Spectral Theory for the Robustness and Dynamical Properties of Complex Networks

by

Dane R. Taylor

M.S., Applied Mathematics, University of Colorado, Boulder, 2012

B.S., Physics, University of Wyoming, 2008

B.S., Electrical Engineering, University of Wyoming, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Applied Mathematics

2013

This thesis entitled:
Spectral Theory for the Robustness and Dynamical Properties of Complex Networks
written by Dane R. Taylor
has been approved for the Department of Applied Mathematics

_____

Prof. Juan G. Restrepo

_____

Prof. François G. Meyer

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Taylor, Dane R. (Ph.D., Applied Mathematics)

Spectral Theory for the Robustness and Dynamical Properties of Complex Networks

Thesis directed by Prof. Prof. Juan G. Restrepo

From biological processes to critical infrastructures and social phenomena, many complex systems may be studied as large networks of interacting components. Research investigating the important role of network topology is therefore of broad interest, where techniques may be developed, for example, to control complex dynamical processes with strategic network modifications. Applications range from mitigating damage incurred to critical infrastructure (e.g., the energy, banking, and transit systems) to controlling spreading processes, including both those that are harmful (e.g., epidemics) and beneficial (e.g., information dissemination). Among the many successful techniques for studying complex networks, spectral graph theory has been shown to be remarkably useful for analyzing and controlling the dynamical and robustness properties of a given network. In this thesis, I discuss my contributions to this field, which explore the following applications: (i) The analysis of a given network's robustness to the strategic removal of nodes and/or links; (ii) The development of techniques to judiciously modify a network to tune its robustness and dynamical properties; and (iii) The introduction and analysis of a network formation process yielding networks that self-organize with enhanced spreading and robustness characteristics.

Dedication

To my parents Rod and Ellen. I am eternally grateful for your unbounded love and support.

Acknowledgements

First and foremost, I would like to thank my advisor Juan G. Restrepo. The work provided in this thesis was conducted under Juan's guidance and his leadership was integral to this research. I also would like to express my gratitude to Daniel B. Larremore, with whom I collaborated for the work provided in chapter 3. Daniel was also a graduate student of Juan's, so again I must thank Juan for steering our thoughts in this direction. I would also like to thank Robert S. Maier for helping with the analysis of the asymptotic behavior of Eq. (4.4) by suggesting the use of a dilogarithm. In addition to these contributors for the work provided in this thesis, I also want to acknowledge others whom I've collaborated with during my graduate education, including Ed Ott, Per Sebastien Skardal, François G. Meyer, Jie (Rio) Sun, and Elana J. Fertig. Our many discussions have been integral for my development as an applied mathematician. A final thanks to the remaining members of my thesis committee, James D. Meiss, Aaron Clauset, and Manuel B. Lladser, as well as my funding sources, which include NSF Grant No. DMS-0908221, the Achievement Rewards for College Scientists (ARCS) Foundation, Lockheed Martin, and the Department of Applied Mathematics at the University of Colorado, Boulder.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1 The Network Takeover

Since the turn of the century, there has been a considerable rise in research studying the interplay between structure and function for very large systems [12, 106, 105]. Due to its interdisciplinary nature, this developing field of network science has had a very broad impact, which was recently referred to as the "network takeover" [10]. A central cause for the rapid rise of this field is that many natural and technological systems are large and complex, requiring special techniques to handle huge amounts of information and complexity. For example, the human brain is composed of billions of neurons, each with local dynamics and interactions with other neurons [30], and a systematic approach is required if we are to understand the mechanistic basis of memory and cognition. A network-based mathematical approach is extremely appealing for such endeavors as it is a formalism allowing one to connect microscopic and macroscopic dynamics. This is essential for *complex systems*, which exhibit emergent macroscopic behavior not easily predicted from microscopic properties [9, 137]. The following is a sample of some of the applications in which networks are becoming increasingly important.

Biological Systems

Large systems composed of interacting units are ubiquitous in nature. For example, the dynamics of ecological systems are often studied as large food webs of interacting species, where the structure of interactions can play a vital role [113]. The structure of interactions between individu-

als within a single species also outlines collective behaviors, ranging from fish and avian motion [48] to insect mating [29]. The dynamics of large physiological systems are also naturally manifest as networks, including neuronal dynamics [30], genetic expression [114], protein interactions [28, 66], and metabolic reactions [97]. The dynamics of each of these systems, as well as their interconnectivity, has indicated network theory as a central direction of research for the development of future medicine [13, 73, 163].

Social Networks

Within the social sciences, network theory has had an unprecedented impact [83, 130, 156]. For example, social network analysis (e.g., studying the topology of face-to-face or technological human interaction) is essential for understanding spreading processes, which range from the dissemination of information [75] to harmful epidemics [37, 155]. Theory for engineering social networks may also suggest strategies for improving the performance of various systems, such as corporate hierarchies [108, 119], political infrastructure [99, 115], or economies [149].

Critical Infrastructures

One area of national importance includes critical infrastructures, or the "backbone" networks on which society depends. These networks include transit systems such as the roads [14, 15, 67] and airlines [16], communication systems used for the internet [46] and mobile phones [154], as well as energy systems such as the power grid [49, 50, 148]. Structural analyses for such critical infrastructures can highlight bottleneck elements that are susceptible to attack and failure [31, 98, 121], while also suggesting modifications for improvement [129, 141, 144].

1.2 Networks with Complex Structural Properties

For many applications, the system may be formulated as a set of N nodes $\{n\}_{n=1}^{N}$ (i.e., vertices) and a set of M links $\{l\}_{l=1}^{M}$ (i.e., edges) connecting the nodes. For undirected networks a link corresponds to an unordered pair of node indices, whereas for directed networks it corresponds

to an ordered pair. The structure described by a given set of nodes and links (and possibly link weights) defines a network, or graph $\mathcal{G}$. For unlabeled networks, $\mathcal{G}$ may be thought of as an equivalence class such that a given graph corresponds to all node and link sets that may be obtained under node index permutations (e.g., relabeling node 1 to node 2 and node 2 to node 1 does not change the network's structure). In addition to defining a graph by sets of nodes and links, $\mathcal{G}$ may be encoded by its network adjacency matrix A such that each entry A_{nm} is nonzero if and only if a link exists from node n to node m. Entries A_{nm} may be weighted, but are often taken to be unitary. Note that A is a symmetric matrix for undirected networks.

Because networks are essentially topological structures, they can be characterized by different structural properties. These properties allow for the comparison and classification of networks, which can range from structures that are homogeneous and simple to those that are heterogeneous and complex. The following are some classes of networks, which are listed in order of increasing structural complexity: regular, low-dimensional lattices such as square and triangular lattices [24]; low dimensional, random geometric graphs (i.e., manifold graphs) [112]; random graphs such as small world networks [157]; random graphs with many complex properties [12, 106, 105]. The following observation helps explain why the recent rise in network theory has occurred despite the field of graph theory having a very long history [25, 41]: Whereas most theory has been developed for networks within the first few classes, most real-world networks belong to the last class [12, 106, 105], which are referred to as *complex networks.* The following are a few of most significant complex network properties.

- **Heterogeneity** - A major difference between simple networks and complex networks is that real-world networks are typically composed of nodes with heterogeneous connections. Given a node n, its degree is defined as the number of links connecting it to other nodes, $d_n = \sum_m A_{nm}$. A node n in a directed network will have an in-degree $d_n^{in} = \sum_m A_{mn}$ and out-degree $d_n^{out} = \sum_m A_{nm}$, which respectively denote the number of incoming and outgoing links. In complex networks the distributions of degrees $P(d)$ is often heavy tailed,

indicating that the nodal degrees can vary by orders of magnitude [11, 107]. One such distribution is given by $P(d) \propto d^{-\gamma}$ for $\gamma > 0$ [although typically $\gamma \in (2, 3)$], which is often referred to as a scale free distribution [11].

- **Degree Correlations** - Correlations between nodal degrees are also common in real-world networks and can have a significant effect on network processes [104]. Degree correlations between neighboring nodes (i.e., *Markovian correlations*) may be measured by a Pearson's correlation coefficient

$$r = \frac{\sum_{nm} A_{nm} d_n d_m - M^{-1} \left(\sum_{nm} A_{nm} d_n\right)^2}{\sum_{nm} A_{nm} d_n^2 - M^{-1} \left(\sum_{nm} A_{nm} d_n\right)^2}, \tag{1.1}$$

  which is shown for unweighted, undirected networks and is often referred to as the assortativity coefficient [104]. Measuring correlations allows networks to be classified into two groups: assortative networks in which neighbors tend to have similar properties ($r > 0$) and disassortative networks in which neighbors tend to have dissimilar properties ($r < 0$). Interestingly, biological networks tend to be disassortative while social networks tend to be assortative. Degree correlations between non-neighboring nodes (i.e., *non-Markovian correlations*) can also be significant [145].

- **Hierarchical Community Structure** - Real-world networks often organize naturally into sets of nodes in which there are many links between nodes within the same set and few links between nodes in different sets [45, 58, 99]. The presence of these sets, which are referred to communities, or modules (typically used when referring to biological networks), can greatly affect various dynamics on networks [135, 160]. Moreover, the communities themselves are often organized into large scale communities, resulting in a hierarchical organization of the nodes [43, 44].

- **Subgraph Motifs** - In most real-world networks there are small structures, or subgraphs, that are much more prevalent than statistically expected [6, 93]. These subgraph motifs have been shown to play an important role in the functioning networks [69, 118] and

have even been referred to as the "building blocks" of biological networks [94]. The most commonly studied subgraphs are small loops. The prevalence of loops of size three is often referred to as clustering, or transitivity [64], whereas the prevalence of loops of size two in directed networks is often referred to as reciprocity [130, 156].

1.3 Spectral Graph Theory: An Introduction

Given an adjacency matrix A, a central pursuit in the field of spectral graph theory is to study the set of eigenvalues $\{\lambda_i\}$ and eigenvectors $\{u^{(i)}\}$ and $\{v^{(i)}\}$ corresponding to the eigenvalue equations $Au^{(i)} = \lambda_i u^{(i)}$ and $A^T v^{(i)} = \lambda_i v^{(i)}$ [41]. Typically the eigenvectors are assumed to be normalized with $||u^{(i)}||_2 = ||v^{(i)}||_2 = 1$. Note that A is symmetric for undirected networks and hence $v^{(i)} = u^{(i)}$ and $\lambda_i \in \mathbb{R}$ for each i (otherwise $\lambda_i \in \mathbb{C}$). Various pursuits range from predicting the distribution of eigenvalues in the limit $N \to \infty$ to studying the variability of a specific eigenvalue [40]. Of particular interest is theory describing the largest eigenvalue λ and its corresponding left and right eigenvectors, $v = [v_1, v_2, \ldots, v_N]^T$ and $u = [u_1, u_2, \ldots, u_N]^T$. For networks with positive links $A_{nm} \geq 0$ one can show that u_n, v_n, and $\lambda \geq 0$ [151]. If the network contains at least one loop then $\lambda > 0$ [151]. If the network is also strongly connected (i.e., for any nodes n and m there exists a path $n \to m$ of some length p), the Perron-Frobenius theorem for irreducible nonnegative matrices further implies u_n, v_n, and $\lambda > 0$ [8]. For networks lacking strong community structure, λ is often well separated from the remaining eigenvalues [125], otherwise one can expect a large eigenvalue to correspond to each macroscopic community [35] [1] . The dominant eigenspace given by λ, u, and v is particularly important as it has a very intuitive interpretation for networks. Specifically, if we temporarily restrict our interest to unweighted networks (i.e., $A_{nm} \in \{0, 1\}$) and denote $L_n^{o,p} = \sum_m (A^p)_{nm}$ and $L_n^{t,p} = \sum_m (A^p)_{mn}$ to be the number of paths of length p originating from and terminating at node n, respectively, and $L_p = \sum_{nm} (A^p)_{nm}$ to be the total number of paths of length p, then these quantities satisfy $||L_n^{o,p}||_2^{-1} L_n^{o,p} \to u_n$, $||L_n^{t,p}||_2^{-1} L_n^{t,p} \to v_n$, and $L_{p+1}/L_p \to \lambda$ as

[1] The one-to-on correspondence between the number of communities and large eigenvalues additionally assumes that all communities have a similar structure (i.e., similar nodal heterogeneity).

$p \to \infty$ [144]. Therefore, with respect to long paths in the network, u_n and v_n respectively describe expansion and contraction at node n and λ is a global measure for expansion, or connectivity.

While u_n, v_n and λ provide considerable insight toward the structure of a given network, importantly, they also naturally reflect many complex network properties. For example, one can accurately predict $u_n \propto d_n^{out}$ and $v_n \propto d_n^{in}$ for random networks lacking degree correlations. Intuitively this says that the number of very long outgoing and incoming long paths is well approximated by the number of incoming and outgoing paths of length one (i.e., the number of links). With respect to the largest eigenvalue, for random networks one obtains the following approximation [40, 125]

$$\lambda \approx \max\left\{\frac{\langle d^{in} d^{out}\rangle}{\langle d\rangle}, \sqrt{\max_n d_n}\right\}, \tag{1.2}$$

where $\langle\cdot\rangle$ denotes the average over nodes. This result essentially states that if a single node dominants the network then λ is approximately given by the square root of its degree, otherwise λ is well predicted by the ratio of the number of paths of length two versus the number of paths of length one (i.e., L_2/L_1). These approximations for u_n, v_n and λ for uncorrelated, random networks account for nodal heterogeneity and are often referred to as the mean field approximations. Additional approximations may be developed for the spectral properties of networks with degree correlations [125], subgraph motifs [116], and community structure [35].

Because many complex properties are naturally present in network spectra, spectral theory describing the robustness and dynamical properties of networks will often be inherently robust to the presence of complex structural properties. This is a major benefit of this approach to network analysis over non-spectral techniques (e.g., generating functions leading to mean field equations [106]), which typically require new theory for each added type of structural complexity! However, we point out that these techniques may be more generally classified as either network-specific (e.g., spectral theory for a given network matrix) or ensemble-based (e.g., mean field equations), and therefore the two approaches are complementary. For example, for most applications one is either interested in the properties of a single network or in the properties of a class, or ensemble, of networks (often in the asymptotic limit $N \to \infty$). For the first type of problem it is beneficial to

utilize the actual network topology (e.g., as encoded by its adjacency matrix) to develop analysis. Alternatively, if a given network is assumed to only be representative from an ensemble, then it may be beneficial to develop analysis using its general properties rather than its particular structure. However, because the effects of complex properties on network spectra can often be accounted for [35, 40, 116, 125], the spectral approach has a clear advantage.

1.4 Spectral Theory for Dynamics on Networks: A Survey

A central theme of this thesis will be that judiciously tuning the dominant eigenspace (i.e., the largest eigenvalue and its corresponding eigenvectors) of the network adjacency matrix is a fundamental technique for controlling complex systems. To help motivate this line of research, in this section we summarize recent results showing that the analysis of a broad class of dynamics, which we will refer to as *connectivity governed dynamics*, depends strongly on the dominant eigenspace. In the following chapter we present a spectral theory for how network robustness also depends on this eigenspace. Together these results highlight the broad importance of spectral theory for the robustness and dynamical properties of complex networks and indicate the necessity of analysis for the perturbation and evolution of spectra (the topics of Ch. 3 and Ch. 4). We now briefly discuss four such types of dynamics on networks: (i) the synchronization of heterogeneous phase oscillators, (ii) epidemic spreading and the dissemination of information, (iii) cascades of neuronal excitation, and (iv) the stability of boolean models of gene expression.

Synchronization of Heterogeneous Phase Oscillators

Network coupled Kuramoto phase oscillators are widely studied as a paradigmatic model for understanding collective behavior and synchronization phenomena [140]. Among the many applications, several include the flashing of fireflies [29], clapping audiences [102], neuronal synchrony [88, 131], and animal motion [84, 100]. Given a set of oscillators $\{n\}_{n=1}^{N}$ with phases $\{\theta_n\}_{n=1}^{N}$ coupled by a network with weighted adjacency matrix A, the evolution of each oscillator phase θ_n

is given by [78]

$$\dot{\theta} = \omega_n + k \sum_m A_{nm} \sin\left(\theta_m - \theta_n\right), \tag{1.3}$$

where the scalar k denotes the coupling strength and ω_n is oscillator n's natural frequency, which is taken from a unimodal distribution $\Omega(\omega)$. Letting $re^{i\psi} = N^{-1} \sum_n e^{i\theta_n}$ denote the mean field, one obtains the order parameter $r \in [0, 1]$ measuring the extent of synchrony and the phase $\psi \in [0, 2\pi]$ at which the oscillators' phases cluster. Interestingly, for small coupling strength $k \leq k_c \equiv \frac{2}{\pi \Omega(0) \lambda}$ the oscillators do not synchronize (where λ is the largest eigenvalue of A), whereas a synchronized cluster emerges for large coupling strength $k > k_c$ [120, 124]. Moreover, when synchrony occurs the local extent of synchronization at a node n $r_n = \sum_m A_{nm} r_m$ is proportional to the right eigenvector entry, $r_n \propto u_n$ [120, 124]. Therefore, the characteristics of synchrony for a given network may be deduced by examining the network's dominant eigenspace. If, however, one assumes that the network is a representative of an ensemble of networks, mean field predictions may be applied for random networks lacking degree correlations to obtain, $k_c \approx \frac{2\langle d \rangle}{\pi \Omega(0) \langle d^2 \rangle}$ [65]. See Ref. [125] for further discussion on the assumptions required for the applicability of the mean field approach. Generalizing Eq. (1.3) to more realistic and complex dynamics remains an ongoing area of research [136, 137, 143].

Epidemic Spreading and the Dissemination of Information

Although there are many competing models for spreading processes on networks, we focus on the *susceptible-infected-susceptible* (SIS) model as it has received considerable attention in literature, and it allows for the modeling of both epidemic spreading [155] and the dissemination of information [75]. The SIS model is a continuous time process in which each node may either be (i) susceptible to infection or (ii) infected. Each infected node may infect each of its susceptible network neighbors at rate α, and each infected node may also spontaneously heal and return to being susceptible at rate β. The network state in which no nodes are infected and all nodes are susceptible is a fixed point of the collective dynamics, but this fixed point may not be stable to perturbation (i.e., a small fraction

of nodes being infected by some external agent). For many topologies of connected networks in which a fraction of nodes are initially infected, the expected steady state fraction of infected nodes f may either be zero (no infections, stable fixed point) or nonzero (endemic infection, unstable fixed point), depending on whether α/β surpasses the epidemic threshold λ^{-1} [34, 110, 152, 155]. Moreover, at the onset of infection one can estimate that the steady state probability of infection for a given node n is proportional to the left eigenvector entry v_n [152]. As before, the mean field approximations for network spectra may be applied to predict the epidemic characteristics for an ensemble of random, uncorrelated networks with a given degree distribution [87, 110].

Cascades of Neuronal Excitation

Restricting our attention to the discrete time Kinouchi-Copelli model for neuronal activity [71], the state $x_m(t)$ of a given node, or neuron, n at time t may be in one of $m+1$ states: the resting state $x_n(t) = 0$; the active state $x_n(t) = 1$; or one of the $m-1$ refractory states $x_n(t) = 2, 3, \ldots, m$. In discrete times enumerated $t = 1, 2, \ldots$, the states are simultaneously updated via the following: (i) If a node n is in the resting state (i.e., $x_n(t) = 0$), then it can be excited either by an external stimulus η or by a neighboring active node m (i.e., $x_m(t) = 1$) with probability $0 \leq A_{nm} < 1$, so that $x_n(t+1) = 1$. (ii) If a node n is excited (i.e., $x_n(t) = 1$), then it will deterministically transition to the first refractory state $x(t+1) = 2$. (iii) If a node n is in any refractory state except the last (i.e., $x_n(t) \in \{2, \ldots, m-1\}$), then it will deterministically transition to the next refractory state $x(t+1) = x(t)+1$. (iv) If a node n is in the last refractory state (i.e., $x_n(t) = m$), then it will deterministically transition to the resting state (i.e., $x_n(t+1) = 0$). It follows that the interactions between neurons are encoded by the weighted adjacency matrix A, whose entries denote the probability that one neuron excites another. In Ref. [82] it is shown that the steady state fraction of active nodes either has no activation or nonzero activation as a fixed point, depending on whether the largest eigenvalue λ of A is respectively less than or greater than one. It was found that $\lambda = 1$ is a critical case in which the network is at criticality. Various in-vitro and in-vivo experiments have suggested that actual neuronal tissues maintain this critical state as it maximizes

the dynamic range, or diversity of signals, that may be uniquely sensed [82]. Further study of the interplay between network topology and the dynamics of this model is presented in Refs. [80, 81].

Stability of Gene-Expression

Although models of varying complexity have been proposed to study dynamic gene expression, we will focus on boolean models in which the state $\sigma_n(t)$ of gene n at time $t = 1, 2, \ldots$ may be in one of S_n states, enumerated $\sigma_n \in \{1, \ldots, S_n\}$ [114]. As before we can embed the interactions between genes with the adjacency matrix A, such that A_{nm} is nonzero only for genes $\{m\}$ that influence gene n. The discrete dynamics of the state of a given gene n is thus given by the update function $\sigma_n(t+1) = f_n(\{\sigma_m(t)|A_{nm} \neq 0\})$. Rather than assume a specific functional form for f_n, if one instead considers the probability q_n that a change in input $\{\sigma_m(t)|A_{nm} \neq 0\}$ alters f_n, it follows that the dynamics are governed by the dominant eigenvalue λ_Q of the matrix defined by $Q_{nm} = q_n A_{nm}$ [114]. Specifically, for a given trajectory $\{\sigma_n(t)\}$, one can study whether a perturbed trajectory $\{\tilde{\sigma}_n(t)\}_{n=1}^N$ converges to $\{\sigma_n(t)\}_{n=1}^N$ or diverges as $t \to \infty$. To measure whether or not the perturbation grows, one can consider the difference vector $y_n(t) = \langle\langle|\sigma_n(t) - \tilde{\sigma}_n(t)|\rangle\rangle_\epsilon$, where $\langle\langle\cdot\rangle\rangle_\epsilon$ denotes the average over all realizations in which ϵN of the states $\sigma_n(t)$ are randomly switched. For $\lambda_Q < 1$, $y_n = 0$ remains a stable fixed point of the system, and perturbations decay with time. For $\lambda_Q > 1$ perturbations grow, and a given state trajectory is unstable [114]. Therefore λ_Q indicates a critical value marking the boundary between stable and unstable trajectories of gene expression. Moreover, these dynamics have been recently shown to map to weighted percolation in networks [139], the topic of the following chapter.

1.5 Overview of Thesis

- **Chapter 1** - In Ch. 1 we have presented an overview of a broad field sometimes referred to as network science. We have introduced several important complex properties that are commonly found in real-world networks and provided a brief introduction to spectral theory for complex networks. A survey of dynamics on networks was provided to illustrate that

spectral theory is a crucial area of research for this developing field. Importantly, because many robustness and dynamical properties depend directly on network spectra, one can infer the effects of complex network properties simply by observing their effects on spectra (e.g., the dominant eigenspace).

- **Chapter 2** - In Ch. 2 we introduce our spectral theory analyzing the robustness of networks with bidirectional links subject to biased strategies of node and/or link removal (i.e., weighted percolation). Much of this work was published in Ref. [145]. Among many important conclusions, a central result is that our theory estimates the percolation threshold, which is a natural measure for network robustness. Similar to results presented in Sec. 1.4 for dynamical systems, we find that the dominant eigenspace greatly determines the robustness of a given network. In particular, we find that large λ often indicates robustness.

- **Chapter 3** - Motivated by results in Ch. 1 and Ch. 2 indicating that the dominant eigenspace governs the robustness and dynamical properties for a broad range of network types, in Ch. 3 we study the perturbation of λ and its eigenvectors for a network undergoing structural modification. Specifically, techniques are introduced to optimize the addition of a secondary network, or module, to control a given network's dynamical and robustness properties. Much of this work was published in Ref. [144].

- **Chapter 4** - Utilizing perturbation results from Ch. 3, in Ch. 4 we introduce a network formation process called *Social Climber* (SC) attachment that iteratively maximizes network connectivity (i.e., as measured by λ) during formation. Much of this work was published in Ref. [142]. Besides leading to networks for which λ has maximal scaling $\lambda \sim \mathcal{O}(N^{1/2})$ for a network with $\mathcal{O}(N)$ links, a novel phase transition is observed by adopting λ as the order parameter for link percolation processes. Finally, the broad importance of studying evolving matrix spectra is discussed.

- **Chapter 5** - Conclusions are drawn and several directions of future work are indicated.

Chapter 2

The Robustness of Complex Networks

The widespread importance of networks for science, technology, and society has established the study of network robustness as a vital direction of research. In fact, the broad diversity of applications has led to various definitions and techniques for studying robustness, which can include investigating the resilience of a network against random failure [95, 96, 98] or strategic attack [57, 121]. However, in general one is typically interested in the fragmentation of a network as it is dismantled. This has many applications, which may be organized into two groups: applications in which it is beneficial to design networks that withstand fragmentation [92, 129, 144], and applications involving the development of targeted strategies for fragmenting a network [57, 121]. One of the main applications in the first group is the design of critical infrastructures that are resilient to random failure and/or targeted attack [129]. These can include transit systems such as the airline flight network [16], communication systems such as mobile phones [154], the internet's routing infrastructure [132], as well as utilities such as the power grid [148]. The second group is perhaps more diverse and can include the development of military strategies to fragment critical infrastructures and social networks [61], the development of immunization strategies to fragment social networks to prevent epidemic spreading [47, 87, 111], and the development of novel medicines that utilize network modifications to disrupt biological processes related to disease [13].

2.1 Background Information

The robustness of a network can be measured through its *percolation threshold*, which may be interpreted as the critical number of link and/or node removals that must occur for the network to shift from a regime in which it is only slightly perturbed to a regime in which it is fragmented into small disconnected clusters. We introduce this concept for an undirected network subject to the uniform removal of nodes at probability p, a process referred to as *unweighted node percolation*. We will later consider the removal of links as well as the simultaneous removal of nodes and links.

Consider an undirected, unweighted network with nodes $n = 1, 2, \ldots, N$ and links $l = 1, 2, \ldots, M$. The network may be encoded by the unweighted network adjacency matrix A ($A_{nm} = A_{mn} = 1$ if there is a link from node n to node m and $A_{nm} = A_{mn} = 0$ otherwise). Given such a network, consider the resulting network after each node is removed with uniformly probability p. The resulting network may be described by a new adjacency matrix B given by: $B_{nm} = A_{nm}$ if neither node n or node m is removed and $B_{nm} = 0$ otherwise. Note that when a node is removed, all of the links attached to the node are also removed. A single realization of this probabilistic removal process is referred to as a percolation trial, and in general we are interested in the expected structure of the retained network after a single realization of such a removal process. Note that after such a process, qN nodes are expected to remain in the network, where we refer to $q = 1 - p$ as the uniform, or unweighted, retention probability.

Of the many network properties of interest for such retained networks, the field of percolation theory investigates the size of *clusters*, or disconnected subgraphs. For undirected networks with nodes $\{n\}_{n=1}^{N}$, depending on the network topology these nodes may be allocated into disjoint sets $\{\Omega_k\}_{k=1}^{K}$ with $1 \leq K \leq N$. Here $K = 1$ indicates that all nodes are connected in a single cluster, and $K = N$ indicates that all nodes are disconnected and there are no links. In order for a subset Ω_k to correspond to the nodes in a given cluster k, it must satisfy the following two properties: (1) for any two nodes n and m within a given cluster, there exists an undirected path from node n to node m of some length L. More precisely, if the remaining network is described by the adjacency matrix

B, then this requires that for each n and m in Ω_k there exists some $L > 0$ such that $[B^L]_{nm} \neq 0$. (2) For any node $n \in \Omega_k$, all nodes that may be reached via the network from node n are also in its cluster k (i.e., $\{m : [B^L]_{nm} \neq 0 \text{ for some } L\} \subseteq \Omega_k$). In summary, any undirected network may be split up into subnetworks, or clusters, such that the nodes within a cluster are connected by an undirected path and there are no links between clusters. They are disjoint and unconnected.

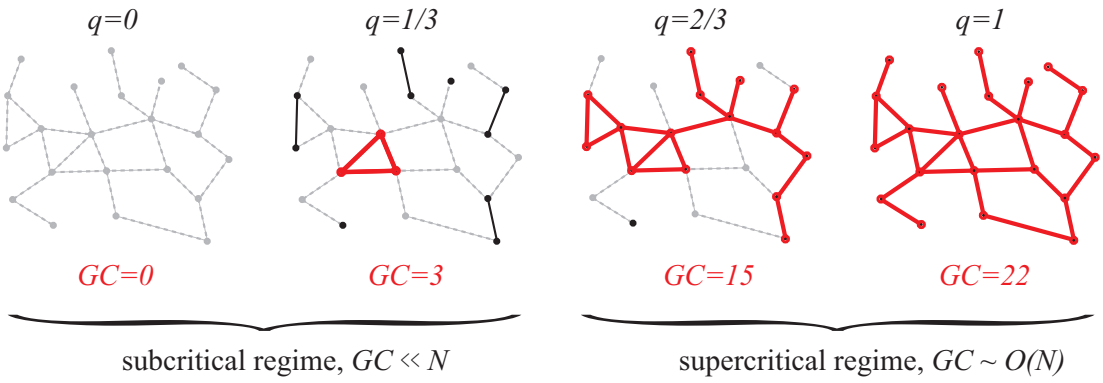


Figure 2.1: The resulting networks are shown for a network subject to probabilistic node removal with increasing retention rate q. Gray nodes and links indicate removed network structure, while black nodes and links indicate retained structure. If these retained nodes and links happen to be in the largest cluster, then they are colored red. Note that the size of the largest cluster GC increases with q. Comparing GC to the network size N leads to two fragmentation regimes: $GC \ll N$ in the subcritical regime and $GC \sim \mathcal{O}(N)$ in the supercritical regime. The critical fraction of node removals q^* serving as the boundary between these two regimes is referred to as the *percolation threshold.*

In Fig. 2.1 we show a small undirected network in which each node is retained with uniform probability q. Example networks obtained after a percolation trial are shown for increasing q, $q \in \{0, 1/3, 2/3, 1\}$. Removed nodes and their adjoined links are shaded gray. Black nodes and

links indicate network structure that is retained after a percolation trial, while those that are highlighted red indicate the largest cluster in each remaining network. We denote the size of a cluster k, which contains nodes Ω_k, as $|\Omega_k|$. If the size of a cluster k is of the same order as the network size, i.e., $|\Omega_k| \sim \mathcal{O}(N)$, then cluster k is referred to as a giant component. Therefore, with slight abuse of notation, we denote the size of the largest cluster by $GC = \max_k |\Omega_k|$. For the network fragmentation processes considered in this thesis, we will assume that all nodes are in the initial network's giant component, and thus any network to be studied will contain only one cluster prior to removals [i.e., $GC = N$ for $q = 1$].

Although the size of the largest cluster of a network obtained from a percolation trial is expected to increase from 0 to N as the retention probability q increases from 0 to 1, unlike the size of the retained network (which varies linearly with q and has expectation qN), the expected size of the largest cluster has two regimes. These regimes are separated by the *percolation threshold* q^* and are defined as follows: For $q < q^*$, with high probability GC is much smaller than the network size ($GC \ll N$) and the retained network is said to be in the subcritical regime. For $q > q^*$, with high probability GC is on the same order as the network size [$GC \sim \mathcal{O}(N)$] and the retained network is said to be in the supercritical regime. Because the percolation threshold q^* indicates the expected fraction of nodes that must be retained for the existence of a giant component in the network, it indicates precisely when a network is expected to fragment and is thus a natural measure of network robustness.

