ABSTRACT

Title of dissertation: DYNAMICS ON AND OF COMPLEX NETWORKS: FUNCTIONAL COMMUNITIES AND EPIDEMIC SPREADING

Sanjeev Kumar Chauhan
Doctor of Philosophy, 2012

Dissertation directed by: Professor Michelle Girvan
Department of Physics

The work presented in this thesis focusses on two topics: functional communities and epidemic spreading on dynamic networks. The first part of the thesis focuses on a functionally-based definition of community structure for complex networks. In particular, we consider networks whose function is enhanced by the ability to synchronize and/or by resilience to node failures. For networks whose functional performance is dependent on these processes, we propose a method that divides a given network into communities based on maximizing a function of the largest eigenvalues of the adjacency matrices of the resulting communities. We also explore the differences between the partitions obtained by our function-based method and the structure-based modularity approach. A major finding is that, in many cases, modularity-based partitions do almost as well as the function-based method in finding functional communities, even though modularity does not specifically incorporate consideration of function. We also discuss the spectral properties of the

networks with community structure, relevant for the case of functional communities studied in this thesis.

In the second part of the thesis, we study a discrete time SIR model on dynamic networks. In our dynamic network model, we consider the case where the nodes in the network change their links both in response to the disease and also due to social dynamics. We assume that the individuals trying to make new connections mix randomly, and, with a certain probability, we also allow for the formation of new susceptible-infected links. We find that increasing the social mixing dynamics inhibits the disease's ability to spread in certain cases. This occurs because susceptibles who randomly disconnect from infected individuals preferentially reconnect to other susceptibles, inhibiting the disease spread. Finally, we also extend our dynamic network model to take into account the case of hidden infection. Here we find that, as expected, the disease spreads more readily if there is an initial time period during which an individual is infectious but unaware of the infection.

DYNAMICS ON AND OF COMPLEX NETWORKS: FUNCTIONAL COMMUNITIES AND EPIDEMIC SPREADING

by

Sanjeev Kumar Chauhan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Michelle Girvan, Chair/Advisor
Professor Edward Ott, Advisor
Professor Rajarshi Roy
Professor Thomas M. Antonsen
Professor Thomas E. Murphy, Dean's Representative

To my family.

Acknowledgments

So a long journey comes to an end. But an end of a journey means the beginning of a new one. There are many people who helped me get to this point. Unfortunately, it would not be possible to mention all of them. But I owe them a great debt of gratitude.

I am very previlidged to have worked with my advisors, Dr. Ott and Dr. Girvan. I appreciate the valueable opportunity they have given to me to work with them. Many thanks to them for being exceptional advisors, teachers and mentors. Despite them being very busy, they had always been there whenever I needed them. They taught me many thing about doing research. Thanks to both of them for their continued support, help and guidance through out these years.

I would also like to thank Prof. S. H. Patil at the Physics Department, IIT Bombay for his help and guidance during my undergraduate years. Thanks to Geet Duggal who had an infulence on me during my initial years in my current research group. I could not thank enough Wai-Shing Lee for all the useful discussions I had with him on a wide range of topics. He had been a true friend, and I learned a great deal from him.

I am also greatful to all the group members, and Professors and students in the Physics Departement who made my stay at Maryland a pleasing experience. Aditionally, I would like to thank the staff at Maryland, especially Jan Hessing, for their many favors and assistance.

Finally, thanks to my parents and my sisters, Sheetal, Shikha and Prerna,

for thier love, support and encouragement. You have been more patient than me. Thank you for believing in me and for all the sacrifices you made. This thesis is dedicated to you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many dynamical processes occur on top of networks, where the networks provide the underlying topology representing the interaction pattern of the fundamental units called nodes. In a broad range of systems, the changing structure of the networks over time (dynamics of the networks), by itself, is also an important feature. When a system includes interplay between both dynamics *on* the networks and dynamics *of* the network, its characteristics become intricate but show rich behavior. The problem is then not only to determine how the structure influences dynamics, but also how the dynamical criterion constrain the topological properties of the network. We expect these co-evolving networks to show behavior that may be quite different from the models without co-evolution.

In many studies, static topology of the networks is considered. While it can be argued that almost all real-world networks are co-evolving to a certain extent, considering network connectivity to be static can be a good assumption in various cases. An extensive research effort is focused on the effect of static connectivity pattern on the dynamical processes occurring over the networks and linking network topology to some dynamical or functional criterion.

The first part of this thesis considers a dynamical definition of communities in which we consider static networks and assume that the communities form such

that they have a functional (and hence dynamical) meaning. In general, a community is understood as a group of network nodes that "interact" more strongly with each other than with nodes outside their community. Our problem is to identify and characterize communities based on a functional criterion. The second part of this thesis focuses on epidemic spreading over complex network. Here, we consider dynamic networks with infection spreading over them. Thus, in this case, network structure is dynamically changing which influences and is influenced by the disease spread on the network.

1.1 Functional Communities

As discussed later in Chap. 3, the usual methods to define and identify communities are based on the structural criterion of the networks. A widely used example of this case is modularity [1, 2], which bases its definition of communities on the assumption that a community is significant when the nodes belonging to a community have more links to nodes in their own community than what is expected when the connections in the network form at random. While there has been a significant research effort devoted to defining communities using structural methods, there may be cases where the communities form to maximize a dynamical function. In this thesis, we specifically consider the case where the communities are thought to form so that they have better synchronizability and/or robustness to random node failures. In previous work, it has been shown that the largest eigenvalue of the network adjacency matrix determines the onset of both synchronization [3,4] and percolation

transitions [5]. Specifically, it has been shown that higher the largest eigenvalue, better the synchronizability of the network and/or robustness to random node failures. We use this observation in Chap. 3 to identify communities by maximizing a function of the largest eigenvalues of their adjacency matrices.

In our method to find functional communities, the number of communities is an input to the algorithm. We determine the number of communities in the networks from the eigenspectra of their adjacency matrices. A consequence of Perron-Frobenious theorem [6] for non-negative matrices is that for connected undirected networks and directed networks with a strongly connected component, the eigenvalue of largest magnitude of the network adjacency matrix is real and positive. We use this result in Chap. 2 to argue that, in many cases, the number of communities in a network can easily be determined from the eigenspectrum of its adjacency matrix. Specifically, when a network has N_c number of communities, the eigenspectrum shows N_c eigenvalues that are substantially larger than the bulk of the rest of the eigenvalues.

1.2 Epidemic Spreading on Dynamic Networks

Real human contacts change over time which can have significant influence on the spread of diseases. To study the effect of time varying contacts, we need to consider dynamic network models in which links are constantly being formed and dissolved. The model considered in this thesis tries to capture some of the essential features found in real social networks. In our model, we allow the nodes in the

network to change their links both in response to the disease (evasion of infected individuals by the susceptible individuals) and also due to social dynamics (individuals delete their links independent of their disease status and seek new partners). We study discrete time SIR model where R is considered removed from the process. In our studies, we focus on how network dynamics influence the threshold value of transmission probability above which the disease infects a finite fraction of the population in the limit of infinite system size.

We also extend the above model to take into account the case of hidden infection. In the case of hidden infection, for some period of time, a newly infected individual is not known to have the disease although he or she is already able to infect susceptible individuals. Thus, the evasion of infected individuals in such a state can not occur and hence the disease spreads more aggressively through the population.

1.3 Outline of Thesis

Each of the main chapters in this thesis (Chaps. 2, 3 and 4) are self contained and can be read independently. In Chap. 2, we discuss the eigenspectra of networks with community structure. In Chap. 3, we discuss the function based communities. In Chap. 4, we discuss epidemic spreading on our dynamic network model.

These chapters are based on the following published, submitted and in preparation articles:

- Chapter 2: S. Chauhan, M. Girvan and E. Ott, *Spectral properties of networks*

with community structure, Phys. Rev. E 80, 056114 (2009).

- Chapter 3: S. Chauhan, M. Girvan and E. Ott, *Detecting functional communities in complex networks*, under review.

- Chapter 4: S. Chauhan and M. Girvan, *Epidemic spreading in dynamic social networks*, in preparation.

- Appendix B: S. Chauhan, *Estimating the maximum and minimum degree correlations in large directed networks*, in preparation.

Chapter 2

Spectral Properties of Networks with Community Structure

2.1 Introduction

Many real complex networks are characterized by the presence of community structure; *i.e.*, there are groups of network nodes that have relatively stronger relationship with nodes in their own group than with nodes outside. Such structures can have significant influence on the functional characteristics of the network. There has been considerable research on developing techniques for finding community structure [7–9], and this continues to be an active area of research. Many community finding algorithms are based on the concept of modularity [10–13] which divides a network into communities by maximizing this quantity.

Spectral properties of the Laplacian matrix of networks with communities have also been studied quite intensively. These properties can be used to detect community structure in complex networks [14, 15]. There has been work that uses synchronization dynamics to find community structure and relate it to the spectral information of the Laplacian matrix [16, 17]. The eigenspectra of undirected "real-world" networks without community structure has been studied in Ref. [18, 19]. Here, by eigenspectrum of a network we mean the spectrum of its adjacency matrix. Farkas *et al.* [18] studied the spectral density of the sparse uncorrelated random graphs, the small-world graph and the scale-free graph, and their deviation from

the well know semi-circle law [20, 21]. Goh *et al.* [19] analyzed the eigenspectra and eigenvectors of the evolving Barabasi-Albert scale-free networks [22–24]. Random uncorrelated graphs have been used by physicists to study various physical phenomenon, and much work has been done exploring the spectral properties of such matrices [25].

To our knowledge, the eigenspectra of networks, directed or undirected, with community structure has gained little or no attention. The objective of this chapter is to study the spectral properties of network adjacency matrix with community structure. In particular, we propose a method for finding the *number* of communities in a network from the eigenspectrum of the network adjacency matrix.

Any given network can be represented by its adjacency matrix, A. In the case of unweighted networks treated here, $A_{ij} = 1$ if there is a link from node j to node i, and $A_{ij} = 0$ otherwise, where $i, j = 1, 2, ..., N$, and N is the number of network nodes. In the case of directed (undirected) networks, in general, $A_{ij} \neq A_{ji}$ ($A_{ij} = A_{ji}$). Our interest is primarily in the case where N is large and A is sparse. As we shall show, the eigenspectrum of the adjacency matrix of a network with communities has the interesting property that it has multiple eigenvalues that are well separated from the rest of the eigenvalues. Our main point in this chapter is that in many cases, the number of such eigenvalues often gives a clear indication of the number of communities in the network.

The organization of this chapter is as follows. As background, in Sec. 2.2, we discuss the pattern formed by plots in the complex plane of the eigenvalues of the adjacency matrix of a network with no communities, illustrating the generic oc-

currence of a cloud of $(N-1)$ eigenvalues of magnitude substantially less than the maximum eigenvalue which is real and positive. Section 2.3 discusses the eigenspectra of networks with communities. We show how the number of communities can be obtained from the eigenspectra of the network adjacency matrix. In Sec. 2.4 we apply our method to some real world networks. In Sec. 2.5 we discuss limitations of our method.

2.2 Eigenvalue Spectra of Networks without Communities

2.2.1 The Perron-Frobenius Eigenvalue

The Perron-Frobenius theorem for matrices with non-negative entries implies that the eigenvalue of A of largest magnitude, here denoted λ_*, is real and positive [6]. As an example, Fig. 2.1(a) shows a plot of the location of all the eigenvalues of the adjacency matrix of a $N = 500$ node Erdős-Rényi directed network with $\langle d^{in} \rangle = \langle d^{out} \rangle = 20$, where $\langle ... \rangle$ denotes the average over all nodes ($i = 1, 2, 3, ..., 500$) and d_i^{in} (d_i^{out}) denotes the number of incoming (outgoing) network links at node i [these numbers are also called the in-degree (out-degree) of node i]. Note that since every out-link originating from a node is also an in-link for some other node, we necessarily have $\langle d^{in} \rangle = \langle d^{out} \rangle$; thus we use the notation $\langle d \rangle$ to denote both $\langle d^{in} \rangle$ and $\langle d^{out} \rangle$. For the example in Fig. 2.1(a), we have taken d^{in} and d^{out} at a node to be uncorrelated. By uncorrelated in/out degrees, we mean that the joint in-degree/out-degree probability distribution function, $\tilde{P}(d^{in}, d^{out})$, giving the

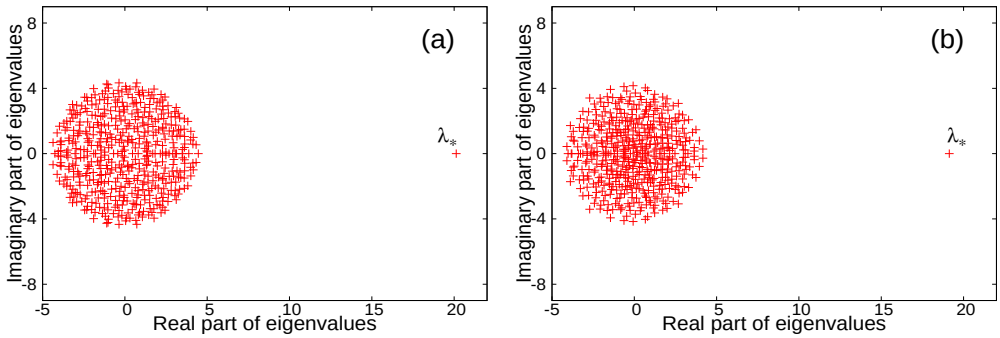


Figure 2.1: Plots of the real and imaginary parts of the adjacency matrix eigenvalues for computer generated directed networks with no community structure. The largest eigenvalue, λ_*, can be seen outside the cloud of the rest of the eigenvalues. (a) Erdős-Rényi network with N=500, $\langle d\rangle$=20. (b) scale-free network with N=500, γ=2.5, $\langle d\rangle$=20.

probability of (d^{in}, d^{out}) at a randomly chosen node, factors

$$\tilde{P}(d^{in}, d^{out}) = P_{in}(d^{in})P_{out}(d^{out}), \tag{2.1}$$

and as a consequence $\langle d^{in}d^{out}\rangle = \langle d^{in}\rangle\langle d^{out}\rangle = \langle d\rangle^2$.

We see in Fig. 2.1(a) that there is a single real positive eigenvalue $\lambda_* \cong 20$, while all the other 499 eigenvalues fall in a circular cloud centered approximately at the origin and entirely enclosed within a radius, denoted λ_0, of about 4, which is substantially less than the maximum eigenvalue $\lambda_* \cong 20$. Thus, there is a large gap between the Perron-Frobenius eigenvalue, λ_*, and the other eigenvalues. Assuming that, aside from the in-degree/out-degree correlation at a node, the network correlations are otherwise random, the mean field approximation to λ_* is (see Ref. [26])

$$\lambda_* \cong \frac{\langle d^{in}d^{out}\rangle}{\langle d\rangle}. \tag{2.2}$$

For an uncorrelated case, *i.e.*, $\langle d^{in}d^{out}\rangle = \langle d\rangle^2$, as in Fig. 2.1, the mean-field approximation to λ_* is $\lambda_* \cong \langle d\rangle$, in agreement with the numerically found value. On the other hand, as shown in Sec. 2.2.2 the root mean square radius of the

cloud has an upper bound given by $\langle d \rangle^{1/2}$ thus explaining the separation of λ_* from the other eigenvalues. Figure 2.1(b) is a plot similar to that in Fig. 2.1(a), but for the case of a scale-free network with degree distribution as in Eq. (2.1) with $P_{in}(d) = P_{out}(d) \sim d^{-2.5}$; as for the case illustrated in Fig. 2.1(a) the network is again randomly connected with $N = 500$, $\langle d \rangle = 20$. Again we see a strong separation between the Perron-Frobenius eigenvalue and the cloud formed by the other 499 eigenvalues.

In networks that are undirected (*i.e.*, $A_{ij} = A_{ji}$), all eigenvalues are real, but a similar result still often applies: All the non-Perron-Frobenius eigenvalues lie in an interval approximately centered at zero with root mean square radius which, as shown in next subsection, scales no stronger than $\langle d \rangle^{1/2}$, and $\lambda_* - \lambda_0$ can be large. As can be seen from Eq. (2.2),

$$\lambda_* \cong \frac{\langle d^2 \rangle}{\langle d \rangle} \tag{2.3}$$

for undirected networks. Note that $\frac{\langle d^2 \rangle}{\langle d \rangle} > \langle d \rangle$ (by the Schwartz inequality). Ref. [18, 19] have also given some results concerning separation between the largest eigenvalue and the bulk of eigenvalue cloud for certain undirected networks.

2.2.2 Size of the Cloud of non-Perron-Frobenius Eigenvalues

In case of undirected Erdős-Rényi networks, the semi-circle law predicts the size of the eigenvalue cloud as $\sim 2\sqrt{Np(1-p)}$ [18], where p is the probability of connection between two nodes. The distribution of eigenvalues in the cloud in this case is symmetric. For undirected scale-free networks, the spectral density deviates

from the semi-circle law. It resembles a symmetric triangle like distribution with power law tail of the density of the eigenvalues [18, 19].

For any given network, directed or undirected, we now show that the root mean square radius of the cloud of non-Perron-Frobenius eigenvalues has an upper bound given by $\langle d \rangle^{1/2}$, independent of whether the degrees are correlated or not. Since A has entries either 1 or 0 for the edges, the trace of $A^T A$, where A^T is the transpose of A, is equal to the total number of directed edges, say M, in the network,

$$\mathrm{Tr}(A^T A) = M. \tag{2.4}$$

The matrix A can be expressed in Schur decomposition form [27] as

$$A = UQU^*, \tag{2.5}$$

where U is a unitary matrix and U^* denotes its conjugate transpose. Q is an upper triangular matrix which can be written as $D+T$, where D is a diagonal matrix with the eigenvalues of A being the diagonal entries, and T is a strictly upper triangular matrix. From this, since A is real,

$$A^* = A^T = UQ^*U^*. \tag{2.6}$$

Thus for $\mathrm{Tr}(A^T A)$, we obtain

$$\mathrm{Tr}(A^T A) = \mathrm{Tr}(UQ^*U^*UQU^*) = \mathrm{Tr}(Q^*Q), \tag{2.7}$$

where we have used the fact that trace is invariant under a similarity transformation and U is unitary.

In Eq. (2.7), $\mathrm{Tr}(Q^*Q)$ is equal to $\mathrm{Tr}(T^*T) + \sum_{k=1}^{N} |\lambda_k|^2$. Since $\mathrm{Tr}(T^*T)$ is real and positive, Eq. (2.4) yields

$$\sum_{k=1}^{N} |\lambda_k|^2 \leq M. \tag{2.8}$$

For large N,

$$\langle |\lambda_k|^2 \rangle_{k \neq 1} \leq (M - \lambda_*^2)/N, \tag{2.9}$$

where $\langle ... \rangle_{k \neq 1}$ denotes the average over all eigenvalues with $\lambda_1 \equiv \lambda_*$ not included. The equality holds when the network is undirected. Since, in large sparse networks, $M >> \lambda_*$ and $M = N\langle d \rangle$, we get an upper bound on the root mean square radius of the eigenvalue cloud as

$$\langle |\lambda_k|^2 \rangle_{k \neq 1}^{1/2} \leq (M/N)^{1/2} = \langle d \rangle^{1/2}. \tag{2.10}$$

Figure 2.2 shows a plot of the largest eigenvalue, λ_*, and the actual radius of the cloud, λ_0, with changing network sizes for random computer-generated, directed, in/out-degree-uncorrelated networks. Plots for both Erdős-Rényi and scale-free networks are shown. Figure 2.2(a) is for the case where $\langle d \rangle = 20$ is held constant as N increases. Figure 2.2(b) is for the case where $\langle d \rangle / N = 1/20$ is held fixed as N increases. The upper solid lines in Fig. 2.2(a) and Fig. 2.2(b) correspond to $\langle d \rangle$, while the lower ones correspond to $\langle d \rangle^{1/2}$. We see that $\lambda_* \cong \langle d \rangle$ for uncorrelated directed networks, in agreement with Eq. (2.2). The actual radius of the cloud (not the root mean square radius), on the other hand, for this particular case of uncorrelated directed networks is approximately equal to $\langle d \rangle^{1/2}$. Thus we see that, for the cases shown, the largest eigenvalue is well-separated from the cloud of the rest of the eigenvalues, and, as the average degree of the network increases, the separation

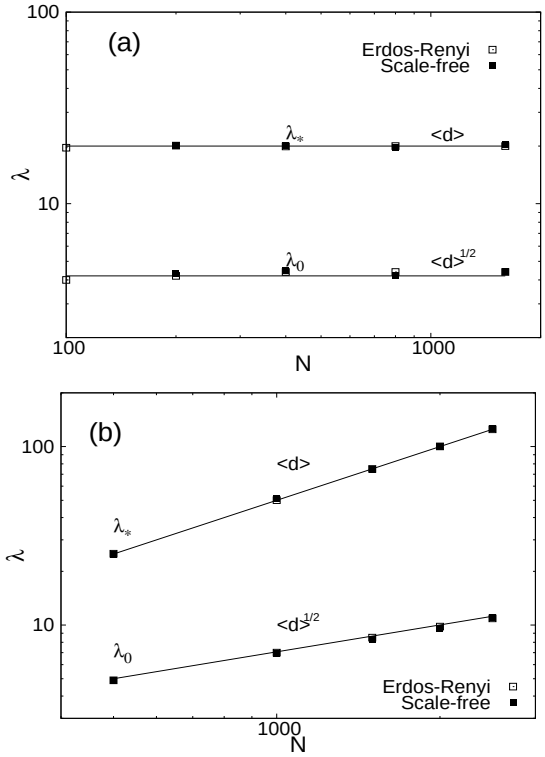


Figure 2.2: Plot of the largest eigenvalue, λ_*, and the actual radius, λ_0, of the eigenvalue cloud for networks with no communities versus the number of nodes in the network. All networks are directed with no degree correlations. (a) Erdős-Rényi and scale-free networks with constant degree, $\langle d \rangle = 20$. (b) Erdős-Rényi and scale-free networks with degree increasing in proportion to N such that $\langle d \rangle / N = 0.05$. In plots (a) and (b), the data points for Erdős-Rényi and scale-free networks overlap.

between them increases. Figure 2.3 shows a similar plot for undirected networks. In this case too, we see the large separation between λ_* and λ_0. All scale-free networks considered in Fig. 2.2 and Fig. 2.3 have degree distribution, $P_{in}(d) = P_{out}(d) \sim d^{-\gamma}$, with the exponent $\gamma = 2.5$.

