RANDOM WALKS AND EXTREME EVENTS ON COMPLEX NETWORKS



A thesis submitted towards partial fulfilment of BS-MS Dual Degree Programme

by

NEHA BORA

under the guidance of

DR. M. S. SANTHANAM

Indian Institute of Science Education and Research Pune

Certificate

This is to certify that this thesis entitled Random Walks and Extreme Events on Complex Networks submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research Pune represents original research carried out by Neha Bora at Indian Institute of Science Education and Research Pune, under the supervision of Dr. M. S. Santhanam during the academic year 2014-2015.

Student

NEHA BORA

Supervisor

DR. M.S. SANTHANAM

Declaration

I hereby declare that the matter embodied in the report entitled **Random** Walks and Extreme Events on Complex Networks are the results of the investigations carried out by me at the Department of Physics, Indian Institute of Science Education and Research Pune, under the supervision of Dr. M. S. Santhanam and the same has not been submitted elsewhere for any other degree.

Student

NEHA BORA

Supervisor

DR. M.S. SANTHANAM

Acknowledgements

First and foremost I would like to thank my supervisor Dr. M. S. Santhanam for his unceasing support and guidance. I want to thank him letting me work at my own pace, without setting any deadlines and at the same time for keeping me from losing track during the project. I have enjoyed our discussions: academic and otherwise.

This year wouldn't even be bearable without my friends. One person I cannot thank enough is Aashay. I want to thank him for being with me through everything. He has been my friend, my family, my mentor and even my student sometimes (though that was just an extreme event) :P. Next, I would like to thank my closest friend Kshiti(Kishti ben, Shakti, Kriti, Kruti and many names she is known by). Coming back to the hostel wouldn't have been the same if it were not for you. I would like to thank my "obsessed with trivia" "classy" friend Pranav for making those coffee sessions "fun";). I would like to thank Vishnu(VVK) for being absolutely awesome and coming up with amazing conversations. Those were my true stress relievers. Additionally, I would like to thank Akshay, Sukriti, Sagar, Sharath, Anurag (Pandu), Meghana, Abhilesh (Leshu), Neha M, Reshma, Charu and all my friends for making this year a memorable one. I would like to thank my cousin, Dolly for feeding me home-food at short notice and for always being there for me in her own way.

I would like to thank thank Mr. Prabhakar for helping us out with all administrative issues so that we could work at our highest efficiency. I would also like to thank DST for the INSPIRE fellowship. I would like to thank Mr. Prabhas from the accounts section for making sure that reimbursements from my INSPIRE contingency grant made a quick way to my bank account.

Lastly, I would like to dedicate this thesis to my mother.

Abstract

Dynamics on complex networks, such as traffic on roads or information packets on network of routers, display a variety of collective and emergent properties. Of practical interest are congestion and extreme events phenomena, which ultimately control the smooth functioning of networks. To get a deeper understanding of these phenomena, we employ a continuous-time random walk model with probabilistic routing protocol for traffic flows in complex networks such that it contains the most relevant characteristics of real-world systems. We study the collective behavior through phase transitions in congestion and individual behavior of nodes through extreme events. We observe that increasing the outgoing flux enlarges the free-flow region in the parameter space. Moreover, a degree-dependent outflux can completely eradicate the congested state in the parameter space. In accordance with the previous results for a parameter-free model, we see that in most cases, small degree nodes are more prone to experience extreme events than the hubs. We also notice a striking relation between the flux fluctuations and extreme event probability on nodes.

Contents

1	Introduction	2
	1.1 Networks	3
	1.2 Dynamics on networks	6
2	Model of Dynamics	9
	2.1 Properties of the Model and Analysis of the dynamics	12
3	Extreme Events	16
	3.1 Extreme Events in Conservative Model	18
	3.2 Extreme events in Non-conservative Model	21
4	Phase Transitions in Non-Conservative Model	27
5	Conclusion	31
	References	33

Chapter 1

Introduction

Networks are all around us - in the form of road networks, communication networks and the Internet to name a few. Networks are formed by us - network of friends, family members and colleagues. Networks are even inside us - the network of veins and arteries, neurons etc. We rely on networks at every moment, in one way or the other. Be it finding the optimal way to go from place A to B, gossiping about a friend's friend, or effective blood circulation to all parts of the body, smooth flow of traffic on networks is of critical importance to us. Traffic jams, unresponsive webpages, cardiac arrest due to artery blockage are just a few examples of effects of traffic disruption on networks. Like all complex systems, networks too, are inherently and unavoidably vulnerable to failures. They undergo many local failures. However, a global failure, such as congestion, occurs when small, seemingly innocuous local failures unite leading to an organized accident. The largest power blackout that occurred in India in 2012, pushing half the nation's population into darkness and incurring huge financial losses is a burning example of power-grid failures due to congestion in networks. Thus, to avoid such losses, it is vital to study traffic flow on networks. One of the widely studied transport phenomena is congestion on networks[1]. Congestion is generally studied by defining a routing protocol for random walkers generated on the network and capacity of the nodes to service them. Thus, a node is said to be in a congested state when it exceeds its capacity to service the incoming flux of walkers. Using this framework, several results on stability of networks, cascading failure to congestion transition have been obtained. A recently studied phenomena on networks is extreme events (EE)[2]. Unlike congestion, an extreme event may not be related to the handling capacity of the nodes, but is defined as exceedance over a prescribed quantile. As opposed to congestion, EE arise from natural fluctuations in the traffic passing through a node and not due to constraints imposed by its capacity. Theoretically, congestion can be completely removed from the system by fine-tuning appropriate parameters. However, EE can never be absolutely eradicated as fluctuations are inherent to any system which is probabilistic in nature. As congestion and EE are different in definition, their effect on the network is also different. A single congested node drives the entire network in a state of congestion. However, no studies as of now suggest individual EE can cause a global failure. EE have so far been studied using a random walk model[2]. The simple model claims, contrary to popular intuition, that lower degree nodes would experience relatively more number of EE than the higher degree nodes. Again, this is opposite to how nodes get congested, the hubs being the first. Hence, network design principles so far have mostly focused on hubs. However, such an approach might fail in context of EE. Are smaller degree nodes vulnerable to attacks even in a realistic frameworks? Motivated by this unique and non-trivial behavior, we aim to study EE and congestion-phase transitions on a simple and realistic traffic model on complex networks.

In the following sections, we discuss relevant preliminary concepts of graph theory, characteristics of real-world networks and give a brief overview of dynamics on networks.

1.1 Networks

Many situations in daily life can be described by means of a diagram consisting of a set of points together with lines joining certain pairs of these points [3]. The set of points is called nodes or vertices denoted by N and the pairwise connections are called edges or links denoted by E. A set of nodes and edges define a graph, for example G = (N, E). Graphs are generally referred as networks in non-mathematical communities. When edges have certain value associated with them, they are called weighted networks. Links/edges may also have directions associated with them. When all the links have no directions or point in both the directions, it is called an undirected graph/network. When all the links point in only one direction, it is called a directed network. A graph can be defined using an adjacency matrix or adjacency lists. We use a $N \times N$ adjacency matrix A to represent a graph G(N, E). The elements of the matrix are defined as,

$$A_{ij} = \begin{cases} 1, & \text{if i and j share an edge} \\ 0, & \text{otherwise} \end{cases}$$
 (1.1)

such that

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

The degree of the i^{th} node, denoted by $k_i(c)$, is defined as the number of non-zero elements in that row. Degree of a node represents the number of links connected to that node. The adjacency matrix of an undirected graph is a symmetric matrix, i.e. $A_{ij} = A_{ji}$ and that of a directed graph is antisymmetric, i.e. $A_{ij} \neq A_{ji}$. An interesting property of the adjacency matrix is that A^n gives the number of paths of length n in between the nodes. We demonstrate an example of an adjacency matrix of a graph G(N,E) with N = 5 nodes and E = 7 edges [4].

