Optimal Graph Filter Design for Large-Scale Random Networks

Submitted in partial fulfillment for the requirements for
the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Stephen M. Kruzick

B.A., Mathematics, Rice University
B.S., Electrical and Computer Engineering , Rice University
M.S., Electrical and Computer Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

May 2018

Acknowledgements

There are many individuals for whose support I am especially grateful, from both the Department of Electrical and Computer Engineering (ECE) at Carnegie Mellon University (CMU) and from other aspects of my life, deserving of acknowledgement here. First and foremost, I would like to express gratitude for my advisor José M. F. Moura for guiding throughout my study at CMU. He introduced me to the graph signal processing problems addressed in this thesis and provided me with years of advice, support, and patience that were vital both to completing this work and to my development as a researcher.

I thank the thesis committee members Jelena Kovačević, Soummya Kar, Giulia Fanti (from the Carnegie Mellon University ECE department) as well as Chai Wah Wu (from IBM Research in New York) for setting aside the time to attend the thesis proposal and thesis defense. The insightful questions, comments, and discussion they provided have influenced the path of this research and will continue to shape future related research.

Additionally, I thank Carol Patterson, Claire Bauerle, Sherri Ferris, Christina Cowan, Nathan Snizaski, and many other members of the ECE department staff for working to support the department and its students.

I also thank all my friends, colleagues, officemates, and fellow research group members from the department: Aliaksei Sandryhaila, Augusto Santos, Aurora Schmidt, Dusan Jakovetic, Evgeny Toporov, Jian Du, Joao Saude, Joel Harley, John Shi, Jonathan Mei, Joya Deri, June Zhang, Kyle Anderson, Liangyan Gui, Mark Cheung, Nick ODonoghue, Rajshekhar Das, Satwik Kottur, Sérgio Pequito, Shanghang Zhang, Subhro Das, Yuan Chen, Andrew Hsu, Anit Sahu, Brian Swenson, Javad Mohammadi, Jonathan Donadee, JY Joo, Kyri Baker, Nikos Arechiga, Rohan Chabukswar, Matthew Baron, Sam Fazel-Sarjui, Umang

Bhatt, Alexander Sugar, and many others. Their companionship through many hours of work and uplifting company over lunches, games, and conferences has been greatly appreciated.

Furthermore, I would like to express gratitude for my friends from outside the department, many of whom are members of the CMU Ballroom Dance Club. While I cannot list all of them here, each member contributed to an experience filled with many good times. Particularly from this group, I would like to thank my roommates over the past few years for their great company: Bill Derocha, Victor Wang, Yimeng Xu, Helen Chao, Kristen Clark, Moses Cruz, Gihyun Kim, and Jinhee Lee. Most of all, I would like to thank Krystle Koe for always being there for me.

Finally, I would especially like to thank my family, including my parents George Kruzick and Victoria Kruzick, brother Daniel Kruzick, sister Amy Kruzick, and all of my aunts, uncles, and cousins for all the love and support over the years.

This work was supported by the Benjamin Garver Lamme/Westinghouse Graduate Fellowship, the National Defense Science and Engineering Graduate Fellowship (NDSEG), and NSF Grant # CCF 1513936.

Abstract

Graph signal processing analyzes signals supported on the nodes of a network with respect to a shift operator matrix that conforms to the graph structure. For shift-invariant graph filters, which are polynomial functions of the shift matrix, the filter response is defined by the value of the filter polynomial at the shift matrix eigenvalues. Thus, information regarding the spectral decomposition of the shift matrix plays an important role in filter design. However, under stochastic conditions leading to uncertain network structure, the eigenvalues of the shift matrix become random, complicating the filter design task. In such case, empirical distribution functions built from the random matrix eigenvalues may exhibit deterministic limiting behavior that can be exploited for problems on large-scale random networks.

Acceleration filters for distributed average consensus dynamics on random networks provide the application covered in this thesis work. The thesis discusses methods from random matrix theory appropriate for analyzing adjacency matrix spectral asymptotics for both directed and undirected random networks, introducing relevant theorems. Network distribution properties that allow computational simplification of these methods are developed, and the methods are applied to important classes of random network distributions. Subsequently, the thesis presents the main contributions, which consist of optimization problems for consensus acceleration filters based on the obtained asymptotic spectral density information. The presented methods cover several cases for the random network distribution, including both undirected and directed networks as well as both constant and switching random networks. These methods also cover two related optimization objectives, asymptotic convergence rate and graph total variation.

Table of Contents

List of Figures

Chapter One

Introduction

1.1 Research Motivation

Networks feature prominently in signal processing problems that arise from increasingly interrelated people, devices, and systems of the modern connected world. Practical applications often come with challenging aspects that can include large-scale size, uncertain structure, and changing conditions. Graph signal processing extends methods from classical signal processing to address data with underlying relationships described by a graph, whether resulting from a physical network or some intangible connection [1]. The associated methods perform analysis on signals defined as functions on the nodes of a graph with respect to a shift operator defined as a matrix that conforms to the graph structure. Typically, the shift matrix is the adjacency matrix, the Laplacian matrix, or a normalized variant thereof [1–5]. Analogies to classical signal processing concepts are drawn in terms of the shift matrix. Importantly, decomposition of the signal vector in terms of the basis of shift matrix eigenvectors (for diagonalizable shift matrices) provides the Graph Fourier Transform [2, 3], where the eigenvectors are pure-frequency signals [4]. Because of their connection to graph total variation [5], the eigenvalues of the shift matrix can be interpreted as frequencies [4]. Polynomial functions of the graph shift matrix perform shift-invariant filtering [2], and the filter polynomial value at each shift matrix eigenvalue provides the filter response in the graph frequency domain [4]. Consequently, filter design optimization problems for graph signal processing require knowledge regarding the eigendecomposition of the shift matrix. Random graphs with random shift matrix eigenvalues must deal with uncertainty by accounting for the eigenvalue statistics, a significantly more complicated task.

Large-scale network size can be exploited to deal with uncertainty in network structure for suitable random network models. While the random matrices that respect the structure of random networks have random eigenvalues with typically intractable joint statistics, the empirical distribution built from the set of eigenvalues sometimes approaches a deterministic function that can be used as an approximation for large-scale networks. Many such limiting theorems appear in random matrix theory literature. Well known examples that cover simple, unstructured cases include the Wigner semicircular law [6], the Marchenko-Pastur law [7, 8], and the Girko circular law [9]. For use with complex networks, results are required that allow for non-identically distributed random matrices with some entries explicitly zero. This thesis discusses several results by Girko [9] that are useful for this reason.

Distributed average consensus provides an example application considered in this thesis for which graph filters can be designed that result in accelerated convergence. In distributed average consensus, network nodes compute the mean of data initially distributed among the nodes over several iterations in which they communicate with neighboring nodes [10]. Each node incorporates the data received from neighbor nodes according to a consensus iteration matrix $W\left(\mathcal{G}\right)$ that respects the network structure $\mathcal{G}$ and that describes the dynamics of the consensus process, producing convergence under appropriate conditions [11]. There are several filtering schemes available in the literature that accelerate convergence [12–28], and one particular approach updates the state with a linear combination of the past d states [16]. For constant network topology, this can be interpreted as a shift-invariant graph filter. Well designed filters minimize the graph filter response magnitude at all eigenvalues of $W\left(\mathcal{G}\right)$ except for the one that preserves the averaging eigenspace [16]. Thus, the spectral asymptotics of large-scale random networks provides relevant information for consensus acceleration filter design, as described in this work.

This thesis focuses on connecting the deterministic spectral asymptotics of random matrices to optimal graph filter design for large-scale random networks, specifically developing methods for consensus acceleration filter design and related graph signal processing problems. The main contributions of the thesis are summarized in Section 1.2. The overall structure of the thesis is described with previews of each chapter in Section 1.3.

1.2 Main Contributions

The main contributions of this thesis are a series of optimization problems for graph filter design for several categories of large-scale random network distributions. The developed methods are specifically applied to consensus acceleration filters and related design objectives. These optimization problems each follow a common thread by leveraging the predictable limiting spectral asymptotics of large-scale network consensus iteration matrices to approximate optimal filter design, substituting this information for true knowledge of the consensus iteration matrix eigenvalues. Broadly, the novel contributions of the thesis can be divided into two categories: results for constant (fixed with respect to time) random networks presented in Chapter 4 and results for switching (varying with respect to time) random networks presented in Chapter 5.

For suitable large-scale constant random networks, the support of the approximate spectral densities derived through the methods presented in Chapter 3 provide a good approximation of the true set of eigenvalues for the random consensus iteration matrix. The following associated novel contributions appear in Chapter 4.

- Periodic consensus acceleration filters are designed to minimize the convergence rate through proposed optimization problems that bound the filter response over the support of the approximate empirical spectral distribution. For undirected networks (with consensus iteration matrices that have real eigenvalues), this results in a linear program (LP). For directed networks (with consensus iteration matrices that have real eigenvalues), this results in a quadratically constrained linear program (QCLP).

- Motivated by two related filter design objectives, weighted variants of the above consensus acceleration linear program and quadratically constrained linear program are proposed for undirected and directed networks, respectively. This generalization enables filter design with respect to worst case graph total variation and expected case graph total variation for both undirected networks and for directed networks.

For large-scale constant switching random networks, approximate spectral densities derived through the methods presented in Chapter 3 provide a good approximation for the moments of the true empirical spectral distributions for the random consensus iteration matrix. The following associated novel contributions appear in Chapter 5.

- An optimization problem to minimize the expected consensus error norm after one filter application is posed as a linearly constrained quadratic program (LCQP). The objective function matrix is the expected Gram matrix of consensus error vectors from the filtering terms with respect to the sequence of matrices and the input error vector.
- An approximation for the expected Gram matrix in the above quadratic program is derived based on stated simplifying assumptions regarding the random eigenvectors. The resulting approximation depends only on the switching probability and on the approximate spectral distribution moments. When possible, the approximate is derived via the methods in Chapter 3. When not possible, a slightly different formulation based on the simulated expected spectral distribution is employed.

1.3 Thesis Overview

This section describes the overall organization of the thesis and briefly describes the topics contained in each remaining chapter. The content of this thesis is organized as follows:

- **Chapter 2** provides background context and notation referenced throughout the thesis. Essential terminology regarding graphs and networks are introduced in Section 2.2. Random matrix theory concepts, definitions, and classic results appear in Section 2.3, which Chapter 3 further supplements with discussion of Girko's methods and their application to networks. The motivating application for this work, distributed average consensus on networks, is described in Section 2.4. The graph signal processing framework provides an essential perspective for the work in this thesis, as introduced in Section 2.5. In particular, consensus acceleration filters designed to produce faster convergence of distributed average consensus dynamics can be viewed as graph filters, as discussed in Section 2.6. Several filter design tasks in this thesis are Chebyshev approximation problems, with background described in Section 2.7.

- **Chapter 3** introduces detailed background on Girko's stochastic canonical equation methods [9] for approximating the spectral distribution of large-scale matrices. The chapter additionally provides an in-depth tutorial describing the application of these methods to the adjacency matrices of random networks. The discussion covers three theorems relevant to random matrices associated with random networks: Girko's K1 theorem, Girko's K27 theorem, and Girko's K25 theorem. Girko's K1 theorem appears as Theorem 3.1 in Section 3.2 and pertains to random symmetric matrices with independent entries (except as related by symmetry) and thus to undirected link-percolation networks. Girko's K27 theorem appears as Theorem 3.2 in Section 3.3 and pertains to random symmetric matrices with independent block submatrices (except as related by symmetry) and thus to undirected link-percolation networks with localized link dependencies. Girko's K25 theorem appears as Theorem 3.3 in Section 3.4 and pertains to random potentially non-Hermitian matrices with independent entries and thus to directed link-percolation networks.

- **Chapter 4** addresses filter design problems for constant (not time-varying) large-scale random networks using spectral asymptotics. Filter design methods for accelerated consensus on large-scale undirected random network models appear in Section 4.2, which poses a linear program (LP) to minimize the worst case filter response. Filter design methods for accelerated consensus on large-scale directed random network models appear in Section 4.3, which poses a linearly constrained quadratic program (LCQP) to minimize the worst case filter response. Both sections provides supporting simulation results. Filter design methods for closely related objectives (worst case graph total variation and expected graph total variation) on large-scale random network models (both directed and undirected) appear in Section 4.4, which poses weighted variants of the LP (for the undirected case) and the LCQP (for the directed case). For the weighted cases, supporting simulation results appear in Section 4.5.

- **Chapter 5** addresses filter design problems for constant (not time-varying) large-scale undirected random networks using spectral asymptotics. Specifically, the filters design method presented in this chapter approximately minimize the expected consensus error with respect to the input consensus error and with respect to the consensus iteration matrix sequence. The resulting optimization problem for the expected norm squared of the consensus error becomes a linearly constrained quadratic program (LCQP), in which the matrix in the objective function can be approximated in terms of approximate spectral density moments. For the unnormalized Laplacian-based iteration matrix $W = I - \alpha\mathcal{L}$, Section 5.2 accomplishes this approximation in terms of the expected spectral density as described in Proposition 5.1. For the row-normalized Laplacian-based iteration matrix $W = I - \alpha\mathcal{L}_R$, Section 5.2 accomplishes this approximation in terms of the approximate spectral density (computed using Girko's methods from Section 3.2 and Section 3.3 of Chapter 3) as described in Proposition 5.2. Supporting simulation results appear in each section.

- **Chapter 6** concludes the thesis by summarizing the work presented in the preceding chapters, reiterating the main contributions of the thesis. Potential topics for continuations of this work are described to guide future efforts.

Chapter Two

Background and Notation

2.1 Introduction

Before proceeding to present spectral distribution approximation methods for random network consensus iteration matrices and the proposed consensus acceleration filter design methods, this chapter provides background context. Section 2.2 defines graph concepts and notation used throughout the thesis, including basic definitions, random graph models, and useful notions of symmetry. Section 2.3 considers tools from random matrix theory to describe the spectral behavior of large-scale random matrices along with classic example results, which will be furthered by discussion of Girko's methods in Chapter 3. Section 2.4 introduces distributed average consensus, for which accelerated convergence of the state dynamics motivates this work. Section 2.5 discusses graph signal processing terminology so that consensus acceleration filters can be understood from the graph signal processing perspective. Section 2.6 describes the consensus acceleration filter design problem, including analysis of related literature and existing approaches. Finally, Section 2.7 discusses Chebyshev filter design methods, of which the filter design methods presented in this thesis are modified versions. Finally, Section 2.8 summarizes the presented background concepts.

2.2 Graphs and Networks

This thesis studies spectral distribution approximation and graph filter design problems for networks described by graphs that can be undirected or directed, deterministic or random, constant or time-varying, and with distribution having or lacking various types of symmetries. This section introduces basic concepts, definitions, and notation regarding graphs and networks that will be used throughout the document.

An undirected graph (undirected network) consists of a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which $\mathcal{V}$ denotes a set of nodes (vertices) and $\mathcal{E}$ denotes a set of undirected links (undirected edges) that are unordered pairs $\{v_i, v_j\} = \{v_j, v_i\}$ of nodes $v_i, v_j \in \mathcal{V}$. Defined similarly, a directed graph (directed network) consists of a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which $\mathcal{V}$ again denotes a set of nodes (vertices) and $\mathcal{E}$ instead denotes a set of directed links (directed edges) that are ordered pairs (v_i, v_j) of nodes $v_i, v_j \in \mathcal{V}$, with v_i as the tail node and v_j as the head node. Note that self-loops, links formed as (v_i, v_j) for $i = j$, are typically excluded from these definitions. A subgraph $\mathcal{G}_{\text{sub}}$ of graph $\mathcal{G}$ is a graph $(\mathcal{V}_{\text{sub}}, \mathcal{E}_{\text{sub}})$ with $\mathcal{V}_{\text{sub}} \subseteq \mathcal{V}$ and $\mathcal{E}_{\text{sub}} \subseteq \mathcal{E}$, and $\mathcal{G}$ is said to be a supergraph of $\mathcal{G}_{\text{sub}}$.

For undirected graphs, the degree of a node $d(v)$ is the number of undirected links incident to that node. For directed graph nodes, the in-degree $d_{\text{in}}(v)$ of a node is the number of directed links with that node as head and the out-degree $d_{\text{out}}(v)$ of a node is the number of links with the node as tail. Let $\mathcal{D}(\mathcal{G})$ be the diagonal matrix of node degrees for undirected graphs. Let $\mathcal{D}_{\text{in}}(\mathcal{G})$ and $\mathcal{D}_{\text{out}}(\mathcal{G})$ be the diagonal matrices of node in-degrees and out-degrees, respectively, for directed graphs.