We note that, although we have only presented illustrative numerical results for random networks, we have also conducted extensive tests for networks with other structures (e.g., assortative and disassortative networks) obtaining similar results.

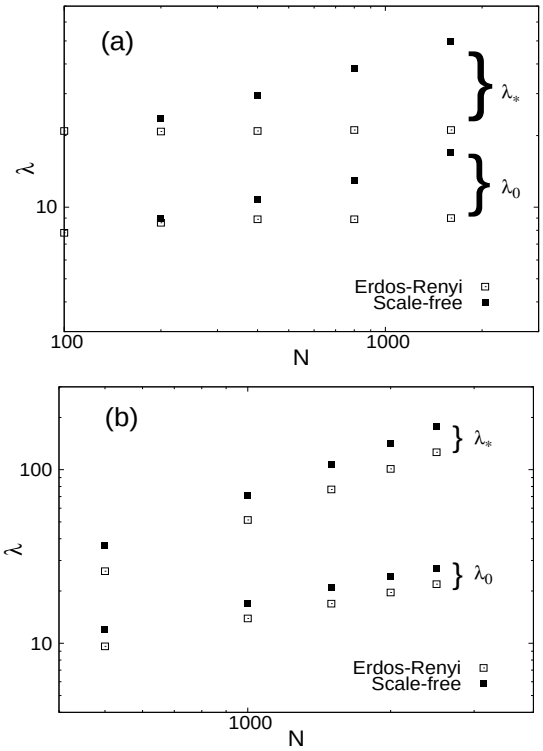


Figure 2.3: Plot of the largest eigenvalue and the actual radius of the cloud for networks with no communities versus the number of nodes in the network. All networks are undirected. (a) Erdős-Rényi and scale-free networks with constant degree, $\langle d \rangle = 20$. (b) Erdős-Rényi and scale-free networks with $\langle d \rangle / N = 0.05$.

2.2.3 Shape of the Cloud of non-Perron-Frobenius Eigenvalues

For a network with zero or few number of bidirected edges, the cloud of non-Perron-Frobenius eigenvalues is circular. Here by a bidirected edge we mean a pair of directed edges corresponding to $A_{ij} = A_{ji} = 1$ for nodes i and j. However, for a network where the number of bidirected edges is comparable to M, numerical computations show that the cloud shape becomes elliptical. In the limiting case where we have all bidirected edges, *i.e.* the case of undirected networks ($A_{ij} = A_{ji}$), the cloud collapses to a line interval on the real axis. This transition from circle to ellipse to line interval can be understood by considering the trace of A^2 which

is equal to the sum of the squares of the eigenvalues of A. Topologically, the trace of A^2 is equal to the number of directed cycles of length two, which in turn equals twice the number of bidirected edges in the network. Thus,

$$\sum_{k=1}^{N}\{[\Re(\lambda_k)]^2 - [\Im(\lambda_k)]^2\} = 2M_b. \tag{2.11}$$

where M_b is the number of bidirected edges in the network, and $\Re(.)$ and $\Im(.)$, respectively, denote the real and imaginary parts of their arguments. In above equation, we have used the fact that complex eigenvalues occur in conjugate pairs. Now for the networks with no self loops, $\langle\Re(\lambda_k)\rangle = \langle\Im(\lambda_k)\rangle = 0$, since $\mathrm{Tr}(A) = 0$. Thus for $M_b >> \lambda_*^2$, the difference in the spread of real and imaginary parts of the eigenvalues in the cloud is given by

$$\sigma^2[\Re(\lambda_k)]_{k\neq 1} - \sigma^2[\Im(\lambda_k)]_{k\neq 1} \cong \frac{2M_b}{N}, \tag{2.12}$$

where $\sigma^2[.]$ denotes the variance of the corresponding entries. The size of the term on the right hand side of Eq. (2.12) determines the ellipticity of the eigenvalue cloud for networks with zero or very small number of self loops. Thus, the ellipticity of the eigenvalue cloud measures the number of pairs of nodes in the network that have direct mutual relationship with each other (*i.e.*, are joined by bidirected links). In the normalize form, for the large sparse networks, we can write the ellipticity of the eigenvalue cloud as $\frac{2M_b}{M}$ which has the property that $0 \leq \frac{2M_b}{M} \leq 1$.

In general, the distribution of eigenvalues in the cloud of non-Perron-Frobenius need not be symmetric and the cloud may be asymmetric. This happens when the

odd moments ($\mathcal{M}_j, j = 3, 5, 7, ...$) of the graph spectral density are non-zero, where

$$\mathcal{M}_j = \frac{1}{N} \sum_{k=1}^{N} \lambda_k^j = \frac{1}{N} \mathrm{Tr}(A^j). \tag{2.13}$$

Topologically, $\mathrm{Tr}(A^j)$ counts the number of $j-$hop closed paths in the network. Farkas *et al.* [18] consider the case of undirected small-world networks in which $\mathcal{M}_3$ is high because of high value of clustering (density of graph triangles). Accordingly, they find high skewness in the spectral density of the small-world graphs.

2.3 Networks with Communities

In order to see how the phenomenon of Fig. 2.1 (*i.e.*, the appearance of λ_* well outside the cloud of other eigenvalues) is affected by the presence of community structure, we give several numerical examples in Sec. 2.3.2. Analytical results describing the behavior of largest eigenvalues observed in Sec. 2.3.2 are given in Sec. 2.3.3. Before presenting our numerical results in Sec. 2.3.2, we give our method of generating directed networks with community structure.

2.3.1 Generating directed networks with communities

In our numerical experiments in Sec. 2.3.2, we consider two types of networks. One of them is the Erdős-Rényi type directed network with communities with random placement of both within community and between community links. The second type of network is the scale-free network with communities with power law degree distribution. To generate Erdős-Rényi type directed networks with communities, we divide the N nodes in the network into the desired number of communi-

ties, say N_c. Communities could have equal or unequal number of nodes as required. Elements A_{ij} of the adjacency matrix corresponding to links between nodes within the same community are set to 1 with some chosen probability (else they are zero), while elements corresponding to links between nodes in different communities are made 1 with some other, smaller, probability. By changing these probabilities we can tune the strength of community structure and the average degree in the network.

To generate scale-free directed networks with community structure, we again start by dividing the nodes into the desired number of communities. For making connection between nodes in the same community, we generate power law degree distribution, $P(d) \propto d^{-\gamma}$, for both the in-degrees and the out-degrees of the nodes in the community. Say the k^{th} community has N_k nodes. We generate N_k numbers using the formula [28]:

$$b(m + m_0 - 1)^{-1/(\gamma-1)} \tag{2.14}$$

for $m = 1, 2, 3, ..., N_k$. Here, the constants b and m_0 determine the maximum degree and node averaged degree. We randomly assign these N_k numbers to the N_k nodes in community k and call these assigned numbers the target within community in-degree of the corresponding node i. We denote this number $t^{in}_{i,k}$. We then repeat the random assignment of these numbers and call the result the target within community out-degree of node i, $t^{out}_{i,k}$. Note that $t^{in}_{i,k}$ and $t^{out}_{i,k}$ are assigned independently at random, so that they are uncorrelated. From these target degree sequences, we obtain the $(i, j)^{th}$ entry of the adjacency matrix A, where i and j are in community k, by

setting $A_{i,j} = 1$ with probability

$$p_{ij}^{k} = \frac{1}{M_k} t_{i,k}^{in} t_{j,k}^{out}, \tag{2.15}$$

where M_k is the target number of edges between nodes in community k. Note that $M_k = \sum_i t_{i,k}^{in} = \sum_i t_{i,k}^{out}$. Links between communities are assigned in a similar manner. For example, say we want to generate links pointing from nodes in community l to nodes in community k. For N_k nodes in community k, we generate N_k numbers using Eq. (2.14). We assign these N_k numbers to nodes in community k, and call them the target in-links from nodes in community l to nodes in community k, $t_{i,kl}^{in}$ for the i^{th} node in community k. We repeat this procedure to get target out-links from nodes in community l to nodes in community k, $t_{j,kl}^{out}$ for the j^{th} node in community l. For a link from node j to node i we then use the probability,

$$p_{ij}^{kl} = \frac{1}{M_{kl}} t_{i,kl}^{in} t_{j,kl}^{out}, \tag{2.16}$$

where M_{kl} is the target number of between community links pointing from nodes in community l to nodes in community k. While generating these target degrees, we choose our constants b and m_0 in Eq. (2.14) such that $M_{kl} = \sum_i t_{i,kl}^{in} \approx \sum_j t_{j,kl}^{out}$. We repeat this procedure for all pairs of communities. While assigning the target values for the number of links to each node, we assign higher $t_{i,kl}^{in}$ and $t_{j,kl}^{out}$ to nodes with higher $t_{i,k}^{in}$ and $t_{j,l}^{out}$, respectively. Similarly, nodes with smaller within community target links get smaller between community target links. Using this procedure, we get power law distribution for both within community and between community in-degrees and out-degrees.

2.3.2 Numerical results

In this subsection, we will verify numerically that, when the network has N_c communities, the eigenvalue plot shows N_c eigenvalues outside the cloud of non-Perron-Frobenius eigenvalues. We consider two cases of networks with $N = 2000$ nodes consisting of four communities:

case (i): The communities have different sizes; $N_c = 700, 600, 400$ and 300.

case (ii): All the communities are of equal size; $N_c = 500$ for each of the four communities.

For the case where the average degree of nodes in a community is proportional to the number of nodes in a community, case (i) leads to the situation where the largest eigenvalues of communities that are 'disconnected' (*i.e.*, there are no between community links) are non-degenerate, while for case (ii) the largest eigenvalues will be approximately degenerate. Figure 2.4 shows the eigenvalue plot for a computer generated Erdős-Rényi type network and for a scale-free network for case (i). Figure 2.4(a) is for the Erdős-Rényi type network, and Fig. 2.4(b) is for the scale-free network with $\gamma = 2.5$ in Eq. (2.14). For the Erdős-Rényi type network used to get the eigenvalue plot in Fig. 2.4(a), the probability of connection between pairs of nodes within same community was 0.04. With this, the average degree of nodes in a community is proportional to the number of nodes in the community. For between community edges, the probability of connection between pairs of nodes was 0.015. With these parameters, the sum of the number of edges within all communities equals the number of all between community edges. The average degree of nodes in

the network is $\langle d \rangle \approx 44$. For generating the scale-free network for the plot in Fig. 2.4(b) the number of edges within communities and between pairs of communities was the same as the number of edges for the Erdős-Rényi type network described above. For within community links, the maximum degree in the sequence from Eq. (2.14) was one fifth of the total number of nodes in the community. For between community links for a pair of communities, the maximum degree was one tenth of the number of nodes in the smallest community from the pair.

In both cases in Fig. 2.4, it is evident that there are four real positive eigenvalues that occur outside a circular shaped cloud formed by the remaining 1996 non-Perron-Frobenius eigenvalues. For comparison, we indicate by vertical dashed lines the four largest (real) eigenvalues that would result if the between community links of these networks were removed. For the smallest community, the number of in-links (and also out-links) from other communities was approximately twice the number of within community links. In this case, we still see the perturbed largest eigenvalue of this community outside the cloud of non-Perron-Frobenius eigenvalues.

Figure 2.5 shows the eigenvalue plot of a computer generated Erdős-Rényi type (Fig. 2.5(a)) and a scale-free (Fig. 2.5(b)) network with $\gamma = 2.5$ in Eq. (2.14) for case (ii). For Erdős-Rényi type and scale-free networks, the network generation parameters are chosen such that the nodes on an average have 20 within community in/out links and 20 in/out links to nodes not in their community. Again, it is clearly evident that there are four eigenvalues occurring outside the cloud of 1996 non-Perron-Frobenius eigenvalues. If the between community links of these networks are removed, the four eigenvalues are nearly degenerate with an average

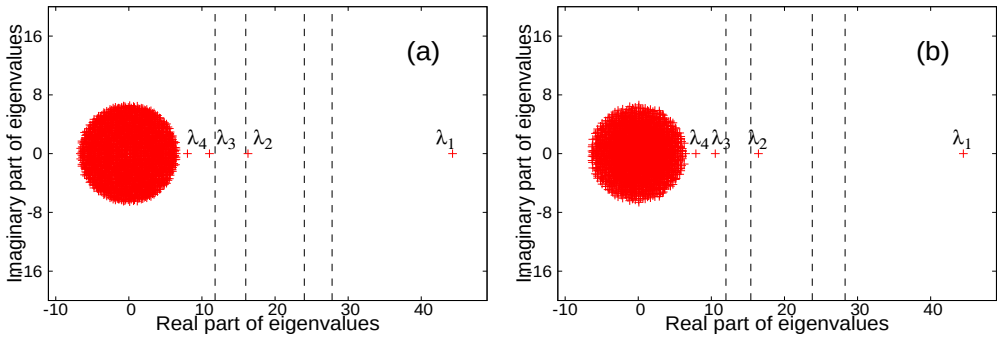


Figure 2.4: Plot of real and imaginary parts of eigenvalues of computer generated directed networks with unequal sized communities. (a) Erdős-Rényi type network and (b) scale-free network. The average number of within community and between community links are equal in the two cases. We see four eigenvalues corresponding to four communities outside the cloud of rest of the eigenvalues.

value indicated by the vertical dashed line. In both Figs. 2.5(a) and 2.5(b) we see that three of the eigenvalues outside the cloud cluster tightly together, while the larger of the four eigenvalues outside the cloud has a substantially bigger value. This largest eigenvalue is always real and positive (by Perron-Frobenius theorem). The triplet of other three larger eigenvalues, in general, could have a complex conjugate pair. Furthermore, when we take the average of these four eigenvalues, this average turns out to be very nearly equal to the degenerate value obtained with the between community connections removed. This observed structure will be explained further in our analysis in Sec. 2.3.3.2.

2.3.3 Perturbation theory

As verified numerically in the subsection above, when the network has N_c communities, the eigenvalue plot shows N_c eigenvalues outside the cloud formed by the rest of the eigenvalues. In order to understand this, consider the simple limiting

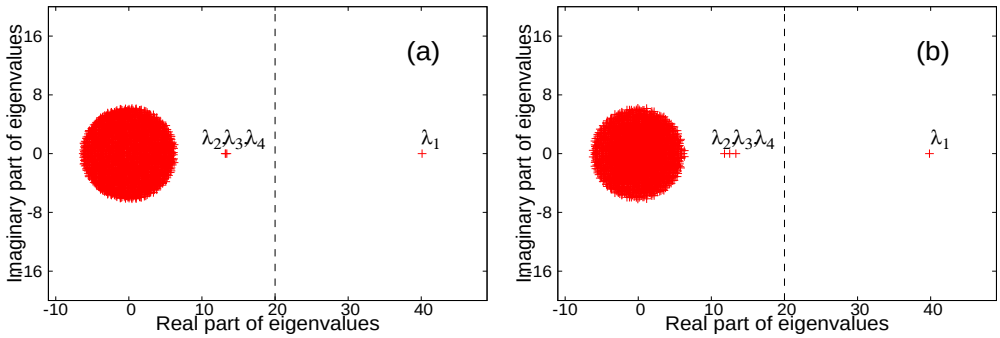


Figure 2.5: Plot of the real and imaginary parts of the eigenvalues of the adjacency matrix of computer generated directed networks with four equal sized communities. The four largest eigenvalues can be seen outside the cloud formed by of the rest of the eigenvalues. (a) Erdős-Rényi type network and (b) scale-free network.

case of a directed network with multiple communities where all the links exist within the communities, and there are no links between communities. In this case, with the proper labeling of the nodes, the adjacency matrix shows block diagonal structure (*i.e.*, there are N_c blocks along the matrix diagonal with $A_{ij} \equiv 0$ for (i, j) not in a block). Thus, the eigenvalues of the adjacency matrix are simply the union of the eigenvalues of the individual blocks. Hence, a plot of the real and imaginary parts of the eigenvalues of the adjacency matrix then has the largest eigenvalues of each of the communities outside the cloud of its other eigenvalues. In addition, these eigenvalues outside their community clouds are all positive and real. In the case where the smallest community Perron-Frobenius eigenvalue exceeds the largest of the radii of the community clouds, the adjacency matrix of the whole network will have N_c Perron-Frobenius eigenvalues outside the aggregate cloud formed by the individual community clouds. Furthermore, we claim that when links between communities are added, provided that the number of added links is not too great, the eigenspectrum still shows that the number of eigenvalues outside the cloud

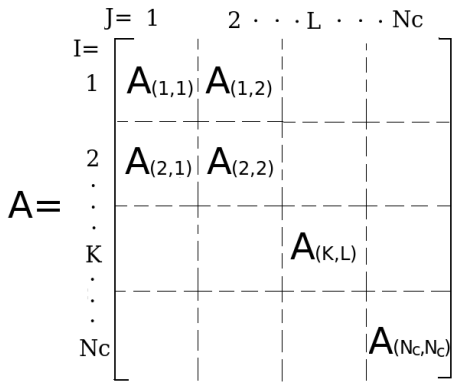


Figure 2.6: Adjacency matrix of a network with N_c communities, in block matrix form. Each diagonal block corresponds to the adjacency matrix of a community, while the off diagonal blocks correspond to links between communities.

corresponds to the number of communities N_c.

In order to analytically address the above claim, we will use perturbation theory by considering links between communities as a perturbation to the adjacency matrices of originally disconnected communities. First we consider the case of networks that have non-degenerate largest eigenvalues of disconnected communities, which corresponds to case (i) in Sec. 2.3.2. Following that, we consider networks that have degenerate (or nearly degenerate) largest eigenvalues of disconnected communities, which corresponds to case (ii) in Sec. 2.3.2.

2.3.3.1 The non-degenerate case

In this subsection, we analyze the case of networks that have non-degenerate largest eigenvalues of disconnected communities. We will show that the largest eigenvalues of the disconnected communities have lowest nonzero perturbative correction of second order when addition of between community links is treated as a perturbation.

Consider the case of networks that have N_c unequal sized communities, each having unequal (*i.e.*, non-degenerate) largest eigenvalues when the communities are disconnected. Let A denote the adjacency matrix of such a network. With proper labelling of the nodes, the matrix A will have block matrix structure with $N_c \times N_c$ number of blocks. Blocks on the diagonal correspond to the adjacency matrices of the individual communities, while the off-diagonal blocks correspond to the perturbation (connections between communities). Let us denote by (I, J) the block of A (Fig. 2.6). When $I = J$, the block is the adjacency matrix of community I, while if $I \neq J$ then $A_{(I,J)}$ corresponds to the block of the adjacency matrix in which links pointing from community J to community I are stored. Now, let us write A as

$$A = A_0 + \delta A, \tag{2.17}$$

where A_0 is a matrix whose diagonal block elements are the diagonal block elements of A and whose off-diagonal block elements are zero. δA is a matrix with zeros on its diagonal blocks, and with off-diagonal block elements being the off-diagonal blocks of A. For the case where between community connections are sufficiently sparser than within community link, we regard δA as a perturbation to A_0.

We denote the N_c non-degenerate largest eigenvalues of A_0 by λ_{*k}, where $k = 1, 2, ..., N_c$. Let U_k be the right eigenvector of A_0 corresponding to the eigenvalue λ_{*k}, where entries in U_k are zero except for those elements corresponding to community k in A_0. Write the perturbations of U_k and λ_{*k} due to δA in Eq. (2.17), as

$$U'_k = U_k + \delta U_{k,1} + \delta U_{k,2}, \tag{2.18}$$

$$\lambda'_{*k} = \lambda_{*k} + \delta\lambda_{*k,1} + \delta\lambda_{*k,2}, \tag{2.19}$$

where the subscripts 1 and 2 denote first and second order corrections. Letting V_k denote the left eigenvector of A_0 corresponding to the eigenvalue λ_{*k}, and multiplying $AU'_k = \lambda'_{*k}U'_k$ from the left by V_k, we obtain

$$\delta\lambda_{*k,1} + \delta\lambda_{*k,2} = V_k\delta A\delta U_{k,1}, \tag{2.20}$$

where we have made use of $V_k\delta AU_k = 0$ which follows from the facts that δA is zero on its diagonal blocks, while both V_k and U_k are non-zero only for their entries corresponding to community k. Since we assume δA to be small, the right side of Eq. (2.20) is of second order, and hence, $\delta\lambda_{*k,1}$ is zero. Therefore, the lowest nonzero correction to the largest eigenvalues is of second order,

$$\delta\lambda_{*k,2} = V_k\delta A\delta U_{k,1}. \tag{2.21}$$

This shows that the largest eigenvalues of disconnected communities that have non-degenerate largest eigenvalues are perturbed more weakly than the perturbation applied.

First order correction, $\delta U_{k,1}$, to the eigenvector U_k is given by [29]

$$\delta U_{k,1} = \sum_{r\neq k}^{N} \frac{(V_r\delta AU_k)}{(\lambda_{*k} - \lambda_r)}U_r, \tag{2.22}$$

where U_r and V_r are, respectively, the right and left eigenvectors of A_0 corresponding to its eigenvalue λ_r. Here, V_r is the row vector and U_r is the column vector with the normalization condition $V_rU_r = 1\ \forall\ r$ and $V_rU_s = 0$ for $r \neq s$ $(r, s = 1, 2, ..., N)$.

We tested our calculations, specifically Eq. (2.21), by comparing with actual eigenvalues of some computer generated Erdős-Rényi type directed networks with four unequal sized communities. In Fig.2.7, we show comparison between actual and

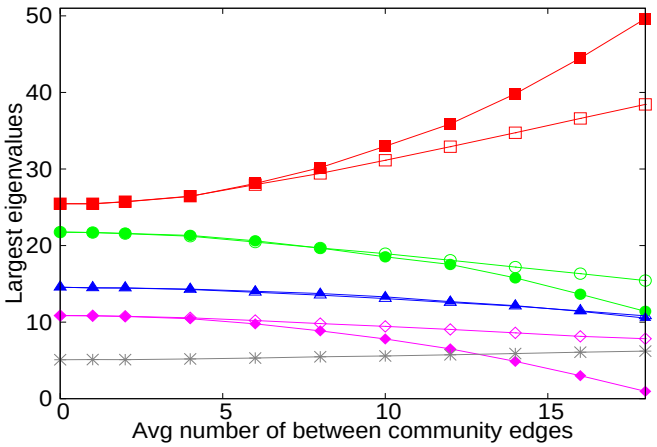


Figure 2.7: Comparison of the actual and predicted four largest eigenvalues with increasing between community edges for Erdős-Rényi type directed networks with four unequal sized communities. Squares ($\blacksquare, \square$) correspond to λ'_{*1}, circles ($\bullet, \circ$) to λ'_{*2}, triangles ($\blacktriangle, \triangle$) to λ'_{*3} and diamonds ($\blacklozenge, \lozenge$) correspond to λ'_{*4}. Open symbols correspond to actual values while the filled ones are the estimated values calculated using second order perturbation theory. The symbol $\divideontimes$ shows the actual radius of the non-Perron-Frobenius eigenvalue cloud. All data points are averaged over 20 simulated networks. Error bars are smaller than the symbol sizes. Lines are just a guide to the eye.

predicted four largest eigenvalues of the network adjacency matrix with increasing between community links. The networks have $N = 2000$ with 700, 600, 400 and 300 nodes in each community. The probability of connection between pairs of nodes in the same community was 0.037 which gives $\langle d \rangle \approx 20$ for the whole network when there are no between community links. When there are non-zero links between communities, to get an estimate of the four perturbed largest eigenvalues, we numerically calculate the four largest eigenvalues of the disconnected communities and add the lowest order correction given by Eq. (2.21). As can be seen, our perturbation calculation predicts the four largest eigenvalues well when the number of between community links is small. The radius of the cloud (the symbol $\divideontimes$) given in Fig.2.7 is the actual radius of the disc of the non-Prrron-Frobenius eigenvalues found by

numerically calculating all the eigenvalues of the network adjacency matrix. Figure 2.7 also shows that when the number of between community links is large, we can still see the actual perturbed largest eigenvalue of the smallest community outside the cloud of the non-Perron-Frobenius eigenvalues.