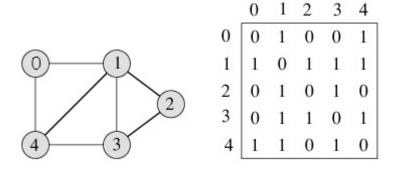


Figure 1.1: Graph G(N, E) with N=5 and E=7 and Adjacency Matrix A

Henceforth, for all our calculations, we use an undirected graph represented by an adjacency matrix.

Many complex systems such as communication, transportation and social systems can be modeled using networks. Empirical studies of real-world networks show that these networks have certain characteristic properties which seem to arise from their topology. The two extremes of network models are random and regular networks. In a regular network, all the nodes have same degree whereas for a random network [5], we start with N nodes and then connect every pair of nodes with probability p, creating a graph with approximately pN(N-1)/2 edges distributed randomly. Real-world networks differ from random and regular networks due to their peculiar characteristics.

The three peculiar characteristics of real-world networks are small world property, clustering co-efficient and degree distribution. According to the small world property, even if the size of the network is large, the distance between nodes is relatively short. The most popular demonstration of the

small-world property is the "six degrees of separation" concept [6], unveiled by Stanley Miligram, who concluded that there was a path of acquaintances with a typical length of about 6 between most pairs of people in the United States [7]. Mathematically, a network is called small world if the average distance scales logarithmically or slower with node number N. Random graphs also posses this property. On the other hand, in regular graphs, the average distance increases linearly with number of nodes N. Another important property is clustering in real-world networks, especially observed in social networks. It is basically the formation of triangles in a network. An example of clustering is a set of acquaintances in which every member knows every other member. In most real networks, the value of the clustering co-efficient [8] is much larger than it is in a random network with the same number of nodes and edges. The property which makes real-world networks robust to random attacks is the probability distribution of degrees of nodes of the network. Degree distribution of a random network is a Poisson distribution with a peak at $P(\langle k \rangle)$, where $\langle k \rangle$ is the average degree of the network. In most large networks, including the World Wide Web(www)[9], the degree distribution has a power law tail.

$$P(k) \sim k^{-\gamma} \tag{1.3}$$

Such networks are called scale-free networks. Barabasi and Albert (1999)[10] devised an algorithm to construct a scale-free network. The key concepts of this algorithm are growth and preferential attachment. Prior to this algorithm, all network models started with a fixed number of nodes N which were then randomly connected or rewired keeping N constant. Unlike these models, real-world networks are actually formed by attaching new nodes over time to a nucleus of nodes existing at time t=0. For example, the www grows exponentially in time by adding new web pages. Secondly, while constructing a random graph, two nodes are connected randomly, independent of their degree. However, most real networks show preferential attachment where the chances of connecting to a node depends on the degree of the node. Thus, a network built using these two principles evolves into a scale-invariant state with the probability that a node has a degree k follows a power law with an exponent $2 < \gamma < 3$. Like real-world networks, scale-free networks are also robust towards failure. As scale-free networks are closest to the realworld networks, we use a scale-free network with N=1000 nodes as the underlying network for our analysis.

1.2 Dynamics on networks

Random walk is a rudimentary model in statistical physics which has numerous applications in a wide variety of fields. Simply put, random walk can be defined as a series of steps taken at random. The path traced by a drunkard or a molecule in the air or animals foraging in a grassland are classic examples of random walk. Consider a walker standing at the origin of the number line. At each time-step, it moves to the right or left with probability p or (1-p) respectively. This is a simple random walk model in one dimension. Similarly, we can construct a random walk on a lattice or any network where each walker chooses to jump to one of its nearest neighbors with equal probability. Random walks are studied in both discrete and continuous space and time. A continuous time random walk(CRTW) can be defined as a random walk where a walker waits for a random amount of time between successive jumps. Weiner process, which characterizes Brownian motion [11], is a popular example of CTRW in which jumps are continuous and normally distributed whereas waiting times follow an exponential distribution. Random walk is one of the simplest ways to explore a network and study its properties. However, the accumulation of walkers/packets on a node is not involved in the random walk model. Hence, it does not reflect a real traffic system completely. In the basic traffic dynamics model [12] [13], an iterative process, all the nodes can generate and deliver walkers/packets. At each time step, packets are generated with a probability p on each node and assigned a random destination node, different from its source. Each packet is delivered from source to destination following a particular routing strategy. Each node is endowed with a queue which may follow first-in-first-out or some other rule to send packets at each time step. Each node has its own delivering capacity or outflux, C, which is the maximum number of packets it can deliver in one time step. Once a packet reaches its destination, it is removed from the network.

The basic model has been altered to make it realistic, by making the packet-generating and packet-delivering capacity dependent on the degree of the node[14][15] or selecting the source and destination in-homogeneously. The most famous routing strategy is the shortest path routing strategy (SPRS), where packets are transferred along the shortest path between the source and the destination. However, networks cannot handle heavy traffic if packets always follow SPRS [14] [16] [17]. An improvement to SPRS is the traffic awareness strategy where waiting time and queue length at neighboring nodes are considered along with shortest path length from neighboring node to the destination [18] [19]. In a heterogeneous network, traffic awareness strategy shortens the maximal time that is required for a packet to reach

its destination and enhances the traffic capacity, the quantitative measure of congestion, compared to SPRS. However, in SPRS the transition from freeflow to jammed state as number of packets generated increase is smooth, unlike in the traffic awareness strategy where the transition is discontinuous. Thus, delay in congestion due to traffic awareness strategy comes at a cost of abrupt congested state without prior warnings. Many strategies, such as above, have been proposed considering global topological information, which are mostly pragmatic for small networks. However, for large networks, like the Internet, strategies based on local information are favored due to the heavy cost incurred in searching global information on networks. Thus, based on these results, it was concluded [13] that under the conditions of heterogeneous delivering capacity of each node, the random walks strategy is the best strategy for routing packets. Random walk on complex networks has been studied with special attention to scale-free networks to see how the underlying network topology can affect the asymptotic diffusion of walker [20]. On a heterogeneous network, random walkers are not distributed uniformly over time, but are concentrated on the higher degree nodes. Thus, higher degree nodes are more prone to congestion. However, higher degree nodes are less susceptible to EE [2], contrary to our intuition. To explore this contrasting behavior, we adopt the random walk model and modify it by adding a probabilistic routing protocol to study EE and congestion on

Many empirical studies have revealed that traffic is not only dependent on the characteristics of the underlying network structure, but is also significantly affected by routing strategies. In order to improve transport performance on real networks, the following strategies are predominantly used:

- 1. **Soft** strategies, where the focus is to design efficient routing strategies.
- 2. **Hard** strategies, where the focus is to make appropriate changes in the underlying network structure. They are called hard strategies because making topological changes is expensive.

The strategies we have discussed so far fall into the realm of soft strategies. Hard strategies are costlier than soft strategies, as adding or rewiring links consumes finances, manpower among other things. However, removing links from network can be implemented at a low cost. Studies show that removing links could alleviate or even mitigate cascades of overloading on networks[21] and may help to enhance synchronization in complex networks of dynamical systems [22]. For example, in biology, removing metabolic reactions, which represent links in the metabolic network, could improve metabolic performances and rescue defective metabolic networks [23].

