Directed complex networks and their structure and dynamics under link deletion

A thesis
submitted in partial fulfillment of the requirements
of the degree of
Doctor of Philosophy

by

G.A.R.S.R.K Kashyap
20132018

IISER PUNE

I dedicate this thesis to my parents, G. Chandra Shekhar Rao and B. Indira Kumari, for their infinite love and endless support.

Certificate

I certify that the thesis entitled "Directed complex networks and their structure, dynamics and function under link deletion" presented by Mr. G.A.R.S.R.K Kashyap represents his original work which was carried out by him at IISER Pune, under my guidance and supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution.

Prof. G. Ambika
(Supervisor)

Date: April 3, 2019

Declaration

I declare that this written submission represents my idea in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

G.A.R.S.R.K Kashyap
(20132018)

Date: April 3, 2019

Acknowledgements

Abstract

Complex networks, with their ability to model the essential elements of a system of interacting entities, have become a versatile tool with immense scope for application. Network science, as a tool, has found relevance in studying physical, biological and chemical systems, transport systems, social systems, medicine, finance and risk etc. While there are many different aspects to a complete and systematic study of complex networks, one of the more important properties is their robustness to perturbation. In this thesis, we are particularly interested in those types of perturbations that result in loss of connectivity of the network. We focus our study on the effects of loss of links in directed complex networks. Investigating the performance of a network, when it is subject to loss of links, helps to design precautionary measures, introduce fail-safes and implement alternate solutions, thus resulting in improved overall performance. Alternatively, it can also be used to design effective strategies to maximally disrupt the connectivity in a network. With this very concise motivation, we set out to explore the structural and dynamical response of directed networks to failures (random loss of links) and attacks (targeted removal of links).

The organization of the thesis is as follows: In chapter-1, we provide a brief introduction to the field of complex networks, discuss their wide range of applications and ever-increasing relevance. From local to intermediate to global scales, we define the most important and relevant network properties and their corresponding measures. Based on these properties, we discuss some standard models of network generation, their advantages and limitations. We finish with a brief mention of a couple of dynamical processes taking place on networks.

In chapter 2, we introduce methods, based on degree preserving rewiring (DPR), to tune clustering and degree-correlations in random and scale-free directed networks. They provide null-models to investigate the role of the mentioned properties along with their strengths and limitations. For both clustering and degree-correlations, we first discuss how they get redefined for the case of directed networks. As we tune clustering, we qualitatively analyze the results of rewiring, as a function of the respective topological parameters. In both random and scale-free networks, we fix the parameter values so that they have the same average degree, and then study the effect of topology itself on the rewiring process. We rewire the two types of networks, for both assortative and dissortative correlations, and study the role of topology, as well as the effects of the respective topological parameters. To explain the observations in scale-free networks, we explore the role of 1-node correlations for both assortative and dissortative rewiring. In both topologies, we study the behavior of the limiting value of dissortative degree-correlation, as a function of parameters of the networks.

Besides the intended purpose of tuning network properties, the proposed mechanisms are also used as a tool to reveal structural relationships and topological constraints. We take a deeper look at the effects of DPR mechanisms that we have introduced and treat the process of rewiring as exploring the space of second-order maximally random directed graphs. Based on this, we make analytic arguments to show that the DPR mechanisms, designed to tune specific type of correlations, also uniquely affect the clustering coefficients. We provide numerical corroboration and present explanations for the same. Consequent to the explanation, we expect to see changes in sizes of the connected components, specific to the type of correlations being tuned.

In chapter 3, we look at the structural response of directed networks to link deletion. We conduct a systematic and detailed analysis of the robustness of the networks under random and targeted removal of links. We work with a set of network models of random and scale-free type, generated with specific features of clustering and assortativity. Besides random deletion, we define strategies based on global (Edge Betweenness Centrality) and local (Edge Degree) properties and use them to breakdown the networks by targeted removal of links. The robustness of the networks to the sustained loss of links is studied in terms of the sizes of the connected components and the inverse path lengths. The effects of clustering and 2-node degree correlations, on the robustness to attack, are also explored. We provide specific illustrations of our study on three real-world data-sets: protein-protein interaction networks, road-network based on the city of Austin TX and the worldwide airport network.

In chapter 4, we explore the dynamical response to link deletion. We study the effects of random and weighted link deletion on a 1-to-1 transmission process, taking place simultaneously on the network. By characterizing the performance of the transmission process in terms of the average number of successful transmissions and the average transmission time, we analyze the results, first by topology and then by strategy, and also study the role of assortative and dissortative correlations. We provide qualitative arguments to show that the behaviours of probability of successful transmission and normalized transmission times are captured by the fractional size of the strongly connected component (SCC) and the average path-length of the SCC respectively, and also provide numerical evidence. In the context of a specific road-network, derived from road intersections in the city of Barcelona, we study the roles of various process parameters like request-rate, probability of transmission and probability of deletion, and interpret the results in the context of traffic flow. We conclude the chapter by proposing and detailing a method, based on weighted selection of nodes, to partially mitigate the loss of transmission.

In the concluding chapter, we compile all the important results from earlier chapters and present them in a broader perspective.

Publications

The work presented in this thesis has appeared in the following publications/conference presentations.

List of publications

1. Kashyap, G. and Ambika, G. (2019). Loss of transmission on directed networks due to link deletion. arXiv preprint arXiv:1907.04007.

2. Kashyap, G., Bapat, D., Das, D., Gowaikar, R., Amritkar, R.E., Rangarajan, G., Ravindranath, V. and Ambika, G., (2019). Synapse loss and progress of Alzheimer's disease-A network model. *Scientific Reports*, 9(1), p.6555.

3. Kashyap, G. and Ambika, G. (2019). Link deletion in directed complex networks. *Physica A: Statistical Mechanics and its Applications*, 514, 631-643.

4. Kashyap, G. and Ambika, G. (2017). Mechanisms for tuning clustering and degree-correlations in directed networks. *Journal of Complex Networks*, 6(5), 767-787.

List of conference presentations

1. Effects of degree preserving rewiring in directed networks – MPIDS, Dec. 2017, Göttingen, Germany.

2. Link deletion in directed complex networks – Dynamics Days Europe (DDE), May 2017, Szeged, Hungary.

3. Tuning clustering and correlations in directed scale-free networks – Conference on Nonlinear Systems and Dynamics (CNSD), Dec. 2016, Kolkata, India.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

Complex networks, as a field, has captured the interest of the research community in a big way in the last couple of decades. Many real-world systems can be modelled based on the relationship between the underlying topological structures, their interactions and the resulting emergent dynamics. The ability of complex networks to capture the essential elements of such systems of interacting entities and their dynamics makes them an indispensable tool of systemic analysis [2–5]. The versatile nature of this toolkit has enabled applications in a wide range of problems cutting across many branches of science, engineering, commerce, management and sociology. They have found relevance in physical, biological [6–9] and chemical systems [10, 11], transport systems [12–14], social systems [15–17], medicine [18, 19], financial risk analysis [20, 21] etc. Depending on the problem at hand, networks have been used for visualization [22–24], robustness/vulnerability analysis [25–34], community-detection [35–42], link-prediction [43–45], information-spreading [46–49] etc. The computerization of data-acquisition and the consequent availability of large amounts of data have also been important contributing factors. With the focus now on interdisciplinary research, network science stands out as a very natural and generic tool of choice.

Further, in this chapter, we start with a description of some very basic network properties and their corresponding metrics. This is followed by a discussion of standard models of network generations and finally we detail a couple of important processes on networks.

1.2 Properties and Metrics

1.2.1 Degree

In a network with N nodes and m edges, two nodes that are connected by an edge are referred to as neighbours. The **degree** of a node is defined as the number of neighbours of that node and can be calculated as

$$k_i = \sum_{j=1}^{N} A_{ij} \tag{1.1}$$

where A is the adjacency matrix of the network and $A_{ij} = 1$ if nodes i and j are connected and 0 otherwise. The **degree distribution** $p(k)$ is defined for the network and gives the fraction of nodes with degree k. It can also be interpreted as the probability

that a randomly chosen node from that network has a degree k. In this case, an edge between two nodes indicates bidirectional connectivity and therefore A is a symmetric matrix and the network is said to be undirected.

When the edge between i and j is unidirectional, say from i to j, it results in a directed network and the corresponding adjacency matrix is not symmetric. The elements of the adjacency matrix, A_{ij} now indicate connection only from i to j. As a result, each node has an $in-degree$ and an $out-degree$. The $in-degree$ (k^{in}) is the number of edges that arrive to the node and the $out-degree$ (k^{out}) is the number of edges that depart from the node. Accordingly, the networks is characterized by a joint degree distribution $p(k^{in}, k^{out})$ which the probability that a randomly chosen node has an in-degree k^{in} and out-degree k^{out}.

$$k_i^{out} = \sum_{j=1}^{N} A_{ij} \qquad\qquad k_i^{in} = \sum_{j=1}^{N} A_{ij}^T \qquad (1.2)$$

This edge-attribute provides an important classification scheme, bisecting all networks into undirected and directed types. If a network contains both types of edges, it can be represented as a directed network with the undirected edges split into two directed edges pointing in opposite directions. Besides the directionality of the edges, the form of the degree distribution itself provides another important classification method. In an undirected network, when nodes are connected by an edge uniformly randomly, the resulting degree distribution takes a binomial form and in the limit of large network size, it becomes a poisson distribution

$$p(k) = e^{-\overline{k}} \frac{\overline{k}^k}{k!} \qquad (1.3)$$

where $\overline{k} = m/N$ is the average degree of the network. This class of networks are commonly referred to as **random networks** or **Erdos-Renyi (ER) networks** [50, 51], named after the two scientists who made significant contributions to their understanding. However, it has been observed that the degree distributions of most real-world networks deviate significantly from that of random networks. Instead, they follow an asymptotic power-law with the scaling exponent α, in the large N limit. These networks are also popularly known as **scale-free (SF) networks** and their degree distribution takes the form

$$p(k) = Ck^{-\alpha} \qquad (1.4)$$

1.2.2 Assortativity

An important property of any network is the correlation between properties of connected nodes. The extent of (dis)similarity of the properties of connected nodes is given by the **assortativity** of the network [51]. If the property is enumerative (like nationality, race, gender etc.) and takes a finite number of discrete values, then the assortativity is quantified using **modularity**, which is defined as

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \qquad (1.5)$$

where $\delta(a, b)$ is the *kronecker* delta. On the other hand, if the properties are scalar and take continuous values (like age, income etc.), the assortativity is calculated using
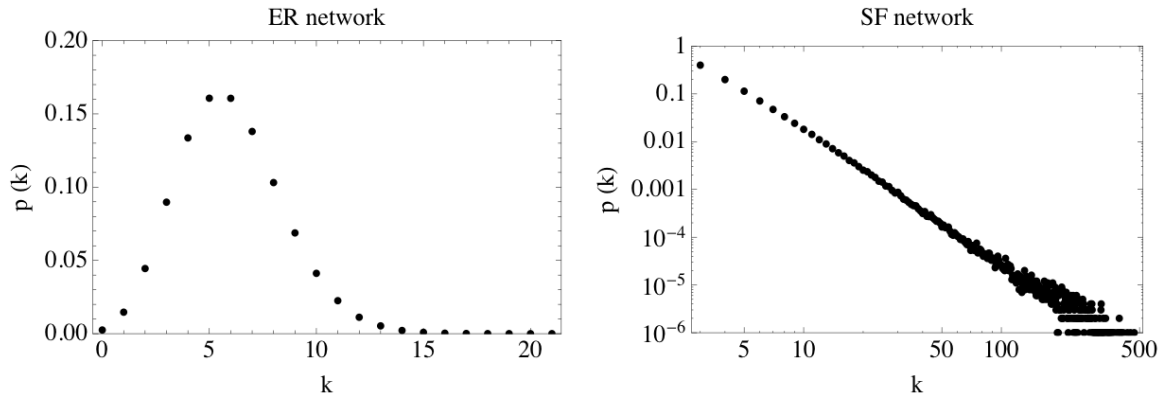
Figure 1.1: Degree distributions of an ER network and SF network for $N = 10^4$ and $m = 6 \times 10^4$

the **assortativity coefficient**, derived from the Pearson's correlation coefficient and takes a value between -1 to 1 [52, 53].

$$r = \frac{\sum_{ij} \left(A_{ij} - k_i k_j / 2m\right) x_i.x_j}{\sum_{ij} \left(k_i \delta_{ij} - k_i k_j / 2m\right) x_i.x_j} \qquad (1.6)$$

A particularly interesting case arises when we calculate the extent to which nodes with (dis)similar degrees connect to each other. These are referred to as **degree correlations** and gain importance because of the interplay between the two structural properties involved.
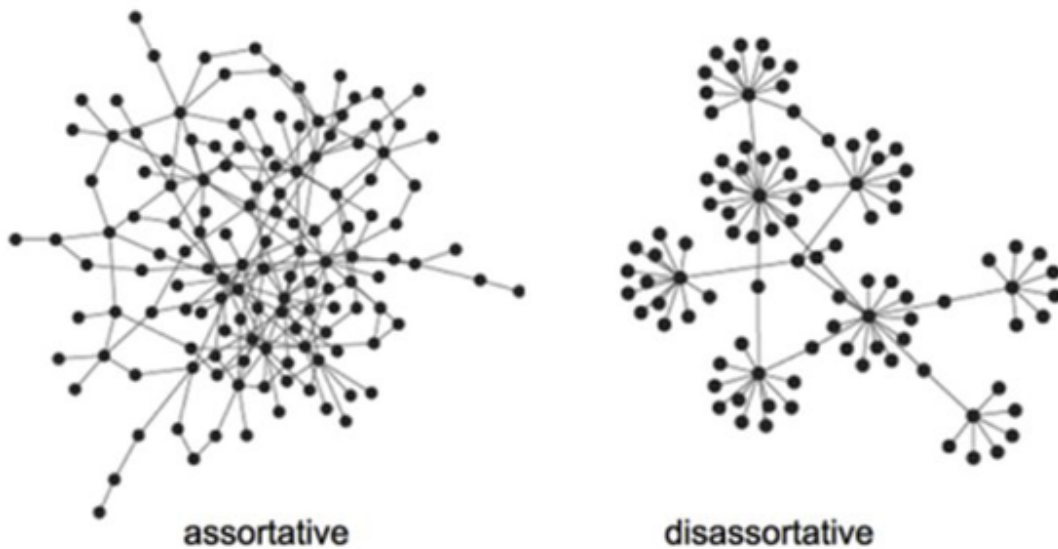
Figure 1.2: Schematic visualization of assortative and dissortative networks. *Source* - Quora.com

$$r = \frac{\sum_{ij} \left(A_{ij} - k_i k_j / 2m\right) k_i.k_j}{\sum_{ij} \left(k_i \delta_{ij} - k_i k_j / 2m\right) k_i.k_j} \qquad (1.7)$$

This pattern of connections is also the basis for classification of networks. When nodes of similar degrees (property) are more likely to connect to each other, the value of r takes a positive value and the networks are said to be **assortative**. Typical examples include social networks, collaboration networks, citation networks etc. A value of $r < 0$

indicates that nodes with dissimilar degrees (property) are more likely to be joined by an edge and such networks are called **dissortative** and examples include internet, World Wide Web (WWW) and most biological networks [52, 53].

1.2.3 Clustering

Clustering is defined as the tendency of nodes to organize into well-knit neighbourhoods or form cliques and is quantified by the clustering coefficients (CC). They play an important role in the spreading of epidemics in large communities and also affect flow in local neighbourhoods.

Given a node i with degree k_i, the **local clustering coefficient** (c_i) of the node is defined as the ratio of number of pairs of connected neighbours of i to the number of pairs of neighbours of the node [51].

$$c_i = \frac{[A^3]_{ii}}{k_i(k_i - 1)} \tag{1.8}$$

where A is the adjacency matrix of the undirected network. Alternatively, it quantifies the extent to which the neighbours of i form closed triplets with i. The **mean clustering coefficient** is defined as the average of local clustering coefficients of all nodes in the network.

$$\overline{C} = \frac{1}{N} \sum_i c_i \tag{1.9}$$

The **global clustering coefficient** is defined as the ratio of number of closed triplets to the number of connected triplets in the network.

$$C = \frac{\sum_{i=1}^{N} [A^3]_{ii}}{\sum_{i,j} [A^2]_{ij} - \sum_{i=1}^{N} [A^2]_{ii}} \tag{1.10}$$

1.2.4 Paths

So far, we have defined local network properties like degree, degree-correlations, clustering etc., that are all associated with a single node/edge or within their immediate neighbourhoods. Now, we move on to another important set of properties, based on the notion of *paths*, whose understanding provides the basis for defining global network properties.

A **path**, in a network, is defined as a sequence of nodes where each consecutive pair of nodes in the sequence is connected by an edge [51]. While in undirected networks, a path from node i to node j implies a path from j to i, in directed network, they constitute two unique paths. The *length* of a path is the number of edges traversed along the path. This may include edges that are visited more than once and are counted separately for each visit. In a simple network, represented by adjacency matrix A, the number of paths of length l from node i to node j is given by the lth power of A.

$$N_{ij}^{(l)} = [A^l]_{ij} \tag{1.11}$$

A path is said to become a **loop** when it terminates at the node where it originated. The number of loops of length l is given by the trace of the lth power of A.

$$L^{(l)} = \sum_{i=1}^{N} [A^l]_{ii} = Tr(A^l) \tag{1.12}$$

Paths that visit each edge in a network exactly once are referred to as **eulerian paths** and those that visit each node exactly once are called **hamiltonian paths**. The study of these paths has had historic significance and continues to play an important role in current-day applications like vehicle-routing, parallel-processing, job-sequencing etc.

It is often the case that two nodes in a network are connected by more than just one unique path. For any two given nodes, the shortest path between is called the **geodesic path** or simply the **shortest path** and the length of this path is called the **geodesic length** (d_{ij}). The geodesic path between any two nodes is not necessarily unique and by convention, if a pair of nodes are not connected by atleast one path, the distance is taken to be ∞. The largest geodesic distance between all pairs of nodes in the network, for which paths exist, is defined as the **diameter** of the network.

$$diameter = max(d_{ij}) \qquad \forall i, j \ni d_{ij} \neq \infty \qquad (1.13)$$

However, the more interesting quantity is the **average path-length** of the network, defined as the mean of distances between all pairs of nodes in the network.

$$APL = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N-1} \sum_{j=1, j \neq i}^{N-1} d_{ij} \qquad (1.14)$$

The above definition presents an obvious mathematical issue when the network is very sparse or there exist isolated nodes ($d_{ij} = \infty$). To work around this problem, the **average inverse path-length** or **efficiency** of the network was defined as the mean of inverse distances between all pairs of nodes in the network. In this case, when there are disconnected pairs of nodes, the inverse distance becomes 0 and the average value remains well-defined.

$$AIPL = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N-1} \sum_{j=1, j \neq i}^{N-1} \frac{1}{d_{ij}} \qquad (1.15)$$

If we calculate the mean distance of all shortest paths associated with node i, the median value of all such distances gives the **characteristic path-length** of the network.

1.2.5 Connected Components

As mentioned earlier, based on the existence of paths, it is possible to define global metrics for a network. The set of all nodes in a network, such that every node is connected to every other node in the set, is called a **connected component**. It can happen that there are multiple distinct connected components in the same network, in which case, the network is said to be **disconnected** and the size of the largest connected component becomes of primary interest. When the size of the largest connected component scales as a function of the network size, it is called the **giant component** of the network [51]. It is a measure of the extent of connectivity in the network and plays an important role during any spreading process.

In a directed network, the set of all nodes, for which there exists a directed path from every node in the set to every other node in the set, is defined as the **strongly connected component (SCC)**. A similar set of nodes, for which there exists an undirected path between all pairs of nodes in the set, is defined as the **weakly connected component (WCC)**. Together, the SCC and WCC give rise to a macroscopic organization referred to as the *bow-tie* structure [54,55]. This is again a sub-class in a broader

set of *core-periphery* structures [55] where the core is dense and strongly-connected while the periphery consists of nodes with sparse connectivity. The study of these connected components provides crucial understanding of the large-scale organization in the networks and is also at the heart of percolation studies on networks.
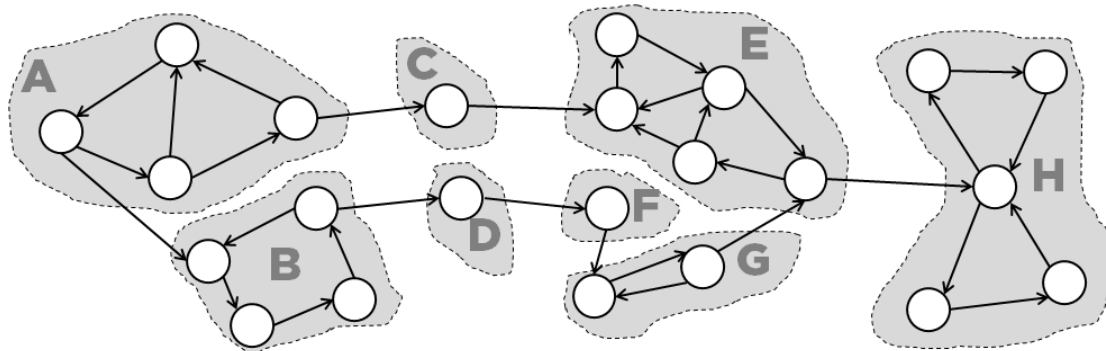
Figure 1.3: Schematic depiction of strongly and weakly connected components in directed network. In this figure, the entire set of nodes constitute the WCC and each highlighted sub-set of nodes represent a SCC. *Source* - Google Images (NOI.PH/training)

1.2.6 Centrality

A centrality measure is used to characterize and rank the most important nodes/edges in a network w.r.t one or more properties. Depending on the definition of *important*, centrality can be defined in many different ways [51].

We first discuss two types of centralities that are defined based on the involvement of a node in the overall cohesiveness of the network. The simplest and most straightforward characterization of the importance of a node is its degree, also known as the **degree centrality**. In directed networks, corresponding to in and out degrees of a node, separate values of centrality may be defined or they could be combined to generate a single value. Despite its apparent simplicity, this measure can capture the influence of a node in its immediate neighbourhood and has found application in citation networks, social networks, collaboration networks etc.

The degree centrality of a node, while measuring the number of neighbours, makes an underlying assumption that any two neighbours are of equal importance. This is not realistic, since each neighbour may have a different degree and therefore varying amounts of influence. Therefore, the **eigenvector centrality** of a node assigns a score proportional to the sum of scores of its neighbours. For example, in the WWW, a web-page that is not important by itself, could gain importance by connecting to few but very popular web-pages. Mathematically, the eigenvector centrality of a node i is calculated as

$$x_i = \frac{1}{\kappa_1} \sum_{j=1}^{N} A_{ij} x_j \tag{1.16}$$

where κ_1 is the largest eigenvalue of the adjacency matrix. From the above equation, it is clear that the centrality of i depends on the centrality of its neighbours, x_j and also on the number of its neighbours via A. Extensions, for the case of directed networks, have led to more accurate modifications, leading to the definitions of Katz centrality, Pagerank etc [51, 56, 57].

The next couple of centrality measures are based on the contribution of a node to the overflow in the network and effectively depend on the geodesic paths in the network. For each node i, the mean of the geodesic path-lengths from i to every other node in the network, is defined as its **closeness centrality** l_i. If d_{ij} is the shortest path-length from i to j, then the closeness centrality of i is given by

$$l_i = \frac{1}{N-1} \sum_{j=1, j \neq i}^{N} d_{ij} \tag{1.17}$$

where we do not consider the influence of i on itself. An interesting property of the closeness centrality is that, in contrast to most other centrality measures, lower values of l_i indicate higher influence. For this reason, in the study of social networks, l_i is calculated using the inverse path-lengths $1/d_{ij}$. Despite its simplicity, the closeness centrality spans a very small dynamics range, because of the logarithmic scaling of average path-length with network size, thereby making it hard to distinguish between nodes with high and low centralities. Also, when $d_{ij} = \infty$, this measure becomes ill-defined, hence making it valid only for strongly connected networks [51].

The last type of centrality, that we will discuss, is the **betweenness centrality** [51, 58], which measures the extent to which a vertex lies on the paths between other vertices. The betweenness centrality of a node gives the fraction of shortest paths, between any pair of nodes in the network, which pass through that node. For node i, if n_{st}^i is the number of shortest paths from s to t that pass through i and n_{st} is the total number of shortest paths from s to t, then the betweenness centrality x_i is calculated as

$$b_i = \sum_{st} \frac{n_{st}^i}{n_{st}} \tag{1.18}$$

Nodes with high betweenness have a strong influence on any spreading processes taking place on the network and consequently, their removal results in maximal disruption. This measure has found extensive application in the areas of social sciences, epidemic spreading, information routing in the internet and WWW etc [51]. The definition of betweenness centrality can also be extended to rank edges.

1.3 Models of network generation

So far, in this chapter, we have looked at properties of networks, from the local to the global scale. But throughout this process, we have taken for granted that the underlying network structure, that resulted in those properties, was given. In this section, we will focus our attention on the underlying structures of these networks and the models used to generate them, based on empirically observed characteristics. In particular, we will discuss the Erdos-Renyi (ER) model, Barabasi-Albert (BA) model and the watts-Strogatz (WS) model. These models attempt to capture the properties like scale-free nature, high-clustering, finite degree-correlation and small average path-length, that are commonly observed in real-world networks.

1.3.1 Erdös-Rényi (ER) model

The Erdös-Rényi model generates a random network, $G(N, p)$, with N disconnected nodes initially [50]. Then, every pair of nodes is connected with a probability p. In the infinite size limit, the resulting network has a poisson degree distribution, characterized by a fixed average degree ($\approx pN$). Alternatively, if $\bar{k}$ is a desired average degree, then
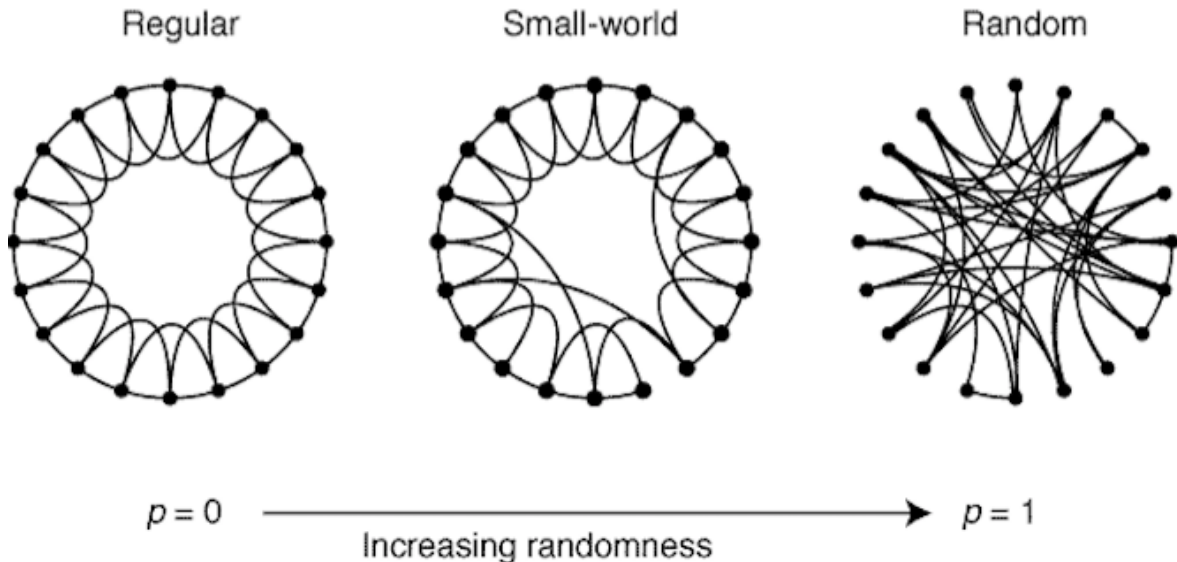
Figure 1.4: A schematic representation of the Watts-Strogatz model. Image taken from [1]

kN edges can be placed among N nodes uniformly randomly. This gives a network $G(N, kN)$ with a fixed number of edges kN. As $N \to \infty$, the mean clustering coefficient, global clustering coefficient and assortativity coefficient go to 0 and the average path-length scales logarithmically with the network size.

However, most real-world networks are found to exhibit an asymptotic power-law degree distribution and finite clustering and/or degree correlations. Therefore, networks generated using this model do not accurately represent observed networks. But being the first model to be studied, this model holds historical significance. Also, being a class of maximally random networks, their simplicity allows for calculation of many exact analytical results. They also act as a good null-model to compare against other empirically observed properties.

1.3.2 Watts-Strogatz (WS) model

As mentioned earlier, the networks generated by the ER model exhibit negligible amount of clustering but still capture the small-world nature observed in many real-world networks. In an attempt to generate networks with both high clustering and small average path-length, Watts and Strogatz put forward the Watts-Strogatz (WS) model [1].

The construction process starts with a k-regular ring lattice with N nodes (fig-1.4). For every node, each associated edge is rewired to a randomly chosen, with a rewiring probability p_r, while avoiding multi-edges and self-loops. For $p_r = 1$, the rewired network is completely random. Initially, the ring lattice has very high clustering and high average path-length. But the rewiring process introduces long-range connections which lead to an exponential decrease in path-length while the global clustering still remains relatively high. Thus, for an optimal value of p_r, the process generates networks with high clustering and small average path-lengths, typically referred to as small-world networks. While this model captures the clustering and average path-length, its degree distribution is quite unrealistic and assortativity goes from exactly 0 (for $p_r = 0$) to zero on average (for $p_r = 1$).

1.3.3 Barabasi-Albert (BA) model

The Barabasi-Albert model is designed to explain the occurrence of a heavy-tail in the degree distribution, as observed in many real-world networks. The presence of these nodes, with disproportionately high degrees (hubs), holds important consequences for network control, robustness, flows etc. The model is based on the mechanisms of growth and preferential attachment, both of which are crucial to real-world networks.

In this model [59], the growth-process starts with a small number of connected nodes (n_0). At every successive time-step, a single node is added to the network until a predetermined size N is reached. The incoming node connects with $m < n_0$ existing nodes with a probability proportional to their respective degrees. As $N \to \infty$, the process generates a power-law degree distribution with scaling exponent equal to 3. The presence of hubs results in a very small average path-length, that scales with the network size as $ln(N)/ln(ln(N))$. The assortativity coefficient is again 0 and the clustering goes to 0 asymptotically, although slower than in ER networks. Despite its partial shortcomings, this model has gained extreme importance and many current and more accurate models of real-world networks use the basic features of this model.

While the ER model, WS model and BA model capture essential properties like degree distribution and clustering, they are by no means representative of the vast number and types of generative methods. While these methods laid out the foundation, many more models have been built upon similar principles. Non-linear preferential attachment models, vertex-copying models, random-walk models etc. have essentially broadened the usage of preferential attachment along with growth. Similarly, static methods have also been developed, like exponential random graph models, stochastic block models, latent space models models etc. However, an interesting set of methods have been developed using the idea of dynamic edge distributions. These methods, also popularly known as rewiring methods, are of particular interest in this work. We develop such methods to preserve degree distribution, while adjusting the values of properties like correlations and clustering.

1.4 Processes on network

In this section, we are take quick overview of two kinds of processes that take place on complex networks. This aspect of network science is focused on understanding the interrelationship between the structure and function of complex real-world networks. Although there are many different types of processes observed to be taking place on networks, like virus-spreading in a computer-network, rumour-spreading in a community, neuronal firing in the human-brain, synchronization of dynamical systems on a network and many more, we will discuss only two such processes in further detail.