In Fig. 2.2 we plot the size of the largest cluster GC normalized by the network size $N = 2000$ for a network in which nodes are retained with increasing probability q. We will further refer to such plots as $GC(q)$ curves, where for fixed q a particular $GC(q)$ value can reflect a single percolation trial or a value averaged over several percolation trials. The network under consideration in Fig. 2.2 was obtained by retaining the largest component of a network generated using the Erdős-Rényi (ER) network formation process [55] with the addition of $M \approx N$ undirected links (i.e., which were iteratively added by randomly selecting two nodes to join). This yielded an average nodal degree of $\langle d \rangle \approx 2$ and the percolation threshold $q^* = \langle d \rangle^{-1} = 0.5$ [22]. Note that the curve shown

in Fig. 2.2 appears to be "noisy" due to finite size fluctuations and the fact that the GC values reflect a single percolation trial for each q. These fluctuations may be "smoothed out" by averaging many percolation trials for each q or by considering the asymptotic limit $N \to \infty$. For the limit $N \to \infty$, the curve $GC(q)$ limits to a continuous function and the network is said to undergo a second order, or continuous, phase transition at q^*. For example, if we had considered the entire network obtained from the ER process rather than just its giant component (i.e., which ensured that $GC = N$ for $q = 1$), then in the limit $N \to \infty$ the fraction of nodes in the giant component, $G = GC/N$, would be given by [22]

$$G = \begin{cases} 0 & , \quad q \leq \langle d \rangle^{-1} \\ q(1 - e^{-\langle d \rangle G}) & , \quad q > \langle d \rangle^{-1} \end{cases} . \tag{2.1}$$

This result follows from applying the original results of Erdős and Rényi [55] to the retained network, which follows the same statistics as the ER network but has new average degree $q\langle d \rangle$ and size qN.

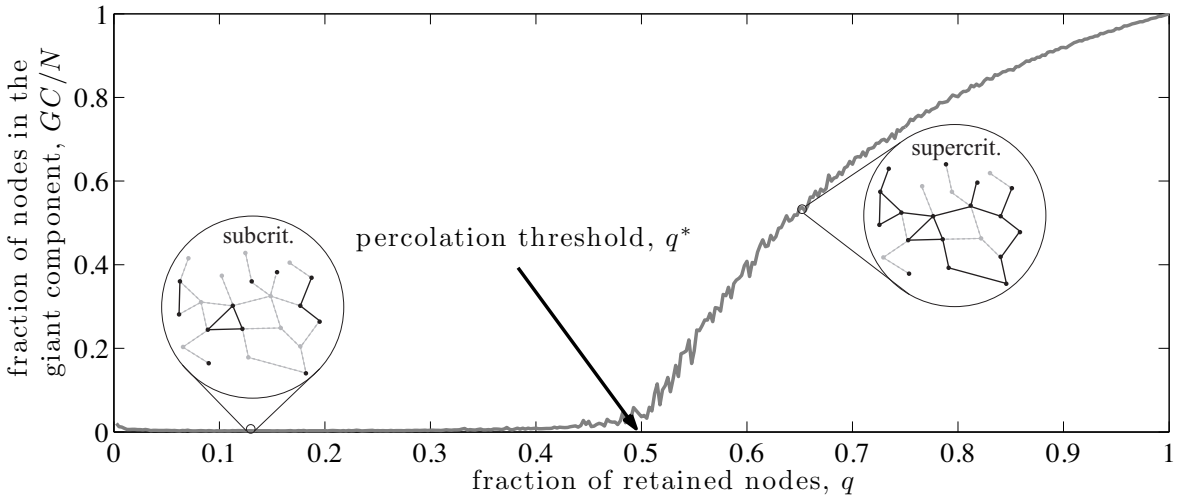


Figure 2.2: A continuous phase transition observed through the order parameter GC (the size of the giant component) occurs at the percolation threshold q^* (the critical fraction of nodes removed that separates the subcritical and supercritical regimes).

So far we have introduced the percolation threshold for a fragmentation processes for undirected networks in which nodes are removed with a uniform probability p, a process referred to as unweighted node percolation [126]. However, the percolation threshold is a general measure of robustness that is relevant to many network fragmentation processes. This includes fragmentation of networks in which nodes and/or links are removed at uniform probabilities (i.e., unweighted percolation [31, 46]), biased probabilities (i.e., weighted percolation [126]), or deterministic removal processes such as strategic attacks [129]. It follows that the percolation threshold q^* allows one to compare the robustness of different networks, as well as their respective robustnesses against various processes of fragmentation. Specifically, for a given network subject to a specific fragmentation process, small q^* indicates robustness whereas large q^* indicates fragility.

As shown in Fig. 2.3, the concept of a percolation threshold also generalizes to networks with directed links. However, several types of giant components must be considered. Specifically, one can study the giant strongly-connected component (GSCC, defined as the largest subgraph Ω_{GSCC} such that there exists a path between all ordered pairs $(n, m) \in \Omega_{GSCC} \times \Omega_{GSCC}$), the giant in-component (GIC, defined as the largest subgraph for which all nodes $n \in \Omega_{GIC}$ have a directed path to at least one node in the GSCC), the giant out-component (GOC, defined as the largest subgraph for which there is a path from at least one node in the GSCC to a node in Ω_{GOC}), and the giant weakly-connected component GWCC, defined as the union of the GIC, GOC, and GSCC). Defined in this way, all four giant components simultaneously appear at a single percolation threshold for directed networks [1] [20].

[1] Note that our definition of the GWCC is slightly different than that of Ref. [20]. Specifically, if one instead defines the GWCC as the largest cluster neglecting link direction, then the GWCC emerges at the percolation threshold of the corresponding undirected network rather than the percolation threshold of the directed network.

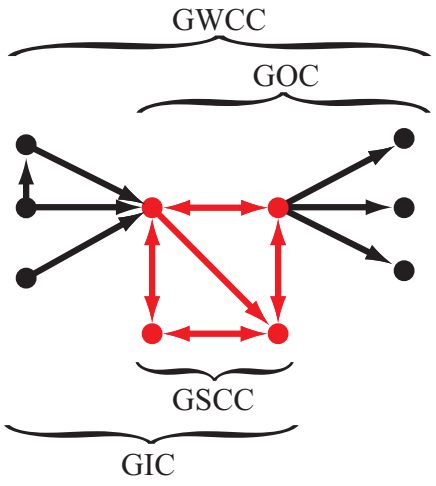


Figure 2.3: The concept of a cluster generalizes for directed networks. Restricting our attention to the largest cluster, or giant component, it consists of a giant strongly-connected component (GSCC), giant in-component (GIC), giant out-component (GOC), and giant weakly-connected component (GWCC).

Before introducing our theory studying the size of clusters and q^* for weighted percolation in networks with bidirectional links, we briefly summarize several major findings regarding q^* for complex networks:

- (i) Scale-free networks (i.e., those with a power-law degree distribution, $P(d) \propto d^{-\gamma}$ for $\gamma > 0$) can lack a percolation threshold for unweighted percolation. Specifically, it has been shown that $q^* \to 0$ for undirected and directed networks in the asymptotic limit $N \to \infty$ [31, 46].

- (ii) Scale-free networks are robust to unweighted percolation yet are fragile to targeted, deterministic attacks. Specifically, although they may lack an unweighted percolation threshold (i.e., $q^* \to 0$), the existence of nodes with very large degree (i.e., hubs) makes them susceptible to deterministic removal of these important nodes [11].

- (iii) Assortative mixing by degree (i.e., the tendency of neighboring nodes to have similar degrees) increases the robustness of networks by decreasing the unweighted percolation

threshold [104].

- (iv) Networks that are robust to deterministic removals targeting hubs often have an "onion structure" [129]. This "onion structure" can be characterized as *very strong assortativity* in which the hubs only link to other hubs and nodes with small degree only link to other nodes with small degree [141].

In the analysis to follow, we will show that these important results describing the robustness of networks to probabilistic and deterministic removals may be understood more generally under the framework of spectral graph theory, and in particular, through the dominant eigenspace of the network adjacency matrix A.

2.2 Weighted Percolation in Complex Networks

Several studies have proposed techniques to estimate the percolation threshold of a network for various situations [20, 31, 46, 57, 104, 107, 126]. These studies typically use *ensemble approaches*, where one studies the typical behavior of a set of networks satisfying some set of properties. Common ensembles include: (i) networks with fixed degree sequence generated from, for example, the configuration model [17, 18]; (ii) networks with an expected degree sequence generated with, for example, the Chung-Lu model [38]; and (iii) *Markovian* networks with correlations between nearest neighbors, where correlations may be captured by the probability $P(\mathbf{d}'|\mathbf{d})$ that a node with degree $\mathbf{d}$ is connected to a node of degree $\mathbf{d}'$. Here $\mathbf{d} = (d^{in}, d^{out})$ denotes the the number of incoming (d^{in}) and outgoing (d^{out}) links at a given node. While significant progress has been made in the study of such ensembles [20, 31, 46, 57, 104, 107], we note two important limitations of the ensemble approach:

- Given a single network, it is not clear what ensemble should be selected to capture the properties of the network. Typically, networks found in applications contain various structural properties (e.g., correlations [104], clustering [63, 157], and community structure [45, 58]) that are not always accounted for in the ensembles. A related problem is that it has been

recently observed [32] that, given an ensemble of networks, some network properties can vary significantly within the ensemble. Thus, it is not clear that ensemble approaches give the best description of a single network found in practice.

- Some ensemble theories are impractical when applied to individual real networks. For example, theories to estimate the percolation threshold in Markovian ensembles [20] require the estimation of a potentially dense and very large, $(d^{in}_{max}d^{out}_{max}) \times (d^{in}_{max}d^{out}_{max})$ matrix $P(\mathbf{d}|\mathbf{d}')$, which is difficult to obtain from a single network.

Therefore, in this section we develop a network-specific theory for weighted percolation in complex networks with bidirectional links. The majority of this work was recently published in Ref. [145]. In contrast to previous approaches to network-specific estimates of the percolation threshold, which were restricted to purely directed networks [126], our work has a broader range of applicability as it can be applied to directed networks with bidirectional links and undirected networks. Our method is based on an analysis of the network's adjacency matrix, which is often known or can be estimated in important applications (e.g., the power grid [148] and air transportation networks [16]). Besides relaxing the ensemble assumptions of previous research (e.g., that the network is strictly Markovian), one significant advantage of this approach is that it can easily account for arbitrary strategies of node/link removal. Network-specific approaches are therefore well suited for developing network-specific attack/defense strategies, immunization techniques, etc. In addition to estimating the percolation threshold, we predict the expected number of nodes accessible to each node after the network disintegrates. This has various applications such as predicting the outbreak size of an epidemic [4]. We finally show that our method may be used to study the fragmentation of a network subject to either probabilistic or deterministic attack.

2.2.1 Analysis

We formalize weighted percolation (i.e., in which nodes and/or links are retained with arbitrary probabilities) as follows: for a network with N nodes described by a possibly asymmetric

adjacency matrix A ($A_{nm} = 1$ if a link exists from node n to node m and $A_{nm} = 0$ otherwise), node n is retained with independent probability q_n, and the directed link from node n to node m is retained with probability p_{nm}. Note that these probabilities are independent. Letting $q = N^{-1} \sum_n q_n$ denote the *average node retention probability*, qN nodes are expected to remain after a realization of this process (referred to as a *percolation trial*). Unweighted node percolation corresponds to $q_n = q < 1$ and $p_{nm} = 1$, while unweighted link percolation corresponds to $q_n = 1$ and $p_{nm} = p < 1$. For our analysis, it is useful to introduce a matrix $\hat{A}$ with entries defined by $q\hat{A}_{nm} = q_m p_{nm} A_{nm}$, which represents the probability that a link exists from node n to node m, given that node n is retained. Because our analysis depends only on the matrix $\hat{A}$, it is applicable to link, node, and mixed (i.e., simultaneous link and node) percolation. However, for the remainder of this chapter we consider only node percolation, $p_{nm} = 1$.

For a given node-targeting strategy, defined as a set of retention probabilities $\{q_n\}$, we are interested in the size GC of the *largest strongly-connected component* (LSCC), the largest subset of nodes so that any node in the subset is reachable from any other node in the subset. The percolation threshold q^* is defined as the value of q such that $GC \ll N$ for $q < q^*$ (the *subcritical regime*) and $GC \sim N$ for $q > q^*$ (the *supercritical regime*). Note that for simultaneous removals of nodes and links the percolation threshold would need to be defined using averages values for both q_n and p_{nm}. For a particular removal strategy, the subcritical regime may be analyzed by noting that after a percolation trial, only a fraction of the network is reachable from a given node n following directed links. As discussed in the previous section, because the emergence of all giant components is simultaneous, we will develop analysis studying the largest out component. Following Ref. [107] we define, for a given percolation trial, the *out-component* of node n as the set of nodes that may be reached from node n via the remaining network (including node n) and define s_n^{out} as the size of the out-component of node n averaged over many percolation trials.

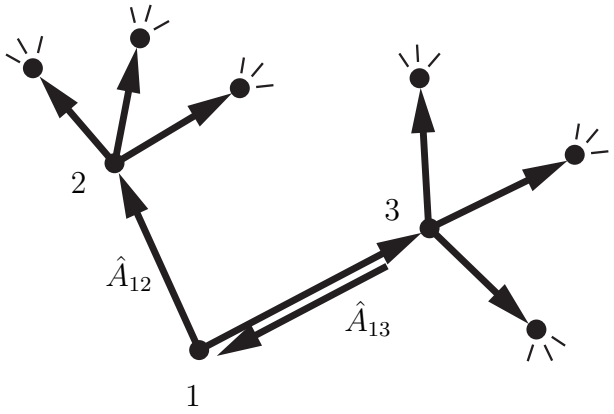


Figure 2.4: When computing s_1^{out}, to compensate for the over-counting of nodes due to the bidirectional link $1 \leftrightarrow 3$, β_{13} reduces the contribution of s_3^{out} on s_1^{out}. We approximately have $\beta_{13} \sim 3/4$ and $s_1^{out} = 1 + q\hat{A}_{12}s_2^{out} + q\hat{A}_{13}\beta_{13}s_3^{out}$.

To motivate subsequent analysis, consider first the case when the network is a directed tree (i.e., the network is a perfect branching structure in which there are no loops and the in-degree of each node is either 1 or 0). In that case s_n^{out} satisfies the relation $s_n^{out} = 1 + \sum_m q\hat{A}_{nm}s_m^{out}$, where the right hand side counts the nodes reachable from node n by counting the nodes reachable from its neighbors, and adds 1 to account for node n itself. The same expression approximately applies to directed networks that are locally tree-like (e.g., networks with few short loops [90]) and leads to the results in Ref. [126], $q^* \approx \hat{\lambda}^{-1}$, where $\hat{\lambda}$ is the principal, or Perron-Frobenius, eigenvalue of $\hat{A}$ (i.e., $\hat{A}\hat{\mathbf{u}} = \hat{\lambda}\hat{\mathbf{u}}$). When links are allowed to be bidirectional, however, the expression above overestimates the size of s_n^{out}, since the terms s_m^{out} on the right hand side might include nodes that are reachable by following links back into node n (see Fig. 2.4).

To correct for this over-counting of nodes, we heuristically modify the contribution of s_m^{out} on the right hand side by a factor β_{nm} (to be determined),

$$s_n^{out} = 1 + \sum_m q\hat{A}_{nm}\beta_{nm}s_m^{out}. \tag{2.2}$$

To determine a self-consistent expression for β_{nm}, we note that from Eq. (2.2) the relative contribution of the out-component of node m on s_n^{out} is $q\hat{A}_{nm}\beta_{nm}s_m^{out}/s_n^{out}$. Therefore, to reduce the contribution of s_m^{out} on s_n^{out} to account for the branch returning to node n (if present), we let

$\beta_{nm} = 1 - q\hat{A}_{mn}\beta_{mn}s_n^{out}/s_m^{out}$. Inserting here the corresponding expression for β_{mn} we obtain

$$\beta_{nm} = 1 - q\hat{A}_{mn}\left(1 - q\hat{A}_{nm}\beta_{nm}\frac{s_m^{out}}{s_n^{out}}\right)\frac{s_n^{out}}{s_m^{out}}. \tag{2.3}$$

Finally, after solving for β_{nm}, we obtain

$$\beta_{nm} = (1 - q^2\hat{A}_{nm}\hat{A}_{mn})^{-1}\left(1 - \frac{q\hat{A}_{mn}s_n^{out}}{s_m^{out}}\right). \tag{2.4}$$

After substitution of Eq. (2.4) into Eq. (2.2), we find

$$\mathbf{s}^{out} = [I - D(q)]^{-1}\mathbf{y}, \tag{2.5}$$

where $\mathbf{s}^{out} = [s_1^{out}, ..., s_N^{out}]^T$, I is an identity matrix of size N, $\mathbf{y}$ is a vector with entries

$$y_n = \left[1 + q^2\sum_k \hat{A}_{nk}\hat{A}_{kn}(1 - q^2\hat{A}_{nk}\hat{A}_{kn})^{-1}\right]^{-1}, \tag{2.6}$$

and $D(q)$ is a matrix with entries

$$D_{nm}(q) = q\hat{A}_{nm}y_n(1 - q^2\hat{A}_{nm}\hat{A}_{mn})^{-1}. \tag{2.7}$$

Given a removal strategy, Eq. (2.5) can be solved to obtain the expected out-component size for each node. To obtain an estimate for the percolation threshold, note that Eq. (2.5) requires the invertibility of the matrix $I - D(q)$. A sufficient condition for the invertibility of this matrix is that $\lambda_{D(q)} < 1$, where $\lambda_{D(q)}$ is the principal eigenvalue of $D(q)$. As $\lambda_{D(q)} \to 1^-$ the out-component sizes diverge as $s_n^{out} \sim [1 - \lambda_{D(q)}]^{-1}w_n$, where $\mathbf{w}$ is the principal eigenvector of $D(q)$. A similar argument can be made for the divergence of the in-component sizes. Since the LSCC above the percolation threshold can be thought of as the set of vertices with infinite in- and out-components [20], we predict the percolation threshold as

$$q_D^* = \min_{q\in[0,1]}\{q : \lambda_{D(q)} = 1\}. \tag{2.8}$$

We note that if there are no bidirectional links, $\hat{A}_{nm}\hat{A}_{mn} = 0$ and $D(q) = q\hat{A}$, and the results of Ref. [126] are recovered.

While one may solve Eq. (2.8) numerically, it is both practical and insightful to approximate Eqs. (2.4)-(2.8) for large s_n^{out} and small q, a situation that is expected for large networks with heavy-tailed degree distributions. Letting $s_n^{out} \gg 1$ and $\beta_{nm} \sim 1$ in Eq. (2.2) yields the approximate eigenvalue problem $s_n^{out} \approx q \sum_m \hat{A}_{nm} s_m^{out}$. It follows that $\mathbf{s}^{out} \propto \hat{\mathbf{u}}$ and $q \sim \hat{\lambda}^{-1}$ under these conditions, yielding to first order $\beta_{nm} \approx 1 - \hat{\lambda}^{-1} \hat{A}_{mn} \hat{u}_n / \hat{u}_m$. Defining

$$C_{nm} = \hat{A}_{nm} \left(1 - \frac{\hat{A}_{mn} \hat{u}_n}{\hat{\lambda} \hat{u}_m} \right), \tag{2.9}$$

with principal eigenvalue equation $C\mathbf{z} = \lambda_C \mathbf{z}$ and using $\mathbf{y} \approx \mathbf{1} = [1, 1, \ldots, 1]^T$, we obtain the predictions

$$\mathbf{s}^{out} \approx (I - qC)^{-1} \mathbf{1}, \tag{2.10}$$

$$q_C^* \approx \lambda_C^{-1}. \tag{2.11}$$

In addition to offering simplified predictions for s_n^{out} and q^*, for unweighted percolation (i.e., $\hat{A} = A$ and $\hat{\mathbf{u}} = \mathbf{u}$) these estimates allow us to bound λ_C using the principal eigenvalue λ of the network adjacency matrix A (e.g., $A\mathbf{u} = \lambda \mathbf{u}$). Direct application of the Bauer-Fike Theorem [60] for the limiting case of an undirected network yields $|\lambda_C - \lambda| \leq ||\lambda^{-1} U A U^{-1}||_2 = 1$, where $U = \text{diag}[u_1, \ldots, u_N]$. Finally, considering $\mathbf{1}^T C \mathbf{u}$ and using $\mathbf{z} \sim \mathbf{u}$ yields $\lambda_C \approx \lambda - 1$ for unweighted percolation on undirected networks. One implication of these results is that $q^* \to 0$ for large λ, which is consistent with the lack of an unweighted percolation threshold for well-connected networks such as scale-free networks [31, 46]. In addition, because increasing or decreasing the assortativity of a network respectively increases or decreasing the largest eigenvalue of the network adjacency matrix λ [27, 125], our results agree well with [104] by indicating that an increase or decrease in assortativity is expected to increase or decrease the network's robustness.

Before continuing with numerical experiments, we note that while our heuristic approach compensates for the overcounting of nodes caused by bidirectional links, it does not account for higher order effects caused by loops of size larger than 2. However, for the large class of locally tree-like networks [90], we have found that these effects are small compared to the effect of bidirectional

links. We finally note that Eqs. (2.10) and (2.11) are expected to be in best agreement with Eqs. (2.5) and (2.8) near $q = q^*$ and when the network is strictly undirected or strictly directed.

2.2.2 Numerical Experimentation

We now provide results for several experiments to illustrate the accuracy of our theory for (i) unweighted percolation on synthetic networks, (ii) weighted percolation on synthetic networks, and (iii) weighted percolation on real-world networks. In these respective sections we have chosen to include experiments to (i) motivate the need for our theory offering network-specific analysis for the robustness of networks with bidirectional links, (ii) highlight that our results are robust to various removal strategies and complex network properties, and (iii) highlight that our predictions are accurate for real-world networks.

However, an inherent problem naturally arises when testing percolation theory for a single, given network: although one can clearly deduce the percolation threshold q^* for a class, or ensemble, of networks by considering their scaling properties as N increases, q^* is not well defined for a single network with fixed size. We overcome this obstacle by using extrapolation to numerically infer a network's percolation threshold q^* from a $GC(q)$ curve. An example of this extrapolation is shown in Fig. 2.5. For the numerical experiments to follow, experimental values for q^* were found using extrapolation over 6 overlapping intervals of width 0.03 that span $[0.03, 0.13]$ to find the intercept with the horizontal axis. This yielded a mean and standard deviation (which is only shown when significant). We finally note that while q^* can sometimes be extracted by examining the size of the second-largest cluster [22], this approach was observed to significantly overestimate q^* for the relatively small networks considered here, which contain $N < 10^5$ nodes. This alternative method is primarily used for studying the asymptotic limit $N \to \infty$ for an ensemble of networks [22].

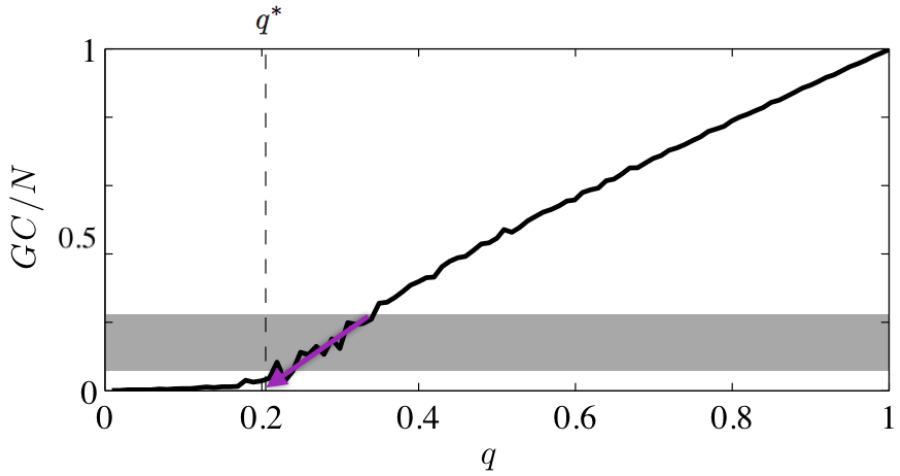


Figure 2.5: Consider a $GC(q)$ curve for a given network subject to a given attack strategy $\{q_n/q\}$. We numerically extract the percolation threshold q^* using extrapolation. Specifically, we consider small values $GC(q)$ within some range (e.g., shaded area) and extrapolate the intercept with the horizontal axis for these points. In practice, several such ranges are considered leading to a mean and variance for q^*.

2.2.2.1 Unweighted Percolation in Synthetic Networks

Our first numerical experiments explore unweighted (i.e. $q_n = q$) node percolation in synthetic, or user-generated, networks. These experiments are designed to (i) highlight the need for a network-specific percolation theory for networks with bidirectional links, (ii) show that our results are in good agreement with previous research, and (iii) illustrate the accuracy of our results for specific types of structural complexities, including degree-degree correlations and clustering.

Network-Specific vs. Ensemble-Based Theory

We begin with an example highlighting the need for a network-specific method for undirected networks with bidirectional links and show that, unlike the ensemble approach, a network-specific method captures variability in q^* across an ensemble. We consider percolation in an uncorrelated, random network formed by retaining the giant component from an Erdős-Rényi process [55] (henceforth referred to as an ER network) with $N = 10^4$ nodes and $3N$ links.

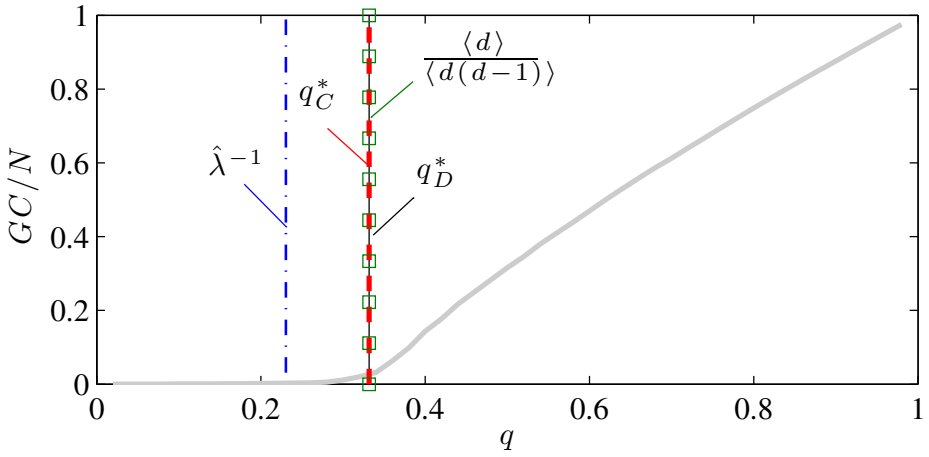


Figure 2.6: The relative size of the giant component, GC/N (solid grey line), is shown for unweighted percolation on an ER network [55] (see text). Our predictions to q^* given by Eq. (2.8) and Eq. (2.11) are given by vertical solid black and dashed red lines and are in good agreement with the undirected ensemble result [107] (green squares). As expected for this undirected network, the network-specific prediction for directed networks, $\hat{\lambda}^{-1}$[126] (blue dot-dashed line), is shown to be inaccurate.

In Fig. 2.6 we show the average fraction of retained nodes in the LSCC, GC/N, as a function of q (solid grey curve). Our predictions for q^* given by Eq. (2.8) (black line) and Eq. (2.11) (red dashed line) and the undirected ensemble theory [107] (green squares) work well, whereas the network-specific theory for directed networks, $\hat{\lambda}^{-1}$ [126] (blue dot-dashed line), does not, as expected since the network is undirected.

In Fig. 2.7 we show numerically observed values for the percolation threshold, q^*, versus our prediction using Eq. (2.8) for an ensemble of uncorrelated networks obtained by rewiring the ER network while retaining a fixed degree sequence, $[d_1, d_2, \ldots, d_N]$. The rewiring process is similar to that in Refs. [104, 121], except an additional step is taken to ensure the resulting network contains all nodes in its LSCC. Note that whereas our network-specific method naturally accounts for variability in q^* across the ensemble, the ensemble approach (horizontal blue line) cannot as it predicts the percolation threshold $q^* \approx 0.33$ for all members of the ensemble.

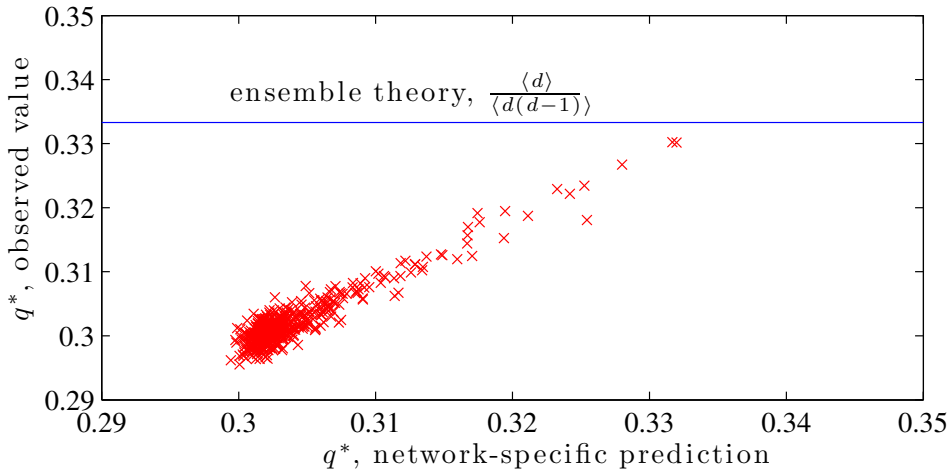


Figure 2.7: Numerically observed values of the percolation threshold, q^*, are shown for many realizations of unweighted percolation in many ER networks with $N = 10^4$ nodes and $3N$ undirected links (red crosses). Whereas variation in q^* is naturally accounted for by our network-specific approach, Eq. (2.8), the ensemble approach offers a single approximation to q^* for all networks within the ensemble (horizontal blue line).

Recovering Previous Results

Our next numerical test illustrates that our analysis for networks with bidirectional links is in good agreement with previous results for unweighted percolation in networks that contain either directed links, undirected links, or a combination of both. Therefore, in the following experiment we consider an undirected ER network and iteratively replace undirected links with directed links until the entire network is directed. Results are shown in Fig. 2.8. However, we point out that although our results are in good agreement with Ref. [20] for unweighted percolation, Ref. [20] is not applicable to networks with degree-degree correlations or to networks undergoing weighted percolation.

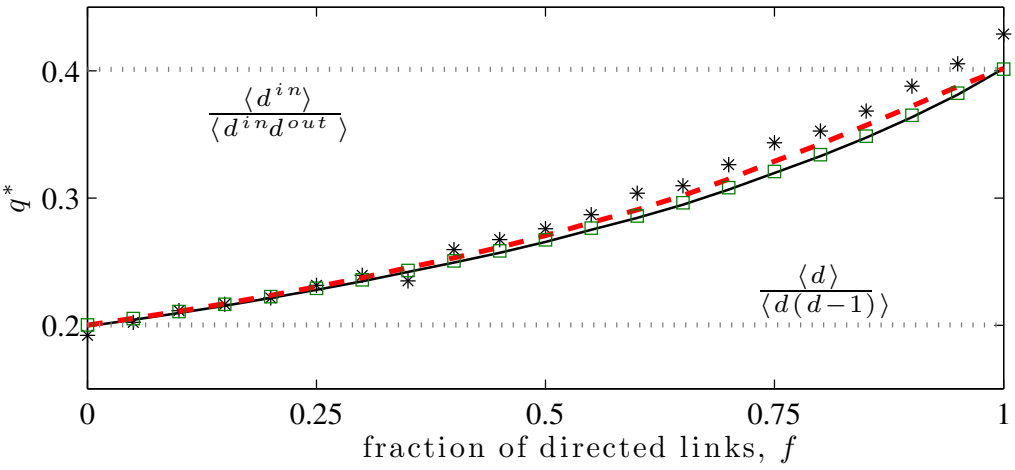


Figure 2.8: Our results are in good agreement with previous results for unweighted node percolation in a random network that is rewired from being strictly undirected ($f = 0$) to strictly directed ($f = 1$). Specifically, Eq. (2.8) (black solid line) and Eq. (2.11) (red dashed line) agree well with Ref. [20] (green squares) for variable f. For $f = 0$, our result is in good agreement with previous results for undirected networks (lower horizontal line [31, 46]) and for $f = 1$ our theory is in good agreement with previous results for directed networks (upper horizontal line [107]). Stars indicated numerically observed values of q^* via extrapolation.

Letting f denote the fraction of directed links, we begin with an undirected (i.e., $f = 0$) ER network with $N = 10^4$ nodes and 5N links and iteratively replace randomly chosen undirected links with directed links in a random orientation until the network is strictly directed (i.e., $f = 1$). One can observe in Fig. 2.8 that our predictions by Eqs. (2.8) and (2.11) agree well with the ensemble result [20] and experimental values. Note that the predictions of Eq. (2.11) agree with those of Eq. (2.8) very well for $f = 0$ or 1 and that maximal disagreement occurs near $f \sim 0.7$. We note that one disadvantage of extrapolating on a fixed range (e.g., $GC/N \in [0.03, 0.13]$) while varying q^* is that a drift occurs due to the combination of finite-size affects and the fact that the slope of an $GC(q)$ curve just above q^* varies monotonically with q^*. This is one cause for the disagreement

between theory and experiment in Fig. 2.8 for large f.