For both undirected and directed graphs, the graph adjacency matrix $\mathcal{A}(\mathcal{G})$ encapsulates the graph information with $\mathcal{A}(\mathcal{G})_{ij} = 1$ if $(v_j, v_i) \in \mathcal{E}$ and $\mathcal{A}(\mathcal{G})_{ij} = 0$ if $(v_j, v_i) \notin \mathcal{E}$. For undirected graphs, normalized versions of the adjacency matrix can variously be defined, such as the row-normalized adjacency matrix $\mathcal{A}_R(\mathcal{G}) = \mathcal{D}(\mathcal{G})^{-1} \mathcal{A}(\mathcal{G})$, the column-normalized adjacency matrix $\mathcal{A}_C(\mathcal{G}) = \mathcal{A}(\mathcal{G}) \mathcal{D}(\mathcal{G})^{-1}$, the symmetrically-normalized adjacency matrix $\mathcal{A}_S(\mathcal{G}) = \mathcal{D}(\mathcal{G})^{-1/2} \mathcal{A}(\mathcal{G}) \mathcal{D}(\mathcal{G})^{-1/2}$, and the spectrally normalized adjacency matrix $\mathcal{A}_E(\mathcal{G}) = \mathcal{A}(\mathcal{G}) / \rho(\mathcal{A}(\mathcal{G}))$ where ρ is the spectral radius. For directed graphs, the definitions for these normalized adjacency matrices must be adjusted in terms of the in-degrees and out-degrees, with the row-normalized adjacency matrix $\mathcal{A}_R(\mathcal{G}) = \mathcal{D}_{\text{in}}(\mathcal{G})^{-1} \mathcal{A}(\mathcal{G})$, the column-normalized adjacency matrix $\mathcal{A}_C(\mathcal{G}) = \mathcal{A}(\mathcal{G}) \mathcal{D}_{\text{out}}(\mathcal{G})^{-1}$, and the spectrally normalized adjacency matrix $\mathcal{A}_E(\mathcal{G}) = \mathcal{A}(\mathcal{G}) / \rho(\mathcal{A}(\mathcal{G}))$ where ρ is the spectral radius.

For undirected graphs, the Laplacian matrix closely relates to the adjacency matrix and degree matrix by $\mathcal{L}(\mathcal{G}) = \mathcal{D}(\mathcal{G}) - \mathcal{A}(\mathcal{G})$. The Laplacian matrix has the vector of ones $\mathbf{1}$

as both a left and right eigenvector of $\mathcal{L}(\mathcal{G})$ with eigenvalue 0. For this reason consensus iteration matrices for undirected networks are sometimes defined from the Laplacian by $W(\mathcal{G}) = I - \alpha\mathcal{L}(\mathcal{G})$ for some α because $W(\mathcal{G})\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^\top W(\mathcal{G}) = \mathbf{1}^\top$, an important property for consensus iteration matrices. For directed graphs, directed versions of the Laplacian matrix are sometimes defined using the in degrees $\mathcal{L}_{\text{in}}(\mathcal{G}) = \mathcal{D}_{\text{in}}(\mathcal{G}) - \mathcal{A}(\mathcal{G})$ or using the out-degrees $\mathcal{L}_{\text{out}}(\mathcal{G}) = \mathcal{D}_{\text{out}}(\mathcal{G}) - \mathcal{A}(\mathcal{G})$. These each have $\mathbf{1}$ as either a left-eigenvector or a right-eigenvector. As with the adjacency matrix, normalized versions of the Laplacian can be defined, with the row-normalized Laplacian especially important to this thesis work. For undirected graphs these are given by the row-normalized Laplacian matrix $\mathcal{L}_R(\mathcal{G}) = I - \mathcal{A}_R(\mathcal{G}) = I - \mathcal{D}(\mathcal{G})^{-1}\mathcal{A}(\mathcal{G})$, the column-normalized Laplacian matrix $\mathcal{L}_C(\mathcal{G}) = I - \mathcal{A}_C(\mathcal{G}) = I - \mathcal{A}(\mathcal{G})\mathcal{D}(\mathcal{G})^{-1}$, the symmetrically-normalized Laplacian matrix $\mathcal{L}_S(\mathcal{G}) = I - \mathcal{A}_S(\mathcal{G}) = I - \mathcal{D}(\mathcal{G})^{-1/2}\mathcal{A}(\mathcal{G})\mathcal{D}(\mathcal{G})^{-1/2}$. For directed graphs, the definitions for these normalized Laplacian matrices must be adjusted in terms of the in-degrees and out-degrees, with the row-normalized Laplacian matrix $\mathcal{L}_R(\mathcal{G}) = I - \mathcal{A}_R(\mathcal{G}) = I - \mathcal{D}_{\text{in}}(\mathcal{G})^{-1}\mathcal{A}(\mathcal{G})$ and the column-normalized Laplacian matrix $\mathcal{L}_C(\mathcal{G}) = I - \mathcal{A}_C(\mathcal{G}) = I - \mathcal{A}(\mathcal{G})\mathcal{D}_{\text{out}}(\mathcal{G})^{-1}$.

For random networks, the associated graph has uncertain structure described by a graph-valued random variable. When the network exists over several time iterations, it may either remain constant or vary over time. In addition to constant (non-time-varying) random networks, this thesis also considers independently switching networks, a simple time-varying network model. In such switching models, the network can change at each iteration according to the result of an independent Bernoulli trial, with the new network formed as a new sample from the random network distribution independent from the previous network value.

An example class of random network models, Bernoulli link-percolation models include links from a supergraph that describes all potential connections according to independent random variables for each supergraph link. The simplest of these, the Erdős-Rényi model, consists of a network with N nodes in which links between each pair of nodes are included according to independent trials with some probability θ [29]. Because all links are possible, this represents a Bernoulli percolation model with a complete supergraph [30]. The adjacency matrices of Erdős-Rényi networks are closely associated with Wigner matrices and are well studied in random matrix theory [31]. They also have important asymptotic connectedness properties [32]. Stochastic block models represent a generalization of Erdős-Rényi networks. In random networks described by stochastic block models, each node belongs to a population, and links between each pair of nodes are included according to independent trial with some probability that depends on the two node populations [33]. Due to the availability of compatible random matrix theory results [9], stochastic block models with various properties will provide the example distribution for many of the simulations included later in the thesis.

For the analysis of random matrices arising from random graphs, symmetry properties of the random graph distribution will be important for computational simplification. A graph automorphism is a permutation ϕ of the graph nodes that preserves the graph links, such that $(v_i, v_j) \in \mathcal{E}$ if and only if $(\phi(v_i), \phi(v_j)) \in \mathcal{E}$. A graph is said to be node-transitive if for every pair of nodes $v_i, v_j \in \mathcal{V}$ there is a graph automorphism ϕ such that $\phi(v_i) = v_j$. Automorphisms and node-transitivity can be naturally extended to random graph distributions, with an automorphism defined as a permutation of the random graph nodes that preserves the random graph distribution. Define a random graph distribution to be node-transitive if for every pair of nodes $v_i, v_j \in \mathcal{V}$ there is an automorphism of the random graph distribution such that $\phi(v_i) = v_j$. For directed graphs, node-transitive symmetry of a random graph distribution must be distinguished from link-reversability-symmetry of a random graph distribution, which corresponds to transpose-symmetry of the associated random adjacency matrix distribution. Of course, all undirected random graph distributions are automatically transpose-symmetric.

2.3 Random Matrix Theory

Because graph signal processing on random networks requires a random shift matrix respecting the random network topology, results from the mathematical theory of random matrices have relevance for the filter design work in this thesis. This section introduces background context for random matrix theory definitions, transforms, and examples in a general sense, without examining methods specifically applicable to matrices arising form random networks. Chapter 3 discusses random matrix theory results by Girko that this thesis applies to random network models, relying on the notation discussed in this section.

While the joint distribution of eigenvalues would be useful, it is unobtainable in practice for most random matrix distributions. An exception, the Gaussian ensembles, which are among the simplest random matrices, provide models for which the joint distribution of the eigenvalues may be explicitly found [34]. However, the joint distribution of the eigenvalues is inaccessible for most other random matrix models. Nevertheless, for suitable random matrix distributions, the asymptotic behavior of the eigenvalues as the matrix size increases may be described through the empirical distribution and empirical density functions built from the set of random matrix eigenvalues.

For a $N \times N$ matrix Ξ_N with real eigenvalues (e.g., Hermitian), the empirical spectral distribution F_{Ξ_N} and corresponding empirical spectral density f_{Ξ_N} are defined as

$$F_{\Xi_N}(x)=\frac{1}{N}\sum_{i=1}^{i=N}\chi(x\leq\lambda_i(\Xi_N)) \tag{2.1}$$

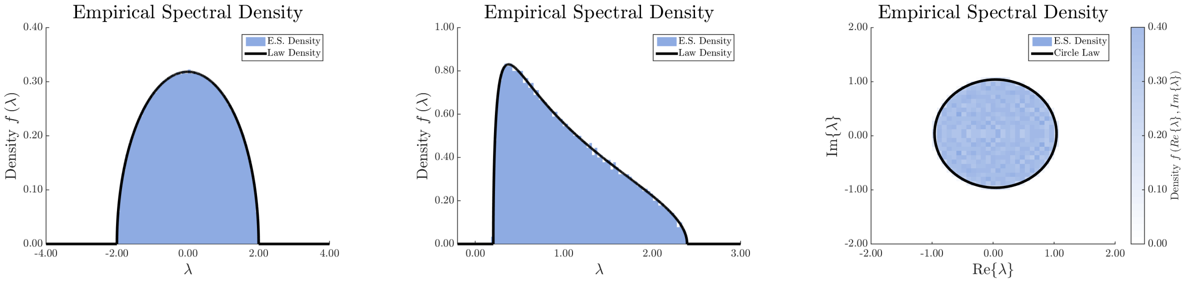
$$f_{\Xi_N}(x)=\frac{1}{N}\sum_{i=1}^{i=N}\delta(x-\lambda_i(\Xi_N)) \tag{2.2}$$

where χ is the indicator function and δ is the Dirac delta function. For a $N \times N$ matrix Ξ_N with potentially complex eigenvalues, the empirical spectral distribution F_{Ξ_N} and corresponding empirical spectral density f_{Ξ_N} are defined as

$$F_{\Xi_N}(x,y)=\frac{1}{N}\sum_{i=1}^{i=N}\chi(x\leq\mathrm{Re}\{\lambda_i(\Xi_N)\},y\leq\mathrm{Im}\{\lambda_i(\Xi_N)\}) \tag{2.3}$$

$$f_{\Xi_N}(x,y)=\frac{1}{N}\sum_{i=1}^{i=N}\delta(x-\mathrm{Re}\{\lambda_i(\Xi_N)\},y-\mathrm{Im}\{\lambda_i(\Xi_N)\}) \tag{2.4}$$

where χ is the indicator function and δ is the two-dimensional Dirac delta function. Although the empirical spectral distribution and density are random functions depending on the random eigenvalues, for large-scale random matrices these functions may approach a



(a) Wigner Semicircular Law (b) Marchenko-Pastur Law (c) Girko Circular Law

Figure 2.1: The above figures show example random matrix asymptotic ESD laws. Each plot shows the deterministic law (black curve/contour) and the empirical eigenvalue density histogram for a single sample of the random matrix (blue shaded).

deterministic limit. Well known examples that cover simple, unstructured cases include the Wigner semicircular law [6], the Marchenko-Pastur law [7, 8], and the Girko circular law [9]. Each of these cases will be briefly described below to provide context.

The Wigner semicircular law describes the limiting spectral distribution of Wigner matrices, Hermitian matrices Ξ_N such that the entries $\Xi_{N,ij}$ for $i \leq j$ are independent and identically distributed with mean $E\left[\Xi_{N,ij}\right] = 0$ and variance $E[|\Xi_{N,ij}|^2] = 1/N$ [6]. Asymptotically as N increases, the sequence of empirical spectral distributions F_{Ξ_N} for a sequence of Wigner matrices almost surely converges in measure to a deterministic distribution with density function [6]

$$f(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & x \in \mathbb{R}[-2,2] \\ 0 & x \notin \mathbb{R}[-2,2] \end{cases}. \tag{2.5}$$

For a Wigner matrix with independent, identically distributed entries selected with equal probability from points $\pm\sqrt{1/N}$, Figure 2.1a shows the Wigner semicircular law density (black curve) along with the binned empirical spectral density of a simulated matrix of size $N = 10^4$ (blue shaded).

The Marchenko-Pastur law describes the limiting spectral distribution of certain $N \times N$ Gram matrices $\Xi_N = X_{N,M(N)}X^*_{N,M(N)}$ where the dimensions $N \times M(N)$ of $X_{N,M(N)}$ satisfy

$$\lim_{N\to\infty} \frac{N}{M(N)} = c > 0 \tag{2.6}$$

and where the entries $X_{N,M(N),ij}$ are independent and identically distributed with mean $E\left[X_{N,M(N),ij}\right] = 0$ and variance $E\left[\left|X_{N,M(N),ij}\right|^2\right] = 1/M(N)$ [8]. Asymptotically as N increases, the sequence of empirical spectral distributions F_{Ξ_N} for a sequence of these matrices almost surely converges in measure to a deterministic distribution with density function

$$f(x) = \begin{cases} (1{-}c^{-1})\delta(x){+}... & c > 1{,}... \\ \frac{1}{2\pi cx}\sqrt{(x{-}a)(b{-}x)} & x \in \mathbb{R}[a{,}b] \\ & c < 1{,}... \\ \frac{1}{2\pi cx}\sqrt{(x{-}a)(b{-}x)} & \\ & x \in \mathbb{R}[a{,}b] \\ 0 & x \notin \mathbb{R}[a{,}b] \end{cases} \tag{2.7}$$

where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$ [8]. For a Gram matrix with $N/M(N) = 0.3$ and where the independent, identically distributed entries of $X_{N,M(N)}$ are selected with equal probability from points $\pm\sqrt{1/M}$, Figure 2.1b shows the Marchenko-Pastur law density (black curve) along with the empirical spectral density of a simulated matrix Ξ_N formed from $X_{N,M(N)}$ of size $N = 3 \times 10^3$, $M = 10^4$ (blue shaded).

The Girko circular law describes the limiting spectral distribution of non-Hermitian matrices Ξ_N such that the entries $\Xi_{N,ij}$ for all i, j are independent and identically distributed with mean $E\left[\Xi_{N,ij}\right] = 0$ and variance $E[|\Xi_{N,ij}|^2] = 1/N$ [8]. Asymptotically as N increases, the sequence of empirical spectral distributions F_{Ξ_N} for a sequence of the above matrices almost surely converges in measure to a deterministic distribution with density function

$$f(x_R, x_I) = \begin{cases} \frac{1}{\pi} & |x_R + x_I i| \le 1 \\ 0 & |x_R + x_I i| > 1 \end{cases}, \tag{2.8}$$

which is a uniform distribution on the unit disk. For a matrix with independent, identically distributed entries selected with equal probability from points $\pm\sqrt{1/N}$, Figure 2.1c shows the outline of the Girko circular law disk (black closed curve) along with the binned empirical spectral density of a simulated matrix of size $N = 10^4$ (blue shaded).

The preceding limiting spectral distribution laws all describe random matrices with independent, identically distributed entries, which do not describe random network adjacency matrices apart from the Erdős-Rényi case. Two methods suitable for analysis of adjacency matrices of large-scale random undirected networks are presented in Chapter 3, namely Girko's K1 equation for symmetric matrices with independent entries (except as related by symmetry), which is introduced in Section 3.2, and Girko's K27 equation for symmetric matrices with independent block submatrices (except as related by symmetry), which is introduced in Section 3.3. A method suitable for analysis of adjacency matrices of large-scale random directed networks is also introduced in Chapter 3, namely Girko's K25 equation for non-Hermitian matrices with independent entries in Section 3.4. Each of these methods requires solving a potentially large system of non-linear equations involving the matrix distribution parameters. Chapter 3 also enhances this background by discussing computational simplifications that enable faster numerical computation of approximate empirical spectral distributions and providing detailed application to example random network models.

Analysis of the empirical spectral distribution for Hermitian matrices often involves the Stieltjes transform [8]

$$S_F\left(z\right) = \int_{-\infty}^{\infty} \frac{1}{x-z} dF\left(x\right), \quad \operatorname{Im}\left\{z\right\} \neq 0. \tag{2.9}$$

The Stieltjes transform can be inverted to obtain the corresponding distribution and density by computing the following expressions [8].

$$F\left(x\right) = \lim_{\epsilon \to 0^+} \frac{1}{\pi} \int_{-\infty}^{x} \operatorname{Im}\left\{S_F\left(\lambda + \epsilon i\right)\right\} d\lambda \tag{2.10}$$

$$f\left(x\right) = \lim_{\epsilon \to 0^+} \frac{1}{\pi} \operatorname{Im}\left\{S_F\left(x + \epsilon i\right)\right\} \tag{2.11}$$

For the empirical spectral distribution F_{Ξ_N} of an $N \times N$ Hermitian matrix Ξ_N, the Stieltjes function computes to

$$S_{F_{\Xi_N}}(z) = \frac{1}{N}\operatorname{tr}\big((\Xi_N - zI_N)^{-1}\big), \ \operatorname{Im}\{z\} \neq 0, \tag{2.12}$$

the normalized trace of the resolvent $(\Xi_N - zI_N)^{-1}$ [8]. Girko's methods for symmetric matrices compute the Stieltjes transform of a deterministic approximation to the empirical spectral distribution, and the inversion formula in (2.10)-(2.11) must be used to obtain the approximate distribution and density functions.