1.4.1 Network Resilience and Percolation

The first process of interest is the problem of percolation, which was first proposed in the 1950s and immediate application to the problem of disease-spreading [60, 61]. For the same reason, it gained popularity when real-world data-sets became available and relevant [62]. A **percolation process** is one in which nodes or edges in a network are randomly labelled as *occupied* or *unoccupied* and the properties of the resulting patterns of occupancy of nodes/edges are studied. When the entity of interest is a node, we have *site* percolation and if it an edge, it is called *bond* percolation. Although the early motivation for percolation studies was disease-spreading, we limit this discussion

to the context of network resilience.

Network resilience can be measured in different ways but a simple indicator is the behaviour of nodes in the largest connected component (giant component) of the network. For example, if we consider a hypothetical communication network, in which a path connecting two nodes means that the nodes can communicate with each other, then those nodes that belong to the giant component can communicate with a large fraction of nodes in the network, and those in small components can only communicate with a small number of nodes. Following initial numerical works on subsets of the WWW [25, 27, 54, 63], it became quite clear that the problem of random loss of nodes could be mapped to the problem of site percolation on the network. Nodes are randomly occupied (working) or unoccupied (not working), and the number of remaining nodes that can communicate with each other is exactly the size of the giant component in the corresponding percolation model.

The study of percolation on networks, generated by the configuration model, has produced many analytic results. Cohen et. al [27] used a configuration model for a general degree distribution $p(k)$, in which a randomly chosen fraction q of nodes were occupied, and studied the behaviour of the giant component. They found that in the case of power-law degree distributions ($p(k) \sim k^{-\alpha}$), for $\alpha \leq 3$, the critical value q_c at which the giant component disappears is zero. This implies that such networks, with a diverging second moment of $p(k)$ always have a giant component. Callaway et. al [63] took this model further by introducing a probability of failure, that was proportional to the degree of the nodes. This made it possible to study the effects of targeted attacks on nodes instead of just random failures. It was found that networks with power-law degree distribution were particularly susceptible to this type of attacks [64, 65]. These works were extended by Moreno and Vasquez [66] to the case of networks with degree-correlations, where they have shown that the critical fraction q_c is zero even for networks with finite second moments, if they are assortative.

Besides random failure and targeted attacks on nodes, many more works have dealt with modifications or extensions of these problems. For example, this theory has been extended to the case of directed networks as also to study the problem of bond percolation, which mimics the random/targeted loss of edges in the network. Another closely related problem that has received a lot of attention is the cascading failure in networks [29, 67–70]. In this problem, each node has a *load-capacity* and is involved in distributing load across the network. If the load on the node exceeds its capacity, then the node fails and the load is redistributed to its neighbours, causing them to possibly get overloaded. This has profound implications in the context of power-grids.

1.4.2 Spreading processes on networks

While there are many types of spreading processes that can occur on networks, depending on the type of application, like information, computer-viruses, rumors etc [71], one of the more important and well-studied is that of epidemic spreading.

The study of epidemic spreading is based on compartmental models, where the population of interest is divided into mutually exclusive subsets based on their disease state [72]. Although many different compartmental models have been developed, one model in particular captures the most basic and fundamental aspects of common infections/diseases, namely the SIR model. In this model, each individual (node) can be in either *susceptible* (S), *infected* (I) or *recovered* (R) state. A susceptible individual is healthy but gets infected with some probability of infection, if exposed to an infected individual. An infected individual stays infected for some time, before changing

to a recovered/removed state with some probability of recovery. Once an individual recovers, the state remains constant and cannot become susceptible again.

The long-term evolution of this model reflects the contrasting dynamics w.r.t the other compartmental models. In the SIR model, the recovered population does not become susceptible again and therefore the disease eventually dies out. In this case, the quantity of interest is the fraction of population that is infected after the susceptible population completely dies down. The probability that the initial fraction of susceptible individuals lead to an epidemic is proportional to the ratio of probability of infection to probability of recovery. The critical value of this ratio, below which the probability of an epidemic is zero and above which there is a finite probability of an epidemic, is called the *epidemic transition*. In all compartmental models, there is an implicit assumption of homogeneous mixing meaning that an infected individual can interact with any randomly chosen individual in the entire population and potentially infect them [72, 73]. While this makes it possible to write down a system of analytically solvable differential equations, it also overlooks or averages over all local details and is very unrealistic. But, over the years, these models have become the starting points for more advanced models that include details like the underlying network structure [74], non-zero correlation of interactions between individuals, community-structure, age-classes, spatial-data etc.

In the limit of infinite size, the epidemic threshold is analogous to the critical fraction q_c and the fraction of infected population to the size of the giant component while the homogeneous mixing hypothesis is equivalent to the mean-field treatment. The SIR model on a network, with fixed probabilities of recovery and infection, can be exactly mapped to the problem of bond percolation on the same network. Based on this mapping, many results on the size of the epidemic have been derived in various networks with arbitrary degree distributions. In the absence of degree-correlations and for fixed average degree $< k >$, the epidemic threshold can be derived as $< k > / < k^2 >$, below which the there is no epidemic and over which the disease affects a finite size of the population. In networks with a diverging second moment ($< k^2 > \rightarrow \infty$), the threshold always has a finite value and the disease persists but in networks with fluctuating second moment (scale-free with exponent between 2 and 3), the disease completely vanishes as $N \rightarrow \infty$. Many more important and interesting results have been derived for more advanced models and also include more realistic details like degree-correlations, clustering, memory-based infection etc [71, 72, 75, 76]. Complimentary to this approach, methods based on generating functions have also been studied to develop exact solutions for the SIR model. Another important set of results that has emerged is the study of immunization techniques [77, 78], and their effectiveness on networks with various topologies and combinations of properties, which can be applied to develop preventive and curative strategies.

Before we conclude this chapter, we take a quick look at some of the other important processes that have been studied on complex networks. Considerable efforts have been devoted to understanding the process of diffusion on networks [79–82]. Watts and Strogatz studied models of opinion formation and voter models as cellular automaton processes on networks [1, 83]. The role of network topology on the progression of iterated games has also generated considerable interest [1, 84, 85]. But perhaps the largest body of literature delves into the emergent behaviour of dynamical systems on networks [86–89].

1.5 Summary

In this chapter, we have briefly touched upon complex networks, their characteristics, the processes taking place on networks and their applications. However, we have mostly limited our discussion to the case of undirected networks, which have been extensively studied and for which vast amounts of literature is already available. In contrast, the focus on the current thesis will be on directed networks. This is mainly motivated by their abundant application to important real-world systems and the obvious gaps in existing literature. Since most properties do not automatically extend to the case of directed networks, we redefine them as and when relevant. In the chapters ahead, we present the results of a study on degree preserving rewiring methods in directed networks, the effects of random and targeted link deletion on the structural properties and the changing dynamics of a transmission process occurring on the networks.

Chapter 2

Degree Preserving Rewiring in directed networks

2.1 Introduction

Technological advancements have made it possible to amass vast amounts of data from many real-word systems and this is redefining the way we approach complex networks. Analysis of such data-sets has opened up access to a host of properties associated with the respective networks and has made it clear that an understanding of such properties, their origins and relationships is of utmost importance in order to fully understand the underlying systems. Although the origins of these properties are not well understood, we are nevertheless interested in exploring their relationships and also their influence on the overall structure and functioning of the network.

In this regard, considerable work has been done on generating model networks, using both growth [59, 90–96] and static [1, 50, 97, 98] methods. These models have also been modified to incorporate further properties like 2-node degree-correlations and clustering [52, 53, 99–107]. However, the vast majority of this literature is limited to the case of undirected networks and cannot be broadened to incorporate the case of directed networks [108–112]. Most often, directed networks have been studied by transforming them into undirected networks.

The contents of this chapter include our results from the study of network rewiring mechanisms on directed networks [113]. Two important properties, namely *clustering* and *degree-correlations*, are tuned using these mechanisms. After generating directed ER and SF networks using existing models, we tune the properties of interest in these networks using our proposed Degree Preserving Rewiring (DPR) mechanisms. By isolating the role of the degree distribution and the intrinsic structural properties, these mechanisms allow us to focus on the unique role played by such properties [112, 114–118]. We study each mechanism on two network topologies, the resulting effects and cross-effects and compare their performances.

2.2 Clustering in directed networks

In complex networks, clustering is defined as the tendency for node to organize into well-connected neighbourhoods. it is quantified by any of the three types of clustering coefficients, as discussed in sec-1.2.3. In this chapter, we prefer to use the mean clustering coefficient (MCC), rather than local clustering coefficients (LCC) of the network. This is because it produces a single number that is representative of the average local

connectivity in the network. This makes it possible to effectively and easily compare results from different networks, without bypassing any important information.
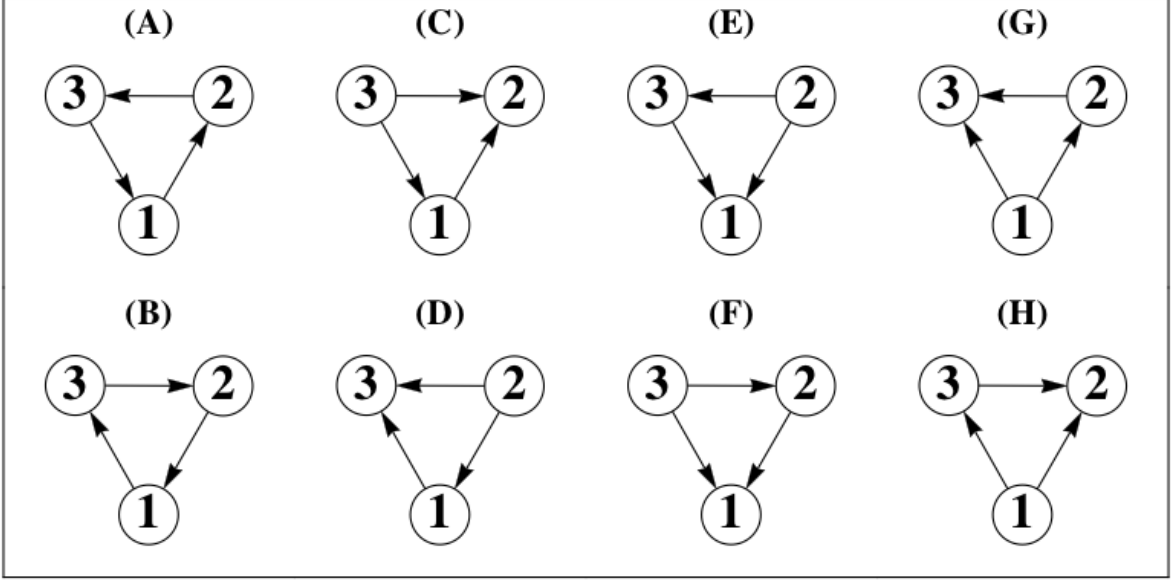
Figure 2.1: The four types of closed triplets that are formed in directed networks: Cycles - (A,B), Mids - (C,D), Intri - (E,F) and Outri - (G,H). In each case, the triplet of interest is shown w.r.t node 1.

In directed networks, owing to the directed nature of edges, there are four possible types of simple closed triplets [108]. The four types of closed triplets, namely, Cycles, Middleman-triangles (Mids), In-triangles (Intri) and Out-triangles (Outri) are shown in fig-2.1.

Consider a network of size N. For any node i in the network, let d_i^{out}, d_i^{in}, $d_i^{\leftrightarrow}$ and c_i represent the out-degree, in-degree, bidirectional edges and LCC of i respectively. If this network is denoted by an adjacency matrix A where $A_{ij} = 1/0$ indicates presence/absence of an edge from node i to node j, then the MCC for each type of triplet can be calculated as follows:

$$\overline{C}_{cyc} = \frac{1}{N} \sum_{i=1}^{N} c_i^{cyc} = \frac{1}{N} \sum_{i=1}^{N} \frac{(AAA)_{ii}}{d_i^{in} d_i^{out} - d_i^{\leftrightarrow}} \tag{2.1}$$

$$\overline{C}_{mid} = \frac{1}{N} \sum_{i=1}^{N} c_i^{mid} = \frac{1}{N} \sum_{i=1}^{N} \frac{(AA^T A)_{ii}}{d_i^{in} d_i^{out} - d_i^{\leftrightarrow}} \tag{2.2}$$

$$\overline{C}_{int} = \frac{1}{N} \sum_{i=1}^{N} c_i^{int} = \frac{1}{N} \sum_{i=1}^{N} \frac{(A^T AA)_{ii}}{d_i^{in} (d_i^{in} - 1)} \tag{2.3}$$

$$\overline{C}_{out} = \frac{1}{N} \sum_{i=1}^{N} c_i^{out} = \frac{1}{N} \sum_{i=1}^{N} \frac{(AAA^T)_{ii}}{d_i^{out} (d_i^{out} - 1)} \tag{2.4}$$

From the above equations, it is clear that for a node i to have a well-defined c_i with respect to in-triangles and out-triangles, the respective in and out degrees of i must be greater than 1. In order for it to have a well-defined c_i w.r.t cycles or middleman triangles, it is sufficient to have both in-degree and out-degree atleast equal to 1, provided they are not associated with the same neighbour of i.

2.2.1 DPR mechanisms

To tune the amount of clustering in a network, our mechanisms identify a suitable chain of connected nodes. The edges within this set of nodes are reorganized in such a way that a desired type of closed triplet is generated, all the while preserving the degrees of the nodes involved. For all variants of clustering, schematic representations of the rewiring processes are given in fig-2.2 and the corresponding algorithms are given in appendix-A. While the mechanisms do appear very similar to some extent, they have minor but critical differences, which we will discuss in greater detail as we go ahead. In order to focus on these minor variations, we do not attempt to generalize the procedures. In all the different mechanisms, there is some ambiguity associated with the choice of node i, whose neighborhood we seek to modify. There are two main schemes for the choice of i. It can be chosen either randomly or in some ordered fashion, where the ordering may be based on c_i, in-degree of i, out-degree of i etc. While a weighted choice seems intuitively effective, we find that the change in clustering is comparable in either scheme for a given number of rewiring steps. Therefore, for computational efficiency, in all the procedures, we choose i uniformly randomly.
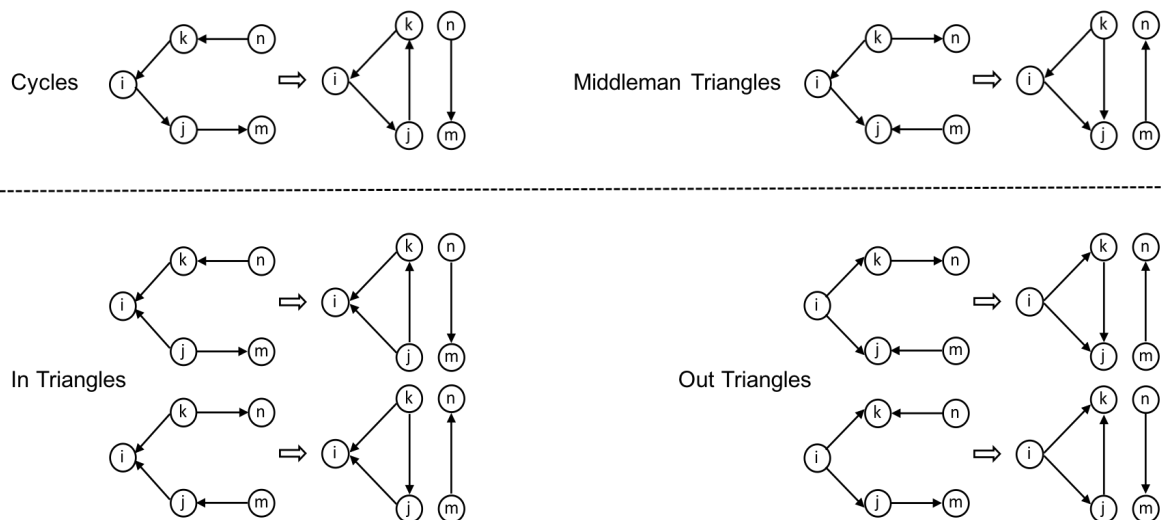
Figure 2.2: An illustration of the degree preserving rewiring mechanisms for increasing the clustering, in directed networks, w.r.t each of the four types of triangles.

The choice of second neighbors of i (in steps 11 and 12 of alg-**??**) is the most important difference between the rewiring schemes for cycles and middleman-triangles. When rewiring for cycles, for the incoming neighbour k of i, we choose one of the incoming neighbor n of k and for the outgoing neighbor j of i, we choose an outgoing neighbor m of j. In the case of middleman-triangles, for the incoming neighbor k of i, we choose an outgoing neighbor n of k and an outgoing neighbor j of i, we choose an incoming neighbor m of j. Besides this crucial difference and the calculation of the relevant clustering coefficient, the algorithms for the cycles and middleman-triangles are exactly the same. For each type of triangle involved, there is a unique selection of neighbours. For both cycles and middleman-triangles, a successful rewiring step can generate only one triangle w.r.t i.

The algorithm for tuning clustering w.r.t In-Triangles is given in alg-**??**. The same algorithm can be used to explain the rewiring procedure for out-triangles by bringing out the important differences. While tuning for Out-Triangles, in step-7 of alg-**??**, we ensure that $d_i^{out} > 1$ and if so, in step-8, we select two distinct nodes, j and k, that are

outgoing neighbors of i. Also, throughout the procedure, we work with $\overline{C}_{out}^{init}$, $\overline{C}_{out}^{final}$ and $\overline{C}_{out}^{des}$ instead of $\overline{C}_{int}^{init}$, $\overline{C}_{int}^{final}$ and $\overline{C}_{int}^{des}$. An important difference that is immediately obvious, is that the rewiring procedures for intri/outri generate one of two possible in/out-triangles w.r.t i for every successful rewiring step. The exact type of triangle is dependent on the actual choice and availability of second neighbours of i. Also, the in-degree and out-degree of i must be greater than 1 when tuning for in and out triangles respectively. The choice of second neighbors also determines the type of complimentary triangles that are formed. For example, in the schematic for in-triangles (top) in fig-2.2, we have an out-triangle w.r.t j and middleman-triangle w.r.t k, while it is the other way around in the bottom schematic. If j and k do not have the same in and out degrees (which is normally the case), then the LCCs w.r.t the two nodes, in the two cases are different. This is also reflected in the respective MCCs.

In the algorithms, at the end of every rewiring step, a new value of MCC is calculated and compared to the old value, to decide if the new arrangement should be retained or rejected. To make the algorithm more efficient, it would seem that calculation of the LCCs of the affected nodes, instead of the MCC, should be sufficient as the rewiring process affects only a local change. This is, however, not possible for two main reasons. First, besides the nodes constituting the triangle of interest, the second and possibly third neighbors of i may also be affected. Second, we find that there is no unique way to identify this set of affected nodes. It is dependent on the number of shared neighbors of different pairs of nodes in the neighbourhood of i. The entire argument narrows down to a trade-off between recalculating all the LCCs or first identifying the affected nodes and calculating only their LCCs. We go with recalculating all the LCCs.

Before we move into the results of rewiring, we will quickly discuss the complexity of the algorithms. The evaluation of MCC is the time-dominant step in the entire algorithm. This calculation is crucial to the performance of the algorithm because the algorithm does not ensure an monotonic change in clustering. The rejection of a particular rewiring is dependent on this step. The calculation happens in $O(N)$. We convert the input edge-list to an adjacency list at the very beginning, to make the calculation of MCC more efficient. This step has a complexity $O(M)$, where M is the number of edges in the network. Having the adjacency list is an added advantage because random-sampling from the lists of neighbors can now be completed in constant time. If, in the worst case, we do not reach the desired value of clustering and therefore have to go through the fixed number of rewiring steps, the number of times MCC is evaluated and the number of sampling operations and comparisons is dependent on the fraction of rewiring steps in which the conditionals are satisfied. In the limiting case, when the number of rewirings is much greater than the network size N, each of the conditionals are satisfied a fixed number of times (equal to an integral multiple of the network size). Overall, we find that the algorithm has a cumulative time-complexity of the order of $O(N^2 + M)$. Between the algorithms for in/out-triangles and cycles/middleman triangles, the actual running times differ only by an additive term.

2.2.2 Role of topology

We test the performance of the proposed mechanisms on directed ER and SF networks and also their effectiveness in generating the desired properties. Directed random networks are generated using the Erdos-Renyi model [50] and the model by Bollobas et al. [93] is used to generate directed SF networks. The rewiring mechanisms are first tested for fixed parameter values in both topologies. This is followed by further

analysis to understand the role of changing parameter values. In the model for SF networks, the link-density ($\bar{k}$) can be parameterized by β. In the limit of infinite size, β and $\bar{k}$ are related as $\bar{k} = 1/(1-\beta)$. Knowledge of his relationship allows us to compare the performance of ER and SF networks having approximately the same link-density.

Before we go any further, we present a brief discussion of the model, for the generation of SF networks, as proposed in [93]. We also derive the relationship between $\bar{k}$ and β that we stated earlier. This generative model is based on a growth mechanism combined with preferential attachment. It is controlled by five non-negative real-valued parameters $\alpha, \beta, \gamma, \delta^{in}$ and δ^{out}, and $\alpha + \beta + \gamma = 1$. Different properties of the resulting network can be manipulated using one or more combinations of parameter values. For example, an appropriate choice of α and δ^{in} can fix the value of the in-exponent (X^{IN}), and the value of the out-exponent (X^{OUT}) can be fixed by choosing γ and δ^{out}. Similarly, the combination of α and γ controls the number of nodes in the network at any given time-step and the density of links in the network can be changed using β. From sec-3 in [93], assuming $t_0 = n_0 = 0$, the average number of nodes in the network at a given time t is approximated by $(\alpha + \gamma)t$. A single link is added to the network at every time-step and therefore, the total number of links in the network at time t is t. Using this information, the link-density at time t can be estimated by

$$\bar{k} = \frac{t}{(\alpha + \gamma)t} = \frac{1}{(1 - \beta)} \tag{2.5}$$

Since the entire process of generation is completely probabilistic, all the results hold true, on average, and in the $t \to \infty$ limit. Simultaneously tuning five parameters can quickly get exhausting and overwhelming, therefore, we fix the values of all parameters except β. The values are decided in such a way that the resulting networks exhibit some desired characteristics. Having the freedom to manipulate β, allows us to compare the results of rewiring in SF networks and ER networks, on the basis of link-density. In our work, we fix $X^{IN} = X^{OUT} = 2.5$. The value of β is decided based on the desired value of link-density. To keep everything simple and easy to understand, we also make $\alpha = \gamma = ((1 - \beta)/2)$. Having fixed the values of α, β and γ, the values of δ^{in} and δ^{out} are calculated, such that $X^{IN} = X^{OUT} = 2.5$. Using these values for the parameters, we implement the construction process as described in [93].

To study the working of our proposed rewiring mechanisms, we start by generating two ensembles, one each for ER and SF networks, and each consisting of 100 networks. Both ensembles consist of networks with 1000 nodes. The SF networks are generated with $\beta = 0.8$ and the average degree of the ER networks is equal to 5. For each variant of clustering, the respective procedure is applied iteratively to both ER and SF networks. The results, taken as ensemble averages, as given in fig-2.3 for ER networks and fig-2.4 for SF networks. We observe, from fig-2.3 and fig-2.4, that for a given network size N having approximately the same link-density and for the same number of rewiring steps, the mechanisms introduce a substantially greater amount clustering in the ER network as compared to the SF network. In the case of $\overline{C}_{out}$, this difference is particularly pronounced. The DPR mechanism introduces twice the amount of change in ER networks than in SF networks. For all MCCs, we observe a rapid initial increase an this is followed by a state where they remain more or less constant or show a very slow increase.

We also observe certain structural side-effects that manifest themselves, independent of topology. When we rewire the network to increase $\overline{C}_{out}$, we also observe an increase in $\overline{C}_{int}$ and $\overline{C}_{mid}$. Looking at the schematic for out-triangles in fig-2.2, we see that the rewiring of edges between the first and second neighbours of i generates
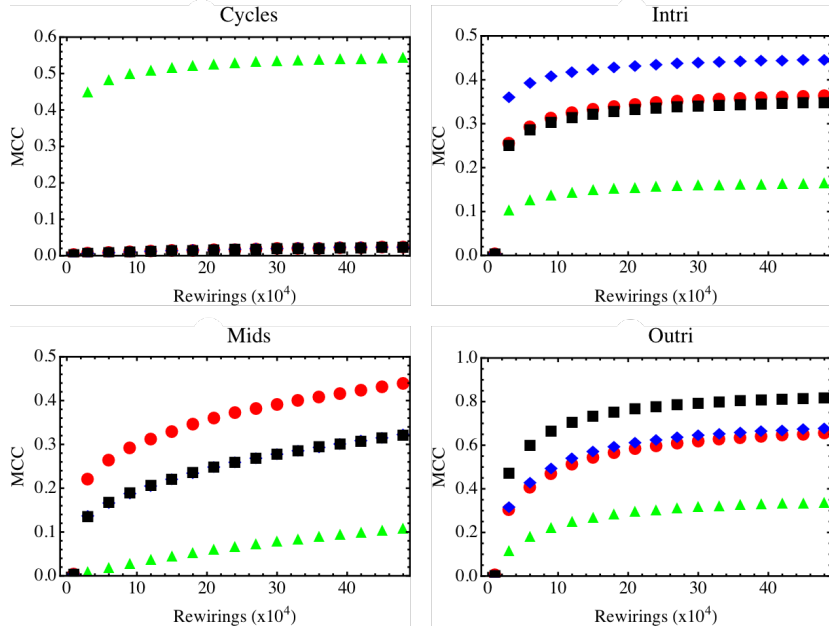
Figure 2.3: Behaviour of MCC values as a function of rewiring steps, in directed ER networks, while being tuned for the various types of clustering. The specific type of clustering, that is being tuned in each case, is indicated by the respective subplot label. $\overline{C}_{cyc}$ - green-triangles, $\overline{C}_{mid}$ - red-circles, $\overline{C}_{int}$ - blue-diamonds and $\overline{C}_{out}$ - black-squares.

two complimentary triangles, one each w.r.t the nearest neighbours of i, which explain the increase in $\overline{C}_{int}$ and $\overline{C}_{mid}$. We observe a similar increase in $\overline{C}_{int}$ and $\overline{C}_{out}$ when rewiring for $\overline{C}_{mid}$ and in $\overline{C}_{mid}$ and $\overline{C}_{out}$ when rewiring for $\overline{C}_{int}$. However, in the case of cycles, we do not observe any such occurrence because the complimentary triangles are also cycles.
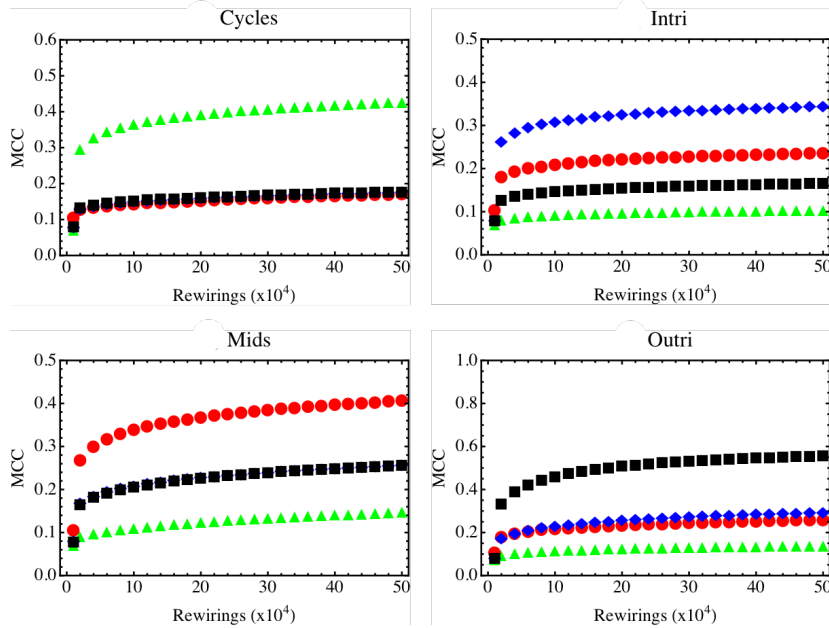
Figure 2.4: Behaviour of MCC values as a function of rewiring steps, in directed SF networks, while being tuned for the various types of clustering. The specific type of clustering, that is being tuned in each case, is indicated by the respective subplot label. $\overline{C}_{cyc}$ - green-triangles, $\overline{C}_{mid}$ - red-circles, $\overline{C}_{int}$ - blue-diamonds and $\overline{C}_{out}$ - black-squares.

2.2.3   Role of topological parameters
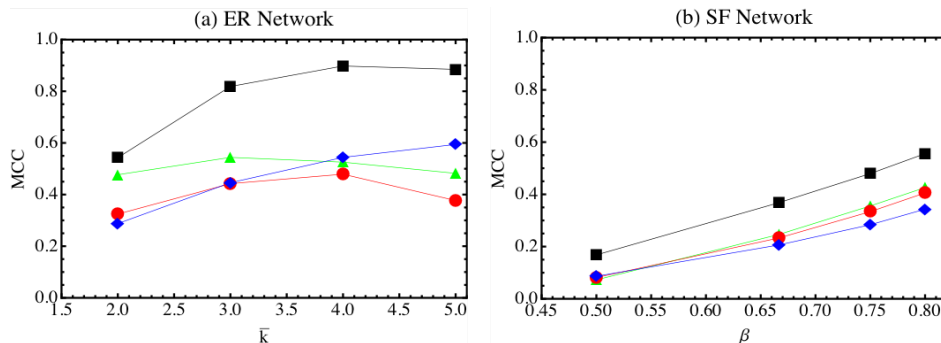
Figure 2.5: Behaviour of MCC values as a function of the corresponding topological parameter values in (a) ER networks and (b) SF networks. In each case, the results are presented for 10^6 rewiring steps. For plot-legends, refer to the caption in fig-2.3.

In SF networks (fig-2.5), we observe that for different MCCs, the qualitative behaviour remains the same while the quantitative behaviours vary as the link-density is increased. For the same number of rewiring steps, different MCCs reach different values. Further studies are required to confirm if this is a true structural limitation. While $\overline{C}_{mid}$ and $\overline{C}_{cyc}$ go as high as 0.4 for $\beta = 0.8$, $\overline{C}_{out}$ reaches a maximum value of 0.6 and $\overline{C}_{int}$ approaches 0.3. For all four types of MCCs, the qualitative behaviour remains the same as β is increased. In ER networks, the smooth variation of MCCs, as seen in SF networks, is evidently absent. The values of $\overline{C}_{cyc}$ and $\overline{C}_{mid}$ are clustered close together for $\bar{k} = 2, 3, 4, 5$ while the values for $\overline{C}_{int}$ and $\overline{C}_{out}$ show a consistent variation. Overall, a comparison of results on both ER and SF networks reveals that the topology does affect the DPR mechanisms to some extent in both cases.

2.3   Degree-correlations in directed networks

In a network with a given degree distribution, the tendency of similar/dissimilar nodes to form connections among themselves is measured by degree-correlations. Whether the role of such correlations is causal or consequential is not completely understood but their effect on the network structure and function is well-accepted. While correlations can be studied with respect to any enumerative (color, race) or scalar (income, age) property, the study of correlations between node-degrees takes priority because of the interplay between the two structural properties involved.
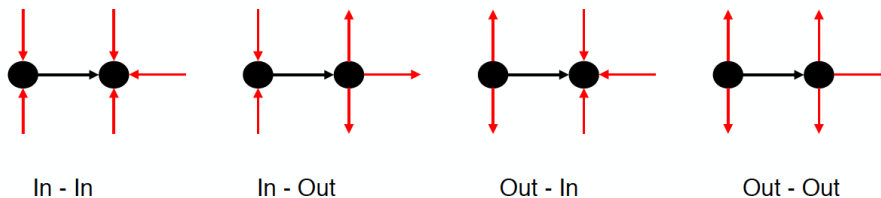
Figure 2.6: The four different types of 2-node degree-correlations in directed networks: $In - In$, $In - Out$, $Out - In$ and $Out - Out$.

