{9}$$

The mean and variance of the flux passing through the given node is

$$\langle f\rangle = W\frac{\phi}{\sum_{l=1}^{N}\phi_l},$$
$$\sigma^2 = W\frac{\phi}{\sum_{l=1}^{N}\phi_l}\left(1 - \frac{\phi}{\sum_{l=1}^{N}\phi_l}\right). \tag{10}$$

Note that the results in Eqs. (9) and (10) depend only on the generalized strength ϕ that characterizes a node including its neighborhood. It does not depend on the large scale connectivity pattern. Hence, these results will hold good for any network, such as scale free, random, or small world, irrespective of its degree distribution. Further, in the cases for which $\sum_{l=1}^{N}\phi_l \gg \phi$, we obtain the approximate relation $\sigma \approx \langle f\rangle^{1/2}$. This relation can be thought of as a generalization of a similar relation for the unbiased random walks reported in Ref. [14]. However, the exponent 1/2 is not universal and instead depends on details such as the fluctuation in number of walkers and sampling resolution of the flux [19]. The distribution of random walkers on two nodes with different degrees, $k = 4$ and $k = 234$, is plotted in Fig. 2. The biased random walk simulations were performed on a scale-free network with 5000 nodes with 19 915 links and 39 830 walkers. Initially, at time $n = 0$, the walkers are randomly distributed on N nodes. The simulation results presented in Fig. 2 have been obtained after averaging over 100 realizations with different initial conditions of random walkers. The simulation results, the solid lines in Fig. 2, show a good agreement with the analytical distribution given by Eq. (9).

III. PROBABILITY FOR EXTREME EVENTS

We take an extreme event to be the one for which the probability of occurrence is small and is typically associated with the tail of the probability distribution function for the events. We extend this principle to the events on the nodes of a network [14]. Given that the number of walkers w passing through a node with generalized strength ϕ follow the binomial distribution, if more than q walkers pass through the node, then it is an extreme event for the node. Then, the probability for the occurrence of extreme event is

$$F_i = \sum_{w=q_i}^{W} \binom{W}{w} p_i^w (1-p_i)^{W-w} \tag{11}$$

$$= I_{p_i}(\lfloor q_i \rfloor + 1, W - \lfloor q_i \rfloor), \tag{12}$$

where $\lfloor u \rfloor$ is the floor function defined as the largest integer not greater than u and $I_z(a,b)$ is the standard incomplete beta function [20]. In this form, the extreme event probability will depend on the choice of threshold q_i. First, we consider the case of constant threshold. If $q_i = 0$, using Eq. (11) we obtain $F_i = 1$ for all the nodes on the network. Thus, all the nodes will experience extreme events all the time. On the other hand, if we set $q_i = W$, then we obtain

$$F_i = p_i^W. \tag{13}$$

Since $p_i \ll 1$, we get $F_i \approx 0$ for all the nodes, implying that there are no extreme events anywhere in the network. Hence, these choices of threshold values are not physically interesting cases. Any other arbitrary choice such as $q_i = q_0$, where q_0 is a constant, will predominantly lead to some nodes encountering extreme events nearly all the time and others having no events at all. This too is not an interesting case. The foregoing arguments imply that an interesting scenario would arise if the threshold is chosen to be proportional to the natural variability of the flux passing through a node. Thus, we choose the threshold for extreme events to be [14]

$$q_i = \langle f_i \rangle + m\sigma_i, \tag{14}$$

where $m \geqslant 0$. The mean flux $\langle f_i \rangle$ and standard deviation σ_i are given by Eq. (10). Substituting q_i in Eq. (12), it is clear that the probability for the occurrence of extreme events is dependent only on the generalized strength ϕ of the node. In Fig. 3, we show the simulation and analytical results for the probability of extreme events as a function of ϕ for several choices of α. The numerical results are based on simulations with $W = 39\,380$ walkers on a scale-free network with $N = 5000$ nodes evolved for 10^7 time steps. An unusual feature is that F_i predicts a higher probability of occurrence of extreme events, on average, for nodes with small values of generalized strength ϕ than for the nodes with higher values of generalized strength ϕ. For instance, in Fig. 3(a), the probability of extreme-event occurrence is generally higher for nodes with $\phi < 10^{-5}$ than for nodes with $\phi > 10^{-3}$. A similar effect is seen in Figs. 3(b)–3(e). Even though nodes with higher generalized strength ϕ attract more walkers as given by Eq. (5), this

