1
Introduction

Networks can be seen in real-world complex systems such as biological networks [Girvan and Newman, 2002] [Assenov *et al.*, 2008], transportation networks [Bell and Iida, 1997], communication networks [Schwartz, 1977] and many more [Kelly, 2011]. These networks can be analyzed under two different dimensions: first is the *structural dynamics* and second is the *diffusion dynamics* on networks. An inter-dependency between the two is observed and can be utilized in many ways. This thesis contributes in both the dimensions.

1.1 MOTIVATION AND APPLICATIONS

1.1.1 Network modeling

Understanding and modeling of complex network systems have received a lot of interest in recent time due to its applications in many real world systems including transportation and communication systems, WWW, Internet, power grid, etc. [Wang *et al.*, 2011a; Albert *et al.*, 1999; Caldarelli, 2007; Albert *et al.*, 2000, 2004; Arianos *et al.*, 2009; Newman, 2010]. Real networks such as Internet, social networks and transportation networks are large growing complex networks with non-trivial connectivity pattern and structural properties. Collective dynamics of a network is governed by local evolution processes. In order to model a real growing network, it is essential to adapt the evolving process of real networks which possibly can be captured by model parameters. Reasons for development of parametric network models include the following.

1. The structure of a real network is the outcome of processes which drive the structural evolution of the network. A given real network can be possibly estimated by a combination of different stochastic processes on it. Thus a network model could inherit properties of real networks where the parameters involved in the model signify the contribution of individual local stochastic processes.

2. Parametric models allow to infer the hypotheses regarding complex stochastic processes associated with network evolutions. Correlation among the structural properties of real networks and model networks verifies the novelty of the model and correctness of the adopted hypotheses.

3. A parametric model can answer the question that how localized stochastic processes and structures are responsible for global network patterns. In particular, whether such localized processes are sufficient to obtain collective properties of real networks.

4. In general, it is believed that a deterministic approach to model real networks is not a good idea to capture statistical properties of real complex networks. A more flexible and traceable approach can work well in this context. A parametric model can be of great value if it is able to provide closed form expression of structural properties of evolving networks. Algebraic relations between the parameters can justify influence of these parameters in order to determine specific structural properties in the corresponding model networks.

5. A parametric model can work in two ways: firstly it can be used to produce the networks of desired properties by setting the particular values of parameters, and secondly it can be used for structural reconstruction of a real network by suitably choosing the parameter values utilizing a structural property of the real network.

1.1.2 Structural reconstruction of networks

The structural reconstruction of a real network is concerned with reconstruction of a real network by using both a network model and limited information of the real network. By reconstruction, we mean that the model network should possess some collective structural and/or spectral properties of the real network. For a parametric network model the question is how to choose the values of the model parameters using limited information about the real network such that for those values, the model network can capture collective properties of the real network.

A few reconstruction methods are introduced recently in literature for structural reconstruction of networks. For example, a reconstruction method is proposed using betweenness centrality in [Comellas and Paz-Sánchez, 2008], spectral reconstruction of complex network is introduced in [Comellas and Diaz-Lopez, 2008], and evolutionary reconstruction of networks is considered in [Ipsen and Mikhailov, 2002]. Indeed a common feature in these proposed approaches is the consideration of an initial (random) network and then the given network is reconstructed by rewiring the edges of the initial network. Thus its application is limited since in general, real networks are growing networks. These methods are also ad-hoc and lack intuition about the evolution process during its formation as a growing network to achieve a particular structure in the desired network.

As mentioned above, a pertinent question in the study of real networks is how to reconstruct a real network when limited information about the network is given. If a method can be accomplished for such a task then the problem of storing of large networks can also be addressed. Indeed, the standard procedure to store a network is by storing the adjacency matrix or edge list associated with the network having space complexity $O(n^2)$ and $O(\overline{k}_n n) = O(n^{1+\delta})$, respectively, where $1 > \delta > 0$, n is the size of the network and $\overline{k}_n$ is the average degree of the network. Instead, a suitable reconstruction method can reduce the space complexity by storing only a statistical property (of space complexity $O(n)$) of the given network while preserving other statistical properties. If parametric model networks can capture statistical properties of real networks then any such model network can be used as a platform to analyze diffusion dynamics on the corresponding real network. For example, information diffusion and epidemic spreading on a real social network can be speculated by defining different protocols on the corresponding model networks.