In directed networks, we can define four different types of 2-node degree-correlations. Due to the distinct in-degree and out-degree associated with each node, we have $In - In$,

$In - Out$, $Out - In$ and $Out - Out$ degree correlations. Going ahead, we use $p - q$ to represent a general type of correlation, where $p, q \in \{In, Out\}$ and p and q are associated with the source and target nodes respectively. Traditionally, the Pearson correlation coefficient, r_q^p, has been the metric of choice to quantify correlations [52].

$$r_q^p = \frac{\sum_e k_e^p k_e^q - M^{-1} \sum_e k_e^p \sum_{e'} k_{e'}^q}{\sqrt{\left[\sum_e (k_e^p)^2 - M^{-1} \left(\sum_e k_e^p \right)^2 \right] - \left[\sum_e (k_e^q)^2 - M^{-1} \left(\sum_e k_e^q \right)^2 \right]}} \qquad (2.6)$$

where e runs over the edge-list. But, as was shown in [119–121], in networks with heavy-tailed degree distributions, the value of r_q^p scales with the network size N. In the infinite size limit, it converges to a non-negative value. For this reason, r_q^p is not a good metric to study very large SF networks, especially if they are dissortative. And so we turn to the Spearman's rank correlation (eqn-2.7), to quantify these dependencies.

$$\rho_q^p = \frac{12 \sum_e R_e^p R_e^q - 3M(M+1)^2}{M^3 - M} \qquad (2.7)$$

where R_e^p and R_e^q are the ranks of source and target nodes associated with edge e, based on their p and q-degrees, respectively, and ρ_q^p is referred to as the Spearman's Rho [121].

Using the following rule, we assign ranks, R_e^p and R_e^q, to all the nodes in the network. Given a sequence of n distinct values $\{x_i : 1 \leq i \leq n\}$, $R(x_i)$ denotes the rank of x_i and is given by $R(x_i) = \{j : x_j \geq x_i\}, 1 \leq i \leq n$. If x_i belong to a degree sequence, we have multiple duplicate values due to the presence of multiple neighbours. As a result, the assigned ranks are not unique. Either the ties can be broken in a uniformly random manner or the tied ranks can share an average rank [121]. If we consider an example of a sequence with repeated values, $x_i = \{3, 1, 10, 4, 3\}$, the resulting ranks with ties would be $\{3, 1, 5, 4, 3\}$. If we resolve the ties uniformly randomly, we can get either $\{3, 1, 5, 4, 2\}$ or $\{2, 1, 5, 4, 3\}$. Basically, a sequence of repeated ranks are assigned unique ranks with equal probability. On the other hand, if we use an average rank to replace the tied ranks, we end up with $\{2.5, 1, 5, 4, 2.5\}$. A proof for the equivalence of the two methods is detailed and discussed in [121]. For the purpose of this study, the ties are broken uniformly at random.

Both ρ_q^p and r_q^p take values in the range [-1,1]. When edges are placed preferentially between nodes of similar degree, the values of correlation is positive and it is negative when they are placed between nodes of dissimilar degrees. In the absence of net bias, the value is close to zero.

2.3.1 DPR mechanisms

To tune 2-node degree correlations in directed networks, we propose mechanisms based on the rewiring rule in [114]. We alter the rule to adapt it to the case of directed edges and ensure that the identities of the source and target nodes are preserved. We introduce the following modifications to the procedure in [114].

Two edges, (a, b) and (c, d), are chosen at random from the edge-list. Given that $p, q \in \{in, out\}$ and a and c are the source nodes, we select the node with the higher value of p-degree. Similarly, from the target nodes, b and d, we select the node with the higher value of q-degree. An edge is placed between the two nodes after ensuring that they are not identical and are not already connected. A second edge is placed between the two rejected nodes given that they also satisfy similar constraints. Simultaneously, we delete the original edges (a, b) and (c, d). The constraints are put in place to avoid
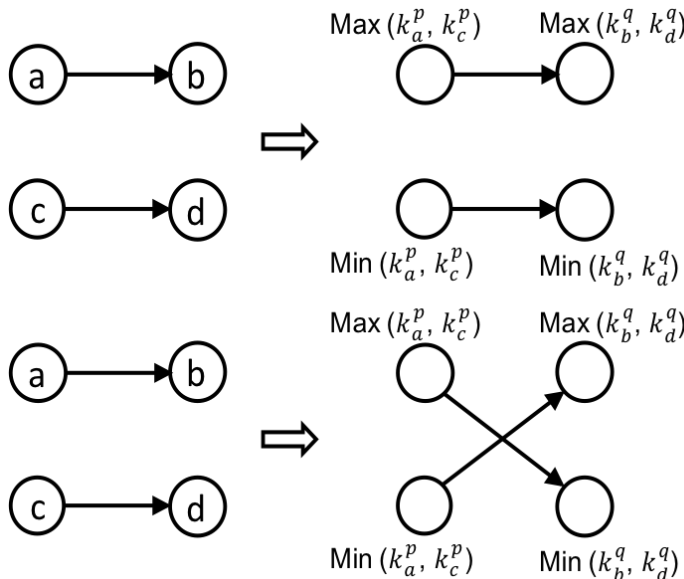
Figure 2.7: An illustration of the DPR mechanisms used for tuning assortative correlations (Top) and dissortative correlations (Bottom) on directed networks.

self-loops and multi-edges. The above sequence of steps represents a single rewiring step (top part of fig-2.7). When this rewiring step is repeatedly applied to the edge-list, the resulting network has assortative p-q type of correlation. When a desired value of ρ_q^p is reached or if we finish a predetermined number of iterations, the process is terminated. The algorithm for implementation is given in the appendix as alg-**??**.

In order to tune dissortative $p - q$ correlations in a network, we again start by randomly choosing two edges (a, b) and (c, d) from the edge-list. Between nodes a and c, the one with higher value of p-degree is chosen. But from the set of target nodes, we choose the node with the lower value of q-degree. Similarly, the rejected nodes also form a pair. If both the selected pairs of nodes satisfy the constraints mentioned earlier, two edges are placed between them while the original edges are simultaneously deleted. Once again, we ensure that the network remains simple. Repeated iteration of the above rewiring step (bottom part of fig-2.7) generates a network with dissortaive $p - q$ correlations. Once again, the procedure is terminated when a desired value of correlations is reached or after a predetermined number of iterations are run.

The algorithm for tuning degree correlations has an overall time-complexity of $O(N^2)$. We find this by using an approach similar to the one discussed in the context of clustering. In this case, the time-consuming step is the evaluation of ρ_q^p that runs in $O(N)$. Because we do not use an adjacency list, there is no $O(M)$ term.

2.3.2   Role of topology

The results of DPR mechanisms, are analyzed on both ER and SF networks. For each topology, we generate an ensemble of 100 networks, each with $N = 10^4$ nodes. ER networks are generated with $\bar{k} = 5$ and SF networks with $\beta = 0.8$. On each ensemble, we iterate the rewiring mechanisms, for both assortative and dissortative correlations. The results are averaged over the ensemble and are shown in fig-2.8 and fig-2.9, for ER and SF networks respectively.

In ER networks, the results of the tuning mechanisms are along expected lines (fig-2.8). For each mechanism, only the relevant $p - q$ type of correlation ($p, q \in \{in, out\}$)
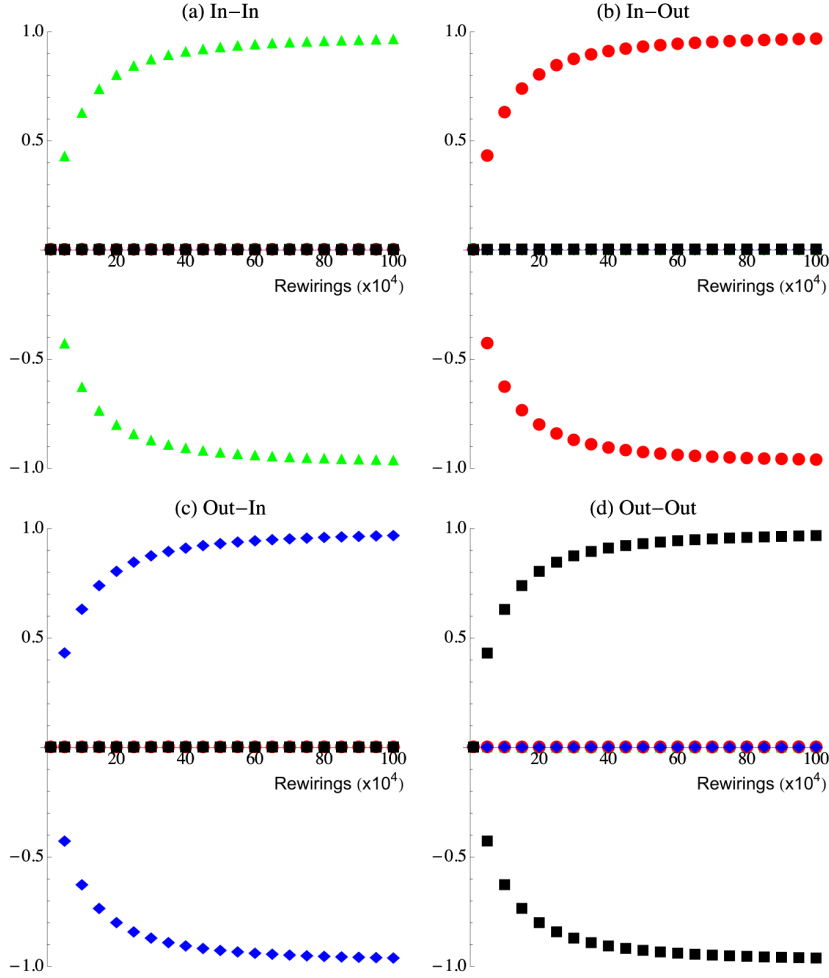
Figure 2.8: Performance of the rewiring mechanisms as a function of rewiring steps, for all four variants of degree-correlations, in directed ER networks. The results of assortative and dissortative rewiring, for each variant, are plotted together for easy comparison. In each case, the specific type of correlation that is being tuned is indicated the corresponding subplot label. ρ_{in}^{in} - green-triangles, ρ_{out}^{in} - red-circles, ρ_{in}^{out} - blue-diamonds and ρ_{out}^{out} - black-squares.

shows an increase while the change in all the other variants remains negligible. However, the results of rewiring in SF networks present a far more interesting scenario (fig-2.9). When tuning for $p - q$ type of correlations, the desired variant is introduced to the largest extent, along with small but considerable increases in all the other variants. Since this is not observed in ER networks, we conclude that this behaviour is not a result of the rewiring mechanisms but of some other topological property.

On further analysis, we find that besides the degree distributions, the two types of networks are differentiated by exactly one other property, which is 1-node correlations (r_{in-out}). The correlation between the in and out degrees of a given node is negligible in ER networks while it is strongly positive in SF networks. A quick recap of the model in [93] reveals that it is indeed an artefact of the generative process, where nodes appearing early in the growth process tend to have high in and out degrees while those appearing in the later stages have lower degrees. We introduce 1-node correlations into ER networks, in a controlled manner, to confirm our argument on the role of r_{in-out}. Given a network of size N with an in-degree sequence d^{in} and an out-degree sequence d^{out}, r_{in-out} is calculated as the Pearson correlation coefficient of d^{in} and d^{out} (eqn-2.8).
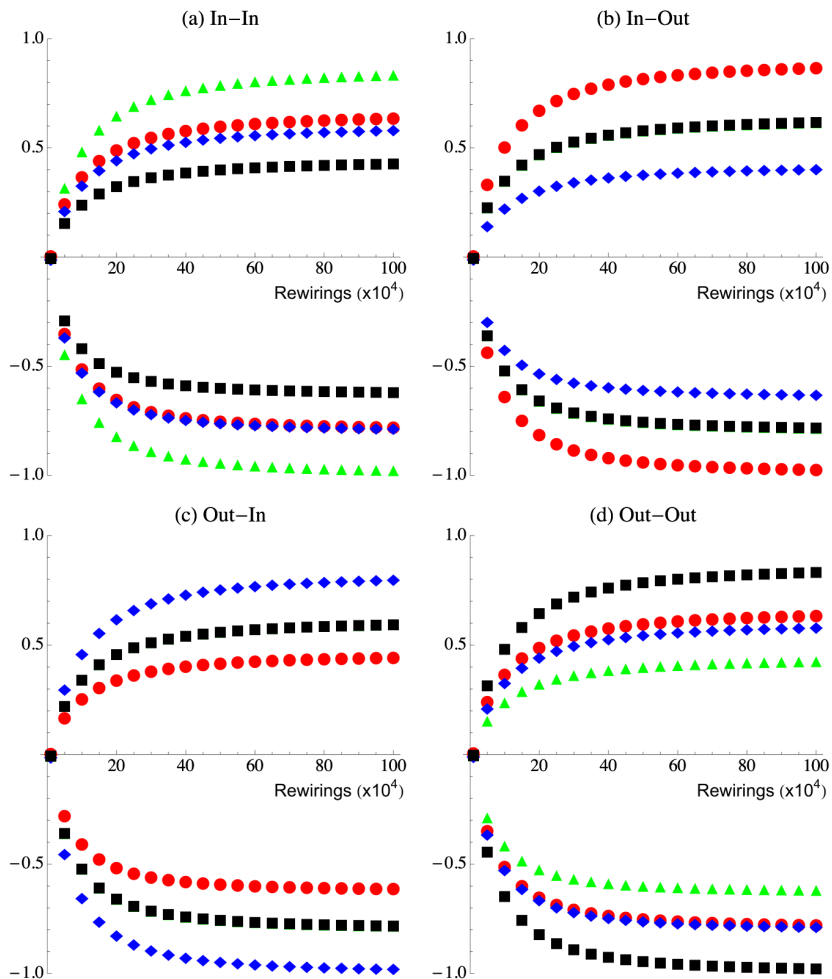
Figure 2.9: Performance of the rewiring mechanisms as a function of rewiring steps, for all four variants of degree-correlations, in directed SF networks. (Refer to the caption of Fig.2.8 for plot-legends)

$$r_{in-out} = \frac{1}{N} \frac{\sum_{i=1}^{N}(d_i^{in} - \bar{k})(d_i^{out} - \bar{k})}{\sigma_{in}\sigma_{out}} \tag{2.8}$$

where σ_{in} and σ_{out} are the standard deviations of d^{in} and d^{out} respectively and $\bar{k}$ is the average degree of the network. We also mention that the use of r_{in-out}, as compared to Spearman's Rho, is completely justified. As such, the Pearson correlation coefficient and Spearman's Rho do not take the exact same value. But the absence of a heavy-tail in the degree distribution of ER networks means that we do not observe any scaling effects with network size. To confirm this, we analyzed an ensemble of ER networks, for which we calculated both r and ρ and found them to be closely distributed around the same average value. For the specific case of Out-Out correlations, we study the effect of increasing values of r_{in-out} on the behaviour of 2-node correlations. Although the results presented in fig-2.10 are for Out-Out correlations, they hold true for any general type of 2-node correlation.

In a SF network (fig-2.9), for the same number of rewiring steps, dissortative rewiring results in a higher absolute value of ρ_q^p as compared to assortative rewiring. We find that, when the network is rewired for asymmetric types of correlations, both assortative and dissortative types result in qualitatively similar behaviour. All four types of

correlations take only three distinct values. However, when the networks are rewired for symmetric types of correlations, this behaviour is still present but only in the case of dissortative rewiring. During assortative rewiring, we observe a further splitting, so that all four variants of correlations take distinct values. This raises important question about the role of 1-node correlations and its effect on the rewiring process.
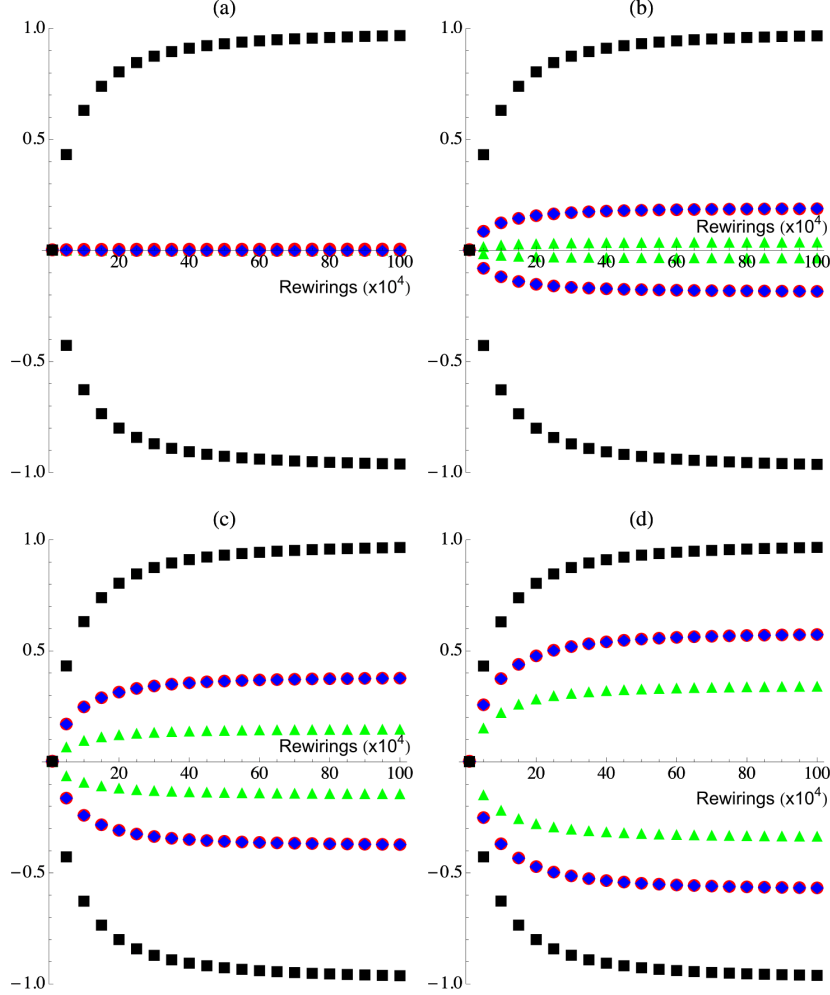
Figure 2.10: Performance of the tuning mechanisms for 2-node Out-Out correlations in ER networks, as r_{in-out} is gradually increased to take values of (a)-0, (b)-0.2, (c)-0.4 and (d)-0.6. The results extend without loss of generality to other variants of 2-node correlations. (Refer to the caption of Fig.2.8 for color-codes)

2.3.3 Role of topological parameters

In ER networks, we study the effects of assortative and dissortative rewirings for $\bar{k} = 2, 3, 4, 5$. In both cases, we see that the behaviour is identical, with slower rate of change for higher values of link-density (fig-2.11 - left). For all values of $\bar{k}$, ρ_q^p appears to converge to 1, provided there are sufficiently large number of rewiring steps. In SF networks, as we increase value of β, the results for dissortative rewiring remain similar to those in ER networks (fig-2.11 - right). However, when the network is rewired for assortative correlations, we observe that the value of ρ_q^p does not quite converge to 1 as β is increased. Instead, for the same number of rewiring steps, the value at which ρ_q^p saturates is lower for higher values of β.
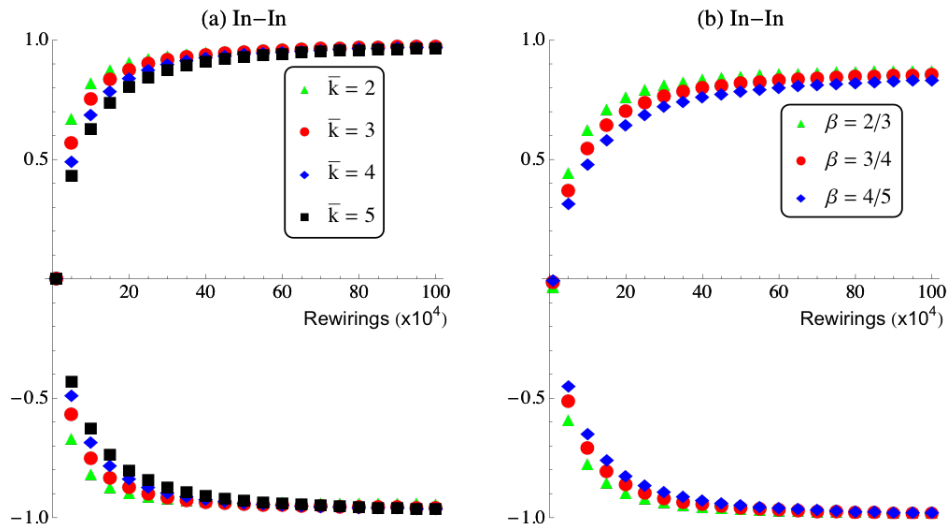
Figure 2.11: Role of link-density on the performance of DPR mechanisms for In-In correlations in ER networks (a) and SF networks (b). Rewiring for In-Out, Out-In and Out-Out correlations also show similar qualitative and quantitative behaviors.

The behaviour of the limiting value of ρ_q^p, when a network is rewired for dissortative correlations, is our next topic of interest. In [122], the authors provide a theoretical estimation and numerical corroboration of the behaviour of the limiting value. But the results are limited to undirected networks. In [122], a Dissortative Graph Algortihm (DAG) is used to generate maximally dissortative graphs and the behaviour of the limiting value of dissortativity is studied as a function of the topological parameters in ER and SF networks. However, in our work, a neutral network is made incrementally dissortative using a rewiring process. But in both cases, the degree sequence of the network is preserved. In an attempt to compare the limiting values obtained by DPR and DGA, we try to extend the DGA to directed networks. With the exception of Out-in correlations, we find that such an extension is not possible. Nevertheless, we still study the effect of topological parameters on the limiting values of correlations in directed ER and SF networks.
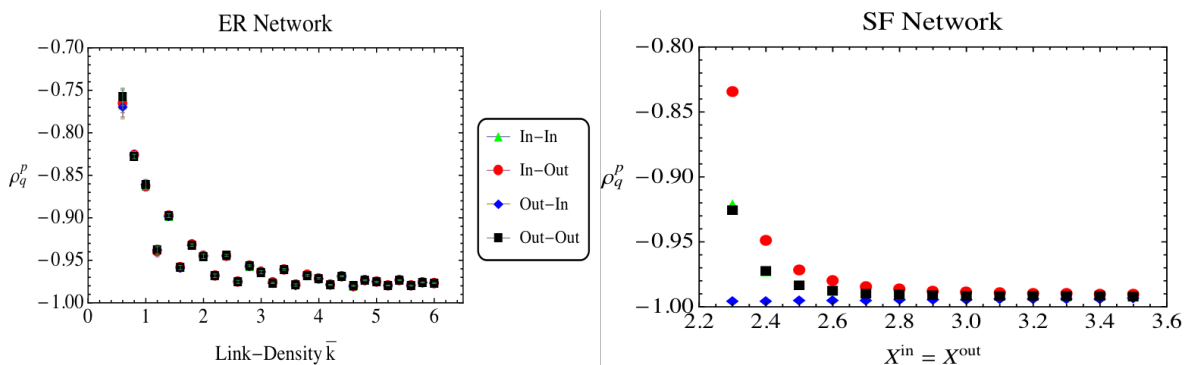
Figure 2.12: Behaviour of the limiting values of all the variants of degree-correlations as a function of link-density in ER networks and as a function of the power-law exponents in SF networks. For the case of SF networks, the in and out exponents have been constrained to stay equal.

In ER networks, we observe that the behaviour of the limiting value is exactly same for all four types of correlations (fig-2.12 - left). In all cases, the values approach -1 very rapidly as the value of link density is increased. This is also consistent with the observations in undirected networks in [122]. To study the behaviour of the limiting
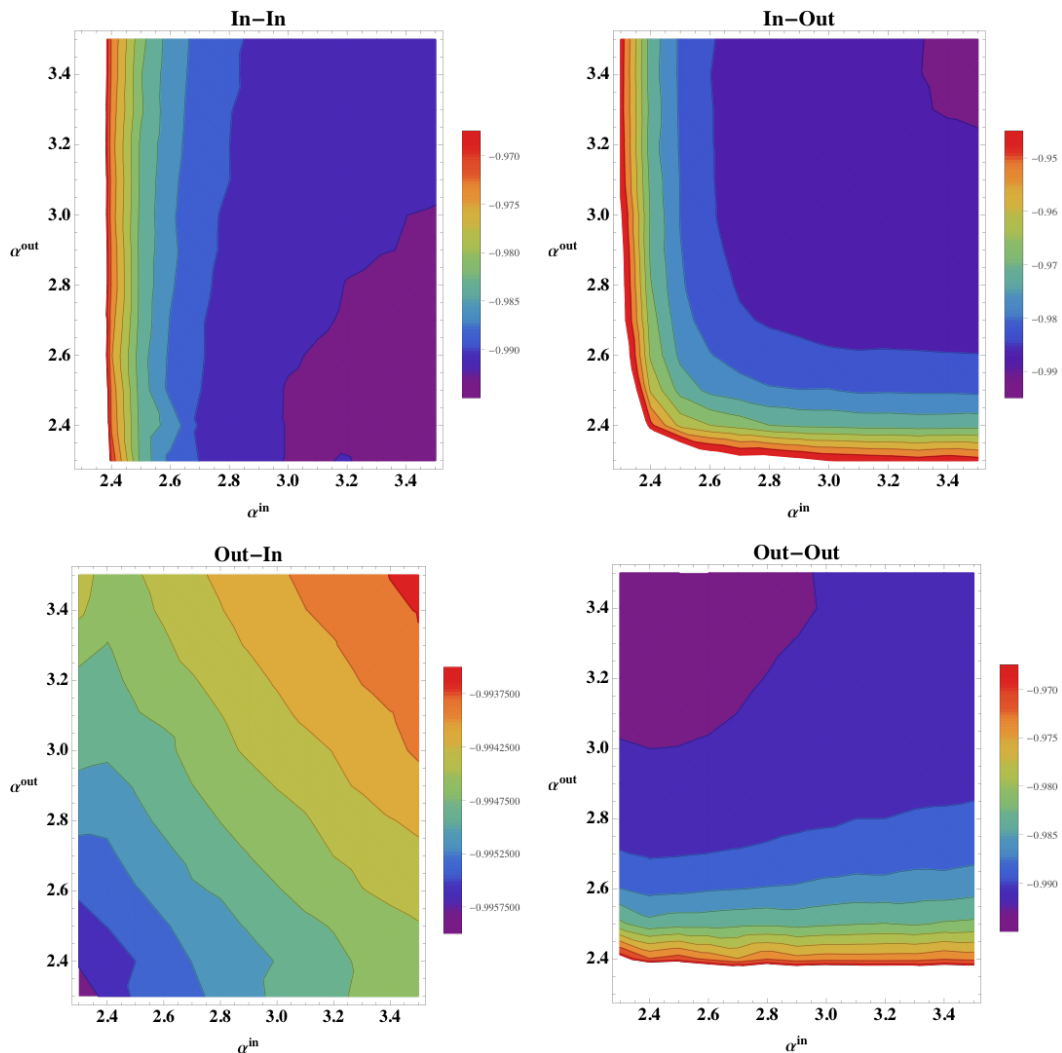
Figure 2.13: Contour plots of all four variants of degree-correlations as we simultaneously vary the values of both in-exponent and out-exponent. In each sub-figure, the color-coded values indicate the limiting values of ρ_q^p in the limit of large number of rewirings.

values as a function of the scaling exponents in SF networks, we start by setting $X^{in} = X^{out}$. The limiting values tend to -1 as the exponents become more negative. The single exception to this is the case of Out-In correlations, where the limiting value remains quite constant and very close to -1 for all values of X^{in}. The results are also in complete contrast to the observations in [122] where the authors show that the limiting values increase as the exponents becomes more negative. Next, we independently vary X^{in} and X^{out}, and for various types of correlations, we study the changing contours of the limiting values 2.13. The observations are qualitatively similar for symmetric correlations (In-In and Out-Out). For In-In correlations, ρ depends only on X^{in} and similarly only on X^{out} for Out-Out correlations. In both cases, as the relevant exponent becomes more negative, the limiting values rapidly approaches -1. In the case of In-Out correlations, the limiting value has a direct dependence on both the exponents. The behaviour is exactly opposite in the case of Out-In correlations where the limiting values move away from -1 as the exponents become more negative. Although this result is qualitatively consistent with [122], the actual amount of change is quite negligible.

2.4 Cross-effects of rewiring

During the rewiring process, we are basically exploring the space of second-order maximally random graphs. This means that we are considering the set of all graphs with fixed average degree as well as a fixed pattern of connectivity, in form of the degree distribution. In this section, we will look at the behaviour of some higher-order structural properties, when directed ER and SF networks are rewired for tuning degree-correlations under the above constraints.

To study these cross-effects, we revert to the use of the Pearson Correlation Coefficient. This is because the formula retains the information about the actual degrees of the nodes unlike the rank correlations, where the heterogeneity of the degrees is replaced by the uniform spacing of ranks. The primary reason we expect to see a change in other structural properties is due to the constraint imposed by the degree distribution. For example, the degree distribution determines the number of possible paths of length three in a given network and the clustering coefficient depends on the fraction of such paths that are closed. However, the number of paths of length three can also be treated as a higher-order degree-correlation between degrees of nodes connected by a path of length three, which is the multiplicative result of degree-correlations across each of the individual edges. Therefore, when the rewiring process changes the pattern of correlations across a single edge, we expect to see the effects propagate to higher order structures. Since the extent of change decreases significantly for higher orders, we expect maximum change in the next higher order and consequently on the clustering coefficient.

2.4.1 Rewiring in undirected networks

In [114], the authors transform the traditional definition of assortativity (r) into a form comprised of the sum of elements of powers of the adjacency matrix A.

$$r = \frac{N_1 N_3 - N_2^2}{N_1 \sum_{i=1}^{N} d_i^3 - N_2^2} \tag{2.9}$$

where $N_k = \sum_{i=1}^{N} \sum_{j=1}^{N} (A^k)_{ij} = u^T A^k u$ and u is the vector of all-ones. In terms of the degree vector $d = Au$, we have $N_3 = d^T A d$, $N_2 = d^T d$ and $N_1 = Total(d)$ = total number of edges in the network (M). The process of DPR does not affect d and therefore we find that when DPR methods are used to increase assortativity in a network, the change can be narrowed down to the increase in N_3. Since N_3 gives the total number of paths of length three in the network, a special subset of which is the set of 3-cycles, an overall increase in N_3 should result in an increase in the number of 3-cycles. In fact, we observe that there is a preferential increase in closed paths of length three as compared to open paths. In other words, we should observe an increase in the clustering coefficients of the network. We observe that this is indeed the case and the results for undirected ER and SF networks are shown in fig-2.14.