The next chapter contains a detailed description of the model adopted from [1]. It is basically a modified random walk model where we add correlations to make the model realistic in nature. We discuss the behavior of individual nodes and the system as a whole, based on probability of generating and rejecting walkers. In chapter 3, we discuss probability for occurrence of EE in a conservative model where the total number of walkers are constant and in the free-flow state of non-conservative model where the total number of walkers is oscillating. We observe the effect of rejection probability and outgoing flux from a node on EE. In chapter 4, we study the effect of outgoing flux on phase transitions. In Chapter 5, we conclude by summarizing the results and outlining the further scope of this work.

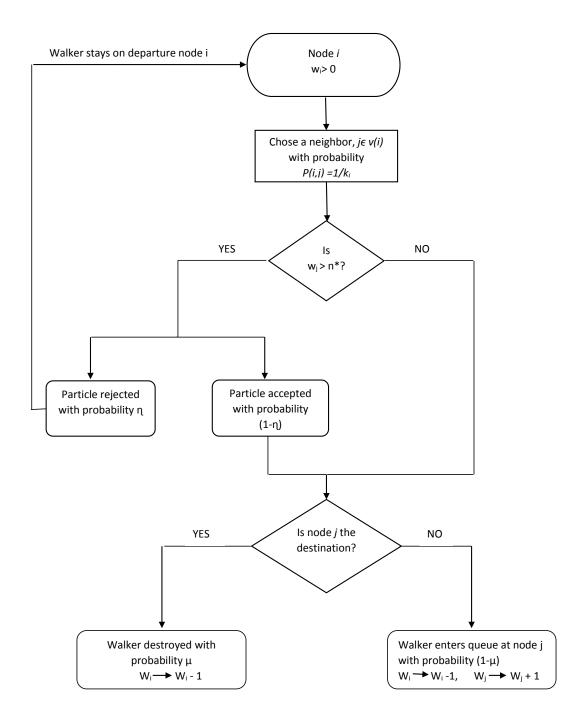
Chapter 2

Model of Dynamics

We use a standard random walk model along with a probabilistic routing rule [1] to study traffic dynamics on a network. We describe the dynamics of non-interacting walkers by adopting a routing protocol in a continuous-time random walk model on a network. Each walker has a source and a destination. Once a walker reaches its destination, it is thrown out of the system, i.e it is destroyed. Each node is endowed with a queue of infinite capacity in which walkers are stored. However, we try to impose constraints to restrict the unbounded increase in queue length. Consider a node i with a set of neighbors v(i). A walker at node i jumps in the queue of a randomly chosen neighbor $j \in v(i)$. Now the destiny of the walker depends on a) if node j is its destination and b) on the state of congestion of node j, in which case it may reject the walker and send it back to node i. We model both of these as probabilistic events.

In our model, we consider an undirected network of N nodes and E edges. We represent the network with an adjacency matrix, A, where A(i,j) = 1 when node i and node j share an edge, otherwise it is 0. The degree of a node i is represented as k_i . A walker is produced at a node i with probability p_i at time t. At any time t, certain number of walkers at every node attempt to jump onto one of the neighboring nodes at random. Every neighbor is chosen with an equal probability, $P_{i,j} = A_{i,j}/k_i$. Let w_i be the number of walkers on node i. We assume that node j rejects a walker with a probability $\overline{\eta}(w_j)$, which is non-decreasing with w_j . We also assume that μ_j is the probability that j is the destination node of a walker. If $w_i \geq 0$, a walker follows the following probabilistic jump rule: With probability $\overline{\eta}(w_j)$ the walker is rejected from the arrival node and remains on the departure node. Otherwise, walker is either killed during the jump, with probability μ_j , or with probability $(1 - \mu_j)$ enters the queue on node j. Figure 2.1 shows the routing algorithm for a walker at node i.

Figure 2.1: Routing algorithm for a walker on a node i



In this study, we focus on a simple case where $\mu_i = \mu$, $p_i = p$ and $\overline{\eta}(w_i) = \eta \theta(w_i - n^*)$ where $\theta(x)$ indicates the Heaviside step function for all node i. Thus, as per the routing protocol, node i refuses to accept a walker with probability η if $w_i \geq n^*$. Thus, here n^* is the threshold queue length for a node.

The statistics of the length of the queues $\{w_i\}$ is used to analyze the model. We do not need to use the information about trajectories of walkers, which lets us ignore the fate of individual walkers. This justifies our decision to use probabilistic modeling for the birth and death processes. Walkers are killed only during the jumping process and not when they are waiting in the queue. Again, this supports our probabilistic model as a walker should be removed from the system as soon as it reaches its destination. We have analytically computed the transition probability for a walker to go from node i to j, $j \in v(i)$, which is given by,

$$P_{i,j} = \frac{A_{i,j}(1 - \eta\Theta(w_j - n^*))}{k_i - \sum_{j \in v(i)} \eta A_{i,j}\Theta(w_j - n^*)}$$
(2.1)

The asymptotic behaviour of the system is determined by the relation between birth (p) and death (μ) probabilities. If $p >> \mu$, incoming flux is greater than outgoing flux, resulting in an over-loaded network. As p is independent of time, the system eventually moves in a non-equilibrium steady state in which queues increase with constant speed. We call this state of the system as a congested state. To quantify the state of the system, we introduce following observable [24]

$$\rho = lim \frac{\aleph(t+\tau) - \aleph(t)}{\tau Np}$$

where ρ is called the order parameter, $\aleph(t) = \Sigma_i w_i(t)$ is the total number of walkers at time t on the network, $p = N^{-1}\Sigma_i p_i$ is the average birth probability and τ is the observation time. To check the state of congestion of any node i, we can compute local order parameter by replacing $\aleph(t)$ by $w_i(t)$ and p by p_i . A system (node) is congested, in the steady state, if the (local) order parameter is greater than 0. Higher the order parameter, higher are the number of individual nodes in a congested state. Another important parameter in our model is the outgoing flux from a node, Ω , at any time t. Our primary aim is to study EE on a real-world system. However, to study EE, we need the system to be in a steady-state.

We consider two scenarios with the above model where we study EE:

- 1. Conservative system: No birth and death processes. n^* is degree-dependent.
- 2. Non-conservative system: Non-zero birth and death probabilities.

2.1 Properties of the Model and Analysis of the dynamics

We discuss the non-conservative model in detail. To begin with the simplest case, let us consider $\mu=0.2$, $\overline{\eta}=0$ and $\Omega=1$ walker. In this case, we allow only one walker to make a jump from each node at any time t. Also, $\overline{\eta}=0$ implies no routing protocol. The simulations are performed on a scale-free network with 1000 nodes, with degree distribution $P(k) \propto k^{-\gamma}$ with $2 < \gamma < 3$. Figure 2.2 displays two typical time series of the total number of walkers $\aleph(t)$ in a free-flow and congested states which are obtained by varying the birth/death probability ratio p/μ . In the free-flow state, the total number of particles fluctuates around a stationary value whereas in the congested state, the number of particles constantly increase with time. In case of a regular network, all nodes have same degree and congestion phase appears around $p_c = \mu$ i.e when $p/\mu = 1$. From the figures, we can clearly see that the threshold $p_c = \mu$ holds true only for homogeneous networks (regular lattices) and in the absence of large degree fluctuations (scale-free graph).

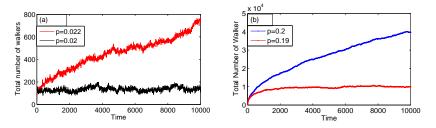


Figure 2.2: Number of Walkers as a function of time on a scale-free network (left) and regular network (K=8)(right) of 1000 nodes, $\mu = 0.2$. Free-flow state occurs at p = 0.02 for scale-free network and p = 0.19 for regular network. Scale-free network gets congested at p = 0.022 and regular network at p = 0.2.