Here we mention that network reconstruction (NR) methods are proposed in the context of accessing the dynamics on the nodes of a network. This approach approximately determines the interaction pattern between the nodes during a dynamical process on the network [Angulo *et al.*, 2015; Shen *et al.*, 2014]. However, the authors in [Angulo *et al.*, 2015] discussed the fundamental limitations of NR. The structural reconstruction of a network is significantly different from accessing interaction pattern between the nodes of a real network during the dynamic process on it. In fact it can address the lack of adequate generative models for complex networks in literature aiming to capture multiple structural properties of a given real-world network simultaneously. Often the existing popular network generation models focus on creating mathematical models for the generation of networks having community structure and/or heavy-tailed degree distribution such as power-law without addressing its potential to mimic structural properties of a given real network. In this context, we mention that the Graph500 benchmark [Murphy *et al.*, 2010; Jose *et al.*, 2013] attempted to generate synthetic networks considered as graph benchmarks in high-performance computing. For example, the Recursive Matrix (R-MAT) model [Chakrabarti *et al.*, 2004] is used to generate benchmark graphs. However it fails to achieve realism [Kolda *et al.*, 2014]. This calls for the development of new parametric network generation models and reconstruction methods that can produce model networks which replicate the structure of a given real-world network.

1.1.3 Relation between Diffusion dynamics and network evolution

Most often the structural organization of a real network is influenced by a diffusion phenomena on the network. As a large class of real networks are scale-free networks, it is of paramount interest to investigate what type of diffusion dynamics gives rise of a power-law in the degree distribution of a network. Random walk dynamics is often used to study different diffusion dynamics. Dynamics of a random walker has been investigated in many contexts [Nash-Williams, 1959; Tetali, 1991], for example, navigation and centrality of networks [Perra *et al.*, 2012; Starnini *et al.*, 2012], routing of packets in the Internet, and diffusion in communication networks. Traffic in transportation networks can also be represented as phenomena of random walkers [Wang *et al.*, 2006]. Random walk dynamics has been applied to explore the dynamics of wealth's distribution in economic networks, gene expression pathways in biological networks and search (navigation) strategies in Internet [Adamic *et al.*, 2001; Tadić and Rodgers, 2002; Tadić and Thurner, 2004; Kim *et al.*, 2002; Rosvall *et al.*, 2005; Germano and de Moura, 2006]. In random walks, a walker positioned at a node i can move to any node j which is linked to the node i in the network with equal probability. Thus the selection of a node for a move is uniform among the neighbors of the current node occupied by a random walker [Akyildiz *et al.*, 2000; Noh and Rieger, 2004].

Navigation or searching dynamics in a network can be modelled as random walker dynamics in which a walker can travel from one node to other connected nodes via links. The performance of the searching can be measured in the form of mean-first-passage-time (MFPT). MFPT of a node is the average number of clicks to reach that node while searching started from a random node. A node of lower MFPT value gets more visits in searching which increase the importance of the node. *Dynamics on the networks* affects the structural *Dynamics of the networks* and vice versa to make the networked system sustainable [Aoki *et al.*, 2015]. Real networks, for example, WWW network exhibits scale-free behavior under preferential or biased growth (proposed by Barabasi and Albert) which should be justified by random walk dynamics, as the interdependency between diffusion and structural dynamics provide a boost to think in this direction. Now the questions are following: What is the motivation behind the preferential or biased growth of the networks like Internet or WWW? How can one justify the preferential or biased growth of the networks using random walk dynamics, if *Dynamics on the networks* and *Dynamics of the networks* are interdependent?

Assume that real-world networks follow random walk based diffusion dynamics and grows preferentially. As it is observed that *Dynamics on the networks* and *Dynamics of the networks* affect each other, so existing random walk dynamics should justify the preferential or biased growth of the networks. Consider WWW network as an example which follows preferential attachment [Barabási *et al.*, 2000] and searching dynamics in the network is modeled as random walk phenomena [Adamic *et al.*, 2001]. In the preferential attachment, a node prefers to get linked with highly connected nodes in the network [Barabási *et al.*, 2000] while according to random walk dynamics, average searching time also known as mean-first-passage-time (MFPT), of a node in a network is inversely proportional to the degree of that node [Redner, 2001]. Thus low degree node has high MFPT which does not justify the preferential growth of the network because if a node has degree 1 (initial degree) then whether it is connected with a high degree node or low degree node does not matter, it will have high searching time. We can conclude that degree based preferential or biased growth and the standard random walk based searching are not consistent to explain the dynamics of real-world networks which follow power-law degree distribution. Other existing studies on biased random walk dynamics are also not able to justify the preferential or biased growth of the networks. From the above discussion, it is concluded that there is a hidden dynamics of diffusion which can not be captured by the existing random walk dynamics (based or unbiased)[Redner, 2001; Fronczak and Fronczak, 2009]. This calls for investigation of biased navigation or random walk dynamics which can able to provide a possible justification of the preferential or biased growth of the WWW and Internet.