We can also make a qualitative argument in support of the above observation. When a network is assortatively rewired, the DPR process forces the nodes with similar degrees to come together and attempts to form complete sub-graphs, thus leading to an increased number of cycles. Clearly, the size of such sub-graphs depends on the degrees of the nodes involved. This becomes important in the case of scale-free networks, where there is an abundance of nodes with low-degrees and also a statistically relevant number of nodes with very high-degrees. While nodes with smaller degrees are able to come together and form tightly knit sub-graphs leading to increased local clustering,
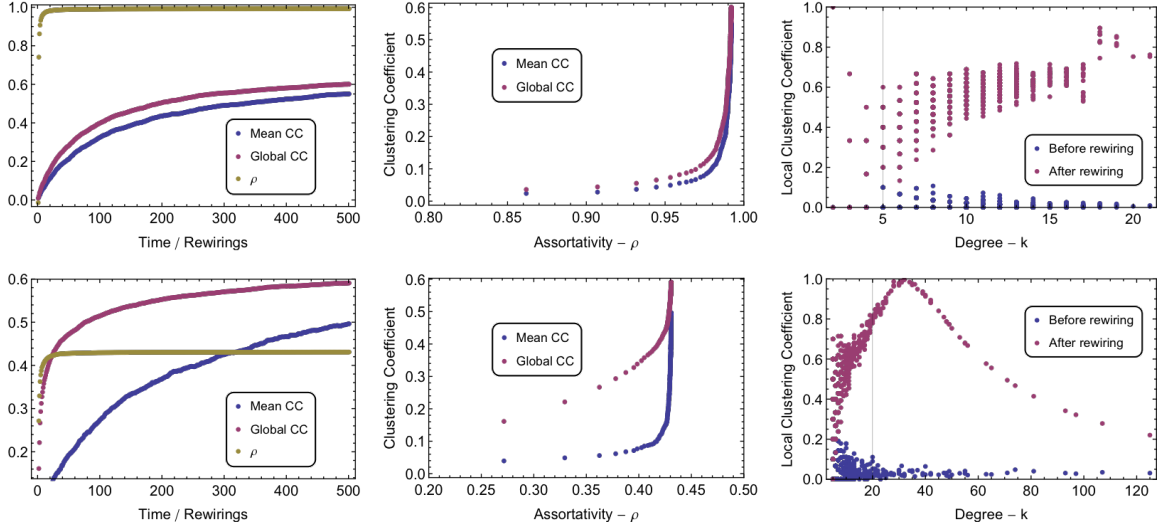
Figure 2.14: Results for DPR in undirected ER networks (Top) and SF networks (Bottom). The SF networks are generated using the Barabasi-Albert model and all the results are averaged over 50 realizations of the network, each generated with $N = 1000$ and $< k > \sim 5$ and undergoing 10^6 rewiring steps. *Left* - The behavior of Assortativity (ρ), Mean Clustering Coefficient (MCC) and Global Clustering Coefficient (GCC) as a function of rewiring steps. *Center* - Plot of MCC and GCC against ρ. *Right* - Plot of initial and final values of the Local Clustering Coefficients (LCCs) against their respective node degree.

the nodes with high-degrees are unable to form such tightly knit neighbourhoods. This is because there are not enough nodes to simultaneously satisfy the condition of assortativity and also form cliques. While a considerable fraction of their edges are shared among themselves, the rest of the edges connect them to nodes with intermediate degrees. For this reason, we observe that in SF networks, the LCC of nodes initially increases with degree, then reaches a maximum at some intermediate value and drops considerably for high-degrees. This behavior does not arise in ER networks because of the absence of nodes with very high-degree.

2.4.2 Rewiring in directed networks

In directed networks, owing to the distinction between in and out degrees of a node, we have four types of 2-node degree-correlations: $In-In, In-Out, Out-In$ and $Out-Out$. Earlier, we have presented DPR mechanisms to tune each type of correlation either assortatively or dissortatively. Using these tuning mechanisms and quantifying the change in correlations in terms of the Pearson Correlation Coefficient (r_q^p), so as to relate to other network properties such as clustering coefficients, we study the effects of tuning specific correlations on specific types of clustering coefficients.

As we have seen in the case of undirected networks, the change in clustering is proportional to the change in correlations. We expect qualitatively similar behaviour in directed networks. Since the initial networks have very low (or negligible) values of clustering coefficients, further decreasing the value of correlations by dissortative rewiring cannot have any substantial impact of the clustering coefficients. However, increasing the correlations by assortative rewiring can be expected to increase the amount of clustering. Therefore, we only focus on rewiring the networks for assortative correlations.

Given a directed network $G(N, M)$, r_q^p is defined as the pearson correlation coefficient of the p-degrees and q-degrees of the source and target nodes, respectively, in the

edge-list, where $\overline{k}^p = \frac{1}{M}\sum_{e=1}^{M} k_e^p$ and $\overline{k}^q = \frac{1}{M}\sum_{e=1}^{M} k_e^q$; A is the adjacency matrix of the network and $p, q \in \{in, out\}$. This is a slightly modified form of eqn-2.6.

$$
\begin{aligned}
r_q^p &= \frac{1}{M\sigma_{k^p}\sigma_{k^q}} \sum_{e=1}^{M}(k_e^p - \overline{k}^p)(k_e^q - \overline{k}^q) \\
&= \frac{1}{M\sigma_{k^p}\sigma_{k^q}} \sum_{e=1}^{M} k_e^p k_e^q - \overline{k}^p k_e^q - \overline{k}^q k_e^p + \overline{k}^p \overline{k}^q \\
&= \frac{1}{M\sigma_{k^p}\sigma_{k^q}} \sum_{i,j=1}^{N} A_{ij}(k_i^p k_j^q - \overline{k}^q k_i^p - \overline{k}^p k_j^q + \overline{k}^p \overline{k}^q) \\
&= \frac{1}{M\sigma_{k^p}\sigma_{k^q}} \left[\sum_{i,j=1}^{N} A_{ij} k_i^p k_j^q - M\overline{k}^p \overline{k}^q\right] \\
&= \frac{1}{M\sigma_{k^p}\sigma_{k^q}} \left[\mathbf{k}^{pT}\mathbf{A}\mathbf{k}^q - M\overline{k}^p \overline{k}^q\right]
\end{aligned}
\tag{2.10}
$$

When we substitute for p and q in eqn-2.10, the individual formulae come out to be the following:

$$
\rho^{in-in} = \frac{1}{M\sigma_{k^{in}}\sigma_{k^{in}}}[\mathbf{k}^{in^T}\mathbf{A}\mathbf{k}^{in} - M\overline{k}^{in}\overline{k}^{in}] = \frac{1}{M\sigma_{k^{in}}\sigma_{k^{in}}}[\mathbf{u}^T(\mathbf{A}\mathbf{A}\mathbf{A}^T)\mathbf{u} - M\overline{k}^{in}\overline{k}^{in}]
\tag{2.11}
$$

$$
\rho^{in-out} = \frac{1}{M\sigma_{k^{in}}\sigma_{k^{out}}}[\mathbf{k}^{in^T}\mathbf{A}\mathbf{k}^{out} - M\overline{k}^{in}\overline{k}^{out}] = \frac{1}{M\sigma_{k^{in}}\sigma_{k^{out}}}[\mathbf{u}^T(\mathbf{A}\mathbf{A}\mathbf{A})\mathbf{u} - M\overline{k}^{in}\overline{k}^{out}]
\tag{2.12}
$$

$$
\rho^{out-in} = \frac{1}{M\sigma_{k^{out}}\sigma_{k^{in}}}[\mathbf{k}^{out^T}\mathbf{A}\mathbf{k}^{in} - M\overline{k}^{out}\overline{k}^{in}] = \frac{1}{M\sigma_{k^{out}}\sigma_{k^{in}}}[\mathbf{u}^T(\mathbf{A}^T\mathbf{A}\mathbf{A}^T)\mathbf{u} - M\overline{k}^{out}\overline{k}^{in}]
\tag{2.13}
$$

$$
\rho^{out-out} = \frac{1}{M\sigma_{k^{out}}\sigma_{k^{out}}}[\mathbf{k}^{out^T}\mathbf{A}\mathbf{k}^{out} - M\overline{k}^{out}\overline{k}^{out}] = \frac{1}{M\sigma_{k^{out}}\sigma_{k^{out}}}[\mathbf{u}^T(\mathbf{A}^T\mathbf{A}\mathbf{A})\mathbf{u} - M\overline{k}^{out}\overline{k}^{out}]
\tag{2.14}
$$

where we use the matrix relations for the in and out degree-vectors, given as $\mathbf{k}^{in} = \mathbf{A}^T.\mathbf{u}$ and $\mathbf{k}^{out} = \mathbf{A}.\mathbf{u}$ and $\mathbf{u}^T = [1, 1, ..., 1] \in \mathbf{R}^{1 \times N}$.

Since any DPR mechanism preserves the values of $\overline{k}^{in}$, $\overline{k}^{out}$, $\sigma_{k^{in}}$ and $\sigma_{k^{out}}$, when the network is assortatively/dissortatively rewired, the increase/decrease in r_q^p can only correspond to an increase/decrease in the positive term. The various product combinations of A and A^T in eqn - 2.11 to 2.14, give the total number of different types of paths of length three. Thus, the diagonal terms of the resulting matrices give the number of various closed triplets and can each be matched to a particular type of clustering coefficient as given in eqn-2.1 to 2.4.
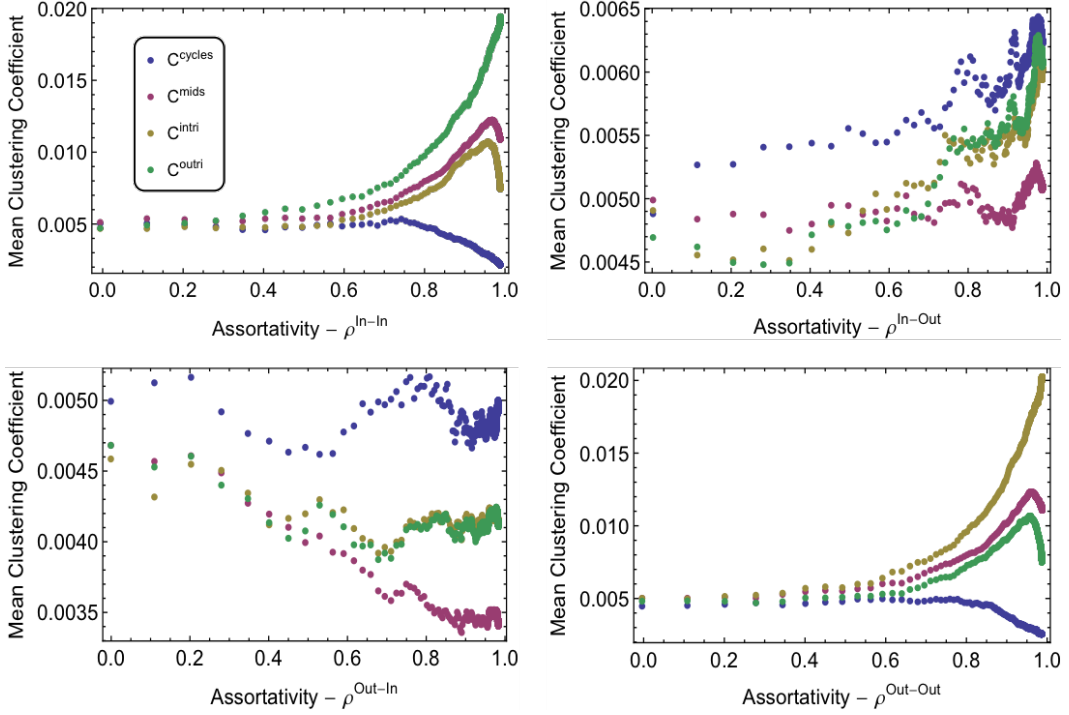
Figure 2.15: Behaviour of various clustering coefficients when the directed ER network is rewired to increase each of the different types of degree correlations.
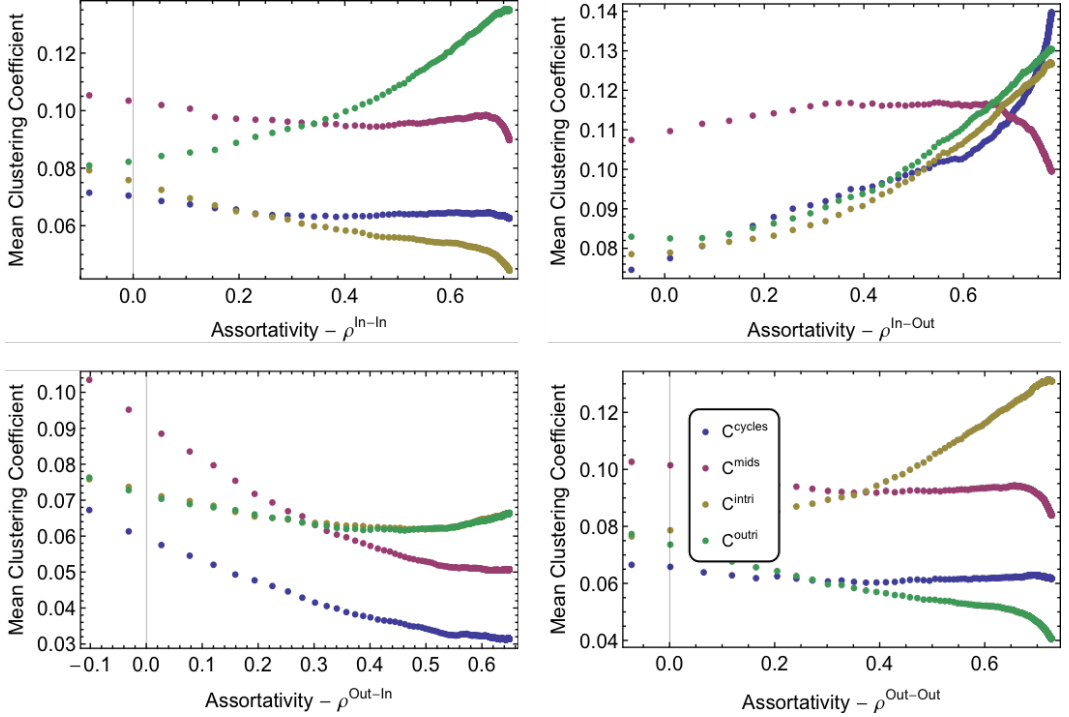
Figure 2.16: Behaviour of various clustering coefficients when the directed SF network is rewired to increase each of the different types of degree correlations.

Therefore, we expect to observe that changes in r_{in}^{in}, r_{out}^{in}, r_{in}^{out} and r_{out}^{out} lead to changes in $\overline{C}_{out}$, $\overline{C}_{cyc}$, $\overline{C}_{mid}$ and $\overline{C}_{int}$ respectively. From fig-2.15 (for ER) and fig-2.16 (for SF), we find that it is indeed the case but only for the symmetric type of correlations, r_{in}^{in} and r_{out}^{out}. In the case of r_{out}^{in} and r_{in}^{out}, there is no distinct change in $\overline{C}_{cyc}$ and $\overline{C}_{mid}$

respectively. In both ER and SF networks, we observe a substantial increase in $\overline{C}_{out}$ and $\overline{C}_{int}$ whereas in the case of r_{out}^{in}, there appears to be an increase in $\overline{C}_{cyc}$ in the limiting case, which leads us to believe that this effect of DPR is most likely independent of topology. In both ER and SF networks, $\overline{C}_{mid}$ decreases with increase in r_{in}^{out}.

In order to explain the observed results, we take a more detailed look at the local mechanisms by which the rewiring process may or may not lead to an increase in clustering. This is based on the premise that a sequence of rewiring steps involving a common set of nodes results in closure of triads. When we rewire a network to tune degree-correlations, we are interested in the processes that take place in the common neighbourhood of 2 nodes on either ends of an edge. Broadly speaking, these processes can be of two types.

In the first type, the rewiring process results in an edge being placed between two nodes that do not have a common neighbour. In the case of rewiring for symmetric type of assortative correlations r_p^p, where $p \in \{in, out\}$, if an edge is placed between nodes i and j, then they must have similar values of p-degree. In some successive step, another edge may be placed either reciprocally or from one of the nodes, say j to another node k with a similar p-degree. Based on the constraint of the rewiring process, there is a high probability that an edge may be placed between i and k, since they must be of similar p-degree, thus resulting in a closed triplet. In general, this process can lead to formation of closed loops of any length, where the probability of loops decreases with loop-length. As a result, the network reorganizes into cliques of various sizes, thus resulting in densely connected neighbourhoods for sets of nodes of comparable degree, consequently increasing the the value of the respective clustering coefficients. Additionally, the presence of reciprocal edges adds to the efficiency of triadic closure. In the case of asymmetric type of correlations r_q^p, where $p, q \in \{in, out\}$ and $p \neq q$, if an edge is placed from i to j and the next edge is placed from k to j, the rewiring process mandates that i and k have the same type of degree. Therefore, an edge between i and k is not preferred, resulting in loss of triadic closure. As a result, rewiring for asymmetric type of correlations do not show any substantial increase in the respective type of clustering.

The second type of process is essentially an extension of the first. During rewiring for symmetric type of correlations, if the nodes already have a valid (comparable degree) common neighbour or have attained one from the first process, then two triads are simultaneously setup, thus increasing the probability of triadic closure. The high possibility of valid reciprocal edges further improve the chances of increased clustering. In case of asymmetric correlations, no reciprocal edges are possible and no triads are formed, resulting in negligible increase in clustering.

2.4.3 Effects on connected components

As discussed above, rewiring for assortative symmetric type of correlations results in an increase in clustering coefficient. This is due to nodes of similar degree-type and degree-value coming together to form smaller complete sub-graphs. For example, when we rewire for assortative in-in correlations, a large number of in-links get concentrated among these smaller sets of nodes, thus leading to lesser heterogeneity in edge distribution. Consequently, the fractional size of the in-component decreases significantly. A similar observation is made for the out-component when rewiring for assortative out-out correlations. Since the size of the strongly connected component (SCC) depends on the intersection of the in and out components and in the limiting case, the upper bound on its size is given by the smaller of the two component sizes. When we

numerically estimate the fractional size of the SCC, we find this is indeed true in both ER and SF networks as seen in fig-2.17.
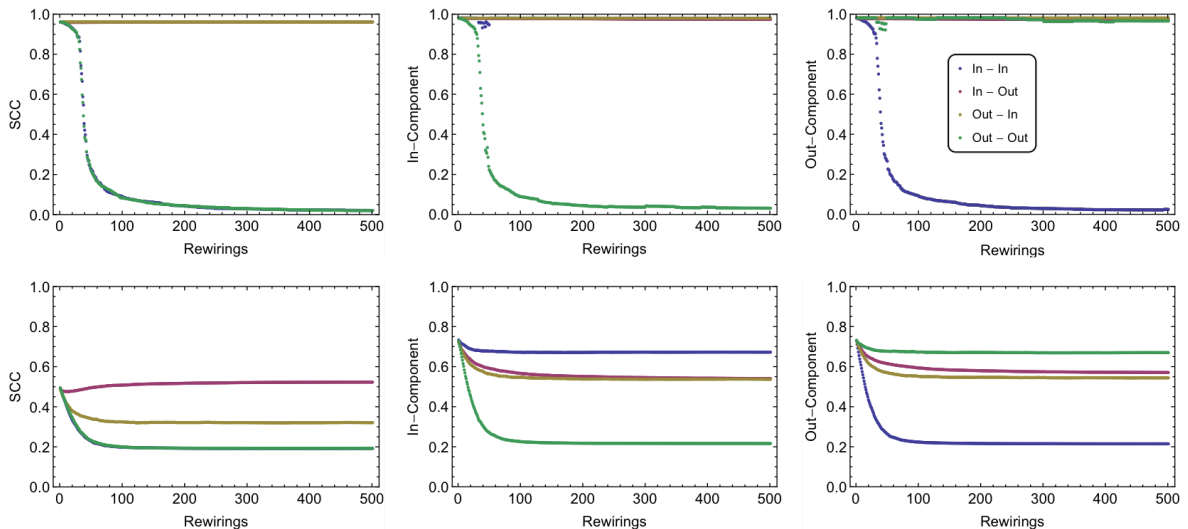
Figure 2.17: Changes in the fractional sizes of the SCC (left), In-Component (center) and Out-Component (right), plotted as a function of rewiring steps, during assortative rewiring in ER networks (top) and SF networks (bottom).

During asymmetric type of correlations, we have seen that there is negligible change in clustering and therefore the heterogeneity of the edge distribution is maintained. This leaves behind a much larger in/out component and consequently a much larger SCC, as shown in fig-2.17 (top). In the case of SF networks, the presence of large 1-node correlations results in a small decrease in component size even for asymmetric correlations (fig-2.17 bottom).

2.5 Summary

In this chapter, we propose DPR mechanisms to tune 2-node degree-correlations and clustering in directed networks and present the results of our study on ER and SF networks. Using these mechanisms, we are able to explore the roles of certain structural properties, independent of the topology of the network and also understand their role in the overall organization of the network. In the absence of viable growth mechanisms, these mechanisms provide alternate ways to introduce and tune network properties. They also play an important role when there is a lack of understanding about the underlying processes or mechanisms that give to these properties in the first place. The ability to artificially tune the amount of clustering in a network, makes it easier to study its role in information dissemination in computer or social networks and rumor-spreading in online or offline communities. It also helps in designing and testing efficient mitigation strategies to contain epidemics in contact networks. Being able to control the amount of degree-correlations also plays an important role in this context. It helps us to explore the role of correlations, on the robustness of networks to failure, in a systematic and controlled manner.

When tuning either correlations or clustering, the link-density of the network does not affect any qualitative change in the working of the mechanisms. The effect of rewiring is in general lower for higher link densities. We also conclude that the topology itself affects the qualitative behavior of the mechanisms. We find that some properties cannot be tuned in isolation due to inherent structural relationships that are built into

the definitions of the properties. This is especially relevant in the case of directed SF networks. We also present a brief discussion on the running-times of the algorithms. For both clustering and correlations, the algorithms have approximately quadratic time-complexity.

In the case of degree-correlations, for sufficiently large number of rewirings steps ($>> N$), the behavior of the limiting values of ρ_q^p is studied as a function of the parameters of the networks. In both ER and SF networks, we observe that the limiting values of ρ_q^p rapidly approach -1 for increasing values of parameters of the networks.

Finally, we find that the same mechanisms that were designed to tune correlations and clustering in directed networks can also be used as tools to explore deep structural relationships in the same networks. We find that rewiring for specific types of degree-correlations theoretically result in change in specific types of clustering coefficients but in practise, it only holds true for symmetric types of correlations but not for asymmetric types. We propose mechanisms by which this could most likely be occurring and find that they predict a decrease in connected component sizes, which we verify numerically.

Chapter 3

Structural response to link deletion

3.1 Introduction

Having established the importance of studying complex networks and proposing DPR mechanisms to tune and control clustering and degree-correlations in directed networks, from this chapter onwards, we will focus on the process of link deletion and its effects on various aspects of these networks. Specifically, in this chapter, we present our studies related to the structural robustness to link deletion. It is one of the more important aspects and measures the ability of the network to withstand and (or) adapt to internal and (or) external changes. A string of recent global events like electricity blackouts [123–127], stock-market crashes [128–131], large-scale email attacks [132] and global internet failures have spiked the interest in network robustness and brought attention to the structural integrity of such systems.

Loss of nodes and/or links can compromise the structural integrity of a network. However, considerable efforts have already been dedicated to understanding the role of node deletion [25–32, 64, 133–136]. These studies are also much more extensive in the case of undirected networks. Therefore, only the problem of link deletion, especially in the case of directed networks has remained relatively unexplored [63, 137–143]. Even the existing studies on the robustness of directed networks [33, 34], are very limited in scope. They deal with very specific examples or applications and the results do not extend to the general case of directed networks. Although many real-world networks are directed in nature, like biological networks, technological networks, transport networks etc., we do not find any proportionate effort to understand their properties and behaviour. Due to the lack of literature concerning the robustness of standard models of directed networks or common strategies for loss of nodes/edges and resulting response, we find this area to be promising and worth exploring.

While link deletion and node deletion may initially seem very similar, this notion changes when we realize that link deletion allows us to manipulate the network connectivity in a much more controlled manner. Also, during link deletion, the nodes remain completely functional with the possibility of reconnection. As a result, removal of links can be used to improve and/or impede the efficiency of flow on a network. An interesting example is the case of the power-grid. When transmission lines are disconnected, either by natural or man-made causes, there is loss of transmission. This has a 2-fold effect on the grid. Besides a decrease in overall transmission efficiency, the loss of a single transmission line can result in overloading of neighbouring nodes and links, thus leading to cascade of failures. In this case, isolating the unstable node from the rest of the network by link deletion, can help prevent a breakdown salvage the remainder of the network. Similar arguments can be made in other applications like computer

networks (to control the spread of a virus), epidemic networks (prevent interactions to control the spread of a disease) and road/rail networks (how will a damaged road/track or a traffic jam effect the movement of traffic?). The study of neuro-degenerative disorders like Alzheimer's Disease (AD) is another important application of link deletion, where the loss of connectivity between neurons leads to lowered efficiency of signal transmission in the nervous system. Alternatively, the trimming of links can also be used to streamline the flow in a network. During the process of breakdown, if the behaviour of the network is known, it helps us to pinpoint the structural and functional vulnerabilities, which in turn helps to design robust systems that are resilient to such specific types of losses. Precautionary measures, that come into play in the event of breakdown, can also be deployed.

In this chapter, we present random and targeted strategies of link deletion and their effects on directed networks. The studies are conducted on networks with random and scale-free topologies, exhibiting other properties like correlations and clustering [144]. We start by selecting a set of network models and then outlining strategies for deletion, based on local and global information. Initially, we work with a set of network models rather than actual real networks. This is done to isolate the roles of specific properties to specific strategies of deletion. This is followed by a discussion of the results of link deletion in the selected networks models and the role of clustering and degree-correlations. Finally, the responses of three real-world networks to the various types of deletion are studied and compared with the results on network models.

3.2 Network models

Despite the wide range of applications that can be modelled by a network based approach, many of the resulting networks tend to exhibit some predominantly common characteristics. As was mentioned in chapter-1, a common occurrence among many real-world networks is the heavy-tailed degree distribution and the resulting scale-free behaviour, that is well-captured by the power-law distribution. The property of clustering, is also observed in many of these networks. Some classes of networks like animal brain networks and social networks, including facebook, twitter, quora etc., show very high values of clustering. In other biological networks like protein interaction networks, gene regulatory networks, metabolic networks and also most technological networks, the clustering is quite negligible. Degree-correlations, based on both enumerative and scalar properties, are also observed in many networks and are an important classification factor. Most prominently, large positive correlations (assortativity) have been observed in social networks with respect to many factors like age, sex, hobbies etc. On the other hand, networks like the World Wide Web (WWW) and the internet show negative correlations (dissortativity). Some networks also exhibit other higher order properties like long-range correlation, occurrence of communities, hierarchical community structure etc.

The process of modelling and studying the response of a network external perturbation becomes a challenging task, due to the wide range of network properties. The complex nature of network organization and the co-occurrence of multiple properties means that a systematic and comprehensive analysis of networks requires us to isolate the role of individual properties. This can also provide insight into the relationship between various network properties. In order to carry out such an analysis, we generate ensembles of networks that exhibit only select properties while others become statistically negligible. These ensembles are generated using existing network models and the desired network properties are introduced by way of degree preserving rewiring mech-

anisms. Specifically, we focus the roles of 2-node degree-correlations and clustering. A brief description of the relevant network models is provided below.

3.2.1 Directed Erdös-Rényi (ER) network

The first network of interest is the directed ER model [50], that was earlier discussed in section-1.3.1. In contrast to what was earlier presented, in this case M directed edges are placed uniformly randomly between N nodes. The in and out degrees have a poisson distribution, with fixed average degree, negligible clustering and correlations and the average path-length scales as the logarithm of N. This model is very simplistic and exhibits no intrinsic structural biases. In the current study, we generate ensembles of ER networks of size $N = 2500$ and average degree equal to 5 and subject them to multiple strategies of link-deletion.

3.2.2 Directed scale-free (SF) networks

Despite its historical importance, the ER model has some important shortcomings, most notably in the degree distribution. Unlike the poisson distribution observed in ER networks, most real world networks exhibit a power-law degree distribution. To capture the role of this power-law distribution, on the robustness to loss of links, we employ the model proposed by Bollobás et. al. in [93]. A detailed discussion of this model is presented in section-2.2.2. The model uses a growth mechanism, along with preferential attachment, to generate directed scale-free networks. In the infinite size limit, the degree-correlations and clustering become negligible, and this allows us to study the effects of scale-free nature in isolation. An ensemble of networks, with $N = 2500$, ~ 9000 links and exponents equal to 2.5, are used in this study.

3.2.3 Degree-correlated ER and SF networks

An important property observed in many real world networks is that of 2-node degree correlations. In some categories of networks, like social networks, the correlations are positive and in others, like technological and computer networks, they are negative. To investigate their contribution, to the resilience of networks to link deletion, we generate neutral ER and SF networks and then introduce degree-correlations into them, using the degree preserving rewiring mechanisms presented in section-2.3.1. Neutral ER and SF networks are generated using ER model and Bollobás model respectively. The directed nature of links gives rise to four different types of 2-node degree correlations, namely $In - In$, $In - Out$, $Out - In$ and $Out - Out$. In the present study, both ER and SF networks are generated with $N = 2500$ and are tuned for all four types of correlations an dto the same extent ($\rho_q^p \approx \pm 0.4$). Spearman's rank correlation, as defined in eqn-2.7, is used to calculate the value of correlation.

3.2.4 Clustered ER and SF networks

Another important property, that is not captured in the ER model or SF model [93], but is observed in many real networks (social networks in particular), is clustering. The ability of a network to organize into well-knit neighbourhoods, it is mathematically defined as shown in eqn-(2.1 - 2.4).

We first use the Watts-Strogatz (WS) model [1], adapted to the case of directed networks, to explore the role of clustering in the absence of scale-free behaviour. The initial network is 1-D lattice of size $N = 1000$ nodes where every node is connected

to three immediate neighbours on each side via in and out links. We rewire each link with a rewiring probability $p_r = 0.1$, so that the resulting network has a clustering coefficient ~ 0.4 and the degree distributions move closer to the poisson distribution. These networks are then subjected to various strategies of deletion and the results are analyzed. For the case of SF networks with clustering, we generate the initial network using [93], and then increase the clustering using degree preserving rewiring mechanisms introduced in section-2.2.1.

3.2.5   Real-world networks

Until this point, we have only discussed standard models that generate networks with some desired properties. But the resulting networks are not always accurate representations of real networks. However, investigation of these model networks allows us to study the specific roles of various network properties on the robustness to loss of links. This can then help us to identify and analyze important results in real networks, where one or many properties can coexist. To this end, we now introduce three real networks, on which we study the problem of link deletion. These data-sets have been chosen to highlight the role of aforementioned properties like SF nature, degree-correlations and clustering. Below, we provide a brief description of each data-set.

Austin Road Network: The first data-set is an example of road transport network, and represents the flow of traffic between road intersections in the city of Austin, Texas. Nodes represent the intersection of two or more roads and the flow of traffic between such nodes constitutes the links. Although the original data-set, sourced from [145], contains detailed metadata, we concern ourselves with only the edge-list of the network. With 7388 intersections and 18596 connections, the degree distribution of this network is quite homogeneous and close to a poisson distribution. As expected of any city transport network, it is highly assortative with respect to all four types of degree correlations but shows negligible clustering. The connected components also span almost the entire network. As a transport network, it lends itself as an extremely relevant system for the study of link deletion. The process of link deletion can be used to model many problems like shutting down of certain roads due to emergencies, unavailability of roads due to traffic jams, closure for maintenance purposes etc. Since this network dominantly exhibits only assortative degree-correlations and no other major attributes, we can expect to study the singular effect of positive degree-correlations on link deletion.