We now recover the ensemble (i.e., mean field) predictions for strictly undirected and strictly directed networks shown by horizontal lines in Fig. 2.8. For $f = 0$, our result $\lambda_C \approx \lambda - 1$ and the undirected mean-field (MF) result $\lambda \approx \langle d^2 \rangle / \langle d \rangle$ lead to $q^* \approx \langle d \rangle / \langle d(d-1) \rangle$ [31, 46]. For $f = 1$ we have $A_{nm}A_{mn} = 0$ and $C = A$, which recovers $q^* \approx \lambda^{-1}$ [126]. Again, the MF result $\lambda^{-1} \approx \langle d^{in} \rangle / \langle d^{in} d^{out} \rangle$ [126] recovers the result of Ref. [107]. We further note that while these MF results have been shown to be asymptotically accurate as $N \to \infty$, for finite sized networks they describe the expected behavior of a network drawn from an ensemble. Such an ensemble may include, for example, all networks with a given degree sequence $\{d_n\}_{n=1}^N$ (e.g., as obtained by the configuration model [17, 18]) or all networks described by an expected degree distribution $P(d)$ (e.g., as obtained by the Chung-Lu method [38]).

Networks with Degree-Degree Correlations

As mentioned in Sec. 2.1, the percolation threshold q^* of a network can depend strongly on whether or not degree-degree correlations are present. Specifically, the percolation threshold has been shown to increase for networks with positive correlations (i.e., assortative networks) and decrease for networks with negative correlations (i.e., disassortative networks) [104]. For both cases, as the correlations strengthen, the change in q^* increases. It follows that positive correlations increase robustness while negative correlations decrease robustness (with respect to unweighted and random removals). However, this previous result is restricted to describing *Markovian* correlations between neighboring nodes. In the following example we confirm this result and show that our theory also applies to *non-Markovian* correlations existing between non-neighboring nodes. Specifically, we investigate correlations between nodes that are separated by paths of length two.

In Fig. 2.9 we plot q^* for undirected ER networks that are rewired to have assortative and disassortative degree-degree correlations [104]. These Markovian correlations existing between neighboring nodes may be measured by the assortativity coefficient r, given by Eq. (1.1). As shown in Fig. 2.9(a), as expected, both the ensemble theory [104] (blue squares) and our result, Eq. (2.8)

(black line), accurately predict numerically observed values for the percolation threshold q^* (black stars). Note that q^* respectively increases and decreases as r is increased and decreased. Degree correlations were varied while keeping the degree distribution constant following the algorithm in Refs. [104, 121].

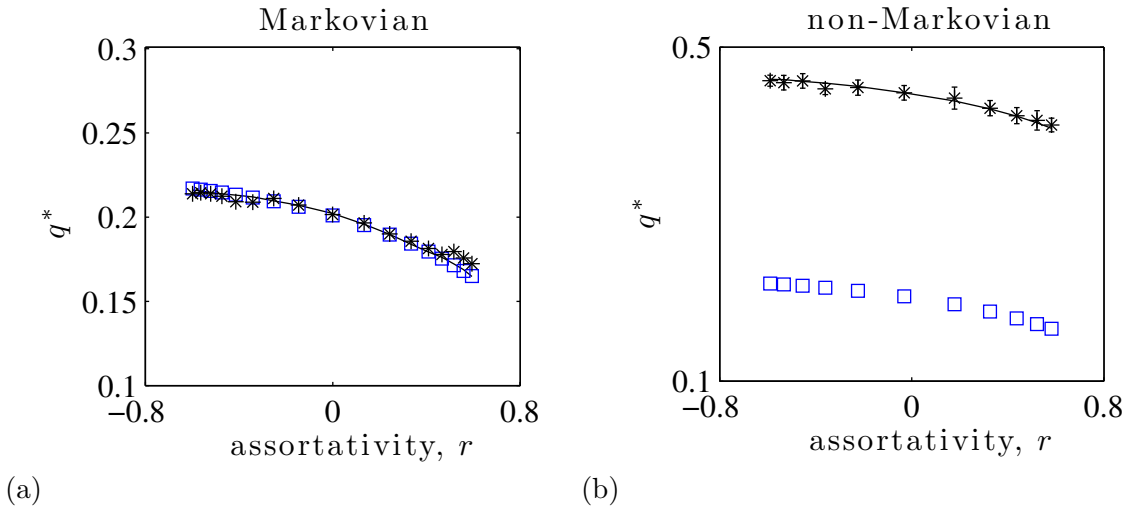


Figure 2.9: The percolation threshold q^* is accurately predicted by our theory, Eq. (2.8) (solid black line), for undirected networks with correlations between the nodal degrees, as measured by the assortativity coefficient r, given by Eq. (1.1). Star symbols indicate numerically observed values, whereas the blue squares indicate the Markovian ensemble theory for undirected networks [104]. (a) Both Eq. (2.8) and the ensemble theory [104] offer excellent approximations to q^* for networks with Markovian correlations. (b) While Eq. (2.8) also accurately describes q^* for networks with non-Markovian correlations, the Markovian ensemble theory [104] is not applicable.

Turning to degree-correlations of the non-Markovian type in undirected networks, a case which no previous theories can handle, we consider the set of networks used to produce Fig. 2.12(a) but subject them to the following rewiring process: each link $n \leftrightarrow m$ is replaced by two new links and a new node j, $n \leftrightarrow j \leftrightarrow m$, resulting in correlations across paths of length two. In Fig. 2.12(b),

whose horizontal axis is carried over from Fig. 2.12(a), we show that the network-specific prediction Eq. (2.8) (solid line) agrees with the observed values of q^* (stars) for unweighted percolation on these non-Markovian networks. For comparison, direct application of the Markovian ensemble method [104] (squares) does not give good results, as expected.

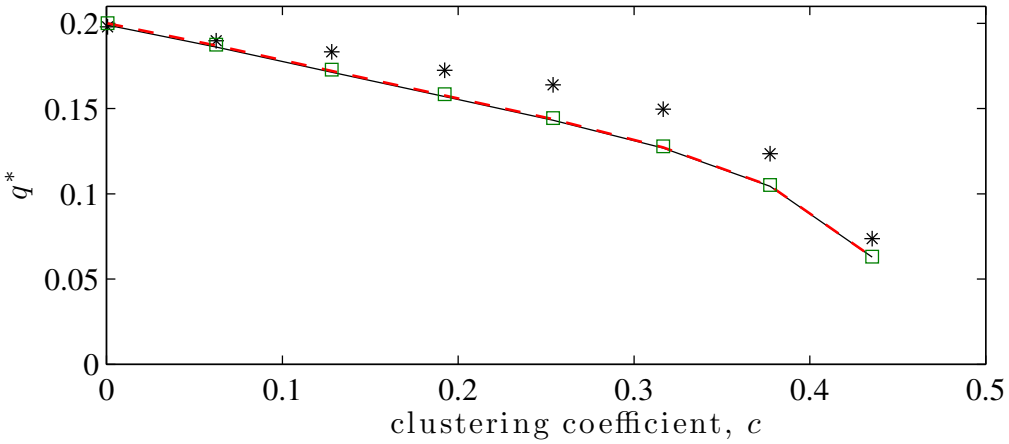


Figure 2.10: Predictions to the percolation threshold q^* and experimental values (stars) for unweighted percolation on an ER network with $N = 10^4$ nodes and $5N$ links are shown as a function of variable clustering coefficient, c. Note that the lack of degree correlations yields good agreement between our results, Eq. (2.8) (solid black line), Eq. (2.11) (dashed red line), and the ensemble result for undirected networks [107] (green squares).

Networks with Clustering

We now show that our results remain accurate for relatively large clustering coefficient, c [63, 157]. In this experiment links were iteratively added to an ER network by finding paths of length two and completing the triangles. In Fig. 2.10 we show our predictions for and the observed value of q^* as a function of c. The solid black and dashed red lines respectively indicate Eq. (2.8) and Eq. (2.11), while the black stars indicate numerically observed values using extrapolation. Note

that because these networks lack degree-degree correlations, the ensemble result for undirected networks [107] (green squares) is in good agreement with our predictions.

As a final note, we point out that the analysis we have presented in the previous section may be interpreted as a first-order correction to the existence of loops of size two. Generalizing our methods to higher-order corrections [i.e., loops of size $k > 2$] remains an open direction of research. Clustering corresponds to loops of size three and is thus expected to cause error if it is sufficiently prevalent. However, our results indicate that this higher-order effect, which is $O(q^3)$ for retention probability q, may in some cases not be significant even with moderate clustering, $c \sim 0.2$.

2.2.2.2 Weighted Percolation on Synthetic Networks

Having shown the accuracy of our theory for unweighted percolation in synthetic networks, we now study weighted percolation in synthetic networks. Specifically, we will consider networks in which each node n is retained with probability based on its degree,

$$q_n \propto d_n^l, \text{ for } n = 1, 2, \ldots, N. \tag{2.12}$$

Here $d_n = (d_n^{in} + d_n^{out})/2$ denotes the degree of node n averaged over incoming and outgoing links. Preferentially removing nodes with large degrees (i.e., $l < 0$) can model probabilistic attacks [57] and biased infrastructure failure [98, 95, 96], whereas preferentially removing nodes with small degrees ($l > 0$) may represent a non-obtrusive degradation. We note that Ref. [57] provides an ensemble approach for a similar removal strategy. However, their analysis is restricted only to scale-free networks lacking degree-degree correlations and cannot handle general forms of weighted percolation.

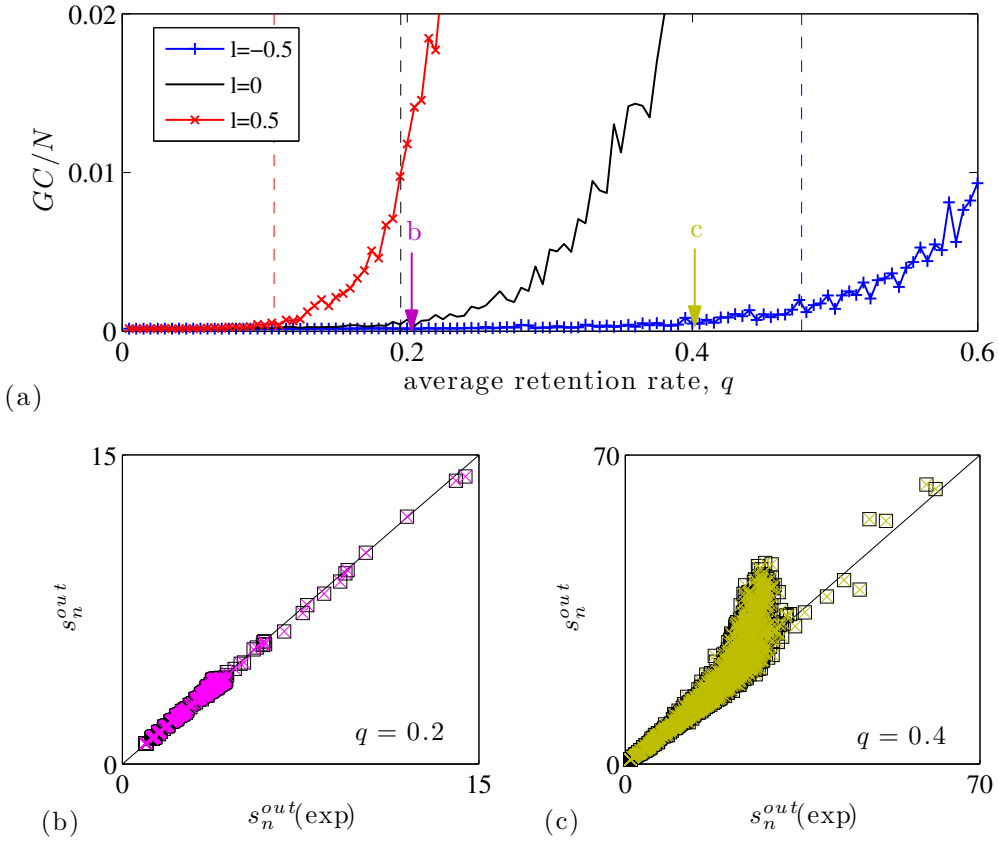


Figure 2.11: (a) Weighted percolation on a random network with $N = 5000$ nodes and $5N$ undirected links. Nodes are removed with probability $q_n \propto d_n^l$ for $l = 0.5$ (red line with x's), $l = 0$ (solid black line), and $l = -0.5$ (blue line with crosses). For each l, a $GC(q)$ curve is shown to grow after q surpasses the respective percolation threshold q^*, as predicted by Eq. (2.8) (vertical dashed lines). (b-c) For $l = -0.5$, predicted values for the out components given by Eq. (2.5) (squares) and Eq. (2.10) (x symbols) are shown to be in good agreement with experimentally observed values s_n^{out}(exp) for $q = 0.2$ (b) and $q = 4$ (c). Note in (c) that the predicted values start to diverge, which is expected as q approaches q^*.

Cluster Size Distributions

We now confirm the accuracy of Eqs. (2.5) and (2.10) which predict the size of the out-components for the nodes in networks undergoing weighted percolation. In this example s_n^{out} can

be interpreted, for example, as the expected number of nodes infected by a computer virus released by user n and preferentially targeting nodes with large degree [4]. Results are shown for the ER network with $N = 5000$ nodes and $5N$ links. In Fig. 2.11(a) we show the size of the giant component GC for variable average retention rate q for three values of l, $l \in \{-0.5, 0, 0.5\}$. Note that compared to unweighted percolation ($l = 0$), preferentially removing nodes with large degree ($l = -0.5$) or small degree ($l = 0.5$) respectively increases or decreases the percolation threshold q^*. As expected, when q surpasses q^* given by Eq. (2.8) (vertical dashed lines) a giant component forms for each l value. In Figs. 2.11(b) and 2.11(c) we show that Eq. (2.5) (squares) and Eq. (2.10) (x symbols) are in good agreement for experimentally observed values $s_n^{out}(\text{exp})$. Figs. 2.11(b) and 2.11(c) show s_n^{out} for $l = -0.5$ and $q = 0.2$ and $q = 0.4$, as indicated by arrows in Figs. 2.11(a). Note that deviation between theory and experiment is expected in Fig. 2.11(c) since s_n^{out} is predicted to diverge at $q^* \sim 0.49$, but experimental values are bounded by the finite network size N, so the predicted value must become larger than the observed value as $q \to q^*$.

Networks with Degree-Degree Correlations

In the next example we demonstrate the robustness of our results for weighted percolation (e.g., $q_n \propto d_n^l$) on complex networks with degree correlations. Again we first consider Markovian degree-degree correlations between the neighboring nodes [104], which can be characterized by the assortativity coefficient $r \in [-1, 1]$ [see Eq. (1.1)]. Recall that $r > 0$ and $r < 0$ respectively indicates that nodes tend to connect to other nodes with similar and different degrees.

In Fig. 2.12(a) we show the effect of degree correlations on q^* for an ER network with $N = 10^4$ nodes and $5N$ undirected links that is rewired to have correlations. Values obtained from numerical experiments are shown for $l = -1$ (gold x's), $l = 0$ (blue stars), and $l = 1$ (violet crosses). Solid lines indicate the prediction of Eq. (2.8), which was found to coincide with that of Eq. (2.11). Degree correlations were varied while keeping the degree distribution constant following the algorithm in Refs. [104, 121]. Note that while assortativity promotes robustness for unweighted percolation ($l = 0$) by reducing q^*, we find that its effect is reduced (amplified) for $l < 0$ ($l > 0$). For example,

the percolation threshold is largely unaffected by degree correlations for $q_n \propto d_n^{-1}$ for this network [see x symbols in Fig. 2.12(a)].

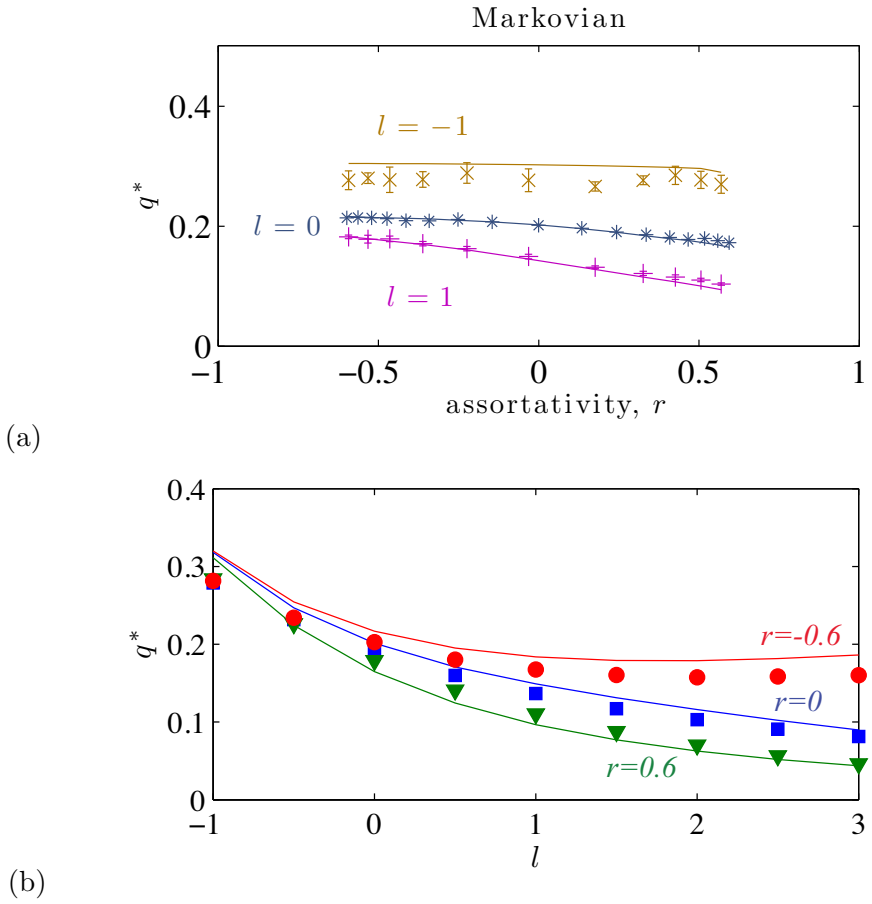


Figure 2.12: (a) Values of q^* predicted from Eq. (2.8) (solid lines) and observed in numerical experiments (symbols) for percolation on an ER network having $N = 10^4$ nodes and $5N$ undirected links rewired with correlations (measured by assortativity coefficient r). The three curves correspond to unweighted percolation ($l = 0$, middle) and weighted percolation ($l = -1$, top, and $l = 1$, bottom). (b) The percolation threshold is also shown for variable l for networks with three values of assortativity: $r = -0.6$ (red line and circles), $r = 0$ (blue line and squares), and $r = 0.6$ (green line and triangles). Note that the curves for all three lines appear to converge as l decreases.

This is further illustrated in Fig. 2.12(b), where we plot q^* as a function of l. Observed (symbols) and predicted (lines) values are shown for three choices of assortativity, $r \in \{-0.6, 0, 0.6\}$. Note that as l decreases all three curves appear to merge, indicating that assortativity may not affect network robustness as nodes with larger degree are preferentially removed. We note that disagreement between experiment and theory is exaggerated in Fig. 2.12(b) due to using a single extrapolation range for all experiments.

2.2.2.3 Weighted Percolation on Real-World Networks

We now illustrate our theory on real-world networks undergoing weighted node percolation with $q_n \propto d_n^l$. We will consider real-world networks for which previous methods are either impractical or not justified.

Our first example illustrates Eqs. (2.5) and (2.10) in a peer-to-peer (P2P) network of file downloads [127] with $N = 6301$ nodes for weighted node percolation with $q_n \propto d_n^{-0.5}$. In this example s_n^{out} can be interpreted as the expected number of nodes infected by a computer virus released by user n and preferentially targeting nodes with large degree [4]. In Fig. 2.13 we show a sample of the values of s_n^{out} predicted by Eq. (2.5) (x's) [which in this case agrees with Eq. (2.10)] versus their experimental values, s_n^{out} (exp). The prediction given by Eq. (2.5) (x's) is very accurate for the average removal probability $q = 0.2$ shown in Fig. 2.13(a). For the larger value $q = 0.4$, our prediction deviates somewhat from the observed values as shown in Fig. 2.13(b). This is expected because s_n^{out} is predicted to diverge at $q^* \sim 0.476$, but experimental values are bounded by the finite network size N.

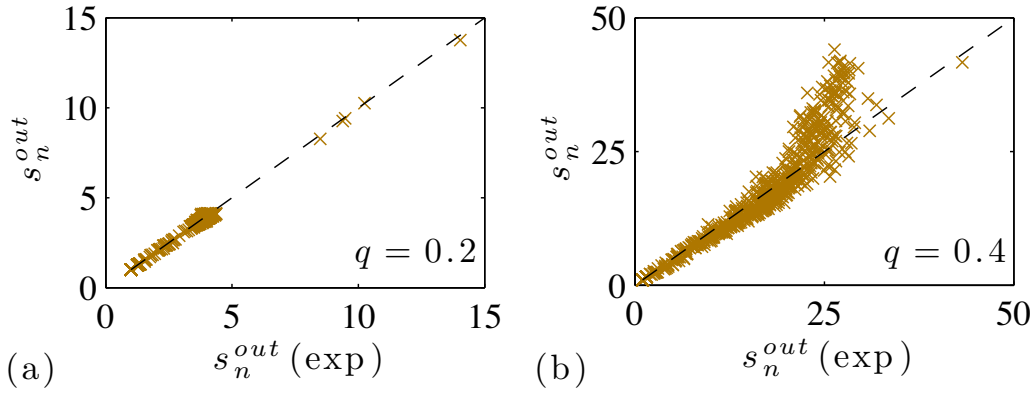


Figure 2.13: (a) For weighted percolation on a P2P network [127] with retention probability $q_n \propto d_n^{-0.5}$ and average retention rate $q = 0.2$, the values of s_n^{out} predicted by Eq. (2.5) (x's) agree well with experimental values, s_n^{out} (exp), which were averaged over 2^{16} percolation trials. (b) Similar results are shown for $q = 0.4$, where expected deviations appear (see text).

In Fig. 2.14(a) we show the relative size of the giant component, $\phi = GC/(qN)$, for weighted percolation on a directed Word-Association (WA) Network [103]. Vertical dashed lines and x's represent the predictions of Eq. (2.8) and Eq. (2.11). The inset shows this graph with the horizontal axis respectively normalized by the value of q^* found from Eq. (2.8) for each targeting strategy, where a giant component of size $GC \sim N$ appears at $q/q^* \sim 1$ for all curves.

In Fig. 2.14(b), we show the prediction by Eq. (2.8) (solid lines) and observed (symbols) values of q^* as a function of l, normalized by the value of q^* at $l = 0$, for the WA network [103] and an undirected network of Facebook (FB) friendships at Caltech [147]. Being a well-connected network with power-law degree distribution, the FB network has a very small unweighted percolation threshold [$q^*(0) \sim 0.01$, not shown]. However, preferentially removing nodes with large degree can yield nontrivial thresholds with nearly a ten-fold increase, $q^*(-1) \sim 0.1$.

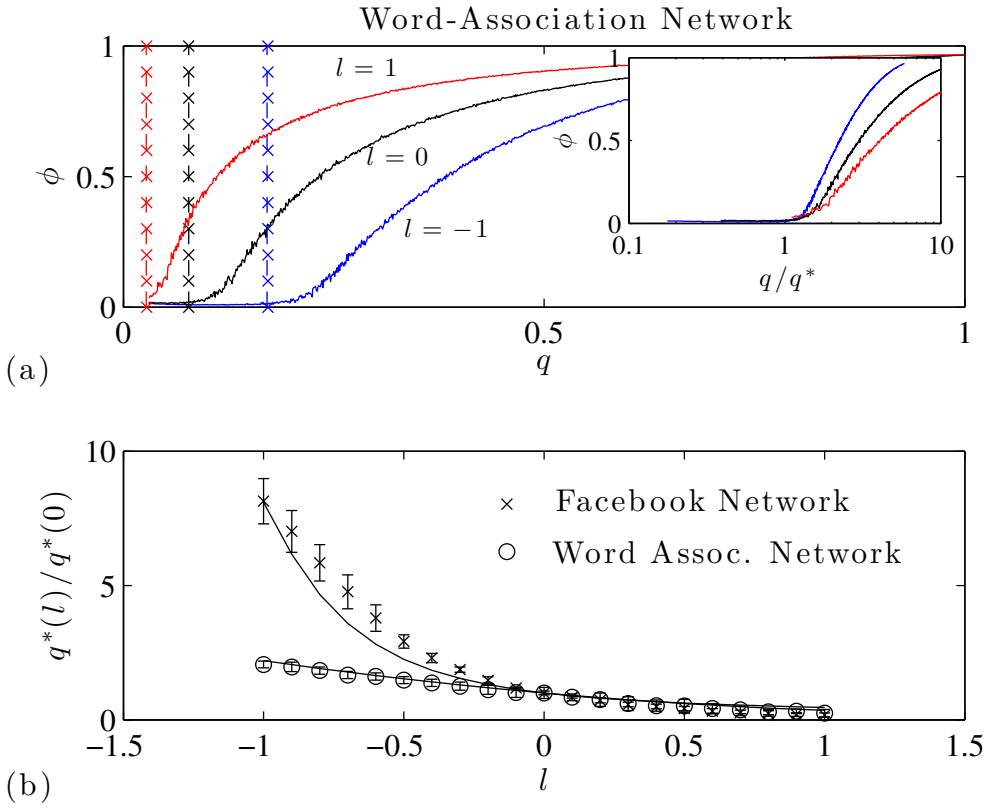


Figure 2.14: (a) Average values for $\phi = GC/(qN)$ over 32 trials are shown for weighted percolation with $q_n \propto d_n^l$ on the scale-free WA network [103]. Vertical dashed lines and x's represent the predictions of Eq. (2.8) and Eq. (2.11). The inset shows the same quantities with the q-axis normalized by Eq. (2.8) for each l value. (b) The dependency of q^* on variable l is shown for both the Word Association [103] (circles) and Facebook networks [147] (x symbols). Unweighted node percolation corresponds to $l = 0$. Targeting the probabilistic removal of nodes with large (i.e., $l < 0$) and small (i.e., $l > 0$) degrees respectively increases and decreases the percolation threshold q^*.

2.3 Deterministic Network Fragmentation

Before concluding, we illustrate the utility of our theory for studying networks subject to deterministic removal processes, such as malicious attacks. We consider three undirected networks, a network describing protein-protein interactions in yeast [28], a power grid [148], and an airline transit network [16]. Note that knowledge of how these networks fragment is essential for taking measures for protection. Simulated attacks were implemented by iteratively removing the node n corresponding to the largest dynamical importance, $\mathrm{DI}_n \propto u_n v_n$ [121], where $\mathbf{u}, \mathbf{v}$ are the right and left principal eigenvectors of A. For these undirected networks, $\mathrm{DI}_n = u_n^2$. After each node removal, components fragmented from the LSCC were also removed. In the upper panels of Fig. 2.15, we show the relative size GC/N of the giant component after the removal of k nodes (red lines), whereas in the lower panels we show both λ (dashed red lines) and λ_C (solid red lines), where C is given by Eq. (2.9). For comparison, blue lines indicate the the same variables for random node removal. Note that the disappearance of the giant component corresponds to $\lambda_C \approx 1$. For these undirected networks this corresponds to $\lambda \approx 2$, whereas for directed networks we recover the result of Ref. [126], $\lambda \approx 1$.

In light of recent research studying the fragmentation of networks subject to deterministic removals, our results offer several important insights. For example, Fig. 2.15 suggests that an effective approach for mitigating the damage incurred from deterministic attacks may be obtained by attempting to delay the k/N value at which λ_C decreases to 1 (or λ decreases to 2). Because λ increases monotonically with assortativity [27, 125] and it has been found that networks with extreme assortativity have an onion-like structure that is robust to deterministic removals targeting nodes with large degree [129, 141], tuning λ and its corresponding eigenvectors may offer a fundamental route for protecting networks. This will be explored in later research.

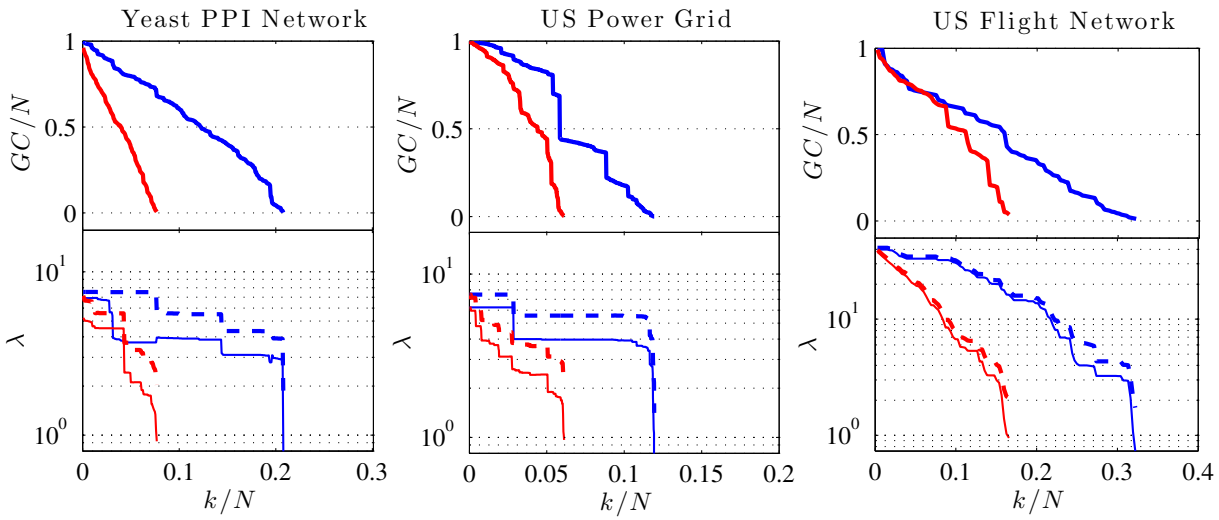


Figure 2.15: Deterministic malicious attacks were simulated on three undirected networks: (a) a protein-protein interaction network for yeast [28], (b) a power grid for the U.S. [148], and (c) a flight network for US Airlines [16], where the node with largest dynamical importance [121] was iteratively removed (along with nodes disconnected from the LSCC after its removal). The relative size of the GC is shown in the top panels (red lines) as a function of the fraction of removed nodes, k/N. The relevant eigenvalues for the networks' adjacency matrices, i.e., λ (dashed red lines) and λ_C (solid red lines) are shown in the bottom panel as a function of the fraction k/N of removed nodes. Note that the disappearance of the GC corresponds to $\lambda_C \approx 1$ and $\lambda \approx 2$ for these undirected networks. Blue lines indicate the same variables under random node removals.

2.4 Discussion

In Sec. 2.2 we presented a *network-specific* approach to weighted percolation in undirected networks and directed networks with bidirectional links (previously open problems). As opposed to most previous theory dealing with network ensembles, our spectral method predicts unique percolation characteristics for each unique network. While ensemble and network-specific methods offer complementary strategies which may lead to different insights, here we highlighted several benefits of a network-specific approach. (i) Ensemble methods cannot account for variability across

networks within the ensemble (see Fig. 2.6 and Ref. [32]). (ii) Application of any ensemble approach to a real-world network requires *a priori* assumptions about that particular network (e.g., that it either lacks complexity not accounted for in the ensemble or that its effects are small). (iii) Our approach naturally accounts for degree correlations of both the Markovian and non-Markovian types. (iv) Arbitrary targeting for node and/or link removal can be easily handled with our approach. (v) As shown in Sec. 2.3, a network-specific analysis allows one to additionally study deterministic attacks.

We have also provided several examples showing that our results are robust to various attack strategies, degree distributions, degree correlations, and moderate clustering. However, when strong community structure is added we expect our analysis to breakdown as occurs with the ensemble approaches [7]. Finding a spectral theory accounting for communities remains open for future research. Another direction of future work includes applying our techniques to more complicated percolation problems such as core percolation [51, 85] and Achlioptas [2] processes. However, efforts following the techniques provided here may also be restricted to the subcritical regime.

2.5 Conclusions

A central topic of this chapter has been the percolation threshold q^*, which indicates the critical fraction of nodes that must be retained such that the size of largest cluster of the remaining network is of the same order as the original network. The percolation threshold is a natural measure of a network's robustness (i.e., small q^* indicates robustness), and it allows one to compare both the robustness of different networks as well as the robustness of a single network subject to different fragmentation processes. Because q^* is generally inversely related to the largest eigenvalue λ of the network adjacency matrix, our results help explain why strategically decreasing and increasing λ offer fundamental strategies to respectively attack [121] and protect [92, 144] networks. This important result (along with the analysis for dynamics on networks presented in Ch. 1) motive the remaining topics of this thesis, the perturbation (Ch. 3) and evolution (Ch. 4) of the dominant eigenspace of the network adjacency matrix.