2.4 Distributed Average Consensus

Distributed average consensus refers to a network agreement task in which each network node has an initial scalar data element and must compute the average of these data elements through a distributed process [10]. Through several network iterations, the nodes update their data according to a linear combination of their local data and data received by communicating with neighboring nodes [11]. Thus, for constant networks, this process can be described by the linear dynamics

$$\mathbf{x}_n = W\left(\mathcal{G}\right)\mathbf{x}_{n-1} \tag{2.13}$$

where the node states at iteration n are collected in vector $\mathbf{x}_n$, the initial node data elements are collected in vector $\mathbf{x}_0$, and the network iteration matrix $W\left(\mathcal{G}\right)$ respects the local structure of the network as described by graph $\mathcal{G}$. The state vector asymptotically converges to the average consensus value

$$\lim_{n\to\infty}\mathbf{x}_n = J_{\boldsymbol{\ell}}\mathbf{x}_0 = \left(\frac{\boldsymbol{\ell}^\top\mathbf{x}_0}{\boldsymbol{\ell}^\top\mathbf{1}}\right)\mathbf{1} \tag{2.14}$$

if the consensus iteration matrix satisfies the conditions

$$W\left(\mathcal{G}\right)\mathbf{1} = \mathbf{1},\ \ \boldsymbol{\ell}^\top W\left(\mathcal{G}\right) = \boldsymbol{\ell}^\top,\ \ \rho\left(W\left(\mathcal{G}\right) - J_{\boldsymbol{\ell}}\right) < 1, \tag{2.15}$$

where $\boldsymbol{\ell}$ is the left eigenvector of $W\left(\mathcal{G}\right)$ corresponding to eigenvalue $\lambda = 1$, $J_{\boldsymbol{\ell}} = \mathbf{1}\boldsymbol{\ell}^\top/\boldsymbol{\ell}^\top\mathbf{1}$ is the $\boldsymbol{\ell}$-weighted average consensus matrix, and ρ is the spectral radius. That is, the state at each node approaches the average of the initial data elements weighted according to the corresponding entries of $\boldsymbol{\ell}$. The ability to reach distributed average consensus has practical relevance in applications such as processor load balancing [35], sensor data fusion [36], coordination of multi-agent systems [37], and distributed inference [11].

Faster convergence of the system (2.13) implies that a given level of accuracy can be achieved in fewer iterations, thereby reducing communication cost, or that improved accuracy can be achieved in a given number of iterations. Under the conditions (2.14), the system converges exponentially at a rate governed by $\ln\left[\rho\left(W\left(\mathcal{G}\right)-J_{\boldsymbol{\ell}}\right)\right]$ as can be seen by applying Gelfand's formula to the asymptotic per-iteration convergence rate.

$$\lim_{k\to 0}\left\|W\left(\mathcal{G}\right)^{k}-J_{\boldsymbol{\ell}}\right\|^{1/k}=\lim_{k\to 0}\left\|\left(W\left(\mathcal{G}\right)-J_{\boldsymbol{\ell}}\right)^{k}\right\|^{1/k}=\rho\left(W\left(\mathcal{G}\right)-J_{\boldsymbol{\ell}}\right) \tag{2.16}$$

Therefore, methods to achieve fast convergence focus on minimizing this spectral radius. For a given network topology $\mathcal{G}$, [38] designs the optimal iteration matrix for fast convergence. For a given iteration matrix scheme, $W\left(\cdot\right)$ and some constraints on the network topology, [39] designs the optimal topology. In contrast, the main results of this thesis (related to the published works [22–28]) design consensus acceleration filters that can be applied at each node to achieve faster convergence, as will be described with greater depth in Section 2.6.

2.5 Graph Signal Processing

Much of the increasingly vast amount of data available in the modern day exhibits nontrivial underlying structure that does not fit within classical notions of signal processing. The theory of graph signal processing has been proposed for treating data with relationships and interactions best described by complex networks. While the consensus acceleration filter design work presented in this thesis could be described without the language of graph signal processing, the periodic consensus acceleration filters designed in this thesis have a natural interpretation as shift invariant graph filters, with the consensus iteration matrix as shift matrix. Furthermore, much of the research on consensus acceleration filters is done by members of the graph signal processing research community. Therefore, this section introduces concepts from graph signal processing useful for interpreting the results of this thesis.

Within the graph signal processing framework, signals manifest as functions on the nodes of a network. The shift operator used to analyze these signals is provided by a matrix related to the network structure, such as the adjacency matrix, Laplacian matrix, or normalized versions thereof [1–5]. Decomposition of a signal according to a basis of eigenvectors of the shift operator serves a role similar to that of the Fourier Transform in classical signal processing [3]. In this context, multiplication by polynomial functions of the chosen shift operator matrix performs shift invariant filtering [2].

The eigenvalues of the shift operator matrix play the role of graph frequencies and are important in the design and analysis of graph signals and graph filters. If W is a diagonalizable graph shift operator matrix with eigenvalue λ for eigenvector $\mathbf{v}$ such that $W\mathbf{v} = \lambda\mathbf{v}$ and, for example, if a filter is implemented on the network as $p\left(W\right)$ where p is a polynomial, then $p\left(W\right)$ has corresponding eigenvalue $p\left(\lambda\right)$ for $\mathbf{v}$ by simultaneous diagonalizability of powers of W [13]. The framework of graph signal processing regards $p\left(\lambda\right)$ as the frequency response of the filter [4]. Hence, knowledge of the eigenvalues of W informs the design of the filter p when the eigenvalues of $p\left(W\right)$ should satisfy desired properties. Furthermore, the eigenvalues relate to a notion of signal complexity known as the signal total variation, which has several slightly different definitions depending on context [1, 3–5]. For purposes of motivation, taking the shift operator to be the row-normalized adjacency matrix $\mathcal{A}_R$, define the l_p total variation of the signal $\mathbf{x}$ as

$$\mathrm{TV}_{\mathcal{G}}\left(\mathbf{x}\right) = \left\|\left(I - \mathcal{A}_R\right)\mathbf{x}\right\|_p^p = \left\|\mathcal{L}_R\mathbf{x}\right\|_p^p \tag{2.17}$$

which sums over all network nodes the pth power of the absolute difference between the value of the signal $\mathbf{x}$ at each node and the average of the value of $\mathbf{x}$ at neighboring nodes [4, 5]. Thus, if $\mathbf{v}$ is a normalized eigenvector of the row-normalized Laplacian $\mathcal{L}_R$ with eigenvalue λ, $\mathbf{v}$ has total variation $|\lambda|^p$. The eigenvectors that have higher total variation can be viewed as more complex signal components in much the same way that classical signal processing views higher frequency complex exponentials.

As an application, consider a connected, undirected network on N nodes with row-normalized Laplacian $\mathcal{L}_R$ with the goal of achieving distributed average consensus via a graph filter. For such a network, there is a simple Laplacian eigenvalue $\lambda_1(\mathcal{L}_R) = 0$ corresponding to the averaging eigenvector $\mathbf{v}_1 = \mathbf{1}$ and other eigenvalues $0 < \lambda_i(\mathcal{L}_R) \leq 2$ for $2 \leq i \leq N$. Any filter p such that $p\left(0\right) = 1$ and $|p(\lambda_i(\mathcal{L}_R))| < 1$ for $2 \leq i \leq N$ will asymptotically transform an initial signal $\mathbf{x}_0$ to a weighted average consensus signal upon iterative application [16]. If the eigenvalues $\lambda_i(\mathcal{L}_R)$ for $2 \leq i \leq N$ are known, consensus can be achieved in finite time by selecting p to be the unique polynomial of degree $N-1$ with $p\left(0\right) = 1$ and $p(\lambda_i(\mathcal{L}_R)) = 0$ [13]. Note that the averaging eigenvector has total variation 0 by the above definition (2.17) and that all other, more complex eigenvectors are completely removed by the filter. Thus, a finite time consensus filter represents the most extreme version of a non-trivial lowpass filter. With polynomial filters of smaller fixed degree d, knowledge of the eigenvalues can be used to design filters of a given length for optimal consensus convergence rate, as will be discussed further in Section 2.6. This can also be attempted for situations in which the graph is a random variable leading to uncertainty in the eigenvalues, which represents the main focus of this thesis.

2.6 Consensus Acceleration Filters

This section describes background information and literature pertaining to consensus acceleration filters, which can be applied to the record of states at each node in order to achieve faster convergence. One such approach periodically applies a filter to the states every d iterations, where d is the filter degree, in addition to the typical consensus dynamics.

$$\mathbf{x}_n := \sum_{k=0}^{k=d} a_k \mathbf{x}_{n-d+k}, \qquad n \equiv 0 \pmod{d} \tag{2.18}$$

For constant network topologies, this can be understood as a graph filter since the resultant transformation is the polynomial

$$p\left(W\left(\mathcal{G}\right)\right) = \sum_{k=0}^{k=d} a_k W\left(\mathcal{G}\right)^k \tag{2.19}$$

in the iteration matrix. Note that $p\left(1\right)=1$ to satisfy the modified consensus convergence conditions.

$$p(W(\mathcal{G}))\mathbf{1}{=}\mathbf{1},\ \boldsymbol{\ell}^{\top}p(W(\mathcal{G})){=}\boldsymbol{\ell}^{\top},\ \rho(p(W(\mathcal{G})){-}J_{\boldsymbol{\ell}}){<}1 \tag{2.20}$$

The convergence rate of the modified system is governed by

$$\lim_{k\to\infty}\left\|p\left(W(\mathcal{G})^{k}\right){-}J_{\boldsymbol{\ell}}\right\|_{2}^{1/k}{=}\lim_{k\to\infty}\left\|\left(p(W){-}J_{\boldsymbol{\ell}}\right)^{k}\right\|_{2}^{1/k} \tag{2.21}$$

which equals $\rho\left(p\left(W\right)-J_{\boldsymbol{\ell}}\right)$ by Gelfand's formula [40]. Hence,

$$\frac{1}{d}\ln\left[\rho\left(p\left(W\left(\mathcal{G}\right)\right)-J_{\boldsymbol{\ell}}\right)\right] \tag{2.22}$$

gives the per-iteration exponential convergence rate. By the spectral mapping theorem [41], the eigenvalues of $p\left(W\left(\mathcal{G}\right)\right)$ are $p\left(\lambda\right)$ for every eigenvalue λ of $W\left(\mathcal{G}\right)$. Subtracting $J_{\boldsymbol{\ell}}$ eliminates the $\lambda=1$ eigenvector, so the spectral radius is the maximum response magnitude of the polynomial at the eigenvalues of $W\left(\mathcal{G}\right)$.

Exact finite time consensus provides the most extreme example of a consensus acceleration filter when the graph $\mathcal{G}$ and iteration matrix $W\left(\mathcal{G}\right)$ are known by achieving this exactly with a degree $K-1$ filter where K is the number of distinct eigenvalues of W [19]. Of course, this method designs very large degree filters when K is large, which is likely undesirable. Furthermore, it depends on exact knowledge of the eigenvalues, making it unsuitable for network models with probabilistic uncertainty. Similarly, given a known graph $\mathcal{G}$ and consensus iteration matrix $W\left(\mathcal{G}\right)$, the optimal acceleration filter of fixed degree $d\leq K-1$ can be found. This problem was posed in [16], which provides the following semidefinite program to minimize the spectral radius (where $\preccurlyeq$ and $\succcurlyeq$ denote Loewner ordering [42]).

$$\begin{aligned}
\min_{\eta,p}\quad & \eta\\
\text{s.t.}\quad & p\left(W\right)-J \preccurlyeq \quad \eta I\\
& p\left(W\right)-J \succcurlyeq -\eta I\\
& p\left(W\right)\mathbf{1} \quad = \quad \mathbf{1}
\end{aligned} \tag{2.23}$$

This approach thoroughly and adequately addresses the case in which W is deterministic and constant. Additionally, [16] also attempts to address optimal filter design for network stochastic processes. The paper proposes two methods, one based on Newton's interpolating polynomial and one based on the above semidefinite program applied to the mean matrix. However, it can be shown that both of these methods can produce undesirable results in some network model cases.

The first method designs a Newton interpolating polynomial that is zero valued and is as smooth as possible at a selected point λ^*, typically a lower bound for the least eigenvalue but possibly some other spectrum point. While satisfying the constraint $p(1) = 1$, this method determines the polynomial of degree d that has a zero at λ^* and that has $d-1$ derivatives that are zero at λ^*. This tends to produce polynomials with all zeros closely concentrated around the selected point λ^*, and, therefore, produces results with negligible improvement over increasing degree. The paper [16] claims this method shows good robustness to changing network topologies. However, this is likely due to the fact that it behaves similarly to fixed iteration consensus because all roots are near λ^*.

The second method applies the above semidefinite program to the mean of the network stochastic process, which is assumed to be mean stationary. Noting that minimizing the expected spectral radius would be a convex problem, the paper [16] also notes that this would require extensive simulation. Using Jensen's inequality, [16] provides intuition as to why the mean matrix is used to approximate the true matrix in the semidefinite program. This method takes into account more information regarding the random graph distribution than the Newton polynomial method and, thus, sometimes achieves better results. By using the mean to approximate the iteration matrix W in the semidefinite program, the method essentially approximates the eigenvalues of the matrices by the eigenvalues of the mean matrix $\mathrm{E}[W]$.

However, there are problems with this method, some of which are noted in [16]. The paper [16] observes that the filter can lead to results that diverge in switching networks with high switching rate. However, it can also be shown that anomalous and suboptimal

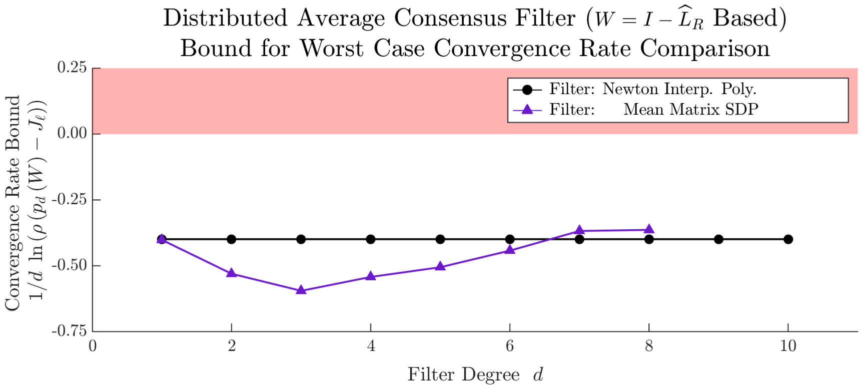


Figure 2.2: The plot shows the expected convergence rate (on a per iteration basis) for filters of varying degree based on the Newton interpolating polynomial with $\lambda^* = 0$ [16] (black circle) and based on the semidefinite program [16] (purple triangle) in (2.23) for a 3-D lattice stochastic block model network with $3 \times 3 \times 3$ populations and percolation probabilities $(0.10, 0.60, 0.70, 0.80)$. This network model allows nodes to connect if the lattice coordinates of their populations differ by at most one symbol. The results show flat behavior with increasing filter degree for the Newton polynomial method and slowing convergence rate with increasing filter degree for the semidefinite program method, displaying poor performance.

behavior can be produced in static random networks. In reality, the true eigenvalues spread from those of $\mathrm{E}\left[W\right]$, leading to poor results for some network models, especially for high filter degrees relative to the number of distinct eigenvalues of $\mathrm{E}\left[W\right]$. For instance, Figure 2.2 shows a case in which the convergence rate actually slows with increasing filter degree for a constant, random network model. For additional examples, see Figures 4.16-4.23.