Protein Interaction Network: This data-set comprises a network of interacting human proteins and was put together by U. Stelzl et al. [146,147]. This was done in an effort to improve the understanding of the organization of the human proteome. The network is constructed using 1706 proteins (nodes) undergoing 6171 unique interactions (links) and the data was obtained by high-throughput Y2H (Yeast Two-Hybrid) screening. The network has very high 1-node correlation and both the in and out degree distributions follow power-laws. The clustering in the network is negligible and the strongly connected component includes 1493 nodes. All four types of 2-node degree-correlations, in this network, are strongly negative. Hence, this network is a good representative of a dissortative SF network and is perfectly suited for us to explore the effects of link deletion on such a network. The process of link deletion can be used to model the problem of failed interactions between proteins, which could happen due to various reasons like random failure, structural mismatch, prohibitive media etc.. We

Table 3.1: Summary of the network models and real networks used in the study.

Network	Nodes	Links	WCC	SCC	AIPL	Assortativity	Clustering
Erdös Rényi (ER)	2500	12500	2500	2465	0.20	-	-
Scale-Free (SF)	2500	8750	2500	975	0.12	-	-
Assortative ER	2500	12500	2500	2462	0.20	0.30 ± 0.001	-
Assortative SF	2500	8750	2500	970	0.12	0.35 ± 0.002	-
Dissortative ER	2500	12500	2500	2466	0.20	-0.30 ± 0.001	-
Dissortative SF	2500	8650	2500	1080	0.14	-0.40 ± 0.001	-
Watts-Strogatz	1000	6000	1000	1000	0.18	-	0.45 ± 0.002
Clustered SF	1000	4000	1000	500	0.17	-	0.23 ± 0.016
Austin Road	7388	18956	7388	7381	0.02	0.23 ± 0.012	0.009 ± 0.0009
Protein Interactions	1706	6171	1615	1492	0.19	-0.13 ± 0.008	0.005 ± 0.0005
Airport	3425	37594	3397	3354	0.25	0.047 ± 0.0005	0.45 ± 0.018

study the resilience of this protein interaction network and try to present biological interpretation wherever relevant.

Airport Network: Using data obtained from the *openflights.org* dataset [148], the network represents the patterns of flight connectivity between various airports across the world. The airports represent nodes in the network and any two airports are said to be connected by a link if there is atleast one flight connecting the source to the destination airport. The constructed network has 3425 airports that are connected by 37594 flights. Although the original network has multi-edges, because of multiple flights connecting the same two airports, for the purpose of this study, we approximate it to a binary, simple graph. Since there is almost always to and fro connectivity between airports, the network has a very high 1-node correlations. The largest strongly connected component in the network spans 3354 airports. The amount of 2-node degree-correlations is negligible but the network is strongly clustered with respect to all four types of triangles. Therefore, this network lends itself as a good example of a scale-free clustered network. By studying the problem of link deletion on this network, we can model the effects of specific flight cancellations, happening on select important routes, and study the effect on dynamics of passenger traffic and the extent of loss of connectivity.

In Table - 3.1, we present a list of models and real networks that are used in this study and the relevant properties for each network are also summarized. For each network, we calculate the average value of all types of degree-correlations and present a single value under the column for assortativity. Similarly, the values given in the column for clustering are averages of MCCs of all types of closed triplets.

3.3 Strategies for link deletion

When designing a deletion strategy, the list of possibilities is quite open-ended. It is only limited by the actual type/amount of information available to carry out a particular type of attack and the purpose of the deletion itself. We choose strategies that independently utilize both local and global information about the network. Whenever we implement a particular strategy, it is assumed that all the necessary information is available. This is not always a very realistic assumption. For example, in many cases, knowledge of the network is limited/incomplete and we cannot accurately calculate the node/edge properties. This could happen due to many reasons including very large sizes or faulty data acquisition methods. Either way, strategies may need to be imple-
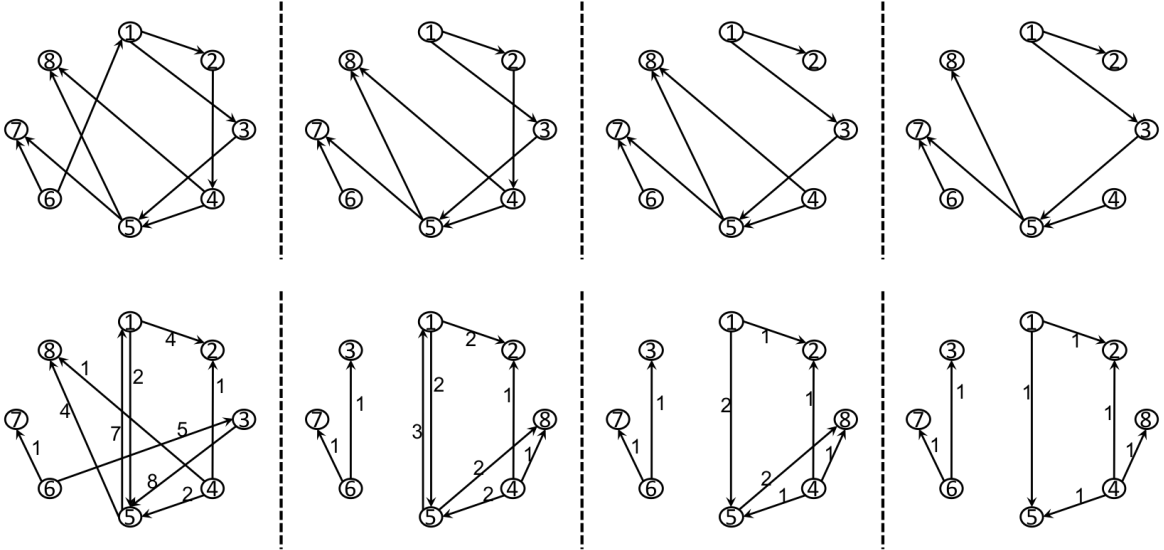
Figure 3.1: We depict the early stages of the link deletion using a schematic representation of the process (Left to Right). *Top* - During random deletion, every link is equally important and they all have the same fixed weight and so we do not specify any particular value. At any given step, each link is equally likely to be deleted. *Bottom* - For the case of targeted deletion, in this specific example, we provide the EBC values. In general, however, it could be any property of the edges. In any instance of deletion, the link with the highest weight (EBC) is deleted. If there is more than one edge with the highest value, then one of the edges is selected at random for deletion. Also, after every instance of deletion, the edge-weights are recalculated before the process is repeated.

mented with incomplete information. Some strategies may also involve the calculation of computationally intensive metrics. Centrality based strategies are an immediate example. The implementation of such strategies is extremely time-consuming and not viable for very large networks.

In our study, we consider both random and targeted deletion of links. When implementing targeted deletion, properties like Edge-Degree (local-information) and Edge Betweenness Centrality (global-information) are used to determine weights for the links. We also consider deletion based on ascending and descending orders of weights separately. Next, we discuss each strategy in greater detail. This is followed by an analysis of the relationship between Edge-Degree and Edge Betweenness Centrality and its importance.

3.3.1 Random Deletion (S1)

When it comes to real-world networks, most often, links are lost in an unforeseen and random manner. For example in the case of the power-grid, transmission could be lost due to inclement weather, or in the case of peer-to-peer networks, wear and tear of cables leads to loss of communication between routers. Even in the brain, the decay of synapses is very unpredictable and leads to neuro-degenerative disorders. In the case airport networks, flights may be cancelled due to unforeseen circumstances and it affects the performance of airports across the world. The unpredictable nature of this kind of loss of connectivity is best captured by random deletion of links. In this study, at every step, we choose a random link and delete it from the network. The resulting structural changes are studied as a function of loss of links. Although in the strict sense of the term, random deletion does not constitute an attack strategy, it still acts as a good null-model for other attack strategies. In fact, it is more correctly interpreted as

a failure of links. We refer to this type of random deletion of links as strategy S1.

3.3.2 EBC based deletion (S2,S3)

The betweenness centrality for links is a direct extension of the corresponding definition for nodes (eqn - 1.18). If n_{ij}^e is the number of shortest paths between nodes i and j, that pass through link e and n_{ij} is the total number of shortest paths between i and j. Then, the EBC of e is defined as the ratio of n_{ij}^e to n_{ij}, summed over all possible pairs of nodes in the network. But this results in a single real number that has limited use when comparing different networks with varying sizes. Therefore, we normalize this number by the total number of vertex pairs, and the resulting value of EBC is in the range from 0 to 1.

$$EBC_e = \frac{1}{N(N-1)} \sum_{i,j} \frac{n_{ij}^e}{n_{ij}} \tag{3.1}$$

When using strategy S2, the links are arranged in decreasing order of EBC value. At every step of deletion, the link with highest value is removed. In any network, the links with the highest EBC values are the ones that play the role of $'bridges'$ and connect groups of nodes that would otherwise be disconnected. Since they connect large groups of nodes, the loss of such links results in groups of nodes getting disconnected and loss of multiple shortest paths in a single step of deletion. In undirected networks, betweenness centrality based deletion has been explored earlier and has been shown to be very effective in breaking down a network. In strategy S3, the link with the lowest value of EBC is deleted. These are typically the links that connect already well-connected nodes or groups of nodes. Deleting these links would, at worst, result in longer path-lengths and lower efficiency but will not immediately create disconnected nodes or groups of nodes. After every deletion, we allow the network to reorganize and recalculate the values of EBC. Measures of robustness, degree-correlations, clustering coefficients are also evaluated at regularly timed intervals.

3.3.3 ED based deletion (S4,S5)

While the idea of degree of a node comes quite intuitively, the idea of a degree of an edge is not equally intuitive. In the context of undirected networks, the edge-degree of an edge is defined as the product of degrees of nodes on either ends of the edge. This was first introduced in [143]. In the present study, we adapt this definition to the case of directed networks. The edge-degree of a directed edge is defined as the product of in-degree and out-degree of the source-node and target-node respectively. Later, in this section, we justify the choice of this definition. If k_i^{in} and k_j^{out} are in and out degrees of source-node i and target-node j, respectively, then ED is given by eqn-3.2.

$$ED_e = k_i^{in} k_j^{out} \tag{3.2}$$

The primary motivation behind defining an edge-degree is to assign a property to a link such that it can be evaluated using only local information concerning the link. Therefore, with eqn-3.2 as the working definition of ED, we assign weights to the links and order them in increasing/decreasing order of their weights.

In strategy S4, the links are ordered in descending order of ED values and the link with the highest value of ED is deleted. The network is then allowed to reorganize and the values of ED and other network properties are recalculated. Based on the definition of ED, during S4 type of deletion, links connecting in-hubs and out-hubs are deleted

first. As a result, we can expect the network to split into multiple strongly connected components. Meanwhile, S5 targets links that connect low degree nodes and should therefore result in chipping away of the network structure.

3.3.4 Relationship between EBC and ED

As mentioned earlier in this section, there are two important requirements to designing an effective deletion strategy, namely the type and amount of information available about the network. Besides random deletion (S1), all other strategies utilize some knowledge about the network and accordingly assign weights to the edges. However, the difference between ED based strategies (S4,S5) and EBC based strategies (S2,S3) lies in the type of information being used. Strategies S2 and S3 use global information (betweenness centrality) while S4 and S5 use only local information (node-degree). We attempt to extend the traditional definition of ED to the case of directed networks, such that we can also establish a relationship between ED and EBC. By identifying this connection and also comparing the results of S2 and S4 types of deletion, we investigate the possibility of finding an ED based strategy that can replicate the results of an EBC based strategy.

$$
\begin{aligned}
k_i^{tot} k_j^{tot} &= (k_i^{in} + k_i^{out})(k_j^{in} + k_j^{out}) \\
&= k_i^{in} k_j^{in} + k_i^{in} k_j^{out} + k_i^{out} k_j^{in} + k_i^{out} k_j^{out}
\end{aligned}
\tag{3.3}
$$
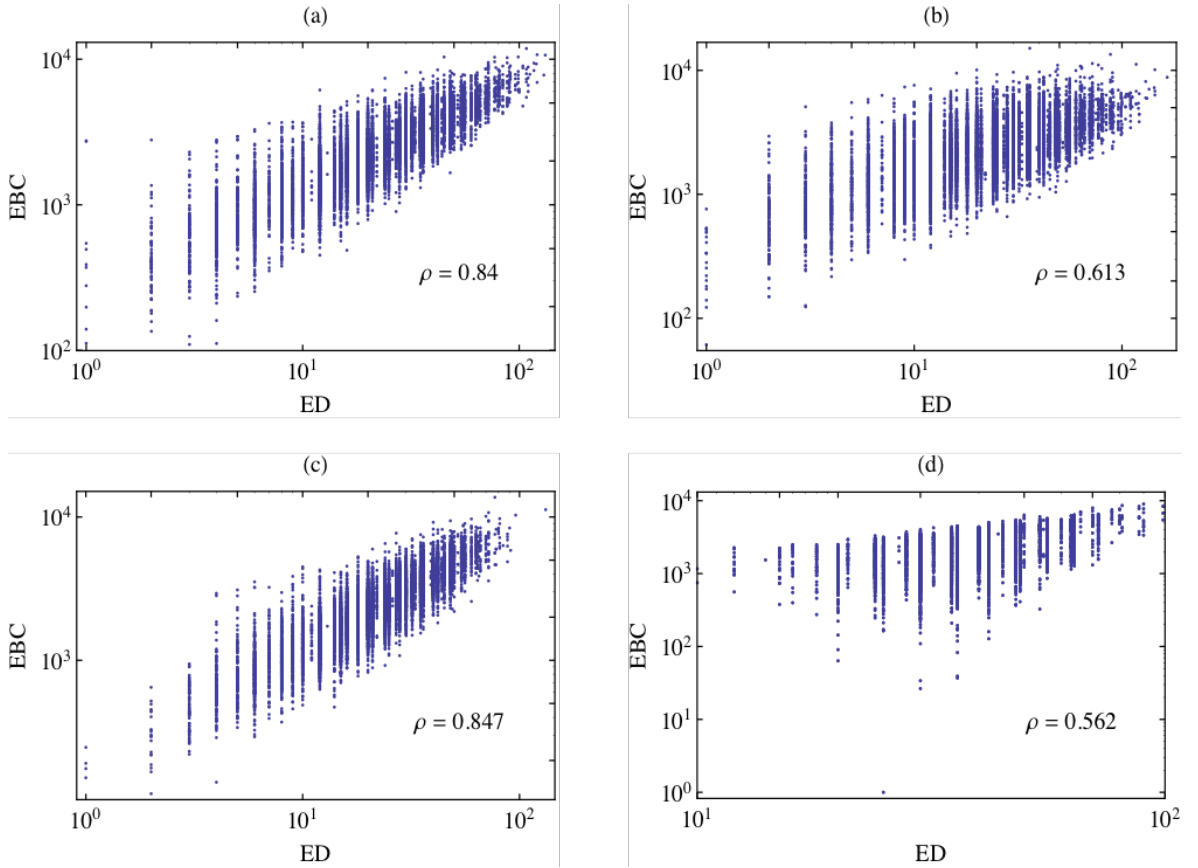
Figure 3.2: Scatter plots of EBC vs ED for (a)-ER (b)-Assortative ER (c)-Dissortative ER (d)-Directed WS. The values of the correlation coefficient are inset in the respective subfigures.
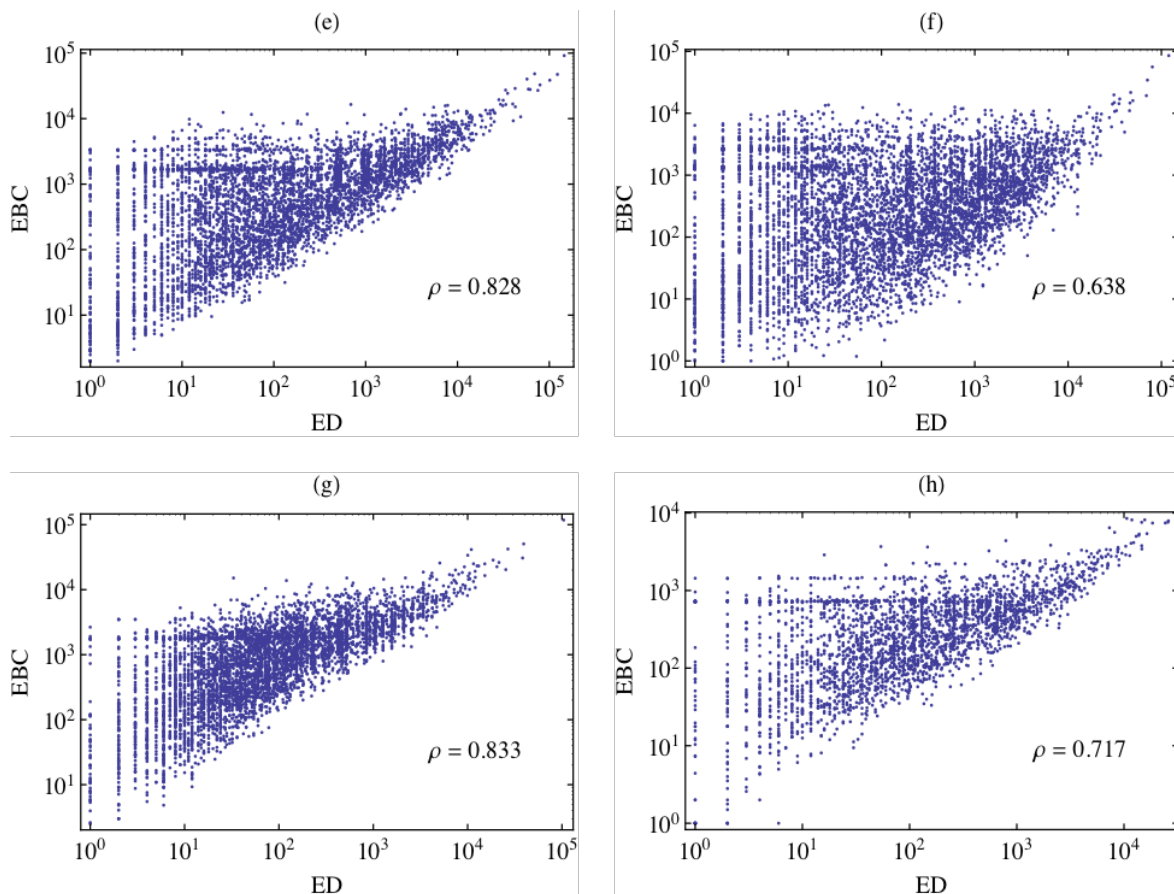
Figure 3.3: Scatter plots of EBC vs ED for (e)-SF (f)-Assortative SF (g)-Dissortative SF and (h)-Clustered SF networks. The values of the correlation coefficient are inset in the respective subfigures.

Given a link e, we start by defining ED as the product of total-degrees of the source-node i and target-node j. A study of the scatter plots of EBC vs ED does not yield any interesting relationship. We then decompose the product of total-degrees into sum of products of in and out degrees of i and j (eqn-3.3). We then study scatter plots of EBC versus each of the terms in eqn-3.3. While there is no discernible dependence for three terms, one of the terms, $k_i^{in}.k_j^{out}$, shows a strong positive correlation with EBC, especially for the case of links that connect hubs. The Pearson correlation coefficient of the two quantities is calculated as a measure of this dependence.

In fig-3.2 and fig-3.3, the scatter plots of EBC vs $k_i^{in}.k_j^{out}$ for the various networks discussed in the previous section are shown. Fig-3.2 (a) and Fig-3.3 (e) show the plots for ER and SF networks respectively. We see that there is a strong positive correlation in both cases but it is slightly higher in ER networks. In SF networks, we see that the links connecting nodes of low degrees show a wide range of EBC values and are spread over three orders of magnitude. In fig-3.2 (b,c), we look at the role of positive and negative degree correlations in ER networks on the EBC vs $k_i^{in}.k_j^{out}$ scatter plots. The same for SF networks is shown in Fig-3.3 (f,g). We find that, in both ER and SF networks, assortative rewiring leads to lower values of correlation coefficient, particularly for links connecting nodes of low degrees, while dissortative rewiring does not have any significant effect. From Fig-3.3 (h), we see that rewiring for clustering in SF networks lowers the value of ρ to 0.5. In the directed WS model, the dependence is further skewed but still quite positive as seen from the value of $\rho = 0.6$, and clearly visible for links connecting nodes with relatively high degrees.

This dependence between EBC and $k_i^{in}.k_j^{out}$, justifies the definition of ED given in eqn-3.2. This correlation between EBC and ED of a link is very important because it is much easier to deal with local properties, whether it is data collection or computational ease. Having this relationship gives us the freedom to design strategies based on local properties that are easy to execute and are also as effective as the EBC based strategies.

3.4 Measures

The process of link deletion brings about various changes in the network structure. In order to quantify these changes and understand the behaviour of the network, it is important that we choose a set of relevant measures that can capture the changes and also provide meaningful interpretations. To this end, we work with a set of reliable metrics, which clearly map the changes, their relationships and also bring out the physical implications.

3.4.1 Largest Connected Components

To understand the effects of link deletion on the macroscopic organization of the network, we study the fractional sizes of largest weakly connected component and largest strongly connected component. These metrics were discussed earlier in sec-1.2.5 but we quickly recap the definitions in directed networks. In directed networks, the set of all nodes, for which there exists a directed path between all unique pairs of nodes, is defined as the Strongly Connected Component (SCC). And the Weakly Connected Component (WCC) is defined as the set of all nodes for which there exists atleast an undirected path between all pairs of nodes in the set. The set of all nodes that can be reached by following the out-links from all nodes in the SCC is called the Out-Component (OUT) of the network. Similarly, the In-Component (IN) is obtained by following the in-links from all nodes in the SCC. The set of all nodes, that is the union of SCC, IN and OUT components is defined as the Weakly Connected Component (WCC) of the network. These different types of connected components come together to give rise to a macroscopic organization referred to as the *Bow-Tie* structure [54, 55]. This is a subclass in a broader set of core-periphery structures [55] where the core consists of high-degree nodes (hubs) and is very dense and strongly connected while the periphery is comprised mostly of low-degree nodes with sparse connectivity among themselves.

We will see that this core-periphery organization plays an important role in how various types of link-deletion play out in different types of networks. The analysis of the behaviour of these connected components leads to an understanding of the macroscopic organization in networks and is also at the heart of percolation studies on networks. In this work, we restrict our study to the fractional sizes of the largest weakly connected component (W) and the largest strongly connected component (S). We do not focus on the sizes of the associated IN and OUT components since they behave in the same manner as S.

3.4.2 Efficiency

While the behaviour of connected components helps us to study the structural changes and also the extent of flow/spread in the network, it does not shed any light on the rate of a spreading process or information flow in the network. In order to understand the rate, we calculate the Average Inverse Path Length (AIPL), also referred to as

Efficiency. It is defined as the average of the inverse geodesic lengths between all pairs of nodes in the network [149] (eqn-1.14).

$$Efficiency = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N-1} \sum_{j=1, j \neq i}^{N-1} \frac{1}{d_{ij}} \qquad (3.4)$$

From the previous three sections, it is clear that we are dealing with substantial set of results involving five different strategies of deletion on eleven types of networks. In order to compare results on the various networks, we normalize all values of the metrics by their initial values, so that they are all bound between 0 and 1. The normalized measures are represented by W/W_0, S/S_0 and E/E_0, where W_0, S_0 and E_0 are the initial values of largest weakly connected component, largest strongly connected component and efficiency respectively.

3.5 Results

In this section, we present the results of link deletion on the various network models and real-world networks discussed in sec-3.2. The networks are subject to various types of random and targeted deletion strategies and their performance is analyzed using the metrics described in sec-3.4. For each network model, subject to each type of deletion strategy, the network metrics are calculated after each event of deletion. The final results are averaged over an ensemble of 500 networks. We study the plots of each of the measures as a function of fraction of total links lost $(1 - M/M_0)$. Starting with neutral ER and SF networks, we go on to study the role of assortative and dissortative degree-correlations in each topology. This is followed by a study of effect of clustering. Finally, we take up three real world networks and study their performance under various kinds of deletion and analyze the results.

3.5.1 Link deletion in ER networks

For all strategies of deletion, the fractional size of the weakly connected component W/W_0 in ER networks, shows strong robustness (fig-3.4 (a - c)). Atleast 60% of the links are lost before W/W_0 shows any substantial decrease. However, different strategies have different effects on the behaviour of the largest strongly connected component (S). S/S_0 is initially robust to random deletion of links (S1), and breaks down rapidly after more than half the initial number of links are removed and it ceases to exist for (1-$M/M_0) \geq 0.8$. Also, $(1-M/M_0) = 0.8$ matches the known critical value of $1/N$, for structural phase-transition in random networks. The network is extremely vulnerable to S2 and S4 types of deletion and we observe instant deterioration of the strongly connected component. During S2 type of deletion, loss of less than $\sim 20\%$ of the most central links leads to complete breakdown of the SCC while it is only slightly better, at $\sim 30\%$, in the case of S4 type of deletion. The behaviour of efficiency (E/E_0) is qualitatively similar to that of SCC, for all strategies of deletion. E/E_0 and S/S_0 show almost identical behaviour, for S2 and S4 strategies, except at the point of transition to complete breakdown.

In strategies S3 and S5, we see that the SCC remains quite robust until $\sim 40\%$ of the links are lost while the efficiency starts to decrease immediately. This implies that the nodes are initially strongly connected via small-paths but with the initial loss of links, the paths get longer and less efficient. When, approximately half of the links are removed, nodes begin to get disconnected from the network and connecting

paths do not exist anymore, to or from such nodes. This results in an accelerated decrease in efficiency around $M/M_0 \approx 0.5$. This is further supported by the behavior of W/W_0, which also begins to depart from 1 at approximately the same value of $1 - M/M_0$, suggesting the appearance of disconnected nodes. We also note that the SCC undergoes accelerated breakdown, during S2 and S4 types of deletion, while the decrease in efficiency is quite constant.

Figure 3.4: Performance of WCC (W/W_0), SCC (S/S_0) and Efficiency (E/E_0) as function of the fraction of removed links (1 - M/M_0), during various strategies of link deletion, in directed ER networks (a - c) and SF networks (d - f). S1 - red circles, S2 - blue up-triangles, S3 - green down-triangles, S4 - brown squares and S5 - black diamonds.

3.5.2 Link deletion in SF networks

The WCC is quite robust to all strategies of link deletion in SF networks (fig-3.4 (d - f)). During S1 type of deletion, the decrease in WCC is initially slow and continuous but becomes more rapid during the later stages of deletion. It is more resilient to S2 and S4 strategies as compared to random deletion. The reason for this is the well-emphasized core-periphery structure that is observed in SF networks. Since most of the hubs are contained within the core, the links in the core tend to have very high values of ED and these are targeted during S4 type of deletion. Due to the large number of redundant links in the core, the network stays connected for a longer period ($1 - M/M_0 \sim 0.4$) before it ultimately begins to break down into disconnected components. Removal based on strategy S5, leads to considerable decrease in W/W_0 from the very beginning. This is because S5 targets the links in the sparsely connected periphery,creating isolated nodes, while the densely connected core remains robust. As the core-periphery structure is more emphasized in SF networks, the response to S5 is also clear-cut.

The effects of strategies S2 and S4, on the behaviour of SCC, are not very different in ER and SF networks. With loss of just 10% and 30% of initial number of links, S2 and S4 respectively lead to complete breakdown of the SCC. However, the effect of random deletion is quite different. In this case, the SCC undergoes a continuous decrease and does not exhibit any threshold behaviour. This is also consistent with results from percolation studies on networks.

Unlike ER networks, the SCC shows strong initial robustness during S5 type of

deletion. This leads back to the core-periphery structure where the low-degree nodes lie in the periphery and the hubs constitute the strongly connected core. Because of the structural bias present in the generated SF networks, where in and out degrees of a node are highly correlated, the links connecting nodes in the periphery will have the smallest values of ED. Therefore, when these links are removed, the network structure gets chipped away at the periphery, creating isolated nodes, while the core remains quite intact. The value of $1 - M/M_0 \approx 0.4$, at which the SCC begins to fall apart, is also markedly visible in the behavior of WCC as the point at which the rate of decrease abruptly slows down.

The change in efficiency follows along the lines of WCC. For S2 and S4 strategies, there is rapid initial decrease in efficiency until about 40% of the initial number of links are lost. Due to the large number of redundant links in the core, during the initial stages of deletion, the network remains well-connected while progressively becoming less efficient. This is observed until the first isolated nodes begin to appear, which also corresponds to the sharp decrease in WCC. During random deletion, we observe a monotonic decrease with a constant rate. During S5, we observe a rapid initial decrease which slows down around $1 - M/M_0 \approx 0.4$. This is the same point near which the behavior of SCC and WCC also undergoes a qualitative change. Following from our earlier explanation (for SCC), the nodes with low-degrees get isolated in the early stages, resulting in decrease in the overall number of shortest paths, which has a greater effect on efficiency compared to increased path lengths. The transition point, where the SCC and WCC also begins to breakdown, indicates the beginning of the decomposition of the core. During this phase, well-connected nodes become not-so-well-connected (rather than disconnected) and therefore the rate of change of E/E_0 decreases.

3.5.3 Role of assortative rewiring in ER and SF networks

Figure 3.5: Performance of WCC (W/W_0), SCC (S/S_0) and Efficiency (E/E_0) as function of the fraction of removed links (1 - M/M_0), during various strategies of link deletion, in directed ER networks (a - c) and SF networks (d - f) with assortative degree-correlations. S1 - red circles, S2 - blue up-triangles, S3 - green down-triangles, S4 - brown squares and S5 - black diamonds.

Next, we study the response of networks to link deletion in the presence of 2-node degree correlations. We first consider ER and SF networks with assortative correla-

tions. The rewiring process reorganizes the network so that links connect nodes of similar degrees, while preserving the degree distributions. The resulting configuration, where low degree nodes connect to other nodes with low-degrees and hubs connect to other hubs, effectively enhances the core-periphery structure of the network. The core and periphery, each become well-defined, while the connections between the two are few. The effects of strategies S4 and S5 are strongly affected as a result of heterogeneity introduced into the distribution of ED values. In ER networks, the absence of heterogeneity in the degree distributions ensures that there is minimal effect on the large-scale organization of the network due to rewiring. As a result, the effects of link deletion in assortative ER networks are similar to those of neutral ER networks (fig-3.5 (a - c)). The rewiring process further enhances the already prominent core-periphery structure found in SF networks and the ED based strategies elicit a strong response in accordance with this. S4 and S5 affect a qualitative change in the responses of SCC and WCC, while the effects of random and centrality based strategies do not vary considerably from that of neutral SF networks.

The WCC shows some initial robustness to S4 type of deletion but as much as the neutral SF networks. This is due to the high concentration of links in the core. As more links with high values of ED are lost, links with intermediate values of ED (that connect the core and periphery) become more relevant. The rapid collapse seen in the intermediate stages of deletion can be attributed to the loss of these links. When the deletion process reaches the low ED edges in the periphery, there are fewer links and the breakdown is more controlled. During the early stages of link removal, the WCC and SCC show comparatively less resilience to S5 and in the later stages, the rate of change slows down. In neutral SF networks, S/S_0 remains equal to 1 until about 40% of the links are lost while in the assortative SF network, the SCC begins to breakdown after only 20% of the links are lost. In the later stages, when the process of deletion reaches the strongly connected core, the rate of breakdown decreases. This is seen in the behavior of W/W_0, S/S_0 and E/E_0 after about half the links have been lost. This is because, in the early stages, S5 targets and removes the links in the periphery, leading to isolation of those nodes. But the process of breakdown decelerates, in the later stages, when the attack reaches the densely connected core.