To understand the dynamics at each node when the system is in a free-flow state or a congested state, we look at the local order parameter. When the system is in a free-flow state, $\rho = 0$ and the local order parameter, ρ_i for every node i is also zero. However, when the system is in a congested state, all its nodes may not be congested. We expect each node to fall in one of the following categories when $\eta > 0$:

1. Free nodes with exponentially decreasing distribution of walkers: The number of walkers on free nodes are always less than the

- threshold n^* . Therefore, $w_{freenodes} < n^*$. Additionally, the local order parameter $\rho_i = 0$, otherwise they will eventually get congested.
- 2. Congested nodes with non-normalizable distribution: The number of walkers on congested nodes are always greater than the threshold n^* . Therefore, $w_{congested nodes} > n^*$ Additionally, the local order parameter $\rho_i > 0$ for all the congested nodes.
- 3. Fickle nodes/unstable nodes with a distribution that peaks around $w_i = n_i^* = n^*$: Fickle nodes are trickier than the others. The probability that their queue will be empty decreases exponentially with n^* and disappears only for $n^* \to \infty$

We see the time series and distribution of walkers on a node for each of the above categories in Figure 2.3 and Figure 2.4

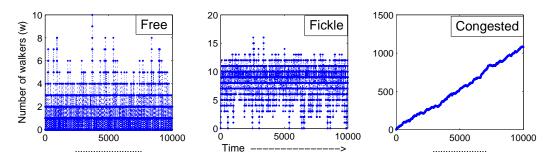


Figure 2.3: w(t) Timeseries of a free node, fickle node and a congested node.

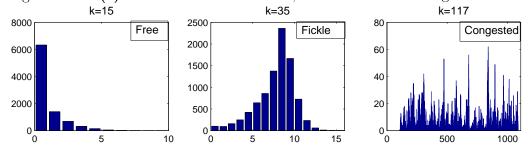


Figure 2.4: **Distribution of walkers** for nodes with k=15 (free node), k=35 (fickle node) and k=117 (congested node) on a scale-free network of 1000 nodes, $\mu = 0.2$, p = 0.07, $\eta = 0.75$ and $\rho = 0.0014$.

A node *i* is congested, if $\rho_i > 0$. As discussed earlier, higher the order parameter, more the number of congested nodes. However, order parameter/congestion parameter doesn't give any information about the time required for the system to enter a congested state. It is known that the local order parameter varies linearly with k [1] Thus, higher degree nodes are

more likely to get congested. Therefore, for every p, there exists a real-valued threshold $k^*(p)$ such that all nodes with $k > k^*$ are congested whereas nodes with $k < k^*$ are not congested. In a heterogeneous network, not all nodes get congested at the same time. The birth probability p, at which a node gets congested depends on its degree, beginning with the highest degree hub. As we can see in Figure 2.5, as we increase p, more number of nodes are congested i.e $\rho_i > 0$. We can also see that the hubs are the ones to get congested first

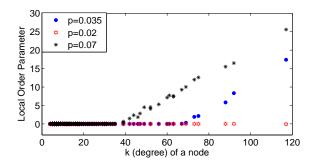


Figure 2.5: Local Order Parameter ρ_{k_i} for birth probabilities p = 0.02, p = 0.035 and p = 0.07 on a scale-free network of 1000 nodes; $\mu = 0.2$, $\eta = 0.0$.

Next, we introduce the rejection probability $\bar{\eta}(w_i) = \eta \theta(w_i - n^*)$. The effect of introduction of rejection probability becomes clear upon observing behavior of the order parameter as a function of birth probability. From Figure 2.6, we can clearly see that increasing rejection probability changes the nature of phase transition from a continuous to a discontinuous one.

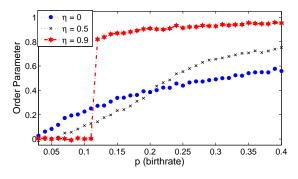


Figure 2.6: Order Parameter as a function of birth probability on a scale-free network of 1000 nodes. $\mu = 0.2$, $n^* = 10$, Continuous to discontinuous phase transition as $\eta = 0$, $\eta = 0.5$ and $\eta = 0.9$.

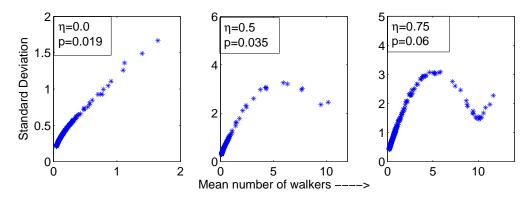


Figure 2.7: Standard Deviation Vs Mean for $\eta = 0$, $\eta = 0.5$ and $\eta = 0.75$.

Macroscopically, rejection probability changes the nature of phase transition. Microscopically, its effect can be seen when the system is in a free-flow state. Barabasi observed that many real-world networks are bound by a universal relation between the flux fluctuations σ and average flux < f > on individual nodes, given by

$$\sigma \sim < f >^{\alpha}$$

where $\alpha=1/2$ or $\alpha=1$ [10]. The universality of this relation has been challenged over the past few years, leading to system-dependent scaling laws [25] [26]. In our system, varying the rejection probability gives us the demonstration of the changing dependence between flux fluctuations and average flux, shown in Figure 2.7.

To summarize, we have analyzed a simple model used to demonstrate congestion phenomena on a scale-free network. We discuss the properties of this model in detail, and examine how various parameters of the model affect the individual nodes as well as the system as a whole.

Chapter 3

Extreme Events

As discussed in the previous chapter, a node experiences congestion when it is so overloaded with walkers/packets that quality of service deteriorates, resulting in queuing delay or packet loss. Extreme event, as the name suggests, is defined as exceedance above a prescribed quantile. Unlike congestion, it is not necessary that extreme events (EE) be related to the node capacity. EE arise from the natural fluctuation in the traffic passing through a node and not due to external constraints, such as capacity of a node. EE in a parameter-free model have been thoroughly studied [2]. The transport model used is random walk on complex networks [20]. Using this simple random walk model, probabilities of occurrence of EE on the nodes are obtained. Contrary to expectation, it is observed that the lower degree nodes are more susceptible to EE than the higher degree ones. The extreme event threshold for a node i, τ_i , is defined as

$$\tau_i = \langle w \rangle_i + q\sigma_i \tag{3.1}$$

where $q \in \Re$, mean number of walkers < w > and standard deviation σ are calculated from the values obtained from the simulations when the system attains a steady state. In order to define EE, it is imperative for the system to be in an equilibrium stationary state. The threshold τ , used to define EE, can be arbitrarily chosen. It was shown in [2] that the extreme event probability scales with the choice of threshold τ or q. As value of τ decreases, the difference in the probability of EE for smaller degree nodes and hubs decreases. As threshold τ decreases, more number of events are qualified as EE, thereby resulting in a higher probability of occurrence of EE. These results hold true not only for scale-free graphs but also for random and small-world networks. To check the robustness of the result, random walk with a constraint on the path chosen by the walker to go from node i to node j is implemented. The walker has to follow the shortest path between the source

and destination. Even with the shortest path routing algorithm, it is found that the smaller degree nodes are more susceptible to EE than the hubs. This result not only holds for a system with constant number of walkers, but also for a system where number of walkers fluctuate uniformly about a mean value.