Finally, it is only possible to apply the mean matrix semidefinite program method when designing filters up to degree $K - 1$ where K is the number of distinct mean matrix eigenvalues. To show this, consider the following argument. There is a unique polynomial p^*_{K-1} of degree $K - 1$ with $p^*_{K-1}\left(1\right) = 1$ and $p^*_{K-1}\left(\lambda\right) = 0$ for each distinct eigenvalue $\lambda \neq 1$ of the mean iteration matrix $\mathrm{E}\left[W\right]$, such that

$$p^*_{K-1}\left(\mathrm{E}\left[W\right]\right) = J. \tag{2.24}$$

When designing filters of degree $d > K - 1$, there are multiple solutions to the semidefinite program with the same objective function value. Namely, for any polynomial $q\left(\lambda\right)$ of degree at most $d - (K-1)$ with $q(1) = 1$, the polynomial

$$p_d^*(\lambda) = q(\lambda)\, p_{K-1}^*(\lambda) \tag{2.25}$$

is a solution for which $\eta = 0$ because the linear matrix inequalities in (2.23) become zero on the left when applied to $\mathrm{E}[W]$. However, the performance of these filters will vary significantly when applied to the true random iteration matrices, in which the eigenvalues spread from the mean matrix eigenvalues. The mean matrix semidefinite program method does not specify how to select within this class, so any results from a solver for degree $d > K-1$ would be unpredictable. Therefore, this work does not consider it in comparisons beyond degree $d \leq K-1$. This can be very restrictive when the mean matrix has few distinct eigenvalues. For instance, this method could only produce a first degree filter for an Erdős-Rényi model or for a random geographic network model, which only have two mean matrix eigenvalues due to network distribution symmetries.

The mean matrix semidefinite program method of [16] provides a reasonable approach to the problem that can be accomplished with relatively little computation. However, the mean iteration matrix eigenvalues are a coarse approximation of the potential true eigenvalue set, especially when the mean matrix has relatively few eigenvalues. Noting this, our method is inspired by [16] to optimize with respect to a notion of width or spread about each of the mean eigenvalues to better approximate the support intervals. Hence, a more complete characterization of random graph matrix eigenvalues should inform the filter design process. The requisite information may sometimes be obtained through the methods of random matrix theory, which describes asymptotic behavior of the empirical distribution of the eigenvalues for suitable random graph models [6, 8, 9, 43, 44]. The methods presented in this thesis for constant random network topologies, both undirected and directed, then employ this characterization of the empirical distribution of the eigenvalues in the optimal design of consensus acceleration filters of given degree d.

Before concluding this section, some additional related literature regarding accelerated consensus algorithms will be explored. An alternative approach to filtering in which random link rewiring in small world graphs was shown to improve the convergence rate was developed in [45]. A Chebyshev polynomial based filtering scheme was detailed in [17], in which a sequence of iteration matrix polynomials are computed through the Chebyshev recurrence to produce a polynomial of increased degree at each iteration. The parameters of this algorithm define an elliptical convergence region for the iteration matrix eigenvalues, govern the convergence rate, and have optimal values determined by the largest and smallest eigenvalues. The autoregressive moving average filters designed in [12] perform filtering of graph signals by approximating specified filter responses, although the target responses are formed separately from information about the graph. More recently, some additional information regarding the eigenvalue support has been incorporated, such as in [20] where two spectral clusters are formed from knowledge of the smallest and second largest eigenvalue modulus and then used to design the filter response. Also some contemporary works also use asymptotic results for random matrices, such as [18] which approaches design for Erdős-Rényi networks by different tools than used in this thesis work.

2.7 Chebyshev Approximation

Function approximation by a polynomial of given degree on some interval of interest frequently appears in signal processing filter design problems. When the worst case absolute error gives the criterion for choosing the approximating function, the solution is known as the minimax polynomial approximation [46]. For space of polynomials P_d of degree d approximating target function f over the real interval X, the problem may be expressed as

$$\min_{p \in P_d} \max_{x \in X} \left| f\left(x\right) - p\left(x\right) \right|. \tag{2.26}$$

Furthermore, with the additional consideration of a positive, continuous weight function w, a weighted version of the problem may be posed as

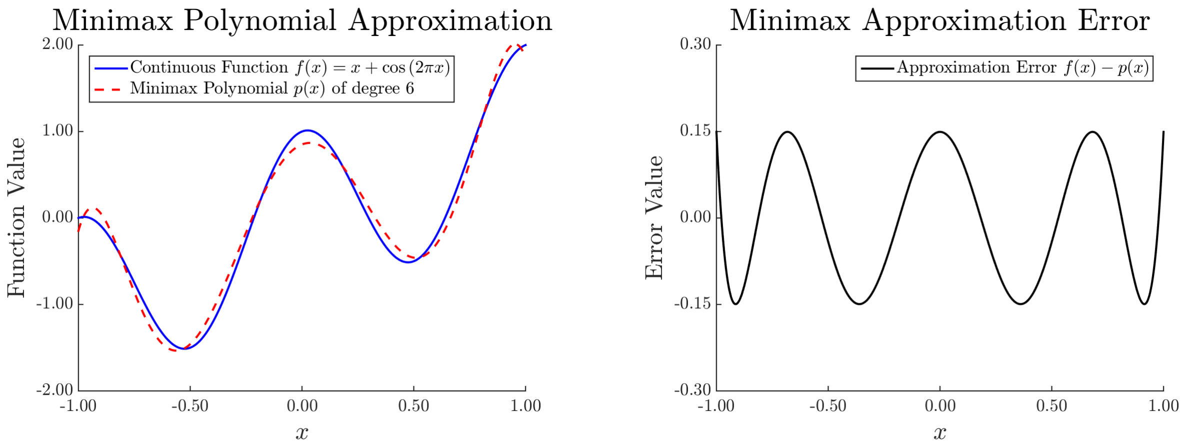
$$\min_{p \in P_d} \max_{x \in X} \left| w\left(x\right) \left(f\left(x\right) - p\left(x\right) \right) \right|. \tag{2.27}$$

The work presented in this thesis employs this optimal polynomial problem to design filters for graph signal processing, specifically for consensus.

Minimax polynomial approximations to continuous functions on an interval exhibit an important property that allows the optimal solution to be identified. The Chebyshev equioscillation theorem states that $p \in P_d$ provides the minimax approximation of degree d to continuous function f on interval X if and only if there are $d+2$ points $x_0, \ldots, x_{d+1} \in X$ ordered such that $x_0 < \ldots < x_{d+1}$ where the approximation errors $f(x_n) - p(x_n)$ are equal in magnitude and alternating in sign [46]. Furthermore, given an approximating polynomial $q \in P_d$ and set of $d+2$ points $x_0, \ldots, x_{d+1} \in X$ ordered such that $x_0 < \ldots < x_{d+1}$ where the approximation errors $f(x_n) - q(x_n)$ alternate in sign, the De la Vallée-Poussin theorem establishes that $\min_n |f(x_n) - q(x_n)|$ provides a lower bound for the minimax error [46].

Based on these theorems, the Remez exchange algorithm iteratively finds the minimax optimal polynomial approximating a continuous function on an interval [46]. More generally, the above two theorems apply not only to polynomial approximation using the monomial basis but also to approximation over any space with a basis that satisfies the Haar condition, known as a Chebyshev system. This allows the Remez algorithms to apply to other approximation schemes, such as those using Chebyshev polynomials and those using rational functions [46], which are useful for IIR filters. The algorithm begins with a set of $d+2$ ordered reference points on the interval of interest and iterates over two steps. The first of these steps computes the coefficients of the degree d polynomial with minimum absolute error at the set of reference points such that the sign of the error alternates, which is the solution to a linear system of equations. The second step selects a new set of reference points by selecting the $d+2$ local extrema of the approximation error. If the error values at these local maxima are alternating in sign and, within tolerance, equal in value, the algorithm terminates with the minimax polynomial solution [46]. With careful attention to numerics, very large degree filters can be designed, such as done in [46] with the barycentric Remez algorithm. Figure 2.3 shows the resulting polynomial approximation to a function, with oscillating error function demonstrating equal local extrema.

Alternatively, linear programming may be employed to solve minimax polynomial ap-



(a) Minimax Approximation

(b) Approximation Error

Figure 2.3: The plot in Figure 2.3a (left) shows the unweighted minimax polynomial approximation of degree $d = 6$ (red dash) to $f(x) = x + \cos(2\pi x)$ (blue solid) on the interval $\mathbb{R}[-1, 1]$. The approximation error (black solid) in Figure 2.3b (right) demonstrates the equiripple property, with the $d + 2 = 8$ error local maxima equal in value and alternating in sign.

proximation problems. Linear program formulations are sometimes preferable as they permit additional constraints to be applied to the optimization problem [46]. Given a region X on which the function f should be approximated and a subset of points $X_S \subseteq X$, the following linear program computes the coefficients $\{a_i\}_{i=1}^{i=N}$ corresponding to the basis functions $\{\phi_i\}_{i=1}^{i=N}$ that provide the minimax approximation to f on the sample points X_S.

$$\begin{aligned} \min_{\{a_n\},\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & f(x_i) - \sum_{j=1}^{N} a_j \phi_j(x_i) < \epsilon \\ & -f(x_i) + \sum_{j=1}^{N} a_j \phi_j(x_i) < \epsilon \\ & \text{for all } x_i \in X_S \end{aligned} \tag{2.28}$$

For the constant random network filter design problems in this thesis, this approach will be applied at filtering regions defined by the support of deterministic approximations to the empirical spectral distribution.

2.8 Summary

In summary, this chapter presented background concepts to provide context and notation necessary to discuss graph filter design for accelerated convergence of distributed average consensus dynamics. The terms, definitions, and random matrix models introduced for graphs and networks will appear throughout the thesis. The random matrix theory definitions will also be referenced throughout the document, especially in Chapter 3 which will introduce methods that can handle random matrix models that are more complex and relevant to networks than the classic examples displayed in this chapter. The material discussed for distributed average consensus, graph signal processing, and consensus acceleration filters, and Chebyshev filter design directly ties into the filter design methods for consensus acceleration on large-scale random networks that are presented in Chapter 4 and Chapter 5.

Chapter Three

Spectral Asymptotics: Girko's Methods

3.1 Introduction

For graph filter design on large-scale random networks, methods suitable for approximating the empirical spectral distribution of random symmetric matrices that respect the network structure provide useful information. This chapter introduces random matrix theory results by Girko that provide suitable methods to derive deterministic approximations for the empirical spectral distributions of certain random graph adjacency matrices, discussing properties of random networks that allow computational simplification. These methods are then applied to the adjacency matrices of example network models, which will later be used in the numerical simulations for consensus acceleration filter design in Chapter 4 and Chapter 5.

Specifically, this chapter discusses two results by Girko [9] that handle symmetric random matrix cases, Girko's K1 equation for random symmetric matrices with independent entries in Theorem 3.1 of Section 3.2 and Girko's K27 equation for random symmetric matrices with independent block submatrices in Theorem 3.2 of Section 3.3. Additionally, this chapter discusses a result, Girko's K25 equation, that handles random, potentially non-Hermitian matrices with independent entries in Theorem 3.3 of Section 3.4. Unlike the symmetric case, there is not an analogous result contained in [9] for matrices organized into independent block submatrices. Finally, Section 3.5 summarizes the chapter in conclusion.

Importantly, unlike many random matrix theory methods, these theorems place few conditions on the entries of the matrix, allowing non-identically distributed entries and explicitly zero entries (with some justification necessary for Girko's K25 equation), which are both important for random networks. These theorems are applicable to scaled adjacency matrices of suitable undirected link-percolation networks, random networks in with links are chosen from a supergraph according to some random variables. Note that the presentation of Theorem 3.1, Theorem 3.2, and Theorem 3.3 in this thesis differs from Girko's presentation in [9] to improve clarity and avoid notation conflicts.

3.2 Symmetric Matrices: Girko's K1 Method

For random symmetric matrices with independent entries except as related by symmetry, Girko's K1 equation, presented below as Theorem 3.1, computes an approximate empirical spectral distribution with approximation error to the empirical spectral distribution converging to zero for large-scale matrices almost surely at almost all points [9]. The result applies to a family Ξ_N of random matrices indexed by size N, provided the total absolute mean in (3.1) and total variance in (3.2) along any row or column have finite supremum and provided that the Lindberg-like condition in (3.3) that prevents the variance from concentrating in too few entries is satisfied. Under such conditions, the Stieltjes transform of a deterministic equivalent distribution can be found by solving the system of equations in (3.6). The theorem guarantees existence of a unique solution to this system of equations where the solution function has imaginary component of the function value matches the imaginary component of the input parameter. Under the additional condition (3.8), the supremum converges [9].

Theorem 3.1 (Girko's K1 Equation [9]) *Let Ξ_N be a family of symmetric, real-valued $N \times N$ random matrices indexed by size N with expectation $B_N = \mathrm{E}\left[\Xi_N\right]$, centralization $H_N = \Xi_N - B_N$, and entry variance $\sigma^2_{N,ij} = \mathrm{E}[(H_N)^2_{ij}]$. Furthermore, let Ξ_N have independent entries, except as related by symmetry, that satisfy the following regularity conditions.*

$$\sup_N \max_i \sum_{j=1}^{j=N} \left|(B_N)_{ij}\right| < \infty \tag{3.1}$$

$$\sup_N \max_i \sum_{j=1}^{j=N} \mathrm{E}\left[(H_N)^2_{ij}\right] < \infty \tag{3.2}$$

$$\lim_{N\to\infty}\max_{i}\sum_{j=1}^{j=N}\mathrm{E}\left[(H_N)_{ij}^2\,\chi\left(\left|(H_N)_{ij}\right|>\tau\right)\right]=0 \textit{ for all } \tau>0 \tag{3.3}$$

Then for almost all x,

$$\lim_{N\to\infty}\left|F_{\Xi_N}(x)-\widehat{F}_{\Xi_N}(x)\right|=0 \tag{3.4}$$

almost surely, where $\widehat{F}_{\Xi_N}$ is the distribution with Stieltjes transform

$$S_{\widehat{F}_{\Xi_N}}(z)=\frac{1}{N}\sum_{k=1}^{k=N}C_{kk}(z),\quad \mathrm{Im}\left\{z\right\}\neq 0 \tag{3.5}$$

and the functions $C_{kk}(z)$ satisfy the system of equations

$$C_{kk}(z)=\left[\left(B_N-zI_N-\left(\delta_{\ell j}\sum_{s=1}^{s=N}C_{ss}(z)\mathrm{E}\left[(H_N)_{js}^2\right]\right)_{\ell,j=1}^{\ell,j=N}\right)^{-1}\right]_{kk} \tag{3.6}$$

for $k=1,\ldots,N$. The notation $(\cdot)_{\ell,j=1}^{\ell,j=N}$ indicates a matrix built from the parameterized contents of the parentheses, such that $M=(M_{\ell j})_{\ell,j=1}^{\ell,j=N}$, and $\delta_{\ell j}$ is the Kronecker delta function. There exists a unique solution $C_{kk}(z)$ for $k=1,\ldots,N$ to the system of equations (3.6) *among*

$$L=\{X(z)\in\mathbb{C}|X(z) \textit{ analytic}, \mathrm{Im}\{z\}\mathrm{Im}\{X(z)\}>0\}. \tag{3.7}$$

Furthermore, if

$$\inf_{i,j}N\,\mathrm{E}\left[H_{ij}^2\right]\geq c>0, \tag{3.8}$$

then

$$\lim_{N\to\infty}\sup_{x}\left|F_{\Xi_N}(x)-\widehat{F}_{\Xi_N}(x)\right|=0 \tag{3.9}$$

almost surely, where $\widehat{F}_{\Xi_N}$ is defined as above.

For every point x at which the value $\widehat{f}_{\Xi_N}(x)$ is required, the solution to (3.6) must be found at $z=x+\epsilon i$ for a small $\epsilon>0$, such that the density can be approximately computed by approximately inverting the Stieltjes transform according to (2.11). For clarity, define

$C\left(z\right)$ as the full-matrix solution to (3.6) of which $C_{kk}\left(z\right)$ is the kth diagonal entry.

$$C(z){=}\left(B_N{-}zI_N{-}\left(\delta_{ij}\sum_{s=1}^{s=N}C_{ss}(z)\mathrm{E}\left[\left(H_N\right)^2_{js}\right]\right)_{\ell,j=1}^{\ell,j=N}\right)^{-1} \tag{3.10}$$

For any random matrix model satisfying the conditions, the solution $C\left(z\right)$ can be found through brute force by performing an iterative fixed point search for the unique solution guaranteed to exist by Theorem 3.1 from an initial candidate values for the entries $C_{kk}\left(z\right),\ k = 1,\dots N$. However, this involves repetitively inverting $N \times N$ matrices for large N, a computationally costly step.