2.3.3.2 The degenerate case

We now consider the case of networks that have N_c initially disconnected equal sized communities, each with N/N_c nodes and with similar number of within community edges. In this case, each of these disconnected communities will have approximately equal largest eigenvalues. We denote this approximately common eigenvalue by λ_*. As the perturbation is applied by adding between community links, we find that $(N_c - 1)$ of the N_c perturbed largest eigenvalues will become approximately equal and smaller than the remaining perturbed largest eigenvalue (as an example see in Fig. 2.5 for the case $N_c = 4$). The perturbation of these eigenvalues is such that the mean distance of all these N_c largest eigenvalues from their initial value is zero. The adjacency matrix A will have block matrix structure with $N_c \times N_c$ number of blocks of equal sizes of dimension $N/N_c \times N/N_c$. As before, we write $A = A_0 + \delta A$, with A_0 and δA being same as described in Eq. (2.17).

Let us write a right eigenvector, say U', of A which corresponds to one of the perturbed largest eigenvalues as

$$U' = \sum_{k=1}^{N_c} \alpha_k U_k + \delta U, \tag{2.23}$$

where α_k are the coefficients to be determined, δU is a higher order correction,

and U_k denotes the right eigenvector of the block matrix A_k corresponding to its maximum eigenvalue. All the blocks in matrix A_k are zero except for the diagonal block corresponding to community k. As a consequence, the entries in U_k will be zero except for those elements corresponding to community k in A_0. We regard δU as small since we regard the perturbation to be small. Note that as in Eq.(2.18), the perturbed eigenvector U' in Eq.(2.23) does not have subscript k corresponding to community k because we will have N_c such eigenvectors for different sets of coefficients α_k. We again denote by V_k the left eigenvector of A_0 corresponding to the maximum eigenvalue of A_k and assume that the eigenvectors of A_0 are normalized such that $V_k U_k = 1$.

Multiplying $A_0 + \delta A$ from right by U' and from left by V_l, and keeping terms up to first order we get

$$\sum_{k \neq l} y_{lk} \alpha_k = \alpha_l (\lambda'_* - \lambda_*), \tag{2.24}$$

where λ'_* is the perturbed eigenvalue and $y_{lk} = V_l \delta A U_k$. For N_c different V_l eigenvectors, we will have N_c such equations corresponding to $l = 1, 2, ..., N_c$ in Eq. (2.24).

For the case in which we have equal sized communities that have similar number of within and between community links with the same degree distribution (similar perturbation for all the communities), all the y_{lk} coefficients are approximately the same. Thus, to simplify our calculation and to get qualitative results, we assume that $y_{lk} = y \ \forall \ l, \ k$ with $y > 0$. Equations (2.24) are an eigenvalue problem of the

form $C\alpha = \lambda'_* \alpha$ with $\alpha = [\alpha_1, \alpha_2, ..., \alpha_{N_c}]^T$, and

$$C = y\tilde{\mathbf{1}}_{N_c} + (\lambda_* - y)\mathbf{1}_{N_c}, \tag{2.25}$$

where $\tilde{\mathbf{1}}_{N_c}$ is a $N_c \times N_c$ matrix all of whose entries are ones, while $\mathbf{1}_{N_c}$ is the $N_c \times N_c$ identity matrix. The eigenvectors of C are thus the eigenvectors of $\tilde{\mathbf{1}}_{N_c}$. One such eigenvector is $[\ 1\ 1\ 1\ ....\ 1\ ]^T$ corresponding to an eigenvalue of C equal to $\lambda_* + (N_c - 1)y$. The other $N_c - 1$ eigenvectors of $\tilde{\mathbf{1}}_{N_c}$ correspond to the $N_c - 1$ dimensional space of vectors $[\alpha_1, \alpha_2, ..., \alpha_{N_c}]^T$ such that $\sum_i \alpha_i = 0$. For all these vectors, the eigenvalue of $\tilde{\mathbf{1}}_{N_c}$ is zero, corresponding to $N_c - 1$ degenerate eigenvalues of C given by $\lambda_* - y$. This suggests that there is a largest perturbed eigenvalue, approximately given by $\lambda_* + (N_c - 1)y$, which is larger than the rest of the $N_c - 1$ degenerate eigenvalues, approximately given by $\lambda_* - y$, which are clumped together (as can be seen in Fig. 2.5 for $N_c = 4$). Note that the average of the N_c perturbed eigenvalues is λ_* (the unperturbed degenerate eigenvalue). It can be shown with simple argument that y scales no stronger than N.

Figure 2.8 gives comparison between actual eigenvalues and our calculations of this subsection. The networks considered are the Erdős-Rényi type directed networks with four equal sized communities. The networks have $N = 2000$ with 500 nodes in each community. The within community link probability is 0.04 which gives $\langle d \rangle \approx 20$ when the communities are disconnected. The estimate of the perturbed four largest eigenvalues was calculated by numerically finding the unperturbed largest eigenvalues and using the estimate of y, which is calculated by averaging over all the 12 possible y_{lk}'s. As can be seen in Fig. 2.8, our calculations agree well with

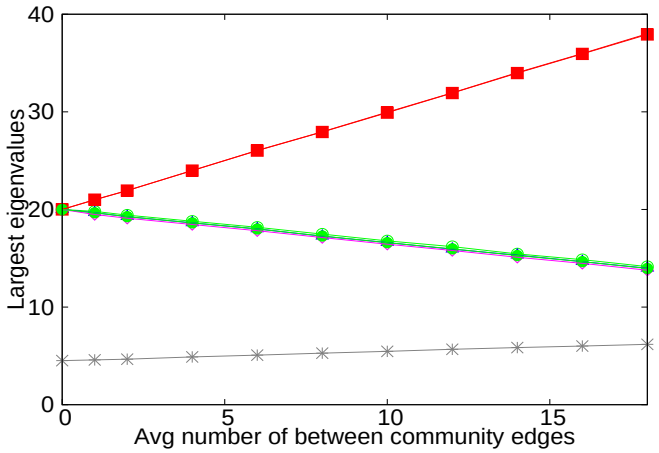


Figure 2.8: The actual and predicted four largest eigenvalues with increasing between community edges for Erdős-Rényi type directed networks with four equal sized communities. Squares (■ for predicted, □ for actual) correspond to λ_1. Data points for predicted values overlap with the actual ones. The symbols ⋇ show the actual radius of the non-Perron-Frobenius eigenvalue cloud. Rest of the data points correspond to to λ_2, λ_3 and λ_4 which are all approximately equal. Here, the data points for predicted λ_2, λ_3 and λ_4 lie on top of each other and overlap with the actual ones. All data points are averaged over 20 simulated networks. Error bars are smaller than the symbol sizes. Lines are just a guide to the eye.

the actual values.

Thus our numerical results of Sec. 2.3.2 seem to be quite well explained by our perturbation results of the present subsection (Sec. 2.3.3) even though the 'perturbations' for the numerical examples of Sec. 2.3 are not small (*e.g.*, Fig. 2.4 and Fig. 2.5).

2.3.4 Discussion

Based on our perturbation analysis, one might suspect that, since the unperturbed eigenvectors U_k and V_k have non-zero element values only for their community entries, it might be possible to use the eigenvectors of the adjacency matrix A to obtain the communities, and not just their number. In initiating our research re-

ported here, we were originally motivated by this possibility. However, as described below, we found this to be problematic.

In Sec. 2.3.3.1, the eigenvectors of the matrix A corresponding to the largest eigenvalues of communities are denoted by U'_k. When δA is small, the entries in eigenvector U'_k that are labeled by nodes belonging to nodes in community k will have larger magnitude compared to entries labeled by nodes not in community k. For a given node i, by comparing entries labeled by node i in eigenvectors U'_k, for $k = 1, 2, ..., N_c$, we can assign node i to the group of nodes that have largest magnitude of the corresponding entry in the same eigenvector.

Our experimentation with this method on some computer generated networks shows that the method works pretty well when the eigenvalues of disconnected communities are non-degenerate and the perturbation is not too large. This method fails, however, when the maximum eigenvalues of disconnected communities are too close and the perturbations are too large. When the maximum eigenvalues of disconnected communities are nearly degenerate, an indication of the difficulty is provided by Eq. 2.23 which shows that the perturbed eigenvector can have almost equal contribution from all the unperturbed U_k eigenvectors ($k = 1, 2, ..., N_c$).

2.4 Application to Real Networks

We now test our prediction on two real networks for which we show eigenvalue plots in Fig. 2.9. The networks considered are the political books network [30] and the political blogs network [31]. These two examples are convenient for our purpose

since we naturally have the division of the network into two major groups based on left/liberal or right/conservative orientation of the book or the blog. The political books network is an undirected network. The nodes represent books on politics available from the online retailer Amazon.com. There is an edges between two nodes when the same buyer(s) buys books represented by the nodes. The Political blogs network, on the other hand, is the compilation of network data on US political weblogs as recorded by Adamic and Glance [31] in 2005. It is a directed network where the edges represent hyperlinks between the weblogs on US politics.

The total number of nodes in the political books network is 105. We show the eigenvalue plot of the adjacency matrix of the political books network in Fig. 2.9(a). Since it is an undirected network, the adjacency matrix is symmetric, and all eigenvalues are consequently real. We 'estimate' the size of the eigenvalue cloud by the magnitude of the most negative eigenvalue. The vertical dashed line in Fig. 2.9(a) corresponds to this value. Consistent with the prediction in this chapter, we see that there are two eigenvalues substantially to the left of this dashed line (λ=11.9, 11.6).

The political blogs network is a relatively larger network as compared to the political books network. It is a directed network with 1224 nodes. The eigenvalues of the adjacency matrix of this network are in general complex since this is a directed network (Fig. 2.9(b)). The cloud of eigenvalues is substantially contracted towards the real axis. For this network we have $M_b = 2307$ and $M = 19022$. The difference in the spread of real and imaginary parts of eigenvalue cloud (left hand side of Eq.(2.12) with two largest eigenvalues excluded) is 2.22. Again we 'estimate' the cloud size

from the magnitude of the most negative eigenvalue (vertical dashed line). The two eigenvalues of magnitude 34.5 and 26.9, corresponding to the two communities, can be seen separated from the rest of the cloud by a large amount.

In Fig.2.9(a) and (b), we see that there are few eigenvalues that lie just outside (to the right of) the vertical dashed line. These eigenvalues lying close to the vertical dashed line can not be said to belong to any particular community with any degree of certainty. For networks where the eigenvalue cloud is symmetric, as can be seen for the computer generated networks considered in this chapter, the size of the cloud can be well estimated by looking at the eigenvalue of largest magnitude with negative real part. However, for many real networks, as discussed in subsection 2.2.3, the eigenvalue cloud may not be symmetric. For the political books network we calculated the clustering coefficient, given in Ref. [32], which we found to be relatively high (a value of 0.348). For the political blogs network, we found relatively high values of first few odd moments of the spectral density, an order of magnitude higher, compared to the randomly generated scale-free networks with similar degree distribution and two communities. These findings suggest that the clouds are right-skewed and should actually extend past the vertical dashed line.

2.5 Limitations in determining the number of communities

The method we propose in this chapter for finding the number of communities works best when the node average degrees within communities are of same order. Limitation to this method occurs when one or more of the communities are much

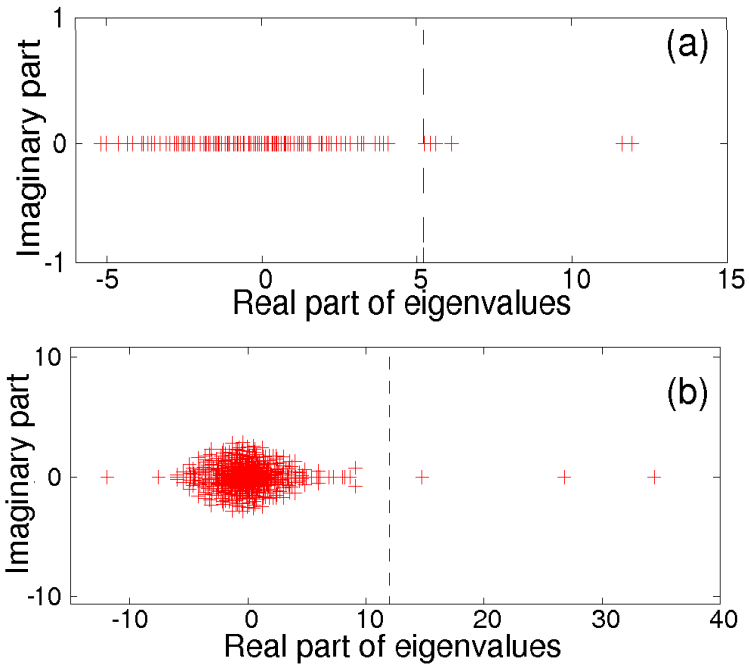


Figure 2.9: Plots of the real and imaginary parts of the eigenvalues of adjacency matrix of real networks. (a) Political books network. (b) Political blogs network.

smaller compared to the largest community, or when a community has sparser within community connections compared to other communities. In particular, even in the absence of perturbation ($\delta A = 0$ in Eq. (2.17)), the maximum eigenvalue of the smaller community can lie inside the cloud of non-Perron-Frobenius eigenvalues of the largest community. This puts a limitation on the sizes of the communities that can be detected using our method. For example, in the simplest case where the in and out-degrees are uncorrelated and $\delta A = 0$, this happens when, $\langle d \rangle_s \leq \langle d \rangle_l^{1/2}$, where $\langle d \rangle_s$ is the average degree of a smaller community and $\langle d \rangle_l$ is the average degree of the largest community. In the case of network communities where the average degree of nodes is proportional to the number of nodes within communities, this condition roughly translates to the statement that when $N_s \leq N_l^{1/2}$, we will not

be able to detect smaller communities with N_s nodes when the number of nodes in one of the largest community is N_l.

As discussed in subsection 2.2.3, in case of networks that have non-zero odd moments of the spectral density, the cloud of non-Perron-Frobenius eigenvalues may not be symmetric. As can happen in small-world networks without community structure with large clustering, the largest eigenvalue of the network adjacency matrix may not be well separated from the cloud of the non-Perron-Frobenius eigenvalues [18]. In case of networks with community structure, the skewed eigenvalue cloud may even overlap with the largest eigenvalues of the smaller communities. Thus, we may not be able to see them well separated from the eigenvalue cloud.

2.6 Conclusions

We studied the eigenspectra of adjacency matrix of large sparse networks. The eigenspectrum gives a clear indication of the number of "dominant" communities in the networks in certain cases. Here, by dominant we mean the communities whose eigenvalues lie outside the cloud of the non-Perron-Frobenius eigenvalues. We examine the eigenvalues of the network adjacency matrix and infer the number of communities by finding the number of eigenvalues falling outside a typically occurring dense cloud of eigenvalues. For the example of uncorrelated in/out-degrees, we argued that there is a large gap between the non-Perron-Frobenius eigenvalues and the Perron-Frobenius eigenvalue. Owing to this large gap (also seen more generally with in/out-degree correlation and assortative/disassortative networks), we

can determine the number of communities in a network, even when the community structure is not strong.

In this chapter, we have not specified exactly the radius of the eigenvalue cloud. While there are results on the spectral density of the eigenvalue cloud for Erdős-Rényi and scale-free undirected networks when the distribution of eigenvalues is symmetric, we still need to deal with the case when the odd moments of the spectral density are non-zero resulting in an asymmetric eigenvalue distribution.

Finding the number of communities from the eigenvalue plot could be helpful in some community finding algorithms (as in Chap. 3), where the number of communities is an input to the algorithm. The method has a limitation based on the relative sizes of the communities, and, in general, it may miss smaller or weaker communities (Sec. 2.5). Further limitations in determining the number of communities from the eigenvalue plot may occur when the eigenspectra is highly skewed because of non-zero odd moments.

Chapter 3

Detecting Functional Communities in Complex Networks

3.1 Introduction

Complex networks have received much attention from researchers in diverse disciplines. Networks serve as the structural underpinning of models for understanding properties of many real complex systems. They provide insight into the dynamical behavior and functional attributes of such systems. Over the last decade, interest in networks has grown substantially, partly spurred by the discovery of previously under-appreciated properties seen in real-world networks, e.g., small world behavior [33], scale-free degree distribution [22], assortative mixing [34], etc. The properties of networks have been studied at all levels, ranging from microscopic to mesoscopic to macroscopic. At the mesoscopic level, one potentially important property of networks is community structure. Roughly, a community can be defined as a group of network nodes that "interact" more strongly with each other than with nodes outside their community. Community structure has been shown to exist in many real networks [7, 31, 35–37]. Such structures can have significant influence on the organization and dynamics of the network as a whole. For example, communities might be substructures that represent functional units, as in some biological systems [38].

Much of the research related to community structure in networks has been

directed toward finding the "best" possible community partition of a network. Direct application of traditional computer science and sociological approaches for finding community structure in complex networks has been shown to be problematic [7, 8, 10]. Various methods have been proposed for detecting community structure in complex networks, *e.g.*, the edge betweenness method [7], the eigenvector method [10], methods based on simulated annealing [12], synchronization dynamics [16, 39], spectral analysis [14, 15], k-clique percolation [40], link communities [41], etc. Many community finding methods are based on modularity [1], which, for a given partition of nodes into communities, gives a structural measure of the goodness of that partition. In the definition of modularity, a community is considered to be a group of nodes within which connections are relatively dense compared to a suitable expectation. Reviews of structural based methods for dividing networks into communities (with most based on modularity) can be found in Refs. [8, 9]. An excellent overall review on community structure can be found in Ref. [42].

Our motivation for this chapter is that, as discussed above, in the past, the definition of a community has often been based on the structural features of networks, *e.g.*, modularity. In this chapter, we will adopt the view that, in many situations, the most appropriate way of defining a community may depend on the application that the resulting division will be used for, which, in turn, depends on the *function* of the network. For instance, we may desire a different definition of community structure if we are trying to find clusters of friends in a social network than if we are trying to find metabolic pathways in a biochemical network. One expects that a method designed for a particular consideration may not necessarily work in other

situations. In this chapter, as an example, we consider a particular network function and propose an alternate definition of communities for this kind of function. Specifically, we consider networks whose function is enhanced when the maximum eigenvalue, λ_*, of their adjacency matrix is large. Examples where this applies include synchronization of network coupled phase oscillators [3, 4] and percolation on directed networks [5]. Although we specifically consider directed networks in this chapter, the method can also be used to find communities in undirected networks.

It is not obvious that the partitions obtained using a structure based method will also correspond to good functional partitions. To analyze this, we explored the difference between the method presented in this chapter and the widely used modularity method, which is based purely on consideration of network structure. Although, we find cases where the two methods yield significantly different results (Sec. 3.5.2), we also find that, in many situations (Sec. 3.5.1), the partitions that maximize modularity also tend to score highly according to our functional measure, which we found rather surprising. Our results suggest that, in many cases, modularity maximization is effective in identifying functional communities.

The organization of this chapter is as follows. In Sec. 3.2, we review the largest eigenvalue of the adjacency matrix of networks without community structure and its relation to network functional properties. In Sec. 3.3, we define a largest-eigenvalue-based measure that can be used to determine community structure in networks. In Sec. 3.4, we describe the method used to detect community structure given our functional definition. The construction of networks with eigenvalue based communities is also discussed. In Sec. 3.5, we give results for the method proposed

in this chapter and compare these results with results from the modularity approach.

3.2 Network functions and the largest eigenvalue of the adjacency matrix

The largest eigenvalue of a network's adjacency matrix in the absence of community structure can be used to characterize both synchronization and percolation phenomenon. In this section, as background, we discuss the significance of the largest eigenvalue of network adjacency matrix for these network functions.

3.2.1 Synchronization

Synchronization is a population effect that emerges in many complex systems composed of a large number of dynamical components [43]. The classical model of Kuromoto describes the synchronization of phase oscillators that are uniformly globally coupled and have natural frequencies drawn from a heterogeneous distribution [44]. In the limit of large network size, a phase transition, separating the synchronized and the unsynchronized states, is observed for the Kuramoto model. For synchronization on networks with large average degree and arbitrary degree distribution, similar results have been reported [3, 4].

For synchronization of phase oscillators in complex networks, the evolution,

$$\dot{\theta}_i = \omega_i + K \sum_{j=1}^{N} A_{ij} \ \sin(\theta_j - \theta_i), \tag{3.1}$$

is considered, where θ_i and ω_i are the phase and intrinsic frequency of the i^{th} oscillator, K is an overall coupling strength, and N is the number of nodes in the network.

Here, A_{ij} is the $(i,j)^{th}$ entry of the adjacency matrix which has value 1 if there is a link from node j to node i; otherwise it is 0. The synchronization of nodes in the network can be characterized by the global order parameter, r, given by

$$r = \left| \frac{\sum_{j=1}^{N} e^{i\theta_j}}{N} \right|. \tag{3.2}$$

Perfect synchronization (typically occurring for $K \to \infty$) corresponds to $r = 1$. For large N, synchronized and unsynchronized behaviors of the system are signified by a value of r significantly above zero and close to zero, respectively.