Chapter 3

Perturbation Theory for Network Modifications

In Ch. 1 we introduced several examples in which the largest eigenvalue λ of a network's adjacency matrix and its corresponding eigenvectors determine various types of dynamics occurring on a complex network. These included the collective behavior of coupled phase oscillators [120, 122, 123, 124], spreading processes for epidemics and information dissemination [34, 152, 155], the stability of gene expression in boolean networks [114, 139], and criticality in network-coupled excitable systems [80, 81, 82]. In addition, the dominant eigenspace is significant to the stability of equilibria for certain systems of network-coupled ordinary differential equations [1] [27, 56, 89, 133]. Moreover, in Ch. 2 we showed that the dominant eigenspace is also significant for the robustness of networks subject to network fragmentation processes. Therefore, it is essential that we fully understand how network perturbations affect the dominant eigenspace of the network adjacency matrix. This perturbation analysis is important not only for predicting the effects of network modifications, but it may also be used to guide network evolution. In particular, tuning network spectra through strategic structural modifications is a fundamental approach for controlling the robustness and dynamical properties of complex networks.

3.1 Background Information

Given the broad importance of λ and its eigenvectors in determining the outcome of so many dynamical processes involving networks, there has been much interest in its perturbation after

[1] Our method is applicable to systems where the eigenvalue of the Jacobian matrix with largest magnitude is real and well separated from the bulk of the spectrum.

structural modification. Here we briefly review several of these important results.

Dynamical Importance

When considering the importance of a node or link with respect to a network, and possibly a dynamical process involving the network, many heuristic metrics have been introduced to measure importance. These include eigenvector centrality, nodal degree, betweenness centrality, etc. [26, 106]. While many such centrality metrics have been shown to correlate well with various types of dynamical processes on networks, in many cases it is beneficial to have a fundamental relation between a given dynamical process and a given node or link's role in the dynamics. For the large class of *connectivity governed dynamics*, which may be described by the dominant eigenspace of the network adjacency matrix, the involvement of a given node or link in the collective dynamics is given by its *dynamical importance* [121]. Dynamical importance is defined as the magnitude of decrease in the dominant eigenvalue λ of the network adjacency matrix upon the respective removal of that node or link. Using perturbation theory, a first order approximation to the dynamical importance of a node n is given by [121]

$$I_n = -\frac{\delta\lambda_n}{\lambda} \approx \frac{v_n u_n}{v^T u}, \tag{3.1}$$

where u and v are the respective right and left eigenvectors corresponding to the principal eigenvalue λ (i.e., $Au = \lambda u$, $v^T A = \lambda v^T$), and $\delta\lambda_n$ is the decrease in the principal eigenvalue that would result from the removal of node n. Similarly, the dynamical importance of a link from node n to node m may be approximated as [121]

$$I_{nm} = -\frac{\delta\lambda_{nm}}{\lambda} \approx \frac{A_{nm} v_n u_m}{v^T u}. \tag{3.2}$$

Importantly, ranking nodes and links by their dynamical importance orders them corresponding to how significant they are for the dominant eigenvalue of the network adjacency matrix and consequently the dynamics governed by this eigenvalue. It follows that nodes and links with the largest dynamical importance are expected to have the largest effect on dynamics when perturbed, whereas nodes and links with the smallest dynamical importance will have the smallest effect on

dynamics when perturbed. This concept is far reaching and can be used, for example, to identify excellent immunization candidates for preventing epidemic spreading [121]. Moreover, because the largest eigenvalue and corresponding eigenvectors are important for the fragmentation of networks, the dynamical importance of nodes and links also relates to their structural importance. For example, it has been shown that a deterministic node removal strategy targeting nodes by dynamical importance often fragments a network more rapidly than targeting nodes by degree [121].

Other Research

While Ref. [121] focuses on the perturbation of λ due to the removal of a node or link, there is a general need for theory describing the addition and removal of sets of nodes, sets of links, subgraphs, etc. Reference [92] extended these results by finding perturbative expressions for the change in eigenvalue due to the removal of sets of nodes as well as the addition or deletion of sets of links. Reference [35] considered a perturbative approach to studying the spectrum of networks with community, or modular, structure. Unlike many of the other complex properties of complex networks, community structure can significantly affect a network's spectra by causing, for example, localization of the eigenvectors (see Ref. [146] for results regarding eigenvectors of the closely related Laplacian matrix). In general, network perturbation may be studied following the general principles of matrix perturbation [60].

3.2 The Effect of Module Addition

In this section, we consider the effect on the largest eigenvalue of a network's adjacency matrix due to the addition of a secondary network (referred to as the module or community). The majority of this work was recently published in Ref. [144]. As opposed to previous work [35, 92, 121], we explicitly consider the effect of the module's topology on the resulting eigenvalue and use this information to discuss how one can make optimal connections to either maximize or minimize the effect on λ. There are many applications where smaller groups adhere to a larger network in social and economical networks [101] (e.g., the merging of corporations or markets) and

biological networks (e.g., modifying a system of biochemical reactions with a drug [62, 74, 86, 163] or the merging of ecosystems [134]). For example, recent studies have shown that the effect on the largest eigenvalue of the Jacobian matrix describing interactions in an ecological network due to the addition of a species may be integral to the formation of ecological communities [53, 159]. Moreover, our results offer new insight regarding the prevalence of subgraph motifs (recurrent subgraphs having a frequency higher than expected). While motifs have been cited as essential building blocks in biological networks [6, 93], their role is not fully understood. For example, in contrast to several studies indicating that the global stability and robustness of a system is strongly influenced by the structure of motifs [68, 69, 118, 162], our study suggests that "how" a motif is connected to the remaining network may be as significant as its structure (see Fig. 3.8).

The remainder of this chapter is organized as follows. In Sec. 3.2.1 we present perturbative approximations for $\delta\lambda$ in terms of spectral properties of A, the links to be made, and the secondary network, or module, to be added. In Sec. 3.2.2 we test these approximations on several real networks. In Sec. 3.3 we discuss how our results can be used to optimize the connections between the original network and module. In Sec. 3.4 we discuss our results and cite several applications of how they may be used to control dynamics on complex networks. These results have broad application, range from those in which just a single merger needs to be optimally designed to cases where a large number of small additions need to be optimized to quickly evolve λ to a desired value. Conclusions are drawn in Sec. 3.5.

3.2.1 Eigenvalue Perturbation Analysis

We consider the addition of a secondary network, or module, to an existing network, as shown schematically in Fig. 3.1. The original network of size n is described by an $n \times n$ weighted adjacency matrix A such that its entries A_{ij} satisfy $A_{ij} \neq 0$ if there exists a link from node i to node j and $A_{ij} = 0$ otherwise. Another network of size m (described by an $m \times m$ adjacency matrix S) is to be connected to the original network. We will refer to this secondary network as the *module*. In what follows, we will sometimes refer to both the original network and the module by their respective

adjacency matrices, A and S.

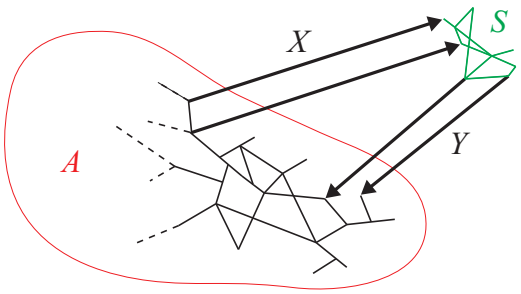


Figure 3.1: A module (described by matrix S) is connected to the original network (described by matrix A) using directed connections (described by the matrices X and Y).

Assuming that the original eigenvalue problems $Au = \lambda u$ and $v^T A = \lambda v^T$ have been solved, the modified eigenvalue problem after module addition may be formulated as

$$\begin{bmatrix} A & X \\ Y^T & S \end{bmatrix} \begin{bmatrix} u + \Delta^U \\ \Delta^L \end{bmatrix} = (\lambda + \delta\lambda) \begin{bmatrix} u + \Delta^U \\ \Delta^L \end{bmatrix}, \tag{3.3}$$

where we use the following definitions: (i) $\delta\lambda \geq 0$ denotes the shift in the largest eigenvalue; (ii) matrix X (Y) is size $n \times m$, has positive entries, and describes all directed links from A to S (S to A); (iii) Δ^U is a vector of length n which represents the shift in eigenvector u; and (iv) Δ^L is a vector of length m which represents the new eigenvector components. For the typical case in which no negative weights are allowed (i.e., $A_{ij} \geq 0$), the principal eigenvalues $\{\lambda, \lambda+\delta\lambda\}$ and all entries in $\{u, u+\Delta^U, \Delta^L\}$ are guaranteed to be nonnegative by the Perron-Frobenius theorem for nonnegative matrices [8, 151]. Although in this paper we only consider matrices with positive entries so that the Perron-Frobenius theorem can be applied, in general our analysis only requires that the eigenvalue with largest magnitude λ is real and well separated from the remaining eigenvalues. While this is typical for networks with positive links [35], it is also observed for networks with negative links provided that they represent a small fraction of the number of links (e.g., see Fig. 6 in [124]).

We restrict our analysis to cases where the effect of the module addition is small, which will allow us to study its effect as a perturbation to the original eigenvalue problem. This restriction is applicable to describing heavy-sided mergers and applications for which a network is modified gradually, such as the expansion of infrastructure. Considering the upper and lower blocks of Eq. (3.3) independently and after left-multiplying the top block by the left principal eigenvector v^T (i.e., $v^T A = \lambda v^T$), we obtain

$$\begin{aligned} \delta\lambda &= \frac{v^T X \Delta^L}{v^T(u+\Delta^U)}, \\ \Delta^L &= [(\lambda+\delta\lambda)I - S]^{-1} Y^T (u+\Delta^U), \end{aligned}$$

where I is the identity matrix of size m.

Assuming that the effect of the module addition is small, we have $\delta\lambda \ll \lambda$ and $v^T\Delta^U \ll v^T u$. We also assume that $\delta\lambda \ll |\lambda - \lambda_S|$, where λ_S is the largest eigenvalue of the module. In particular, we do not consider the degenerate case in which $\lambda_S = \lambda$. To first order, we find

$$\delta\lambda \approx \frac{1}{\lambda v^T u} v^T X K^S Y^T u, \tag{3.4}$$

$$\Delta^L \approx \lambda^{-1} K^S Y^T u, \tag{3.5}$$

where we have defined the $m \times m$ matrix

$$K^S \equiv (I - S/\lambda)^{-1}. \tag{3.6}$$

These expressions relate the change in the dominant eigenvalue $\delta\lambda$ to the topology of the added module S, the spectral properties of the original networks (u, v, and λ), and the way in which the module is coupled to the network by matrices X and Y. When the module contains few nodes, approximating $\delta\lambda$ by inverting an $m \times m$ matrix is significantly more efficient and, as we will see, offers more insight than solving the original $(m+n) \times (m+n)$ eigenvalue problem. Using $v = u$ and $X = Y$ for undirected networks, Eq. (3.4) simplifies to

$$\delta\lambda \approx \lambda^{-1} (X^T u)^T K^S (X^T u). \tag{3.7}$$

If the connections between the module and original network are made randomly, we can use Eq. (3.4) to estimate average values of $\delta\lambda$. Suppose that the entries of the matrix X are independent random variables such that $X_{ij} = 1$ with probability $x/(nm)$ and 0 otherwise, so that the expected number of links from the original network to the added module is x. Similarly, we assume that the entries of Y are independent random variables which are 1 with probability $y/(nm)$ and 0 otherwise. By averaging Eq. (3.4) and using the independence of X and Y, we find

$$\langle\delta\lambda\rangle = \frac{\bar{u}\bar{v}}{\lambda v^T u}\left(\frac{x}{m}\right)\left(\frac{y}{m}\right)\sum_{i,j} K^S_{ij}\ , \tag{3.8}$$

where $\bar{u} = n^{-1}\sum_{j=1}^{n} u_j$ and $\bar{v} = n^{-1}\sum_{j=1}^{n} v_j$. Thus, in addition to properties dependent on the original network, $\langle\delta\lambda\rangle$ is proportional to the product of the relative number of connections to and from the module (xy/m^2) and on the sum of elements in the matrix K^S. Moreover, for large λ/λ_S we have $\sum_{ij} K^S_{ij} \approx m$, the number of nodes in S. While this expression provides us with the average $\langle\delta\lambda\rangle$ when X and Y are chosen randomly, as discussed in Sect. 3.3, strategically selecting connection matrices (X,Y) (e.g., to maximize $\delta\lambda$) can lead to significant variations in $\delta\lambda$ for a given module.

For the optimization objectives explored later in Sec. 3.3, as well as situations in which computing K^S is inconvenient, it is useful to represent Eqs. (3.4) and (3.5) using a series expansion for K^S. For $\lambda_S < \lambda$, we have $K^S = (I - S/\lambda)^{-1} = \sum_{j=0}^{\infty}(S/\lambda)^j$. Using the truncated series $K^S \approx \sum_{j=0}^{k}(S/\lambda)^j$, we define the k-th order approximations as

$$\delta\lambda_k = \frac{1}{\lambda v^T u}\sum_{j=0}^{k-1}\lambda^{-j} v^T X S^j Y^T u, \tag{3.9}$$

$$\Delta^L_k = \sum_{j=0}^{k-1}\lambda^{-(j+1)} S^j Y^T u. \tag{3.10}$$

Because the matrices (X,Y) are often sparse and the module is often much smaller than the network, Eq. (3.9) is typically very computationally efficient. We note that for large enough k, the error introduced by neglecting higher order terms in Eq. (3.4) dominates the error of series truncation in Eqs. (3.9) and (3.10). No gain was found by using $k > 4$ in the experiments that are to follow.

Network and reference	N	$\langle d \rangle$	λ	λ_2
Neural network of C. elegans [158]	297	7.9	9.2	5.7
Network of political blogs [3]	1490	12.8	34.4	26.8
Yeast PPI network [28]	2361	5.6	12.1	9.4
Word association network [103]	5018	12.7	13.4	10.2

Table 3.1: Test networks used and their characteristics: number of nodes N; mean degree $\langle d \rangle$; largest eigenvalue λ; and second largest eigenvalue λ_2.

3.2.2 Numerical Experimentation

We test our approximations by considering module additions to four networks: a neural network of C. elegans [158]; a network of political blogs [3]; a network of protein-protein interactions in the organism S. cerevisiae (i.e., brewers/bakers yeast) [28]; and a network of associations between words [103]. Their characteristics are summarized in Table 3.1.

Average Behavior for Randomly Selected Links

We now investigate our results for the average effects of adding a module with random connections [Eq. (3.8)]. First, matrices were constructed by randomly selecting 10 entries in X and Y to be 1 and the rest to be 0 (i.e., in the notation introduced before Eq. (3.8), $x = y = 10$). For each realization $\delta\lambda^{\text{act}}$, the actual eigenvalue shift [i.e. solving Eq. (3.3)], was compared to our approximations given by Eqs. (3.4) and (3.9). In the top panel of Fig. 3.2, Eq. (3.8) (stars) is shown to accurately predict the numerically-observed average $\langle \delta\lambda \rangle$ (solid line) for connecting the modules to the directed neural network of C. elegans using 10^4 realizations of (X,Y). Average values for the relative error $\epsilon = (\delta\lambda - \delta\lambda^{\text{act}})/\delta\lambda^{\text{act}}$ are plotted in the bottom panel for both the neural network (circles) and a network of political blogs (triangles) (see Table 3.1) for all 13 non-isomorphic, directed modules of size 3. (Results for the other networks were found to be similar and are omitted for clarity.) Solid lines correspond to Eq. (3.4), while dotted (dashed) lines correspond

to Eq. (3.9) with $k = 1$ ($k = 2$).

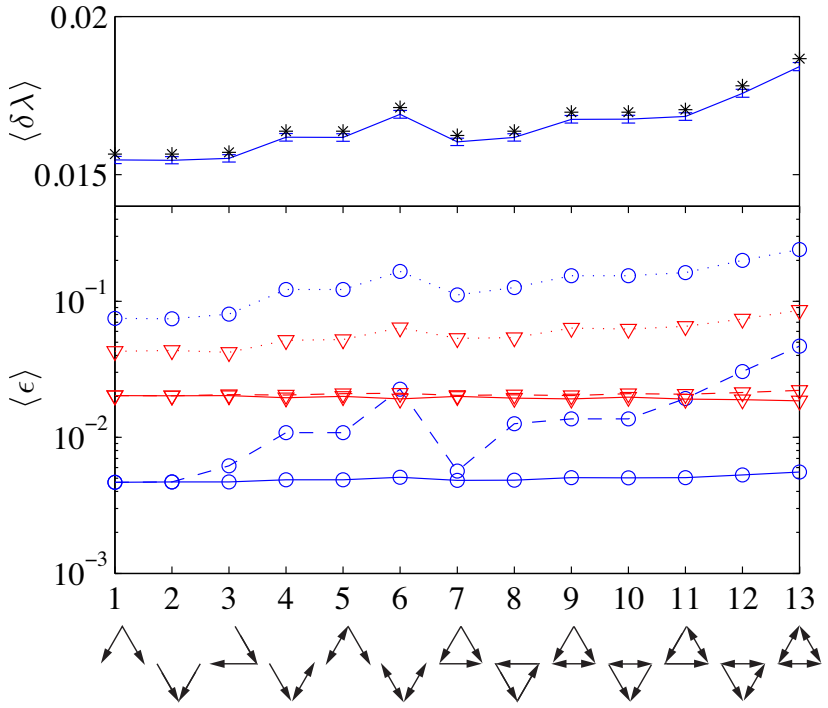


Figure 3.2: $\delta\lambda$ and approximation errors, ϵ, were averaged over 10^4 realizations of connecting a three-node module to the networks in Table 3.1 using 10 random links (see text). Equation (3.8) (stars) is shown to be accurate in the upper panel for typical results for the neural network of C. elegans. The average relative error $\langle \epsilon \rangle$ for the neural network of C. elegans (circles) and a network of political blogs (triangles) are given in the lower plot, where Eq. (3.4) (sold lines) and Eq. (3.9) with $k = 1$ (dotted) and $k = 2$ (dashed) are shown.

It can be observed in Fig. 3.2 that $\langle \delta\lambda \rangle$ changes substantially for the different three-node modules (for all networks, $\langle \delta\lambda \rangle$ typically increased $\sim 20\%$ from module 1 to module 13). Observe that the average error $\langle \epsilon \rangle$ of Eq. (3.9) when the module structure is not used [$k = 1$ (dotted lines in lower plot)] is strongly correlated with $\langle \delta\lambda \rangle$ (upper plot). This is to be expected as the error from

neglecting module structure should be related to that structure's ability to modify λ. Note that for the political blog network (triangles), using $k = 2$ in Eq. (3.9) is nearly as accurate as directly using Eq. (3.4). As previously mentioned, for large enough k the dominant source of error comes from neglecting higher orders of $\delta\lambda/\lambda$ in the derivation of Eq. (3.4) [as opposed to series truncation in Eq. (3.9)].

Strategic Link Selection

Having confirmed the accuracy of our results for randomly selected connections between the original network and module, we now turn to applications in which links between the module and network are chosen with some strategy. The validity of our approximations for specific connections is shown by considering the addition of two bidirectionally-linked nodes ($m = 2$) to an undirected protein-protein interaction (PPI) network and a directed network of word associations (see Table 3.1). In order to illustrate the dependence of $\delta\lambda$ on the matrices X and Y, we will consider two connection strategies: connecting the module to nodes with either (A) increasing nodal degrees or (B) increasing eigenvector entries. For strategy A, the nodes in the original network are ordered so that the in-degrees monotonically increase: $d_1^{in} \leq d_2^{in} \leq \cdots \leq d_N^{in}$. Then, considering k as a parameter, for $k \in \{1, 2, ..., N-20\}$, we establish a directed link from nodes $\{k, k+1, \ldots, k+20\}$ to both nodes in the module. The nodal out-degrees are then ordered such that $d_{i_1}^{out} \leq d_{i_2}^{out} \leq \cdots \leq d_{i_N}^{out}$, and links are made to nodes $\{i_k, i_{k+1}, \ldots, i_{k+20}\}$ from both nodes in the module. The case $k = 0$ corresponds to connecting the network nodes with smallest d^{in} to both module nodes, which in turn connect to the nodes with smallest d^{out}; whereas the case $k = N - 20$ corresponds to connecting the nodes with largest d^{in} to both module nodes, both of which in turn connect to the nodes with largest d^{out} (shown schematically in Fig. 3.3).

For strategy B, we now order the nodes in the original network in order of increasing entries of the left eigenvector v so that $v_1 \leq v_2 \leq \cdots \leq v_N$. As before, for $k \in \{1, 2, ..., N-20\}$, we connect nodes $\{k, k+1, \ldots, k+20\}$ in the network to both module nodes, both of which in turn connect to nodes $\{i_k, i_{k+1}, \ldots, i_{k+20}\}$, where the indices i_j now correspond to the ordering of the

right eigenvector entries such that $u_{i_1} \leq u_{i_2} \leq \cdots \leq u_{i_N}$. For both strategies, the indices simplify for undirected networks, for which we have $u = v$, $d^{out} = d^{in}$, and $i_k = k$.

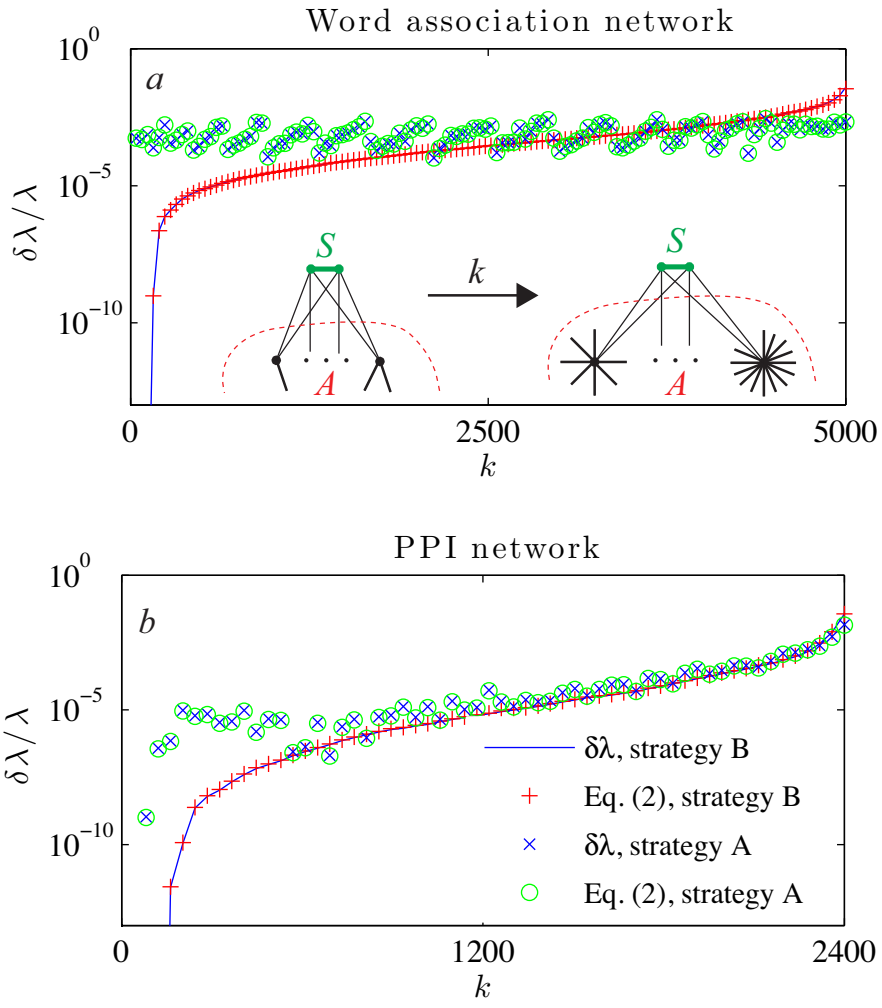


Figure 3.3: Eigenvalue shift $\delta\lambda$ for connecting a two-node module to (a) the word association network and (b) the PPI network. Equation (3.4) (crosses) agrees well with actual values $\delta\lambda^{\text{act}}$ (solid line) for strategy B. The x's and circles show the same respective quantities for strategy A. As indicated by the drawing, increasing k corresponds to connecting the module to nodes with increasing degrees (strategy A) or eigenvector entries (strategy B).

In Fig. 3.3, $\delta\lambda$ is plotted for strategies A and B as a function of the parameter k for both (a) the directed word-association network and (b) the undirected PPI network. For strategy B, the crosses show the approximation given by Eq. (3.4) and the solid line shows the numerically-calculated value from directly solving the eigenvalue problem Eq. (3.3). The x's and circles respectively show the same quantities for strategy A. The first observation is that the approximation for $\delta\lambda$ works well, with only a small deviation as the perturbation becomes large (not shown). One can observe that strategy B is superior for yielding either large or small $\delta\lambda$ for both networks. However, the two strategies are similar for producing large $\delta\lambda$ for the PPI network in Fig. 3.3b. This is expected when the first-order approximations to the eigenvectors ($u_i \propto d_i^{out}$ and $v_i \propto d_i^{in}$ [121]) are valid. The results of this experiment suggest that it may be useful to devise connection strategies to systematically maximize (or minimize) $\delta\lambda$.

3.2.3 Eigenvector Perturbation Analysis

We now study the perturbation of the dominant eigenvectors corresponding to λ, which are important in many applications. For example, in Ch. 2 we found that for an uncorrelated, random network subject to unweighted node percolation, the expected sizes of the in-component and out-component of a given node n may be respectively approximated by its corresponding left and right eigenvector entries, $s_n^{in} \propto v_n$ and $s_n^{out} \propto u_n$. As discussed in Sec. 1.4, similar estimates may be made for various types of dynamics.

We begin by solving the upper block of Eq. (3.3) for Δ^U,

$$\Delta^U = \frac{1}{\lambda + \delta\lambda}\left(I - \frac{A}{\lambda + \delta\lambda}\right)^{-1}\left(X\Delta^L - \delta\lambda u\right), \tag{3.11}$$

which uses that $\delta\lambda > 0$ for networks with positive links to guarantee invertibility. In fact, the inverse term can be replaced by its series expansion

$$\Delta^U = \frac{1}{\lambda + \delta\lambda}\sum_{j=0}^{\infty}\left(\frac{A}{\lambda + \delta\lambda}\right)^j\left(X\Delta^L - \delta\lambda u\right). \tag{3.12}$$

At this point it is convenient to separate vector $X\Delta^L$ into its projection on the subspace spanned by u, $u^T X\Delta^L u$, and the orthogonal subspace, $b = X\Delta^L - u^T X\Delta^L u$ (i.e., note that $u^T b = 0$). It

follows that the above equation may be simplified as

$$\Delta^U = \frac{1}{\lambda + \delta\lambda} \sum_{j=0}^{\infty} \left(\frac{A}{\lambda + \delta\lambda} \right)^j b + \left(\frac{u^T X \Delta^L}{\delta\lambda} - 1 \right) u, \tag{3.13}$$

where the first term indicates perturbation out of the subspace spanned by u and the second term denotes perturbation along this subspace. Note that this result has not required any approximations.

Because we are primarily interested in the perturbation of the dominant eigenspace away from the subspace spanned by u, we will study the projection $\tilde{\Delta}^U = \Delta^U - u^T \Delta^U u$ so that $u^T \tilde{\Delta}^U = 0$. Using our estimates for $\delta\lambda$ and Δ^L given by Eqs. (3.4) and (3.5) [or equivalently Eqs. (3.9) and (3.10) for sufficiently large k], we now introduce an iterative method estimating $\tilde{\Delta}^U$ using the above result. Letting $\tilde{\Delta}_0^U = \vec{0}$, we define

$$\tilde{\Delta}_{j+1}^U = \frac{1}{\lambda + \delta\lambda} A \tilde{\Delta}_j^U + \frac{1}{\lambda + \delta\lambda} b. \tag{3.14}$$

Note that because b is orthogonal to u, $\tilde{\Delta}_j^U$ converges to an estimate of $\tilde{\Delta}^U$ with a rate given by $\lambda_2/(\lambda + \delta\lambda)$, where λ_2 is the second largest eigenvalue of A [60]. Also note that this convergence is expected to be fast as typically one has $\lambda_2 \ll \lambda$ for large networks lacking strong macroscopic community structure [125].

Motivated by the observation that the largest entries in the vector $\tilde{\Delta}^U$ typically correspond to nodes that are in close proximity to the added subgraph (e.g., either they, or their neighbors, link directly to the module), we additionally present a sparse approximation for this local perturbation. Neglecting the requirement that $\tilde{\Delta}^U$ must be orthogonal to u, the largest entries may be estimated by

$$\widehat{\Delta}_{j+1}^U = \frac{1}{\lambda + \delta\lambda} A \widehat{\Delta}_j^U + \frac{1}{\lambda + \delta\lambda} X \Delta^L, \tag{3.15}$$

with $\widehat{\Delta}_0^U = \vec{0}$. Note that because X contains at most x nonzero entries, for a network with average nodal degree $\langle d \rangle$, each estimate $\widehat{\Delta}_j^U$ will contain on average $x\langle d \rangle^{j-1}$ nonzero entries for $j > 0$.

3.2.4 Numerical Experimentation

In this section we illustrate the accuracy of our predictions to $\tilde{\Delta}^U$. Specifically, we consider the addition of a module to several real-world networks via either randomly selected or strategically selected links.

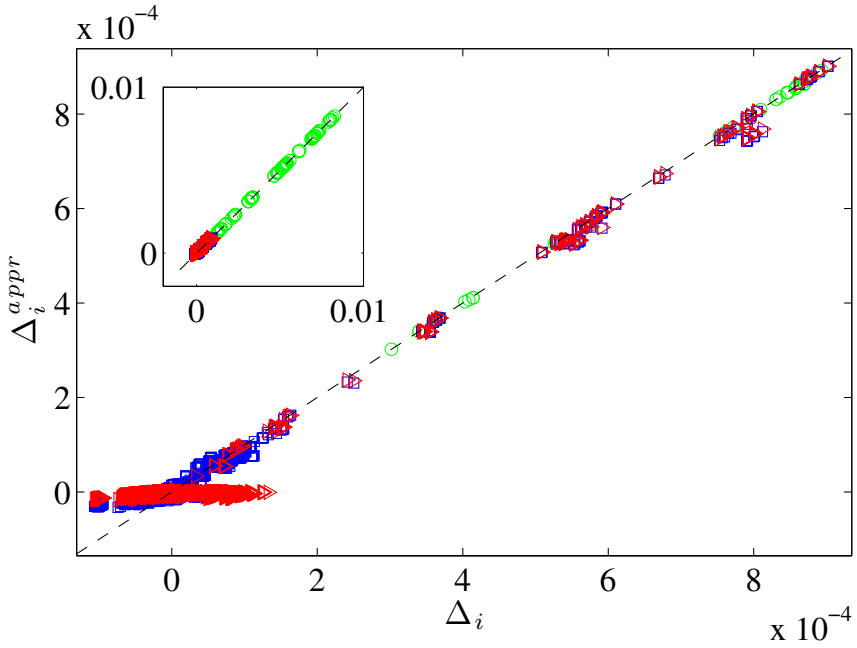


Figure 3.4: Our prediction for $\tilde{\Delta}^U$ given by Eq. (3.14) with $j = 1$ (red triangles) and $j = 2$ (blue squares) is shown for the addition three node modules to the directed C. elegans neural network [158] via randomly selected links. Here i indicates nodal index. The subplot shows a larger plot range to emphasize the fact that entries Δ^L (green circles) are much larger that those in $\tilde{\Delta}^U$.

Random Link Selection

In this first experiment we consider the accuracy of Eq. (3.14) for the neural network of C. elegans [158] in which directed modules with three nodes are added using between $x = 1$ and $x = 5$

randomly selected, undirected links. In Fig. 3.4 we compare the actual perturbations $\tilde{\Delta}_i^U$ versus our prediction $\tilde{\Delta}_i^{appr}$ using Eq. (3.14) with $j = 1$ (red triangles) and $j = 2$ (blue squares). In addition to plotting the entries of vector $\tilde{\Delta}^U$, we also compare actual values of Δ^L versus our prediction given by Eq. (3.10) (green circles). The inset shows a larger plot range to emphasize that the entries Δ^L are much larger than the entries Δ^U. Also note that while most entries Δ^U are very small (see the clump in the low left corner), a few entries can be large. These large entries correspond to the network nodes that link to the module and their neighbors.