1.2 THESIS OVERVIEW AND CONTRIBUTIONS

In this thesis we propose a few parametric network formation models and develop a reconstruction method for structural reconstruction of real scale-free networks by utilizing the degree sequence/degree distribution of the real network. We also introduce a biased random walk on networks and show its potential to justify preferential growth of the networks during diffusion on the networks. Diffusion protocols on both static and dynamic networks are proposed that can be used for link failure detection for real networks. Finally, a reconstruction technique for real networks is described by utilizing diffusion dynamics on the networks.

Thesis consists of theoretical contributions in theory of network models including application in network reconstruction from statistical properties of scale-free networks, and diffusion on networks including applications such as link failure detection, and inference of network structure from diffusion data. Chapters $3-6$ deal with different models of networks with different perspective such a context dependency in selection, nucleation as network formation, choice and chance based link formation, and finally in Chapter 6, it is shown that the proposed parametric model for complex networks can be utilized for structural reconstruction of scale-free real networks. Chapters $7-9$ are focused on study of diffusion dynamics on real networks and its applications for detection of link failure in real networks, justification for heterogeneous growth of real networks, and reconstruction of networks using diffusion data. Chapter 7 raised a question about the existing random walk diffusion protocols which are not able to justify the growth of networks and a new biased random walk is proposed that justifies the observed property (biased growth or preferential growth) of real networks. Thesis contributes in the domain of network modeling and diffusion on networks. A brief overview of the thesis (chapter by chapter) is as follows:

1.2.1 Network Models

Parametric network generation model: Context dependent preferential attachment (Chapter 3)

In this chapter we propose a parametric growing network generation model based on context dependent preferential attachment. It can be observed in real social networks that a newly joined node forms a link with an existing node based on contexts. Consider an example of a matrimonial site. While selecting a bride or groom, one looks for multiple attributes of a lass or lad, for example, family background, job, physical appearance, etc. Similarly, in the case of buying a product, a person thinks about the price and quality. So we can observe that our selection process is not one-dimensional but we consider multiple dimensions. We adopt context dependent node selection for growth or link formation among the participating nodes. We propose "Context dependent preferential attachment model" (CDPAM) for generation of growing random networks which inherit observed properties of real networks.

In the proposed network model, the preference of a new node to get attached with an old node is determined by the local and global property of the old node. We consider that local and global properties of a node as the degree and relative average degree of the node respectively. We prove that the degree distribution of complex networks generated by CDPAM follow power law with exponent lies in the interval [2, 3] and the expected diameter grows logarithmically with the size of new nodes added to the initial small network. Numerical results show that the expected diameter stabilizes when alike weights to the local and global properties are assigned by the new nodes. Computing various measures including clustering coefficient, assortativity, number of triangles, algebraic connectivity, spectral radius, we show that the proposed model replicates properties of real networks when alike weights are given to local and global properties. Finally, we observe that the BA model [Barabási and Albert, 1999a] is a limiting case of CDPAM when new nodes tend to give large weight to the local property compared to the weight given to the global property during link formation.

Parametric model for signed networks (Chapter 4)

Often modeling of networks are considered in which all the connected pair of nodes have similar kind of relations, for example, friendship. However, it is well known that there are multiple examples of real-world networks which have different type of relations among the participating nodes. Broadly, three types of relationships are categorized: one is positive, for example friendship, second is negative, for example, enmity, and third is neutral that can be seen as no relationship at all. These type of relations are represented by $+1$, -1 and 0, respectively. A network in which existing links are denoted by positive and negative edges, for example, friendship and enmity respectively, is called a signed network.