The computational cost can be considerably reduced if the symmetry group of the random matrix model with respect to equal row and column permutations acts transitively on $\{1,\dots,N\}$. Under this condition, it is clear that $C_{kk}\left(z\right) = C_{\ell\ell}\left(z\right) := c\left(z\right)$ and that $\sum_{s=1}^{s=N}\mathrm{E}\left[H_{ks}^2\right] = \sum_{s=1}^{s=N}\mathrm{E}\left[H_{\ell s}^2\right]$ for all $k,\ell = 1,\dots,N$ by symmetry. Therefore, the last term in (3.10) becomes a scalar matrix $c\left(z\right)\left(\sum_{s=1}^{s=N}\sigma_{js}^2\right)I_N$. Note that $c\left(z\right) = \frac{1}{N}\operatorname{tr}C\left(z\right)$. By applying the trace function to each part of (3.10) and writing the trace of the right expression as the sum of eigenvalues, for such matrices the Stieltjes transform of the approximate empirical spectral distribution can be found via iterative search using the following modified equation.

$$c\left(z\right) = \frac{1}{N}\sum_{k=1}^{k=N}\frac{1}{\lambda_k\left(B_N\right) - z - c\left(z\right)\sum_{s=1}^{s=N}\sigma_{sj}^2} \tag{3.11}$$

For random network models that are node-transitive in distribution, the associated random adjacency matrices will have this property and can be analyzed through (3.11) (if they satisfy the conditions of Theorem 3.1). That is, no node is statistically distinguishable from any other node in the random network. (Note that the actual networks sampled from the distribution need not be node-transitive.) The iterative process for finding the solution to this system of equations does not depend on the matrix size but only on the number of distinct mean matrix eigenvalues, as equal terms in the sum can be grouped and multiplied by algebraic multiplicity.

For use in the filter design problems for undirected networks in Section 4.2, the empirical spectral distribution of the row-normalized adjacency matrix must be approximated. In order to apply Girko's method, a random matrix with independent entries must be defined such that the spectral density will be a good approximation. For this purpose, initial work [22–28] used the scaled adjacency matrix $\Xi'_N = \frac{1}{\gamma_1}\mathcal{A}(\mathcal{G}_N)$ where $\gamma_1 = \lambda_{\max}(\mathrm{E}[\mathcal{A}(\mathcal{G})])$, which is also the expected row sum if all rows have equal expected sum like in node-transitive distributions. However, this introduces a bias in the expected largest eigenvalue if $\mathrm{E}[\lambda_{\max}(\mathcal{A}(\mathcal{G}_N))] > \lambda_{\max}(\mathrm{E}[\mathcal{A}(\mathcal{G}_N)])$ (i.e., $\mathrm{E}[\rho(\mathcal{A}(\mathcal{G}_N)/\gamma_1)] = \mathrm{E}[\rho(\mathcal{A}(\mathcal{G}_N))]/\rho(\mathrm{E}[\mathcal{A}(\mathcal{G}_N)]) \geq 1$), which should be exactly $\lambda_{\max} = 1$ for the row-normalized Laplacian. This must be true for node-transitive distributions since $\lambda_{\max}(\mathcal{A}(\mathcal{G}_N)) \geq d_{\text{avg}}(\mathcal{A}(\mathcal{G}_N))$ where $d_{\text{avg}}(\mathcal{A}(\mathcal{G}_N))$ is the average degree [47], so it follows that $\mathrm{E}[\lambda_{\max}(\mathcal{A}(\mathcal{G}_N))] \geq \mathrm{E}[d_{\text{avg}}(\mathcal{A}(\mathcal{G}_N))] = \lambda_{\max}(\mathrm{E}[(\mathcal{A}(\mathcal{G}_N))])$. (Note that previous results are not invalidated as for large matrices this bias is observed to be asymptotically eliminated. However, it is problematic for large but finite network size.) To partially correct this bias without affecting the variance, define

$$\Xi_N = \frac{1}{\gamma_1}(\mathcal{A}(\mathcal{G}_N) - \mathrm{E}[\mathcal{A}(\mathcal{G}_N)]) + \frac{1}{\gamma_2}\mathrm{E}[\mathcal{A}(\mathcal{G}_N)] \tag{3.12}$$

where $\gamma_1 = \lambda_{\max}(\mathrm{E}[\mathcal{A}(\mathcal{G}_N)])$ and $\gamma_2 = \mathrm{E}[\lambda_{\max}(\mathcal{A}(\mathcal{G}_N))]$. That this reduces the bias is unproven, but the results match significantly better in practice.

Example 3.1: Undirected SBM with Transitive Population Structure

Stochastic block models (SBMs) are random network models that generalize Erdős-Rényi networks and are used to describe networks with a population structure [33]. Each pair of nodes potentially form a link according to an independent Bernouli trial with link probability Θ_{ij} that depends on the populations containing the two nodes [33]. Note that a stochastic block model is a node-transitive random network distribution if the population graph (showing populations where nodes connect with non-zero probability), with nodes labeled with intra-population link probabilities and edges labeled by inter-population link probabilities, is a node-transitive labeled graph and the populations are all equal size. Large-scale stochastic block models can be grown by increasing the node population sizes. Therefore, the adjacency matrices of such networks provide convenient examples for which to derive spectral density approximations.

Consider an undirected stochastic block model with $M_1 \times M_2$ node populations each of size S with $N = M_1 M_2 S$ nodes in total, and label each population by an ordered pair (m_1, m_2). Suppose that two populations are adjacent, allowing their nodes to randomly connect, if they differ by at most one in each symbol, with the last symbol considered adjacent to the first symbol. Suppose that random intra-population links occur independently with probability $\theta_{N,0}$. Also suppose that random inter-population links occur independently with probability $\theta_{N,1}$ if the populations differ by one in the first symbol with equal second symbol, $\theta_{N,2}$ if the populations differ by one in the second symbol with equal first symbol, $\theta_{N,3}$ if the populations differ by one in both symbols.

Clearly, this example random network model is node-transitive. The link probability between nodes (i, j) are described by the entries of the following matrix where $\mathcal{C}_n$ is the adjacency matrix of a length n directed cycle and $1_{n\times n}$ is the $n \times n$ matrix of ones.

$$\begin{aligned}\Theta_N = &\theta_{N,0} \left(I_{M_1}\right) \otimes \left(I_{M_2}\right) \otimes \left(1_{S\times S} - I_S\right) \\ &+\theta_{N,1} \left(\mathcal{C}_{M_1} + \mathcal{C}_{M_1}^\top\right) \otimes \left(I_{M_2}\right) \otimes \left(1_{S\times S}\right) \\ &+\theta_{N,2} \left(I_{M_1}\right) \otimes \left(\mathcal{C}_{M_2} + \mathcal{C}_{M_2}^\top\right) \otimes \left(1_{S\times S}\right) \\ &+\theta_{N,3} \left(\mathcal{C}_{M_1} + \mathcal{C}_{M_1}^\top\right) \otimes \left(\mathcal{C}_{M_2} + \mathcal{C}_{M_2}^\top\right) \otimes \left(1_{S\times S}\right)\end{aligned} \tag{3.13}$$

Therefore, $E\left[A\left(\mathcal{G}\right)\right] = \Theta_N$ and the eigenvalues of the mean adjacency matrix are easily computed from the eigenvalues of undirected cycles and Kronecker product relationships. The variance of each entry is described by $\Theta_N \circ \left(1_{N\times N} - \Theta_N\right)$.

The approximate empirical spectral distribution for Ξ_N can be computed through the reduced equation (3.11) for Girko's K1 method based on its mean eigenvalues and total row variance, which can be found though appropriate transformation of the adjacency matrix mean eigenvalues and total row variance. Because the approximation will be numerically computed at a large fixed value of N and fixed probability parameters, the conditions (3.1)-(3.3) are not verifiable without making assumptions about how the model would change with increasing N. For a percolation model, such as this example, where each probability paramter scales asymptotically with N in the same way (i.e., maintain a ratio) and where each parameter governs a number of entries in each row proportional to N, the conditions are satisfied for Ξ_N if $N\theta_N \to \infty$. Then the first condition (3.1) is automatically satisfied. The second condition (3.2) is satisfied if each parameter $N\theta_N \not\to 0$. The third condition (3.3) is satisfied if additionally $N\theta_N \to \infty$.

Assume that the network model of interest is a particular point in a larger family of network models indexed by N such that the conditions are satisfied by this family of models. The resulting density will be used to approximate the empirical spectral distribution of the row-normalized adjacency matrix. Simulation results comparing the expected spectral density and approximate spectral density (of a transformation of the row-normalized adjacency matrix) appear in Figure 3.1. Filter design results using the approximate empirical spectral distribution for this model can be found in Figures 4.1-4.3 of Section 4.2 in Chapter 4 for this model with specific numerical parameters.

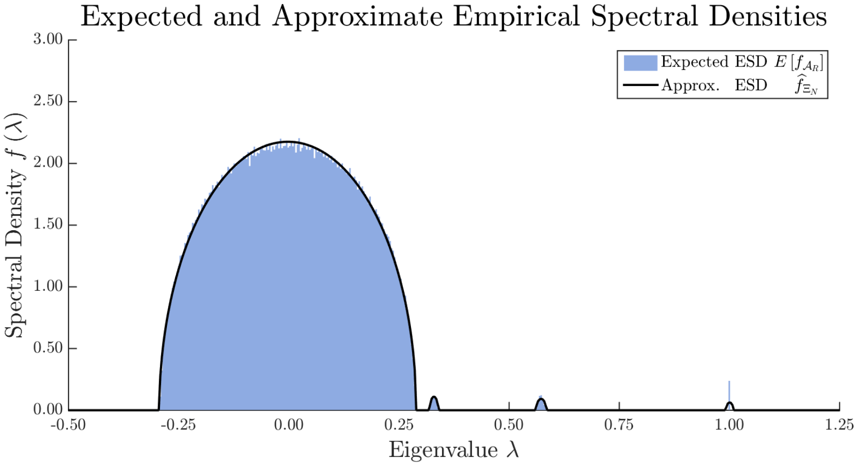


Figure 3.1: Expected empirical spectral distribution $\mathrm{E}\left[f_{\mathcal{A}_R(\mathcal{G})}\right]$ of the normalized adjacency matrix (simulated over 1000 Monte-Carlo trials) compared agains the deterministic approximation $\widehat{f}_{\Xi_N}$ computed using Girko's K1 method ($\epsilon = 10^{-5}$) for the model in Example 3.1 with $M_1 = 3$, $M_2 = 4$, $S = 100$, $\theta_0 = 0.10$, $\theta_1 = 0.07$, $\theta_2 = 0.06$, $\theta_3 = 0.02$

3.3 Symmetric Matrices: Girko's K27 Method

For random symmetric matrices with independent block submatrices except as related by symmetry (but with potentially correlated entries within block submatrices), Girko's K27 equation, presented below as Theorem 3.2, computes an approximate empirical spectral distribution with approximation error to the empirical spectral distribution converging to zero for large-scale matrices almost surely at almost all points [9]. The result applies to a family Ξ_N of random matrices indexed by size N, provided the total 2-norm of mean blocks in (3.14) and Frobenius norm squared of centralized blocks (total entry variance) in (3.15) along any block row or block column have finite supremum and provided that the Lindberg-like condition in (3.16) that prevents the variance from concentrating in too few block submatrices is satisfied. Under such conditions, the Stieltjes transform of a deterministic equivalent distribution can be found by solving the system of equations in (3.19). The theorem guarantees existence of a unique solution to this system of equations where the solution function has imaginary component of the function value matches the imaginary component of the input parameter.

Theorem 3.2 (Girko's K27 Equation [9]) *Let Ξ_N be a family of symmetric, real-valued $N \times N$ random matrices indexed by size N with expectation $B_N = \mathrm{E}\left[\Xi_N\right]$ and centralization $H_N = \Xi_N - B_N$. Let Ξ_N have a block structure with N_1 blocks of size $N_2 \times N_2$ such that $N = N_1 N_2$, and denote the block in position (i,j) by $(\Xi_N)_{[ij]}$. Furthermore, let Ξ_N have independent random matrix blocks, except as related by symmetry, that satisfy the following regularity conditions. Note that the random matrix blocks may have internal statistical dependencies.*

$$\sup_N \max_i \sum_{j=1}^{j=N_1} \left\| (B_N)_{[ij]} \right\|_2 < \infty \tag{3.14}$$

$$\sup_N \max_i \sum_{j=1}^{j=N_1} \mathrm{E}\left[\left\| (H_N)_{[ij]} \right\|_F^2 \right] < \infty \tag{3.15}$$

$$\lim_{N\to\infty}\max_{i}\sum_{j=1}^{j=N_1}\mathrm{E}\left[\left\|(H_N)_{[ij]}\right\|_F^2\chi\left(\left\|(H_N)_{[ij]}\right\|_F>\tau\right)\right]=0 \textit{ for all } \tau>0 \tag{3.16}$$

Then for almost all x,

$$\lim_{N\to\infty}\left|F_{\Xi_N}(x)-\widehat{F}_{\Xi_N}(x)\right|=0 \tag{3.17}$$

almost surely, where $\widehat{F}_{\Xi_N}$ is the distribution with Stieltjes transform

$$S_{\widehat{F}_{\Xi_N}}(z)=\frac{1}{N}\sum_{k=1}^{k=N_1}\operatorname{tr}\left(C_{[kk]}(z)\right),\quad \operatorname{Im}\{z\}\neq 0 \tag{3.18}$$

and the $N_2\times N_2$ matrix functions $C_{[kk]}(z)$ satisfy the system of equations

$$C_{[kk]}(z)=\left[\left(B_N-zI_N-\left(\delta_{\ell j}\sum_{s=1}^{s=N_1}\mathrm{E}\left[(H_N)_{[js]}C_{[ss]}(z)(H_N)_{[js]}^{\top}\right]\right)_{\ell,j=1}^{\ell,j=N_1}\right)^{-1}\right]_{[kk]} \tag{3.19}$$

for $k=1,\ldots,N_1$. The notation $(\cdot)_{\ell,j=1}^{\ell,j=N}$ indicates a matrix built from the parameterized contents of the parentheses, such that $M=\left(M_{[\ell j]}\right)_{\ell,j=1}^{\ell,j=N_1}$, and $\delta_{\ell j}$ is the Kronecker delta function. There exists a unique solution $C_{[kk]}(z)$ for $k=1,\ldots,N_1$ to the system of equations (3.6) *among*

$$L_{N_2\times N_2}=\left\{X(z)\in\mathbb{C}^{N_2\times N_2}\,\middle|\,X(z)\textit{ analytic}, \operatorname{Im}\{z\}\operatorname{Im}\{X(z)\}>0\right\}. \tag{3.20}$$

As before, for every point x at which the value $\widehat{f}_{\Xi_N}(x)$ is required, the solution to (3.19) must be found at $z=x+\epsilon i$ for a small $\epsilon>0$, such that the density can be approximately computed by approximately inverting the Stieltjes transform according to (2.11). For clarity, define $C(z)$ as the full-matrix solution to (3.19) of which $C_{[kk]}(z)$ is the kth diagonal block.

$$C(z)=\left(B_N-zI_N-\left(\delta_{\ell j}\sum_{s=1}^{s=N_1}\mathrm{E}\left[(H_N)_{[js]}C_{[ss]}(z)(H_N)_{[js]}^{\top}\right]\right)_{\ell,j=1}^{\ell,j=N_1}\right)^{-1} \tag{3.21}$$

For any random matrix model satisfying the conditions, the solution $C(z)$ can be found through brute force by performing an iterative fixed point search for the unique solution guaranteed to exist by Theorem 3.2 from an initial candidate values for the block matrix entries $C_{[kk]}(z)\,,\; k=1,\dots N$. However, this involves repetitively inverting $N\times N$ matrices for large N, a computationally costly step.

Computational simplification conditions are more complicated in this scenario. The computational cost can be considerably reduced if the symmetry group of the random matrix model with respect to equal block-row and block-column permutations acts transitively on $\{1,\dots,N_1\}$. Under this condition, it is clear by symmetry that $C_{[kk]}(z) = C_{[\ell\ell]}(z) := C_{\text{blk}}(z)$ and that $\sum_{s=1}^{s=N} \mathrm{E}\left[H_{ks}C_{\text{blk}}(z)\,H_{ks}^{\top}\right] = \sum_{s=1}^{s=N} \mathrm{E}\left[H_{\ell s}C_{\text{blk}}(z)\,H_{\ell s}^{\top}\right] := X\left[C_{\text{blk}}(z)\right]$ for all $k,\ell = 1,\dots,N$. Therefore, the last term in (3.21) becomes the block-scalar matrix $I_{N_1}\otimes X\left[C_{\text{blk}}(z)\right]$. Suppose that B_N is block-diagonalizable with diagonal blocks $\left(\Lambda_{B_N}\right)_{[kk]}$ for $k=1,\dots,N_1$. Suppose that this block diagonalization preserves $I_{N_1}\otimes X\left[C_{\text{blk}}(z)\right]$ and, furthermore, that each $\left(\Lambda_{B_N}\right)_{\text{blk}}$ is simultaneously diagonalizable with $X\left[C_{\text{blk}}(z)\right]$. Then all terms of (3.21) are simultaneously diagonalizable.