In Fig. 3.5 we show the accuracy of our sparse approximation given by Eq. (3.15). Red triangles and blue squares respectively represent Eq. (3.15) with $j = 1$ and $j = 2$. One can observe that the sparse approximation accurately predicts the largest entries $\tilde{\Delta}^U$. Note, however, that these are still much smaller than the entries in Δ^L (green circles).

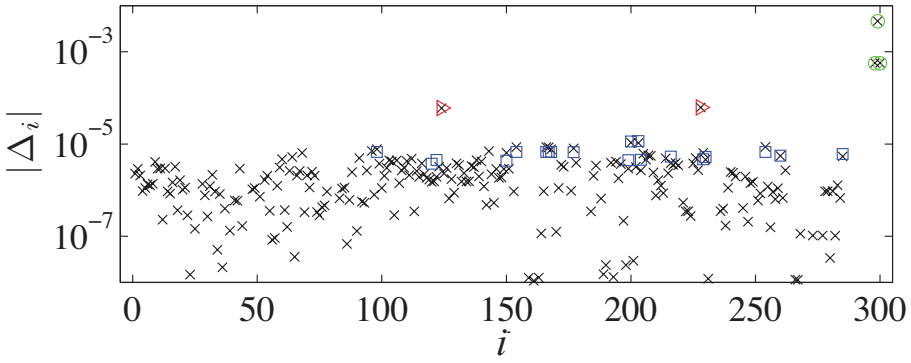


Figure 3.5: Entries Δ_i (black x symbols) are shown versus node index i. The accuracy of Eq. (3.15) for $j = 1$ (red triangles) and $j = 2$ (blue squares) is shown, where it can be observed that these correspond to the largest entries in Δ^U. The accuracy of our prediction to Δ^L using Eq. (3.10) for $k = 4$ is also displayed (green circles).

Strategic Link Selection

We now introduce an experiment in which we heuristically attempt to control the magnitude of the perturbation $\tilde{\Delta}^U$ upon the addition of a single node to a network, which we measure by $||\Delta^U||_2$. Because Eq. (3.15) shows that small perturbations will predominantly have a local effect (i.e., the perturbation $\tilde{\Delta}_i^U$ is expected to decay as λ^{-l}, where l is the smallest network path from the module to node i), one would expect that u can be most perturbed by linking the module to the nodes $\{i\}$ corresponding to the largest entries $\{u_i\}$. For undirected networks this corresponds to the nodes with the largest dynamical importance [121].

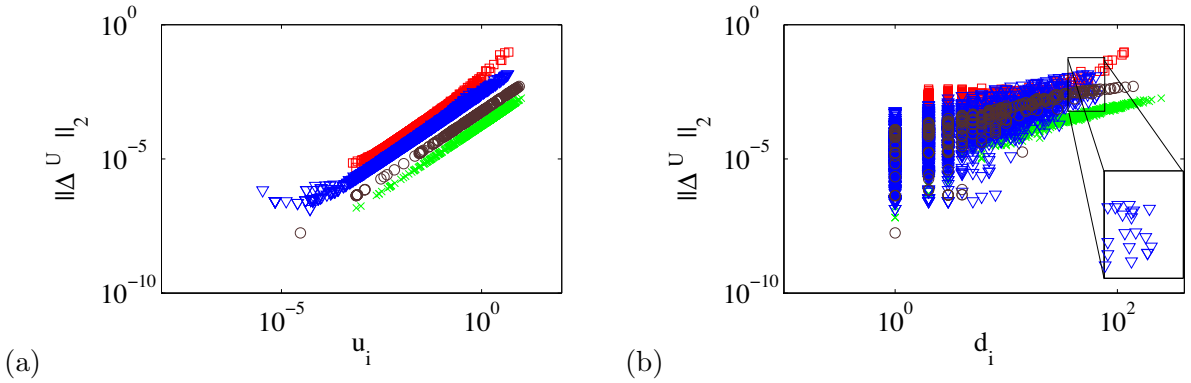


Figure 3.6: For the addition of a single node to node i in an undirected network via one undirected link, the disturbance to the original dominant eigenvector (quantified by $||\Delta^U||_2$) is shown as a function of either (a) the components u_i of the eigenvector, or (b) the nodal degrees d_i. Observe that this disturbance can be maximized and minimized by linking to the node i such that u_i is the largest and smallest entry in u, respectively. In contrast, using nodal degree d_i often does a poor job of controlling the perturbation to the eigenvector. The following test networks were used: Barabási-Albert preferential attachment network [11] (red squares); a network of US Air flights [16] (brown circles); a Caltech Facebook network [147] (green x symbols); and a protein-protein interaction network for yeast [28] (blue triangles).

In Fig. 3.6 we support this claim with several experiments in which a single node was added to test networks with a single undirected link. Four networks were considered: a Barabási-Albert preferential attachment network [11] with $N = 5000$ nodes, and $M = 2N$ undirected links (red squares); a network of US Air flights [16] (brown circles); a Facebook network for Caltech [147] (green x symbols); and a protein-protein interaction network for yeast [28] (blue triangles). Here we plot the perturbation magnitude $||\Delta^U||_2$ versus either (a) the corresponding eigenvector entries u_i, or (b) the corresponding nodal degrees d_i. Note in Fig. 3.6(a) that linking to nodes corresponding to small and large u_i appears to respectively minimize and maximize $||\Delta^U||_2$. In contrast, one can observe in Fig. 3.6(b) that linking to small and large d_i often does a very poor job of controlling the perturbation to the dominant eigenvector u. However, for random networks lacking degree-correlations, these two approaches are expected to be similar due to the mean field approximations $u_i \propto d_i$ [125].

3.3 Optimizing Connections with Strategic Link Selection

The issue of efficiently decreasing λ by removing nodes or links from a network has been recently addressed [121], where it was found that when removing a single node, λ is most decreased by removing the node with largest dynamical importance. We consider a closely related issue: given a module S to be added to a network A with given constraints (such as a fixed number of connections), how should the links between the network and module be chosen to either maximize or minimize $\delta\lambda$? Given some set of constraints and staying within our previous assumptions, we will look for matrices (X,Y) that maximize (or minimize) $\delta\lambda$ in Eq. (3.4). In the examples that follow, it is helpful to assume that the node indices are now ordered such that the left eigenvector entries are in decreasing order: $v_1 \geq v_2 \geq \cdots \geq v_n \geq 0$. In addition, the entries of the right eigenvector are ordered using indices $\{l_i\}$ so that $u_{l_1} \geq u_{l_2} \geq \cdots \geq u_{l_n} \geq 0$. (If A is symmetric, $u = v$ and $l_i = i$.) We will present our optimization methodology for two examples, yet the techniques presented are general and have potential application beyond these particular constraints.

Example I: Multiple Links Per Module Node

In the first example we assume that the number of connections from the original network to the module, x, and the number of connections from the module to the original network, y, are fixed and less than n, the number of nodes in the original network. It is also assumed that all links have strength one (i.e., $X_{ij}, Y_{ij} \in \{0,1\}$) and multiple links per module node are allowed.

The right hand side of Eq. (3.4), which approximates the quantity to be maximized, is proportional to $\sum_{i,j}(X^T v)_i^T K_{ij}^S (Y^T u)_j$. This sum can be maximized by (i) finding indices a and b such that $K_{ab} = \max_{ij}\{K_{ij}\}$ and (ii) choosing X and Y to make $(X^T v)_a^T$ and $(Y^T u)_b$ as large as possible. The scalar $(X^T v)_a^T$ is maximized by placing the x ones in the a-th column of X and in positions $1, 2, \ldots, x$ corresponding to the largest values of v, while $(Y^T u)_b^T$ is maximized by placing the y ones in the b-column of Y and in positions $l_1, l_2, \ldots, l_y$ corresponding to the largest values of u. In this way, $(X^T v)_i^T = \delta_{ia}\sum_{j=1}^{x} v_j$ and $(Y^T u)_i^T = \delta_{ib}\sum_{j=1}^{y} u_{l_j}$, where δ_{ij} is Kronecker's delta. The maximum of Eq. (3.4) is then

$$\delta\lambda^{max} \approx \frac{K_{ab}^s}{\lambda v^T u}\sum_{i=1}^{y} v_i \sum_{j=1}^{x} u_{l_j} \tag{3.16}$$

This result implies $\delta\lambda$ may be maximized for the constraints of example (I) by connecting the x nodes with the largest left eigenvector entries v_i in the original network to a single node in the module (having index a), and by also originating all links from the module to the original network from a single module node (having index b) to the y nodes in the original network with the largest entries of the right eigenvector u. For large values of λ/λ_s, the maximum entry of matrix K^s is typically in its diagonal, yielding $a = b$ as shown in Fig.3.7.

For a heuristic interpretation of this result, let us assume that $A_{ij} \in \{0,1\}$ and denote $L_i^{o,p} = \sum_j (A^p)_{ij}$ and $L_i^{t,p} = \sum_j (A^p)_{ji}$ as the number of paths of length p originating from and terminating at node i, respectively. Thus $L_p = \sum_{ij}(A^p)_{ij}$ is the total number of paths of length p. These quantities satisfy $||L_i^{o,p}||_2^{-1} L_i^{o,p} \to u_i$, $||L_i^{t,p}||_2^{-1} L_i^{t,p} \to v_i$, and $L_{p+1}/L_p \to \lambda$ as $p \to \infty$ [60]. Therefore connecting nodes with large v_i (which receive many paths) to nodes with large u_i (which distribute many paths) will have the largest impact in how L_p grows with p, which determines λ.

We therefore define a node i with large v_i as a *point of contraction* and a node j with large u_j as a *point of expansion*. Therefore our result for example (I) is that the effect of the whole module is to act as a bridge from points of contraction to points of expansion in the original network.

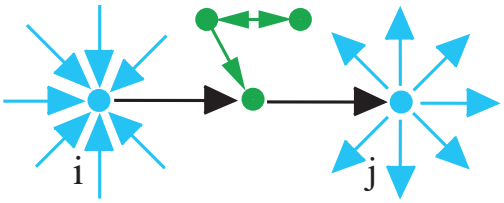


Figure 3.7: A typical optimal connection for example (I): node i (a *point of contraction* with large left eigenvector entry v_i) points to a node in the module, which in turn points to node j (a *point of expansion* with large right eigenvector entry u_j).

Example II: One Link Per Node

In the second example we require that, in addition to a fixed number of links x and y with unity strength, no more than one link can be added to a particular node in the module or original network. Because undirected links are equivalent to two links and violate our constraint, it is reasonable (although not necessary) to assume that the network and module are directed. To treat this case, we maximize successive terms in Eq. (3.9). The first term, $v^T XY^T u/\lambda$, vanishes since any entry of XY^T is nonzero only if there is a module node that has links both to and from the network, a situation which is not allowed by our constraint. Therefore, we maximize the next term, $(X^T v)^T S(Y^T u)/\lambda^2$.

As shown in Fig. 3.8(a), let us denote the set of nodes in the original network that point to the module as NO (network outgoing), the set of nodes in the module that are pointed to by the original network as SI (module incoming), the set of nodes in the module that point to the original

network as SO (module outgoing), and the set of nodes in the original network that are pointed to by the module as NI (network incoming). Because no node can have more than one new link, there is a one-to-one correspondence between nodes in NO and nodes in SI. The index of nodes in SI will be represented as i_j, where node j in NO points to node i_j in SI. We have $(X^T v)^T_{i_j} = v_j$ if $j \in$ NO and $i_j \in$ SI, and 0 otherwise. With a similar notation for SO and NI, we have $(Y^T u)_{m_k} = u_k$ if $m_k \in$ SO and $k \in$ NI, and 0 otherwise. It follows that Eq. (3.9) yields

$$\delta\lambda_2 = \frac{1}{\lambda^2 v^T u} \sum_{j \in \mathrm{NO}} v_j \sum_{k \in \mathrm{NI}} S_{i_j m_k} u_k.$$

This expression is maximized if S contains a directed complete bipartite graph for disjoint subsets SI and SO such that every node in SI points to every node in SO (see Fig. 3.8a). Assuming that one can be found, we may set $S_{i_j m_k} = 1$ and look for sets NO and NI that solve

$$\delta\lambda_2^{max} = \left(\frac{1}{\lambda^2 v^T u}\right) \max_{\mathrm{NO} \cap \mathrm{NI} = \emptyset} \left(\sum_{j \in \mathrm{NO}} v_j \sum_{k \in \mathrm{NI}} u_k \right). \tag{3.17}$$

Let $Q = \{i\}_{i=1}^{x} \cap \{l_i\}_{i=1}^{y}$. If $Q = \emptyset$ then Eq. (3.17) is solved by letting NO $= \{i\}_{i=1}^{x}$ and NI $= \{l_i\}_{i=1}^{y}$, which yields

$$\delta\lambda^{max} \approx \frac{1}{\lambda^2 v^T u} \sum_{j=1}^{x} v_j \sum_{i=1}^{y} u_{l_i}. \tag{3.18}$$

As indicated in Fig. 3.8a, this corresponds to selecting nodes of contraction for NO and nodes of expansion for NI.

The significance of link choices for maximizing $\delta\lambda$ is shown in Fig. 3.8(b), where the module in Fig. 3.8(a) was added to the neural network of C. elegans with constant node selections for NI and NO but using several module orientations (defined as a particular choice for the disjoint sets SO and SI in the module, and shown in the horizontal axis of Fig. 3.8b). The solid lines show $\delta\lambda/\lambda$ found numerically using NI $= \{l_1, l_2\}$ and either N0 $= \{1, 2\}$ (thick) or N0 $= \{2, 1\}$ (thin) (see next paragraph for discussion). Symbols indicate $\delta\lambda/\lambda$ found using Eq. (3.4). One can observe that our maximization strategy for example (II) (Fig. 3.8a) does in fact maximize $\delta\lambda$ (see orientation 6). An important practical issue is that the eigenvectors may be unknown and require estimation using local information. One can observe that attempting to maximize $\delta\lambda$ using

the first-order approximations $v_i \propto d_i^{in}$ and $u_i \propto d_i^{out}$ [121] may also be a good strategy (dashed lines). If necessary, a more refined approximation for the eigenvectors may be sought (e.g., using second-order neighbors [92]).

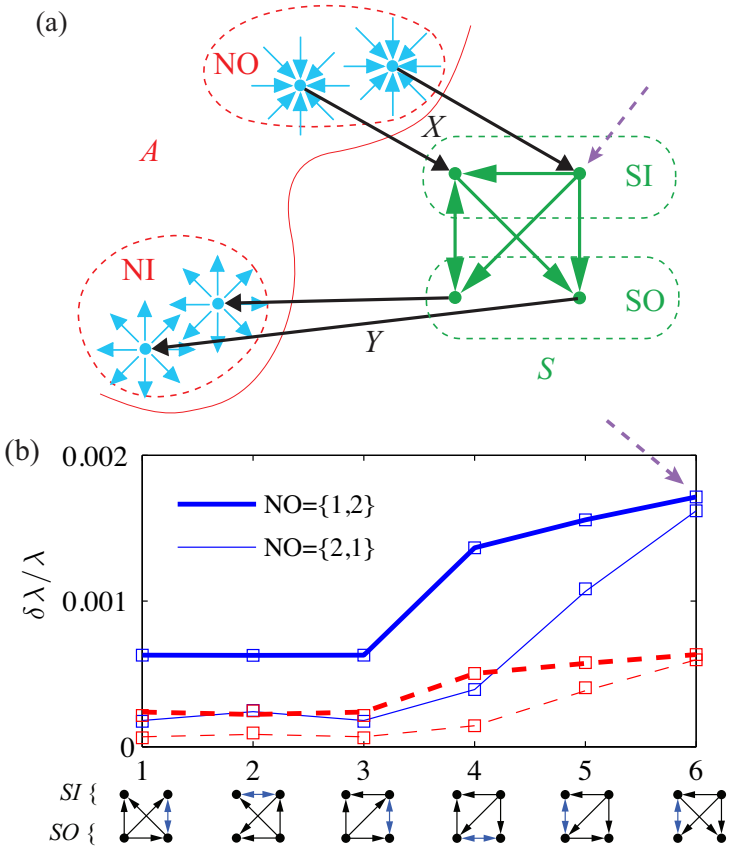


Figure 3.8: (a) Typical optimal link selections for example (II) for $x = y = 2$. Two points of contraction (NO) link to two module nodes (SI) and the remaining two module nodes (SO) link to two points of expansion (NI). The module also contains a directed complete bipartite graph pointing from SI to SO. (b) Under the restrictions of example (II), the module in Fig. 3.8a was connected to the neural network for C. elegans using various orientations. Solid lines indicate letting NI $= \{l_1, l_2\}$ and either NO $= \{1, 2\}$ (thick) or NO $= \{2, 1\}$ (thin). Symbols show Eq. (3.4). Approximating points of contraction (expansion) by nodes with large d^{in} (d^{out}) also offers a decent strategy (dashed).

It is important to note that we have so far neglected higher order terms of Eq. (3.9) in addressing example (II), which are responsible for the difference in $\delta\lambda$ for the permutation NO $= \{1, 2\}$ or NO $= \{2, 1\}$. Attempting to maximize the third term of the series in Eq. (3.9) (which is proportional to $v^T X S^2 Y^T u$) while using the nodes of contraction, 1 and 2 (with $v_1 \approx 0.58$ and $v_2 \approx 0.23$), we see that the more-dominant point of contraction (node 1) should link to the module node indicated by the dashed arrow in Fig. 3.8a. (Note that there is a path of length 2 stemming from this node to each node in SO, whereas there are none for the other node in SI.) Unlike permuting nodes in SI, permuting nodes in SO had little effect for this network since $u_{l_1} \approx u_{l_2} \approx 0.23$.

Up to this point we have assumed $Q = \emptyset$, where Q is defined just after Eq. (3.17); however this may not always be the case. For example, as more links are made (i.e., for increasing x, y), one would expect some nodes to have large values for both v_i and u_i. This may also occur for networks with correlations between d^{in} and d^{out} and, in fact, always occurs for undirected networks where $l_i = i \ \forall \ i$. For these situations, nodes in Q must be allocated to either NO or NI and additional nodes must be selected. Considering the limiting case of an undirected network under the constraints of example (II), maximization of the second-order term in Eq. (3.9) indicates that we should choose NO,NI $\subset \{i\}_{i=1}^{x+y}$. (Recall that the first-order term is zero by our constraints.) The allocation of these indices should then correspond to successively maximizing the third-, fourth-, ..., kth-order terms until all degrees of freedom have been exhausted. While this strategy of successive maximization does not guarantee the optimal connections (which would require considering all possible links between S and A), it is computationally efficient and ensures a near-optimal solution.

3.4 Discussion

While we have presented an efficient strategy for maximizing $\delta\lambda$ for the addition of a module under two examples of constraints, our methodology is general and is thus applicable for many constraints not discussed here. For example, the problem of minimizing $\delta\lambda$ under the constraints of Example II may be solved by minimizing successive terms of Eq. (3.9). Heuristically, this

corresponds to connecting nodes in A with small values of v_n to the module, and then from the module to nodes in A with small values of u_n. The module should also be oriented so as few links as possible point from SI to SO. We now discuss several applications of using module addition(s) to direct dynamics on networks.

Increasing λ has many real-world applications. For example, because λ relates to the ability of network-coupled oscillatory systems to synchronize [120, 122, 123, 124], one or several module additions to increase λ may be useful to promote synchronization in, for example, a biological process or power grid. Moreover, epidemic thresholds of spreading processes on networks are often dependent on λ^{-1} [34, 152, 155]. Increasing λ can increase the connectivity of a network, improving flow and reducing the epidemic threshold. This may be useful, for example, if one wants to improve communication over a social network or routing-system. The related problem of percolation on networks (where nodes and/or links are randomly removed) is also related to λ^{-1} [23, 126, 145]. Increasing λ can increase a network's robustness against network degradation under failure, blackout, jamming, or attack.

For other dynamical systems, it is beneficial to have a small value for λ. For example, the instability of equilibria for a system of network-coupled ordinary differential equations (ODEs) (e.g., interactions in a metabolic network) is related to the largest eigenvalue of a weighted adjacency matrix defined in terms of the system's Jacobian [27, 56, 89, 133] (i.e., if $\lambda < 1$, then the equilibria are stable). When the eigenvalue of the Jacobian matrix with largest magnitude is real and well separated from the bulk of the spectrum, our method is applicable. For example, besides choosing appropriate link weights to keep $\delta\lambda$ small, choosing optimal connections and module orientation (as shown in Sect. 3.3) may also aid in preserving the stability of equilibria for a system undergoing modification. Such analysis may be relevant, for example, in understanding the formation of ecological communities for which the largest eigenvalue of the system's Jacobian has already been suggested to guide the network's evolution under species additions and subtractions [53, 159].

Future applications of our results are also not limited to network dynamics for which the dependency on λ is currently well established. For example, minimizing $\delta\lambda$ for a module addition

may present an effective strategy for minimizing global effects during a network modification. Possible applications may include aiding the development of systems-level drug design by indicating candidate drug targets that are less invasive (e.g., nodes with middle-valued degrees are typical [62]). The importance of developing mathematical approaches for this promising field are often mentioned [74, 86, 163]. Another open question is the implications of our results on the prevalence of subgraph motifs, which have been proposed to be the basic building blocks of biological networks [6, 93]. In contrast to several studies showing that global dynamics of a system can depend on the structure of subgraph motifs [68, 69, 118, 162], our results suggest that "how" motifs are connected in the network may be as important as their structure.

3.5 Conclusions

In this chapter we have studied the perturbation of the dominant eigenspace of the network adjacency matrix for synthetic and real-world networks. Our work was motivated by the observation that the robustness and dynamical properties of complex networks strongly depend on this eigenspace. Because tuning λ and its corresponding eigenvectors therefore has broad applications for controlling complex systems, we presented techniques to optimize network connectivity during the addition of module. Specifically, we introduced techniques to maximize λ during network mergers. In the following chapter, we will build on these results by considering repeated perturbations (i.e., the iterative addition of a link) and the evolution of λ for forming networks.

Chapter 4

Maximizing the Dominant Eigenvalue of Forming Networks

Dynamics on networks has become a research area of broad importance, with considerable effort focused on understanding how dynamics are affected by network structure. Of particular interest are dynamics that depend on global measures of network connectivity, and in particular on the largest eigenvalue λ of the network adjacency matrix A ($A_{ij} \neq 0$ if a link exists from node i to node j). We will refer to this broad class, which includes models for synchronization [120, 122, 123, 124], genetic expression [114, 139], neural excitation [80, 81, 82], and epidemic spreading [34, 110, 152, 155], as *connectivity-governed dynamics* (see Sec. 1.4 for a summary). We note, however, that while analyses of such systems typically assume that the network structure is static and connected (i.e., lacking isolated nodes/clusters), many applications exist for which the network structure is non-static and/or fragmented, such as epidemic spreading with immunization [34, 36, 47, 87, 111], communication and transit systems operating under failure or attack [49, 96, 98], and information processing in the brain [30].

The systems that we categorize as having connectivity-governed dynamics share a common property of being easily manipulated through changes in λ. For example, one can prevent viral spreading in technological and social networks by decreasing λ through immunization [34] or promote the dissemination of information in communication and sensor networks by increasing λ [144]. As discussed in Ch. 3, there has been much interest in studying the effects of topological modification on λ and developing efficient strategies for tuning λ through the addition and/or subtraction of links and/or nodes [92, 121, 144, 153]. However, such perturbation techniques do not address

networks undergoing formation or fragmentation processes, a problem traditionally studied with network percolation theory [24, 52].

From network resilience against targeted attacks and failures to dynamics on networks, which often map spreading dynamics to percolation processes, there are widespread applications for percolation theory in the field of complex networks. However, applying such techniques to dynamics on evolving networks is hindered in that it can require two levels of analysis: theory for the change in network topology and theory for the dynamics. We hypothesize that a central element frustrating the development of this approach is that this field of research has largely focused on studying a phase transition in the size of the largest connected cluster by analyzing cluster aggregation and the emergence of a *giant component* (recall Ch. 2), a cluster whose size is of the same order as the entire system [24, 52]. However, the application of subsequent percolation theory requires information about cluster topology in addition to cluster size, which highlights the need for percolation theory focusing on other cluster properties such as the spectra of clusters (i.e., the eigenvalues of adjacency matrices corresponding to clusters), modularity, assortativity, transitivity, etc. [106, 105].

Here we study the link-percolation phase transition using λ as a novel order parameter, shedding light on a new phase transition in connectivity, corresponding to a poorly connected network becoming well-connected (or vice versa) in terms of the topology's effect on dynamics. The majority of this work was recently published in Ref. [142]. In order to produce such a transition, we introduce a link-formation rule called *Social Climber* (SC) attachment for which we derive the asymptotic scaling behavior of λ for large network size N. We show that networks forming under SC attachment exhibit maximal scaling $\lambda \sim \mathcal{O}(N^{1/2})$ for networks containing $\mathcal{O}(N)$ nodes, indicating that our model may be of broad interest for the design of networks with large λ, a property that is often beneficial [144] and can lead to, for example, excellent robustness against attack and failure [126, 145] (e.g., see Ch. 2) and very good spreading characteristics [34, 80, 81, 82, 110, 152, 155].

It follows that SC attachment is a promising approach for the design of self-organized communication and sensor networks [33, 128] with topologies designed for the rapid dissemination of information. We demonstrate this application by showing that networks formed under SC attach-

ment exhibit enhanced spreading properties with respect to the susceptible-infected-susceptible (SIS) model [34, 110, 152, 155], a contagion model with many applications including the dissemination of information, sometimes referred to as "gossip based" communication or epidemic routing [70, 150, 161]. We note that the development of this potential application may be facilitated by the fact that SC attachment may be combined with arbitrary percolation processes, such as Erdős-Rényi (ER) percolation [55] and Achlioptas processes [2, 21, 54, 138], to independently control cluster aggregation (determined by the percolation process) and connectivity within clusters (determined by SC attachment).

4.1 Background Information

In Ch. 2 we described the robustness of networks undergoing processes of fragmentation. In this pursuit, we described the field of percolation theory in which the sizes of clusters, and in particular the giant component, are studied for a network as it fragments into small, disconnected clusters. Percolation theory also studies the opposite process in which links are added to disconnected nodes until clusters form and a giant component emerges. In fact for many processes, network fragmentation and formation may simply correspond to a time reversal (i.e., removing versus adding links [19]). Therefore, the background information required for our analysis of network formation in this chapter is similar to that presented in Ch. 2. For example, in both chapters we will study clusters (i.e., sets of nodes in which any node may reach any other node), the giant component (i.e., the largest cluster) for undirected networks, and the percolation threshold (i.e., a critical point marking the appearance of a cluster with size comparable to the original number of nodes). Therefore, please refer to Ch. 2 for further introductory material on these topics.

4.2 Social Climber Attachment

A *link-percolation process* begins with N isolated nodes, indexed $n = 1, 2, \ldots, N$. In discrete steps $\hat{t} = 1, 2, \ldots$, a new undirected link between two nodes is selected according to a rule or set of rules, and is then formed. Thus, after $\hat{t}$ steps there will be $\hat{t}$ links in the network, resulting in

clusters of connected nodes, each of whose size (number of nodes in the cluster) may range from one (an isolated node) to N (a cluster that spans the entire network). Depending on the rules used to select links, the evolution of cluster sizes, and in particular the size of the largest cluster, may vary significantly. Social Climber attachment introduces a new link reselection step between link selection and link formation, which we motivate by analogy to a corresponding social process: colloquially, a "social climber" is someone who actively attempts to make powerful friends in order to become more powerful himself. When introduced to a new person, a social climber learns about the relative popularity of the people in that person's clique and eventually befriends whoever is of maximal importance.

With this in mind, SC attachment is a link reselection step during percolation where the proposed link between two nodes is altered by allowing one of those nodes to act like a social climber, choosing to link to the node of maximal importance in the other node's cluster. Therefore, given a link-percolation process, we summarize SC attachment as follows:

- (i) Let x be a proposed undirected link connecting nodes a and b, generated by an arbitrary percolation model. The labeling of nodes a and b is random to account for asymmetric link-percolation processes.

- (ii) Let clusters C_a and C_b be the clusters to which nodes a and b belong, respectively. Then, if $C_a \neq C_b$ the proposed link x is discarded and instead a link y is made between node a and the largest-degree node in C_b, as shown in Fig. 4.1a.

- (iii) If nodes a and b belong to the same cluster, $C_a = C_b$, then the proposed link x is made without modification.

Note that SC attachment does not affect which clusters combine, but does affect the topology of the resulting joined cluster. The SC model chooses a connection to the node of largest degree in a cluster using nodal degree as a proxy for the *dynamical importance* measure, $DI = u_n v_n$ [121], where u and v denote the right and left eigenvectors of A corresponding to λ. For the undirected networks considered here, symmetry of A implies $u = v$, and thus the node with largest eigenvector

entry u_n will have maximal DI in its cluster [121]. We allow the SC model to select nodes based on degree for simplicity and ease of computation. One may equivalently view SC attachment as forming a link to the node with largest degree by using degree centrality as a proxy for eigenvector centrality. Also note that this approach is in good agreement with our results developed in Ch. 3. Specifically, in Sec. 3.3 we showed that for a network resulting from the merging of large and small networks, λ could be maximized by using the smaller network as a bridge between nodes n in the larger network corresponding to large v_n and u_n. SC attachment therefore approximately maximizes λ for the merging of clusters by using that nodes with large u_n are expected have large degree, d_n.

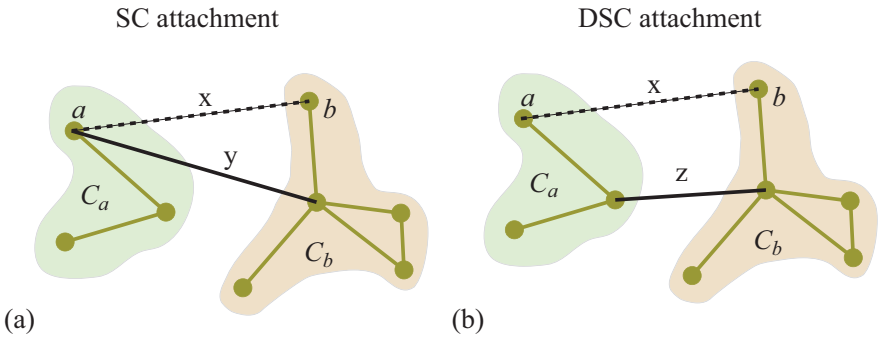


Figure 4.1: Link x is proposed by an arbitrary percolation model (dashed lines) to connect nodes a and b, merging clusters C_a and C_b. Because $C_a \neq C_b$ the proposed link x is discarded and instead one of the following new links is formed. (a) Node a is linked to the largest-degree node in C_b with link y to model SC attachment. (b) The largest-degree nodes from C_a and C_b are linked together with link z to model DSC attachment. These processes may be visualized using free PercoVIS software [79].

In addition to SC attachment, we introduce *Double Social Climber* (DSC) attachment, in which a proposed link between nodes a and b is either replaced by a link between the nodes with maximal degree in each cluster when $C_a \neq C_b$, as shown in Fig. 4.1(b), or is formed between a and

b without modification when $C_a = C_b$. We note that DSC attachment corresponds to maximizing connectivity of the resulting cluster, as measured by λ, whenever the node of maximal degree is also the node of maximal eigenvector entry in each cluster [144]. SC and DSC attachment may be visualized using free PercoVIS software [79].

4.2.1 Analysis for the Largest Eigenvalue

Although we will later generalize our methods to other percolation rules, we first analyze SC and DSC attachment for the well-known Erdős-Rényi (ER) percolation process [55]. The rule for selecting a link in ER percolation is simple: two nodes are chosen uniformly at random and a link is formed if there is not already a link between them. Traditional analysis has focused on the relationship between the number of links added $\hat{t}$ and the size of the largest cluster $\hat{G}(\hat{t})$, called the *giant component* (GC) when $\hat{G}(\hat{t}) \sim \mathcal{O}(N)$ (recall Sec. 2.1). It is convention to rescale both $\hat{t}$ and $\hat{G}$ by N [i.e. $t = \hat{t}/N$ and $G(t) = \hat{G}(t)/N$], where one obtains in the asymptotic limit $N \to \infty$ [55],

$$G(t) = \begin{cases} 0 & , \quad t \leq 0.5 \\ 1 - e^{-2tG(t)} & , \quad t > 0.5 \end{cases} . \tag{4.1}$$

Here, for variable control parameter t, the network undergoes a second order phase transition in cluster size at the percolation threshold $t_c^{ER} = 0.5$, as observed through the order parameter $G(t)$. Because SC attachment affects the topology of clusters and not their sizes, Eq. (4.1) remains valid for ER percolation combined with SC attachment.