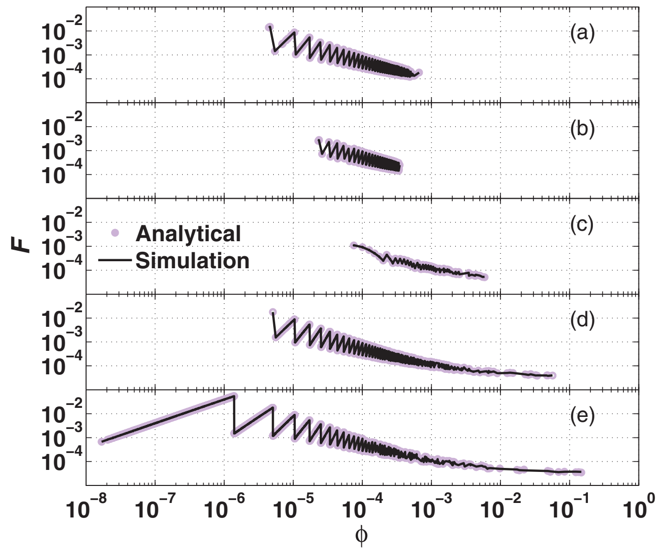

FIG. 3. (Color online) The probability of the occurrence of extreme events plotted as a function of the node generalized strength ϕ (normalized) for different values of bias parameters: (a) $\alpha = -2.0$, (b) $\alpha = -1.0$, (c) $\alpha = 0.0$, (d) $\alpha = 1.0$, and (e) $\alpha = 2.0$. The threshold for an extreme event is $q = \langle f \rangle + 4\sigma$. The circles are from analytical results in Eq. (12), while the solid lines are the simulation results performed on a scale-free network ($N = 5000$, $E = 19\,915$) with $W = 2E$ walkers averaged over 100 realizations with randomly chosen initial positions of walkers.

does not imply that they also have a higher probability for extreme events. This is a generalization of the result obtained in Ref. [14] for the standard random walk on networks which shows that extreme events are more probable for nodes with a small degree than for the ones with a high degree. The local fluctuations seen in Fig. 3 are inherent in the system and not due to insufficient ensemble averaging. Further, notice that Eq. (12) does not depend on the large scale structure of the topology, and hence it is valid for biased random walks on any topology, random or small world or scale free.

However, the local connectivity patterns in the vicinity of any node play a crucial role in the diffusion of an extreme event. Suppose an extreme event takes place at node A at time n; then one interesting question is how probable it is for an extreme event to take place in its immediate neighborhood at time $n+1$, i.e., after the first jump. We call it first-jump probability, and it is similar to the one reported in [21]. In the case of a standard random walk ($\alpha = 0$), our simulations (not shown here) indicate that, in general, if node A is a hub, then the probability to encounter an extreme event in its neighborhood is higher (at least by a factor of 3–4) compared to the case when node A is a small degree node. For biased random walks, the results suggest a higher likelihood for an extreme event to be transferred to its neighborhood in the case when $\alpha < 0$ compared to the case with $\alpha > 0$.

IV. FLUCTUATIONS IN EVENT SIZE

The size of an event is measured in units of the standard deviation σ of the flux passing through a node. In this section, we show that the extreme fluctuations in the flux of walkers are