3.5.4 Role of dissortative rewiring in ER and SF networks

When dissortative correlations are introduced into the networks, links get reorganized to connect nodes of dissimilar degrees. With a greater fraction of links now connecting the core and periphery, the distinction between the two is blurred. This results in a decrease in density of links in the core. As a result, the distribution of ED becomes more homogeneous and is dominated by intermediate values. Once again, in ER networks, the rewiring process does not affect the response to loss of links considerably.

In scale-free networks (fig-3.6 (d - f)), random deletion and centrality based deletion evoke a response similar to that of neutral SF networks. However, the ED based strategies bring out the effects of reduced concentration of links in the core. Both, WCC and SCC show reduced resilience when links are removed in decreasing order of ED values (S4). When the links are deleted in increasing order of ED values (S5), the response is exactly the opposite of that seen in assortative SF networks. W/W_0 shows significant robustness initially and around $1 - M/M_0 \approx 0.6$, it begins to deteriorate more rapidly. The SCC, on the other hand, begins to breakdown very early and maintains a steady rate to the very end. The behavior of E/E_0 closely follows that of S/S_0.
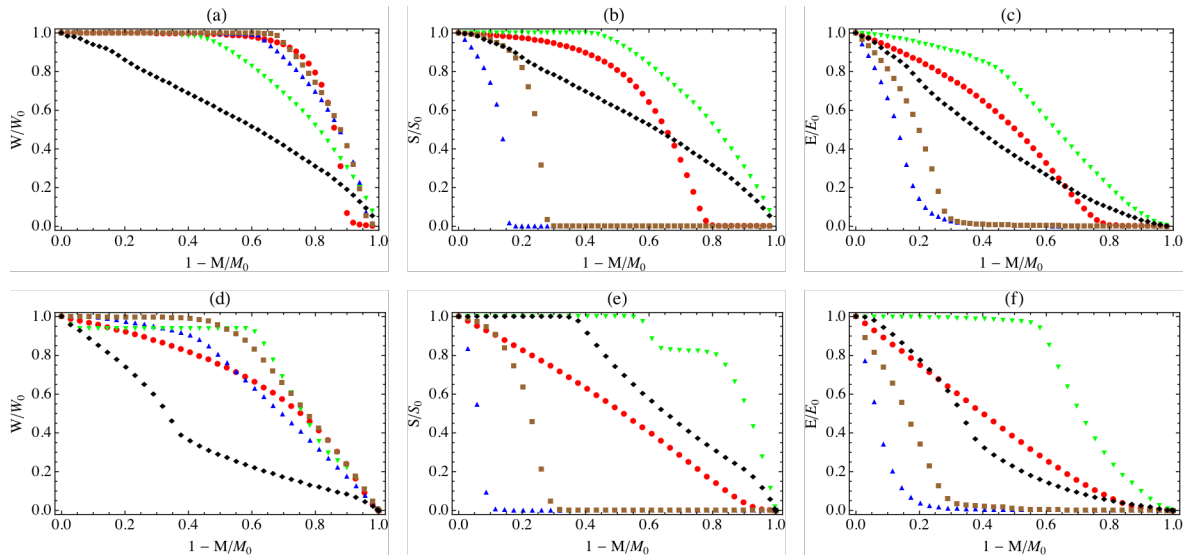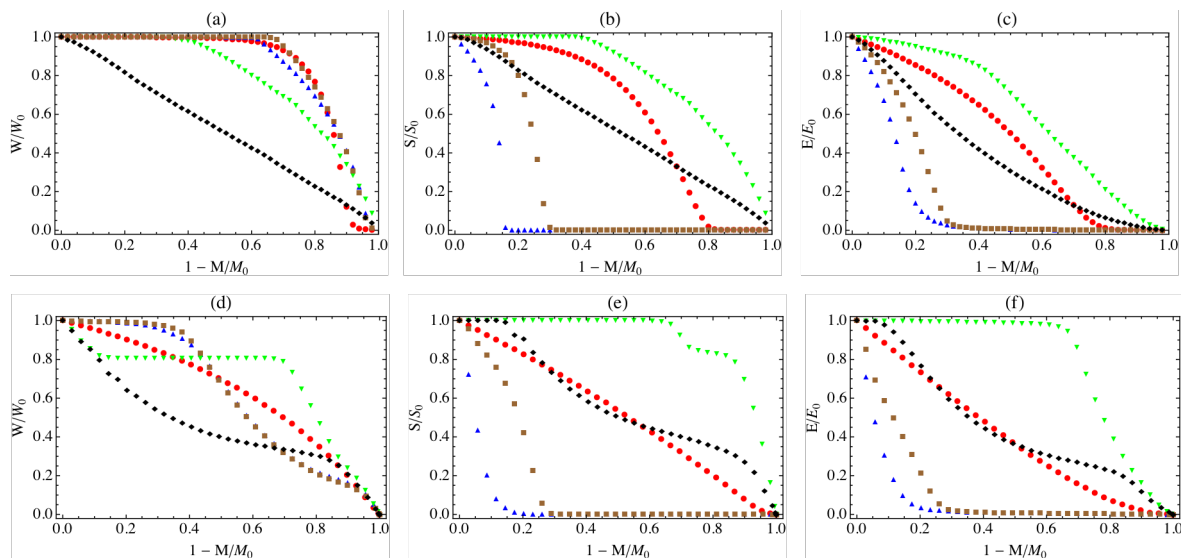
Figure 3.6: Performance of WCC (W/W_0), SCC (S/S_0) and Efficiency (E/E_0) as function of the fraction of removed links (1 - M/M_0), during various strategies of link deletion, in directed ER networks (a - c) and SF networks (d - f) with dissortative degree-correlations. S1 - red circles, S2 - blue up-triangles, S3 - green down-triangles, S4 - brown squares and S5 - black diamonds.

3.5.5 Role of clustering in ER and SF networks

The role of clustering, on the resilience of random networks to loss of links, is studied using the directed WS model (fig-3.7 (a - c)). The network is generated using a 3-regular lattice, in which the links are randomly rewired with a probability $p_r = 0.1$. When we consider an average over an ensemble of such networks, about 10% of the links in each network are rewired. This implies that the network could retain some lattice-like characteristics while simultaneously exhibiting small-world nature. As the probability of rewiring increases, the network becomes more random and the high values of EBC (characteristic of lattice-structures) drop by an order of magnitude. Since we rewire only 10% of the links, the distribution of EBC values retains a substantially large number of links with very high EBC values. As a result, the rewired networks remain prone to EBC based deletion strategies. Consequently, the SCC becomes highly vulnerable and we observe a 90% breakdown with the loss of just 10% of the highly central links. The WCC also begins to deteriorate after losing just 20% of the links. Right from the beginning, there are links with low values of EBC, due to which the there is no pronounced effect of S3 type of deletion.

When it comes to ED based strategies, we observe an increased resilience to deletion. Due to the lattice-like structure, where there is a large redundancy in paths with comparable path-lengths, both SCC and WCC show comparably higher robustness to loss of links with high ED values. However, the response to strategy S4 is less effective when compared to ER networks.

In SF networks, the effect of rewiring for clustering is clearly seen in the change in response to S5 type of deletion (fig-3.7 (d - f)). The WCC becomes significantly more robust while the SCC becomes less so. The behavior of E/E_0 is consistent with that of the SCC. However, during random deletion, the results are consistent with those of neutral SF networks. After rewiring, the network is still vulnerable to S2 and S4 types of deletion. In strategy S2, the SCC breaks down completely with the removal of just 20% of the links and the same happens during strategy S4 but with the loss of 30% of the links.
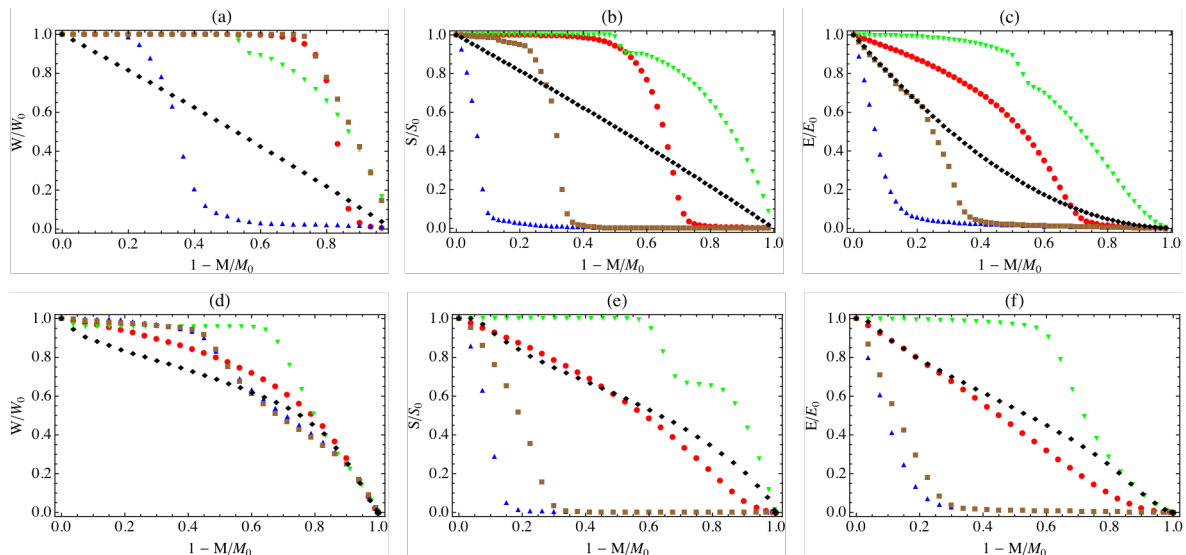
Figure 3.7: Performance of WCC (W/W_0), SCC (S/S_0) and Efficiency (E/E_0) as function of the fraction of removed links ($1 - M/M_0$), during various strategies of link deletion, in directed and clustered ER networks (a - c) and SF networks (d - f). S1 - red circles, S2 - blue up-triangles, S3 - green down-triangles, S4 - brown squares and S5 - black diamonds.

3.5.6   Analysis of real-world networks

So far, we have dealt with the results on network models. Now, we shift our attention to the results of random and targeted link deletion on three real world networks. The first network under consideration is based on traffic flow between road intersections in the city of Austin, TX (fig-3.8 (a - c)). The degree distribution is close to a poisson distribution and the only dominant property is the assortative nature of links. As a transport network, we expect it to be fully reciprocal, but due to the presence of one-way lanes and closed down roads, that assumption is not always true. Another important property is the presence of 1-node correlations as a result of important intersections having large to and fro connectivity.

It is immediately obvious that both the strongly and weakly connected components as well as the efficiency are extremely vulnerable to all strategies of link deletion. The SCC is extremely vulnerable to S2 and completely breaks down after losing only 10% of the links. However, it is more robust to S1 and S4 types of deletion. The behaviour of WCC is not very different that of SCC. It is very vulnerable to EBC based deletion strategies, suggesting that early losses could very well lead to isolated road intersections. The connected components are particularly susceptible to S5 indicating that the roads joining less popular intersections play a pivotal role in the overall connectivity of the city.

An analysis of this nature helps us to tackle any issues arising from the closure of roads and/or rerouting of traffic. Shutting down of select roads for maintenance reasons is an important example in this regard. An analysis based on link deletion studies helps us to identify roads for closure, such that the effect on overall connectivity is minimal. It also helps to plan for rate of progress of maintenance. Alternatively, in a scenario where certain roads become unavailable due to unexpected events like traffic jams or accidents, we can make informed decisions for rerouting traffic based on an understanding of the loss of further connectivity and traffic pile-ups.

Next, we look at the results of link deletion on the Protein Interaction Network. This network is scale-free and dominantly dissortative. As expected, the results on this
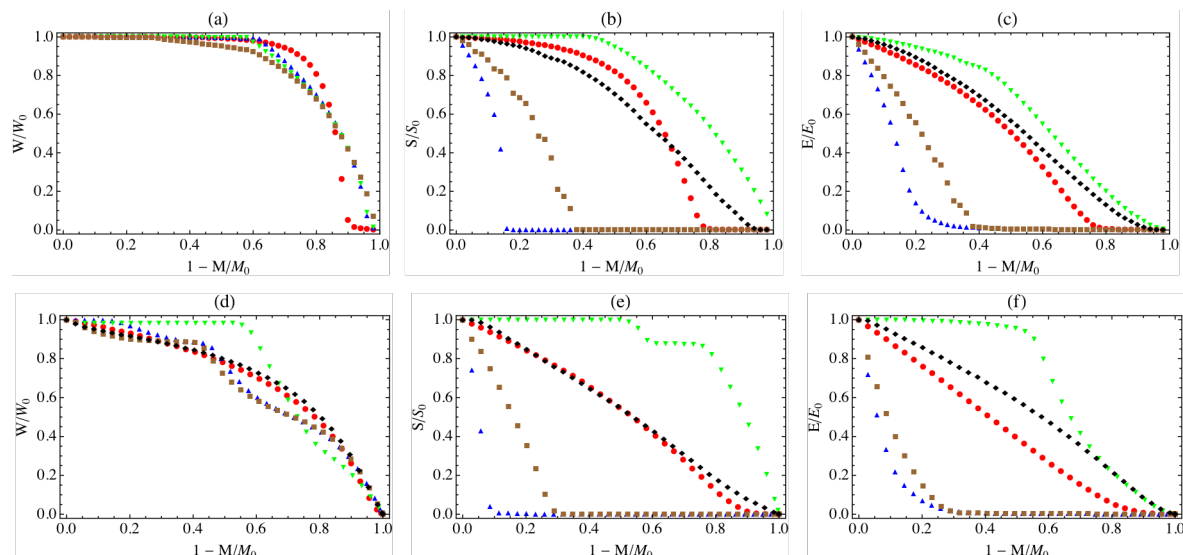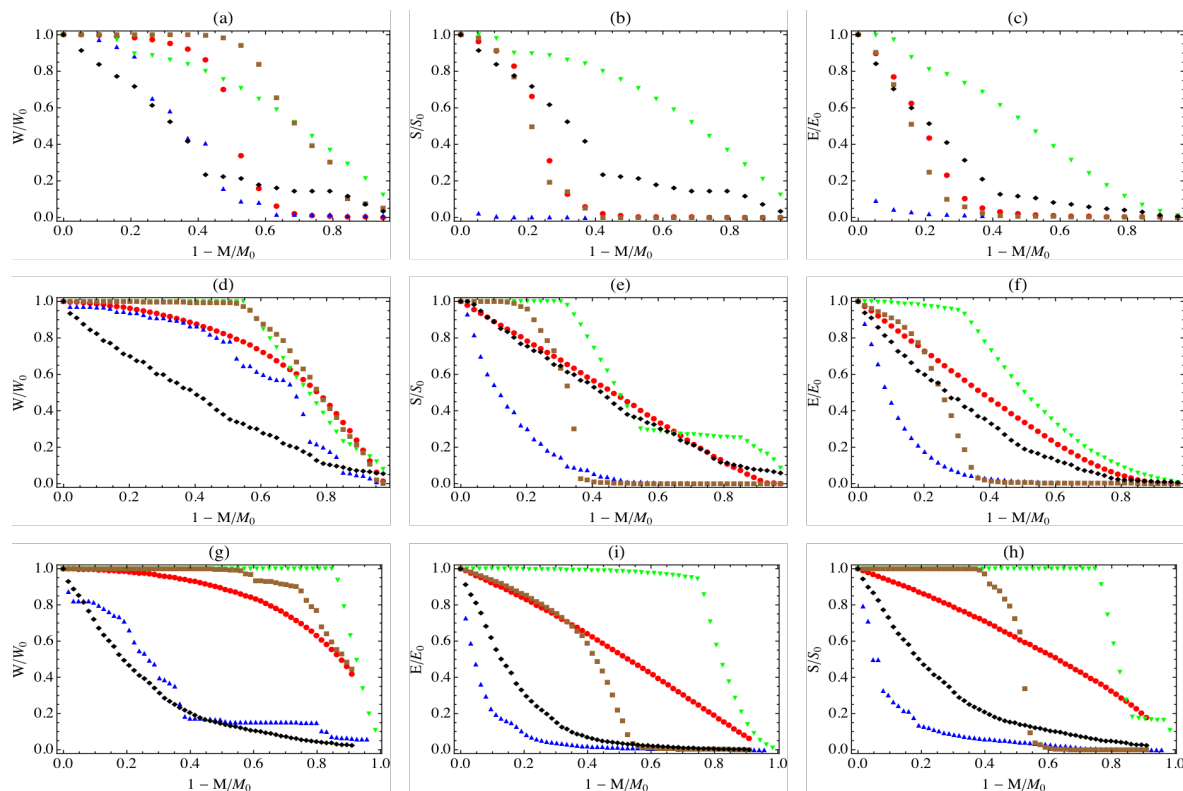
Figure 3.8: Performance of WCC (W/W_0), SCC (S/S_0) and Efficiency (E/E_0) as function of the fraction of removed links (1 - M/M_0), during various strategies of link deletion, in the road-intersection network (a - c), protein-protein interaction network (d - f) and the airport network (g - i). S1 - red circles, S2 - blue up-triangles, S3 - green down-triangles, S4 - brown squares and S5 - black diamonds.

network closely resemble those obtained during deletion in dissortatively rewired SF networks. (fig-3.8 (d - f)). During S1 type of deletion, the WCC disintegrates slowly while the SCC and Efficiency have a steady continuous decline. The decentralized organization in the network is brought out by the behaviour of WCC, which shows considerable resistance to S2 type of deletion. However, the SCC is not similarly resilient to S2 and undergoes an exponential breakdown. Its response to S3 is more resilient compared to its response to S2 and it remains robust until $\sim 40\%$ of the links are lost. The WCC appears to be most vulnerable to S5 type of deletion, where links connecting the nodes with lowest degrees are removed first, but moreresilient to strategy S4. This suggests that the proteins undergoing fewer interactions play an important role in sustaining the range of functional connectivity in the proteome. While the SCC is initially robust to loss of links with high ED values, it undergoes sudden and extensive disintegration soon after. Overall, in this network, the WCC exhibits more robustness as compared to the more vulnerable behaviour of the SCC. The changes in efficiency follow along the lines of SCC.

Finally, we come to the results of our study on the network of airports across the world (fig-3.8 (g - i)). The network is scale-free for both the in and out degree distributions ans also shows substantial clustering. As another example of a transport network, it shows substantial 1-node correlation and very low values of 2-node degree-correlations. As this network is subject to various strategies of link deletion, we analyze the results and attempt to validate them using results obtained earlier on corresponding network models.

During random deletion of links, the WCC and SCC show immediate and continuous

reduction in relative sizes. As seen in earlier results for directed scale-free networks, the connected components show continuous deterioration and the threshold for breakdown is absent. The WCC shows strong resilience to strategy S4 and W/W_0 retains a value very close to 1 until about 80% of the links are lost. The WCC undergoes an exponential decrease, after showing little resilience to centrality based deletion of links. Strategy S5 is much more effective on this network as compared to the clustered SF mode or the WS model. The change in SCC is remarkably close to that of the WCC. The SCC is initially robust to S4 type of deletion but not as much as the WCC. With the loss of links with high ED, the SCC undergoes a sharp transition after loss of 50% of the links. We observe an immediate exponential decline when links with high values of betweenness centrality are lost.

The behaviour of efficiency, most conspicuously brings out the susceptibility of the airport network to targeted removal of links. This observation implies that connectivity is not fatally affected when flights between major cities are cancelled. Rather, it leads to a pileup of passenger volume at airports while making travel less efficient and more cumbersome. On the other hand, cancelling flights between smaller cities, for any reason, presents a strong risk of isolating them from the overall network. In this network, the links connecting the hubs to the smaller airports play the most central role. There is immediate and extensive damage, to both the extent and efficiency of connectivity, when such links are lost.

3.6   Summary

In this chapter, we present the results of a comprehensive study on the effects of link deletion in directed networks. We consider ER and SF networks and subject to them to various strategies of link deletion. In particular, we consider random deletion and EBC and ED based strategies. We also study the roles of 2-node degree-correlations and clustering, on the resilience to link deletion, in both ER and SF networks. These properties are independently introduced into initially neutral networks by means of degree preserving rewiring mechanisms.

As the process of deletion progresses, the changes occurring in the network are analyzed by studying the behaviour of the fractional sizes of the strongly and weakly connected components and the efficiency. Irrespective of the strategy of deletion, ER networks consistently showed more robustness than SF networks. In ER networks, assortative correlations do not make any difference while dissortative correlations lead to a decrease in the robustness of the SCC and an increase in the case of WCC. However, in SF networks, both types of correlations lead to an decrease in the resilience of the connected components. Presence of clustering increases the vulnerability of both ER and SF networks.

We find that the loss of highly central links has the most damaging effect on any network topology while the removal of the least central links is the least effective. The effects of random deletion are quite moderate when compared to ED and EBC based strategies and are different for different topologies. The analyses of the responses of real-world networks, chosen from the broad areas of transportation and biological interactions, have implications on the definitions of strategies for deletion. They also play an important role in the design of alternate measures to counter or repair the damages of loss of connectivity in the networks.

The response of networks to EBC based strategies of deletion is approximately reproduced using only local information (node degrees) about the networks. A combination of degrees of nodes on either ends of a directed edge, given as $k_i^{in}.k_j^{out}$, is found to be

a suitable choice for the definition of ED. For large values of ED and EBC, we observe a strong positive correlation. This is further supported by the response that strategies S2 and S4 evoke in the various network models. Therefore, ED based strategies can be used as a computationally viable alternative to launch highly effective targeted counterattacks on networks to contain unwanted spreading processes.

In this study, while the strategies for deletion are chosen quite arbitrarily, the network models are chosen to represent a broad classification of real-world networks. The results in this chapter are meant to provide a qualitative picture of the responses of networks to link deletion and the roles of various network properties in an idealistic scenario. Exact results for real-world networks require a narrow and focused effort and may diverge considerably from those of network models.

Chapter 4

Dynamical response to link deletion

4.1 Introduction

In the previous chapter, we have dealt with the effects of random and targeted link deletion on the structural aspects of directed networks. In this chapter, we focus on the dynamical response to link deletion. We investigate the effects of loss of links, and the resulting structural responses, on a 1-to-1 transmission process occurring on the network. While the process of 1-to-1 transmission is implemented as a model of ping-generation and propagation, the mechanism can be generalized to study particular kinds of flows in a network. This process is also significantly different from most existing methods that are used to study spreading processes [150–154]. With suitable modifications and interpretation, this model can be used to represent many important applications. For example, the pings could be made to represent packets of data, on the internet, that travel between routers or they could be representing pulses of current flowing along well-defined transmission lines between input and output gates on a digital circuit [155, 156]. Another interesting example is the case of reaction networks, where the pings represent chemicals/compounds, that transform from reactants to products by following a sequence of pre-determined steps [157, 158].

However, in this work, we focus our efforts on the study of traffic flow in transportation networks. In particular, we are interested in traffic flow on road networks [159–162]. In this example, the pings represent traffic (people/vehicles), travelling from origin to destination locations. While there has been substantial work done on transportation networks and their vulnerabilities and robust design, the vast majority of it takes the approach of congestion-analysis caused by overloading of nodes [163–166]. We overlook the problem of congestion and try to focus on the role of the underlying network structure. The aim is to study the changes in the dynamics of the transmission process as a response to the deletion of links. To the best of our knowledge, the proposed kind of study has not been undertaken, especially using models of directed networks. We find that it leads to a better understanding of the relationship between the structure of the network and the dynamics occurring on it.

In this chapter, we start by introducing the two processes occurring on the network and define the respective process parameters. Then, we present the results on some standard network topologies and also discuss the effects of 2-node degree-correlations in each case. To bring out the relationship between the two processes, we provide qualitative arguments and supporting numerical results. We study, independently, the roles of three different process parameters, in the context of a real-world road network. Finally, we propose a scheme, based on weighted selection of nodes, to improve transmission during link deletion.

4.2 Co-occurring processes

Having introduced and motivated the aim of this work and its potential applications, we now discuss each of the co-occurring processes in greater detail and also introduce the terminology and definitions. We first detail the process of link deletion, followed by the transmission process, so as to emphasize the unidirectional influence of the former on the latter. A schematic depiction of the processes is given in fig-4.1.

4.2.1 Link deletion

In this section, we discuss the process of link deletion in greater detail. We start by defining two parameters of the deletion process, namely the rate of deletion (v_D) and the probability of deletion (p_D). The rate of deletion (v_D) is defined as the number of selected links that can be deleted at any time-step and p_D is the probability with which each of the v_D selected links are deleted.

Therefore, at every time-step, v_D number of links are chosen according to one of the three strategies as discussed in the previous chapter. To briefly summarize, links are deleted using one of three strategies: S1 - Random deletion, S2 - Edge Betweenness Centrality (EBC) based deletion and S3 - Edge Degree (ED) based deletion. In strategy S1, v_D number of links are chosen uniformly randomly from the existing pool of links and each of them is deleted with a probability p_D. In strategy S2, all the existing links are ordered in decreasing order of their EBC values and the most central links are selected as v_D and deleted with probability p_D. In strategy S3, v_D links with the highest values of ED are selected and deleted with a probability p_D.

At every time-step, after link deletion has occurred, the change in the networks structure is quantified by calculating the size of the largest strongly connected component (SCC) and the average path-length in the SCC. The size of the SCC gives an estimate of the extent of connectivity in the network while the latter points to the efficiency of connectivity. Further reasons for calculating the average path-length of the SCC as opposed to that of the entire network are discussed in sec-4.4. It is also worth noting that there is no unique way in which the deletion process depends on the transmission process but rather only on the structural properties of the network.

4.2.2 1-to-1 transmission

Occurring simultaneously with link deletion is the process of 1-to-1 transmission on the network. At every time-step t, a fixed number of requests/pings are generated with randomly chosen start and stop nodes. This number is defined as the $request - rate$ r. The shortest (most efficient) path between the start and stop nodes is calculated at t and assigned to the respective pings. If, for any ping, no valid path can be found, it is terminated and characterized as a $failed$ transmission. If multiple such paths exist for a single ping, one of them is assigned at random. This pre-assignment of specified paths to each ping differentiates this study from those based on random-walk models. Once transmission starts, at every successive time-step, the pings can traverse a maximum number of links, given by the $transmission\ velocity$ (v_T), with a $transmission\ probability$ (p_T). The idea behind introducing a separate probability for transmission is to effectively implement traversal of fractional edges.

During transmission, if the subsequent link in the currently assigned path is lost due to link deletion, the ping is rerouted via a new shortest path calculated from the current location of the ping to the stop-node. If no such valid path can be found, the ping is terminated as a $failed$ transmission. A ping is considered $successful$ if it
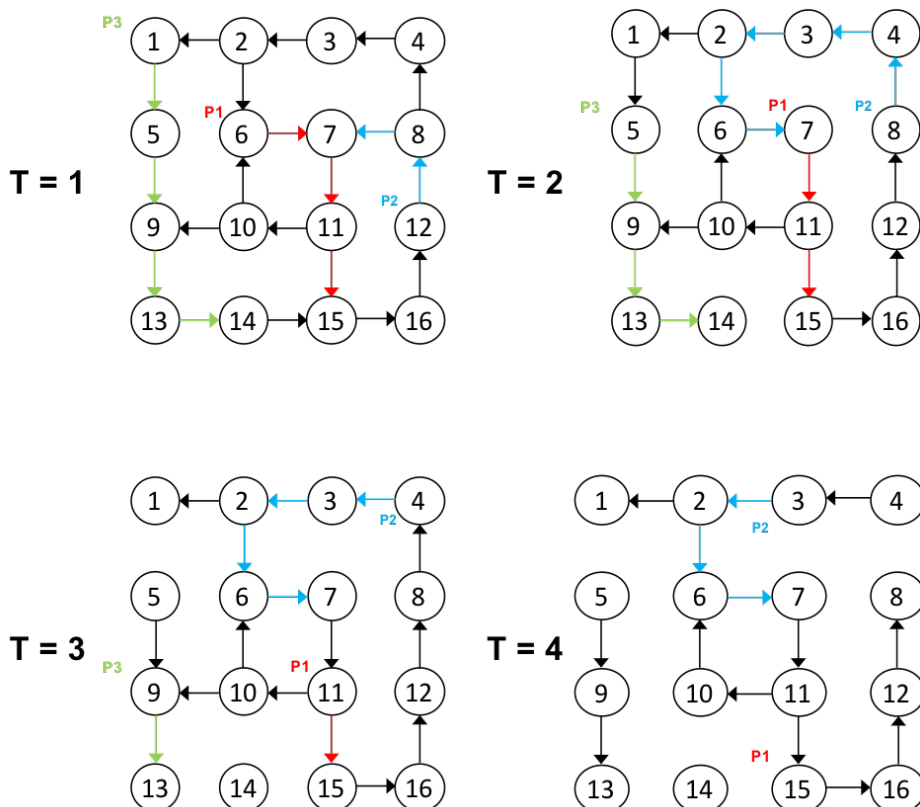
Figure 4.1: Schematic representation of the co-occurring deletion and transmission processes. At $T = 1$, three pings (P1 - red, P2 - blue and P3 - green) are generated and the pre-assigned paths from start to stop are highlighted in the corresponding colors. at $T = 2$, two links are deleted at random and in the process, P2 gets rerouted via a new longer path. At $T = 3$, after two more links are deleted, the path of P3 gets disrupted and no alternate path is available and therefore, it is a failed transmission. At $t = 4$, P1 reaches its destination and is a successful transmission.

gets transmitted from the start-node to the stop-node, either by the originally assigned path or via a rerouted path.

The time taken by a ping to travel from the start-node to the stop-node is defined as the *ping time* and is calculated as the difference between the time it is generated and the time it reaches the stop-node. For successful pings, the *transmission efficiency* is calculated as the ratio of the predicted *ping time* to the actual *ping time* associated with the path assigned during generation. This gives a measure of the extent of rerouting.

4.3   Results on standard network models

In this section, we present results on networks from standard models such as directed Erdös-Rényi (ER) networks, directed scale-free (SF) networks and directed small-world (SW) networks. The process starts with a network size N (= 1000) and average degree equal to 4. At every time-step, links are deleted according to a given strategy. Simultaneously, new pings are generated while existing pings traverse their respective prescribed routes. This sequence of events continues until the number of links remaining is equal to or less than half the number of nodes in the network. The effect of link deletion on the transmission process is characterized in terms of the number of successful pings and the normalized ping times. All results are averaged over 200 realizations. The process parameters v_D, p_D, v_T and p_T are all set to 1 and the *request rate* $r = 5$.

4.3.1   Role of topology and strategy
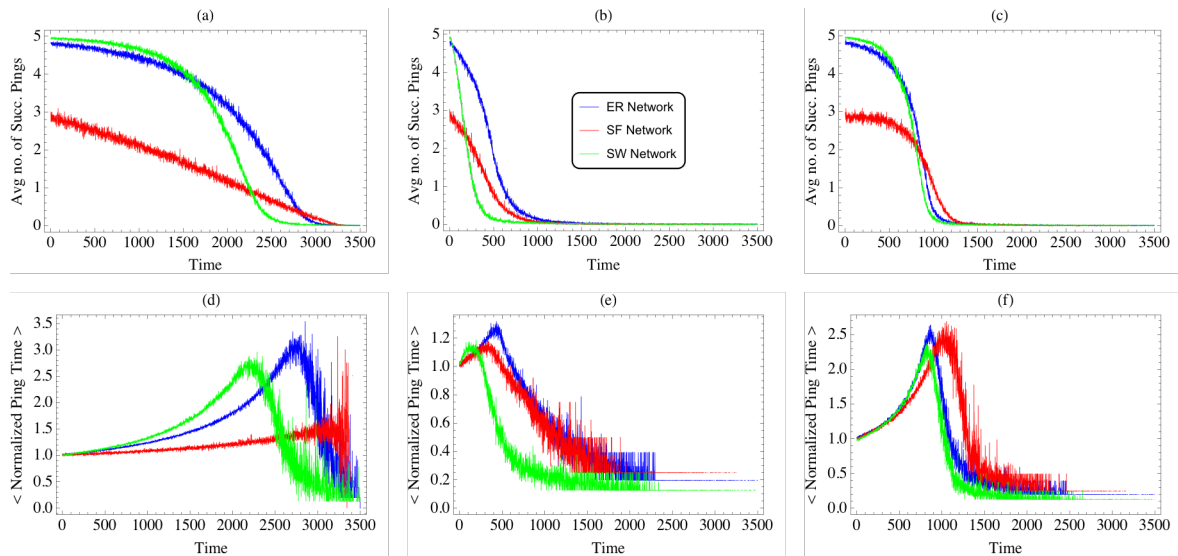
Figure 4.2: *Top* - Decrease in the average number of successful pings with time (or equivalently loss of links) in ER networks (blue), SF networks (red) and SW networks (green). *Bottom* - The change in behaviour of ping-times during loss of links in the three types of directed networks models. The above quantities are studied for S1 - random deletion (a, d), S2 - EBC based deletion (b, e) and S3 - ED based deletion (c, f).