We begin our analysis by studying the emergence of large-degree nodes. For a given time t, consider a large cluster C containing $s \gg 1$ nodes, and let $k^{\max}$ denote the maximal nodal degree in C. By large cluster, we mean that s is with high probability larger than the size of another randomly chosen cluster in the network, and eventually we will consider only the case in which cluster C is the largest cluster in the entire network. We will compute the expected change in $k^{\max}$ for the addition of a single link. When a link is proposed between nodes a and b by ER percolation, $k^{\max}$ will increase by one if: (i) $a \notin C$ and $b \in C$ (depicted in Fig. 4.1a where $C_b = C$), or (ii) $a, b \in C$ and the degree of a or b is $k^{\max}$. Since ER percolation chooses nodes uniformly at

random, the probability that $a \notin C$ is $1 - s/N$, and the probability that $b \in C$ is s/N. Since a and b are chosen independently, the probability of case (i) is $(s/N)[1 - s/N]$. The probability that a randomly chosen node in a cluster of size s has degree $k^{\max}$ is r/s, where r is the number of nodes in that cluster with degree $k^{\max}$. Thus, the probability of case (ii) is, to leading order as $N \to \infty$, $(s/N)[1 - s/N](r/s) + (s/N)^2[2r/s]$, where the first term corresponds to $a \in C$, $b \notin C$ and the second term corresponds to $a, b \in C$. We note that other corrections may be included to address the chance in (i) that the maximal degree of C_a is larger than the maximal degree of C_b, $k^{max} + 1$, but such corrections decay rapidly as the size difference between C_b and C_a increases. Since $s \gg 1$ while $r \sim \mathcal{O}(1)$, case (i) is the dominating process for sufficiently large s, so the expected rate of change of $k^{\max}$ averaged over all possible links is

$$E\left[\frac{dk^{\max}}{d\hat{t}}\right] = \zeta \frac{s}{N}\left(1 - \frac{s}{N}\right). \tag{4.2}$$

Here, $\zeta = 1$ for SC attachment and $\zeta = 2$ for DSC attachment, since for DSC case (i) applies to both $a \notin C$, $b \in C$ and $a \in C$, $b \notin C$. Using $d/d\hat{t} = Nd/dt$, integration of Eq. (4.2) predicts that the largest degree of a node within the GC at time t is

$$E\left[k_{ER}^{\max}(t)\right] = \zeta N \int_0^t \left[G(\tau) - G^2(\tau)\right] d\tau. \tag{4.3}$$

Note that this scaling with N is the largest achievable scaling of a degree. For comparison, in networks with power-law degree distribution, $P(k) \propto k^{-\gamma}$, the expected maximal degree scales as $\mathcal{O}(N^{1/(\gamma-1)})$, approaching $\mathcal{O}(N)$ as $\gamma \to 2^{+}$ [1] .

In order to understand the implications of Eq. (4.3), we use $\lambda \approx \sqrt{k^{\max}}$, an asymptotically ($N \to \infty$) accurate approximation derived in [39, 59] and discussed further in [125]. While the model used to generate this estimate is not equivalent to SC attachment, we find it remains accurate here. Using this estimate in conjunction with Eq. (4.3), and noting that for $t > t_c^{ER}$, the largest eigenvalue of the GC will be larger than the largest eigenvalues of smaller clusters, we obtain the following

[1] For example, this follows from finding $k^{\max}$ such that the probability a node is larger than or equal to $k^{\max}$ is equal to N^{-1}, or $N^{-1} = \int_{k^{\max}}^{\infty} P(k)dk$.

expression for the expected largest eigenvalue for ER percolation with SC attachment,

$$E\left[\lambda_{ER}(t)\right] = \sqrt{\zeta N \int_0^t \left[G(\tau) - G^2(\tau)\right] d\tau}, \tag{4.4}$$

implying that the network undergoes a continuous phase transition in connectivity (i.e., as measured by the magnitude of λ) at precisely the same value $t = t_c^{ER}$ at which a phase transition in cluster size occurs. This result has additionally assumed that $E[\sqrt{k^{\max}}] = \sqrt{E[k^{\max}]}$, which is also believed to be asymptotically accurate as $N \to \infty$ due to the convergence of the $\lambda \approx \sqrt{k^{\max}}$ estimate. In the super-critical regime, λ achieves maximal scaling for a network containing $\mathcal{O}(N)$ nodes, $\lambda \sim \mathcal{O}(N^{1/2})$. For comparison a wheel network, which consists of a single node linked to the remaining $N-1$ nodes via $N-1$ links, has similar scaling, $\lambda = \sqrt{N-1}$.

One value of particular interest is the asymptotic scaling constants of Eq. (4.4) for large t, which may be found by examining $\lim_{t\to\infty} \lambda_{ER}(t)/\sqrt{N}$. These constants may be solved by integrating with respect to G,

$$\lim_{t\to\infty} \int_0^t G(\tau) - G^2(\tau) d\tau = \int_0^1 G(1-G) \left(\frac{d\tau}{dG}\right) dG, \tag{4.5}$$

and using that $\tau = \frac{\ln(1-G)}{-2G}$ and $\frac{d\tau}{dt} = \frac{\ln(1-G)}{2G^2} + \frac{1}{2G(1-G)}$. Finally, using that $-\int_0^1 \frac{\ln(1-G)}{G} dG = \mathrm{Li}_2(1) = \frac{\pi^2}{6}$, where $\mathrm{Li}_2(\cdot)$ is the dilogarithm function [72], one obtains $\lambda_{ER}(t)/\sqrt{N} \to \sqrt{1-\pi^2/12} \approx 0.42$ for SC attachment and $\lambda_{ER}(t)/\sqrt{N} \to \sqrt{2-\pi^2/6} \approx 0.6$ for DSC attachment.

Scaling with network size N is also of particular interest. Because the expected connectivity, measured by λ_{ER}, is a function of $G(\cdot)$, developing scaling arguments for λ_{ER} is straightforward since scaling for $\hat{G}(t)$ is known: $\hat{G} \sim \log(N)$ for $t < t_c^{ER}$ and $\hat{G} \sim N$ for $t > t_c^{ER}$ [55]. Therefore, when SC is used in conjunction with ER percolation, in the limit $N \to \infty$ we have

$$\begin{aligned} \lambda_{ER}(t) &\sim \sqrt{\log N} && t < t_c^{ER}, \\ \lambda_{ER}(t) &\sim \sqrt{N} && t > t_c^{ER}. \end{aligned} \tag{4.6}$$

The methods used to derive Eqs. (4.3)-(4.6), which involved calculating the probability that an isolated cluster attaches to the GC, may be easily adapted to other percolation models. For

example, consider Achlioptas processes [2, 21, 54, 138] for which the merging of clusters depends on cluster size (up to some bound). This class of percolation models has recently received much attention, focusing on analysis of a rapid phase transition in cluster size referred to as "explosive percolation" [2, 54]. In particular, we also study Adjacent-Edge (AE) link percolation [54], which is designed to delay the formation of large clusters via the following process: (i) three nodes are chosen uniformly at random, labeled n_1, n_2, and n_3; (ii) an undirected link is then formed between node n_1 and the node $n \in \{n_2, n_3\}$ that is contained in the smaller cluster; and (iii) the remaining node is unmodified and this process is iterated. Repeating the reasoning process in deriving Eqs. (4.3)-(4.6) for AE percolation, we predict the largest eigenvalue to be

$$E[\lambda_{AE}(t)] = \sqrt{\frac{\zeta N}{2} \int_0^t [G(\tau) + G^2(\tau) - 2G^3(\tau)]\, d\tau}, \tag{4.7}$$

where again $\zeta = 1$ for SC and $\zeta = 2$ for DSC. Note that despite the near-discontinuous phase transition in G [54], maximal scaling $\lambda \sim \mathcal{O}(N^{1/2})$ is still achieved.

4.2.2 Numerical Experimentation

In the numerical examples to follow we will (i) explore the significant effect SC and DSC attachment has on forming networks, (ii) validate the accuracy of our analysis, (iii) show that a variation of our model restricted to local information also yields networks with large λ, and (iv) study the evolution of λ for the formation of a real-world network.

The Effect of SC Attachment on the Largest Eigenvalue

In this first experiment, we illustrate that DSC attachment leads to networks with very strong connectivity, as indicate by very large λ. We also show that for networks forming under link percolation with DSC attachment, a phase transition as observed through λ, occurs simultaneously with the phase transition as observed through the size $G(t)$ of the network's giant component.

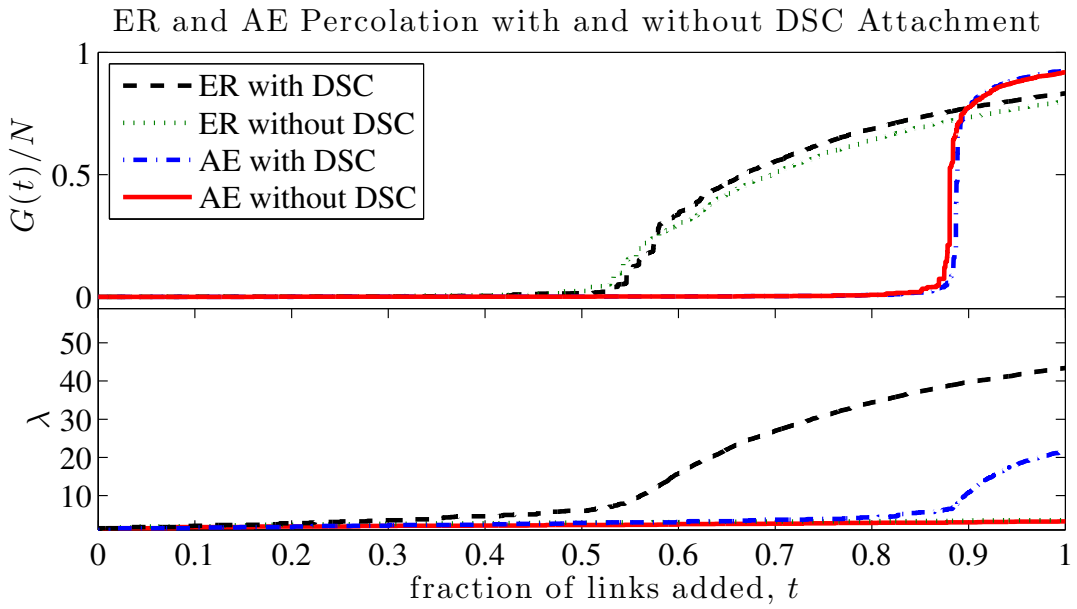


Figure 4.2: (upper panel) The relative size of the giant component is shown for ER and AE link percolation on $N = 10^4$ nodes, both with and without DSC attachment. Note that DSC attachment does not affect the size of a forming network's cluster or its largest cluster. Also note that the transition is much more rapid for AE percolation. (lower panel) The largest eigenvalue of the forming network's adjacency matrix is also shown for ER and AE link percolation. Note that a phase transition may be observed through λ at the same critical t value as the emergence of the giant component for these percolation processes in combination with DSC attachment. However, in the absence of DSC attachment, λ remains small and is largely unaffected by the emergence of the giant component.

In Fig. 4.2 we show numerically observed values of the size of the giant component $G(t)$ and the largest eigenvalue of the network adjacency matrix λ for networks forming under link percolation, both with and without DSC attachment. These are shown for both ER link percolation and AE link percolation, which respective undergo phase transitions at their respective percolation thresholds $t_c^{ER} = 0.5$ and $t_c^{AE} \approx 0.766$. For these processes with DSC attachment, a phase transition may be observed through both $G(t)$ and $\lambda(t)$, and we say that the networks undergo phase transitions both in cluster size and connectivity. For these processes without DSC attachment, a

phase transition only occurs in cluster size as λ is largely unaffected by the emergence of the giant component.

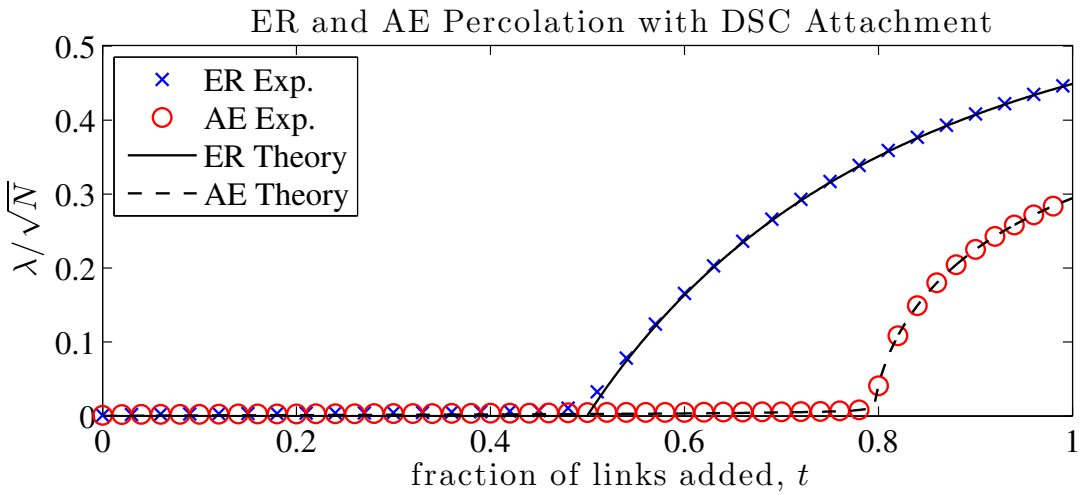


Figure 4.3: Predictions for λ [Eq. (4.4) (solid line) and Eq. (4.7) (dashed line)] are shown to be in excellent agreement with observed values for ER and AE percolation with $N = 10^6$ nodes with DSC attachment [x symbols and circles].

Largest Eigenvalue Predictions

In the next examples, we confirm the accuracy of Eq. (4.4) by direct simulation of our model, shown in Fig. 4.3, which demonstrates excellent agreement between Eq. (4.4) (solid line) and observed values of λ_{ER} for ER percolation with DSC attachment with $N = 10^6$ (X symbols). For integration in Eq. (4.4) we use the asymptotic theoretical value $G(t)$ given by Eq. (4.1). In Fig. 4.3 we also show good agreement between observed values for λ_{AE} (circles) and Eq. (4.7) for DSC (dashed line), where observed values for $G(t)$ were used in Eq. (4.7) as an analytic expression has yet to be developed. Note that $\lambda_{AE}(t) < \lambda_{ER}(t)$, which we attribute to the integrands of Eqs. (4.4) and (4.7), which are maximized at $G = 1/2$ and $G = (1 + \sqrt{7})/6$, respectively, and are

zero at $G = 0$ and $G = 1$. Since AE percolation produces rapid growth in G, the integrand of Eq. (4.7) is not large over a majority of the integration interval, so essentially, the explosive growth in G minimizes the regime during which SC attachment has a large effect on λ.

In Fig. 4.4 we illustrate the accuracy of of Eq. (4.4) for both SC and DSC attachment. Symbols indicate observed values for $\lambda(t)$ for ER link percolation with $N = 10^6$ nodes, while the lines indicate their expected values given by Eq. (4.4).

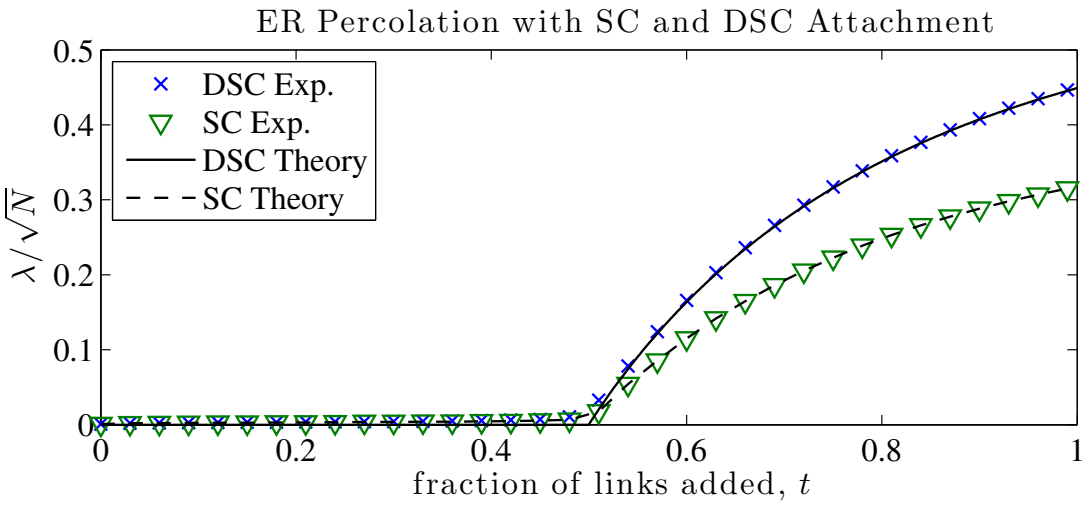


Figure 4.4: The largest eigenvalue of the network adjacency matrix λ is shown for networks forming under ER link percolation, either with SC or DSC attachment. Note the observed values (symbols) are in good agreement with our theory, given by Eq. (4.4) (lines).

Scaling Predictions

In this example we illustrate the accuracy of Eqs. (4.6), which highlights that a phase transition occurs as observed through the order parameter λ_{ER} at t_c^{ER} as it moves from a regime in which $\lambda_{ER} \sim \sqrt{\log(N)}$ to a regime in which $\lambda_{ER} \sim \sqrt{N}$, where N is the number of nodes in the network. To illustrate the accuracy of this scaling argument, we first show that as network

size increases, the system asymptotically approaches our theory given by Eq. (4.4). In Fig. 4.5 we plot our prediction, Eq. (4.4) (solid line) versus observed values of λ_{ER} for several network sizes, $N = 10^4$ (green crosses), $N = 10^5$ (red squares), and $N = 10^6$ (blue x symbols). As expected, $\lambda_{ER}/\sqrt{N}$ is well approximated by Eq. (4.4) for $t > t_c^{ER}$, and $\lambda_{ER}/\sqrt{N} \to 0$ as $N \to \infty$ for $t < t_c^{ER}$. Eq. (4.6) predicts that this convergence is given by $\sqrt{\log(N)/N}$.

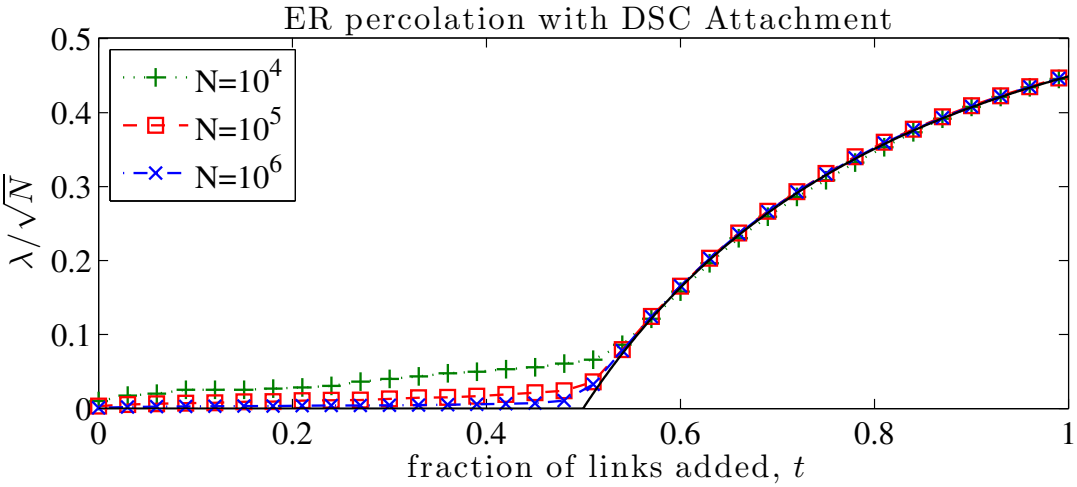


Figure 4.5: For $t > t_c^{ER} = 0.5$, $\lambda_{ER}/\sqrt{N}$ is well approximated by Eq. (4.4). For $t < t_c^{ER}$, $\lambda_{ER}/\sqrt{N} \to 0$ as $N \to \infty$, which is predicted to converge as $\sqrt{\log(N)/N}$.

To validate Eq. (4.6) we estimate the change in λ_{ER} when system size N is increased by defining $\phi_{ER}(t)$ as the ratio of $\lambda_{ER}(t)$ for $N = 10^6$ to $\lambda_{ER}(t)$ for $N = 10^5$. As shown in Fig. 4.6, we predict that $\phi_{ER}(t) \approx \sqrt{6/5}$ for $t < t_c^{ER}$ (dashed line) and $\phi_{ER}(t) \approx \sqrt{10}$ for $t > t_c^{ER}$ (solid line), both of which agree well with $\phi_{ER}(t)$ calculated from a single simulation of each system size (x symbols). In Fig. 4.6 we also show $\phi_{AE}(t)$, the ratio of $\lambda_{AE}(t)$ for $N = 10^6$ and $N = 10^5$, where we observe similar scaling in the subcritical and supercritical regimes as observed and predicted for λ_{ER}. Note that as expected, the shifts between scaling regimes for ER and AE percolation with DSC attachment are found to respectively occur at t_c^{ER} and t_c^{AE}.

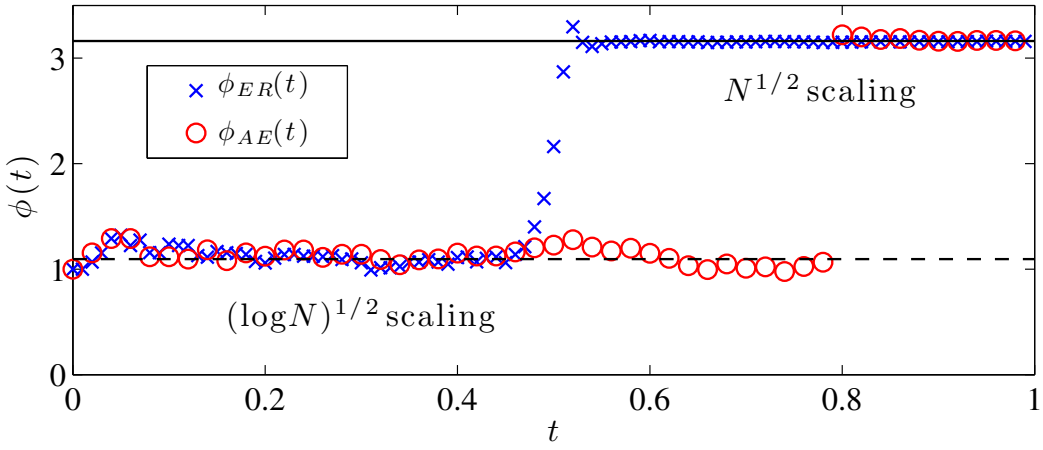


Figure 4.6: Scaling of $\lambda_{ER}(t)$ predicted in Eq. (4.6) is demonstrated using the ratios of $\lambda_{ER}(t)$ for $N = 10^6$ and $N = 10^5$, denoted $\phi_{ER}(t)$. Agreement between the prediction of Eq. (4.6) (lines) and measurement from simulation (crosses) is good. Measurements from AE percolation, $\phi_{AE}(t)$, are also shown (circles).

Local Social Climber Attachment

In the next example, we illustrate that a rapid increase in λ still occurs even if Social Climber attachment is restricted to a local neighborhood. Specifically, we introduce *Local Social Climber* (LSC) attachment, which is similar to SC attachment except that during the link re-selection step, instead of linking the node a in the first cluster, C_a, to the node in the second cluster, C_b, with maximum degree, node a is instead linked to the node in the second cluster C_b that has maximum degree and is in a neighborhood $\mathcal{N}_r(b)$ of geodesic radius r that is centered at the original target node b in the second cluster C_b, defined as $\mathcal{N}_r(b) = \{n \in C_b | b \leftrightarrow n$ for some path of length $p \leq r\}$.

In Fig. 4.7, we plot λ for a network with $N = 10^4$ nodes forming under ER percolation with LSC attachment for several radii. In all cases we find $\lambda \sim \mathcal{O}(\sqrt{\log N})$ for $t < t_c^{ER}$. However, for $t > t_c^{ER}$ we find that large values of λ occur even for very small radii r. For example, for $N = 10^4$ with $r = 3$, λ reaches values on the order of what is reached without the local restriction. This experiment highlights that effective methods for generating networks with very large λ may be

developed even for applications in which global information is not feasible.

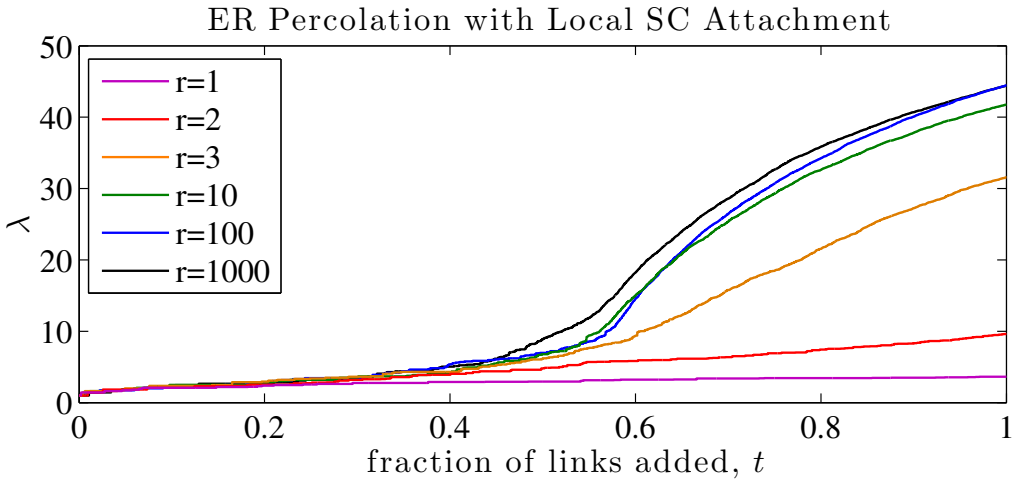


Figure 4.7: The largest eigenvalue λ is shown for $N = 10^4$ nodes undergoing ER percolation with LSC attachment for various radii r. Note that very large λ can still be achieved even for small r.

The Largest Eigenvalue of a Forming Co-Authorship Network

In the experiments so far we have investigated the largest eigenvalue λ of the adjacency matrix for networks forming under various theoretical link percolation processes. In this example, we highlight that that this technique may also be useful for studying the structure of real-world networks that evolve. Specifically, we will study the structure of a forming network in which links indicate scientific co-authorships between authors who've published articles in the field of condensed matter on the arXiv [1] during the years of 1992-1995. Importantly, the actual sequential order of link appearance is utilized.

A co-authorship network was generated by examining the open access publications in the field of condensed matter on the arXiv [1] during the years of 1992-1995. To generate the network, we began with an empty network without nodes or links. Considering each publication in sequence,

nodes were added for each author that lacked a corresponding node and links were added between all pairs of authors for a given publication. Nothing was done if a given link already existed from a prior publication. This processes was iterated for all publications until the full network was obtained, which contained $N = 6432$ nodes and $M = 19867$ links. To map this network formation to a link percolation process, we considered the set of nodes corresponding to the full network in isolation, without links. Then, in enumerated steps $l = 1, 2, \ldots$ we added the links in sequential order until the full network was recovered. A random link order was chosen for publications yielding more than one link.

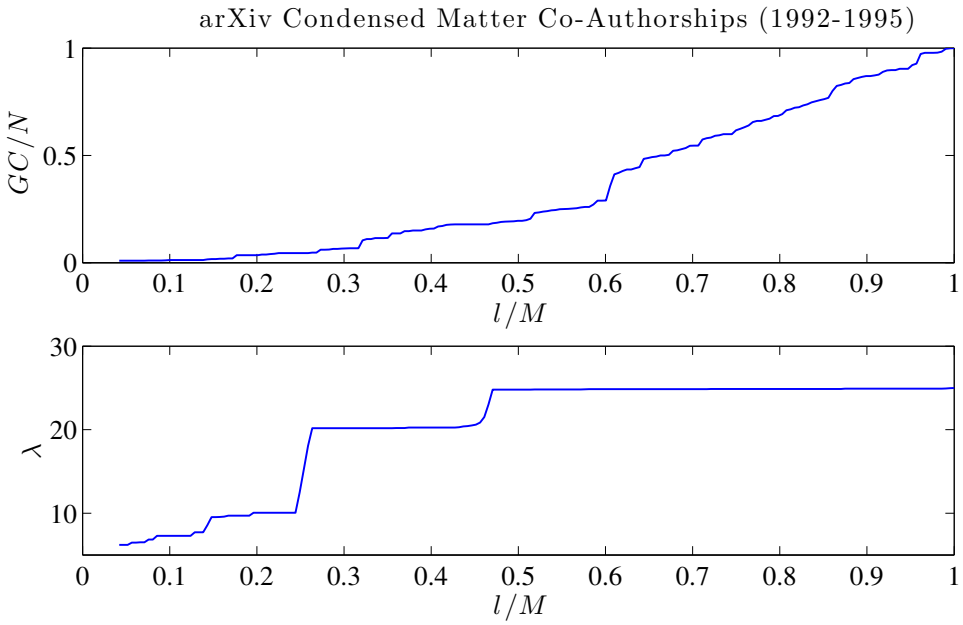


Figure 4.8: The sequential formation of a co-authorship network is shown for condensed matter articles on the arXiv from 1992-1995. (upper panel) The size GC of the giant component is plotted as the number of co-authorship links, l, increases. (lower panel) The largest eigenvalue of the network's adjacency matrix is found to undergo rapid jumps, indicating that network restructuring may be occurring that is not captured by the cluster sizes.

In Fig. 4.8 we depict this network formation process. In the top panel we show the size GC of the giant component versus the fraction of links added. Note that unlike the theoretical link percolation models studied in this chapter, no phase transition may be observed through the GC as asymptotically one expects $q^* \to 0$ for networks with heavy-tailed degree distributions. In the lower panel we plot the largest eigenvalue λ of the network's adjacency matrix. Unexpectedly, sharp jumps may be observed, which suggest that the network's topology is undergoing changes that can't be captured by observing the cluster sizes. This experiment supports our claim that λ can be a very useful alternative order parameter for studying the evolving structure of real-world networks. This example is preliminary and will be further explored in future research.

4.2.3 Analysis for the Distributed Characteristics of Clusters

In this section we develop analysis to describe the distributed properties of clusters that appear in networks forming under ER and AE percolation, either with or without DSC attachment. In the absence of DSC attachment, we analyze the joint distribution $\rho_{k,d}$ that describes the probability that a randomly selected node will be in a cluster of size k and will have a degree given by d. In the presence of DSC attachment, we analyze the joint distribution $\eta_{k,\delta}$ that describes the probability that a randomly selected node will be in a cluster of size k and the maximum degree of the nodes in that cluster is given by δ. Our analysis is shown to be in good agreement with previous results describing the probability x_k that a randomly selected node with be in a cluster of size k, and therefore this work may be thought of as a generalization from distributions for cluster size to joint distributions for cluster size and connectivity.