We consider modeling of social networks as signed networks based on observed social phenomena, preferential attachment and random connection formation with the structural balance of local groups. A multilayer network model based approach is adopted to model it in which two layers are considered, one is of positive edges and another of negative edges having the same number of nodes in both the layers. We show that the degree distribution of the signed networks generated by the proposed model follow power law on aggregation with power-law exponent $\left(\gamma = 1 + \frac{1}{\lambda_1(\mathscr{C})}\right)$ where $\lambda_1(\mathscr{C})$ is the largest eigenvalue of the inter-layer correlation matrix $\mathscr{C}$. The novelty of the model is validated by comparing the properties of model networks and real-world signed social networks using different metrics. The networks generated by the proposed model exhibit the behavior of real-world signed social networks more closely as compared to other considered modeling approaches in the literature. Internal growth and the theory of social balance are also considered while proposing the model.

Parametric network model inspired by nucleation process (Chapter 5)

In the previous chapters, we considered the adaptation of social theories to model the growth processes for networks. In this chapter, the well known nucleation process is observed and studied as network formation which is also adopted for social network modeling. Events of small duration, for example interactions among participants of workshops and conferences can be modelled as a nucleation process in the form of networks.

In nucleation, individual entities move in the solution or vacuum in Brownian motion. In the process, two or more entities come in physical proximity and thereby produce temporal combined units or clusters. Depending on the prevailing physical conditions and the size of the unit, the unit gains in terms of volume free energy and looses surface free energy. Such clusters, called embryos, may also gather additional units by similar way and loose units by dissolution. If they travel across an energy hill by gathering additional units, the volume free energy gain dominates over surface free energy effect after a certain size. At this stage, embryos give birth to a nucleus which is qualified to grow without dissolution. On the other hand, an embryo may travel in the opposite direction in the energy axes and dissolve completely or partially. In other words, the entities in the solution form network, they have reward and penalty on formation of network which depends on the prevailing conditions and the unit size, and an individual embryo behaves stochastically which is similar to a network formation.

We consider the nucleation phenomena in the formation of groups of participants having similar research interest in a workshop organized in an institute. Here, we assume that the nodes represent researchers in the institute and V denotes the set of all researchers among them who are participating in the workshop. Thus, for the network formation, we assume $N(\gg |V|)$ is the total number of researchers available in the institute. We assume that the workshop schedule includes k parallel sessions focused on different sub areas of research in each such session. The speakers in each session at a given instant can be considered as the initiators for the resultant network formation, in which, the closeness of a researcher (node) can be defined by considering his research interest or social acquaintances with fellow researchers. The resultant network can be thought of as an outcome of the crystallization process.

Structural network reconstruction using parametric modeling approach (Chapter 6)

Here, we consider the following problem: given the degree sequence and/or degree distribution of a network G, can one construct a model network which can capture or replicate the structural properties of G? For parametric complex network generation models such as [Pandey and Adhikari, 2015; Dorogovtsev *et al.*, 2000], the question can be posed as follows. How to choose the values of the model parameters that can generate a network with desired structural properties? We call this problem as structural network reconstruction (SNR) problem which can be interpreted as an inverse problem of network generation. In this chapter, we consider SNR for scale-free networks, that is, we develop a parametric growing network generation model which can inherit various properties of a given scale-free network depending on the values of the model parameters.

We use the preferential attachment and random attachment with local growth techniques for the generation of networks. We validate the proposed model for certain real world networks and show that the networks generated by the proposed model can replicate various properties of the given real-world networks. We provide a sufficient condition for the model parameters satisfying which the model can generate networks which follow edge-densification and densification power-law. Computable expressions for the expected number of triangles and expected diameter are also obtained. Finally, we numerically establish that the proposed model can generate networks with shrinking diameter and modular structure when specific model parameters are chosen.

In order to verify the novelty of the proposed parametric model in comparison to other existing parametric network models, we consider solving the SNR problem by using DMS model [Dorogovtsev *et al.*, 2000], context dependent preferential attachment model (CDPAM) [Pandey and Adhikari, 2015], forest fire model (FFM) [Leskovec *et al.*, 2007] and community guided attachment (CGA) model [Leskovec *et al.*, 2007]. The results show that performance of both DMS and CDPAM to solving SNR problem is not satisfactory. FFM and CGA model are able to capture a few properties of the real-world networks for a specific choice of domains of its parameters but have more time complexity compared to the proposed parametric model. Degree distribution, distribution of clustering, hops, triangles, network values, singular values, and algebraic connectivity are considered to verify the performance of different considered parametric models and proposed model to solve the problem of SNR.