Suppose that the form of $C(z)$ is known (inferred from symmetries) with eigenvalues computed from its parameters. Finally, suppose that the eigenvalue of $X\left[C_{\text{blk}}(z)\right]$ that matches to $\lambda_k(B_N)$ and $\lambda_k(C(z))$ can be determined as a function $g_k\left(\left\{\lambda_\ell\left(C(z)\right)\right\}_{\ell=1}^{\ell=N}\right)$ of the eigenvalues of $C(z)$, through the parameters inherited by the diagonal blocks $C_{\text{blk}}(z)$. Then the eigenvalues of $C(z)$ can be written as the solution to the following system of equations.

$$\begin{aligned}\lambda_k(C(z)) = &\frac{1}{\lambda_k(B_N) - z - g_k\left(\{\lambda_\ell(C(z))\}_{\ell=1}^{\ell=N}\right)} \\ &\qquad\qquad\qquad\text{for}\quad k=1,\dots N\end{aligned} \tag{3.22}$$

As before, the solution to this system of equations describing the eigenvalues can be found through an iterative fixed point search. The iterative process for finding the solution to this system of equations would appear to depend on the matrix size N, but only depends on the number of distinct equations, which may be much fewer. Note that the sum of these eigenvalues computes the trace in (3.18), providing the Stieltjes transform of the approximate empirical spectral distribution. While the above conditions are complicated and must be verified on a case-by-case basis, they allow analysis of adjacency matrices for some node-block-transitive random network models with intra-block correlations as shown in Example 3.2.

For use in the filter design problems for undirected networks in Section 4.2, the empirical spectral distribution of the row-normalized adjacency matrix must be approximated. In order to apply Girko's method, a random matrix with independent entries must be defined such that the spectral density will be a good approximation. For this purpose, initial work [22–28] used the scaled adjacency matrix $\Xi'_N = \frac{1}{\gamma_1}\mathcal{A}\left(\mathcal{G}_N\right)$ where $\gamma_1 = \lambda_{\max}\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}\right)\right]\right)$, which is also the expected row sum if all rows have equal expected sum like in node-transitive distributions. However, this introduces a bias in the expected largest eigenvalue if $\mathrm{E}\left[\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right] > \lambda_{\max}\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right)$ (i.e., $\mathrm{E}\left[\rho\left(\mathcal{A}\left(\mathcal{G}_N\right)/\gamma_1\right)\right] = \mathrm{E}\left[\rho\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right]/\rho\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right) \geq 1$), which should be exactly $\lambda_{\max} = 1$ for the row-normalized Laplacian. This must be true for node-transitive distributions since $\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right) \geq d_{\text{avg}}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)$ where $d_{\text{avg}}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)$ is the average degree [47], so it follows that $\mathrm{E}\left[\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right] \geq \mathrm{E}\left[d_{\text{avg}}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right] = \lambda_{\max}\left(\mathrm{E}\left[\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right]\right)$. (Note that previous results are not invalidated as for large matrices this bias is observed to be asymptotically eliminated. However, it is problematic for large but finite network size.) To partially correct this bias without affecting the variance, define

$$\Xi_N = \frac{1}{\gamma_1}\left(\mathcal{A}\left(\mathcal{G}_N\right) - \mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right) + \frac{1}{\gamma_2}\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right] \tag{3.23}$$

where $\gamma_1 = \lambda_{\max}\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right)$ and $\gamma_2 = \mathrm{E}\left[\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right]$. That this reduces the bias is unproven, but the results match significantly better in practice.

Example 3.2: Modified Undirected SBM with Transitive Population Structure and Dependent Block Submatrices

Consider a modified undirected stochastic block model with M populations for which each population is composed of S_1 sub-populations of S_2 nodes, such that each population has $S = S_1 S_2$ nodes and such that the network has $N = M * S$ nodes. Dependencies between random links will be permitted if those potential links join the same unordered pair of sub-populations, such that the resulting random adjacency matrix has independent block structure (except as related by symmetry) with $N_1 = M * S_1$ and $N_2 = S_2$ in Girko's K27 theorem. Specifically, for each unordered pair of sub-populations $\{i, j\}$ with $1\leq i, j \leq N_1$, links will be random variables that are conditionally independent given another random variable (specific to that sub-population pair and independent of the random variable for any other sub-population pair) that controls the common link probability for that sub-population pair. If the two sub-populations are the same (i.e., $i = j$), let the possible link probabilities $\{\theta_{N,0,k}\}$ each be chosen with probabilities $\{\phi_{N,0,k}\}$ (Note: $\sum_k \phi_{N,0,k} = 1$). If the two sub-populations are different but part of the same population (i.e., $i \neq j$ but $mS_1 < i, j \leq (m+1)S_1$ for some integer $m = 0, ..., M-1$), let the possible link probabilities $\{\theta_{N,1,k}\}$ each be chosen with probabilities $\{\phi_{N,1,k}\}$ (Note: $\sum_k \phi_{N,1,k} = 1$). If the two sub-populations are in different populations entirely, let the possible link probabilities $\{\theta_{N,2,k}\}$ each be chosen with probabilities $\{\phi_{N,2,k}\}$ (Note: $\sum_k \phi_{N,2,k} = 1$). In order to achieve good results from Girko's K27 method, note that S_1 in particular must be large so there are a large number of blocks of the intra-sub-population type.

Clearly, this example random network model is node-transitive and node-block-transitive. The mean matrix has the following form.

$$\begin{aligned}
B_N =& \frac{1}{\gamma_2} \operatorname{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right] \\
=& \frac{1}{\gamma_2}\left(\sum_k \phi_{N,0,k}\theta_{N,0,k}\right)\left(I_M\right) \otimes \left(I_{S_1}\right) \otimes \left(1_{S_2\times S_2} - I_{S_2}\right) \\
+& \frac{1}{\gamma_2}\left(\sum_k \phi_{N,1,k}\theta_{N,0,k}\right)\left(I_M\right) \otimes \left(1_{S_1\times S_1} - I_{S_1}\right) \otimes \left(1_{S_2\times S_2}\right) \\
+& \frac{1}{\gamma_2}\left(\sum_k \phi_{N,2,k}\theta_{N,0,k}\right)\left(1_{M\times M} - I_M\right) \otimes \left(1_{S_1\times S_1}\right) \otimes \left(1_{S_2\times S_2}\right)
\end{aligned} \tag{3.24}$$

The eigenvalues of B_N are easily computed from Kronecker relationships. Collect these eigenvalues in a vector $\boldsymbol{\lambda}_B$. (Note that there are at most 8 distinct eigenvalues, so we can assume $\boldsymbol{\lambda}_B$ is 8×1.) By symmetry, it can be inferred that $C(z)$ is of the form

$$\begin{aligned}
C(z) = &\mathbf{c}_1(z)(I_M) \otimes (I_{S_1}) \otimes (I_{S_2}) \\
&+\mathbf{c}_2(z)(I_M) \otimes (I_{S_1}) \otimes (1_{S_2 \times S_2} - I_{S_2}) \\
&+\mathbf{c}_3(z)(I_M) \otimes (1_{S_1 \times S_1} - I_{S_1}) \otimes (I_{S_2}) \\
&+\mathbf{c}_4(z)(I_M) \otimes (1_{S_1 \times S_1} - I_{S_1}) \otimes (1_{S_2 \times S_2} - I_{S_2}) \\
&+\mathbf{c}_5(z)(1_{M \times M} - I_M) \otimes (I_{S_1}) \otimes (I_{S_2}) \\
&+\mathbf{c}_6(z)(1_{M \times M} - I_M) \otimes (I_{S_1}) \otimes (1_{S_2 \times S_2} - I_{S_2}) \\
&+\mathbf{c}_7(z)(1_{M \times M} - I_M) \otimes (1_{S_1 \times S_1} - I_{S_1}) \otimes (I_{S_2}) \\
&+\mathbf{c}_8(z)(1_{M \times M} - I_M) \otimes (1_{S_1 \times S_1} - I_{S_1}) \otimes (1_{S_2 \times S_2} - I_{S_2})
\end{aligned} \tag{3.25}$$

for some parameters $\mathbf{c}_1(z), \ldots \mathbf{c}_8(z)$ collected into vector $\mathbf{c}(z)$. The eigenvalues of $C(z)$, collected into a vector $\boldsymbol{\lambda}_C(z)$ and be written as

$$\boldsymbol{\lambda}_C(z) = T_C \mathbf{c}(z) \tag{3.26}$$

for appropriate transformation T_C computed from (3.25). (Note that there are at most 8 distinct eigenvalues, so we can assume $\boldsymbol{\lambda}_C(z)$ is 8×1 and T_C is 8×8. Also, note that T_C is invertible.) The diagonal blocks arise from the first two terms such that

$$\begin{aligned}
C_{\text{blk}}(z) = &\mathbf{c}_1(z)(I_{S_2}) \\
&+\mathbf{c}_2(z)(1_{S_2 \times S_2} - I_{S_2})
\end{aligned} \tag{3.27}$$

To compute $X[C_{\text{blk}}(z)]$ note that

$$\begin{aligned}
X_{k\ell} = &\mathbf{c}_1(z) \sum_{s=1}^{s=N_1} \sum_{\substack{r,t=1 \\ r=t}}^{r,t=N_2} \mathrm{E}\left[\left((H_N)_{[js]} \right)_{kr} \left((H_N)_{[js]} \right)_{\ell t} \right] \\
&+\mathbf{c}_2(z) \sum_{s=1}^{s=N_1} \sum_{\substack{r,t=1 \\ r\neq t}}^{r,t=N_2} \mathrm{E}\left[\left((H_N)_{[js]} \right)_{kr} \left((H_N)_{[js]} \right)_{\ell t} \right]
\end{aligned} \tag{3.28}$$

There are many cases for $\mathrm{E}\left[\left((H_N)_{[js]}\right)_{kr}\left((H_N)_{[js]}\right)_{\ell t}\right]$ where $H_N = \frac{1}{\gamma_1}\left(\mathcal{A}(\mathcal{G}) - \mathrm{E}\left[\mathcal{A}(\mathcal{G})\right]\right)$. These depend on whether $k = t$ or $k \neq t$, whether $r = s$ or $r \neq s$, and whether $j = s$, $j \neq s$ but j, s are in the same population, or j, s are in different populations. Thus, there are a total of 12 cases, which will not be listed here, but are easily computed. Hence,

$$\begin{aligned} X\left[C_{\mathrm{blk}}(z)\right] = &\left(a_1\mathbf{c}_1(z) + a_2\mathbf{c}_2(z)\right)\left(I_{S_2}\right) \\ &+ \left(a_3\mathbf{c}_1(z) + a_4\mathbf{c}_2(z)\right)\left(1_{S_2\times S_2} - I_{S_2}\right) \end{aligned} \tag{3.29}$$

and the eigenvalues of $I_{N_2} \otimes X\left[C_{\mathrm{blk}}(z)\right]$, collected into a vector $\boldsymbol{\lambda}_{I\otimes X}(z)$ can be computed as

$$\boldsymbol{\lambda}_{I\otimes X}(z) = T_X\mathbf{c}(z). \tag{3.30}$$

All terms are simultaneously diagonalizable. Because of the Kronecker form, it is easy to match corresponding eigenvalues of each term. Assume this has been done properly in the vectors $\boldsymbol{\lambda}_C(z)$, $\boldsymbol{\lambda}_{I\otimes X}(z)$, and $\boldsymbol{\lambda}_B$ and the matrices T_C and T_X. Then the functions g relating the eigenvalues in (3.22), collected into a vector function $\mathbf{g}$, are

$$\mathbf{g}\left(\boldsymbol{\lambda}_C(z)\right) = T_X T_C^{-1}\boldsymbol{\lambda}_C(z) = \boldsymbol{\lambda}_{I\otimes X}(z). \tag{3.31}$$

The approximate empirical spectral distribution for Ξ_N can be computed through the reduced system of equations (3.22) for Girko's K27 method based on the above derivations. Because the approximation will be numerically computed at a large fixed value of N and fixed probability parameters, the conditions (3.14)-(3.16) are not verifiable without making assumptions about how the model would change with increasing N. For a percolation model, such as this example, where each probability paramter scales asymptotically with N in the same way (i.e., maintain a ratio) and where each parameter governs a number of entries in each row proportional to N, the conditions are satisfied for Ξ_N if $N\theta_N \to \infty$. Then the first condition (3.14) is automatically satisfied. The second condition (3.15) is satisfied if each parameter $N\theta_N \not\to 0$. The third condition (3.16) is satisfied if additionally $N\theta_N \to \infty$.

Assume that the network model of interest is a particular point in a larger family of network models indexed by N such that the conditions are satisfied by this family of models. The resulting density will be used to approximate the empirical spectral distribution of the row-normalized adjacency matrix. Simulation results comparing the expected spectral density and approximate spectral density (of a transformation of the row-normalized adjacency matrix) appear in Figure 3.2. Filter design results using the approximate empirical spectral distribution for this model can be found in Figures 4.4-4.6 of Section 4.2 in Chapter 4 for this model with specific numerical parameters.

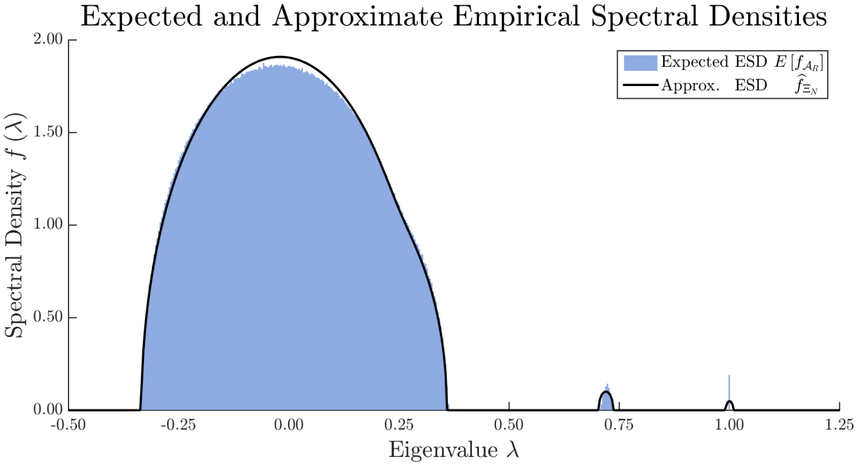


Figure 3.2: Expected empirical spectral distribution $\mathrm{E}\left[f_{\mathcal{A}_R(\mathcal{G})}\right]$ of the normalized adjacency matrix (simulated over 1000 Monte-Carlo trials) compared agains the deterministic approximation $\widehat{f}_{\Xi_N}$ computed using Girko's K27 method ($\epsilon = 10^{-5}$) for the model in Example 3.2 with $M = 5$, $S_1 = 30$, and $S_2 = 10$ and with probabilities $\{\phi_{0,k}\} = \{0.333, 0.333, 0.334\}$, $\{\theta_{0,k}\} = \{0.300, 0.600, 1.000\}$, $\{\phi_{1,k}\} = \{0.333, 0.333, 0.334\}$, $\{\theta_{1,k}\} = \{0.030, 0.060, 0.100\}$, $\{\phi_{2,k}\} = \{0.333, 0.333, 0.334\}$, $\{\theta_{2,k}\} = \{0.003, 0.006, 0.010\}$

3.4 Non-Symmetric Matrices: Girko's K25 Method

For random non-Hermitian matrices with independent entries, Girko's K25 equation, presented below as Theorem 3.3, computes an approximate empirical spectral distribution with approximation error to the empirical spectral distribution converging to zero for large-scale matrices almost surely [9]. The result applies to a family Ξ_N of random matrices indexed by size N, provided the total absolute mean in (3.32) and total variance in (3.33) along any row or column have finite supremum and provided other conditions (3.34)-(3.35) hold. Note that while it would appear that the condition in (3.34) would prevent application of this theorem to network adjacency matrices, this objection can be dismissed for reasons discussed later. Under these conditions, a deterministic equivalent distribution can be found by solving the system of equations in (3.39)-(3.40) and then computing the density function through (3.37)-(3.38). The theorem guarantees existence of a unique solution to this system of equations where the solution function has imaginary component of the function value matches the imaginary component of the input parameter.