We first extend previous results for the evolution of cluster size distributions [54, 138] to the evolution of joint distributions describing cluster sizes and nodal degrees. This is done for both ER [55] and AE [54] link percolation. Beginning with an ER process, in which two nodes are randomly chosen and an undirected link is formed between them, we find the following evolution equation

for $\rho_{k,d}$,

$$\Delta\rho_{k,d} \quad = \quad -2k\rho_{k,d} + 2\sum_{j=1}^{k-1} x_{k-j}[(j-1)\rho_{j,d} + \rho_{j,d-1}], \tag{4.8}$$

which represents a mean-field analysis over all graph evolutions and may be interpreted as the expected change in $\rho_{k,d}$ after the addition of a link. Here the terms represent the following probabilities: (i) either the first or second randomly-selected node is of type (k,d), which happens if any of the k nodes in the cluster are selected; and (ii) a node of type (k,d) is formed by merging clusters of size $k-j$ and j. Note that the node in each cluster receiving the link increases its degree by one (i.e., from $d-1$ to d), whereas the degrees of the remaining $j-1$ nodes remain constant. We point out that since the probability that a randomly selected node is in a cluster of size k is given by $x_k = \sum_d \rho_{k,d}$, summing Eq. (4.8) over d recovers the evolution equation for x_k [138],

$$\Delta x_k = -2kx_k + k\sum_{j=1}^{k-1} x_j x_{k-j}. \tag{4.9}$$

Turning to AE percolation, in which a node and two candidate nodes are randomly selected and a link is formed between the first node and the candidate node in the smaller cluster, we find the following evolution equation for $\rho_{k,d}$,

$$\begin{aligned}\Delta\rho_{k,d} = &-k\rho_{k,d} - k\rho_{k,d}(2s_k - x_k) + \sum_{j=1}^{k-1} x_{k-j}[(j-1)\rho_{j,d} + \rho_{j,d-1}](2s_j - x_j) \\ &+ \sum_{j=1}^{k-1} x_{k-j}(2s_{k-j} - x_{k-j})[(j-1)\rho_{j,d} + \rho_{j,d-1}].\end{aligned} \tag{4.10}$$

Here $x_k = \sum_d \rho_{k,d}$ again denotes the probability that a randomly selected node is in a cluster of size k and $s_k = 1 - \sum_{j<k} x_j$ denotes the probability that a node is in a cluster of size greater or equal to k. From left to right, the terms correspond to the following probabilities: (i) the first node is type (k,d); (ii) the second node is type (k,d), which for AE link percolation requires that a third candidate node is in a cluster that is larger than or equal to k [54]; (iii) the probability that the first node is in a cluster of size $k-j$ and the second node is either type (j,d) or type $(j,d-1)$; and (iv) the probability that the second node is in a cluster of size $k-j$ and the first node is either type (j,d) or type $(j,d-1)$. Note that for $j-1$ nodes their degree d is unchanged, whereas for a the

node that receives the link, its degree increases from $d-1$ to d. As expected, summing Eq. (4.10) over d recovers the the evolution equation for x_k [54],

$$\Delta x_k = -k(x_k + s_k^2 - s_{k+1}^2) + k\sum_{j=1}^{k-1} x_{k-j}(s_j^2 - s_{j+1}^2). \tag{4.11}$$

We now study ER and AE percolation with DSC attachment. Letting $\eta_{k,d}$ denote the probability that a randomly selected node will be in a cluster of size k with maximal degree δ, for ER percolation with DSC attachment we find

$$\Delta\eta_{k,\delta} = -2k\eta_{k,\delta} + k\sum_{j=1}^{k-1}(\eta_{k-j,\delta-1}\eta_{j,\delta-1}) + 2k\sum_{j=1}^{k-1}(\eta_{k-j,\delta-1}y_{j,\delta-1}), \tag{4.12}$$

where $y_{j,\delta} = \sum_{e=0}^{\delta-1}\eta_{j,e}$ denotes the probability that a node is in a cluster of size j and has a maximum degree less than δ. From left to right the terms represent the following: (i) the probability that one of the two nodes selected to receive the link is type (k,δ); (ii) the probability that the two nodes are respectively type $(j,\delta-1)$ and $(k-j,\delta-1)$; and (iii) the probability that one node is type $(k-j,\delta-1)$ and the other node is type (j,e), where $e<\delta-1$. The factor of two accounts for either the maximum degree of the first cluster being smaller than that of the second cluster, or vice versa. Also note that all k nodes increase their cluster's maximum degree and cluster size upon merging to a cluster of size k.

For AE percolation with DSC attachment we instead find

$$\begin{aligned}\Delta\eta_{k,\delta} = &-k\eta_{k,\delta} - k\eta_{k,\delta}(2s_k - x_k) + k\sum_{j=1}^{k-1}\eta_{k-j,\delta-1}y_{j,\delta-1}(2s_j - x_j)\\ &+ k\sum_{j=1}^{k-1}\eta_{j,\delta-1}(2s_j - x_j)y_{k-j,\delta-1} + k\sum_{j=1}^{k-1}\eta_{k-j,\delta-1}\eta_{j,\delta-1}(2s_j - x_j),\end{aligned} \tag{4.13}$$

where the terms represent the following probabilities: (i) the first node is type (k,δ); (ii) the second node is type (k,δ), which again for AE link percolation requires that a third candidate node is in a cluster with size larger than or equal to k; (iii) the first node is type $(k-j,\delta-1)$ and the second node is type (j,e) for any $0\le e<\delta-1$; (iv) the second node is type $(j,\delta-1)$ and the first node is type $(k-j,e)$ for any $0\le e<\delta-1$; and (v) the first node is type $(k-j,\delta-1)$ and the second node is type $(j,\delta-1)$. The agreement of Eqs. (4.12) and (4.13) with previous results and direct numerical simulations will be numerically shown in the following section.

4.2.4 Numerical Experimentation

In this section we illustrate the accuracy of Eqs. (4.8)-(4.13) by comparing our predictions for the distributions $\rho_{k,d}$ and $\eta_{k,\delta}$ with distributions observed from directly simulating the link percolation processes. Recall that these predictions, and hence the simulations to follow, reflect the connectivity of clusters in the subcritical regime of percolation. The experiments to follow will explore (i) the accuracy of our results for ER and AE link percolation without DSC attachment and (ii) with DSC attachment.

ER and AE percolation without DSC attachment

Simulations of ER and AE link percolation without DSC attachment were run for networks with $N = 10^6$ nodes until either $l = 0.5N$ or $l = 0.75N$ undirected links were added, respectively (i.e., $t = 0.5$ or 0.75). At that time the cluster properties were recorded and compared with our predictions given by numerically integrating our evolution equations with a step size $dt = N^{-1}$. Note that all results reflect cluster statistics in the subcritical regime. In Fig. 4.9 we plot observed values of $\rho_{k,d}$ (symbols) versus our theoretical predictions given by numerically integrating either (a) Eq. (4.8) for ER link percolation or (b) Eq. (4.10) for AE link percolation. Note the good agreement between theory and observation of $\log(\rho_{k,d})$ for values greater than $\log(N^{-1}) \approx -13$ (the smallest observable value for a finite network with N nodes). Also note that the distributions shown in Fig. 4.9(a) and 4.9(b) appear to be very similar, indicating that at this scale the structures are quite similar for clusters forming under AE and ER percolation. However, we note that structural differences are expected for large clusters such as the giant components [42].

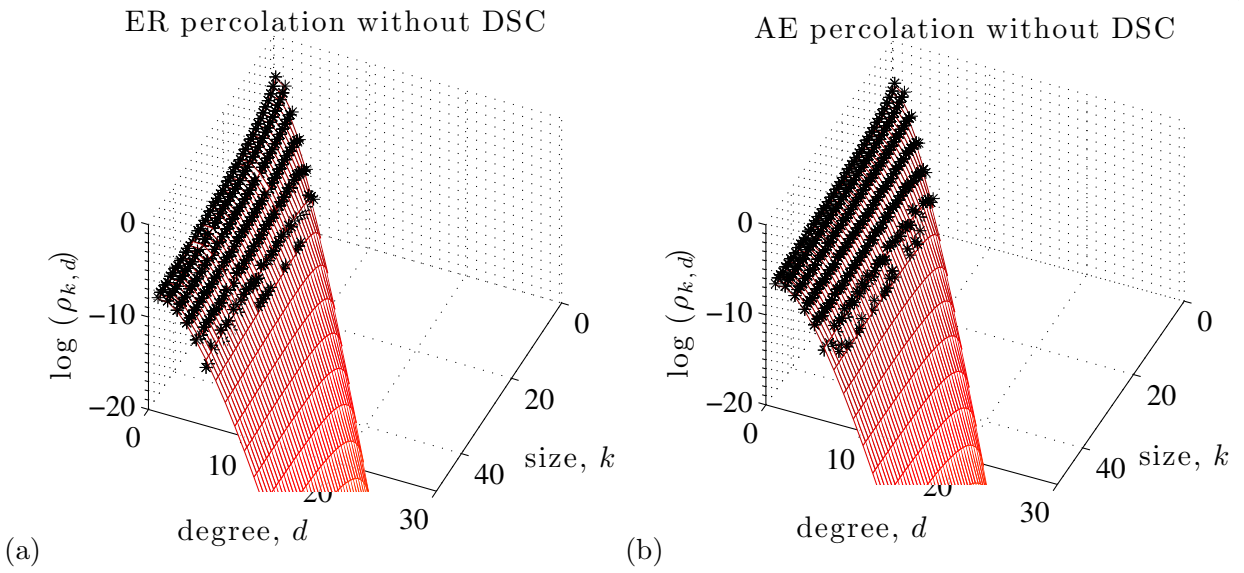
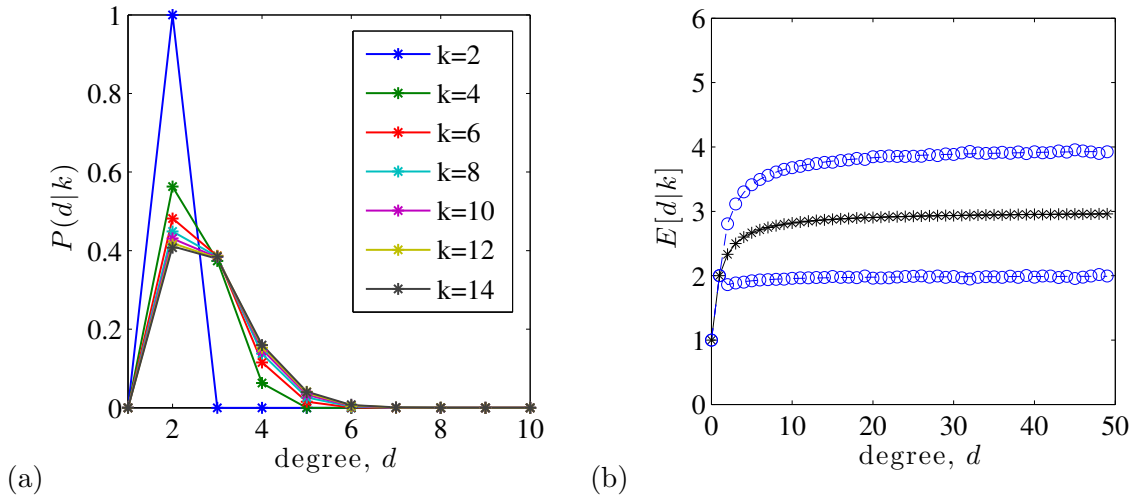


Figure 4.9: Theoretical (surface plots) and observed (symbols) values for $\rho_{k,d}$ are shown to be in good agreement for (a) ER percolation and (b) AE percolation without DSC attachment with $N = 10^6$ nodes. Predicted values were obtained by numerically integrating Eq. (4.8) and Eq. (4.10), respectively, with a time step $dt = N^{-1}$.

To better illustrate the good agreement between the observed and predicted values of $\rho_{k,d}$ for ER and AE percolation, we now present results for the conditional distribution $P(d|k)$ and expected value $E[d|k]$ describing the probability that a randomly selected node in a cluster of size k will have degree d,

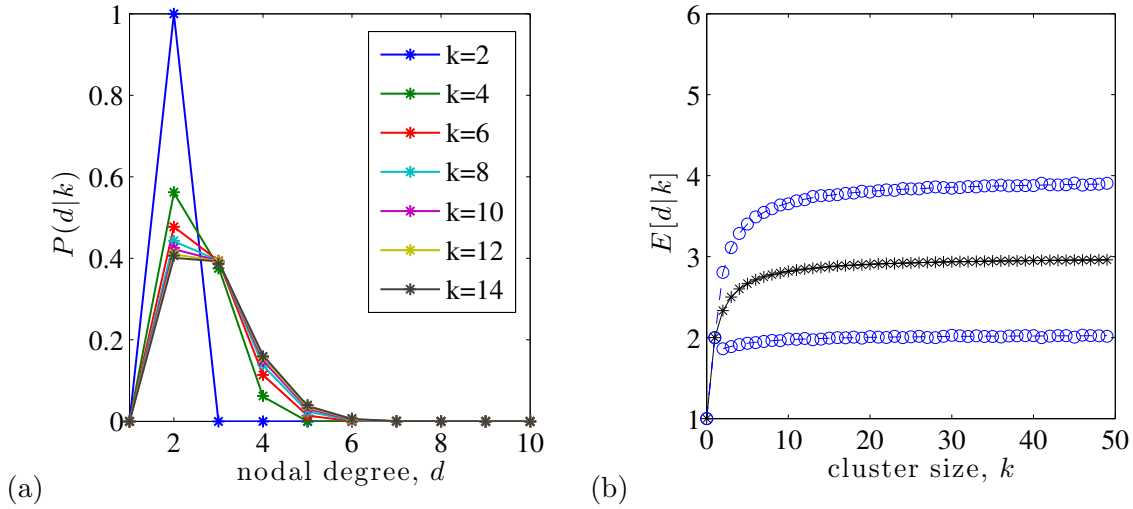
$$P(d|k) = \frac{\rho_{k,d}}{\sum_d \rho_{k,d}}, \qquad E[d|k] = \sum_d P(d|k)d. \tag{4.14}$$

In Fig. 4.10 we show statistics for the connectivity of clusters forming under ER percolation without DSC attachment. In Fig. 4.10(a) we plot $P(d|k)$ for several values k. Observed values (symbols) are shown to agree with our theory using Eq. (4.8) (lines). In Fig. 4.10(b) we plot the expected degree of a node as a function of cluster size k. The black stars and solid line respectively indicate observed values and predicted values for $E[d|k]$, whereas the blue circles and dashed lines respectively indicate the standard deviations obtained from simulation and our theory. Note that for sufficiently large clusters (e.g., $k > 20$) it appears that the statistics limit to a specific distribution with mean $E[d|k] \approx 3$ and standard deviation $\sigma[d|k] \approx 1$.



(a) (b)

Figure 4.10: ER percolation without DSC attachment at $t = 0.5$ for $N = 10^{-6}$. (a) The distribution of nodal degrees in a given cluster of size k, $P(d|k)$. (b) The expected degree of a node in a cluster of size k, $E[d|k]$ (black stars and solid line), and the standard deviation, $\sigma[d|k]$ (blue circles and dashed lines). In both plots predicted and observed values are indicated by lines and symbols.



(a) (b)

Figure 4.11: AE percolation without DSC attachment at $t = 0.75$ for $N = 10^{-6}$. (a) The distribution of nodal degrees in a given cluster of size k, $P(d|k)$. (b) The expected degree of a node in a cluster of size k, $E[d|k]$ (black stars and solid line), and the standard deviation, $\sigma[d|k]$ (blue circles and dashed lines).

In Fig. 4.11 we show statistics for the connectivity of clusters forming under AE percolation without DSC attachment. In Fig. 4.11(a) we plot the distribution of nodal degrees in clusters of size k. Observed values (symbols) are shown to agree with our theory using Eq. (4.10) (lines) for several cluster sizes. In Fig. 4.11(b) we plot the expected degree of a node as a function of the node's cluster size k. The black stars and solid line respectively indicate observed values and predicted values for $E[d|k]$. The blue circles and dashed lines respectively indicate the standard deviations obtained from simulation and our theory. Besides showing good agreement, these plots also suggest that the distribution $P(d|k)$ converges to a distribution as $k \to \infty$, similar to that observed for ER percolation in Fig. 4.10. In fact Fig. 4.10 and Fig. 4.11 almost appear identical for the cluster sizes shown. This indicates that the structure of the clusters is nearly identical for ER and AE percolation for small k.

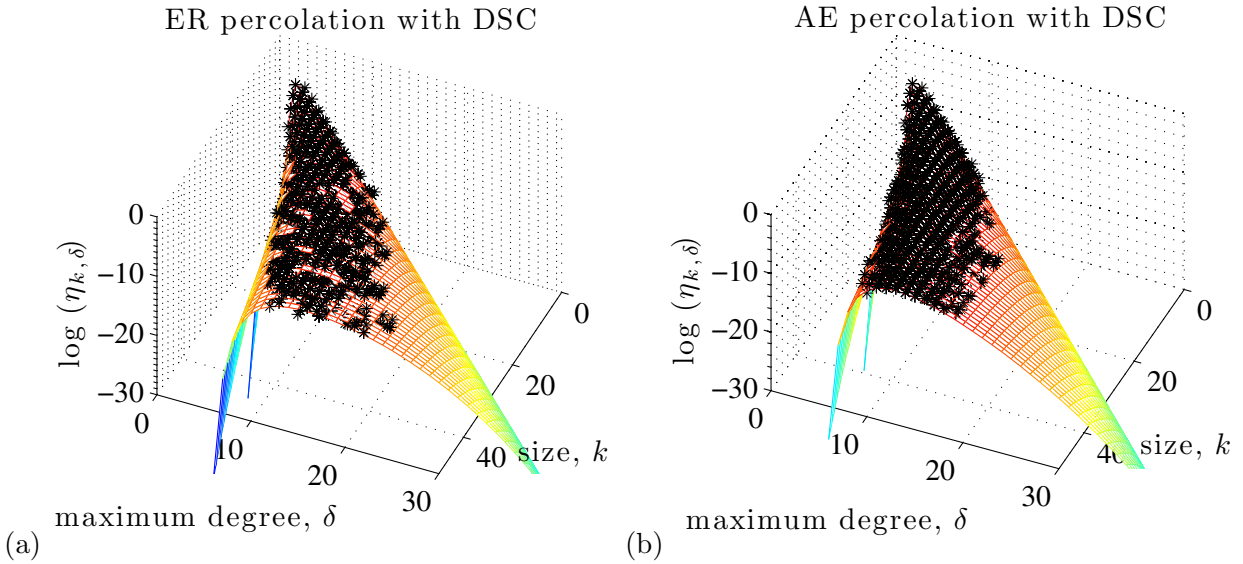


Figure 4.12: Theoretical (surface plots) and observed (symbols) values for $\eta_{k,\delta}$ are shown to be in good agreement for (a) ER percolation and (b) AE percolation with DSC attachment with $N = 10^6$ nodes. Predicted values were obtained by numerically integrating Eq. (4.12) and Eq. (4.13), respectively, with a time step $dt = N^{-1}$.

ER and AE percolation with DSC attachment

We now study the effect of DSC attachment on the connectivity of clusters forming under ER and AE percolation. As before, simulations of ER and AE percolation were run for networks with $N = 10^6$ nodes until either $l = 0.5N$ or $l = 0.75N$ undirected links were added, respectively (i.e., $t = 0.5$ or 0.75). In Fig. 4.12 we plot observed values of $\eta_{k,\delta}$ (symbols) versus our theoretical predictions given by numerically integrating either (a) Eq. (4.12) for ER link percolation or (b) Eq. (4.13) for AE link percolation. Note the good agreement between theory and observation of $\log(\rho_{k,\delta})$ for values greater than $\log(N^{-1}) \approx -13$.

Similar to our results for $\rho_{k,d}$, the agreement between the observed and predicted values for $\eta_{k,\delta}$ may be more easily seen by considering the conditional distribution $P(\delta|k)$ and expected value $E[\delta|k]$, which describe the probability that the maximum degree in a cluster of size k is given by δ,

$$P(\delta|k) = \frac{\eta_{k,\delta}}{\sum_\delta \rho_{k,\delta}}, \qquad E[\delta|k] = \sum_\delta P(\delta|k)\delta. \tag{4.15}$$

In Fig. 4.13 we show statistics for the connectivity of clusters forming under ER percolation with DSC attachment. In Fig. 4.13(a) we plot the distribution of maximum degrees given cluster size k. Observed values (symbols) are shown to agree with our theory using Eq. (4.12) (lines) for several cluster sizes. Note that the presence of DSC attachment causes the clusters' connectivities to depend very strongly on cluster size k. Specifically, as the cluster size increases, the distribution $P(\delta|k)$ shifts to the right and widens. In fact, by considering all possible cluster topologies that may be obtained via DSC attachment for increasing cluster size k, one can deduce that the maximum degree δ within a cluster of size k has the following bounds,

$$\frac{\log(k)}{\log(2)} \leq \delta \leq k - 1. \tag{4.16}$$

It follows that each distribution $P(\delta|k)$ has compact support within this interval. In Fig. 4.13(b) we plot the expected maximum degree in a cluster of size k. The black stars and solid line respectively indicate observed values and predicted values for $E[\delta|k]$, whereas the blue circles and dashed lines respectively indicate the standard deviations obtained from simulation and theory.

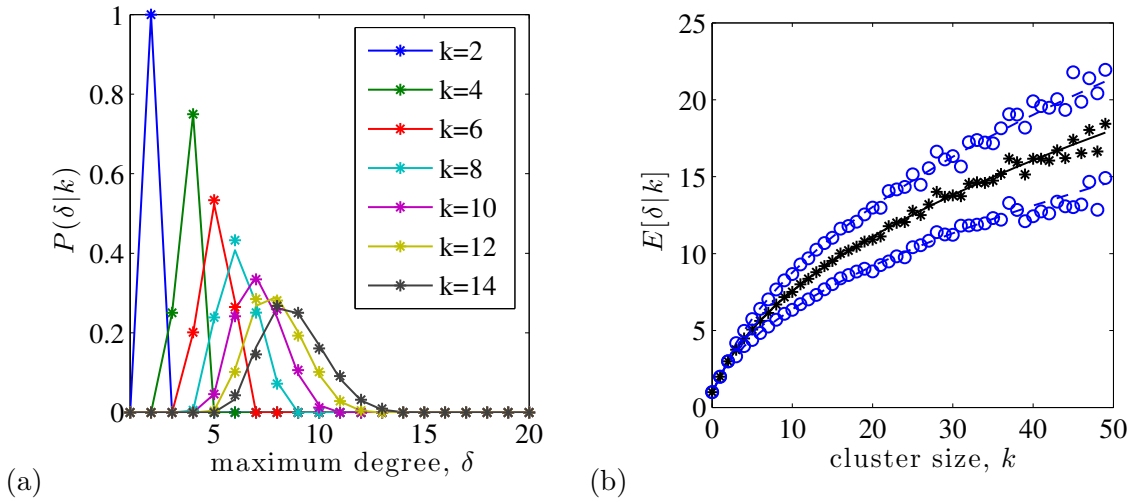


Figure 4.13: ER percolation with DSC attachment at $t = 0.5$ for $N = 10^{-6}$. (a) The distribution of maximum nodal degree in a given cluster of size k, $P(\delta|k)$. (b) The expected maximum degree in a cluster of size k, $E[\delta|k]$ (black stars and solid line), and the standard deviation, $\sigma[\delta|k]$ (blue circles and dashed lines). In both plots predicted and observed values are indicated by lines and symbols.

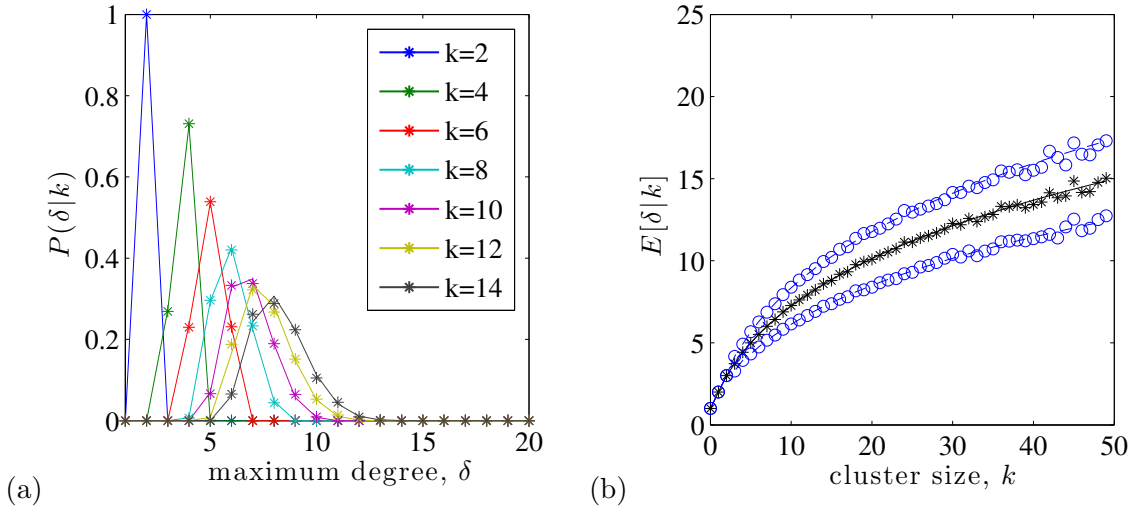


Figure 4.14: AE percolation with DSC attachment at $t = 0.75$ for $N = 10^{-6}$. (a) The distribution of maximum nodal degree in a given cluster of size k, $P(\delta|k)$. (b) The expected maximum degree in a cluster of size k, $E[\delta|k]$ (black stars and solid line), and the standard deviation, $\sigma[\delta|k]$ (blue circles and dashed lines).

In Figs. 4.14 we show similar results for AE link percolation with DSC attachment. In Fig. 4.14(a) we plot the distribution of maximum degrees in clusters of size k, where observed values (symbols) are shown to agree with our theory using Eq. (4.13) (lines). Again the presence of DSC attachment causes the distribution $P(\delta|k)$ describing cluster connectivity to widen and shift to the right as k increases. In Fig. 4.14(b) we plot the expected maximum nodal degree in a cluster of size k. The black stars and solid line respectively indicate observed values and predicted values for $E[\delta|k]$, whereas the blue circles and dashed lines respectively indicate the standard deviations obtained from simulation and our theory. These are also in excellent agreement.

We note that although this technique has been shown to correctly describe the connectivity of clusters forming with and without DSC attachment, due to the computational intensity of numerically integrating Eqs. (4.8)-(4.13) with timestep $dt = N^{-1}$, our predictions have been limited to clusters of size $k \leq 50$. Because excellent agreement was observed, it is expected that our results are also valid for larger k provided that the system remains in the subcritical regime. Although this can be improved with more efficient computational procedures, our results may be interpreted as an initial effort in this direction.

4.3 Spreading Processes on Forming Networks

We now demonstrate the effect of SC attachment on dynamics. Because SC attachment produces networks with maximal scaling of λ, we focus on an application in which large λ is beneficial: the dissemination of information in communication and wireless sensor networks [33, 128], which is often modeled as an epidemic [70, 150, 161]. We note, however, that large λ is not always advantageous. For example, large λ in ecological networks can promote instability and species extinction [109]. See Sec. 3.4 and Ref. [144] for a discussion of applications in which it is beneficial to have either small or large λ. Here we study SIS contagion [34, 110, 152, 155], which has been used to model spreading process from viral propagation in social and technological networks to the dissemination of information such as rumors and data [70, 150, 161].

To briefly review, the SIS model is a continuous time process in which each node may be

susceptible to infection or *infected*. Each infected node may infect each of its susceptible network neighbors at rate α, and each infected node may also spontaneously heal and return to being susceptible at rate β. The network state in which no nodes are infected and all nodes are susceptible is a fixed point of the collective dynamics, but this fixed point may not be stable to perturbation (i.e., a small fraction of nodes being infected by some external agent). For many topologies of connected networks in which a fraction of nodes are initially infected, the expected steady-state fraction of infected nodes f may either be zero (no infections, stable fixed point) or nonzero (endemic infection, unstable fixed point), depending on whether α/β surpasses the epidemic threshold λ^{-1} [155, 110, 152, 34]. Note that endemic infection can be prevented by decreasing λ through immunization until $\lambda^{-1} > \alpha/\beta$. Interestingly, for very large infection rates, reducing λ to prevent endemic infection can require the complete fragmentation of the network. For example, if $\alpha/\beta \geq 0.5$ the prevention of endemic infection requires $\lambda \leq 2$ which guarantees fragmentation of the network [145]. This scenario has been observed experimentally for virus propagation on mobile phone devices [154], where slowly spreading Bluetooth viruses may be inhibited by immunization (i.e., antiviral software) but rapidly spreading messaging viruses are inhibited only by a fragmented network.

We simulated SIS dynamics for moderate α/β on two networks forming under ER percolation, one with SC attachment and the other without, predicting that SC attachment will have a significant impact on the steady state fraction of infected nodes, f. We simulated dynamics with $(\alpha, \beta) = (0.075, 1)$ on $N = 10^5$ nodes at many points t in the percolation process, initially infecting 1% of nodes and then allowing the system to reach a steady state fraction of infected nodes $f(t)$, before allowing percolation to continue to another value of t, where the dynamics were re-initialized, simulated, and so on. The resulting curves $f(t)$, are shown in Fig. 4.15, where the shaded region highlights that networks forming with SC attachment (open squares) have significantly enhanced spreading characteristics compared with networks forming without SC attachment (filled squares). To contrast this result, we also plot $f(t)$ for a large infection rate $(\alpha, \beta) = (0.5, 1)$, where $f(t)$ with SC attachment (open circles) is indistinguishable from that without SC attachment (filled circles). This is not surprising, as one would expect any initial infection to saturate the cluster in which

it begins, in which case $f(t)$ would depend primarily on cluster size, not topology. We thus find two regimes of SIS dynamics on fragmented networks: when α/β is sufficiently large, f depends primarily on network fragmentation (i.e. the size of the GC), but for moderate and small α/β, f depends strongly on the connectivity of clusters (i.e., their respective λ values).

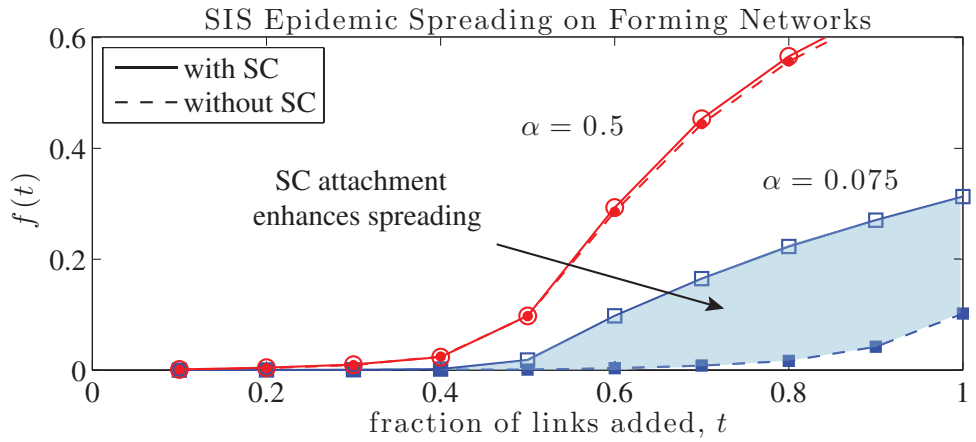


Figure 4.15: SIS epidemics [34, 110, 152, 155] were simulated on a network forming by ER percolation with SC attachment (open symbols) or without (filled symbols), where tN is the number of links added. $f(t)$ is the steady-state fraction of infected nodes when 1% of nodes are initially infected. The shaded region highlights the significant impact of SC attachment for moderate infection rate (squares). For high infection rate, SC attachment has no effect (circles).

Having shown that SC attachment can significantly enhance the spreading characteristics of a self-organizing network, we now highlight that because the network consists of disconnected clusters the steady state size of the infection f is strongly dependent on the size of the initial infection. In Fig. 4.16 we plot $f(t)$ for SIS spreading with healing rate $\beta = 1$ and infection rate of either $\alpha = 0.3125$ (red circles) or $\alpha = 0.075$ (blue squares). Simulations are shown for each α with initial conditions such that either 5% (solid lines) or 95% randomly selected nodes are initially infected. One can observe that for a given network at time t, $f(t)$ depends strongly on initial conditions and

exhibits multi-stability. This extreme sensitivity on initial conditions may be intuitively understood by noting that for small initial infections there exists many clusters with strong enough connectivity to support a nonzero infection, but it can occur that none of their nodes are initially infected.

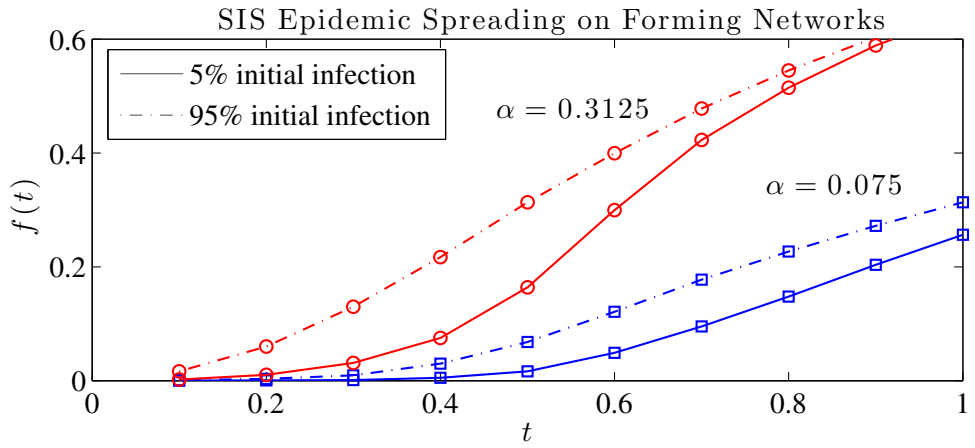


Figure 4.16: Multistability is shown for spreading on networks forming under ER percolation with SC attachment. The steady state fraction of infected nodes $f(t)$ is shown for SIS dynamics with healing rate $\beta = 1$ and two infection rates, $\alpha = 0.3125$ (red circles) or $\alpha = 0.075$ (blue squares). For each α value, simulations are initialized with either 5% (solid lines) or 95% (dot-dashed lines) random nodes being initially infected.