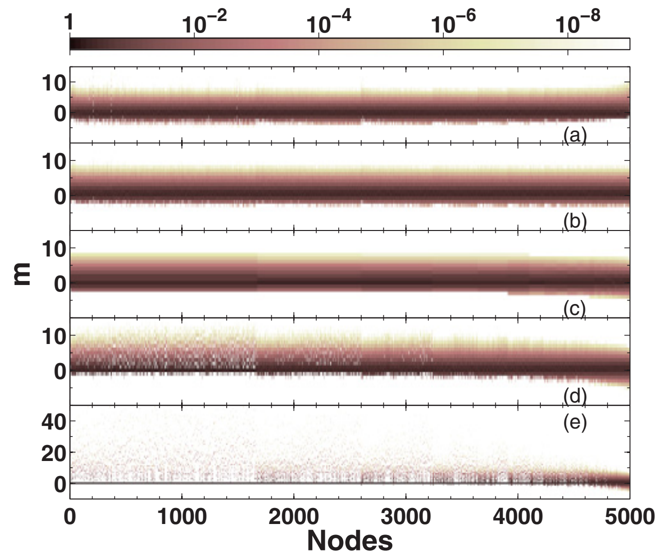

FIG. 4. (Color online) The distribution of event sizes for biased random walks as a function of the node number on the x axis obtained from simulations performed on a scale-free network for different values of the bias parameter: (a) $\alpha = -2.0$, (b) $\alpha = -1.0$, (c) $\alpha = 0.0$, (d) $\alpha = 1.0$, and (e) $\alpha = 2.0$. The nodes are arranged in order of increasing degree. The probability values $\mathcal{P}_m$ are color coded. This should be compared with analytical results in Fig. 5.

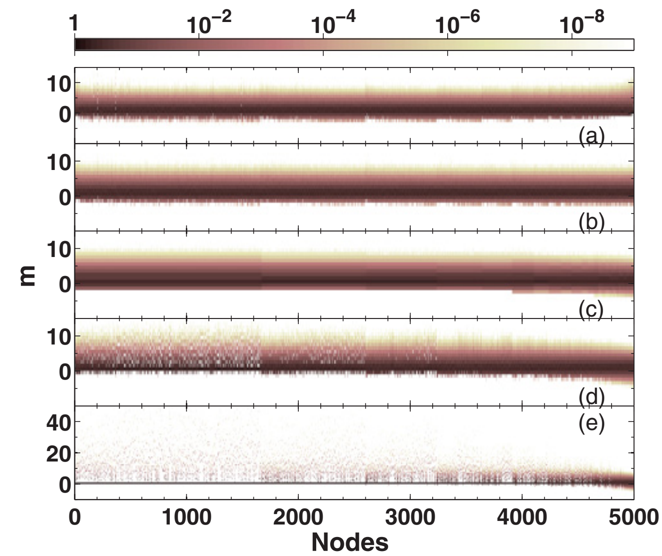

FIG. 5. (Color online) The distribution of event sizes for biased random walks as a function of the node number on the x axis obtained analytically using Eq. (15) for different values of the bias parameter: (a) $\alpha = -2.0$, (b) $\alpha = -1.0$, (c) $\alpha = 0.0$, (d) $\alpha = 1.0$, and (e) $\alpha = 2.0$. The nodes are arranged in the order of increasing degree. The probability values $\mathcal{P}_m$ are color coded.

realized in the case of $\alpha = 2$, which implies that the walkers are biased toward the nodes with a larger generalized strength ϕ (hubs). An event is of size m if $m\sigma \leqslant w - \langle w \rangle < (m+1)\sigma$, where w is the number of walkers on a given node.

Then, the probability for the occurrence of an event of size m can be written down as

$$\mathcal{P}_m = I_p(\lfloor q_m \rfloor + 1, W - \lfloor q_m \rfloor) - I_p(\lfloor q_{m+1} \rfloor + 1, W - \lfloor q_{m+1} \rfloor). \tag{15}$$

To illustrate the result, we show the distribution of event sizes in Fig. 4 for $\alpha = -2,-1,0,1,2$ in a scale-free network obtained from simulations evolved for 10^7 steps and averaged over 100 ensembles. Here, the events with a probability of occurrence of less than 10^{-8} have been discarded to maintain the numerical accuracy. In the case of $\alpha = 0$ (standard random walk), the distribution of events is shown in Fig. 4(c). The events of size $m = 0$ are highly probable with $\mathcal{P}_0 \sim 0.1$. In contrast, the probability for events of size $|m| > 0$ decreases, and thus extreme events of size $m = -2,8$ occur with probability $\mathcal{P}_{-2} \sim \mathcal{P}_8 \sim 10^{-8}$. The limitation on the lower limit of event sizes is restricted by the minimum possible number of walkers on a node, i.e., 0. For lower degree nodes, events of sizes -2σ to 8σ are observed, but in the case of higher degree nodes $k > 100$, event sizes range from -5σ to 6σ only. In the case of a standard random walk, for the whole network, event size m varies from -5σ to 8σ.