In all our calculations, we use q=4. We present one more reason to explain the choice of q=4. According to Chebyshev's inequality [27], for any probability distribution, no more than $1/\lambda^2$ values can lie outside λ times the standard deviation from the mean of the distribution. Quantitatively, almost 90 percent of all the values fall within 3σ from the mean of the distribution. As the name suggests, extreme event is supposed to be an event that happens very rarely. Now, we call the values falling beyond 4σ from the mean of the distribution to be EE. This implies that the tail of the distribution must be long enough to make meaningful conclusions from the results obtained. For a scale-free network with 1000 nodes and 9816 walkers, the probability of EE is shown in Figure [3.1](left). Each point in this figure represents an average over all the nodes with same degree. This is called as average probability of EE.

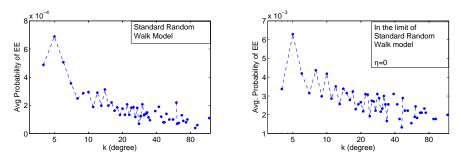


Figure 3.1: **Probability of EE** as a function of degree k, averaged over nodes of same degree on a scale-free network with 1000 nodes, 4908 edges and 10000 walkers.

This model doesn't have any extrinsic parameters apart from q. Hence, it could be called as a parameter-free model. Also, the dynamics is a discrete-time dynamics. Real-world systems are incredibly complex due to correlations and consist of processes that vary on a continuous time-scale. Thus, it is imperative to modify the model by incorporating realistic features. In the following sections, we study EE in a realistic framework on two types of systems: conservative and non-conservative.

3.1 Extreme Events in Conservative Model

In this section, we examine EE probabilities in a model consisting of realistic parameters. The conservative model has fixed number of walkers. Therefore, birth probability and death probability are zero in all the cases to be discussed in this section. Let us consider W walkers randomly distributed over a scale-free network with N=1000 nodes at time t=0. With p=0 and $\mu=0$, we have three more parameters to consider, rejection probability $\bar{\eta}$, threshold queue length n^* and outgoing flux from a node Ω . In this case, we take the threshold queue length, n^* , degree dependent. For the standard random walk problem, the mean number walkers on a node is given by

$$\langle f \rangle = \frac{Wk}{2E} \tag{3.2}$$

where f(w) is the distribution of walkers on a node with degree k. We define $n^* = < f >$. Now, we need a define a way to call an event an extreme event. Therefore, the extreme event threshold τ is defined as

$$\tau = < w > +4\sigma \tag{3.3}$$

where mean, $\langle w \rangle$, and standard deviation, σ , are calculated from the values obtained from simulations, leaving transients. We have W=10000 walkers on a scale-free network with 1000 nodes and 4908 edges.

The number of walkers in the system is conserved. Therefore, we eliminate two parameters viz., birth probability p and death probability μ . We can come closest to the standard random walk by eliminating routing and allowing each walker to make a transition at every time t. We observe in Figure [3.1] (right) that this limiting case shows similar behavior to the standard random walk case.

For a routing-free traffic model on a network, higher degree nodes attract more walkers than the lower degree nodes. The threshold queue length for each node is crucial to route the walkers on the network. In a conservative system, if we use a constant threshold for all the nodes, it might make the results biased. For example, keeping a low threshold might lead to not using the full potential of the higher degree nodes to manage traffic or over-crowding the lower degree nodes. The outcome may also depend on the total number of nodes in the system. Hence, we define $n^* = \langle w \rangle_{srw}$ where $\langle w \rangle_{srw}$ are the mean number of walkers calculated for a standard random walk model with total number of walkers W. The mean and variance for a given node, in the standard random walk model, can be written as

$$< w>_{srw} = \frac{WK}{2E}, \sigma^2 = W\frac{K}{2E}(1 - \frac{K}{2E})$$
 (3.4)

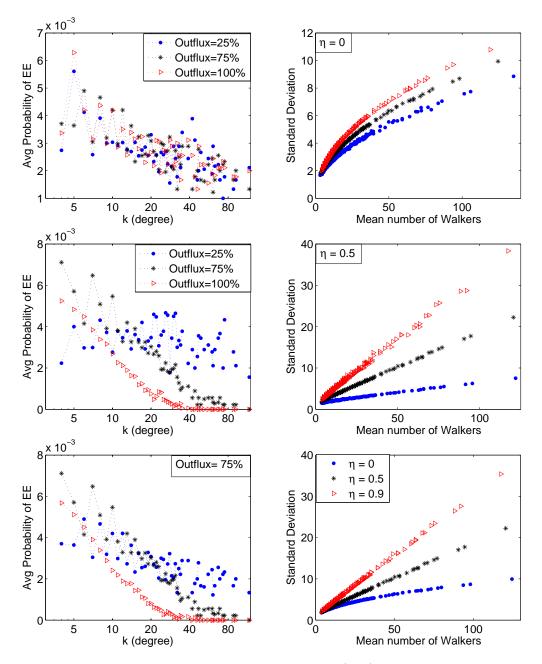


Figure 3.2: Average Probability of EE Vs k (left) and Standard Deviation Vs Mean (right), on a scale-free network with 1000 nodes and 10000 walkers, for various combinations of η and outflux Ω .

In the first set of figures [3.2], $\eta=0$ i.e no routing. Here, we see that irrespective of the number of walkers allowed to hop from each node at any time t, the probability of EE is lower for higher degree nodes than for the

lower degree nodes. Upon careful observation of the right-hand side figures, it is clear that the mean number of walkers on nodes do not vary with outflux or the rejection probability. It is also important to note that mean number of walkers in the conservative case is similar to that in the standard random walk case. An example is provided in Figure[3.3], where $\eta=0.5$ and outflux Ω varies from 25% to 100%. A comparison with the standard random walk case with same number of walkers on a similar network is shown for reference. However, introduction of outflux and rejection probability changes the flux fluctuations in the system. Also, these fluctuations are much lesser than those observed in the standard random walk case. It must also be noted that, reducing the outflux results in lesser relative fluctuations in the system, given by the slope of the curve.

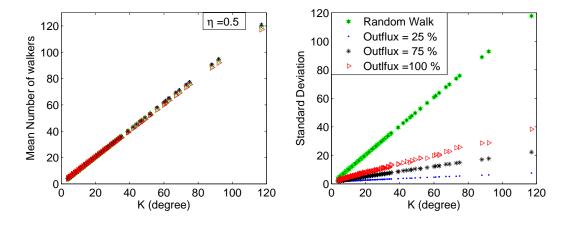


Figure 3.3: Mean < w >, as a function of degree (left) and standard deviation σ , as a function of degree (right) for $\eta = 0.5$ and outflux varying from 25% to 100%. Comparison with the standard random walk is shown in green.

Hence, the threshold for EE decreases and this leads to more EE and hence, increased probability. In the limiting case, all the nodes have similar probability for occurrence for EE. Also, in cases where the fluctuations are very high, no EE are recorded. This could be seen as a continuous-time random walk in some sense, where whether a walker hops onto a neighboring queue depends on the number of walkers in the host's queue and being rejected can be considered as waiting time. Thus, we see a similar trend in occurrence probability of EE in a conservative system with a traffic routing protocol as is seen in the previous work [2].

Thus, we see that the previously obtained results hold true even in the case of a continuous-time random walk, with a routing scheme.