Theorem 3.3 (Girko's K25 Equation [9]) *Let Ξ_N be a family of complex-valued $N \times N$ random matrices indexed by size N with expectation $B_N = \mathrm{E}\left[\Xi_N\right]$, centralization $H_N = \Xi_N - B_N$, and entry variance $\sigma^2_{N,ij} = \mathrm{E}[|\left(H_N\right)_{ij}|^2]$. Furthermore, let Ξ_N have independent entries that satisfy the following regularity conditions.*

$$\sup_N \max_i \sum_{j=1}^{j=N} \left|\left(B_N\right)_{ij}\right| < \infty, \ \sup_N \max_j \sum_{i=1}^{i=N} \left|\left(B_N\right)_{ij}\right| < \infty \tag{3.32}$$

$$\sup_N \max_i \sum_{j=1}^{j=N} \mathrm{E}\left[\left(H_N\right)^2_{ij}\right] < \infty, \ \sup_N \max_j \sum_{i=1}^{i=N} \mathrm{E}\left[\left(H_N\right)^2_{ij}\right] < \infty \tag{3.33}$$

Also let Ξ_N satisfy

$$\mathrm{E}\left[\left|\left(H_N\right)_{ij}\right|^2\right] > c/N \tag{3.34}$$

for some $c > 0$ and

$$\max_{i,j} \mathrm{E}\left[\left|\left(H_N\right)_{ij}\right|^{2+\kappa}\right] \leq c/N^{(2+\kappa)/2} \tag{3.35}$$

for some $\kappa > 0, c < \infty$. Also suppose that the densities of the real or imaginary part of $\sqrt{N} H_{kk}$ exist and are integrable when raised to τ for some $\tau > 1$. Then

$$\lim_{\beta \to 0^+} \lim_{N \to \infty} \left\| F_{\Xi_N}\left(x, y\right) - \widehat{F}_{\Xi_N, \beta}\left(x, y\right) \right\| = 0 \tag{3.36}$$

almost surely, where

$$\frac{\partial^2 \widehat{F}_{\Xi_N,\beta}(t,s)}{\partial x \partial y} = \begin{cases} -\frac{1}{4\pi}\int_\beta^\infty \left(\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial s^2}\right) m_N(u,t,s)du & (t,s)\notin G \\ 0 & (t,s)\in G \end{cases} \tag{3.37}$$

(with the region G defined below) and

$$\begin{aligned} m_N\left(u,t,s\right) = & \tfrac{1}{N}\operatorname{tr}\Big[\big(C_1\left(u,s,t\right)+\ldots \\ & \left(B_N-(t+is)I_N\right)C_2\left(u,s,t\right)^{-1}\left(B_N-(t+is)I_N\right)^*\big)^{-1}\Big] \end{aligned} \tag{3.38}$$

for $u > 0$. The matrices $C_1\left(u,s,t\right)$ and $C_2\left(u,s,t\right)$ are diagonal matrices with entries that satisfy the system of equations

$$\begin{aligned} \left(C_1\right)_{kk}\left(u,s,t\right) = u + \sum_{j=1}^{j=N} \sigma_{N,kj}^2 & \Big[\big(C_2\left(u,s,t\right)+\ldots \\ & \left(B_N-(t+si)I_N\right)^* C_1\left(u,s,t\right)^{-1}\left(B_N-(t+si)I_N\right)\big)^{-1}\Big]_{jj} \end{aligned} \tag{3.39}$$

$$\begin{aligned} \left(C_2\right)_{\ell\ell}\left(u,s,t\right) = 1 + \sum_{j=1}^{j=N} \sigma_{N,j\ell}^2 & \Big[\big(C_1\left(u,s,t\right)+\ldots \\ & \left(B_N-(t+si)I_N\right)C_2\left(u,s,t\right)^{-1}\left(B_N-(t+si)I_N\right)^*\big)^{-1}\Big]_{jj} \end{aligned} \tag{3.40}$$

for $k,\ell = 1,\ldots N$. There exists a unique solution to this system of equations among real positive analytic functions in $u > 0$. The region G is given by

$$G = \left\{ (t,s) \middle| \limsup_{\beta\to 0^+} \limsup_{N\to\infty} \left| \frac{\partial}{\partial\beta} m_N\left(\beta,t,s\right) \right| < \infty \right\}. \tag{3.41}$$

For every point $z = x+yi$ at which the value $\widehat{f}_{\Xi_N}\left(x,y\right)$ is required, the solution to (3.39)-(3.40) must be found at (u,x,y) for a range of u values such that the density can be approximately computed by the numerical integration in (3.37). These u values should range from a small value of $\beta > 0$ (10^{-6} used for simulations) to a large upper limit (10^2 used for simulations). Because the function $m_n\left(u,s,t\right)$ changes quickly for small β within the complement of the region G, logarithmically spaced integration points over the integration interval are recommended. For any random matrix model satisfying the conditions, the solution $C_1(u,s,t), C_2(u,s,t)$ can be found through brute force by performing an iterative fixed point search for the unique solution guaranteed to exist by Theorem 3.3 from an initial candidate values for the entries $(C_1)_{kk}(u,s,t), (C_2)_{kk}(u,s,t)$ for $k = 1,\ldots,N$. However, this involves repetitively multiplying and inverting $N \times N$ matrices for large N, computationally costly steps.

The computational cost can be considerably reduced if the symmetry group of the random matrix model with respect to equal row and column permutations acts transitively on $\{1,\ldots,N\}$ and if the mean matrix B_N is normal. Note that such transitivity does not guarantee normal matrices [48], so normality must be verified for each model with transpose-asymmetric distribution. By transitive symmetry of the distribution with respect to row and column permutations, the diagonal solution matrices must be scalar matrices $C_1(u,s,t) = c_1(u,s,t) I_N$, $C_2(u,s,t) = c_2(u,s,t) I_N$. Furthermore, note that $\sum_{k=1}^{k=N} \sigma_{N,ki}^2 = \sum_{k=1}^{k=N} \sigma_{N,kj}^2$ for all $1 \le i,j \le N$. By summing each expression of (3.39) with respect to k and each expression of (3.40) with respect to ℓ the variance terms no longer depend on j and can be factored from the sum with respect to j. Recognizing the resulting traces and writing them as the eigenvalue sums, this results in the following simplified system of equations where $z = t + si$.

$$\begin{aligned} c_1(u,s,t) = u + \left(\frac{1}{N}\sum_{k=1}^{k=N}\sigma_{N,kj}^2\right)\sum_{r=1}^{r=N}\big(c_2(u,s,t) + \ldots \\ \ldots\ 1/c_1(u,s,t)\,\lambda_r\left[(B_n - zI_N)^*(B_N - zI_N)\right]\big)^{-1} \end{aligned} \tag{3.42}$$

$$\begin{aligned} c_2(u,s,t) = 1 + \left(\frac{1}{N}\sum_{\ell=1}^{\ell=N}\sigma_{N,\ell j}^2\right)\sum_{r=1}^{r=N}\big(c_1(u,s,t) + \ldots \\ \ldots\ 1/c_2(u,s,t)\,\lambda_r\left[(B_n - zI_N)(B_N - zI_N)^*\right]\big)^{-1} \end{aligned} \tag{3.43}$$

Thus, the original system of $2N$ equations (3.39)-(3.40) reduces to the system of 2 equations (3.42)-(3.43) under the transitive symmetry group condition. However, because the eigenvalues of $(B_N - zI)(B_N - zI)^*$ are required for every $z = t + si$ of interest, the problem can still be computationally costly due to the large number of eigenvalue computations required. For random matrix distributions in which the mean matrix B_N is not a normal matrix, these eigenvalues must be recomputed for each z. For random matrix distributions in which the mean matrix B_N is a normal matrix, the eigenvalues can be quickly computed from the eigenvalues of B_N as follows [49].

$$\lambda_r\left(\left(B_N - zI\right)\left(B_N - zI\right)^*\right) = \left|\lambda_r\left(B_N\right) - z\right|^2. \tag{3.44}$$

Therefore, it is only necessary to compute the eigenvalues of B_N once and then determine the eigenvalues of $\left(B_N - zI\right)\left(B_N - zI\right)^*$ through simple arithmetic. Note that this automatically includes all transpose-symmetric distributions and need only be verified for transpose-asymmetric distributions. To simplify (3.42)-(3.43) further, eigenvalues can be grouped according to algebraic multiplicity such that the computation only depends on the number of distinct eigenvalues.

For use in the filter design problems for undirected networks in Section 4.3, the empirical spectral distribution of the row-normalized adjacency matrix must be approximated. In order to apply Girko's method, a random matrix with independent entries must be defined such that the spectral density will be a good approximation. For this purpose, initial work [22–28] used the scaled adjacency matrix $\Xi'_N = \frac{1}{\gamma_1}\mathcal{A}\left(\mathcal{G}_N\right)$ where $\gamma_1 = \lambda_{\max}\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}\right)\right]\right)$, which is also the expected row sum if all rows have equal expected sum like in node-transitive distributions. However, this introduces a bias in the expected largest eigenvalue (though much less prominent than in the undirected case) if $\mathrm{E}\left[\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right] > \lambda_{\max}\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right)$ (i.e., $\mathrm{E}\left[\rho\left(\mathcal{A}\left(\mathcal{G}_N\right)/\gamma_1\right)\right] = \mathrm{E}\left[\rho\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right]/\rho\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right) \geq 1$), which should be exactly $\lambda_{\max} = 1$ for the row-normalized Laplacian. This must be true for node-transitive distributions since $\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right) \geq d_{\text{avg}}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)$ where $d_{\text{avg}}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)$ is the average degree [47], so it follows that $\mathrm{E}\left[\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right] \geq \mathrm{E}\left[d_{\text{avg}}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right] = \lambda_{\max}\left(\mathrm{E}\left[\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right]\right)$. (Note that previous results are not invalidated as for large matrices this bias is observed to be asymptotically eliminated. However, it is problematic for large but finite network size.) To partially correct this bias without affecting the variance, define

$$\Xi_N = \frac{1}{\gamma_1}\left(\mathcal{A}\left(\mathcal{G}_N\right) - \mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right) + \frac{1}{\gamma_2}\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right] \tag{3.45}$$

where $\gamma_1 = \lambda_{\max}\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right)$ and $\gamma_2 = \mathrm{E}\left[\lambda_{\max}\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right]$. That this reduces the bias is unproven, but the results match significantly better in practice.

Before applying this method to example percolation network models, the objection that (3.34) would prevent analysis of random matrices arising from structured networks, where some entries must be dismissed. (Indeed, with the scaling used by Ξ_N, the variance of individual entries falls as approximately $1/\left(N^2\left(\Theta_N\right)_{ij}\right)$, making this problematic regardless of sparsity.) However, Girko also provides an alternate form of this theorem (see Corollary 25.1 on page 379 of [9]) that does not suppose the condition (3.34) but does suppose doubly stochastic variance (which is suitable for node-transitive distributions). Furthermore, the proof is closely related to the theorem for Girko's K7 equation, which does not assume the condition (3.34). Instead, random matrices analyzed through the alternate form of Girko's K25 equation and through Girko's K7 equation must satisfy a Lindberg-like condition. However, statement of the problem according to the alternate version of Girko's K25 equation is significantly more cumbersome than the main formulation of Girko's K25 equation presented in [9]. Therefore, this thesis references the main version of Girko's K25 equation for these problems while dismissing the condition (3.34).

Example 3.3: Directed SBM with Symmetric, Transitive Population Structure

Consider a directed stochastic block model with M node populations each of size S with $N = MS$ nodes in total. Suppose that the populations are arranged in a cycle and that two nodes from the same population or from adjacent populations can potentially connect, with probability that does not depend on cycle direction. Suppose that random intra-population links occur independently with probability $\theta_{N,0}$. Also suppose that random inter-population links occur independently with probability $\theta_{N,1}$ between adjacent populations.

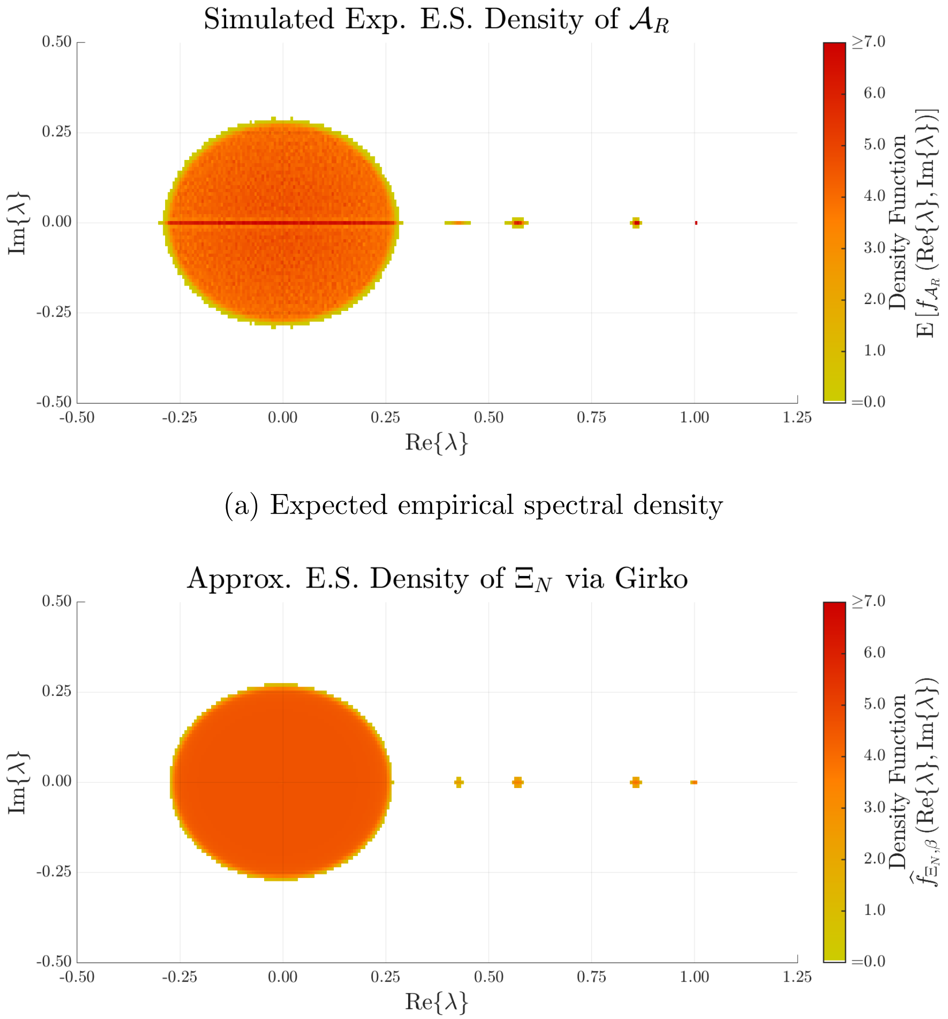
Clearly, this example random network model is node-transitive, and the associated random adjacency matrix distribution is transpose-symmetric. This implies that its mean adjacency matrix is symmetric and, thus, normal. The link probability between nodes (i,j) are described by the entries of the following matrix where $\mathcal{C}_n$ is the adjacency matrix of a length n directed cycle and $1_{n\times n}$ is the $n\times n$ matrix of ones.

$$\begin{aligned}\Theta =&\theta_{N,0}\left(I_M \qquad\qquad\right)\otimes\left(1_{S\times S}-I_S\right)\\ &+\theta_{N,1}\left(\mathcal{C}_M+\mathcal{C}_M^{\top}\right)\otimes\left(1_{S\times S}\qquad\quad\right)\end{aligned}\tag{3.46}$$

Therefore, $E\left[A\left(\mathcal{G}\right)\right]=\Theta$ and the eigenvalues of the mean adjacency matrix are easily computed from the eigenvalues of directed cycles and Kronecker product relationships. Furthermore, the mean matrix is normal. The variance of each entry is described by $\Theta\circ(1_{N\times N}-\Theta)$.

The approximate empirical spectral distribution for Ξ_N can be computed through the reduced equations (3.42)-(3.42) for Girko's K25 method based on its mean eigenvalues and total row variance, which can be found though appropriate transformation of the adjacency matrix mean eigenvalues and total row variance. Because the approximation will be numerically computed at a large fixed value of N and fixed probability parameters, the conditions (3.32)-(3.35) are not verifiable without making assumptions about how the model would change with increasing N. For a percolation model, such as this example, where each probability paramter scales asymptotically with N in the same way (i.e., maintain a ratio) and where each parameter governs a number of entries in each row proportional to N, the conditions are satisfied for Ξ_N if $N\theta_N\to\infty$. Then the first condition (3.1) is automatically satisfied. The second condition (3.2) is satisfied if each parameter $N\theta_N\not\to 0$. The Lindberg-like condition of the modified formulation is satisfied if additionally $N\theta_N\to\infty$.

Assume that the network model is a particular point in a larger family of network models indexed by N such that the conditions are satisfied by this family of models (except condition (3.34) as previously noted). The resulting density will be used to approximate the empirical spectral distribution of the row-normalized adjacency matrix. Simulation results comparing the expected spectral density and approximate spectral density (of a transformation of the row-normalized adjacency matrix) appear in Figure 3.3. Filter design results using the approximate empirical spectral distribution for this model can be found in Figures 4.7-4.10 of Section 4.3 in Chapter 4 for this model with specific numerical parameters.