For networks with large average degree and an arbitrary degree distribution, results based on mean field theory show that the critical value of coupling strength, which separates the synchronized and unsynchronized states, is determined by the first two moments of the degree distribution of the nodes [45, 46]. Restrepo *et al.* obtained better estimates of the critical coupling strength in the case of directed networks [3]. In particular, they show that the critical value of the coupling strength, K_c, is determined by the largest eigenvalue of the network adjacency matrix,

$$K_c = \frac{K_0}{\lambda_*}, \tag{3.3}$$

where K_0 is a constant which depends on the distribution of oscillator frequencies and is independent of the network characteristics. Thus, the higher the largest eigenvalue of the network adjacency matrix, the smaller the value of K needed to attain the phase transition to synchronization for such networks.

3.2.2 Percolation

Percolation is another network property that has been studied extensively. In the percolation model, a phase transition separates two phases characterized by the presence and absence of a giant connected component when nodes (site percolation) or links (bond percolation) are removed from the network. In undirected networks that do not have any degree correlations between linked nodes, the percolation transition has been shown to depend on the second moment of the degree distribution [47]. Percolation in case of directed networks has also been explored (*e.g.* see Refs. [48–51]).

Some approaches focus on a Markovian approach for studying percolation phenomenon [50–53]. Restrepo *et al.* [5] studied the percolation problem without the need of a Markov network model but requiring the knowledge of the network adjacency matrix. For directed networks that are locally tree like, they found that the percolation transition occurs when a fraction of nodes,

$$p_c = 1 - \frac{1}{\lambda_*}, \tag{3.4}$$

have been randomly removed from the network. This indicates that when the largest eigenvalue, λ_*, of the network adjacency matrix is high, the network can tolerate a large number of node deletions before it disintegrates.

3.3 A functional definition of community structure using eigenvalues

As discussed in Sec. 3.2, in the case of directed networks without community structure, larger values of λ_* make the network more resilient to breaking up into

many disconnected pieces when nodes are randomly removed (e.g., due to failure or attack). Furthermore, synchronization in a heterogeneous collection of phase oscillators is promoted by increasing λ_*. This suggests that, if a network's function depends on synchronization of heterogeneous oscillators and/or robustly maintaining connectivity, then consideration of the largest eigenvalues of the adjacency matrices of individual communities may provide a natural basis for a useful functional definition of community structure on such networks.

We propose a measure that is meant to quantify the strength of network division into communities that have better synchronizability and robustness to random node failures. Motivated by the role of the largest eigenvalue in both synchronization and percolation, our measure sums a monotonically increasing function of the largest eigenvalues of the communities. We view this as an example of a functional definition of communities that might be appropriate in some cases, but we also emphasise that other definitions would be appropriate for other purposes.

For clarity, we can write the adjacency matrix, A, of networks with community structure in block matrix form as shown in Fig.2.6. Each diagonal block of A then corresponds to the adjacency matrix of an individual community, while the off diagonal blocks correspond to the links between communities. We propose that, given a network, if we can find a partition of the network into communities that have higher largest eigenvalues of their corresponding diagonal block adjacency matrices, then those communities will have enhanced network functions.

Specifically, the definition of community structure that we study is as follows:

1. Consider a partition of a network into g communities.

2. Calculate the maximum eigenvalues $(\lambda_{*1}, \lambda_{*2}, ..., \lambda_{*g})$ of the adjacency matrices of all the communities. Here, λ_{*k} is the largest eigenvalue of the k^{th} diagonal block in Fig.2.6.

3. Define the "spectral cohesion":

$$\Lambda = \sum_{k=1}^{g} \ln(\lambda_{*k}). \tag{3.5}$$

The spectral cohesion, Λ, provides a functionally based measure of the community strength of a particular partitioning of the network. We can thus define the *best* division into g communities as the one that maximizes Λ, where we think of *best* as being with respect to the enhancement of synchronization or resilience. Note that the definition of communities according to Eq.(3.5) can be used for both symmetric and asymmetric matrices. In Sec. 3.5, we will demonstrate the utility of this definition for directed networks, in particular.

As an aside, we emphasize that our choice of the spectral cohesion function in Eq.(3.5) is somewhat arbitrary; *e.g.*, $\Lambda = \sum f(\lambda_{*k})$ for any function $f(\lambda)$ that is monotonically increasing with λ might alternatively be considered. However, we shall, in all of what follows, use $f(\lambda) = \ln(\lambda)$. This is partly motivated by the analogy to entropy, and by our studies with $f(\lambda) = \lambda^{\beta}$, for $\beta = 1$ and 2, which, for several test networks, yielded results that were very similar to those for $f(\lambda) = \ln(\lambda)$.

While the method of maximizing the spectral cohesion, Λ, gives us the best division into g communities, it does not tell us how to choose the appropriate value

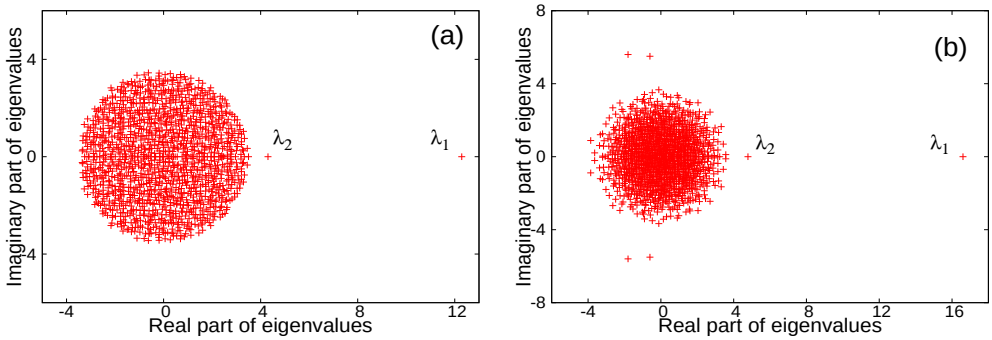


Figure 3.1: Eigenvalue plots for (a) a directed Erdős-Rényi type network , and (b) a directed scale-free network with two equally sized communities. By construction, the nodes in the network have $\langle d \rangle_I = \langle d \rangle_X = 6$, but the communities are defined such that they have maximal directional degree assortativity within them. The networks have $N = 1400$ and $\langle d \rangle = 12$. Here, λ_1 and λ_2 are the largest and the second largest real positive eigenvalues of the network adjacency matrix.

of g, i.e., the number of communities that the network contains. We have addressed this problem in Chap. 2, showing that the number of communities may be obtained from the eigenspectra of the adjacency matrix of the full network. Specifically, we show that for networks with communities, there typically exists a relatively small set of N_c positive real adjacency matrix eigenvalues that are significantly larger than, and well separated from the large number of other eigenvalues. This number of large positive eigenvalues is shown in Chap. 2 to provide an appropriate choice for g. Chapter 2 provides examples of eigenvalue plots for networks with community structure that have high density of links within communities and lower density of links between communities. In Fig. 3.1, we give examples of the eigenvalue plots for other types of networks with community structure. Figure 3.1(a) is for an Erdős-Rényi type directed network, and Fig. 3.1(b) is for a scale-free directed network. Both networks have two communities of equal sizes, $N = 1400$ and $\langle d \rangle = 12$, where $\langle d \rangle$ for a directed network here denotes average in-degree or out-degree, which are

both equal. In order to briefly describe our choice of networks for Fig.3.1, we first note a result for random networks without community structure. In particular, if there is directional degree assortativity (see Eq. (3.13) for the definition), then a mean field theory (described in Sec. 3.4.2) [26] shows that, other things being equal, λ_* is larger for networks with larger directional degree assortativity. The networks used to generate the plots of Fig. 3.1 have the property that the average number of in/out-links that connect a node to nodes in its own community, $\langle d \rangle_I$, is equal to the average number of in/out-links that connect the node to nodes in the other community, $\langle d \rangle_X$, but the communities are defined such that the communities have maximal directional degree assortativity within them (see Sec. 3.4.2 for details). Thus, in the absence of directional degree assortativity, the networks are random networks with no community structure.

3.3.1 Cycles in the graph and the largest eigenvalue

In this chapter, our aim is to find communities that have enhanced network functions which in turn depends on the largest eigenvalues of the community adjacency matrices. Although our choice of Λ in Eq.(3.5) to a certain extent was arbitrary, we can consider a useful interpretation of this function. Eigenvalues of the community adjacency matrices, A_k, are related to the cycles in the communities. The number of cycles of length n inside a community k equals the sum of the diagonal components of A_k^n which in turn equals $\sum_{i=1}^{N_k} \lambda_{ik}^n$, where λ_{ik} is the i^{th} eigenvalue of community k. Thus, for large n, the exponential growth of the number of cycles

with their length is

$$\lim_{n\to\infty} \frac{\ln\ (\sum_{i=1}^{N_k} \lambda_{ik}^n)}{n} = \ln\ \lambda_{*k}, \tag{3.6}$$

when the limit exist. Thus, when we use Λ, we expect to find communities that have high number of cycles within them.

3.4 Methods

3.4.1 Detecting functional communities

Thus far, we have proposed a quantity whose maximization, we hypothesize, should yield a good division of a network into communities for the network functions we are interested in. In this section, we provide an outline for a simulated annealing scheme [54] that finds a desirable division of the network. The advantage of this simulated annealing method is that it can provide a network division whose spectral cohesion is very close to the true maximal value. The disadvantage is that it is computationally quite intensive. In order to fairly compare our results with the modularity approach, we also use simulated annealing to find a network division that maximizes the modularity function for a fixed number of communities. The modularity function is based on a comparison between the number of links connecting nodes in the same community to the number expected in a random network without community structure. For directed networks, the modularity (Q) is defined as [2, 13]

$$Q = \frac{1}{m} \sum_{i,j} \left[A_{ij} - d_i^{in} d_j^{out}/m \right] \delta_{c_i,c_j}, \tag{3.7}$$

where d_i^{in} denotes the in-degree of node i, d_j^{out} denotes the out-degree of node j, and m is the number of edges in the network. c_i and c_j denote the community indices of nodes i and j, and $\delta_{c_i,c_j} = 1\ (= 0)$ if $c_i = c_j\ (c_i \neq c_j)$.

In our simulated annealing scheme, we begin by assigning nodes randomly to N_c different communities, where we find N_c as described in Sec. 3.3 and Chap. 2. We then choose a node at random and pick a random community, to which to consider moving it. If this move would result in an increase in the value of the function we are trying to optimize, say F (which could be either Λ or Q) , we perform the move. If the move would result in a decrease in the value of the function, we perform it with Boltzmann acceptance probability $e^{\Delta F/T}$, where $\Delta F < 0$ is the change in the function F and T is the 'temperature' (this is the basic Metropolis algorithm [55]). For each temperature value, we repeat this process αN^2 times, where N is the number of nodes in the network and α is a chosen factor ≤ 1. After αN^2 iterations, we reduce the temperature by a factor of 0.99. Using the parameters described, the whole process is repeated until an asymptotic value of F is reached.

In the case of the spectral cohesion, there is a caveat to the movement of nodes. In some networks, there are nodes that do not affect the eigenvalue of any of the communities. If such a node is chosen at a given iteration, it is moved to a community that is randomly selected at that iteration if it has more links to that community than its own, without regard to the directionality of the links. If it has fewer links to the randomly chosen community, the move is accepted with a probability which depends on the number of links the node has to both the communities. We expect this strategy to be reasonable only if, as in the numerical examples we

treated, we have a small number of such nodes in the network.

Since we are interested only in the largest eigenvalues of the matrices, we use the power method [56] to calculate these eigenvalues. Assuming that λ_{*k} is well separated from the other eigenvalues, for a dense $N_k \times N_k$ matrix, where N_k is the number of nodes in community k, the needed computational time for this approach to give the dominant eigenvalue of the matrix scales as $O(N_k^2)$. In the case of sparse matrices, the required number of operations needed to compute λ_{*k} scales as $O(M_k)$, where M_k is the number of non-zero entries in the matrix corresponding to community k. Assuming that we are working with sparse networks with $M_k \sim N_k$ (as is the case with many real-world networks) and that the number of required temperature reductions is independent of N (an optimistic assumption), assuming $N_k \sim N$, this yields an algorithm whose required number of operations scales at best as $O(N^3)$.

When maximizing the spectral cohesion, in many cases, run times of our simulated annealing program can be further reduced by using perturbation theory [28] for calculating the estimate of $\Delta\Lambda$ above. We accept or reject a move based on this estimate of $\Delta\Lambda$. When a move is accepted, we calculate only the eigenvalues of the communities involved in the change. When a move is rejected, we go to the next step. This is much less computationally expensive than recalculating the eigenvalues at each step of the simulated annealing procedure.

The use of perturbation theory for calculating the estimate of $\Delta\Lambda$ is explained as follows. When a node, say i, is chosen at random, we consider moving it from its current community, say k, to another community, say l, the estimated change in the

largest eigenvalue of the adjacency matrix of the node's current community, due to the removal of the node, is calculated using the approximation given in Ref. [28]. Let V_k and U_k, respectively, be the left and right eigenvectors of the adjacency matrix of community k corresponding to λ_{*k} and normalized so that $V_k^T U_k = 1$. Let $(V_k)_i$ and $(U_k)_i$ denote the components of the vectors V_k and U_k corresponding to node i. Then for $N_k >> 1$ and $(V_k)_i (U_k)_i << 1$, removal of node i leads to a change in λ_{*k}, which is approximately given by [28]

$$\Delta\lambda_{*k} = -\lambda_{*k}\ (V_k)_i (U_k)_i. \tag{3.8}$$

Similarly, we estimate the increase in the largest eigenvalue of community l, when we consider adding node i to it to be

$$\Delta\lambda_{*l} = \frac{(V_l^T \delta A_l)_i (\delta A_l U_l)_i}{\lambda_{*l}}. \tag{3.9}$$

Here, we assumed $\Delta\lambda_{*l} << \lambda_{*l}$ where λ_{*l} is the largest eigenvalue of the adjacency matrix of community l before addition of node i. In Eq. (3.9), δA_l is the perturbation applied to the adjacency matrix of community l due to the addition of node i, V_l and U_l are the left and right eigenvectors of the adjacency matrix of community l corresponding to λ_{*l} that satisfy the normalization condition $V_l^T U_l = 1$, and $(V_l^T \delta A_l)_i$ and $(\delta A_l U_l)_i$ denote the components of the corresponding vectors corresponding to node i. Note that δA_l is of dimension $(N_l + 1) \times (N_l + 1)$, because when we consider moving node i to community l, the number of nodes in community l becomes $N_l + 1$. All the elements of δA_l are zero except for the row and column corresponding to node i, which has 1's at appropriate locations corresponding to in-links and out-links to and from node i to nodes in community l. The vectors V_l

and U_l are $(N_l + 1)$ dimensional column vectors, with the entry corresponding to node i being zero. Thus, for N_k, $N_l >> 1$, the estimated change in the value of the spectral cohesion Λ, given by Eq.(3.5), due to the movement of the node is

$$\Delta\Lambda = \frac{\Delta\lambda_{*k}}{\lambda_{*k}} + \frac{\Delta\lambda_{*l}}{\lambda_{*l}}. \tag{3.10}$$

The above time-saving scheme is especially useful for large networks. The larger the network, the better the perturbation theory in estimating the change in the spectral cohesion.

Because we use a simulated annealing approach, our method for finding communities is more computationally demanding than many methods that have been proposed that are based on structural definitions of communities. Our goal here, however, is not to introduce the most efficient algorithm for finding community structure, but rather to test the degree to which a functional approach to finding communities may be appropriate in certain cases. We leave the development of fast algorithms that identify functional community structure for future work.

3.4.2 Construction of test networks with eigenvalue-based communities

In this section, we give methods for the construction of networks with eigenvalue-based communities. We will subsequently use these networks for our numerical experiments in Sec. 3.5.1. As preliminary preparation for explaining how we construct networks with eigenvalue-based communities, we first note two results relating λ_* to the topological properties of networks without communities [26].

3.4.2.1 The effect of node in/out-degree correlations

Considering random directed networks without community structure, if the network is characterized by a joint in/out-degree probability distribution $P(d^{in}, d^{out})$, then the expected value of the maximum eigenvalue is [26]

$$\lambda_* = \eta\langle d\rangle, \tag{3.11}$$

where $\langle d\rangle := \langle d^{in}\rangle = \langle d^{out}\rangle$, $\langle ...\rangle$ denotes an average over the network nodes, and η is the in/out-degree correlation coefficient,

$$\eta = \langle d^{in} d^{out}\rangle / \langle d\rangle^2. \tag{3.12}$$

Thus, in/out-degree correlation, $\eta > 1$ (anticorrelation, $\eta < 1$) increases (decreases) λ_*. Note that in the absence of node in/out-degree correlation, we have $\eta = 1$ and $\lambda_* \approx \langle d\rangle$.

In obtaining the estimate in Eq.(3.11), the network is imagined to be constructed by first randomly assigning each node values (d^{in}, d^{out}) according to P, and then randomly linking the nodes accordingly as described for the networks discussed in the Appendix A.

3.4.2.2 The effect of directional degree assortativity

We now consider random directed networks with uncorrelated in/out node degrees in the distribution $P(d^{in}, d^{out})$ that are assortative by degree according to the directed degree assortativity coefficient [26],

$$\rho = \langle d_i^{out} d_j^{in}\rangle_e / \langle d_i^{out}\rangle_e \langle d_j^{in}\rangle_e, \tag{3.13}$$

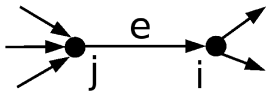


Figure 3.2: An example of one term in the average $\langle ... \rangle_e$ where $d_j^{in} = 3$ and $d_i^{out} = 2$.

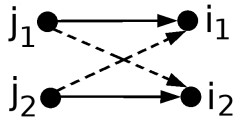


Figure 3.3: Illustration of a destination edge interchange.

where $\langle ... \rangle_e$ denotes the average over all the edges from node j to node i (Fig.3.2), but are otherwise random. In this case, the expected value of λ_* is [26]

$$\lambda_* = \rho \langle d \rangle. \tag{3.14}$$

Thus assortativity (corresponding to $\rho > 1$) increases λ_*, while disassortativity ($\rho < 1$) reduces λ_*.

Here the network is imagined to be constructed in two stages [26]. First a non-assortative and node degree uncorrelated network is randomly constructed (see Appendix A). Such a network will have $\rho \approx 1$ for large N. Next, to increase ρ to any desired target value, we first randomly choose two edges, $(j_1 \to i_1)$ and $(j_2 \to i_2)$ (see Fig.3.3). We then imagine that we interchange the destinations of these two links, thus producing two new links, $(j_1 \to i_2)$ and $(j_2 \to i_1)$. If ρ increases, we implement the change; if ρ decreases, we do not. We then randomly choose two new links, and successively repeat this process until ρ approximately reaches its target value.

Test networks: The above results can be used as a basis for the construction of networks with eigenvalue-based community structure. For example, consider

networks with two nominally equally sized communities. The communities can have any ratio of within to between community links but they also have η_c or ρ_c greater than one. Here η_c and ρ_c are defined by Eqs. (3.12) and (3.13) but with consideration restricted to only those nodes and links that lie within a community under consideration, and thus only using within community node degrees. That is, we produce higher maximum eigenvalues for the communities by increasing the within community in/out-degree correlation or directional degree assortativity. We consider both directed scale-free and Erdős-Rényi type networks of these types. In Sec. 3.5.1, we will use such networks in numerical experiments.

All the test networks of the type described above that are used in Section 3.5.1 have $N = 1400$ with two nominally equally sized communities. In our test networks in Section 3.5.1, we keep $\langle d \rangle_I = 6$ while changing $\langle d \rangle_X$. By doing this we make sure that the communities have same maximal values of directional degree assortativity and node degree correlations within them as we change $\langle d \rangle_X$. For the scale-free networks, the maximal attainable η_c was approximately 2.12, while the maximal attainable ρ_c was approximately 2.05. For the Erdős-Rényi type networks, the corresponding maximal attainable values were approximately 1.16 for both η_c and ρ_c. [1] In Section 3.5.1, we used these maximal situations such that both the communities either have maximal ρ_c and $\eta_c \approx 1$, or have maximal η_c and $\rho_c \approx 1$. In these situations, the values of λ_{*k} are substantially larger than would be obtained for a random partition of the network into two equally sized communities. In addition,

[1] Formulas for estimating the maximum and minimum attainable degree correlations in large directed Erdős-Rényi and scale-free networks are given in Appendix B.

for the networks with maximal directional degree assortativity within communities, in the eigenvalue plots in Fig. 3.1, we see two positive eigenvalues outside the cloud of the rest of the eigenvalues even with $\langle d \rangle_I = \langle d \rangle_X$, indicating the presence of two communities. More details on the methods of constructing test networks with eigenvalue based communities are given in the Appendix A.

3.5 Results

In this section, we report results from using the functionally motivated definition of community structure proposed in this chapter. For comparison, we also present results using the modularity method to find partitions in both artificial and real networks.

3.5.1 Structural identification

Here we present results from applying our eigenvalue-based measure Λ and modularity Q to divisions of test networks into two communities via the simulated annealing procedure described in Section 3.4.1. Results for Erdős-Rényi type networks are shown in Fig. 3.4, while results for scale-free networks are shown in Fig. 3.5. Each data point in Figs. 3.4 and 3.5 represents an average over 20 random network realizations.