We start by looking at the effects of topology (fig-4.2) and strategy of deletion (fig-4.3) on the transmission process. During random deletion, in ER and SW networks, the transmission process exhibits very similar initial behavior, in terms of the number of successful pings. In both networks, there is a high volume of successful pings until about the halfway point, beyond which transmission fails at an accelerated pace in the SW network. As a result, complete failure of the transmission process, occurs earlier in SW networks. This is also reflected in the behaviour of normalized ping-times, which show an exponential increase, with the rate of increase consistently higher in SW networks than in ER networks.

From our understanding of the link deletion process from the previous chapter, we know that two types of changes take place in the network in parallel - (a) Among the set of connected nodes, there is an increase in the path-lengths and (b) appearance of disconnected nodes. Therefore, this increase in ping-times must be indicative of the increasing path-lengths in the networks and implying that while both ER and SW networks sustain a similar number of successful pings initially, the pings are traversing longer paths in the SW network in order to be successful. We see that there is a qualitative change in behaviour of ping-times in the latter half of the deletion process. The exponential increase in ping-times indicates the dominance of increasing path-lengths while the region of exponential decrease marks the dominance of disconnected nodes. Therefore, the occurrence of the peak coincides with the breakdown of the transmission process. Meanwhile, in the SF network, there is relatively smaller number of successful pings but the decrease remains consistently linear until complete breakdown. The normalized ping-times also remain very close to 1, demonstrating the limited role of increasing path-lengths and consequently the limited role of rerouting.

When links are deleted by strategy S2, transmission is affected most aggressively in SW networks where successful transmissions decrease exponentially from the very beginning. In comparison, transmission in ER and SF networks decreases linearly at the

start but later becomes exponential although in SF networks there is a relatively smaller number of successful pings even at the start. Similarly, the normalized ping-times also show very little influence of increasing path-length and switch to exponential decrease at very early stages in the deletion process implying an abundance of disconnected nodes.
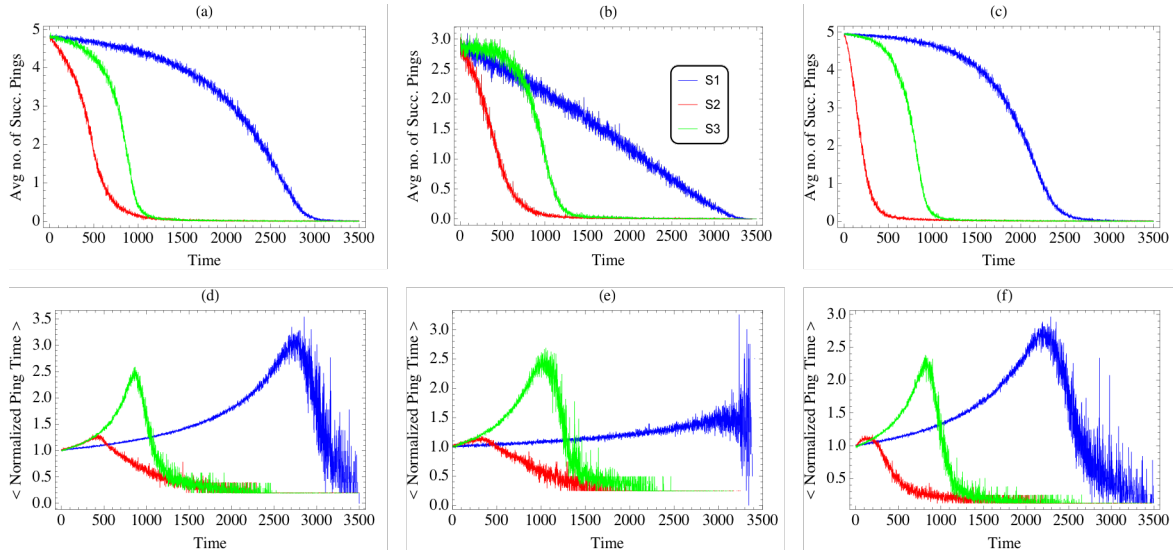
Figure 4.3: *Top* - Decrease in average number of successful pings/transmissions for various types of link deletion (S1 (Blue) - Random deletion, S2 (Red) - EBC based deletion and S3 (Green) - ED based deletion). *Bottom* - The change in behaviour of average ping-times during loss of links for various types of deletion strategies. This is studied in ER networks (a, d), SF networks (b, e) and SW networks (c, f).

During S3 type of deletion, the effect on the number of successful pings and the normalized ping-times is very similar for all three network topologies. In the early stage of deletion, there is a sustained number of successful pings which then undergoes a rapid decrease at almost the same time in all topologies. Even in the case of ping-times, almost equal increase in all the networks. This suggests that the network topology may not play an important role on the effect of ED based deletion on 1-to-1 transmission in the network.

4.3.2 Role of correlations

In this section, we study the effect of 2-node degree-correlations in directed ER and SF networks. Neutral networks with desired topologies are first generated and correlations are introduced into them by process of degree-preserving rewiring (DPR) mechanisms, as proposed in chapter-2. The resulting networks are subject to various strategies of link deletion and the results are presented in fig-4.4.

During random deletion in ER networks, dissortative rewiring does not affect any substantial change in the average number of successful pings or the normalized ping-times, as compared to the neutral case. However, during assortative rewiring, there is visible decrease in the number of successful pings but these pings also travel more efficiently as indicated by the smaller rate of increase in ping-times. In SF networks, assortative rewiring has a qualitatively similar effect but the decrease in number of successful pings is quite significant. Dissortative correlations, on the other hand, lead to a small increase in successful pings but they become less efficient, especially in the later stages of deletion.

61
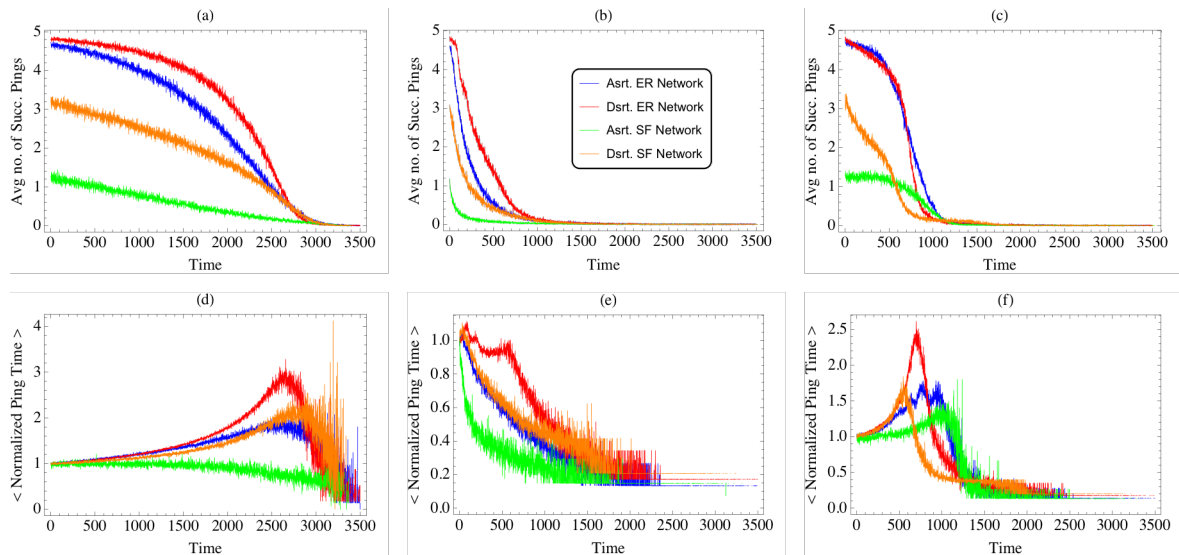
Figure 4.4: Change in average number of successful pings/transmissions (top) and average ping-times (bottom) during S1 - (a, d), S2 - (b, e) and S3 - (c, f) types of deletion strategies in assortatively and dissortatively rewired ER and SF networks.

When links are deleted based on their betweenness centrality values (S2), both assortative and dissortative rewiring result in decreased number of successful pings irrespective of topology but the rate of decrease is faster in case of assortative correlations. In the neutral case, the role of longer, less efficient paths is mildly evident as seen from the small initial increase in the ping-times while in assortative networks, the transmission process is immediately affected by the appearance of isolated nodes. Transmission in dissortative ER networks, however, shows a small initial robustness to deletion.

When links are removed by strategy S3, neither assortative nor dissortative correlations have any significant impact on the number of successful transmissions in ER networks. The normalized ping-times, however, are reasonably affected. There is considerable decrease in assortative ER networks while in dissortative ER networks, the change in ping-times is faster compared to the neutral case. In assortative SF networks, there is considerable reduction in number of successful pings but the behaviour is qualitatively similar to the neutral SF network. The normalized ping-times are also much lower. Dissortative rewiring in SF networks leads to linear decrease in successful pings while the behaviour of ping-times is similar to the neutral SF network.

4.4 Relationship between link deletion and transmission

Since we have two processes taking place on the network; link deletion - which directly affects the network structure and transmission - which is affected by the structural changes in the network, and they are both parameterized by the same time variable, we can study the transmission process as a function of the deletion process to gain insight into the relationship between the two processes. The results are shown in fig-4.5 where we plot the fraction of successful pings and the normalized ping-times as a function of fraction of links removed, which is an indicator of the extent of link deletion.

We first explore the relation between the size of the Strongly Connected Component (SCC) and the number of successful pings. For a ping to be successful, at least one path must exist from the start-node to the stop-node and must continue to exist (subject to
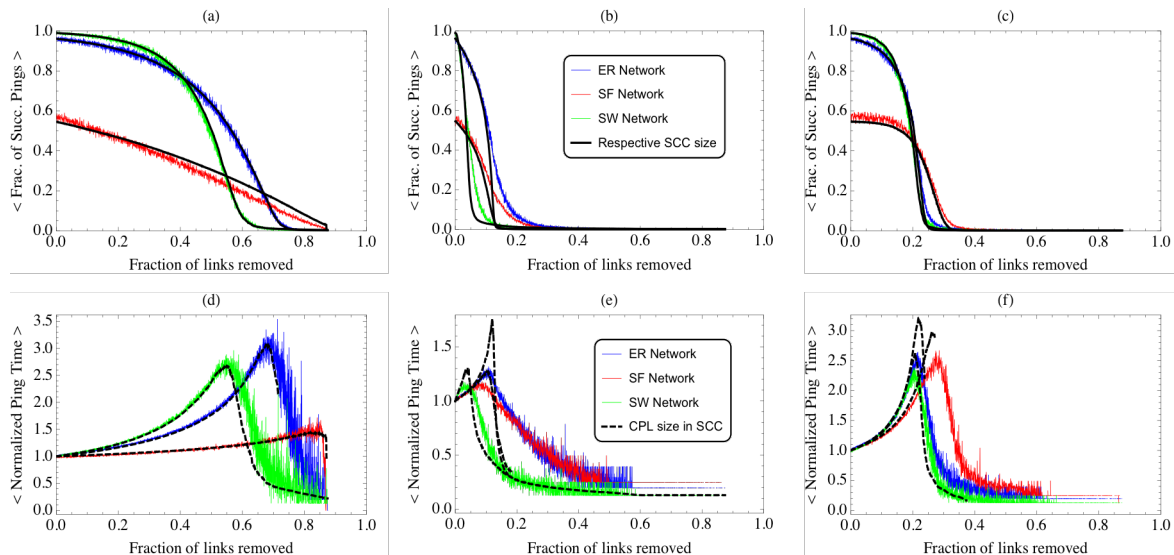
Figure 4.5: Relationship between the link deletion process and the transmission process is explored as the correlation between the average fraction of successful pings and fractional SCC size of the network when plotted as a function of fraction of deleted links (top). We also see that the average path-length of the SCC captures the behaviour of average ping-times as the process of deletion progresses (bottom). This observations remains consistent for all network topology and strategy of deletion.

rerouting) until the stop-node is reached. But a unique path that runs in the opposite direction is not necessary. From a network point of view, such a condition is satisfied when the end-points belong to the Weakly Connected Component (WCC) but not necessarily to the SCC. Therefore, it would seem that the probability of a ping being successful is proportional to the probability that the start and stop nodes belong to the WCC and therefore on the fractional size of the WCC itself. But when we consider the fact that during generation of pings, the start-nodes and stop-nodes are chosen uniformly randomly, it becomes necessary for paths must exist in to and fro directions. This is because it is possible to assign the two unique paths to two unique pings. In this scenario, paths between all possible pairs of nodes must be considered. As a result, the probability of a ping being successful is proportional to the probability that the start-node and stop-node belong to the SCC and consequently on the fractional size of the SCC itself. Therefore, the fractional size of the SCC (which is itself varying with the number of links in the network) captures the average behaviour of the probability of successful transmission. This is true irrespective of network topology and strategy of deletion.

If R is the number of pings generated at time t and S is the number of successful pings, then the probability of success for a ping generated at t is given by

$$\frac{S}{R} = f(t) \tag{4.1}$$

where f is the time-dependent probability and is equal to the fractional size of the SCC. If we consider the pings generated in an interval of time Δt, then we have $S/R = f(\Delta t)$ and $f(\Delta t)$ is the fractional size of the SCC that remains constant in Δt. Eqn - 4.1 can then be written in terms of the *request rate r* as

$$S = f r \Delta t \tag{4.2}$$

In the context of the above argument, we move to infer the relationship between the ping-times and the average path-length (APL) in the SCC. Since the probability of

successful transmission is equal to the fractional size of the SCC, it is evident that the ping-times must correlate to the characteristic length scale of the SCC. Since certain pings may not be contained within the SCC (may lead into or out of the SCC), the average behaviour of normalized ping-times is captured exactly by the normalized average path-length of the SCC. This is also an added advantage because the path-length of the SCC remains well-defined during the deletion process in contrast to the path-length of the whole network. Once again, this result is independent of deletion strategy and network topology but dependent only on the random selection of start and stop nodes.

For a ping generated at time t and travelling a distance d_p, let the time taken be t_p. Then

$$t_p = \frac{d_p}{p_T . v_T} \tag{4.3}$$

If we consider many such pings at t, and given that p_T and v_T are fixed, then the average ping-time is given by

$$<t_p> = \frac{<d_p>}{p_T . v_T} = \frac{<APL_{scc}>}{p_T . v_T} \tag{4.4}$$

4.5 Role of process parameters

We now shift our attention to the roles of the different process parameters involved. As mentioned earlier, the parameters involved are v_D and p_D (associated with the deletion process) and v_T, p_T and r (associated with the transmission process). We focus on the independent study of r, p_T and p_D, since they play the essential role in capturing the dependencies, while v_T and v_D provide for greater control. We do not consider the parameters of the network, like size and density, because the metrics of interest are normalized so as to make the results independent of them.

We also move away from the network models and instead work with a road-network based on the city of Barcelona. The data-set consists of 1020 nodes and 2522 links where nodes represent road intersections and links are the traffic-flows between these intersections. The network is largely lattice-like, moderately assortative with negligible clustering and the largest SCC spans 929 nodes.

The problem of 1-to-1 transmission can be recast as a model for the flow of road traffic in some geographic setting. In our specific case, every trip taken by a vehicle is associated with unique origin and destination. Since the network itself has well-defined nodes, the exact origins and destinations can be approximated to the nearest intersections. The combination of v_T and p_T captures the distance travelled including fractional lengths of connecting roads. Inaccessible roads are modelled as deleted links. This could happen due to various causes, like traffic-jams, repair-works, accidents etc., that are either systematic or random. When a vehicle reaches an inaccessible road, the driver can ideally make a decision to take a possible (and less efficient) detour or cancel the trip. While cancelling an ongoing trip seems unrealistic, it is incorporated to stay consistent with the case that deletion of links is irreversible in the model. We now explore the role of each parameter and interpret the results in the current context.

4.5.1 Role of request-rate

The *request rate* r is defined as the average number of pings being generated in unit time. At time t, $r(t)$ is the average number of trips that start at t. Since, we work

with a simplified model, where there is no constraint on the number of vehicles at a particular intersection or a connecting road, an increase in the number of trips at starting t should result in proportional increase in the number of successful trips while the probability of successful transmission remains constant for fixed r.
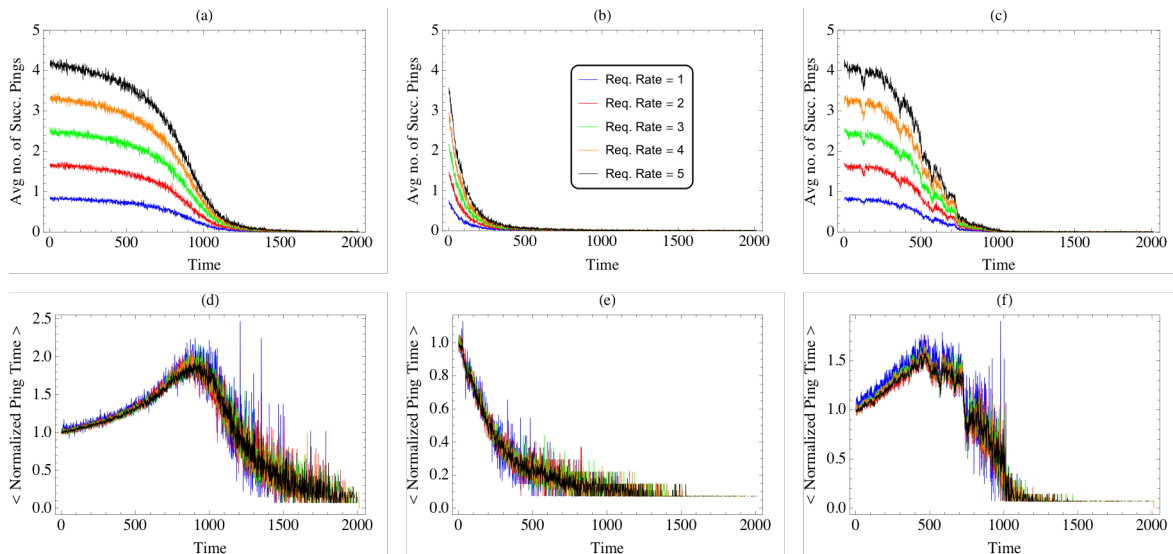
Figure 4.6: Effect of changing request rate on the average number of successful pings (top) and the average normalized ping-times (bottom) during link deletion by various strategies (S1 - a, d; S2- b, e; S3 - c, f) in the Barcelona road network.

From eqn-4.2, for unit time-interval, we have $S(t)$ proportional to r for given t. $S(t)$ is related only to the transmission process and $f(t)$ is related only to the deletion process via the surviving fraction of links ($m(t)$). We take r to be a fixed integer value for the entire duration of deletion but it could also be a set of integers drawn from a poisson distribution with fixed mean. While it mimics a more realistic scenario, it adds no new understanding. Since, r affects only the transmission process and not $f(t)$, the temporal behaviour of $S(t)$ is dependent on $f(t)$ but for a specific t, $S(t)$ is larger by a factor of r. This is seen in fig-4.6 (a - c).

The behaviour of f as a function of fraction of surviving links in the network (m) is specific to the strategy of deletion but the effect of r itself is independent of strategy. For $p_D = v_D = 1$, the qualitative behaviour of the slope of S is governed by the rate of change of $f(m)$, but the exact value is scaled by a factor of r. Therefore, for higher values of r, the rate of decrease of S is proportionally higher as is seen in fig-4.6. From eqn-4.4, we see that the average normalized ping-times do not, in any way, depend on the value of r. While the results are not specified here, we do find that there is no effect of changing r on the behaviour of normalized ping-times.

4.5.2 Role of transmission probability

In the transmission process, each ping travels v_T number of links at every time-step, with a probability p_T. The value of p_T sets a time-scale for the transmission process relative to the deletion process. A value of p_T less than 1 implies that a ping will take more time to traverse the same distance. In the context of traffic-flow on a road network, $p_T < p_D$ is not very realistic and points to possible case of rapid spread of cascading traffic jams. As a result, the time-interval in which f remains constant decreases by p_T relative to the rate of deletion and eqn-4.2 becomes

$$S = fr(p_T \Delta t) \tag{4.5}$$

The above equation can be understood as the decrease in the probability of success of a trip as a result of the network losing links faster than the pings can traverse them. But if we plot the results, as shown in fig-4.7 (a - c), we do not see the expected decrease in the number of successful pings. In fact, it appears like transmission probability does not play a role at all. This is where the effect of rerouting comes into play.

Figure 4.7: Effect of decreased probability of transmission on the average number of successful pings (top) and the average normalized ping-times (bottom) during link deletion by various strategies (S1 - a, d; S2 - b, e; S3 - c, f) in the Barcelona road network.

When a trip is started at t, and $p_T < 1$, there is a higher chance that the assigned path will be disrupted by a deleted link. But this loss of links is well-compensated by the rerouting process with exploits the large number of redundant paths available in the network. Consequently, the trips are successful but they take longer, less efficient paths.

When it comes to ping-times, p_T plays a very direct role on the response to deletion. From eqn-4.4, it is clear that at t, values of $< t_p >$ get scaled by a factor of $1/p_T$, as is shown in fig-4.7 (d - f). This also provides justification for the qualitative argument that trips take longer, less efficient paths but are ultimately successful. From the same set of plots, we also observe that the rate of change of $< t_p >$ is constant for decreasing values of p_T. This becomes clear when we take the time-derivative of eqn-4.4 and see that it is inversely related to p_T. But we also know that the distance travelled by a ping generated at t increases proportional to p_T and therefore, the expression for the slope of $< t_p >$ is independent of p_T. If the ping-times were to be normalized, they would all overlap and this behaviour would not be obvious. For this reason, we work with the actual ping-times.

4.5.3 Role of deletion probability

The probability of deletion (p_D) is the fraction of chosen links (v_D) that are deleted at any given time-step. It can be interpreted as the time-scale of the deletion process w.r.t the transmission process. For $p_D = 1$, all the v_D chosen links are deleted in each time-step. A value less than 1 implies that only a fraction, p_D, of the v_D chosen links

are deleted. Conversely, the time interval required to delete v_D links increases by a factor of $1/p_D$. As a result, eqn-4.2 becomes

$$S = \frac{fr\Delta t}{p_D} \tag{4.6}$$

Consistent with eqn-4.6, we observe an increase in the average number of successful trips, by a factor of $1/p_D$, at any t (fig-4.8 (a - c)). This is true for any strategy of deletion, over the behaviour imposed by the rate of change of $f(m)$. For $r = 5$ and $v_D = 1$, we see that the rate of decrease of $S(t)$ slows down by a factor of p_D and this is observed for all strategies as shown in fig-4.8 (a - c).
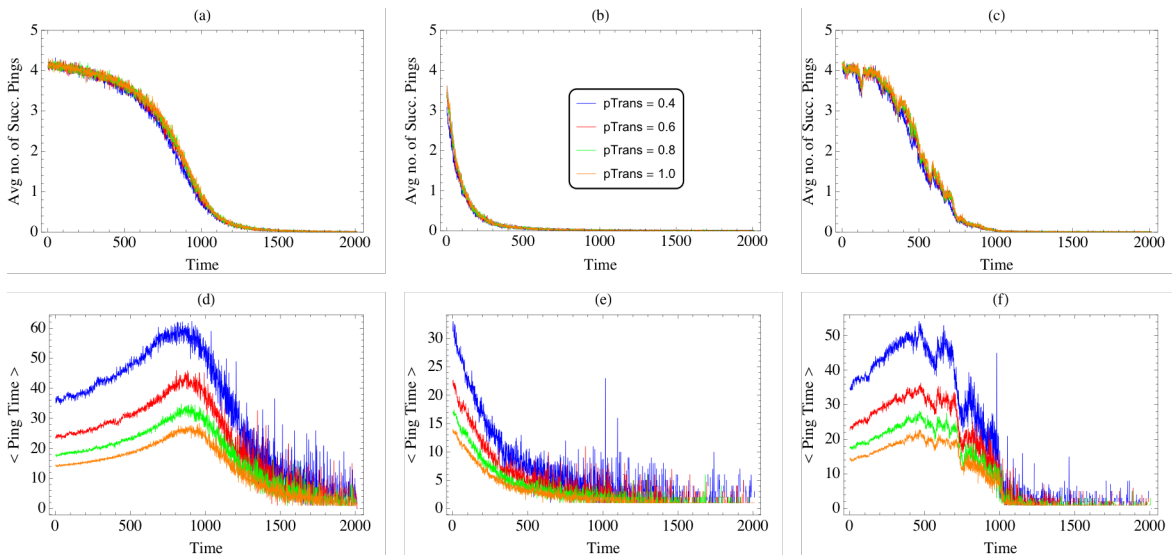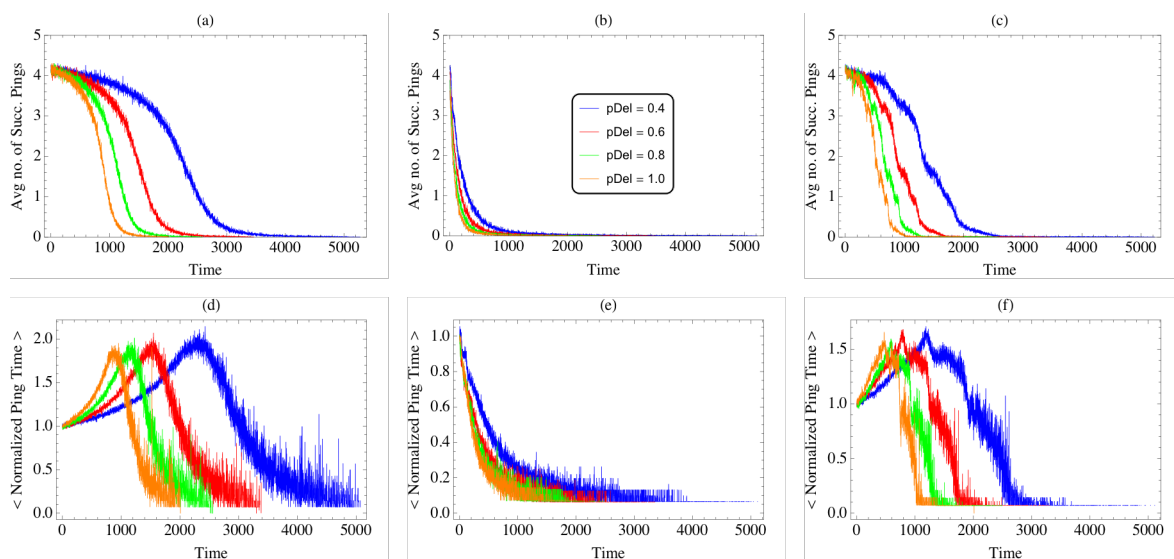

Figure 4.8: Effect of decreased probability of deletion on the average number of successful pings (top) and the average normalized ping-times (bottom) during link deletion by various strategies (S1 - a, d; S2 - b, e; S3 - c, f) in the Barcelona road network.

While a value of $p_D < 1$ gives trips more time to traverse the assigned path, it does not affect the lengths of individual paths in any way. As mentioned earlier, at any t, there is an increase in $S(t)$. The increase is due to those trips that are able to traverse longer rerouted paths and reach their destinations. Consequently, the average distance traversed by all trips starting at t increases and is reflected as an increase in the normalized ping-time at t. This is shown in fig-4.8 (d - f). We also see that the rate of increase of $< t_p >$ is slower for smaller values of p_D. This is because the slow rate of deletion gives the pings more time to travel the shorter paths before they get rerouted through longer less efficient paths. This can also be interpreted from eqn-4.4. The APL of the SCC depends on the density of links in the SCC ($\overline{k}_{scc}$) and also on the size of the SCC itself. But, numerically, we observe that, before the appearance of isolated nodes, the average degree of the SCC remains quite constant and therefore the rate of change of $\overline{k}_{scc} = 0$.

Since the rate of change of $f(m)$ is a monotonically decreasing function, the overall slope becomes positive and is scaled by a factor p_D. This explains the behaviour of normalized ping-times in fig-4.8 (d - f), in the region dominated by increasing path-length. When isolated nodes begin to appear, this argument becomes invalid because the average degree of the SCC does not remain constant anymore.

4.6 Strategy for mitigation - Weighted Selection

In the final section of this chapter, we propose a method to mitigate the effect of deletion on the 1-to-1 transmission process. A reasonable way to arrive at such a method is to first investigate the process by which deletion affects transmission and then try to counteract the effect. Working in this direction, we find that the process of link deletion affects transmission via the elimination of shortest-paths.

Having identified the importance of shortest-paths in sustaining transmission, we need to design a course of action by which the number of these paths or their lengths can be suitably optimized. We find that there are two main ways to ensure this: (a) select an appropriate topology (like SF or SW) so that the average path-length in the network is small thereby increasing the probability of successful transmission and (b) given a network topology, select pairs of nodes such that the number of shortest-paths between them is significantly high. In this section, we work with (b) and discuss the selection of such nodes, the reason such a selection works and the limitations of this method.
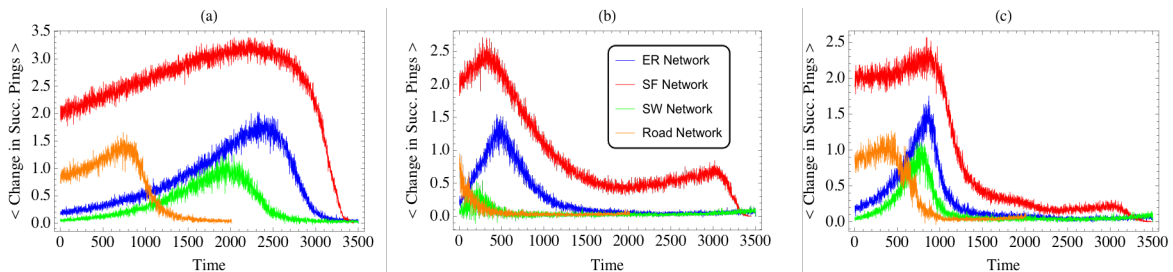
Figure 4.9: *Top* - Average change in the number of successful pings during link deletion in ER networks (blue), SF networks (red), SW networks (green) and Road network (orange) due to weighted selection of endpoints for individual pings.

As mentioned, we make a weighted selection of start and stop nodes in the transmission process. The start-node is selected with a probability proportional to the out-degree of the node and the stop-node is selected with a probability proportional to the in-degree of the node. When interpreted in the context of a traffic-flow, this method of selection is quite relevant. There is heterogeneity in the importance of stops in a road network and therefore trips between any two stops are not equally likely. It is more likely that a trip will happen between two important locations or atleast to/from an important location. Such locations also tend to be better connected. Since we are aware of this observation and are attempting to capture this in a model, we assume that a highly-connected location is most likely any important one and therefore more trips are assigned between such nodes. This method of weighted selection works because nodes with high-degree will tend to have better connectivity just by virtue of high-degree.