1.2.2 Diffusion on Networks

Biased random walk and information diffusion (Chapter 7)

It is well known that the links of real-world networks are dynamic. There are examples of real-world networks in which links between nodes are due to their interactions and the frequency of interactions defines the communicability or transition probability of that link. Consideration of weight of interactions can provide more insight into diffusion dynamics. In this chapter we consider the heterogeneity of link activity to study the diffusion or random walk dynamics. A biased random walk is proposed in this chapter which supports the biased growth of real-world networks. Practically, searching in a network or information diffusion are not continuous long lasting processes. These processes stop after some time and a new thread of similar processes can start in the same network which can be modelled as discontinued truncated random (or biased) walk dynamics.

In this chapter we provide an analytical explanation for preferential growth of networks by defining and analyzing diffusion dynamics on the networks in the form of biased random walks. Diffusion phenomena on networks are studied in many scenarios and contexts such as information diffusion and random walk in searching strategies. We analytically show that the proposed biased random walk supports that *'being a friend of a rich person is beneficial'* in terms of mean-first-passage-time (MFPT). Hence an analytical reasoning in support of the preferential growth of real-world networks is established. Observing the fact that all most all the random walk processes truncate at a finite time, we introduce two discontinued truncated biased random walk processes which are inspired by information diffusion in real world networks. These processes have applications in truncated searching on complex networks. The dynamics of these processes reveal the importance of local structure (degree distribution of first neighbors) in diffusion and searching. Repeated truncated random walk also known as a random walk with jumping phenomena is applied to calculate *PageRank* efficiently.

Diffusion protocols for link failure detection and resource utilization (Chapter 8)

In this chapter we propose a diffusion protocol by utilizing the structural property of the network that can help to identify the structural irregularities of the network. Thus, identification of link failure, distribution of resources, and avoidance of underutilization of resources are few practical problems which can be considered in an effective way.

Several applications in a large network depend on diffusion of information from one node to others. We propose a diffusion protocol for networked multi-agent systems based on both the structure of the network and the priority of agents. The agents (nodes) interact with their neighbors for the diffusion of information based on the weighted difference of the resources available at the neighboring nodes. We observe that the system finally reaches to a weighted agreement which is proportional to the priority of every node, or the degree of the nodes. We also perform the convergence analysis of the proposed protocol. The analysis has been done on both static networks as well as dynamic networks. We propose three different applications where such diffusion protocols can be used: (i) in the detection of link failure, (ii) for ensuring security and utilization of network resources, and (iii) to come up with a static fixed point convergence over a dynamic network. The impact of stability and convergence of the proposed diffusion protocol are also analyzed through simulation results under different test scenarios. Simulated results validate the theoretical findings of the diffusion protocol and its applicability under different applications.

Network reconstruction using diffusion dynamics (Chapter 9)

From various studies; it is evident that diffusion pattern in a network is directed by the diffusion rules and underlying network topology. In this chapter we do analysis of network structure retrieval from the diffusion pattern. In other words, we can say reverse engineering of diffusion dynamics to identify the connectivity pattern of nodes in a network. A framework is proposed to extract network structure from the given data of SIS diffusion dynamics.

Uncovering the heterogeneous connection pattern of the networked systems from the available status-time-series (STS) data of a dynamical process on the network is of great interest in network science and known as a reverse engineering problem. The outcome of the dynamics on a network is related to the structure of the network. Information of the structure of the network can help to devise the control of dynamics on the network. In this chapter, we consider the problem of network reconstruction from the available status-time-series (STS) data using matrix analysis. The proposed method of network reconstruction from the status-time-series data is tested successfully under SIS diffusion dynamics on the real-world and computer-generated benchmark networks. The connection (adjacency) matrix is reconstructed using the corresponding matrix generated by STS data. High accuracy and simplicity of the reconstruction procedure from the status-time-series data define the novelty of the method. Our proposed method outperforms the compressed sensing theory (CST) based method of network reconstruction using STS data. Further, the same procedure of network reconstruction is applied to the weighted networks. The ordering of edges in the weighted network is identified with high accuracy.

…