Figures 3.4 (a) and 3.5 (a) show the spectral cohesion Λ versus $\langle d \rangle_X$ for three different sets of network parameters [ρ_c maximized with $\eta_c \approx 1$; η_c maximized with $\rho_c \approx 1$; and $(\rho_c, \eta_c) \approx (1, 1)$] when Λ is maximized (plotted as solid triangles) and

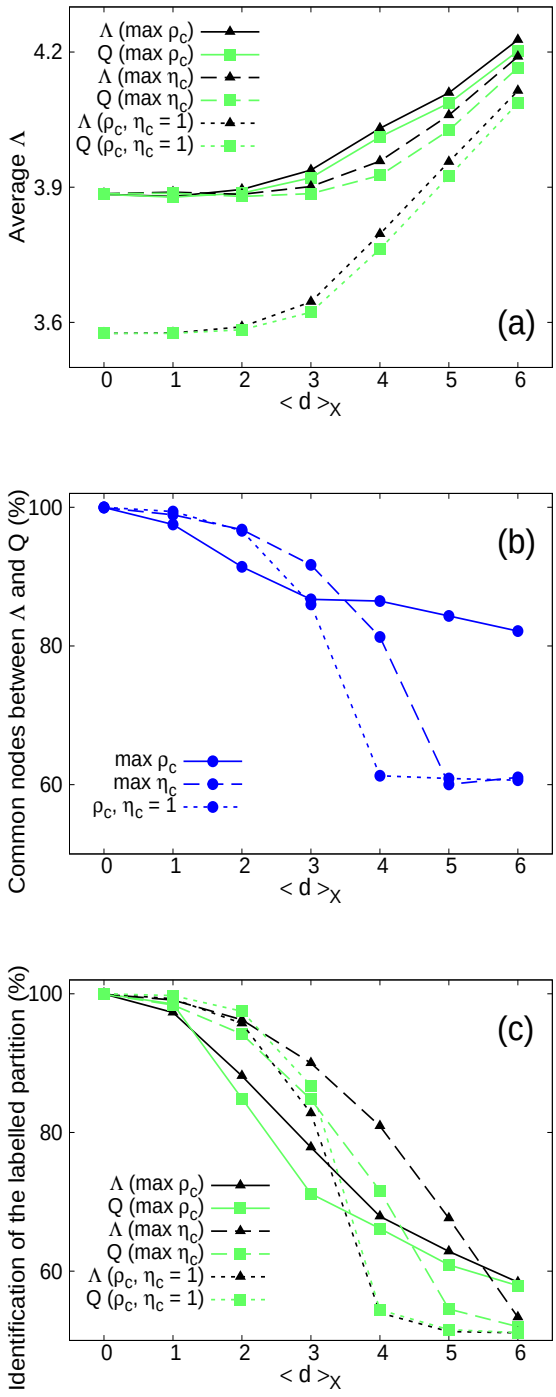


Figure 3.4: Erdős-Rényi type networks with $N = 1400$, $\langle d \rangle_I = 6$ and two communities. (a) The average value of Λ function (Eq. 3.5) for the partitions obtained by maximizing the spectral cohesion function and the modularity function. (b) Average percent of nodes common between communities obtained by maximizing Λ and Q. (c) Average percent nodes of the labelled partition identified by Λ and Q. Dark (black) colored curves are for the spectral cohesion function while the light (green) colored curves are for the modularity function. Data points represent averages over 20 simulated networks.

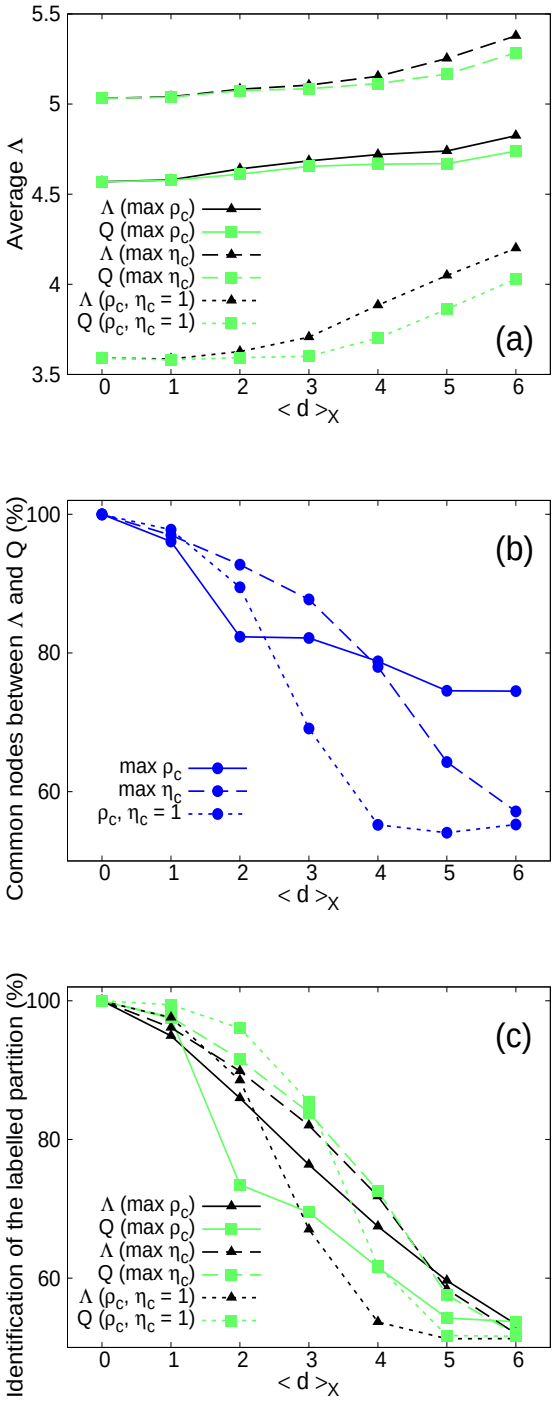


Figure 3.5: Scale-free networks with $N = 1400$, $\langle d \rangle_I = 6$ and two communities. (a) The average value of Λ function (Eq. 3.5) for the partitions obtained by maximizing the spectral cohesion function and the modularity function. (b) Average percent of nodes common between communities obtained by maximizing Λ and Q. (c) Average percent nodes of the labelled partition identified by Λ and Q. Dark (black) colored curves are for the spectral cohesion function while the light (green) colored curves are for the modularity function. Data points represent averages over 20 simulated networks.

when Q is maximized (plotted as solid squares). Figures 3.4 (b) and 3.5 (b) show the percent of nodes that are common between the Λ-based and the Q-based community partitions for the three network parameter sets. Figures 3.4 (c) and 3.5 (c) show the extent to which the Λ-based and the Q-based community divisions correspond to the "labelled partition". By the labelled partition, we mean the partition with two equally sized communities into which we divide the nodes when we generate random networks.

Referring to Figs. 3.4 (a) and 3.5 (a), we take the point of view that, essentially by definition, the Λ-based divisions give the best functional communities. It is notable from these plots that, although Q-based divisions give lower than optimal Λ, the Q-division results for Λ are surprisingly close to optimal throughout the whole range of $\langle d \rangle_X$ plotted. In contrast, Figs. 3.4 (b) and 3.5 (b) show that the percent agreement on nodal divisions between the Λ-based and the Q-based divisions can become substantial at large values of $\langle d \rangle_X$, especially for the networks with η_c maximized and with $(\rho_c, \eta_c) \approx 1$, while agreement is significantly better for networks with ρ_c maximized.

Regarding the difference between Figs. 3.4 (a) and (c) and between Figs. 3.5 (a) and (c), we expect both community finding methods to yield imperfect identification of the labelled partition. For example, this could result because it could happen that, in the random realization of a given network, some nodes with low within-community degrees in the labelled partitions may end up having many links with nodes in the other community or may get linked to high degree nodes in the other community. In the test networks, such nodes would reasonably be

classified as belonging to the community to which they were not originally assigned in the labelled partition.

3.5.2 Networks with biased links between communities

Here, we consider directed networks with two communities of equal sizes. We construct these networks so that, when the directionality of links is neglected, we get undirected random networks without communities. To do this, we start with 32 nodes that are divided into two groups of equal sizes, where each group represent a community. We then create, say, y number of undirected links between the two groups of nodes and $y/2$ randomly oriented directed links within each group. All the undirected links between the two groups are made directed with a bias such that more links point from one group of nodes to the other than the other way around. Thus, when we have x directed links pointing from one group to the other, $y - x$ directed links point in the opposite direction. Varying x gives us networks with a varying degree of community strength. The results for these networks corresponding to $N = 32$ and $N = 64$ are shown in Fig.3.6. At low values of x, when we have more bias, the spectral cohesion does better than modularity. At relatively higher values of x, both functions give similar results. We find that as we increase the number of links in the networks, by increasing the value of y, both the methods show improvement.

Comparing Fig. 3.6 (a) (N=32) and Fig. 3.6 (b) (N=64), we see that increasing the size of the networks keeping the average degree constant, the relative

advantage of the Λ-based partitions as compared to the Q-based partitions increases substantially.

Thus, we see that when functional communities are very strongly dependent on link directionality, modularity may substantially under-perform compared to the spectral cohesion method. We note, however, that so far we have not been successful at finding examples of real networks with this property.

3.5.3 Discovering communities in real world networks

To test how well our method finds communities in real networks, we used the networks of political blogs [31] and jazz bands [37]. The political blogs network is a directed network of weblogs on US politics during the 2004 US presidential elections. The edges are the hyperlinks connecting two blogs. The data for the network of jazz bands was obtained from The Red Hot Jazz Archive digital database. The network consists of bands that performed between 1912 and 1940. In this network, two bands are connected if there is a musician that played in both the bands.

The political blogs network has 1224 nodes with $\langle d \rangle = 15.6$. The eigenvalue plot of the adjacency matrix of this network shows two positive real eigenvalues well separated from the cloud of the rest of the eigenvalues (Chap. 2). This implies that there are two well defined communities in this network. The communities apparently correspond to left/liberals and right/conservatives. We used our simulated annealing procedure to divide the network into two communities by maximizing Λ and Q. Results are shown in Table 3.1, which also gives the percent of nodes common

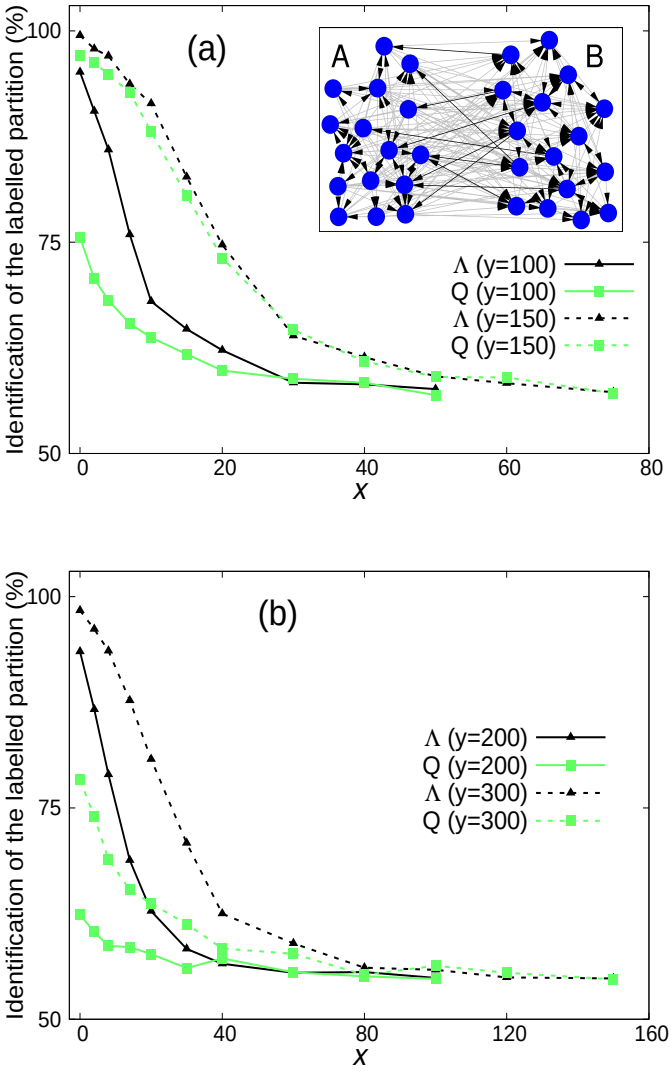


Figure 3.6: Percent of the labelled partition identified by Λ and Q maximization vs 'x' for the computer generated directed networks as explained in the text. (a) Networks with 32 nodes. (b) Networks with 64 nodes. All data points are averaged over 100 network realizations. The inset in figure (a) shows a particular network realization with $N = 32$, $y = 100$ and $x = 10$. For the sake of clarity, the 10 directed links that point from community B to community A are given darker shade. Note that, for the 32 node networks with $y = 100$ and $x = 0$ and the 64 node networks with $y = 200$ and $x = 0$, our Λ-based method does not give 100% identification of the labelled partition. This is because there are a few nodes in our random network realizations that are not part of the giant strongly connected component of either of the communities. For these cases, as we increase the density of links in the networks ($y = 150$ for the 32 node networks and $y = 300$ for the 64 node networks), the probability of such nodes becomes much smaller and the identification rate for the Λ-based method at $x = 0$ becomes close to 100%.

for the spectral cohesion method and the modularity method. The values of Λ for the spectral cohesion method and the modularity method are very close. We believe that this is due to the fact that the two communities have giant strongly connected components that are well separated from each other. Of the nodes that belonged to the giant strongly connected components of the network, there were 97.5 percent nodes common between the two community finding methods. In this network, there were 431 nodes that did not belong to any community's strongly connected component. Thus, in the spectral cohesion method, they were assigned based on the number of links such nodes had to the nodes in the giant strongly connected components of the communities.

The network of jazz bands is an undirected network with 198 nodes and $\langle d \rangle = 27.7$. The eigenvalue plot of this network shows three positive eigenvalues that are well-separated from the bulk of the other eigenvalues, thus indicating three strong communities (Chap. 2). Two strong communities in this network correspond to predominantly the white bands and the black bands, which shows racial segregation. The community of black bands divides further into two groups, the bands that performed in two major US cities, Chicago and New York [37]. Figure 3.7 shows the comparison between the partitions obtained by maximizing Λ and Q. We see that for this network, both the methods yield nearly the same network divisions (also see Table 3.1).

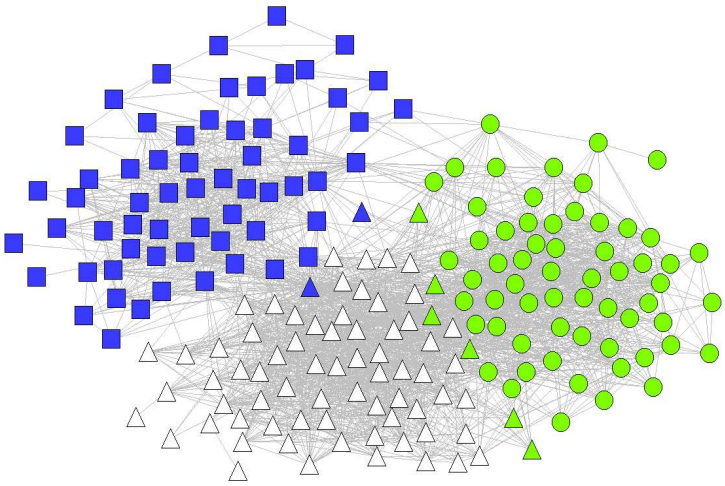


Figure 3.7: Comparison between the spectral cohesion method and the modularity method for the jazz bands networks. Different shapes of the nodes correspond to communities obtained by maximizing Λ, while different colors correspond to communities obtained by maximizing Q.

	Optimizing Λ		Optimizing Q		
Network	Λ	Q	Λ	Q	% common nodes
Political blogs	6.817	0.416	6.816	0.431	94.7
Jazz bands	10.095	0.441	10.084	0.444	96.0

Table 3.1: Function values and percent nodes common for the real networks considered in this section.

3.6 Conclusions

In this chapter, we explored the utility of functional rather than structural definitions of community structure. Specifically, as an example, we considered a definition of communities appropriate to cases where the communities are thought to form to enhance synchronizability and/or robustness to random node failures. Our method is based upon our introduction of the spectral cohesion function Λ (Eq. (3.5)) and is motivated by the role played by the maximum eigenvalue of the adjacency matrix, λ_*, in network functions.

Our study finds, perhaps, the unexpected result that for partitions obtained by maximizing modularity, the spectral cohesion, Λ, values are often close to optimal (Figs. 3.4 (a) and 3.5 (a)) even when the modularity maximized partitions were substantially different from the Λ-maximized partitions (Fig. 3.4 (b) and Fig. 3.5 (b)). Although our eigenvalue based method is computationally intensive, our analysis shows that [except when communities are strongly dependent on link directionality (Sec. 3.5.2)] communities obtained using the modularity-based method often do quite well when evaluated by a functional measure.

Chapter 4

Epidemic Spreading in Dynamic Social Networks

4.1 Introduction

The spread of communicable diseases in human population is a problem of major concern. Significant effort and resources are being devoted to understanding the factors that govern the spread of infections. Modeling is widely employed for this purpose. Using mathematical models, we can estimate quantities such as: the number of individuals getting infected and thus requiring treatment; the maximum number of people needing care at any given time; and the effectiveness of quarantine, vaccination or other control measures.

As a tool in understanding disease spreading phenomena, network based approaches can be used to model a wide variety of infection processes. Networks provide the underlying topology on which epidemics spreads. In a network representation, individuals are nodes connected by links that represent possible contacts through which an infection can propagate from an infected individual to a susceptible individual.

The importance of the contact pattern of individuals on epidemic spreading has long been acknowledged by researchers [57–59] (see Ref. [60] for a review of use of networks in epidemiology). Apart from the problems of defining and documenting contacts that can lead to transfer of infection, and the construction of epidemio-

logically significant networks from the collected data, it is believed that networks represent a more realistic interaction pattern in many cases than a random mixing model. In the absence of actual contact networks for the entire population, simulated and semi-simulated networks are generally used. The advantage of using networks in epidemic spreading is that we can take into account heterogeneity in human contacts. Using networks, the effects of other key aspects of connection patterns, such as correlations in contact patterns can also easily be considered. With the growth of the field of 'complex networks', there has been a surge in the number of papers that studied infection spreading on networks with non-trivial properties [61–68].

Typically, the underlying network in epidemic spreading models is considered to be static. One is then interested in the effect of network structure on disease propagation. While the static network assumption may work well for cases where the topology of the network does not change significantly during the course of spreading process, and it may even be a good approximation in certain cases where contacts are not fixed, in general, we cannot neglect the fact that real human contacts evolve over time. Depending on the disease and the type of contacts we are interested in, the time scales for the rate of change of human contacts and the spread of epidemic may vary. In recent years, there has been an increasing focus on disease spread on networks for which contacts change over time [69–72]. These networks for which the topology changes with time are referred to as *dynamic networks*.

In the epidemiological literature, dynamic models where contacts are constantly being formed and dissolved have been considered. Using models based on partnership formation [67, 73–76], one can study the effect of dynamic contacts due

to mixing. Although these models ignore most of the network features, the effects of concurrency of partnerships, degree based assortative or disassortative mixing, clustering, etc. can be included [67, 74]. These models can be understood as departure from random mixing models in which instantaneous contacts are assumed and the effects of local structures are absent. The above models do not take into account the contact networks behind partnership formation and assume that formation occurs at random. To overcome this 'drawback', a model that considers monogamous partnerships defined by the underlying network was considered by Eames *et al.* [77]. In an another model, Volz and Meyers [72] considered susceptible-infective-recovered (SIR) model on a full network in which each individual's contacts change in time. They provided a low-dimensional system of deterministic equations to predict disease transmission. In their model, referred to as the neighbor exchange model, the dynamics of partnership is implemented through edge swapping between two already connected pairs of nodes. Using a mixing parameter, their model interpolates smoothly between static network models and mass-action models.

Another class of dynamic network models that has been considered in the complex network community are the so called adaptive network models. In these models, a network's contact change in response to the disease. Gross *et al.* [69] studied susceptible-infective-susceptible (SIS) dynamics on adaptive networks where with a given constant probability, susceptibles break their links to infected neighbors and reconnect their broken links to randomly chosen susceptibles. With changing rewiring rates, their model shows interesting behavior where bistable and oscillatory states are possible. Zanette *et al.* [71] also considered SIS dynamics on adaptive

networks where the susceptible-infected links are broken with a constant probability. In different versions of their model, the broken links are either removed or the susceptible nodes reconnect their broken links to randomly chose nodes in the network. In the later version of their model, they also find regions of bistability and other dynamics not found in static network models. Shaw *et al.* [70] analyzed susceptible-infective-recovered-susceptible (SIRS) dynamics on adaptive networks. In their model, susceptibles, after breaking their links to infected neighbors, reconnect their broken links to randomly chosen non-infected nodes in the network. This model is also characterized by the presence of bistability of endemic and disease free states. In this case, the addition of recovered class and resusceptibilty rate allows control of the width of bistable region. They also found that fluctuations of the endemic state near the bifurcation point in SIRS models were significantly larger than in SIS model.

In our study of epidemic spreading, we consider a dynamic network model which has additional features compared to the simpler models discussed above. Although slightly more complicated, our model tries to capture some of the essential features found in the dynamics of real networks. In the model studied in this chapter, there are two types of processes that can lead to change in network topology, *social dynamics* and *evasion*. Social dynamics result in the constant evolution of contacts which is independent of the disease status of the nodes, while evasion causes susceptible nodes to avoid infected nodes. We study discrete time SIR model on such a dynamically changing network where at any given time, the nodes could be in one of the three possible states, namely, susceptible (S), infectious (I) and removed

(R). Here, R represents the state of the node when it is considered removed from the network and no longer takes part in any kind of dynamics. We note that this is different from the case in which R refers to recovered nodes who still maintain connections but can no longer become infected. In our model, we consider the case of *imperfect evasion* where the susceptibles can not avoid infecteds completely when they form new connections. This is in contrast to other models [69–71] where, after the susceptible-infected links are broken, the broken links are either removed or susceptible nodes reconnect the broken links to non-infected nodes in the network.

We also study the case of hidden infection on the dynamic network model discussed above. Hidden infection occurs if there is a period after an individual contracts an infection during which he or she is infectious but asymptomatic. Thus, susceptibles have no way of avoiding infected individuals in such a state. In medical terminology, *occult infection* refers to an infection which presents no clinical signs or symptoms. In our model, we are interested in the case where the infected individual while being asymptomatic, is infective. In dynamic network models, hidden infection effects dynamics on and of the network when we have evasion. The effectiveness of disease intervention measures could also get degraded when they are based on models that do not incorporate hidden infection for diseases with such characteristics.

In the next section, we describe our dynamic network model that takes into account change in network topology due to both social dynamics and evasion. We also describe the model that incorporates hidden infection. In Section 4.3, we give a mean-field analysis for the model without hidden infection to determine the value of transmission probability above which the disease infects a finite fraction of the

population in the limit of infinite system size. In Section 4.4, we give results from computer simulation of our models. Finally, in Section 4.5, we discuss the implications of our findings.

4.2 Description of the models

In our analysis, we imagine all individuals begin in the susceptible state. We then introduce the disease in the network by randomly choosing an individual to infect. Our goal is to characterize the dynamics which follow after the disease is introduced in this manner. Below, we give description of the processes involved in our dynamic models. The set of parameters used in our models are summarized in Table 4.1.

4.2.1 Model A: A dynamic network model with both social dynamics and evasion

As stated above, we consider the acyclic process $S \to I \to R$, where R represent the state of the node where it is considered removed from the network. In our simulations, the dynamics occur in discrete time steps. At any given time step, the total number of nodes in various states is fixed, *i.e.*, $N_S(t) + N_I(t) + N_R(t) = N$, where N is the number of nodes in the network we started with, $N_S(t)$ is the number of susceptible nodes, $N_I(t)$ is the number of infectious nodes and $N_R(t)$ is the total number of removed nodes in the network at time step t. The total number of *active nodes* in the population is given by $N_S(t) + N_I(t) \leq N$.