4.4 Discussion

Motivated by the study of connectivity-governed dynamics on evolving networks, we have developed a percolation theory focusing on the connectivity of clusters, rather than their size. In this pursuit, we have introduced a model, Social Climber attachment, that produces networks with strong connectivity and maximal scaling of λ, and validated our claims using two link-percolation models. While strong connectivity in networks is achievable via other percolation models (e.g., networks with heavy-tailed degree distributions generated by the Chung-Lu model [38]), such methods

typically require that the nodal degrees and network connectivity be defined *a priori*. In contrast, two key properties distinguish SC attachment:

First, SC attachment produces networks with large λ via self-organization. Because λ governs many dynamical processes, SC attachment provides a foundation for designing networks that self-organize with properties linked to large λ such as robustness [126, 145] and the efficient spread of information [144]. Our model is therefore promising as a starting point for the development of self-organized communication networks such as wireless sensor networks [33, 128], where data broadcasting may be modeled by SIS transmission [70, 150, 161]. Development and analysis of this application may be facilitated by the fact that SC attachment does not affect cluster sizes, only their internal topology.

Second, a novel phase transition in connectivity occurs for networks forming under SC attachment, which may have broad applications. For example, $\lambda(t)$ changes most rapidly near the percolation threshold, so creating networks near criticality may offer an effective approach for designing networks on which dynamics can be efficiently controlled by adding or removing a minimal number of links. This approach may therefore aid in the design of critical infrastructure (e.g., the power grid, communication networks, and airline networks) that can be easily switched between topologies designed for high-flow and low-flow conditions.

We conclude by suggesting several possible extensions to this work that may be of interest to readers. First, SC attachment uses complete information about the structures of the clusters that it connects, yet in some applications this information may be difficult or impossible to obtain. Although we presented Local Social Climber attachment, a variant of SC attachment restricted to local information, the effects of incomplete information or noise on SC attachment are as yet unexplored. For example, incorporating a probabilistic (rather than deterministic) link reselection step may be of interest as the resulting process would have some similarity with the preferential attachment network growth model [117, 11]. Second, one may wish to adapt our model to study various real-world networks. While networks formed under SC attachment feature large λ, to incorporate this model for the design of communication networks one would likely need to consider

many other design criteria such as betweenness centrality and software protocols [33, 128]. Finally, we named our model Social Climber attachment to reflect the selfish behavior of individuals in social situations, yet the generation of a network topology similar to that observed in social networks using the SC model would require additional link-formation rules, such as those producing modularity and transitivity [106, 105].

4.5 Conclusions

Building on results from Ch. 3 that illustrated how connectivity may be maximized for merging clusters, in this chapter we have introduced and analyzed network formation processes yielding strongly-connected networks (i.e., characterized by large λ). More generally, we have utilized a perturbation theory for the dominant eigenvalue of the network adjacency matrix to study the evolution of this eigenvalue. Therefore, this work may be interpreted more generally as a contribution to the developing field of spectral theory for evolving networks. In particular, our contribution to this field pertains to applications in which the robustness and dynamical properties of the network is of interest. However, the study of evolving matrix spectra is an ongoing area of research with broad application, ranging from link prediction for information processing [76, 77] to quantum dissipation and financial risk control [5]. Other matrices of particular interest to network science include the Laplacian [91] and modularity [58] matrices. Studying the evolution of spectra for these matrices will have broad importance and may lead to, for example, novel techniques to study and control community structure in evolving complex networks [99].

Chapter 5

Summary

In this thesis, we have illustrated the importance of spectral theory for the robustness and dynamical properties of complex networks and subsequently presented theory for the perturbation and evolution of the dominant eigenspace of the network adjacency matrix. This research has been motivated by the widespread importance of network structure in determining the dynamics of various complex systems, including biological systems [13, 28, 29, 30, 48, 66, 73, 97, 113, 114, 163], social networks [37, 75, 83, 99, 108, 115, 119, 130, 149, 156], and critical infrastructures [14, 15, 16, 46, 49, 50, 67, 129, 141, 148, 154].

In Ch. 2 we presented a spectral approach to studying the robustness of networks by analyzing the open problem of weighted percolation in networks with bidirectional edges [145]. Measuring the robustness of a network through its percolation threshold q^* (i.e., small q^* indicates robustness), we found that often q^* is inversely related to the largest eigenvalue λ of a weighted network adjacency matrix (i.e., large λ indicates robustness). Our analysis was confirmed on both synthetic and real-world networks, where we were able to provide a significant insight toward the effect of degree correlations on robustness. Specifically, while it is well known that assortativity promotes robustness for unweighted percolation [104], we found that the biased removal of nodes with large and small degrees can respectively reduce and enhance this effect. Importantly, assortativity may be negligible if the statistical bias to remove "important" nodes is strong enough. In addition to applications investigating the robustness of a given network, our analysis may also serve as a starting point for the development of improved immunization [47, 111], protection [129, 141, 144],

and attack strategies [98, 121].

Motivated by the importance of the dominant eigenspace of the network adjacency matrix for network robustness [126, 145] (see Ch. 2) as well as for a broad class of dynamics on networks (see Sec. 1.4 for an overview), in Ch. 3 we presented theory for the perturbation of this eigenspace for networks undergoing structural modification. Specifically, we introduced our recent theory [144] describing the perturbation of the largest eigenvalue λ and its corresponding eigenvectors of a network growing undergoing the addition of a module, or community. Besides characterizing the effect of structural modifications on the robustness and dynamical properties of a network, we further showed that our techniques may be utilized to judiciously control these properties. For example, techniques to maximize or minimize λ by optimally connecting the module to the network were presented, the applications of which may range from indicating novel drug targets to designing robust economic, social, and ecological networks.

Utilizing our perturbation and optimization techniques for λ presented in Ch. 3 and Ref. [144], in Ch. 4 we introduced and analyzed a network formation process yielding networks with very strong connectivity (i.e., as indicated by large λ). This network formation process was named *Social Climber* attachment due to its similarity to the well known social process. Specifically, novel attachment rules [142] were combined with standard link percolation models (e.g., Erdős-Rényi percolation) to iteratively maximize λ during network formation. This was found to produce networks with maximal scaling with size N, $\lambda \sim \mathcal{O}(N^{1/2})$, for networks with $\mathcal{O}(N)$ links. The great effect of very large λ on dynamics and robustness was illustrated by showing that spreading processes are significantly enhanced by Social Climber attachment. Among many other applications, enhanced spreading would be beneficial to self-organizing communication networks (e.g., sensors). Finally, by adopting λ as the order parameter measuring network evolution, our analysis led to the observation of a novel phase transition in connectivity.

From a scientific perspective, this work illustrates the broad importance of spectral theory for network science and its many applications. The many advantages of a spectral approach were thoroughly discussed for network-robustness (see Sec. 2.4). However, this discussion also applies

to the study of *connectivity governed* dynamics on networks, which are depend on the dominant eigenspace. In general one finds that network theory can be divided into two groups: (i) network-specific theory studying the properties of a single, given network, and (ii) ensemble-based theory describing the properties of a class of networks in which all networks share a common attribute (e.g., degree distribution). In this work we have shown a clear advantage of the spectral approach, which utilizes the specific topology of a given network as embodied by the network adjacency matrix. Moreover, because many complex network properties (e.g., heterogeneity, degree correlations, clustering, etc. [106, 105]) are naturally manifest in network spectra [35, 116, 125], the spectral approach is applicable for networks containing diverse complex structures. In contrast, competing approaches that describe the expected behavior of a network within an ensemble (e.g., theory yielding mean field equations) typically require new theory for each added type of structural complexity [145]. Moreover, if one is interested in the general effect of a complex network property on a type of dynamics, this may be inferred simply by studying the effect of that property on spectra [27, 89, 125, 144].

From a mathematical perspective, we have contributed to a growing area of spectral graph theory studying the spectra of evolving networks. Our efforts have focused on applications in which the dynamical and robustness properties of evolving networks are of interest. These applications include epidemic spreading in social and technological networks respectively undergoing immunizations and software updates [47, 111], communication and transit systems operating under failure or attack [49, 50, 98, 129], and information processing in the brain [30, 160]. We note that the study of evolving matrix spectra is an ongoing area of research with broad application, ranging from link prediction for information processing [76, 77] to quantum dissipation and financial risk control [5]. In fact, motivated by Ref. [91], which utilizes the network Laplacian matrix for image processing, we are currently studying the effect of noise on the spectra of the Laplacian. Another fruitful area of future research involves studying the perturbation and evolution of the spectra of a network's modularity matrix [58], which is expected to be an effective approach to community structure in evolving networks [99].

Bibliography

[1] http://arxiv.org/.

[2] D. Achlioptas, R. M. D'Souza, and J. Spencer, Explosive percolation in random networks, Science, 323 (2009), pp. 1453–1455.

[3] 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, ACM, 2005, pp. 36–43.

[4] A. Allard, P. Noël, L. Dubé, and B. Pourbohloul, Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics, Physical Review E, 79 (2009), p. 036113.

[5] R. Allez and J.-P. Bouchaud, Eigenvector dynamics: general theory and some applications, preprint arXiv:1203.6228, (2012).

[6] U. Alon, Network motifs: theory and experimental approaches, Nature Reviews Genetics, 8 (2007), pp. 450–461.

[7] J. P. Bagrow, S. Lehmann, and Y.-Y. Ahn, Robustness and modular structure in networks, preprint arXiv:1102.5085, (2011).

[8] R. B. Bapat and T. E. S. Raghavan, Nonnegative matrices and applications, vol. 64, Cambridge University Press, 1997.

[9] Y. Bar-Yam, Dynamics of complex systems, (2003).

[10] A.-L. Barabási, The network takeover, Nature Physics, 8 (2011), p. 14.

[11] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science, 286 (1999), pp. 509–512.

[12] A.-L. Barabási and J. Frangos, Linked: The New Science of Networks, Basic Books, 2002.

[13] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, Network medicine: a network-based approach to human disease, Nature Reviews Genetics, 12 (2011), pp. 56–68.

[14] M. Barthélemy, Spatial networks, Physics Reports, 499 (2011), pp. 1–101.

[15] M. Barthélemy and A. Flammini, Modeling urban street patterns, Physical Review Letters, 100 (2008), p. 138702.

[16] V. Batagelj and A. Mrvar, http://vlado.fmf.uni-lj.si/pub/networks/data/.

[17] A. Békéssy, P. Bekessy, and J. Komlós, Asymptotic enumeration of regular matrices, Studia Scientiarum Mathematicarum Hungarica, 7 (1972), p. 343.

[18] E. A. Bender and E. R. Canfield, The asymptotic number of labeled graphs with given degree sequences, Journal of Combinatorial Theory, Series A, 24 (1978), pp. 296–307.

[19] G. Bizhani, M. Paczuski, and P. Grassberger, Discontinuous percolation transitions in epidemic processes, surface depinning in random media, and hamiltonian random graphs, Physical Review E, 86 (2012), p. 011128.

[20] M. Boguñá and M. Á. Serrano, Generalized percolation in random directed networks, Physical Review E, 72 (2005), p. 016106.

[21] T. Bohman and A. Frieze, Avoiding a giant component, Random Structures, 19 (2001), pp. 75–85.

[22] B. Bollobás, Random graphs, vol. 73, Cambridge university press, 2001.

[23] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan, Percolation on dense graph sequences, The Annals of Probability, 38 (2010), pp. 150–183.

[24] B. Bollobas and O. Riordan, Percolation, Cambridge University Press, 2006.

[25] J. A. Bondy and U. S. R. Murty, Graph theory with applications, vol. 290, Macmillan London, 1976.

[26] U. Brandes, A faster algorithm for betweenness centrality, Journal of Mathematical Sociology, 25 (2001), pp. 163–177.

[27] M. Brede and S. Sinha, Assortative mixing by degree makes a network more unstable, preprint arXiv:cond-mat/0507710, (2005).

[28] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, et al., Topological structure analysis of the protein–protein interaction network in budding yeast, Nucleic Acids Research, 31 (2003), pp. 2443–2450.

[29] J. Buck, Synchronous rhythmic flashing of fireflies. ii., Quarterly Review of Biology, (1988), pp. 265–289.

[30] G. Buzsaki, Rhythms of the Brain, Oxford University Press, USA, 2009.

[31] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, Network robustness and fragility: Percolation on random graphs, Physical Review Letters, 85 (2000), p. 5468.

[32] N. Carlson, D.-H. Kim, and A. E. Motter, Sample-to-sample fluctuations in real-network ensembles, Chaos, 21 (2011), p. 025105.

[33] A. Cerpa and D. Estrin, Ascent: Adaptive self-configuring sensor networks topologies, IEEE transactions on mobile computing, 3 (2004), pp. 272–285.

[34] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, Epidemic thresholds in real networks, ACM Transactions on Information and System Security (TISSEC), 10 (2008), p. 1.

[35] S. Chauhan, M. Girvan, and E. Ott, Spectral properties of networks with community structure, Physical Review E, 80 (2009), p. 056114.

[36] Y. Chen, G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley, Finding a better immunization strategy, Physical Review Letters, 101 (2008), p. 58701.

[37] N. A. Christakis and J. H. Fowler, The spread of obesity in a large social network over 32 years, New England Journal of Medicine, 357 (2007), pp. 370–379.

[38] F. Chung and L. Lu, Connected components in random graphs with given expected degree sequences, Annals of combinatorics, 6 (2002), p. 125.

[39] F. Chung, L. Lu, and V. Vu, Spectra of random graphs with given expected degrees, Proceedings of the National Academy of Sciences, (2003), pp. 6313–6318.

[40] F. Chung, L. Lu, and V. Vu, The spectra of random graphs with given expected degrees, Internet Mathematics, 1 (2004), pp. 257–275.

[41] F. R. Chung, Spectral graph theory, vol. 92, Amer Mathematical Society, 1997.

[42] N. N. Chung, L. Y. Chew, and C. H. Lai, Spectral anlysis on explosive percolation, arXiv preprint arXiv:1301.0399, (2013).

[43] A. Clauset, C. Moore, and M. E. Newman, Structural inference of hierarchies in networks, in Statistical network analysis: models, issues, and new directions, Springer, 2007, pp. 1–13.

[44] ——, Hierarchical structure and the prediction of missing links in networks, Nature, 453 (2008), pp. 98–101.

[45] A. Clauset, M. E. Newman, and C. Moore, Finding community structure in very large networks, Physical Review E, 70 (2004), p. 066111.

[46] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, Resilience of the internet to random breakdowns, Physical Review Letters, 85 (2000), pp. 4626–4628.

[47] R. Cohen, S. Havlin, and D. Ben-Avraham, Efficient immunization strategies for computer networks and populations, Physical Review Letters, 91 (2003), p. 247901.

[48] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, Effective leadership and decision-making in animal groups on the move, Nature, 433 (2005), pp. 513–516.

[49] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization, Chaos, 17 (2007), pp. 026103–026103.

[50] I. Dobson, B. A. Carreras, and D. E. Newman, A loading-dependent model of probabilistic cascading failure, Probability in the Engineering and Informational Sciences, 19 (2005), pp. 15–32.

[51] S. N. Dorogovtsev, A. Goltsev, and J. F. Mendes, K-core organization of complex networks, Physical Review Letters, 96 (2006), p. 40601.

[52] S. N. Dorogovtsev and J. F. Mendes, Evolution of networks: From biological nets to the Internet and WWW, OUP Oxford, 2003.

[53] J. A. Drake, The mechanics of community assembly and succession, Journal of Theoretical Biology, 147 (1990), pp. 213–233.

[54] R. M. D'Souza and M. Mitzenmacher, Local cluster aggregation models of explosive percolation, Physical Review Letters, 104 (2010), p. 195702.

[55] P. Erdős and A. Rényi, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kutató Int. Közl, 5 (1960), pp. 17–61.

[56] J. Feng, V. K. Jirsa, and M. Ding, Synchronization in networks with random interactions: Theory and applications, Chaos, 16 (2006), pp. 015109–015109.

[57] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin, Stability and topology of scale-free networks under attack and defense strategies, Physical Review Letters, 94 (2005), p. 188701.

[58] M. Girvan and M. E. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences, 99 (2002), p. 7821.

[59] K.-I. Goh, B. Kahng, and D. Kim, Spectra and eigenvectors of scale-free networks, Physical Review E, 64 (2001), p. 051903.

[60] G. H. Golub and C. F. Van Loan, Matrix computations, vol. 3, Johns Hopkins University Press, 1996.

[61] A. Gutfraind, Optimizing topological cascade resilience based on the structure of terrorist networks, PloS one, 5 (2010), p. e13448.

[62] T. Hase, H. Tanaka, Y. Suzuki, S. Nakagawa, and H. Kitano, Structure of protein interaction networks and their implications on drug design, PLoS computational biology, 5 (2009), p. e1000550.

[63] P. W. Holland and S. Leinhardt, Transitivity in structural models of small groups., Comparative Group Studies, (1971).

[64] P. Holme and B. J. Kim, Growing scale-free networks with tunable clustering, Physical Review E, 65 (2002), p. 026107.

[65] T. Ichinomiya, Frequency synchronization in a random oscillator network, Physical Review E, 70 (2004), p. 026116.

[66] H. Jeong, S. P. Mason, A.-L. Barabasi, and Z. N. Oltvai, Lethality and centrality in protein networks, preprint arXiv:cond-mat/0105306, (2001).

[67] V. Kalapala, V. Sanwalani, A. Clauset, and C. Moore, Scale invariance in road networks, Physical Review E, 73 (2006), p. 026130.

[68] P. Kaluza, M. Ipsen, M. Vingron, and A. S. Mikhailov, Design and statistical properties of robust functional networks: A model study of biological signal transduction, Physical Review E, 75 (2007), p. 015101.

[69] P. Kaluza, M. Vingron, and A. S. Mikhailov, Self-correcting networks: Function, robustness, and motif distributions in biological signal processing, Chaos, 18 (2008), pp. 026113–026113.

[70] A. Khelil, C. Becker, J. Tian, and K. Rothermel, An epidemic model for information diffusion in manets, in Proceedings of the 5th ACM international workshop on modeling analysis and simulation of wireless and mobile systems, ACM, 2002, pp. 54–60.

[71] O. Kinouchi and M. Copelli, Optimal dynamical range of excitable networks at criticality, Nature Physics, 2 (2006), pp. 348–351.

[72] A. N. Kirillov, Dilogarithm identities, Progress of Theoretical Physics Supplement, 118 (1995), pp. 61–142.

[73] H. Kitano, Systems biology: a brief overview, Science, 295 (2002), pp. 1662–1664.

[74] H. Kitano, A robustness-based approach to systems-oriented drug design, Nature Reviews Drug Discovery, 5 (2007), pp. 202–210.

[75] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Identification of influential spreaders in complex networks, Nature Physics, 6 (2010), pp. 888–893.

[76] J. Kunegis, D. Fay, and C. Bauckhage, Network growth and the spectral evolution model, in Proceedings of the 19th ACM international conference on information and knowledge management, ACM, 2010, pp. 739–748.

[77] J. Kunegis, D. Fay, and C. Bauckhage, Spectral evolution in dynamic networks, Knowledge and Information Systems, (2012), pp. 1–36.

[78] Y. Kuramoto, Chemical oscillations, waves, and turbulence, Dover Publications, 2003.

[79] D. B. Larremore, http://danlarremore.com/.

[80] D. B. Larremore, M. Y. Carpenter, E. Ott, and J. G. Restrepo, Statistical properties of avalanches in networks, Physical Review E, 85 (2012), p. 066131.

[81] D. B. Larremore, W. L. Shew, E. Ott, and J. G. Restrepo, Effects of network topology, transmission delays, and refractoriness on the response of coupled excitable systems to a stochastic stimulus, Chaos, 21 (2011), p. 025117.

[82] D. B. Larremore, W. L. Shew, and J. G. Restrepo, Predicting criticality and dynamic range in complex networks: effects of topology, Physical Review Letters, 106 (2011), p. 58101.

[83] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al., Life in the network: the coming age of computational social science, Science (New York, NY), 323 (2009), p. 721.

[84] N. E. Leonard, T. Shen, B. Nabet, L. Scardovi, I. D. Couzin, and S. A. Levin, Decision versus compromise for animal groups in motion, Proceedings of the National Academy of Sciences, 109 (2012), pp. 227–232.

[85] Y.-Y. Liu, E. Csóka, and M. Pósfai, Core percolation on complex networks, preprint arXiv:1206.2550, (2012).

[86] H. Ma, I. Goryanin, et al., Human metabolic network reconstruction and its impact on drug discovery and development, Drug discovery today, 13 (2008), pp. 402–408.

[87] N. Madar, T. Kalisky, R. Cohen, D. Ben-avraham, and S. Havlin, Immunization and epidemic dynamics in complex networks, The European Physical Journal B, 38 (2004), pp. 269–276.

[88] Y. L. Maistrenko, B. Lysyansky, C. Hauptmann, O. Burylko, and P. A. Tass, Multistability in the kuramoto model with synaptic plasticity, Physical Review E, 75 (2007), p. 066207.

[89] R. M. May, Will a large complex system be stable?, Nature, 238 (1972), pp. 413–414.

[90] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, The unreasonable effectiveness of tree-based theory for networks with clustering, Physical Review E, 83 (2011), p. 036112.

[91] F. G. Meyer and X. Shen, Perturbation of the eigenvectors of the graph laplacian: Application to image denoising, preprint arXiv:1202.6666, (2012).

[92] A. Milanese, J. Sun, and T. Nishikawa, Approximating spectral impact of structural perturbations in large networks, Physical Review E, 81 (2010), p. 046112.

[93] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Superfamilies of evolved and designed networks, Science, 303 (2004), pp. 1538–1542.

[94] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Network motifs: simple building blocks of complex networks, Science Signalling, 298 (2002), p. 824.

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

[96] A. E. Motter, Cascade control and defense in complex networks, Physical Review Letters, 93 (2004), p. 98701.

[97] A. E. Motter, N. Gulbahce, E. Almaas, and A. L. Barabási, Predicting synthetic rescues in metabolic networks, Molecular Systems Biology, 4 (2008), pp. 168–168.

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

[99] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science, 328 (2010), pp. 876–878.

[100] B. Nabet, N. E. Leonard, I. D. Couzin, and S. A. Levin, Dynamics of decision making in animal group motion, Journal of Nonlinear Science, 19 (2009), pp. 399–435.

[101] A. K. Naimzada, S. Stefani, and A. Torriero, Networks, topology and dynamics: theory and applications to economics and social systems, vol. 613, Springer, 2008.

[102] Z. Neda, E. Ravasz, Y. Brechet, T. Vicsek, and A. Barabasi, The sound of many hands clapping, Nature, 403 (2000), p. 849.

[103] D. L. Nelson, C. L. McEnvoy, and T. A. Schreiber, www.usf.edu/freeassociation/.

[104] M. E. Newman, Assortative mixing in networks, Physical Review Letters, 89 (2002), p. 208701.

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

[106] M. E. Newman, Networks: an introduction, Oxford University Press, Inc., 2010.

[107] M. E. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, Physical Review E, 64 (2001), p. 026118.

[108] D. Obstfeld, Social networks, the tertius iungens orientation, and involvement in innovation, Administrative Science Quarterly, 50 (2005), pp. 100–130.

[109] R. K. Pan and S. Sinha, Modular networks emerge from multiconstraint optimization, Physical Review E, 76 (2007), p. 045103.

[110] R. Parshani, S. Carmi, and S. Havlin, Epidemic threshold for the susceptible-infectious-susceptible model on random networks, Physical Review Letters, 104 (2010), p. 258701.

[111] R. Pastor-Satorras and A. Vespignani, Immunization of complex networks, Physical Review E, 65 (2002), p. 036104.

[112] M. Penrose, Random geometric graphs, vol. 5, Oxford University Press Oxford, UK:, 2003.

[113] S. L. Pimm, Food webs, University of Chicago Press, 2002.

[114] A. Pomerance, E. Ott, M. Girvan, and W. Losert, The effect of network topology on the stability of discrete state models of genetic control, Proceedings of the National Academy of Sciences, 106 (2009), pp. 8209–8214.

[115] M. A. Porter, P. J. Mucha, M. E. Newman, and C. M. Warmbrand, A network analysis of committees in the us house of representatives, Proceedings of the National Academy of Sciences, 102 (2005), pp. 7057–7062.

[116] V. M. Preciado and A. Jadbabaie, From local measurements to network spectral properties: Beyond degree distributions, preprint arXiv:1004.3524, (2010).

[117] D. d. S. Price, A general theory of bibliometric and other cumulative advantage processes, Journal of the American Society for Information Science, 27 (2007), pp. 292–306.

[118] R. J. Prill, P. A. Iglesias, and A. Levchenko, Dynamic properties of network motifs contribute to biological network organization, PLoS Biology, 3 (2005), p. e343.

[119] R. Reagans and B. McEvily, Network structure and knowledge transfer: The effects of cohesion and range, Administrative Science Quarterly, 48 (2003), pp. 240–267.

[120] J. G. Restrepo, E. Ott, and B. R. Hunt, Onset of synchronization in large networks of coupled oscillators, Physical Review E, 71 (2005), p. 036151.

[121] J. G. Restrepo, E. Ott, and B. R. Hunt, Characterizing the dynamical importance of network nodes and links, Physical Review Letters, 97 (2006), p. 94102.

[122] J. G. Restrepo, E. Ott, and B. R. Hunt, Emergence of coherence in complex networks of heterogeneous dynamical systems, Physical Review Letters, 96 (2006), p. 254103.

[123] J. G. Restrepo, E. Ott, and B. R. Hunt, Emergence of synchronization in complex networks of interacting dynamical systems, Physica D, 224 (2006), pp. 114 – 122.

[124] J. G. Restrepo, E. Ott, and B. R. Hunt, Synchronization in large directed networks of coupled phase oscillators, Chaos, 16 (2006), p. 015107.

[125] J. G. Restrepo, E. Ott, and B. R. Hunt, Approximating the largest eigenvalue of network adjacency matrices, Physical Review E, 76 (2007).

[126] J. G. Restrepo, E. Ott, and B. R. Hunt, Weighted percolation on directed networks, Physical Review Letters, 100 (2008), p. 58701.

[127] M. Ripeanu and I. Foster, Mapping the gnutella network: Macroscopic properties of large-scale peer-to-peer systems, Peer-to-Peer Systems, (2002), pp. 85–93.

[128] P. Santi, Topology control in wireless ad hoc and sensor networks, ACM Computing Surveys (CSUR), 37 (2005), pp. 164–194.

[129] C. M. Schneider, A. A. Moreira, J. S. Andrade Jr, S. Havlin, and H. J. Herrmann, Mitigation of malicious attacks on networks, Proceedings of the National Academy of Sciences, 108 (2011), pp. 3838–3841.

[130] J. Scott, Social network analysis: A handbook, Sage Publications Limited, 2000.

[131] P. Seliger, S. C. Young, and L. S. Tsimring, Plasticity and learning in a network of coupled phase oscillators, Physical Review E, 65 (2002), p. 041906.

[132] T. C. A. R. D. Set, http://www.caida.org/data/active/as-relationships/., 2006.

[133] S. Sinha, Complexity vs. stability in small-world networks, Physica A, 346 (2005), pp. 147–153.

[134] S. Sinha and S. Sinha, Robust emergent activity in dynamical networks, Physical Review E, 74 (2006), p. 066117.

[135] P. S. Skardal and J. G. Restrepo, Hierarchical synchrony of phase oscillators in modular networks, Physical Review E, 85 (2012), p. 016208.

[136] P. S. Skardal, J. Sun, D. Taylor, and J. G. Restrepo, Effects of degree-frequency correlations on network synchronization: Universality and full phase-locking, EPL (Europhysics Letters), 101 (2013), p. 20001.

[137] P. S. Skardal, D. Taylor, and J. G. Restrepo, Complex macroscopic behavior in systems of phase oscillators with adaptive coupling, Physica D, 10.1016/j.physd.2013.01.012, (2012).

[138] J. Spencer and N. Wormald, Birth control for giants, Combinatorica, 27 (2007), pp. 587–628.

[139] S. Squires, E. Ott, and M. Girvan, Dynamical instability in boolean networks as a percolation problem, Physical Review Letters, 109 (2012), p. 85701.

[140] S. H. Strogatz, From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators, Physica D: Nonlinear Phenomena, 143 (2000), pp. 1–20.

[141] T. Tanizawa, S. Havlin, and H. E. Stanley, Robustness of onionlike correlated networks against targeted attacks, Physical Review E, 85 (2012), p. 046109.

[142] D. Taylor and D. B. Larremore, Social climber attachment in forming networks produces a phase transition in a measure of connectivity, Physical Review E, 86 (2012), p. 031140.

[143] D. Taylor, E. Ott, and J. G. Restrepo, Spontaneous synchronization of coupled oscillator systems with frequency adaptation, Physical Review E, 81 (2010), p. 046214.

[144] D. Taylor and J. G. Restrepo, Network connectivity during mergers and growth: Optimizing the addition of a module, Physical Review E, 83 (2011), p. 066112.

[145] D. Taylor and J. G. Restrepo, A network-specific approach to percolation in complex networks with bidirectional links, EPL (Europhysics Letters), 98 (2012), p. 16007.

[146] K. M. Taylor and F. G. Meyer, A random walk on image patches, SIAM Journal on Imaging Sciences, 5 (2012), pp. 688–725.

[147] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, Comparing community structure to characteristics in online collegiate social networks, SIAM review, 53 (2011), pp. 526–543.

[148] P. Tsaparas, www.cs.helsin-ki.fi/u/tsaparas/macn2006/data-code.html, 2006.

[149] B. Uzzi, The sources and consequences of embeddedness for the economic performance of organizations: The network effect, American sociological review, (1996), pp. 674–698.

[150] A. Vahdat, D. Becker, et al., Epidemic routing for partially connected ad hoc networks, tech. rep., CS-200006, Duke University, 2000.

[151] P. Van Mieghem, Graph spectra for complex networks, Cambridge University Press, 2010.

[152] P. Van Mieghem, J. Omic, and R. Kooij, Virus spread in networks, IEEE/ACM Transactions on Networking, 17 (2009), pp. 1–14.

[153] P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. van de Bovenkamp, D. Liu, and H. Wang, Decreasing the spectral radius of a graph by link removals, Physical Review E, 84 (2011), p. 016101.

[154] P. Wang, M. C. González, C. A. Hidalgo, and A.-L. Barabási, Understanding the spreading patterns of mobile phone viruses, Science, 324 (2009), pp. 1071–1076.

[155] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, Epidemic spreading in real networks: An eigenvalue viewpoint, in Proceedings. 22nd International Symposium on Reliable Distributed Systems, 2003, IEEE, 2003, pp. 25–34.

[156] S. Wasserman and K. Faust, Social network analysis: Methods and applications, vol. 8, Cambridge university press, 1994.

[157] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature, 393 (1998), p. 440.

[158] J. G. White, E. Southgate, J. Thomson, S. Brenner, J. White, E. Southgate, J. Thomson, and S. Brenner, The structure of the nervous system of the nematode caenorhabditis elegans, Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 314 (1986), p. 1.

[159] C. C. Wilmers, S. Sinha, and M. Brede, Examining the effects of species richness on community stability: an assembly model approach, Oikos, 99 (2002), pp. 363–367.

[160] G. Zamora-López, C. Zhou, and J. Kurths, Exploring brain function from anatomical connectivity, Frontiers in Neuroscience, 5 (2011).

[161] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, Performance modeling of epidemic routing, Computer Networks, 51 (2007), pp. 2867–2891.

[162] V. P. Zhigulin, Dynamical motifs: building blocks of complex dynamics in sparsely connected random networks, Physical Review Letters, 92 (2004), p. 238701.

[163] G. R. Zimmermann, J. Lehar, and C. T. Keith, Multi-target therapeutics: when the whole is greater than the sum of the parts, Drug Discovery Today, 12 (2007), pp. 34–42.