3.2 Extreme events in Non-conservative Model

In the non-conservative model, the number of walkers need not be fixed at any time t. Thus, we have non-zero birth and death probabilities. Hence, it may or may not be a finite system. As the system grows with time, it is immaterial what value of n^* is chosen as a threshold value. It doesn't change the behavior of the system qualitatively. Also, the long-term behavior is determined by the relation between birth and death probability, i.e the value of p/μ . We keep $\mu=0.2$ throughout this study and vary the birth probability p, rejection probability $\bar{\eta}$ and outgoing flux Ω . The non-conservative system can be in a free-flow state or a congested state, depending on the value of these parameters. However, the existence of a stationary state is crucial for defining EE. Hence, we study EE probabilities only in the free-flow state of the system. Similar to the conservative case, the EE threshold τ is defined as

$$\tau = < w > +4\sigma \tag{3.5}$$

where mean, $\langle w \rangle$, and standard deviation, σ , are calculated from the values obtained from the simulations, leaving transients.

The number of walkers in the non-conservative model may fluctuate around a finite value or grow with a constant rate. The threshold queue length $n^* = 10$ and death probability $\mu = 0.2$ are kept constant, as varying these parameters does not change the results qualitatively. As discussed already, existence of a stationary distribution is crucial in calculating the EE probabilities. Thus, the EE probabilities cannot be computed for a system with non-equilibrium steady state. However, they can be calculated for a system with small fluctuations in total number of walkers [2]. In a free-flow state, the system attains a stationary state where it is occupied by a finite (oscillating) number of walkers. Hence, we study EE when the system is in a free-flow state.

We look at this system by varying outflux Ω and rejection probability $\overline{\eta}$, ensuring that it is in a free-flow state. For $\Omega=1$ walker, we plot probability of occurrence for EE as a function of degree, averaged over nodes with same degree. For each value of Ω , η varies from 0 to 1. For each of these cases, we plot mean number of walkers against standard deviation to observe how the changing relationship between mean and standard deviation can affect occurrence of EE. We show simulations for $\Omega=1$ walker, $\Omega=2$ walkers , $\Omega=5$ walkers and $\Omega(t)=10\%$ of walkers. All the results in this section are obtained by averaging over 10 realizations with randomly chosen initial conditions over a scale-free network with 1000 nodes and 4908 edges.

1. Outflux: $\Omega = 1$ walker (Figure 3.4) and Outflux: $\Omega = 2$ walkers (Figure 3.6). We see how average number of walkers, their standard deviation and probability of occurrence of EE vary with η .

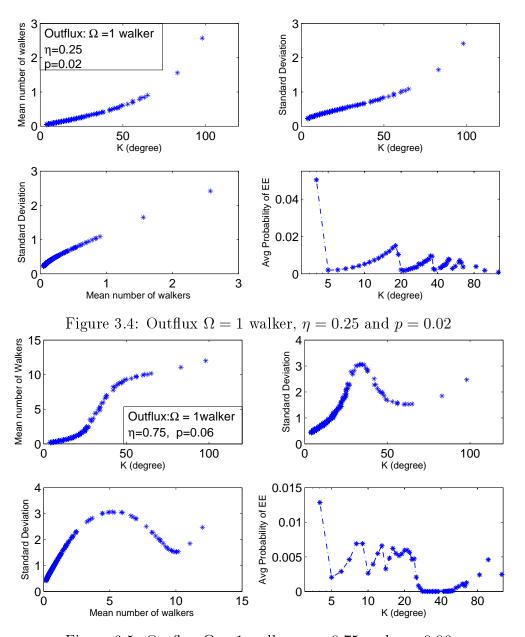


Figure 3.5: Outflux $\Omega = 1$ walker, $\eta = 0.75$ and p = 0.06

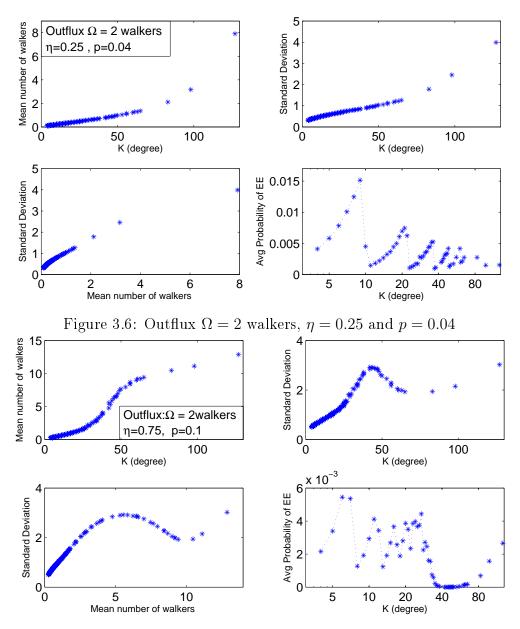


Figure 3.7: Outflux $\Omega = 2$ walkers, $\eta = 0.75$ and p = 0.1

We examine two cases for each value of outflux, in a free-flow state: $\eta=0.25$ and $\eta=0.75$. For both the values of outflux Ω , standard deviation varies linearly with mean number of walkers when $\eta=0.25$ (Figure 3.4, 3.6) whereas the dependence is non-linear for $\eta=0.75$ (Figure 3.5, 3.7). We can notice that, the mean number of walkers for a node of particular degree is higher for higher value of η . When $\eta=0.25$, mean number of walkers is lesser than the outflux Ω for non-hub nodes.

Hence, we can say that all the walkers present on non-hub nodes jump at almost every time-step. In this limit, the system follows a random walk model for non-hubs nodes. Hence, we observe a pattern similar to random walk model for occurrence of EE.

When $\eta=0.75$, mean number of walkers is more than the outflux Ω . Hence, most of the walkers tend to be stagnant at each node, unlike the random walk model. Also, we see a non-linear dependence between flux fluctuations and node degree. We observe that as standard deviation increases, the probability for occurrence of EE decreases and as standard deviation decreases, the probability for occurrence of EE increases.

2. Outflux: $\Omega = 5$ walkers:

We have $\eta=0.25, p=0.13$ (Figure 3.8) and $\eta=0.75, p=0.15$ (Figure 3.9). As we can see, the standard deviation varies linearly with mean number of walkers in both the cases. Here, only hubs have mean number of walkers more than 5. Hence, we can safely say that for non-hubs, almost all the walkers are allowed to hop. We can see that this is similar to the random walk case where all the walkers jump at every time step. Hence, in that limit, we obtain higher probability for occurrence of EE on lower degree nodes as compared to the higher degree nodes.

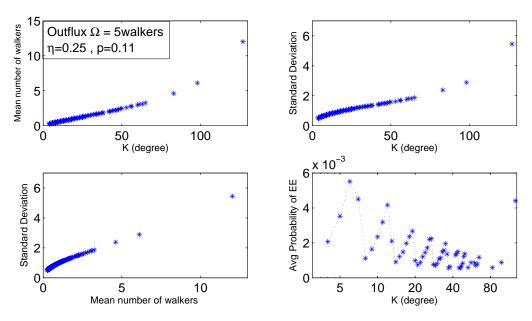


Figure 3.8: Outflux $\Omega = 5$ walkers, $\eta = 0.25$ and p = 0.13

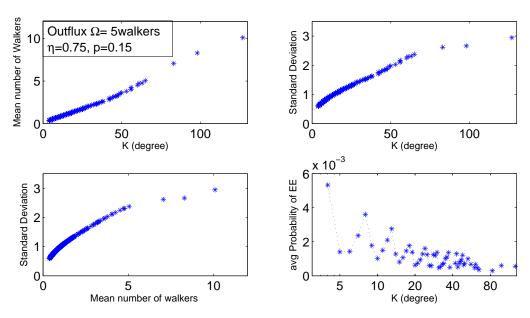


Figure 3.9: Outflux $\Omega = 5$ walkers, $\eta = 0.75$ and p = 0.15

3. Outflux $\Omega = 10\%$ of walkers: An asymptotic distribution of walkers on the network depends on the degree of the node. Hence, it is only reasonable to take a case where outflux is degree dependent. Here, 10% of walkers from each node are allowed to make a transition at every time step. Hence, the nodes with higher occupation will send out more walkers. For this network, the mean and standard deviation show a non-linear dependence for p < 0.12 and behave linearly for values p > 0.12, independent of the rejection probability. The system is always in free-flow or near free-flow state. For p = 0.12, the relation between average flux and fluctuations varies with rejection probability and so does the probability of occurrence of EE. We look at $\eta = 0.25$ (Figure 3.10) and $\eta = 0.75$ (Figure 3.11). We observe a similar behavior as seen for the cases when $\Omega = 1$ walker and $\Omega = 2$ walkers. However, in this case, the outflux is different for each node at every time-step. Hence, we can only say that a linear dependency between mean and standard deviation results in smaller-degree nodes experiencing more EE.