In our model, there are three dynamical processes that are occurring on the network: the evasion dynamics, the social dynamics, and the disease dynamics. At the beginning of each time step, evasion processes and social dynamics occur together, followed by the disease dynamics. The evasion process and the social dynamics constitute the topological dynamics of the network in which nodes delete their old links and form new connections. Due to the evasion of infected individuals by susceptibles, each existing $S-I$ link in the network is deleted with a constant probability, η_1. Due to social dynamics, we assume that all the nodes delete their links independent of their disease status. To account for this, each existing link in the network is deleted with a constant probability, ϕ. Since in our model, both evasion process and social dynamics happen together, each $S-S$ and $I-I$ link is deleted with probability ϕ while each $S-I$ link is deleted with probability $\alpha = \eta_1 + (1-\eta_1)\phi$.

All the nodes have links deleted via the above process search for new partners. This also include nodes from previous time steps that have deleted links due to other reasons discussed below. In our model, individuals attempt to maintain the number of connections specified by their desired degrees (which never changes). Thus the number of partners that an individual seeks is equal to the number of its deleted links (henceforth referred to as *unmatched links*). While searching for new partners, we assume that the unmatched links mix randomly and any unmatched link has an equal chance of meeting any other unmatched link, independent of the disease status of the nodes they belong to. However, we differentiate between unmatched links *meeting* each other and actually forming a new link. When unmatched links of nodes with same disease status meet (*i.e.*, a potential $S-S$ or $I-I$ link), they

form a new link with probability 1. When unmatched links of nodes with different disease status meet (*i.e.*, a potential $S-I$ link), due to the evasion of infected individuals by the susceptible individuals, the link is formed with probability $1-\eta_2$, where η_2 is a second evasion parameter. This is the case of imperfect evasion where the susceptible nodes avoid getting infected by deleting their links with the infected neighbors but while searching for new partners, the susceptible nodes again have the chance of forming links with infected nodes. When the $S-I$ pair of unmatched links is rejected (with probability η_2), the unmatched links of the rejected pair move to the next time step to search for partners.

To keep the dynamics simple, in our simulations we allow the nodes to form self loops and multiple edges. We interpret self loops as reduced interaction of the nodes with other nodes in the population while multiple edges could reflect increased interaction between two nodes.

In each time step, the topological dynamics described above is followed by the disease dynamics. Disease spreads from an infected node to a susceptible node along each existing $S-I$ link with transmission probability λ. Thus, the probability that the susceptible node i with $d_{i,I}$ infected neighbors at a particular time step become infected in that time step is $1-(1-\lambda)^{d_{i,I}}$. Once a susceptible node gets infected, while being infective, it remains in infected state for a fixed period of time, say τ, after which the node moves to the removed class. The nodes in the removed class never change their disease status nor do they take part in topological dynamics. When a node moves to the removed class, the active nodes in the network (nodes with disease status S and I) delete their links with the removed node. These active

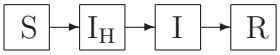
λ	Probability of disease transmission along each $S-I$ link.
ϕ	Social dynamics parameter. Probability with which, independent of the disease status of the nodes, each existing link in the network is deleted.
η_1	Evasion parameter 1. Probability with which an already existing $S-I$ link in the network is deleted.
η_2	Evasion parameter 2. Probability with which a potential $S-I$ link is rejected.
τ_1	Hidden infection period. This is the number of time steps for which a node is infective but is **not** known to have the disease.
τ_2	The number of time steps for which a node is infective and is known to have the disease.
τ	Total period of infectivity. $\tau = \tau_1 + \tau_2$

Table 4.1: Parameters used in the models studied in this chapter. The probabilities are per unit time step.

nodes then have unmatched links that search for new partners in the next time step.

4.2.2 Model B: A dynamic network model with social dynamics, evasion, and hidden infection

Here we extend the model described above to allow for hidden infection. When a susceptible node gets infected, prior to moving to state I, it moves to state I_H. In state I_H an infected individual does not show symptoms of the infection and thus is not known to have the disease. Although nodes in this class are asymptomatic, they are infectious and can infect the susceptible nodes they are connected to. Thus, the risk of infection spreading in the population increases because the susceptibles can not avoid nodes with hidden infection. For this model, the flow of individuals among various disease classes can be depicted as follows:



In the hidden infection model, after spending a fixed amount of time, τ_1, in state I_H, the infected nodes move to state I where they show symptoms of the disease and are also infective. For this model, the amount of time the nodes spend in state I is represented by τ_2. The total period of infectivity, τ, is thus equal to $\tau_1 + \tau_2$. In our model, we assume that the nodes in states I and I_H infect their susceptible neighbors with the same probability λ.

For the purpose of topological dynamics, we treat nodes in states S and I_H equally. Thus, during the link deletion process, $S - S$, $I_H - I_H$, $S - I_H$ and $I - I$ links are deleted with probability ϕ, while $S - I$ and $I_H - I$ links are deleted with probability α. While forming new links, when a potential new link is either $S - S$ or $I_H - I_H$ or $S - I_H$ or $I - I$, the new link is formed with probability 1. When the potential new link is either $S - I$ or $I_H - I$, the new link is formed with probability $1 - \eta_2$.

4.3 Mean-field analysis for Model A

Before we present simulation results, we give a mean-field analysis explaining the behavior of the model described in Sec. 4.2.1. We are interested in finding the transmission probability $\hat{\lambda}$, such for $\lambda \geq \hat{\lambda}$, a significant portion of the population gets infected. More specifically, we want to find the minimum value $\hat{\lambda}$ such that a finite fraction of the population gets infected in the limit of infinite system size. We first present analysis for networks in which each node has degree d_0. We then discuss

how the results can be extended to networks with other degree distributions. We assume that networks are large and sufficiently sparse so that loops can be neglected. In our analysis, we focus on the spreading process during the early stages of the epidemic when the number of susceptible nodes is large.

Before analyzing our model, we first review the case of infinite random static networks with arbitrary degree distributions in cases where loops can be neglected [47]. Let $P(d)$ denote the probability that a randomly chosen node in the network has degree d. Let $Q(d)$ denote the probability density function for the degree of a node at the end of a randomly chosen edge. For random networks, the distribution $Q(d)$ is given by

$$Q(d) = d\ P(d)/\langle d \rangle, \tag{4.1}$$

where $\langle \cdots \rangle$ denotes the expectation over the distribution $P(d)$.

For the SIR model on static networks, assuming that we start with one infected node, we want to find the size of the infected component connected to that individual. Let z_n be the expected number of neighbors getting infected in n steps, we then have

$$z_{n+1} = \lambda z_n \sum_d (d-1)Q(d) = \lambda \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle} z_n. \tag{4.2}$$

The critical value λ_c above which the disease has non-zero probability of propagating across the system is given by $z_{n+1}/z_n = 1$. Thus, for the SIR model on static networks,

$$\lambda_c = \frac{\langle d \rangle}{\langle d^2 \rangle - \langle d \rangle}. \tag{4.3}$$

To analyze the dynamic network model proposed in this chapter, we modify

the above method. Since we start with a sparse network and new connections in our model are formed randomly, we assume that during the early stages of the epidemic, the dynamic network still has locally tree like structure. In the following analysis, we set the period of infectivity of nodes, τ, to be one time step.

We first consider the spreading process on a dynamic large sparse network where each node has degree d_0. Similar to the case of static networks discussed above, to get an estimate of $\hat{\lambda}$ in the mean-field approximation (*i.e.*, neglecting fluctuations), we assume that at the transition point, the expected number of infected individuals at time step t is equal to the number of infected individuals at time step $t-1$. We expect this approximation to be true during the early stages of the epidemic when the number of infected individuals is small compared to the initial size of the population.

Let $M_{SI}(t)$ be the number of $S-I$ links in the network at the beginning of time step t. During the topological dynamics, each $S-I$ link is deleted with probability α. Thus, the number of $S-I$ links that get deleted will be binomially distributed with probability α. In the mean-field approximation, we assume that the expected number of $S-I$ links that get deleted is $\alpha M_{SI}(t)$. Let $I(t-1)$ denote the number of nodes that got infected in time step $t-1$. When the infected nodes from time step $t-2$ move to the removed class in time step $t-1$, the unmatched links of newly infected nodes in time step $t-1$ that were linked to infected nodes from time step $t-2$, search for partners in time step t. Since we assumed tree like structure, the number of such unmatched links is equal to $I(t-1)$. Thus, the number of unmatched links of infected nodes that will search for partners in time

step t is

$$L_I = \alpha M_{SI}(t) + I(t-1). \tag{4.4}$$

Similarly, the number of unmatched links of susceptible nodes is

$$L_S = \alpha M_{SI}(t) + 2\phi M_{SS}(t) + \eta_2 x(t-1) + y(t-1), \tag{4.5}$$

where $M_{SS}(t)$ is the number of $S-S$ links in the network at the beginning of time step t. In the third term, $x(t-1)$ is the number of $S-I$ pairs of unmatched links that meet while searching for partners in time step $t-1$. Out of these, a fraction η_2 of them get rejected. The unmatched links of susceptible nodes from those rejected pairs search for partners in time step t. Finally, the term $y(t-1)$ accounts for the fact that when infected nodes from time step $t-2$ move to the removed class in time step $t-1$, the susceptible nodes previously connected to them have unmatched links looking for partners in time step t. Here, the time dependence of variables L_I and L_S is implied.

When forming new connections, we assume that the unmatched links mix randomly. To keep the dynamics simple, in our simulations, the nodes that disconnected their links from each other in a particular time step are allowed to create new links in the same time step. We also allow for self-loops and multiple edges. Under these assumptions, when we have integer number of unmatched links seeking partners (as happens in computer simulations), say $\tilde{L_I}$ and $\tilde{L_S}$, the probability of $\tilde{x}$ number of susceptible-infected pairs meeting, when $\tilde{L_I} + \tilde{L_S}$ is even, is given by

$$\frac{2^{\tilde{x}}\ \tilde{L_I}!\ \tilde{L_S}!\ \frac{(\tilde{L_I}+\tilde{L_S})}{2}!}{\tilde{x}!\ \frac{\tilde{L_I}-\tilde{x}}{2}!\ \frac{\tilde{L_S}-\tilde{x}}{2}!\ (\tilde{L_I}+\tilde{L_S})!}. \tag{4.6}$$

One can show that over the above distribution, the expected value of $\tilde{x}$ is

$$\frac{\tilde{L_I}\tilde{L_S}}{\tilde{L_I}+\tilde{L_S}-1}. \tag{4.7}$$

Similarly, when $\tilde{L_I}+\tilde{L_S}$ is odd, one can show that the expected value of $\tilde{x}$ is

$$\frac{\tilde{L_I}\tilde{L_S}}{\tilde{L_I}+\tilde{L_S}}. \tag{4.8}$$

We assume that the number of unmatched links seeking partners is large and approximate the expected number of susceptible-infected unmatched links meeting together, $x(t)$, as

$$\frac{L_I L_S}{L_I+L_S}. \tag{4.9}$$

Of the $x(t)$ pairs of susceptible-infected unmatched links that meet, $(1-\eta_2)x(t)$ is the number of pairs that actually form link, while $\eta_2 x(t)$ pairs get rejected. Thus, the number of new $S-I$ links that get created is

$$(1-\eta_2)\frac{L_I L_S}{L_I+L_S}. \tag{4.10}$$

Since we assume that the number of infected nodes is small and that the networks are large, this means that $M_{SS}(t)$ is large. Thus, during the early stages of disease spreading, for $\phi > 0$, at the transition point, L_I and $\alpha M_{SI}(t)+\eta_2 x(t-1)+y(t-1)$ are much smaller than $\phi M_{SS}(t)$. Neglecting terms of higher order, we get the expected number of new $S-I$ connections in time step t as

$$(1-\eta_2)\Big[\alpha M_{SI}(t)+I(t-1)\Big]. \tag{4.11}$$

After the topological dynamics, the total number of $S-I$ links at time step t is

$$(1-\alpha)M_{SI}(t) + (1-\eta_2)\Big[\alpha M_{SI}(t) + I(t-1)\Big]. \tag{4.12}$$

For each of these susceptible-infected links, the probability of disease transfer is λ. Thus, the expected number of new infections is

$$I(t) \approx \lambda\Big[(1-\alpha\eta_2)M_{SI}(t) + (1-\eta_2)I(t-1)\Big]. \tag{4.13}$$

Assuming tree like structure of the network, we have, for large networks, $M_{SI}(t) = (d_0-1)I(t-1)$. At the transition point, we assume $I(t) \approx I(t-1)$ giving us

$$\hat{\lambda} \approx \frac{1}{(d_0-1)(1-\alpha\eta_2) + (1-\eta_2)}. \tag{4.14}$$

In terms of ϕ, η_1 and η_2, we have

$$\hat{\lambda} \approx \frac{1}{(d_0-1)(1-\eta_2\phi-\eta_1\eta_2+\eta_1\eta_2\phi) + (1-\eta_2)}. \tag{4.15}$$

We derived the above expression assuming that the degrees of nodes are constant. For large networks with arbitrary degree distribution, we can arrive at the same expression as above but with d_0-1 in the denominator replaced by $\frac{\langle d^2\rangle - \langle d\rangle}{\langle d\rangle}$. We expect this assumption to work for cases when the degree distribution of the network is not highly skewed, because we assumed that the degree distribution of the network does not change during the early stages of the spreading process. In the case of highly skewed degree distributions, we expect the high degree nodes in the network to get infected early on [78], thus changing the degree distribution of the active nodes in the network.

In Eq. (4.15) ($\tau = 1$ and ϕM_{SS} large), we consider the following special cases:

- $\eta_1 = 0$

$$\hat{\lambda} = \frac{1}{(d_0 - 1)(1 - \eta_2\phi) + (1 - \eta_2)} \tag{4.16}$$

In the absence of evasion by neighbors, $\hat{\lambda}$ increases with increasing ϕ when $\eta_2 > 0$.

- $\eta_2 = 0$

$$\hat{\lambda} = \frac{1}{d_0} \tag{4.17}$$

In the absence of evasion of infected individuals by the susceptible individuals during pair formation, the transition point is independent of ϕ and η_1. Note that this expression differs slightly from Eq. (4.3) because active nodes in the network reconnect their links to removed nodes to form new connections.

4.4 Simulation results

We simulated the models described in Section 4.2 on various types of networks with $N = 10^5$ and $\langle d \rangle$=20. At the beginning of each time step, we start by infecting a randomly chosen node in the population and study the evolution of the system's dynamics. In the results given below, we are primarily concerned with the location of epidemic transition with varying system parameters. The size of the epidemic is measured by the final number of R individuals in the population after the disease has died out.

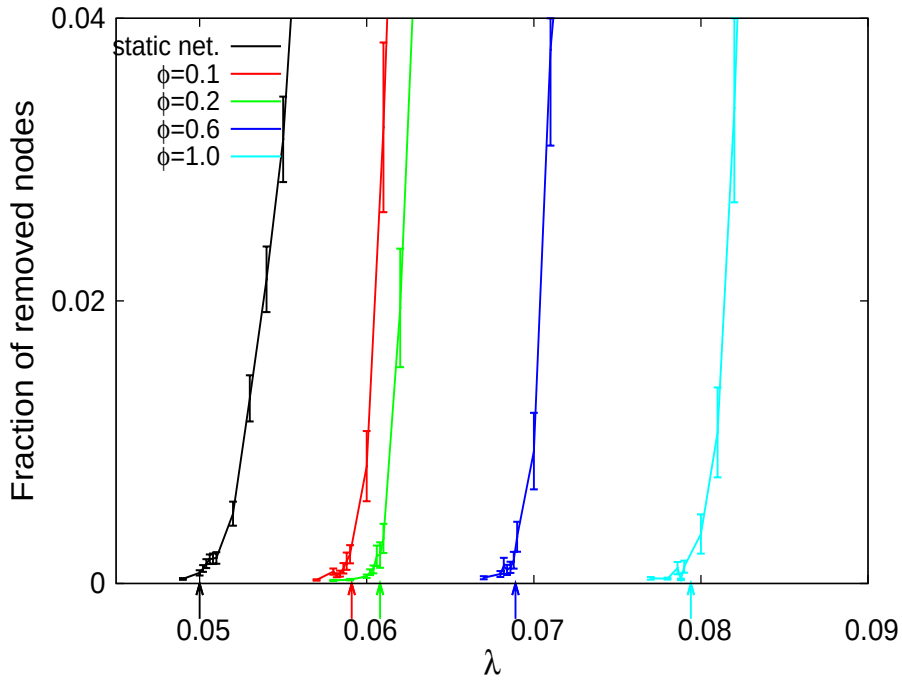


Figure 4.1: Average fraction of removed nodes of the network versus the transmission probability λ with **varying social dynamics parameter** (ϕ) for an **Erdős-Rényi network** with $N = 10^5$ and $\langle d \rangle = 20$. The values of other parameters are $\eta_1 = 0.4$, $\eta_2 = 0.4$ and $\tau = 1$. Each data point is an average over 500 simulation runs for one network realization.

4.4.1 Effect of varying social dynamics and evasion (Model A)

For the results presented in this section we set period of infectivity, τ, to be one time step. Figure 4.1 shows the average fraction of removed nodes (after the disease has died out) versus transmission probability, λ, for an Erdős-Rényi network. Curves for various values of the social dynamics parameter (ϕ) have been plotted with fixed values of η_1 and η_2 (both set to 0.4). For the sake of comparison, results from simulation on a static network are also shown. For the dynamic network model, the arrows on the horizontal axis in Fig. 4.1 show the prediction for the transition point ($\hat{\lambda}$) using Eq. (4.15) but after taking into account the effect of degree distribution. For the static network, the prediction is given by Eq. (4.3). The figure shows that

our simplified mean-field prediction for the onset of epidemics in dynamic networks agrees very well with the simulation results.

Figure 4.1 shows that when the social dynamics parameter ϕ is increased the value of transmission probability at which epidemics first occur also increases. Alternately, note that for a given value of transmission probability λ, as ϕ increases, fewer individuals in the population get infected. This result is contradictory to the intuition that with increased mixing, the disease has a better chance of spreading. In our model, we get this counter-intuitive result because of the presence of the second evasion parameter, η_2. In the model, when ϕM_{SS} is large and $\eta_1 = \eta_2 = 0$, mixing due to social dynamics has no effect on epidemic spreading because $\tau = 1$. For a given value of the evasion parameter $\eta_1 < 1$, as ϕ increases, a larger number of $S - I$ links get deleted due to social dynamics but as the unmatched links of those infected nodes seek new partners, they get rejected with probability η_2 (when η_2 is non-zero), leading to a reduction in the number of $S - I$ links through which the disease can spread. Thus, for $\tau = 1$, when η_2 is non-zero, increasing the social dynamics parameter ϕ actually serves to inhibit the spread of the disease through the population.

Next we examine the case in which ϕ is small (thus we have a small value of ϕM_{SS}) (Fig. 4.2). For this case, we conduct simulations starting with $\phi = 0$ and increasing ϕ. As we increase ϕ, the value of $\hat{\lambda}$ first decreases then increases. To understand this, consider the case when $\phi = 0$. In that case, the number of unmatched links of S and I that seek for partners are of same order. Thus, the unmatched links of I have a good chance of forming $I - I$ links leading to a reduced

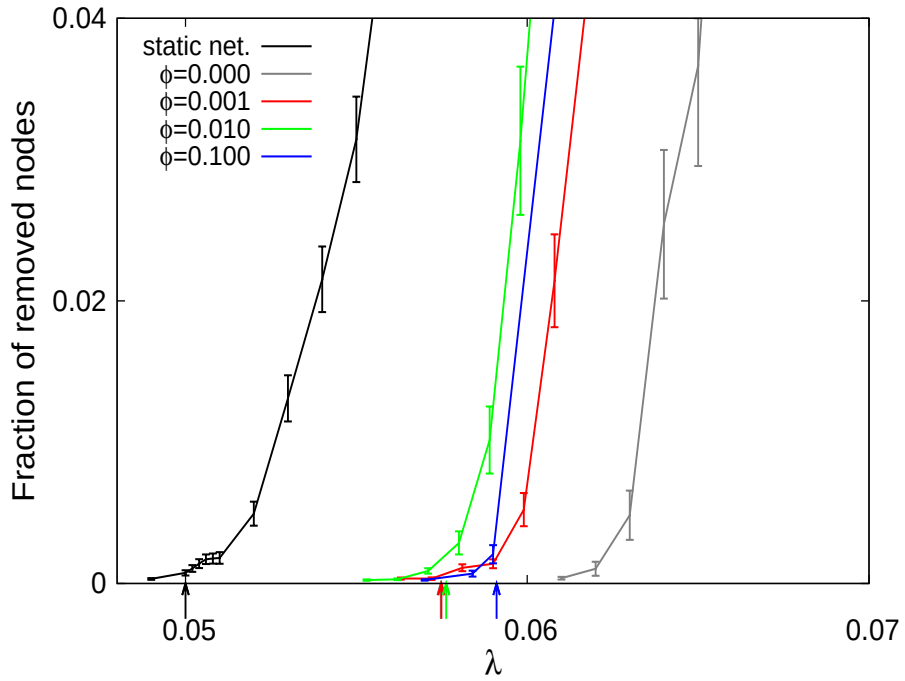


Figure 4.2: Average fraction of removed nodes of the network versus the transmission probability λ with **varying social dynamics parameter** (ϕ) for an **Erdős-Rényi network** with $N = 10^5$ and $\langle d \rangle = 20$. The values of other parameters are $\eta_1 = 0.4$, $\eta_2 = 0.4$ and $\tau = 1$. This figure shows results for low values of ϕ. Each data point is an average over 500 simulation runs for one network realization.

number of $S - I$ links. Further, as ϕ is increased from zero, more unmatched links of S nodes are available to connect with unmatched links of I nodes to form disease transmitting pairs, thus leading to a decrease of the transition value. As ϕ increases further and ϕM_{SS} becomes large, an increase in $\hat{\lambda}$ occurs due to the reason explained in the preceding paragraph. We also note that, as expected, our prediction does not work well for the case when ϕM_{SS} is small, although it works very well for other values of ϕ.

Having explored the role of the social dynamics parameter, ϕ, we now study the effect of varying evasion parameters (η_1 and η_2) with fixed $\phi(= 0.1)$ (Fig. 4.3). The red, green and dark blue curves show the results for a fixed value of $\eta_1(= 0.4)$.