In comparison, for the case of $\alpha = 1$ shown in Fig. 4(d) the events of size 8 have a higher probability of occurrence ($\mathcal{P}_8 \sim 10^{-7}$), and events of even higher sizes are also possible. For $\alpha = 2$, even higher size events, as large as 40, become highly probable for small degree nodes, as seen in Fig. 4(e). Thus, in general, for larger α, larger size events become probable when compared with the case of $\alpha = 0$. Physically, this can be understood as follows. With $\alpha = 0$, the random walkers perform unbiased random walks. However, for $\alpha = 2$, the walkers preferentially choose to hop to nodes with a larger degree (hubs). Since large degree nodes are mostly well connected among themselves, very few walkers reach small degree nodes. Hence the average flux through the small degree nodes becomes so small that even occasional visits by a few walkers lead to extremely large size events. These occasional visits lead to a probability of order 10^{-6} even for events of size 40. Hence, in the case of biased random walks, extremely large fluctuations in event sizes can be observed in small degree nodes. This effect is also seen in the analytical results obtained using Eq. (15), shown in Fig. 5.

On the other hand, for cases $\alpha = -2,-1$ such large fluctuations are not visible in the event sizes in Figs. 4(a) and 4(b). For $\alpha = -1$ in Fig. 4(b), there is a small increase in the event sizes (when compared to $\alpha = 0$) for the small degree nodes, but it is not as large as in the $\alpha = 1$ case. Further, with $\alpha = -1$, it must also be noted that the probability profile remains similar for most of the nodes irrespective of the large differences in their degree. This is because ϕ is an approximate constant for most of the nodes since, in this case, the effect of the bias is balanced by the degree of these nodes. For $\alpha = -2$, the flux is strongly biased toward small degree nodes, and again events of sizes $m = 10$ can be seen in Fig. 4(a), though only on the higher degree nodes. The event sizes for hubs are not as large as observed in the case of $\alpha = 2$ for lower degree nodes. It can be explained as follows: when $\alpha = -2$, the flux preferentially flows through the small degree nodes, which form the bulk in a scale-free network. Most small degree nodes do not have a direct link with other small degree nodes but are connected through a hub. Hence, despite the biased walk favoring the small degree nodes, a sufficiently large flux flows through the hubs as well. Hence, abnormally large event size fluctuations are not seen in hubs for $\alpha = -1, -2$. All these features show a good agreement with the analytical result obtained in Eq. (15) and shown in Fig. 5.

V. DISCUSSION AND SUMMARY

This work is an attempt to understand the extreme events occurring on the nodes due to flow on networks which typically is directed toward or away from the hubs. In this work, we study a biased random walk model in which the traffic preferentially moves either toward or away from the hubs, and we analytically obtain the probabilities for the occurrence of extreme events. In this framework, extreme events are due to inherent fluctuations in the flux passing through any node and is defined as exceedences above a chosen threshold q. The threshold is chosen to be proportional to the natural variability of the node. Each node on the network is characterized by the generalized strength ϕ, which depends on its degree and that of its immediate neighborhood. It is a measure of how much traffic is attracted to the particular node. The larger the generalized strength of a node is, the larger its ability to attract walkers is. In this paper, we have shown that the nodes with a smaller generalized strength, on average, have a higher probability for the occurrence of extreme events when compared to nodes with a higher generalized strength. Further, we have also shown that when the flux is biased toward the hubs, abnormally large fluctuations in event sizes become highly probable. This is one possible signature of the topologically biased flow in a scale-free network.