We repeat the entire procedure, with weighted selection of start-nodes and stop-nodes, in all network topologies and for all strategies of deletion. In fig-4.9, we plot the resulting increase in average number of successful pings as a function of start-time t. First-up, we observe that this method is highly effective during random deletion. In case of targeted deletion, it is more effective in ED-based deletion than in EBC based deletion. During random deletion, there is a maximum increase of more than three successful ping per time-step while in targeted deletion, the maximum increase is a little over two pings.

Irrespective of deletion strategy, weighted selection brings about maximum increase

in SF, ER, SW and the road network, in that order. This brings us to the limitation of this procedure. This method relies heavily on the number of relatively high-degree nodes and therefore on the heterogeneity in the degree distribution. For this reason, we see the most change in SF networks, which basically posses a heavy-tail distribution and the least change in the road network, which is almost lattice-like. When the degree distribution is close to homogeneous, there is minimal scope for picking out high-degree nodes and the result is very close to that of random selection.

While weighted selection of nodes is one mechanism to improve the probability of successful transmission, more methods can be constructed specific to the problem at hand. One such method is to assign a less efficient paths to trips, so that they are not immediately disrupted by the loss of links. Another method would be early rerouting of vehicles, if their paths are disrupted, rather than waiting until the deleted link is reached. Similar methods can be developed based on specific requirements.

4.7   Summary

To summarize, in this chapter, we report the results of our study on the process of 1-to-1 transmission in directed networks and its response to simultaneous deletion of links. In the first part of the work, we studied the transmission process in models of directed networks and analyzed the roles of network topology, deletion strategies and 2-node degree-correlations. We find that EBC based strategy is most disruptive, irrespective of topology, followed by Ed based deletion and random deletion, in that order. Transmission on ER networks is relatively more robust during random deletion but during ED based deletion, it is equally affected in all topologies. While dissortative correlations do not significantly affect transmission, assortative correlations increase the sensitivity to various extents depending on topology. We also present qualitative arguments to deduce the relationship between the two processes by studying the behaviours of the network metrics and transmission quantifiers.

In the second part, we recast the transmission process as problem of traffic-flow on a road-network. Using a network derived from the road-intersection data in Barcelona city, we studied the roles of three process parameters and interpreted the results in terms of the traffic-flow problem. We find that request-rate only affects the number of successful pings but not the normalized ping-times while the transmission probability inversely affects ping-times but not the probability of success. Small values of deletion probability increase the probability of success while reducing the average ping-times. Finally, we propose and detail a mechanism to mitigate the loss of transmission during link deletion based on weighted selection of nodes.

Chapter 5

Conclusion

The aim of this study was to expand the existing body of knowledge concerning directed complex networks. Complex networks have extensive application in a vast range of applications like science, engineering, medicine, finance etc. and the science of networks has become an enviable tool in systemic analysis. In particular, we have focused our efforts on exploring the structural, dynamical and functional response of complex directed networks to random and targeted loss of links. We show that this kind of approach has a two-fold advantage: Understanding the response of a network to loss of links helps us to design precautionary measure and introduce fail-safes; alternatively, investigating the role of deletion strategies makes it possible to design effective strategies for maximal disruption of spreading in the network.

We provide a brief introduction to the field of complex networks, discuss their wide range of applications and ever-increasing relevance. From local to intermediate to global scales, we define the most important and relevant network properties like correlations, clustering, connected components, centrality etc. and their corresponding measures. Based on these properties, we discuss three important standard models of network generation, namely the Erdös-Rényi model, Barabasi-Albert model and Watts-Strogatz model, their advantages and limitations. This is followed by brief discussion of percolation and spreading processes taking place on networks.

We then introduce Degree Preserving Rewiring mechanisms, to tune degree correlations and clustering, in random and scale-free directed networks. Using these mechanisms, we explore all the relevant properties, independent of topology. By artificially tuning the properties, we are able to study the extreme range of values that are not observed in real-world networks.

- We find that, in case of both correlations and clustering, changing the link-density in the network does not affect the working of the DPR mechanisms in any fundamental way.

- In general, we observe that networks with higher average degrees take longer to show the same amount of change in properties and the qualitative behaviour of the mechanisms is altered by the topology itself.

- We find that certain structural side-effects are inherent in the networks, and they restrict some properties from being tuned in isolation. Particularly in SF networks, the structural relationships are highly emphasized.

- After briefly discussing the running times for all the algorithms, we show that the time-complexity is quadratic for both clustering and correlations.

- When tuning degree-correlations, the behaviour of the limiting value of Spearman's Rho is found to differ considerably from the case of undirected networks. We observe that, in both ER and SF networks, the limiting values of correlation approach -1 rapidly for increasing values of parameters of the networks.

Besides the intended purpose of tuning network properties, the proposed mechanisms also reveal structural relationships and topological constraints in directed networks. By treating the process of rewiring as equivalent to exploring the space of second-order maximally random graphs, we are able to explore these relationships.

- We provide analytic arguments to show that the DPR mechanisms, designed to tune specific types of correlations, also uniquely affect the clustering coefficients. This is supported by qualitative arguments and numerical corroboration.

- Consequent to the arguments presented, we predict and numerically observe changes in sizes of the connected components, specific to the type of correlations being tuned.

- Thus, although we set out to find mechanisms to tune clustering and degree-correlations in directed networks, we find that the same mechanisms double up as tools to explore deeper structural relationships in the networks.

Next, we undertake a comprehensive study of the effects of link deletion in directed networks. We consider ER and SF networks and subject them to various strategies of breakdown like random failures and Edge Betweenness Centrality (EBC) and Edge Degree (ED) based deletions. The robustness to breakdown is measured by studying the behaviours of the strongly and weakly connected components and the efficiency of the networks. We also explore the effects of 2-node degree-correlations and clustering, on the robustness to loss of links, in both the network topologies. Our main results are the following:

- Irrespective of the strategy of deletion, ER networks consistently showed more robustness than SF networks.

- In ER networks, assortative correlations do not make any difference while dissortative correlations lead to a decrease in the robustness of the SCC and an increase in the case of WCC.

- In SF networks, both types of correlations lead to a decrease in the resilience of the connected components.

- The presence of clustering increases the vulnerability of ER networks to all strategies and SF networks to ED based strategies.

- Among the deletion strategies, we find that the loss of highly central links has the most damaging effect on any network topology while the removal of the least central links is the least effective.

- Random deletion was moderately effective compared to the EBC and ED based strategies and the network topology played a distinct role.

- We also study the effects of various types of deletion strategies on three real-world networks, chosen from the broad areas of transportation and biological interactions.

The response of networks to EBC based strategies of deletion is approximately reproduced using only local information (node degrees) about the networks. A combination of degrees of nodes on either ends of a directed edge, given as $k_i^{in}.k_j^{out}$, is found to be a suitable choice for the definition of ED. For large values of ED and EBC, we observe a strong positive correlation. This is further supported by the response that strategies S2 and S4 evoke in the various network models. Therefore, ED based strategies can be used as a computationally viable alternative to launch highly effective targeted counterattacks on networks to contain unwanted spreading processes. These results are meant to provide a qualitative picture of the responses of networks to link deletion and the roles of various network properties in an idealistic scenario. Exact results for real-world networks require a narrow and focused effort and may diverge considerably from those of network models.

Having studied the structural effects, we proceed to explore the dynamical response to link deletion in directed networks. We study the dynamics of a 1-to-1 transmission process, taking place simultaneously with random or targeted link deletion, on a network. In the first part of this study, we analyze the transmission process in models of directed networks and analyze the roles of network topology, deletion strategies and 2-node degree-correlations.

- We find the centrality based deletion to be the most disruptive, irrespective of topology, followed by edge-degree based deletion and random deletion, in that order.

- Transmission on ER networks is relatively more robust during random deletion but during ED based deletion, it is equally affected in all topologies.

- While dissortative correlations do not significantly affect transmission, assortative correlations increased the sensitivity to various extents depending on topology.

We also provide qualitative arguments to deduce the relationship between the two processes by studying the behaviours of the network metrics and transmission quantifiers. In the second part of this study, we apply the transmission process on a problem of traffic-flow on a road-network. Using a network derived from the road-intersection data in Barcelona city, we study the roles of three process parameters, namely request-rate, transmission probability and deletion probability and interpret the results in terms of the traffic-flow problem.

- We find that request-rate only affects the number of successful pings but not the normalized ping-times while the transmission probability inversely affects ping-times but not the probability of success.

- Small values of deletion probability result in increased probability of success, while reducing the average ping-times.

- We also propose a detailed mechanism to mitigate the loss of transmission during link deletion based on weighted selection of nodes. The extent of mitigative action is numerically studied.

The results presented in this thesis, while answering many questions, have also brought up some interesting questions. Starting from the results of rewiring mechanisms, future work could put more emphasis on the algorithmic aspects which were not prioritized in this work. The study of dynamical processes, even when limited

to the case of 1-to-1 transmission process, can be extended to include overloading of nodes/edges and further realistic constraints. Also study of the behavior of dynamical systems on directed networks undergoing link deletion would give rise to a deeper understanding of the structure-function relationship in directed networks.

Appendix A

Algorithms

Bibliography

[1] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *nature* **393**, 440 (1998).

[2] Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).

[3] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. Complex networks: Structure and dynamics. *Physics reports* **424**, 175–308 (2006).

[4] Newman, M. E. J. The structure and function of complex networks. *SIAM review* **45**, 167–256 (2003).

[5] Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Reviews of modern physics* **74**, 47 (2002).

[6] Alm, E. & Arkin, A. P. Biological networks. *Current opinion in structural biology* **13**, 193–202 (2003).

[7] Alon, U. Biological networks: the tinkerer as an engineer. *Science* **301**, 1866–1867 (2003).

[8] Pavlopoulos, G. A. *et al.* Using graph theory to analyze biological networks. *BioData mining* **4**, 10 (2011).

[9] Mason, O. & Verwoerd, M. Graph theory and networks in biology. *IET systems biology* **1**, 89–119 (2007).

[10] Wagner, A. & Fell, D. A. The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**, 1803–1810 (2001).

[11] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651 (2000).

[12] Guimera, R., Mossa, S., Turtschi, A. & Amaral, L. N. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences* **102**, 7794–7799 (2005).

[13] Yang, H. & H. Bell, M. G. Models and algorithms for road network design: a review and some new developments. *Transport Reviews* **18**, 257–278 (1998).

[14] Xie, F. & Levinson, D. Measuring the structure of road networks. *Geographical analysis* **39**, 336–356 (2007).

[15] Kumar, R., Novak, J. & Tomkins, A. Structure and evolution of online social networks. In *Link mining: models, algorithms, and applications*, 337–357 (Springer, 2010).

[16] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P. & Bhattacharjee, B. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 29–42 (ACM, 2007).

[17] Leskovec, J., Backstrom, L., Kumar, R. & Tomkins, A. Microscopic evolution of social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 462–470 (ACM, 2008).

[18] Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature reviews genetics* **12**, 56 (2011).

[19] Barabási, A.-L. Network medicine—from obesity to the "diseasome" (2007).

[20] Gai, P. & Kapadia, S. Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **466**, 2401–2423 (2010).

[21] Acemoglu, D., Ozdaglar, A. & Tahbaz-Salehi, A. Systemic risk and stability in financial networks. *American Economic Review* **105**, 564–608 (2015).

[22] Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2010).

[23] Epskamp, S. *et al.* qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* **48**, 1–18 (2012).

[24] Xia, M., Wang, J. & He, Y. Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one* **8**, e68910 (2013).

[25] Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *nature* **406**, 378 (2000).

[26] Crucitti, P., Latora, V., Marchiori, M. & Rapisarda, A. Efficiency of scale-free networks: error and attack tolerance. *Physica A: Statistical Mechanics and its Applications* **320**, 622–642 (2003).

[27] Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Physical review letters* **85**, 4626 (2000).

[28] Solé, R. V., Rosas-Casals, M., Corominas-Murtra, B. & Valverde, S. Robustness of the european power grids under intentional attack. *Physical Review E* **77**, 026102 (2008).

[29] Motter, A. E. & Lai, Y.-C. Cascade-based attacks on complex networks. *Physical Review E* **66**, 065102 (2002).

[30] Wang, J. W. & Rong, L. L. Cascade-based attack vulnerability on the us power grid. *Safety Science* **47**, 1332–1336 (2009).

[31] Crucitti, P., Latora, V., Marchiori, M. & Rapisarda, A. Error and attack tolerance of complex networks. *Physica A: Statistical Mechanics and its Applications* **340**, 388–394 (2004).

[32] Santiago, A. & Benito, R. Robustness of heterogeneous complex networks. *Physica A: Statistical Mechanics and its Applications* **388**, 2234–2242 (2009).

[33] Fang, X., Yang, Q. & Yan, W. Modeling and analysis of cascading failure in directed complex networks. *Safety Science* **65**, 1–9 (2014).

[34] Yan, Y. Z. & Wang, H. X. Random edge deletion in directed scale-free graphs. In *Networks Security, Wireless Communications and Trusted Computing, 2009. NSWCTC'09. International Conference on*, vol. 1, 52–55 (IEEE, 2009).

[35] Duch, J. & Arenas, A. Community detection in complex networks using extremal optimization. *Physical review E* **72**, 027104 (2005).

[36] Lancichinetti, A., Fortunato, S. & Kertész, J. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics* **11**, 033015 (2009).

[37] Newman, M. E. & Girvan, M. Finding and evaluating community structure in networks. *Physical review E* **69**, 026113 (2004).

[38] Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99**, 7821–7826 (2002).

[39] Fortunato, S. Community detection in graphs. *Physics reports* **486**, 75–174 (2010).

[40] Newman, M. E. Fast algorithm for detecting community structure in networks. *Physical review E* **69**, 066133 (2004).

[41] Clauset, A., Newman, M. E. & Moore, C. Finding community structure in very large networks. *Physical review E* **70**, 066111 (2004).

[42] Newman, M. E. Modularity and community structure in networks. *Proceedings of the national academy of sciences* **103**, 8577–8582 (2006).

[43] Lü, L. & Zhou, T. Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications* **390**, 1150–1170 (2011).

[44] Liben-Nowell, D. & Kleinberg, J. The link-prediction problem for social networks. *Journal of the American society for information science and technology* **58**, 1019–1031 (2007).

[45] Martínez, V., Berzal, F. & Cubero, J.-C. A survey of link prediction in complex networks. *ACM Computing Surveys (CSUR)* **49**, 69 (2017).

[46] Guille, A., Hacid, H., Favre, C. & Zighed, D. A. Information diffusion in online social networks: A survey. *ACM Sigmod Record* **42**, 17–28 (2013).

[47] Bakshy, E., Rosenn, I., Marlow, C. & Adamic, L. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, 519–528 (ACM, 2012).

[48] Myers, S. A., Zhu, C. & Leskovec, J. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 33–41 (ACM, 2012).

[49] Yang, J. & Leskovec, J. Modeling information diffusion in implicit networks. In *2010 IEEE International Conference on Data Mining*, 599–608 (IEEE, 2010).

[50] Erdös, P. & Rényi, A. On random graphs i. *Publ. Math. Debrecen* **6**, 290–297 (1959).

[51] Newman, M. *Networks* (Oxford university press, 2018).

[52] Newman, M. E. J. Mixing patterns in networks. *Physical Review E* **67**, 026126 (2003).

[53] Newman, M. E. J. Assortative mixing in networks. *Physical Review Letters* **89**, 208701 (2002).

[54] Broder, A. *et al.* Graph structure in the web. *Computer networks* **33**, 309–320 (2000).

[55] Csermely, P., London, A., Wu, L.-Y. & Uzzi, B. Structure and dynamics of core/periphery networks. *Journal of Complex Networks* **1**, 93–123 (2013).

[56] Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43 (1953).

[57] Brin, S. & Page, L. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems* **30**, 107–117 (1998).

[58] Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* 35–41 (1977).

[59] Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *science* **286**, 509–512 (1999).

[60] Hammersley, J. M. & Handscomb, D. Percolation processes. In *Monte Carlo Methods*, 134–141 (Springer, 1964).

[61] Broadbent, S. R. & Hammersley, J. M. Percolation processes: I. crystals and mazes. In *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 53, 629–641 (Cambridge University Press, 1957).

[62] Newman, M. E. & Watts, D. J. Scaling and percolation in the small-world network model. *Physical review E* **60**, 7332 (1999).

[63] Callaway, D. S., Newman, M. E., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: Percolation on random graphs. *Physical review letters* **85**, 5468 (2000).

[64] Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Breakdown of the internet under intentional attack. *Physical review letters* **86**, 3682 (2001).

[65] Schwartz, N., Cohen, R., Ben-Avraham, D., Barabási, A.-L. & Havlin, S. Percolation in directed scale-free networks. *Physical Review E* **66**, 015104 (2002).

[66] Vázquez, A. & Moreno, Y. Resilience to damage of graphs with degree correlations. *Physical Review E* **67**, 015101 (2003).

[67] Holme, P. Edge overload breakdown in evolving networks. *Physical Review E* **66**, 036119 (2002).

[68] Holme, P. & Kim, B. J. Vertex overload breakdown in evolving networks. *Physical Review E* **65**, 066109 (2002).

[69] Moreno, Y., Gómez, J. & Pacheco, A. Instability of scale-free networks under node-breaking avalanches. *EPL (Europhysics Letters)* **58**, 630 (2002).

[70] Moreno, Y., Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Critical load and congestion instabilities in scale-free networks. *EPL (Europhysics Letters)* **62**, 292 (2003).

[71] Lloyd, A. L. & May, R. M. How viruses spread among computers and people. *Science* **292**, 1316–1317 (2001).

[72] Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control* (Oxford university press, 1992).

[73] Bailey, N. T. *et al. The mathematical theory of infectious diseases and its applications.* 2nd ediition (Charles Griffin & Company Ltd 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975).

[74] Grassberger, P. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences* **63**, 157–172 (1983).

[75] May, R. M. & Anderson, R. M. The transmission dynamics of human immunodeficiency virus (hiv). *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **321**, 565–607 (1988).

[76] May, R. M. & Lloyd, A. L. Infection dynamics on scale-free networks. *Physical Review E* **64**, 066112 (2001).

[77] Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Physical review E* **65**, 036104 (2002).

[78] Dezső, Z. & Barabási, A.-L. Halting viruses in scale-free networks. *Physical Review E* **65**, 055103 (2002).

[79] Jespersen, S., Sokolov, I. & Blumen, A. Relaxation properties of small-world networks. *Physical Review E* **62**, 4405 (2000).

[80] Lahtinen, J., Kertész, J. & Kaski, K. Scaling of random spreading in small world networks. *Physical Review E* **64**, 057105 (2001).

[81] Lahtinen, J., Kertész, J. & Kaski, K. Random spreading phenomena in annealed small world networks. *Physica A: Statistical Mechanics and its Applications* **311**, 571–580 (2002).

[82] Pandit, S. A. & Amritkar, R. Random spread on the family of small-world networks. *Physical Review E* **63**, 041104 (2001).

[83] Watts, D. J. The dynamics of networks between order and randomness. *Small Worlds* (1999).

[84] Kim, B. J. *et al.* Dynamic instabilities induced by asymmetric influence: prisoners' dilemma game in small-world networks. *Physical Review E* **66**, 021907 (2002).

[85] Abramson, G. & Kuperman, M. Social games in a social network. *Physical Review E* **63**, 030901 (2001).

[86] Stauffer, D., Aharony, A., da Fontoura Costa, L. & Adler, J. Efficient hopfield pattern recognition on a scale-free neural network. *The European Physical Journal B-Condensed Matter and Complex Systems* **32**, 395–399 (2003).

[87] Lago-Fernández, L. F., Huerta, R., Corbacho, F. & Sigüenza, J. A. Fast response and temporal coherent oscillations in small-world networks. *Physical Review Letters* **84**, 2758 (2000).

[88] Hong, H., Choi, M.-Y. & Kim, B. J. Synchronization on small-world networks. *Physical Review E* **65**, 026139 (2002).

[89] Barahona, M. & Pecora, L. M. Synchronization in small-world systems. *Physical review letters* **89**, 054101 (2002).

[90] Barabási, A. L., Albert, R. & Jeong, H. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications* **272**, 173–187 (1999).

[91] Dorogovtsev, S. N., Mendes, J. F. F. & Samukhin, A. N. Structure of growing networks with preferential linking. *Physical review letters* **85**, 4633 (2000).

[92] Price, D. d. S. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science* **27**, 292–306 (1976).

[93] Bollobás, B., Borgs, C., Chayes, J. & Riordan, O. Directed scale-free graphs. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, 132–139 (Society for Industrial and Applied Mathematics, 2003).

[94] Krapivsky, P. L. & Redner, S. Organization of growing random networks. *Physical Review E* **63**, 066123 (2001).

[95] Krapivsky, P. L., Redner, S. & Leyvraz, F. Connectivity of growing random networks. *Physical review letters* **85**, 4629 (2000).

[96] Ostroumova, L., Ryabchenko, A. & Samosvat, E. Generalized preferential attachment: Tunable power-law degree distribution and clustering coefficient. In *WAW*, 185–202 (Springer, 2013).

[97] Bollobás, B. Random graphs. In *Modern Graph Theory*, 215–252 (Springer, 1998).

[98] Kim, H., Del Genio, C. I., Bassler, K. E. & Toroczkai, Z. Constructing and sampling directed graphs with given degree sequences. *New Journal of Physics* **14**, 023012 (2012).

[99] Noldus, R. & Van Mieghem, P. Assortativity in complex networks. *Journal of Complex Networks* cnv005 (2015).

[100] Newman, M. E. J. Random graphs with clustering. *Physical Review Letters* **103**, 058701 (2009).

[101] Park, J. & Newman, M. E. J. Solution for the properties of a clustered network. *Physical Review E* **72**, 026136 (2005).

[102] Newman, M. E. J. Properties of highly clustered networks. *Physical Review E* **68**, 026121 (2003).

[103] Szabó, G., Alava, M. & Kertész, J. Clustering in complex networks. *Complex networks* 139–162 (2004).

[104] Klemm, K. & Eguiluz, V. M. Highly clustered scale-free networks. *Physical Review E* **65**, 036123 (2002).

[105] Vázquez, A. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E* **67**, 056104 (2003).

[106] Herrera, C. & Zufiria, P. J. Generating scale-free networks with adjustable clustering coefficient via random walks. In *Network Science Workshop (NSW), 2011 IEEE*, 167–172 (IEEE, 2011).

[107] Krot, A. & Prokhorenkova, L. O. Local clustering coefficient in generalized preferential attachment models. In *International Workshop on Algorithms and Models for the Web-Graph*, 15–28 (Springer, 2015).

[108] Fagiolo, G. Clustering in complex directed networks. *Physical Review E* **76**, 026107 (2007).

[109] Williams, O. & Del Genio, C. I. Degree correlations in directed scale-free networks. *PloS one* **9**, e110121 (2014).

[110] Foster, J. G., Foster, D. V., Grassberger, P. & Paczuski, M. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences* **107**, 10815–10820 (2010).

[111] Mayo, M., Abdelzaher, A. & Ghosh, P. Mixed degree-degree correlations in directed social networks. In *International Conference on Combinatorial Optimization and Applications*, 571–580 (Springer, 2014).

[112] Roberts, E. & Coolen, A. Unbiased degree-preserving randomization of directed binary networks. *Physical Review E* **85**, 046103 (2012).

[113] Kashyap, G. & Ambika, G. Mechanisms for tuning clustering and degree-correlations in directed networks. *Journal of Complex Networks* **6**, 767–787 (2017).

[114] Van Mieghem, P., Wang, H., Ge, X., Tang, S. & Kuipers, F. A. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *The European Physical Journal B-Condensed Matter and Complex Systems* **76**, 643–652 (2010).

[115] Zhou, S. & Mondragon, R. J. Structural constraints in complex networks. *New Journal of Physics* **9**, 173 (2007).

[116] Yang, D., Pan, L. & Zhou, T. Lower bound of assortativity coefficient in scale-free networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 033113 (2017).

[117] Xulvi-Brunet, R. & Sokolov, I. M. Reshuffling scale-free networks: From random to assortative. *Physical Review E* **70**, 066102 (2004).

[118] Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).

[119] Dorogovtsev, S., Ferreira, A., Goltsev, A. & Mendes, J. Zero pearson coefficient for strongly correlated growing trees. *Physical Review E* **81**, 031135 (2010).

[120] Litvak, N. & Van Der Hofstad, R. Uncovering disassortativity in large scale-free networks. *Physical Review E* **87**, 022801 (2013).

[121] van der Hoorn, P. & Litvak, N. Degree-degree dependencies in directed networks with heavy-tailed degrees. *Internet mathematics* **11**, 155–179 (2015).

[122] van der Hoorn, P., Prokhorenkova, L. O. & Samosvat, E. Generating maximally disassortative graphs with given degree distribution. *arXiv preprint arXiv:1607.01742* (2016).

[123] Crucitti, P., Latora, V. & Marchiori, M. Model for cascading failures in complex networks. *Physical Review E* **69**, 045104 (2004).

[124] Pagani, G. A. & Aiello, M. The power grid as a complex network: a survey. *Physica A: Statistical Mechanics and its Applications* **392**, 2688–2700 (2013).

[125] Carreras, B. A., Lynch, V. E., Dobson, I. & Newman, D. E. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos: An interdisciplinary journal of nonlinear science* **12**, 985–994 (2002).

[126] Dobson, I., Carreras, B. A., Lynch, V. E. & Newman, D. E. Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **17**, 026103 (2007).

[127] Carreras, B. A., Lynch, V. E., Dobson, I. & Newman, D. E. Complex dynamics of blackouts in power transmission systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **14**, 643–652 (2004).

[128] Peron, T. D. & Rodrigues, F. A. Collective behavior in financial markets. *EPL (Europhysics Letters)* **96**, 48004 (2011).

[129] Kauê Dal'Maso Peron, T., da Fontoura Costa, L. & Rodrigues, F. A. The structure and resilience of financial market networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **22**, 013117 (2012).

[130] May, R. M., Levin, S. A. & Sugihara, G. Complex systems: Ecology for bankers. *Nature* **451**, 893–895 (2008).

[131] Chi, K. T., Liu, J. & Lau, F. C. A network perspective of the stock market. *Journal of Empirical Finance* **17**, 659–667 (2010).

[132] Zou, C. C., Towsley, D. & Gong, W. Email worm modeling and defense. In *Computer Communications and Networks, 2004. ICCCN 2004. Proceedings. 13th International Conference on*, 409–414 (IEEE, 2004).

[133] Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).

[134] Berche, B., von Ferber, C., Holovatch, T. & Holovatch, Y. Public transport networks under random failure and directed attack. *arXiv preprint arXiv:1002.2300* (2010).

[135] Kawamoto, H., Takayasu, H., Jensen, H. J. & Takayasu, M. Precise calculation of a bond percolation transition and survival rates of nodes in a complex network. *PloS one* **10**, e0119979 (2015).

[136] Iyer, S., Killingback, T., Sundaram, B. & Wang, Z. Attack robustness and centrality of complex networks. *PloS one* **8**, e59613 (2013).

[137] Martin, S., Carr, R. D. & Faulon, J. L. Random removal of edges from scale free graphs. *Physica A: Statistical Mechanics and its Applications* **371**, 870–876 (2006).

[138] Wang, J. W. & Rong, L. L. Edge-based-attack induced cascading failures on scale-free networks. *Physica A: Statistical Mechanics and its Applications* **388**, 1731–1737 (2009).

[139] He, S., Li, S. & Ma, H. Effect of edge removal on topological and functional robustness of complex networks. *Physica A: Statistical Mechanics and its Applications* **388**, 2243–2253 (2009).

[140] Goerdt, A. & Molloy, M. Analysis of edge deletion processes on faulty random regular graphs. *Theoretical computer science* **297**, 241–260 (2003).

[141] Zhang, G. Q., Wang, D. & Li, G. J. Enhancing the transmission efficiency by edge deletion in scale-free networks. *Physical Review E* **76**, 017101 (2007).

[142] Cartledge, C. L. & Nelson, M. L. Connectivity damage to a graph by the removal of an edge or a vertex. *arXiv preprint arXiv:1103.3075* (2011).

[143] Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. Attack vulnerability of complex networks. *Physical review E* **65**, 056109 (2002).

[144] Kashyap, G. & Ambika, G. Link deletion in directed complex networks. *Physica A: Statistical Mechanics and its Applications* **514**, 631–643 (2019).

[145] for Research Core Team, T. N. Transportation networks for research (2017). URL `https://github.com/bstabler/TransportationNetworks`.

[146] Stelzl, U. *et al.* A human protein–protein interaction network: A resource for annotating the proteome. *Cell* **122**, 957–968 (2005).

[147] Human protein (stelzl) network dataset – KONECT (2017). URL `http://konect.uni-koblenz.de/networks/maayan-Stelzl`.

[148] Openflights network dataset – KONECT (2017). URL `http://konect.uni-koblenz.de/networks/openflights`.

[149] Latora, V. & Marchiori, M. Efficient behavior of small-world networks. *Physical review letters* **87**, 198701 (2001).

[150] Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics* **3**, 276 (2007).

[151] Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nature physics* **6**, 888 (2010).

[152] Nowzari, C., Preciado, V. M. & Pappas, G. J. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine* **36**, 26–46 (2016).

[153] Moreno, Y., Nekovee, M. & Pacheco, A. F. Dynamics of rumor spreading in complex networks. *Physical Review E* **69**, 066130 (2004).

[154] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Reviews of modern physics* **87**, 925 (2015).

[155] Oshida, N. & Ihara, S. Packet traffic analysis of scale-free networks for large-scale network-on-chip design. *Physical Review E* **74**, 026115 (2006).

[156] Teuscher, C. Nature-inspired interconnects for self-assembled large-scale network-on-chip designs. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **17**, 026106 (2007).

[157] Kaiser, M. & Hilgetag, C. C. Edge vulnerability in neural and metabolic networks. *Biological cybernetics* **90**, 311–317 (2004).

[158] Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913 (1997).

[159] Sullivan, J., Novak, D., Aultman-Hall, L. & Scott, D. M. Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach. *Transportation Research Part A: Policy and Practice* **44**, 323–336 (2010).

[160] Knoop, V. L., Hoogendoorn, S. P. & van Zuylen, H. J. Approach to critical link analysis of robustness for dynamical road networks. In *Traffic and granular flow'05*, 393–402 (Springer, 2007).

[161] Herty, M. & Klar, A. Modeling, simulation, and optimization of traffic flow networks. *SIAM Journal on Scientific Computing* **25**, 1066–1087 (2003).

[162] Coclite, G. M., Garavello, M. & Piccoli, B. Traffic flow on a road network. *SIAM journal on mathematical analysis* **36**, 1862–1886 (2005).

[163] Zhao, L., Lai, Y.-C., Park, K. & Ye, N. Onset of traffic congestion in complex networks. *Physical Review E* **71**, 026125 (2005).

[164] Holme, P. Congestion and centrality in traffic flow on complex networks. *Advances in Complex Systems* **6**, 163–176 (2003).

[165] Yan, G., Zhou, T., Hu, B., Fu, Z.-Q. & Wang, B.-H. Efficient routing on complex networks. *Physical Review E* **73**, 046108 (2006).

[166] Danila, B., Yu, Y., Marsh, J. A. & Bassler, K. E. Optimal transport on complex networks. *Physical Review E* **74**, 046106 (2006).