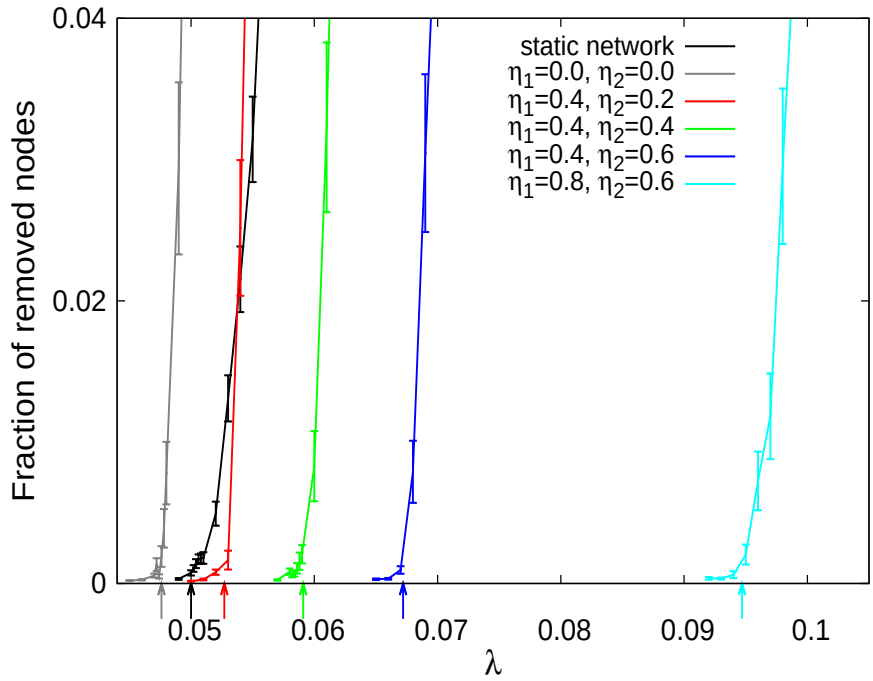


Figure 4.3: Average fraction of removed nodes of the network versus the transmission probability λ for **fixed social dynamics parameter** ($\phi = 0.1$) for an **Erdős-Rényi network** with $N = 10^5$ and $\langle d \rangle = 20$. The period of infectivity, τ, is equal to 1. Red, green and dark blue curves are for $\eta_1 = 0.4$ while dark and light blue curves are for same values of $\eta_2 (= 0.6)$ but different η_1. Each data point is an average over 500 simulation runs for one network realization.

As expected, as η_2 increases, the transition point also increases and disease spread is inhibited. Dark and light blue curves show the effects of increasing η_1 for a fixed value of $\eta_2 (= 0.6)$. For the dynamic network model, the arrows on the horizontal axis correspond to the prediction from Eq. (4.15) after taking into account the effect of degree distribution, and show good agreement with the transition observed in the simulations. The grey colored curve in Fig. 4.3 is for the case $\eta_1 = \eta_2 = 0$ and shows that epidemic transition for our dynamic network model in this case is smaller than that for a static network. This is due to the reconnection of the unmatched links of I nodes, formerly connected to R nodes, to other active nodes in the network.

In Fig. 4.4 we show simulation results for a scale-free network with the degree

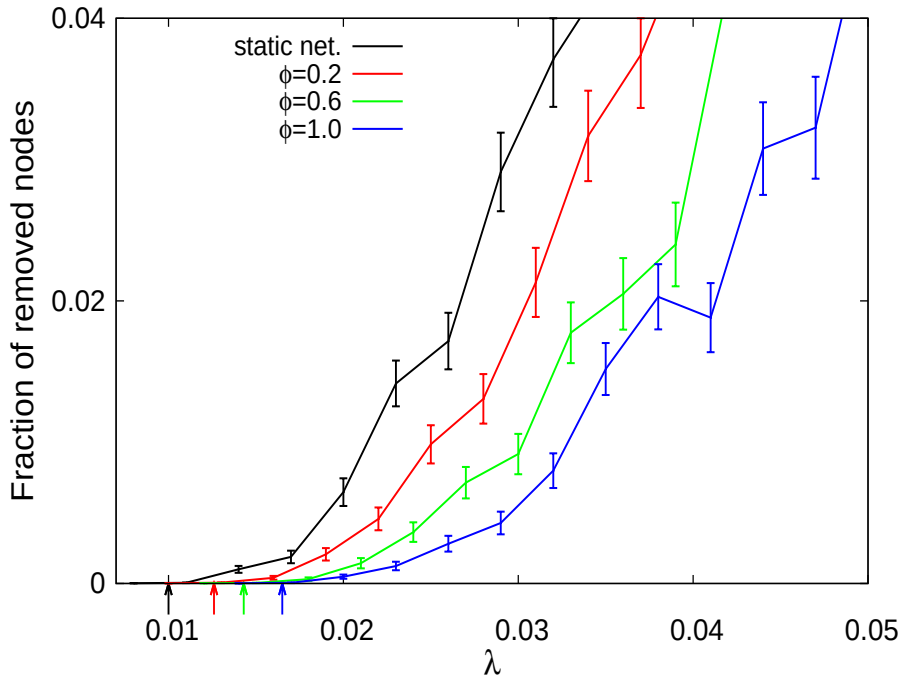


Figure 4.4: Average fraction of removed nodes of the network versus the transmission probability λ with **varying social dynamics parameter** (ϕ) for a **scale-free network** with $N = 10^5$, $\langle d \rangle = 20$, $\gamma = 2.5$ and maximum degree $= 1400$. The values of other parameters are $\eta_1 = 0.4$, $\eta_2 = 0.4$ and $\tau = 1$. Each data point is an average over 500 simulation runs for one network realization.

exponent, γ, equal to 2.5, $N = 10^5$, $\langle d \rangle = 20$ and maximum degree $= 1400$. For the scale-free networks, as is well known for the case of static networks, the transition point is lower than that in Erdős-Rényi networks with same average degree.

To demonstrate the range of applicability of our mean-field predictions, in Fig. 4.5 we plot $\hat{\lambda}$ versus ϕ for both simulations and predictions (Eq. (4.15)). In the system, we define the transition point as the minimum value of λ for which a macroscopic number of nodes get infected (*i.e.*, the epidemic size is a finite fraction of N as $N \to \infty$). We used this to define $\hat{\lambda}$ in simulations as the value of λ for which the average fraction of R nodes, after the disease has died out, become 0.001. This value of the threshold is taken in order to calibrate the results for static networks

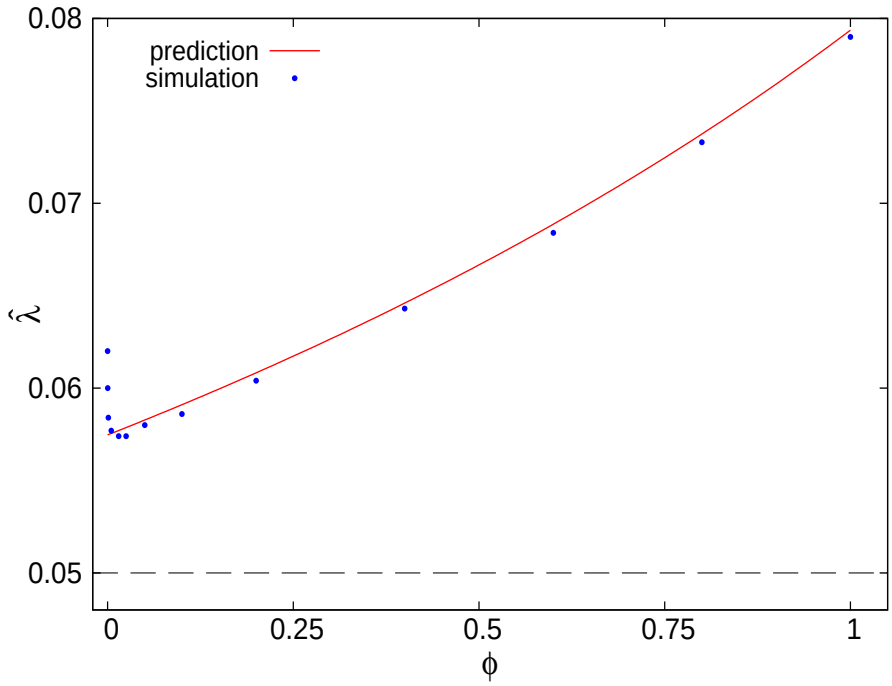


Figure 4.5: $\hat{\lambda}$ versus social dynamics parameter (ϕ) for an **Erdős-Rényi network** with $N = 10^5$ and $\langle d \rangle = 20$. The values of other parameters are $\eta_1 = 0.4$, $\eta_2 = 0.4$ and $\tau = 1$. The dashed line at the bottom shows the prediction for the epidemic transition for static networks.

to the value given by the prediction in Eq. (4.3). The figure shows that predictions for the dynamic network model agree very well with simulations as long as ϕ is not very small.

4.4.2 The case of hidden infection (Model B)

Here we give results for the case of an epidemic with hidden infection on a constant degree network. Newly infected nodes in this case do not show any sign of disease and thus are not identifiable. This means that, the newly infected nodes can not be alienated by susceptible nodes. Thus, for a given value of the transmission probability, we expect the epidemic to infect a higher number of nodes compared to the case when we do not have hidden infection. Fig. 4.6 shows the results with

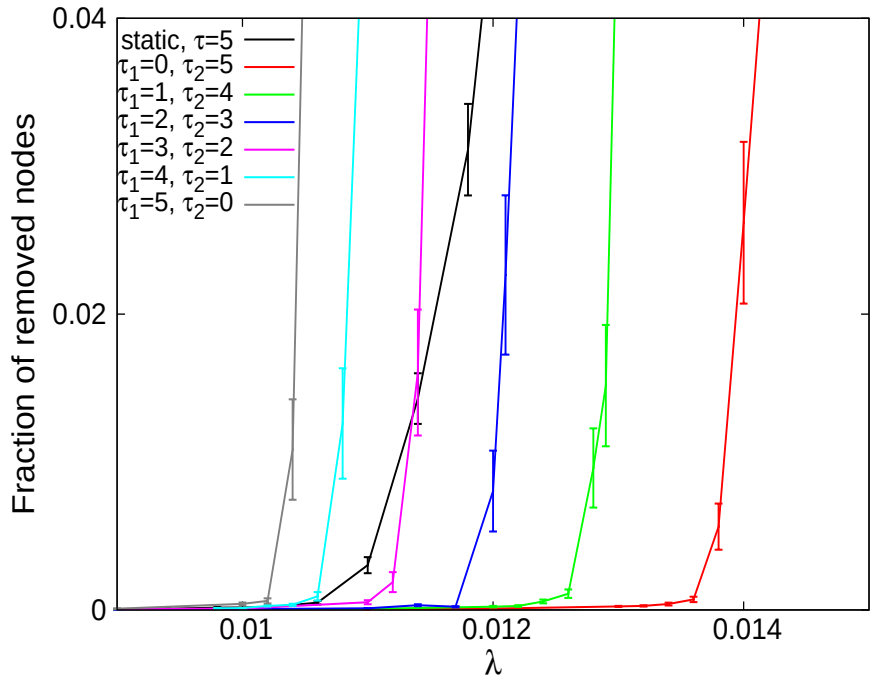


Figure 4.6: The case of hidden infection. Average fraction of removed nodes of the network versus the transmission probability λ with **varying hidden infectivity period** (τ_1) for fixed total infectivity, $\tau = 5$, for a **constant degree network** with $N = 10^5$ and $d_0 = 20$. The values of other parameters are $\phi = 0.2$, $\eta_1 = 0.4$ and $\eta_2 = 0.4$. Each data point is an average over 500 simulation runs for one network realization.

varying hidden period of infectivity (τ_1) but fixed total period of infectivity (τ) which was set to five time steps. We fixed other parameters of the dynamic network model as: $\phi = 0.2$, $\eta_1 = 0.4$ and $\eta_2 = 0.4$. As expected, with increasing τ_1, the transition point shifts towards lower values of λ. For the chosen parameters, when $\tau_1 = 0$, the transition point is larger than the case for a static network but for $\tau_1 = 5$ the transition point in dynamic network is smaller than the static network.

4.5 Discussion and conclusions

Incorporating the effects of changing network topology due to disease independent mixing (social dynamics) as well as disease evasion makes the modeling more complicated but better able to capture important features of disease spread across social networks. In our dynamic network model with both social dynamics and evasion (Model A), we find that the value of transmission probability at which the epidemics first occurs increases with increased mixing. In other words, as social mixing increases, disease spread is inhibited. This occurs due to the avoidance of the infected individuals by the susceptible individuals when new pairs are formed ($\eta_2 > 0$). In the model studied, for $\tau = 1$, disease independent mixing by itself has no effect on the spread of disease, the social dynamics only become important in the presence of evasion dynamics.

We also studied a model incorporating another class of infective individuals, I_H, who are infectious but show no signs of infection (Model B). In static networks, class I_H has no effect on epidemic spreading. In the dynamic network models, the addition of the I_H class has a large effect in the presence of evasion dynamics. As shown in the results, if network connections evolve on the timescale of disease transmission, it becomes necessary to include this new class of infectives when studying a disease with such characteristics.

Appendix A

Generation of Networks with the Eigenvalue Based Communities

A.1 Scale-free networks with nodal in/out-degree correlation within communities

To generate these networks, we start by dividing the nodes into two equally sized communities (labelled by $k = 1, 2$). For nodes in community k, we then generate two degree sequences corresponding to the in-degrees (d_i^{in}) and the out-degrees (d_i^{out}), according to the power law degree distribution, $P(d) \propto d^{-\gamma}$, by using the formula Eq. (2.14). For the test networks used in Chap. 3, we used $\gamma = 2.5$. These N_k numbers, where N_k is the number of nodes in community k, corresponding to both the degree sequences are assigned randomly to the nodes in community k. Since these numbers are assigned independently at random, d_i^{in} and d_i^{out} of the nodes are uncorrelated. We perform this procedure for each of the communities, separately.

For each of the d_i^{in} for node i, we then randomly divide d_i^{in} into two compartments,

$$d_i^{in} = (d_i^{in})_I + (d_i^{in})_X, \tag{A.1}$$

choosing $0 \leq (d_i^{in})_I \leq d_i^{in}$ from a binomial distribution. Here the subscripts I and X signify internal and external, and $(d_i^{in})_I$ signifies the number of links going to node i from nodes in its own community, while $(d_i^{in})_X$ signifies the number of links

going to node i from nodes that are not in its community. In addition, we perform a similar decomposition for d_i^{out},

$$d_i^{out} = (d_i^{out})_I + (d_i^{out})_X. \tag{A.2}$$

In using the binomial distribution in Eq. (A.1) (or Eq. (A.2)), we assume that each of the d_i^{in} links (or d_i^{out} links) has probability $\langle d \rangle_I / \langle d \rangle$ of being internal. We now have associated to each node i the four degrees

$$[(d_i^{in})_I, (d_i^{in})_X, (d_i^{out})_I, (d_i^{out})_X]. \tag{A.3}$$

To create a network with maximal η_c, we now shuffle the node assignments of $[(d_i^{in})_I, (d_i^{in})_X]$ within each community, while keeping the node assignments of $[(d_i^{out})_I$, $(d_i^{out})_X]$ fixed. In particular, if $(d_j^{out})_I$ is the largest internal out-degree of community k, we reassign $[(d_i^{in})_I, (d_i^{in})_X]$ from node i to node j where node i is also the node in community k with the largest value of $(d_i^{in})_I$. We then do the same for the second largest, for the third largest, etc. This leads to new assignments of the four degree quantities (Eq. (A.3)) for each node but only by shuffling degrees of nodes that belong to the same community. Note than, by construction, our reassignment procedure leaves the distribution of the d_i^{in} and d_i^{out} invariant, and that interchanging the roles of "in" and "out" (*i.e.*, preserving the "in" node assignments and shuffling the "out" node assignments) results in an equivalent procedure.

We now construct within community links for each community k. We imagine drawing $(d_i^{in})_I$ in-stubs and $(d_i^{out})_I$ out-stubs at each node i. We then randomly pair the end of an in-stub to the end of an out-stub and connect them with a link.

This is done avoiding repeated links and self-links. Finally, we use the analogous procedure to construct external connections between communities.

When we generate scale-free networks, nodes with high within community in-degrees (out-degrees) will tend to have high between community in-degree (out-degree). Similarly, nodes with low within community degrees will tend to have low between community degrees. We do this with the belief that important nodes that have high number of links within their own community will, in general, have high number of links attached to nodes outside their own community. Due to this, when we generate scale-free networks with maximal node degree correlations within the communities, they also tend to have higher values of node degree correlations for the between community degrees, although for each node, we expect the between community in and out-degrees to be less correlated than the within community in and out-degrees.

When we generate scale-free networks with the maximal node degree correlations within the communities, we find that the values of the directional degree assortativity within the communities become slightly less than 1. Since we are interested in looking at the effect of changing nodal degree correlations within the communities, we use the edge swapping procedure described in Sec. 3.4.2.2 to restore ρ_c to $\rho_c \approx 1$ within the communities.

A.2 Scale-free networks with directional degree assortativity within communities

We construct a directed random node degree uncorrelated scale-free network by using the method given in Appendix A.1 (*i.e.*, without shuffling the degrees of the nodes). We then use the edge swapping procedure given in Sec. 3.4.2.2 to get maximal possible ρ_c within each community by considering only within community degrees, $(d_i^{in})_I$ and $(d_i^{out})_I$.

A.3 Erdős-Rényi type networks with nodal in/out-degree correlation within communities

To get these networks, we first divide the nodes into two equally sized communities and create undirected edges within communities with probability, say p. Considering undirected edges as bidirected links, we randomly reassign links between nodes keeping their degrees constant. This gives us communities in which the in-degree of a node equals the out-degree of the node but the edge degree correlations are absent. Between community directed links are created by creating directed links between pairs of nodes in different communities with some other chosen probability, say q.

A.4 Erdős-Rényi type networks with directional degree assortativity within communities

To generate these networks, we divide the nodes in the network into two equally sized communities. Within communities, we create directed links with probability p while between community directed links are created with some other probability q. We then use the procedure of Sec. 3.4.2.2 to get maximal ρ_c within the communities.

Appendix B

Estimating the maximum and minimum attainable degree correlations in large directed networks

B.1 Introduction

While studying eigenvalue based communities (Section 3.4.2), we were interested in cases where adjacency matrices of communities have high values of their largest eigenvalues (λ_{*k}). Keeping the average degree and the degree distribution inside the communities constant, and restricting ourselves to only those nodes and links that lie within a community, we can increase the largest eigenvalues of the communities by increasing the directional degree assortativity (ρ_k) and node degree correlation (η_k) within communities. In computer simulations, we found that while for communities with power law degree distribution of both in-degrees and out-degrees, the maximum attainable values of ρ_k and η_k were large, for the case of communities that had Poisson degree distribution of both in-degrees and out-degrees, the maximum attainable values of ρ_k and η_k were rather small. We can guess that behavior by considering the nature of the distributions. But, given a network, it can be useful to estimate the maximum and minimum attainable values of degree correlations in the network. In this appendix, we give mean-field formulas for large directed networks for estimating the maximum and minimum attainable

directional degree assortativity, ρ, and node degree correlation, η. We give results for directed networks with Poisson and power law degree distributions. For the scale-free networks, we assume that the degree exponent $\gamma > 2$ and that the degree distribution has a lower cutoff, d_{min}, while the upper degree is unbounded. Although we give results for specific network types, the method, in principle, can be applied to networks with arbitrary degree distributions.

For the sake of interest, we also give mean-field formulas for estimating the maximum and minimum attainable correlation between out-degree of node j and in-degree of node i when there is a link from node j to node i. To characterize this degree correlation (subsequently referred by symbol ν), we will again use the independence measure similar to that we used to define ρ (Eq. (3.13)). This degree correlation is closely related to assortativity by degree that is commonly used in the literature [34] that measures the correlation between excess out-degree of node j and excess in-degree of node i.

B.2 Formulas for directed networks

We consider directed networks that have N nodes and m edges. We assume that N is large and no substructures exist within networks so that mean-field description is applicable. For links from node j to node i, the entry A_{ij} of adjacency matrix of network is non-zero. The in-degree and out-degree of node i are denoted by d_i^{in} and d_i^{out}, respectively. When not referring to specific nodes in the network, we will use the notation $\tilde{d}^{out}$ and $\tilde{d}^{in}$ to refer to the degrees of the nodes that are

connected to each other. The in-degree and out-degree distributions of nodes in the network are denoted by $P_{in}(d^{in})$ and $P_{out}(d^{out})$, respectively. For the sake of simplicity, we will assume that $P_{in}(d^{in}) = P_{out}(d^{out}) = P(d)$.

B.2.1 Node degree correlation

In this section, we will assume that the in and out-degrees of the nodes are assigned such that the node degree correlation, as discussed in Section 3.4.2.1, is given by

$$\eta = \langle d^{in} d^{out} \rangle / \langle d \rangle^2, \tag{B.1}$$

where d^{in} and d^{out} are the in and out-degrees of a node, and $\langle ... \rangle$ denotes an average over the degree distribution in the network.

B.2.1.1 Maximal node degree correlation

When the in and out-degree distributions of the nodes are identical, then for the network that has maximal node degree correlation, $d^{in} = d^{out}$ for a node. For this case, we can write the formula for maximal node degree correlation as

$$\eta_{max} = \frac{\langle d^2 \rangle}{\langle d \rangle^2}. \tag{B.2}$$

For networks with Poisson degree distribution, this gives

$$\eta_{max,PD} = \frac{\langle d \rangle^2 + \langle d \rangle}{\langle d \rangle^2} = 1 + \frac{1}{\langle d \rangle}. \tag{B.3}$$

For scale-free networks with $\gamma > 3$, we get

$$\eta_{max,SF} = \frac{(\gamma - 2)^2}{(\gamma - 1)(\gamma - 3)}. \tag{B.4}$$

For $2 < \gamma \leq 3$, the second moment of degree distribution diverges. In that case, $\eta_{max,SF}$ also diverges.

B.2.1.2 Minimal node degree correlation

The network has minimal node degree correlation when the low in-degree nodes have high out-degrees and vice versa. To get an estimate of minimal η, let d_m be the median degree such that the number of nodes that have degree greater that d_m is equal to the number of nodes that have degree less than d_m.