In general, it is possible to conceive of many ways by which bias can be imparted to independent random walkers on networks. These biasing strategies are motivated by real observations and the quest for efficient search strategies on networks. Various kind of biases based on the local environment, shortest paths, the entropy of random walk, and various adaptive strategies are some examples of biased random walk on networks [21–26]. It will be interesting to study the extreme-event probabilities under such biasing strategies. However, we emphasize that if the stationary probability distribution equivalent to Eq. (5) exists for all the above strategies, then it would be possible to define extreme events and analyze them following the methods presented in this work.

In the context of scale-free network, it has been argued that hubs are important for better functioning of the network. Apart from being responsible for providing better connectivity, the existence of hubs makes the scale-free network robust against random node removal but fragile if the node removal is targeted [27,28]. The results in this paper show that extreme events due to natural fluctuations are more probable on small degree nodes (when compared to the hubs). Hence special attention must be paid to designing the capacity of the small degree nodes so that extreme events can be smoothly handled without leading to disruption of the node. The results in this paper can be used to estimate the capacity a node should possess if it should handle extreme events of size, say, m. If we want the node to handle 4σ events smoothly, then the required capacity can be obtained by inverting Eq. (12). Thus, the numbers so obtained can be useful as an input for arriving at the capacity at which the nodes on a network should be built.

ACKNOWLEDGMENTS

The simulations were carried out on computer clusters at PRL, Ahmedabad, and IISER Pune. V.K. would like to thank IISER Pune for the hospitality provided during his stay in Pune.

[1] *Extreme Events in Nature and Society*, edited by S. Albeverio, V. Jentsch, and Holger Kantz (Springer, Berlin, 2005).
[2] S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, London, 2001).
[3] http://blog.twitter.com/2010/02/measuring-tweets.html.
[4] http://www.comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009.
[5] U.S.-Canada Power System Outage Task Force (2004); see https://reports.energy.gov.
[6] D. De Martino, L. Dall'Asta, G. Bianconi, and M. Marsili, Phys. Rev. E **79**, 015101(R) (2009).
[7] P. Echenique, J. Gomez-Gardenes, and Y. Moreno, Europhys. Lett. **71**, 325 (2005).
[8] K. Kim, B. Kahng, and D. Kim, Europhys. Lett. **86**, 58002 (2009).
[9] B. Tadic, G. J. Rodgers, and S. Thurner, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 2363 (2007).
[10] D. J. Ashton, T. C. Jarrett, and N. F. Johnson, Phys. Rev. Lett. **94**, 058701 (2005).
[11] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).
[12] R. Germano and A. P. S. de Moura, Phys. Rev. E **74**, 036117 (2006).
[13] W. Wang, Z.-X. Wu, R. Jiang, G. Chen, and Y. C. Lai, Chaos **19**, 033106 (2009).
[14] V. Kishore, M. S. Santhanam, and R. E. Amritkar, Phys. Rev. Lett. **106**, 188701 (2011).
[15] A. Fronczak and P. Fronczak, Phys. Rev. E **80**, 016107 (2009).
[16] W.-X. Wang, B. H. Wang, C. Y. Yin, Y. B. Xie, and T. Zhou, Phys. Rev. E **73**, 026111 (2006).
[17] S.-J. Yang, Phys. Rev. E **71**, 016107 (2005).
[18] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).
[19] S. Meloni, J. Gomez-Gardenes, V. Latora, and Y. Moreno, Phys. Rev. Lett. **100**, 208701 (2008).
[20] J. L. Lopez and J. Sesma, Integr. Transforms Spec. Funct. **8**, 233 (1999).
[21] B. Tadic, Eur. Phys. J. B **23**, 221 (2001).

[22] V. Sood and P. Grassberger, Phys. Rev. Lett. **99**, 098701 (2007).

[23] S. Chen, W. Huang, C. Cattani, and G. Altieri, Math. Probl. Eng. **2012**, 732698 (2012).

[24] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, Phys. Rev. Lett. **102**, 160602 (2009).

[25] R. Sinatra, J. Gómez-Gardeñes, R. Lambiotte, V. Nicosia, and V. Latora, Phys. Rev. E **83**, 030103(R) (2011).

[26] A. N. Mian, R. Baldoni, and R. Beraldi, *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops*, *Montreal, QC* (IEEE, Piscataway, NJ, 2009), pp. 153–157.

[27] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).

[28] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).