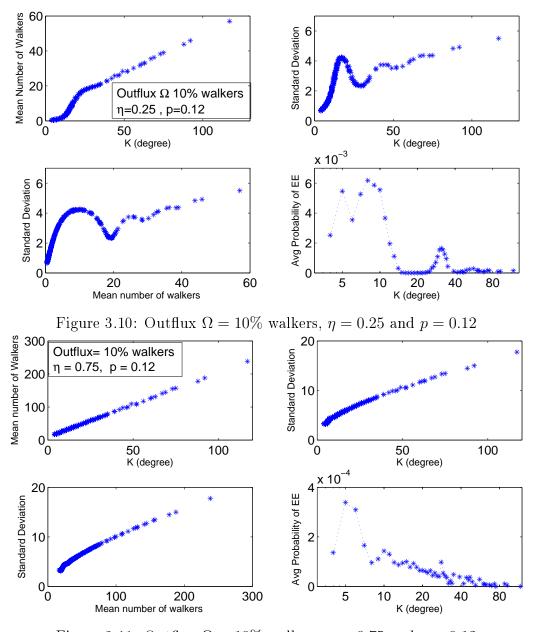


Figure 3.11: Outflux $\Omega = 10\%$ walkers, $\eta = 0.75$ and p = 0.12

Thus, we observe that the when mean and standard deviation have a linear dependence, the smaller degree nodes are more likely to encounter EE than the hubs.

Chapter 4

Phase Transitions in Non-Conservative Model

As we've seen so far, the rejection probability changes the nature of phase transition in our model. Another important parameter of the model is the outflux Ω . In this section, we discuss the effect of Ω on the order parameter and the critical value of the ratio p/μ where the transition changes from a continuous to a discontinuous one. As discussed in chapter 2, higher the order parameter, more is the number of nodes in a congested state. We consider the following four cases for our analysis:

1. Outflux: $\Omega = 1$ walker

2. Outflux: $\Omega = 2$ walkers

3. Outflux: $\Omega = 5$ walkers

4. Outflux: $\Omega = 10\%$ walkers

The first case, $\Omega=1$ walker is nearly similar to model in [1]. At each time t, only a single walker attempts to hop onto its nearest neighbor. For $\eta=0,\ \eta=0.25$ and $\eta=0.5$, we see a continuous transition. As the rejection probability increases, the appearance of congested state is delayed. However, the delay comes at the cost of sudden emergence of the congested state without any warning. Also, as η increases, the transition point shifts to the left, i.e the network gets congested faster, when most of the walkers are rejected as a function of walker-generation probability. This implies that the network can perform its best when η is at neither of the extremes.

As per case 1, influx of walkers at any nodes is unbounded whereas the outflux is restricted to 1 walker. For example, a node with degree k=40

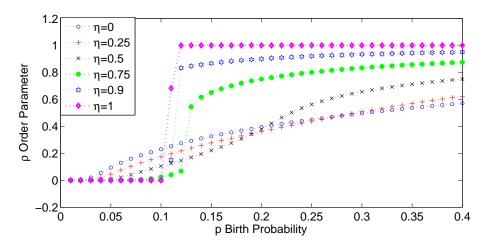


Figure 4.1: **Phase Transition**. Outflux $\Omega = 1$ walker

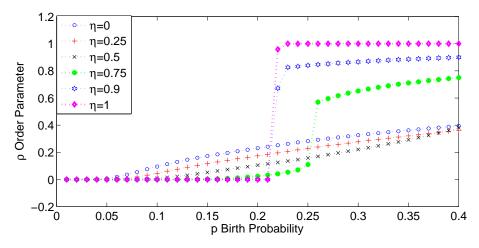


Figure 4.2: **Phase Transition**. Outflux $\Omega = 2$ walkers

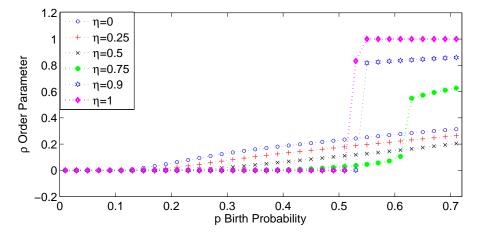


Figure 4.3: **Phase Transition**. Outflux $\Omega=5$ walkers, scale-free network with 1000 nodes, $\mu=0.2,\,n^*=10$ and averaged over 8 realizations in each of the above figures.

may accept 40 walkers at a single time-step while releasing at most 1 walker. Although globally the influx and outflux of walkers may seem to be balanced, it is not the case on an individual node. This biases the system towards accumulating walkers on a particular node. Also, the influx is supported by a time-independent parameter, the birth probability, which keeps generating walkers at a node independent of the queue of walkers at that node. Hence, it only natural for a network to get congested in this framework. To make this model more realistic, we allow at most 2 walkers to hop onto their nearest neighbors from each node at a time t. We consider one more case by allowing at most 5 walkers to hop onto their nearest neighbors from each node at any time t. We observe that as outflux increases, the transition point shifts to the right. Also, the value of the order parameter for every pair (p, η) decreases. Thus, increasing the outflux not only delays the appearance of congested state but also reduces the amount of congestion for a particular value of rejection probability η and ratio of birth to death probability p/μ . Thus, increasing the outflux could make the system more efficient.

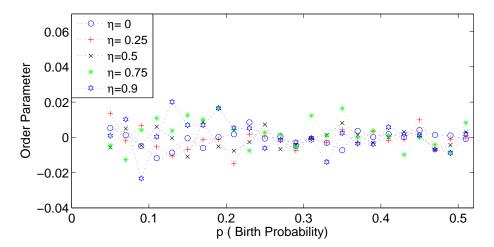


Figure 4.4: **Phase Transition**. Outflux $\Omega = 10\%$ walkers

As we have seen in the previous chapter, average number of walkers on a node increase with its degree. Hence, a node with higher degree is likely to accumulate more walkers than the one with a lower degree. In this light, an outflux independent of degree would result in relatively more outflux from lower degree nodes than higher degree ones, relative to their mean occupation number. Now, we allow at most 10% walkers from each node at any time t to hop onto their nearest neighbor. Even with such a small degree-dependent outflux condition, we notice interesting behavior of the system. The order

parameter lies very close to 0, independent of rejection probability η and ratio of birth to death probability p/μ . Thus, releasing just 10% of walkers makes the network almost congestion-free. Thus, a degree-dependent outflux might help us attain an all-time congestion-free state.

To summarize, we examined four different cases to study congestion phase transitions on a scale-free network. We varied the outflux and studied the phase transition diagrams. We observed that increasing the degree-independent outflux increases the traffic efficiency over the network while making the outflux degree-dependent can potentially lead to a congestion-free state.