The Poisson degree distribution is discrete and is non-symmetric about d_m. Numerically, minimal η can be easily calculated by generating the Poisson degree distribution. To get an analytical estimate for the case when the mean degree of the network is large, we approximate the discrete Poisson distribution by the continuous normal distribution with both mean and variance $\langle d \rangle$ [79]

$$\text{Poisson}(d; \langle d \rangle) \sim \mathcal{N}(d; \langle d \rangle, \langle d \rangle). \tag{B.5}$$

The normal distribution, $\mathcal{N}$, is a symmetric distribution, hence, the median degree in that case is $\langle d \rangle$. Then, for a network with minimal node degree correlation, we can write $d^{out} = 2\langle d \rangle - d^{in}$ for a node. We get this expression by assigning nodes with high out-degree (in-degree) the low in-degree (out-degree), as we move away from the median degree. The approximate minimal η for directed networks with

Poisson degree distribution with large $\langle d \rangle$ is then

$$\eta_{min,PD} \approx \frac{1}{\langle d \rangle^2} \int (2\langle d \rangle - d^{in}) d^{in} \mathcal{N}(d^{in})\ dd^{in}$$
$$= 1 - \frac{1}{\langle d \rangle}. \tag{B.6}$$

For scale-free networks, the median degree is $2^{1/(\gamma-1)} d_{min}$. For the case with minimal η, for any given node, $(d^{out})^{-\gamma+1} = 2(d_m)^{-\gamma+1} - (d^{in})^{-\gamma+1}$. Using Eq. (B.1) and integrating over the scale-free distribution, we get

$$\eta_{min,SF} = \frac{(\gamma-2)^2}{(\gamma-1)^2} \frac{\Gamma(b)}{\Gamma(c)}\ {}_2F_1(a, b; c; 1), \tag{B.7}$$

where ${}_2F_1$ is the Gauss hypergeometric function, $a = \frac{1}{\gamma-1}$, $b = 1 - a$ and $c = 2 - a$. ${}_2F_1$ in this case is convergent because $c - b - a > 0$ [80]. In terms of gamma functions, we can write [80]

$$\eta_{min,SF} = \frac{(\gamma-2)^2}{(\gamma-1)^2} \frac{[\Gamma(b)]^2}{\Gamma(2b)}. \tag{B.8}$$

B.2.2 Directional degree assortativity

In this section, we will assume that the in and out-degrees of the nodes in the network are uncorrelated but the connections between the nodes form such that the directional degree assortativity, as discussed in Section 3.4.2.2, is given by

$$\rho = \langle d_j^{in} d_i^{out} \rangle_e / \langle d_j^{in} \rangle_e \langle d_i^{out} \rangle_e, \tag{B.9}$$

where $\langle ... \rangle_e$ denotes average over all the edges in the network such that the directed link points from node j to node i.

B.2.2.1 Maximal directional degree assortativity

For networks with maximal directional degree assortativity, the nodes with higher in-degree have links pointing to nodes that have higher out-degree (vice versa for low degrees).

For networks with Poisson degree distribution, we write $\tilde{d}^{out} = \tilde{d}^{in}$ where the directed links point from nodes with degree $\tilde{d}^{in}$ to nodes with degree $\tilde{d}^{out}$ (here, $\tilde{d}^{in}$ and $\tilde{d}^{out}$ are degrees of the neighboring nodes). Since we assume that the in and out-degrees of a node are uncorrelated and that the in and out-degree distributions are identical, the average number of out-links from nodes with degree $\tilde{d}^{in}$ should be equal to the average number of in-links to nodes with degree $\tilde{d}^{out}$ when $\tilde{d}^{in} = \tilde{d}^{out}$. For the maximally correlated case, all the out-links from nodes with degree $\tilde{d}^{in}$ point to nodes with degree $\tilde{d}^{out}$. The total number of links emanating from all the nodes that have in-degree d^{in} and out-degree d^{out} is (here, d^{in} and d^{out} are degrees of the same node)

$$NP(d^{in})P(d^{out})d^{out}. \tag{B.10}$$

In Eq. (B.9), averaging over the links in the network (Eq. (B.10)), we get for networks with Poisson degree distribution with maximal directional degree assortativity

$$\rho_{max,PD} = \frac{\frac{1}{m}\sum (d^{in})^2 \quad NP(d^{in})P(d^{out})d^{out}}{\left(\frac{1}{m}\sum d^{in} \quad NP(d^{in})P(d^{out})d^{out}\right)^2}. \tag{B.11}$$

In the expression above, d^{in} and d^{out} are the degrees of the same node. This gives

$$\rho_{max,PD} = \frac{\langle d^2 \rangle}{\langle d \rangle^2} = 1 + \frac{1}{\langle d \rangle}, \tag{B.12}$$

where we used the fact that $m = N\langle d \rangle$.

For scale-free networks, we write

$$\begin{aligned}
\rho_{max,SF} &= \frac{\frac{1}{m}\iint (d^{in})^2 \;\; NP(d^{in})P(d^{out})d^{out}\; dd^{in}\; dd^{out}}{\left(\frac{1}{m}\iint d^{in} \;\; NP(d^{in})P(d^{out})d^{out}\; dd^{in}\; dd^{out}\right)^2} \\
&= \frac{\langle d^2 \rangle}{\langle d \rangle^2} \\
&= \begin{cases} \infty & \text{for } 2 < \gamma \leq 3 \\ \frac{(\gamma-2)^2}{(\gamma-1)(\gamma-3)} & \text{for } \gamma > 3 \end{cases}
\end{aligned} \tag{B.13}$$

B.2.2.2 Minimal directional degree assortativity

In this case, for a link from node j to node i, the in-degree of node j and the out-degree of node i are negatively correlated.

For networks with Poisson degree distribution, as in Section B.2.1.2, we assume that mean degree is large and approximate the Poisson degree distribution by the normal distribution with both mean and variance $\langle d \rangle$. For networks with minimal directional degree assortativity, we write $\tilde{d}^{out} = 2\langle d \rangle - \tilde{d}^{in}$ where the links point from nodes with degree $\tilde{d}^{in}$ to nodes with degree $\tilde{d}^{out}$. The approximate minimal ρ for directed networks with Poisson degree distribution with large $\langle d \rangle$, by averaging over the links in the networks (Eq. (B.10)), is then

$$\begin{aligned}
\rho_{min,PD} &\approx \frac{\frac{1}{m}\iint (2\langle d \rangle - d^{in})d^{in} \;\; N\mathcal{N}(d^{in})\mathcal{N}(d^{out})d^{out}\; dd^{in}\; dd^{out}}{\left(\frac{1}{m}\iint d^{in} \;\; N\mathcal{N}(d^{in})\mathcal{N}(d^{out})d^{out}\; dd^{in}\; dd^{out}\right)^2} \\
&= \frac{1}{\langle d \rangle^2}\int (2\langle d \rangle - d^{in})d^{in}\mathcal{N}(d^{in})\; dd^{in} \\
&= 1 - \frac{1}{\langle d \rangle},
\end{aligned} \tag{B.14}$$

where in the expressions above, d^{in} and d^{out} are the degrees of the same node.

For scale-free networks with minimal ρ, $(\tilde{d}^{out})^{-\gamma+1} = 2(d_m)^{-\gamma+1} - (\tilde{d}^{in})^{-\gamma+1}$ where the links point from nodes with degree $\tilde{d}^{in}$ to nodes with degree $\tilde{d}^{out}$. From this, we get

$$\begin{aligned}\rho_{min,SF} &= \frac{\frac{1}{m}\iint [2(d_m)^{-\gamma+1} - (d^{in})^{-\gamma+1}]^{1/(-\gamma+1)} d^{in} \;\; NP(d^{in})P(d^{out})d^{out}\; dd^{in}\; dd^{out}}{\left(\frac{1}{m}\iint d^{in} \;\; NP(d^{in})P(d^{out})d^{out}\; dd^{in}\; dd^{out}\right)^2} \\ &= \frac{(\gamma-2)^2}{(\gamma-1)^2}\frac{\Gamma(b)}{\Gamma(c)} \;\; {}_2F_1(a,b;c;1) \\ &= \frac{(\gamma-2)^2}{(\gamma-1)^2}\frac{[\Gamma(b)]^2}{\Gamma(2b)}. \end{aligned} \tag{B.15}$$

where, as in Eq. (B.7), $a = \frac{1}{\gamma-1}$, $b = 1-a$ and $c = 2-a$.

B.2.3 Assortativity by degree

To characterize assortativity by degree (ν), we use a measure similar to that used to characterize ρ above,

$$\nu = \langle d_j^{out} d_i^{in}\rangle_e / \langle d_j^{out}\rangle_e \langle d_i^{in}\rangle_e, \tag{B.16}$$

where, as usual, $\langle ...\rangle_e$ denotes average over all the edges in the network such that the directed link points from node j to node i. In this section also, we will assume that the in and out-degrees of the nodes in the network are uncorrelated but the connections between the nodes form such that assortativity by degree is given by the above expression.

B.2.3.1 Maximal assortativity by degree

In this case, for a link from node j to node i, the out-degree of node j and in-degree of node i are positively correlated. Thus, $\tilde{d}^{out} = \tilde{d}^{in}$ where the links point

from nodes with degree $\tilde{d}^{out}$ to nodes with degree $\tilde{d}^{in}$.

For networks with Poisson degree distribution, averaging over the links in the network (Eq. (B.10)), we write

$$\begin{aligned}\nu_{max,PD} &= \frac{\frac{1}{m}\sum (d^{out})^2 \quad NP(d^{in})P(d^{out})d^{out}}{\left(\frac{1}{m}\sum d^{out} \quad NP(d^{in})P(d^{out})d^{out}\right)^2} \\ &= \frac{\langle d\rangle\langle d^3\rangle}{\langle d^2\rangle^2} = 1 + \frac{\langle d\rangle}{(1+\langle d\rangle)^2},\end{aligned} \tag{B.17}$$

where d^{in} and d^{out} are the degrees of the same node. Here, we used $\langle d^3\rangle = \langle d\rangle(1 + 3\langle d\rangle + \langle d\rangle^2)$ for the Poisson distribution.

For the scale-free networks, we write

$$\begin{aligned}\nu_{max,SF} &= \frac{\frac{1}{m}\iint (d^{out})^2 \quad NP(d^{in})P(d^{out})d^{out}\ dd^{out}\ dd^{in}}{\left(\frac{1}{m}\iint d^{out} \quad NP(d^{in})P(d^{out})d^{out}\ dd^{out}\ dd^{in}\right)^2} \\ &= \frac{\langle d\rangle\langle d^3\rangle}{\langle d^2\rangle^2} \\ &= \begin{cases} \infty & \text{for } 2 < \gamma \leq 4 \\ \frac{(\gamma-3)^2}{(\gamma-2)(\gamma-4)} & \text{for } \gamma > 4 \end{cases}\end{aligned} \tag{B.18}$$

B.2.3.2 Minimal assortativity by degree

In this case, for a link from node j to node i, out-degree of node j and in-degree of node i are negatively correlated.

For scale-free networks with minimal ν, to get relationship between $\tilde{d}^{in}$ and $\tilde{d}^{out}$, where links point from nodes with degree $\tilde{d}^{out}$ to nodes with degree $\tilde{d}^{in}$, we equate the number of out-links from nodes at one extreme of the degree distribution to the number of in-links to nodes at the other extreme of the degree distribution

$$\int_{\tilde{d}^{out}}^{\infty} xP(x)\ dx = \int_{d_{min}}^{\tilde{d}^{in}} xP(x)\ dx. \tag{B.19}$$

From this, we get $(\tilde{d}^{in})^{-\gamma+2} = (d_{min})^{-\gamma+2} - (\tilde{d}^{out})^{-\gamma+2}$ for nodes that are connected to each other. Thus

$$\nu_{min,SF} = \frac{\frac{1}{m}\iint[(d_{min})^{-\gamma+2} - (d^{out})^{-\gamma+2}]^{1/(-\gamma+2)} d^{out} \;\; NP(d^{in})P(d^{out})d^{out}\; dd^{in}\; dd^{out}}{\left(\frac{1}{m}\iint d^{out} \;\; NP(d^{in})P(d^{out})d^{out}\; dd^{in}\; dd^{out}\right)^2}, \tag{B.20}$$

where in the expression above, d^{in} and d^{out} are the degrees of the same node. For $\gamma > 3$, this is given by

$$\nu_{min,SF} = \frac{(\gamma-3)^2}{(\gamma-2)^2}\frac{\Gamma(b)}{\Gamma(c)} \;\; {}_2F_1(a,b;c;1), \tag{B.21}$$

where, in this case, $a = \frac{1}{\gamma-2}$, $b = 1-a$ and $c = 2-a$. ${}_2F_1$ is convergent for $\gamma > 3$. In terms of gamma functions, we can write

$$\nu_{min,SF} = \frac{(\gamma-3)^2}{(\gamma-2)^2}\frac{[\Gamma(b)]^2}{\Gamma(2b)}. \tag{B.22}$$

For $2 < \gamma \leq 3$, $\nu_{min,SF}$ is zero.

For networks with Poisson degree distribution with large average degree, when we approximate the degree distribution by the normal distribution, we can not write $\tilde{d}^{in}$ of a node as an explicit function of $\tilde{d}^{out}$ of the node it is connected to, because the relationship between $\tilde{d}^{in}$ and $\tilde{d}^{out}$ comes out in terms of exponentials and integrals of normal distribution.

B.3 Discussions and conclusions

We derived mean-field formulas for estimating the maximum and minimum attainable degree correlations in large directed networks with Poisson and power

law degree distributions. Although one can determine the values of maximum and minimum attainable degree correlations using numerical simulations, we gain some interesting insights when using the above formulas. For example, for infinite scale-free networks, the maximal values of correlations diverge in certain cases. We also find that, in the mean-field approximation, the maximal values of ρ and η are given by the same formulas while maximal ν shows different behavior. Similar observations apply to minimal degree correlations.

Bibliography

[1] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phy. Rev. E*, 69:026113, 2004.

[2] A. Arenas, J. Duch, A. Fernandez, and S. Gomez. Size reduction of complex networks preserving modularity. *New J. Phys.*, 9:176, 2007.

[3] J. G. Restrepo, E. Ott, and B. R. Hunt. Synchronization in large directed networks of coupled phase oscillators. *Chaos*, 16:015107, 2006.

[4] J. G. Restrepo, E. Ott, and B. R. Hunt. Onset of synchronization in large networks of coupled oscillators. *Phy. Rev. E*, 71:036151, 2005.

[5] J. G. Restrepo, E. Ott, and B. R. Hunt. Weighted percolation on directed networks. *Phys. Rev. Lett.*, 100:058701, 2008.

[6] C. R. MacCluer. The many proofs and applications of Perron's theorem. *SIAM Rev.*, 42(3):487, 2000.

[7] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *PNAS*, 99(12):7821, 2002.

[8] M. E. J. Newman. Detecting community structure in networks. *Eur. Phy. J. B*, 38:321, 2004.

[9] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas. Comparing community structure identification. *J. Stat. Mech.*, page P09008, 2005.

[10] M. E. J. Newman. Modularity and community structure in networks. *PNAS*, 103(23):8577, 2006.

[11] J. Duch and A. Arenas. Community detection in complex networks using extremal optimization. *Phys. Rev. E*, 72:027104, 2005.

[12] R. Guimera and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433:895, 2005.

[13] E. A. Leicht and M. E. J. Newman. Community structure in directed networks. *Phy. Rev. Lett.*, 100:118703, 2008.

[14] L. Donetti and M. A. Munoz. Detecting network communities: a new systematic and efficient algorithm. *J. Stat. Mech.: Theor. Exp.*, page P10012, 2004.

[15] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori. Detecting communities in large networks. *Physica A: Statistical Mechanics and its Applications*, 352:669, 2005.

[16] A. Arenas, A. Diaz-Guilera, and C. J. Perez-Vicente. Synchronization reveals topological scales in complex networks. *Phy. Rev. Lett.*, 96:114102, 2006.

[17] J. A Almendral and A. Diaz-Guilera. Dynamical and spectral properties of complex networks. *New J. Phys.*, 9:187, 2007.

[18] I. J. Farkas, I. Derenyi, A.-L. Barabasi, and T. Vicsek. Spectra of "real-world" graphs: Beyond the semicircle law. *Phy. Rev. E*, 64:026704, 2001.

[19] K.-I. Goh, B. Kahng, and D. Kim. Spectra and eigenvectors of scale-free networks. *Phy. Rev. E*, 64:051903, 2001.

[20] M. L. Mehta. *Random Matrices*. Academic, New York, second edition, 1991.

[21] F. Juhasz. *Algebraic Methods in Graph Theory*, page 313. North Holland, Amsterdam, 1981.

[22] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[23] A.-L. Barabasi, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272:173, 1999.

[24] A.-L. Barabasi, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281:69, 2000.

[25] M. Bauer and O. Golinelli. Random incidence matrices: moments of the spectral density. *J. Stat. Phys.*, 103(1/2):301, 2001.

[26] J. G. Restrepo, E. Ott, and B. R. Hunt. Approximating the largest eigenvalue of network adjacency matrices. *Phy. Rev. E*, 76:056119, 2007.

[27] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Jhones Hopkins University Press, 1996.

[28] J. G. Restrepo, E. Ott, and B. R. Hunt. Characterizing the dynamical importance of network nodes and links. *Phy. Rev. Lett.*, 97:094102, 2006.

[29] R. A. Marcus. Brief comments on perturbation theory of a nonsymmetric matrix: The GF matrix. *The Journal of Physical Chemistry A*, 105(12):2612, 2001.

[30] V. Krebs. http://www.orgnet.com/. Unpublished.

[31] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd International Workshop on Link discovery*, page 36, New York, 2005. ACM.

[32] A. Barrat and M. Weigt. On the properties of small-world network models. *Eur. Phys. J B*, 13:547, 2000.

[33] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440, 1998.

[34] M. E. J. Newman. Mixing patterns in networks. *Phy. Rev. E*, 67:026126, 2003.

[35] G. Palla, A.-L. Barabasi, and T. Vicsek. Quantifying social group evolution. *Nature*, 446:664, 2007.

[36] A.-L. Barabasi and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101, 2004.

[37] P. M. Gleiser and L. Danon. Community structure in jazz. *Adv. Complex Syst.*, 6(4):565, 2003.

[38] V. Spirin and L. A. Mirny. Protein complexes and functional modules in molecular networks. *PNAS*, 100:12123, 2003.

[39] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, and A. Rapisarda. Detecting complex network modularity by dynamical clustering. *Phy. Rev. E*, 75:045102(R), 2007.

[40] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814, 2005.

[41] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466:761, 2010.

[42] S. Fortunato. Community detection in graphs. *Physics Reports*, 486:75, 2010.

[43] S. H. Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143:1, 2000.

[44] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer-Verlag, Berlin, 1984.

[45] T. Ichinomiya. Frequency synchronization in a random oscillator network. *Phy. Rev. E*, 70:026116, 2004.

[46] D.-S. Lee. Synchronization transition in scale-free networks: Clusters of synchrony. *Phy. Rev. E*, 72:026208, 2005.

[47] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626, 2000.

[48] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phy. Rev. E*, 64:026118, 2001.

[49] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Giant strongly connected component of directed networks. *Phy. Rev. E*, 64:025101(R), 2001.

[50] N. Schwartz, R. Cohen, D. ben Avraham, A.-L. Barabasi, and S. Havlin. Percolation in directed scale-free networks. *Phy. Rev. E*, 66:015104(R), 2002.

[51] M. Boguna and M. A. Serrano. Generalized percolation in random directed networks. *Phy. Rev. E*, 72:016106, 2005.

[52] A. Vazquez and Y. Moreno. Resilience to damage of graphs with degree correlations. *Phy. Rev. E*, 67:015101(R), 2003.

[53] M. A. Serrano and M. Boguna. Clustering in complex networks. ii. percolation properties. *Phy. Rev. E*, 74:056115, 2006.

[54] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671, 1983.

[55] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087, 1953.

[56] J. D. Hoffman. *Numerical Methods for Engineers and Scientists*. Marcel Dekker, Inc., 2001.

[57] H. W. Hethcote and J. A. Yorke. *Gonorrhea Transmission Dynamics and Control*, volume 56 of *Lecture Notes in Biomathematics*. Springer, Berlin, 1984.

[58] A. S. Klovdahl. Social networks and the spread of infectious diseases: The aids example. *Social Science & Medicine*, 21(11):1203, 1985.

[59] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1992.

[60] M. J. Keeling and K. T. D. Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295, 2005.

[61] C. Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Phys. Rev. E*, 61:5678, May 2000.

[62] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200, Apr 2001.

[63] R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Phys. Rev. E*, 64:066112, Nov 2001.

[64] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, Jul 2002.

[65] M. Boguna and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Phys. Rev. E*, 66:047104, Oct 2002.

[66] V. M. Eguiluz and K. Klemm. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.*, 89:108701, Aug 2002.

[67] M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B*, 266:859, 1999.

[68] K. T. D. Eames and M. J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *PNAS*, 99:13330, 2002.

[69] T. Gross, C. J. D. D'Lima, and B. Blasius. Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.*, 96:208701, May 2006.

[70] L. B. Shaw and I. B. Schwartz. Fluctuating epidemics on adaptive networks. *Phys. Rev. E*, 77:066101, Jun 2008.

[71] D. H. Zanette and S. Risau-Gusman. Infection spreading in a population with evolving contacts. *Journal of Biological Physics*, 34:135, 2008.

[72] E. Volz and L. A. Meyers. Susceptible-infected-recovered epidemics in dynamic contact networks. *Proceedings of the Royal Society B: Biological Sciences*, 274(1628):2925, 2007.

[73] M. Altmann. Susceptible-infected-removed epidemic models with dynamic partnerships. *Journal of Mathematical Biology*, 33:661, 1995.

[74] M. Kretzschmar and M. Morris. Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences*, 133(2):165, 1996.

[75] A. C. Ghani, J. Swinton, and G. P. Garnett. The role of sexual partnership networks in the epidemiology of gonorrhea. *Sexually Transmitted Diseases*, 24:45, 1997.

[76] S. E. Chick, A. L. Adams, and J. S. Koopman. Analysis and simulation of a stochastic, discrete-individual model of std transmission with partnership concurrency. *Mathematical Biosciences*, 166(1):45, 2000.

[77] K. T. D. Eames and M. J. Keeling. Monogamous networks and the spread of sexually transmitted diseases. *Mathematical Biosciences*, 189(2):115, 2004.

[78] M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.*, 92:178701, Apr 2004.

[79] D. G. Rees. *Foundations of Statistics*. Chapman and Hall, London, 1987.

[80] W. N. Bailey. *Generalized Hypergeometric Series*, volume 32 of *Cambridge Tracts in Mathematics and Mathematical Physics*. Stechert-Hafner Service Agency, New York, 1964.