Chapter 5

Conclusion

Complex systems are characterized by large number of elements, non-linear interactions and highly-correlated random variables and display a wide array of emergent properties. Networks are increasingly used to model complex systems as finding interconnections between different components of a system can provide valuable information about the system, which is not captured by studying individual components. In this work, we study extreme events and congestion phenomena, two widely studied phenomena on complex systems, in a network framework, using a modified random walk model. These phenomena are critical for the smooth traffic-flow on a network.

We first examine congestion phase transitions on a scale-free network. Typically, when we introduce birth or death processes in the random walk algorithm, two distinct phases are observed: Free-flow state and congested state. The number of walkers over the network grows steadily in the congested state whereas the free-flow state is characterized by fluctuating, yet finite number of walkers over the network in the steady state. This system exhibits, in many cases, continuous phase transitions from the free-flow to the congested state. Previous studies have shown that introduction of rejection probability, a routing parameter, brings about a discontinuous phase transition in the congestion parameter as a function of walker-generating probability. We introduce an additional parameter in the system, the outgoing flux from each node at every time step. For a degree-independent outflux from each node, we observe that increasing the outflux enlarges the free-flow region in the parameter space while a degree-dependent outflux could make the system congestion-free for all values of other parameters. These results could be used to efficiently route information packets on the Internet or vehicular traffic on the roads to maximize traffic capacity and avoid congestion.

We then examine extreme events on complex networks. Previous studies have used a discrete-time simple random walk model to show that smaller de-

gree nodes have a higher chance to encounter extreme events than the hubs. We modify this simple model by introducing a routing protocol and limiting the outflux from each node, thereby making it continuous in nature. The results obtained are in accord with the previous work, thereby expanding the scope of the model. We then introduce birth and death processes, making the system non-conservative, representing a non-equilibrium process. We observe that the system stays in a free-flow regime when the birth and death probabilities balance each other over the network. We study the dynamics of this non-conservative system by varying rejection probability and outflux. For outflux values of 1 walker, 2 walkers and 10% of walkers, we see that at lower values of rejection probability, smaller degree nodes show higher probability for extreme events than hubs, while for outflux of 5 walkers, smaller degree nodes always have higher probability for extreme events, irrespective of the value of rejection probability. This could be attributed to the linear relationship between the mean and the standard deviation for a given node. Thus, when mean and standard deviation have a linear dependence, we hypothesize that the smaller degree nodes are more likely to encounter extreme events than the hubs.

Hubs are more prone to congestion and smaller degree nodes to extreme events. Most of the network designs concentrate on hubs to increase the traffic capacity and efficiency. However, in the light of these new results, we need to upgrade the existing network design principles to incorporate mechanisms to make the network invulnerable to extreme events, which are unavoidable due to their inherent nature.

Further aim of our work is to find the most critical parameters that characterize the model. We also aim at investigating the reason behind non-linear relation between flux fluctuations and average flux using the diffusion properties of random walkers. We also wish to devise methods to predict possibility of extreme events based only on the relationship between the flux fluctuations and the average flux over a node in the network. We aim to build an analytical framework for the results obtained. For our analysis, we have used an undirected network. However, many real-world systems are directed networks, having nodes which are only used for generating or transferring walkers. Thus, we need to modify our model to obtain results for directed networks.

Our simple model provides a basic framework to study extreme events on complex networks, characterized by correlations of real-world traffic flows.

References

- [1] D. De Martino, L. Dall'Asta, G. Bianconi, M. Marsili, Congestion phenomena on complex networks, arXiv preprint arXiv:0808.0584.
- [2] V. Kishore, M. Santhanam, R. Amritkar, Extreme events on complex networks, Physical review letters 106 (18) (2011) 188701.
- [3] J. A. Bondy, U. S. R. Murty, Graph theory with applications, Vol. 290, Macmillan London, 1976.
- [4] , http://www.geeksforgeeks.org/graph-and-its-representations/.
- [5] P. Erdős, A. Rényi, On random graphs, Publicationes Mathematicae Debrecen 6 (1959) 290–297.
- [6] M. Kochen, The small world, Ablex Norwood, NJ, 1989.
- [7] J. Travers, S. Milgram, An experimental study of the small world problem, Sociometry (1969) 425–443.
- [8] D. J. Watts, S. H. Strogatz, Collective dynamics of âĂŸsmallworldâĂŹnetworks, nature 393 (6684) (1998) 440–442.
- [9] R. Albert, H. Jeong, A.-L. Barabási, Internet: Diameter of the worldwide web, Nature 401 (6749) (1999) 130–131.
- [10] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, science 286 (5439) (1999) 509-512.
- [11] V. Balakrishnan, Elements of Nonequilibrium Statistical Mechanics, Ane Books, 2008.
- [12] R. Guimerà, A. Diaz-Guilera, F. Vega-Redondo, A. Cabrales, A. Arenas, Optimal network topologies for local search with congestion, Physical Review Letters 89 (24) (2002) 248701.

- [13] W.-X. Wang, B.-H. Wang, C.-Y. Yin, Y.-B. Xie, T. Zhou, Traffic dynamics based on local routing protocol on a scale-free network, Physical Review E 73 (2) (2006) 026111.
- [14] Z.-X. Wu, G. Peng, W.-M. Wong, K.-H. Yeung, Improved routing strategies for data traffic in scale-free networks, Journal of Statistical Mechanics: theory and experiment 2008 (11) (2008) P11002.
- [15] H. Zhang, Z. Liu, M. Tang, P. M. Hui, An adaptive routing strategy for packet delivery in complex networks, Physics letters A 364 (3) (2007) 177–182.
- [16] B. Danila, Y. Yu, J. A. Marsh, K. E. Bassler, Optimal transport on complex networks, Physical Review E 74 (4) (2006) 046106.
- [17] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, B.-H. Wang, Efficient routing on complex networks, Physical Review E 73 (4) (2006) 046108.
- [18] P. Echenique, J. Gómez-Gardeñes, Y. Moreno, Improved routing strategies for internet traffic delivery, Physical Review E 70 (5) (2004) 056105.
- [19] P. Echenique, J. Gómez-Gardeñes, Y. Moreno, Dynamics of jamming transitions in complex networks, EPL (Europhysics Letters) 71 (2) (2005) 325.
- [20] J. D. Noh, H. Rieger, Random walks on complex networks, Physical review letters 92 (11) (2004) 118701.
- [21] A. E. Motter, Cascade control and defense in complex networks, Physical Review Letters 93 (9) (2004) 098701.
- [22] T. Nishikawa, A. E. Motter, Network synchronization landscape reveals compensatory structures, quantization, and the positive effect of negative interactions, Proceedings of the National Academy of Sciences 107 (23) (2010) 10342–10347.
- [23] A. E. Motter, N. Gulbahce, E. Almaas, A.-L. Barabási, Predicting synthetic rescues in metabolic networks, Molecular Systems Biology 4 (1).
- [24] A. Arenas, A. Díaz-Guilera, R. Guimera, Communication in networks with hierarchical branching, Physical Review Letters 86 (14) (2001) 3196.

- [25] Z. Zhou, Z.-G. Huang, L. Huang, Y.-C. Lai, L. Yang, D.-S. Xue, Universality of flux-fluctuation law in complex dynamical systems, Physical Review E 87 (1) (2013) 012808.
- [26] S. Meloni, J. Gómez-Gardenes, V. Latora, Y. Moreno, Scaling break-down in flow fluctuations on complex networks, Physical review letters 100 (20) (2008) 208701.
- [27] R. Sheldon, et al., A first course in probability, Pearson Education India,