(a) Expected empirical spectral density

(b) Approximate emprirical spectral density

Figure 3.3: Expected empirical spectral distribution $\mathrm{E}\left[f_{\mathcal{A}_R(\mathcal{G})}\right]$ of the normalized adjacency matrix (simulated over 1000 Monte-Carlo trials) compared agains the deterministic approximation $\widehat{f}_{\Xi_N}$ computed using Girko's K25 method ($\beta = 10^{-6}$, $u_{\max} = 10^2$, 200 logarithmically spaced integration points) for the model in Example 3.4 with $M = 6$, $S = 200$, $\theta_0 = 0.05$, $\theta_1 = 0.01$

Example 3.4: Directed SBM with Asymmetric, Transitive Population Structure

Consider a directed stochastic block model with M node populations each of size S with $N = MS$ nodes in total. Suppose that the populations are arranged in a cycle and that two nodes from the same population or from adjacent populations can potentially connect, with probability permitted to depend on cycle direction. Suppose that random intra-population links occur independently with probability $\theta_{N,0}$. Also suppose that random inter-population links occur independently with probability $\theta_{N,1}$ along one population graph cycle direction and with $\theta_{N,2}$ along the other population graph cycle direction.

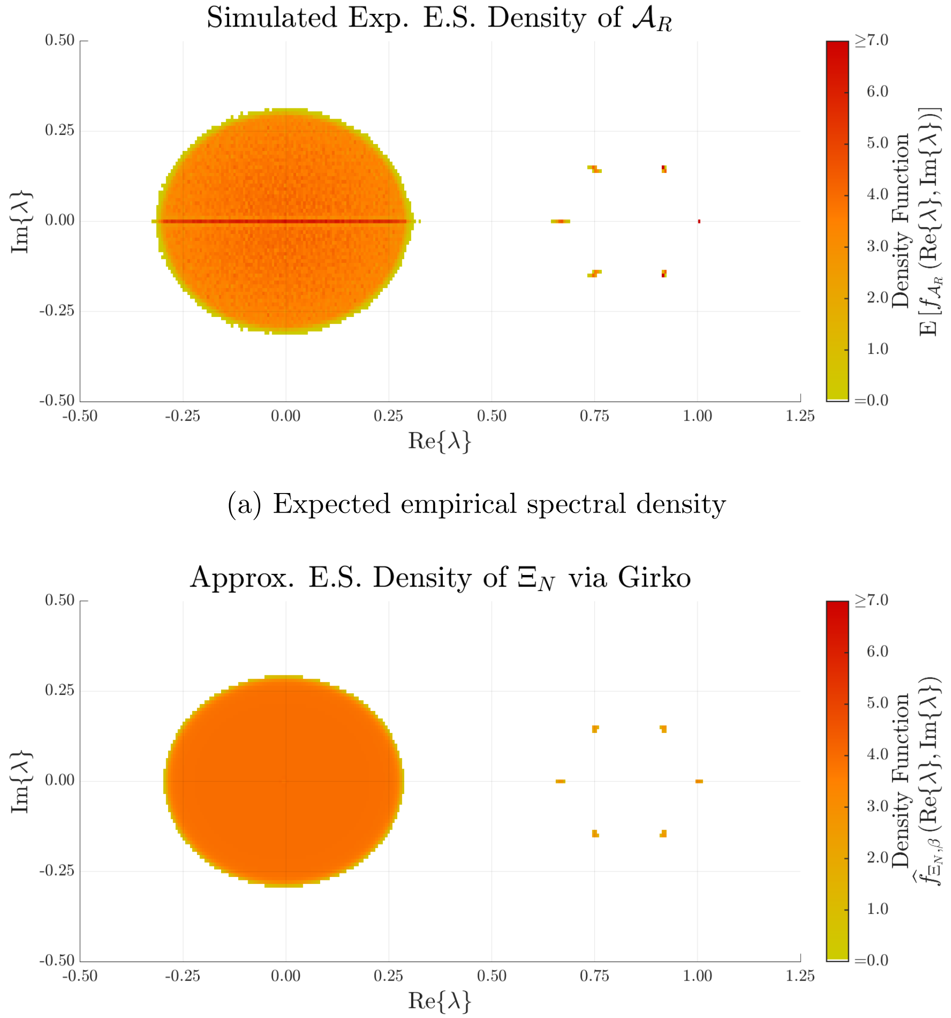
Clearly, this example random network model is node-transitive, but the associated random adjacency matrix distribution is not transpose-symmetric. The link probability between nodes (i, j) are described by the entries of the following matrix where $\mathcal{C}_n$ is the adjacency matrix of a length n directed cycle and $1_{n\times n}$ is the $n \times n$ matrix of ones.

$$\begin{aligned} \Theta_N =& \theta_{N,0} \left(I_M\right) \otimes \left(1_{S\times S} - I_S\right) \\ &+ \theta_{N,1} \left(\mathcal{C}_M\right) \otimes \left(1_{S\times S} \qquad\quad\right) \\ &+ \theta_{N,2} \left(\mathcal{C}_M^\top\right) \otimes \left(1_{S\times S} \qquad\quad\right) \end{aligned} \tag{3.47}$$

Therefore, $E\left[A\left(\mathcal{G}_N\right)\right] = \Theta_N$ and the eigenvalues of the mean adjacency matrix are easily computed from the eigenvalues of directed cycles and Kronecker product relationships. Furthermore, the mean matrix is circulant and, thus, normal. The variance of each entry is described by $\Theta_N \circ \left(1_{N\times N} - \Theta_N\right)$.

The approximate empirical spectral distribution for Ξ_N can be computed through the reduced equations (3.42)-(3.42) for Girko's K25 method based on its mean eigenvalues and total row variance, which can be found though appropriate transformation of the adjacency matrix mean eigenvalues and total row variance. Because the approximation will be numerically computed at a large fixed value of N and fixed probability parameters, the conditions (3.32)-(3.35) are not verifiable without making assumptions about how the model would change with increasing N. For a percolation model, such as this example, where each probability paramter scales asymptotically with N in the same way (i.e., maintain a ratio) and where each parameter governs a number of entries in each row proportional to N, the conditions are satisfied for Ξ_N if $N\theta_N \to \infty$. Then the first condition (3.1) is automatically satisfied. The second condition (3.2) is satisfied if each parameter $N\theta_N \not\to 0$. The Lindberg-like condition of the modified formulation is satisfied if additionally $N\theta_N \to \infty$.

Assume that the network model is a particular point in a larger family of network models indexed by N such that the conditions are satisfied by this family of models (except condition (3.34) as previously noted). The resulting density will be used to approximate the empirical spectral distribution of the row-normalized adjacency matrix. Simulation results comparing the expected spectral density and approximate spectral density (of a transformation of the row-normalized adjacency matrix) appear in Figure 3.4. Filter design results using the approximate empirical spectral distribution for this model can be found in Figures 4.11-4.14 of Section 4.3 in Chapter 4 for this model with specific numerical parameters.



(a) Expected empirical spectral density

(b) Approximate emprirical spectral density

Figure 3.4: Expected empirical spectral distribution $\mathrm{E}\left[f_{\mathcal{A}_R(\mathcal{G})}\right]$ of the normalized adjacency matrix (simulated over 1000 Monte-Carlo trials) compared agains the deterministic approximation $\widehat{f}_{\Xi_N}$ computed using Girko's K25 method ($\beta = 10^{-6}$, $u_{\max} = 10^2$, 200 logarithmically spaced integration points) for the model in Example 3.4 with $M = 6$, $S = 200$, $\theta_0 = 0.05$, $\theta_1 = 0.01$, $\theta_2 = 0.00$

3.5 Summary

In summary, this chapter introduced random matrix theory methods suitable for analysis of the spectral asymptotics of large-scale random network row-normalized adjacency matrices. Two methods for symmetric random matrices were introduced, namely Girko's K1 equation, which describes the spectral asymptotics of matrices with independent entries (except as related by symmetry), and Girko's K27 equation, which describes the spectral asymptotics of matrices with independent block submatrices (except as related by symmetry). One method for potentially non-Hermitian random matrices was introduced, namely Girko's K25 equation, which describes the spectral asymptotics of matrices with independent entries. For each method, conditions on the network distribution that enable computational simplification were discussed, and the reduced systems of equations to provide the approximate density were derived. For Girko's K1 equation, node-transitivity yields significant computational reduction. For Girko's K27 equation, node-transitivity with additional simultaneous diagonalizability properties and known eigenvalue relationships yields significant computational reduction. For Girko's K25 equation, node-transitivity with a normal mean matrix yields significant computational reduction. The random matrix Ξ_N with independent entries (except as related by symmetry for the undirected network case) used to approximate the row normalized adjacency matrix was posed as a transformation of the adjacency matrix. For each case, application to one or more example random network distribution was demonstrated. Approximate spectral densities obtained using the methods in this chapter inform filter design problems for constant random network models in Chapter 4 as well as filter design problems for switching random network models in Section 5.3 of Chapter 5.

Chapter Four

Consensus Filter Design: Constant Networks

4.1 Introduction

Distributed average consensus refers to the network agreement task in which network nodes each begin with state set to an initial data element and must compute the average of all network data through local communications and iterative linear updates to the local state [10]. Under the correct conditions, the resulting dynamic system has each state converge to the (potentially weighted) mean of the initial state values at a rate governed by the spectral radius of the difference between the iteration matrix and the averaging matrix, which describes the convergence of the worst case error vector [11]. Section 2.4 describes distributed average consensus in greater detail. For constant networks, the convergence rate can be accelerated by periodically updating the state value at each node every d iterations by applying a degree d filter to previous state values at that node. This filter design problem can be understood in terms of graph signal processing, where the response magnitude of the filter to the eigenvalues of the consensus iteration matrix $W(\mathcal{G})$ (other than $\lambda(W(\mathcal{G})) = 1$) should be minimized. Section 2.6 describes consensus acceleration filters and related literature more fully. For random matrices, uncertainty in the eigenvalues complicates the design process. One existing method [16] attempts to handle the random network case through the mean iteration matrix, but this can produce poor or even diverging results when the true eigenvalues differ significantly from the mean eigenvalues. A more complete understanding of the eigenvalue spread should therefore inform the filter design process, which this thesis obtains for large-scale random matrices from the spectral density approximation methods in Chapter 3.

This chapter presents the proposed optimization methods for periodic consensus acceleration filter design and other related graph filtering problems on large-scale constant (not time-varying) random networks along with supporting simulation results. Section 4.2 examines undirected random network models, posing a linear program for consensus acceleration filter design based on the real empirical spectral density approximations obtained using the methods in Section 3.2 and Section 3.3 for symmetric matrices. Simulation results for the large-scale constant, undirected random network case are also provided. Section 4.3 examines directed random network models, posing a quadratically constrained linear program for consensus acceleration filter design based on the complex empirical spectral density approximations obtained using the methods in Section 3.4 for non-Hermitian matrices. Simulation results for the large-scale constant, directed random network case are also provided. Section 4.4 presents filter design methods for worst case and expected total variation minimization for both the undirected case and directed case. These optimization problems are variants of those presented in Section 4.2 and Section 4.3 with weighted filter response. Section 4.5 provides simulations for these weighted problems. Finally, Section 4.6 summarizes the chapter in conclusion.

4.2 Constant, Undirected Random Networks

For large-scale constant (not time-varying) undirected random networks, this section introduces an optimization problem for consensus acceleration filter design based on spectral asymptotics. Let $\mathcal{G}_N$ be the random graph with N nodes describing the random network, where $\mathcal{A}\left(\mathcal{G}_N\right)$ is the unnormalized adjacency matrix, $\mathcal{D}\left(\mathcal{G}_N\right)$ is the diagonal matrix of node degrees, and $\mathcal{L}\left(\mathcal{G}_N\right)=\mathcal{D}\left(\mathcal{G}_N\right)-\mathcal{A}\left(\mathcal{G}_N\right)$ is the unnormalized Laplacian matrix. Additionally, let $\mathcal{A}_R\left(\mathcal{G}_N\right)=\mathcal{D}\left(\mathcal{G}_N\right)^{-1}\mathcal{A}\left(\mathcal{G}_N\right)$ be the row-normalized adjacency matrix and $\mathcal{L}_R\left(\mathcal{G}_N\right)=I-\mathcal{D}\left(\mathcal{G}_N\right)^{-1}\mathcal{A}\left(\mathcal{G}_N\right)$ be the row-normalized Laplacian matrix. One common choice for consensus iteration matrix is the Laplacian weights $W\left(\mathcal{G}\right)=I-\alpha\mathcal{L}\left(\mathcal{G}\right)$ which produces an unweighted average consensus for suitable values of α. However, for this filter design problem, the consensus iteration matrix will be derived from the row-normalized Laplacian matrix.

$$W_N = W\left(\mathcal{G}_N\right) = I - \alpha \mathcal{L}_R\left(\mathcal{G}_N\right) \tag{4.1}$$

While this selection typically results in a weighted average, advantages include the ability to approximate the spectral density using Girko's methods (as discussed in Section 3.2 and Section 3.3 of Chapter 3) and good spectral density localization properties for many large-scale percolation matrices. Furthermore, the left eigenvector of $W\left(\mathcal{G}_N\right)$ dictating the weights is $\boldsymbol{\ell}_k = \mathbf{d}_k\left(\mathcal{G}_N\right)$ where $\mathbf{d}_k\left(\mathcal{G}_N\right)$ is the degree of the kth node. Note that if each node knows the average degree

$$d_{\text{avg}} = \frac{\mathbf{d}\left(\mathcal{G}_N\right)^\top \mathbf{1}}{\mathbf{1}^\top \mathbf{1}} \tag{4.2}$$

of the network, the weighted average can be corrected through pre-multiplication of the initial data $\mathbf{x}_0$ by $d_{\text{avg}}\left(\mathcal{G}_N\right) \mathcal{D}\left(\mathcal{G}_N\right)^{-1}$ producing an unweighted average. That, is

$$J_{\mathbf{1}} = J_{\boldsymbol{\ell}}\left(d_{\text{avg}}(\mathcal{G}_N)\, D\left(\mathcal{G}_N\right)^{-1}\right). \tag{4.3}$$

In order to analyze the spectral asymptotics of W_N, consider the random matrix

$$\Xi_N = \frac{1}{\gamma_1}\left(\mathcal{A}\left(\mathcal{G}_N\right) - \mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right) + \frac{1}{\gamma_2}\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right] \tag{4.4}$$

where $\gamma_1 = \rho\left(\mathrm{E}\left[\mathcal{A}\left(\mathcal{G}_N\right)\right]\right)$ and $\gamma_2 = \mathrm{E}\left[\rho\left(\mathcal{A}\left(\mathcal{G}_N\right)\right)\right]$. A deterministic approximation for the empirical spectral distribution of Ξ_N can be used to approximate the empirical spectral distribution of $\mathcal{A}_R\left(\mathcal{G}_N\right)$. Note that

$$\mathcal{A}_R\left(\mathcal{G}_N\right) = \frac{1}{\alpha}(W\left(\mathcal{G}_N\right) - I) + I \tag{4.5}$$

relates $W\left(\mathcal{G}_N\right)$ to $\mathcal{A}_R\left(\mathcal{G}_N\right)$. Therefore, an approximate empirical spectral density $\widehat{f}_{W_N}$ for W_N relates to an approximate empirical spectral density $\widehat{f}_{\Xi_N}$ for Ξ_N via the following expression.

$$\widehat{f}_{W_N}(x) = \frac{1}{\alpha}\widehat{f}_{\Xi_N}\left(\frac{x-1}{\alpha} + 1\right) \tag{4.6}$$

The approximate density $\widehat{f}_{\Xi_N}$ may be obtained through methods such as from Section 3.2 and Section 3.3 for suitable random network distributions. To optimize the convergence rate of the filtered consensus process, the following optimization minimizes the filter response magnitude over regions where the approximate density is nonzero.

$$\begin{aligned} &\min_{p \in P_d} \max_{\lambda \in \Lambda_{\kappa,\tau}} |p(\lambda)| \\ &\text{s.t.} \qquad p(1) = 1 \\ \Lambda_{\kappa,\tau} = &\left\{ \lambda \in \mathbb{R} \middle| 1 - \lambda > \kappa, \widehat{f}_{W_N}(\lambda) > \tau \right\} \end{aligned} \tag{4.7}$$

where P_d is the space of polynomials of degree at most d. The set $\Lambda_{\kappa,\tau} \subseteq \mathbb{R}$ defines the filtering regions, where κ and τ are small constants (e.g., $\tau = 10^{-3}$, $\kappa = 10^{-2}$ for the first simulation in this section). The constant κ serves as a transition distance designed to exclude the region containing $\lambda = 1$ where the equality constrain must be satisfied, since the approximate density generally does not perfectly localize the density at $\lambda = 1$. The constant τ serves as a threshold for detecting non-zero density content, as numerically evaluated approximate densities arising from Girko's equations are small nonzero in the excluded regions.

While it would be possible to apply methods like the Remez algorithm to optimize over this set, optimizing the response at a suitable discretization can provide a good approximation. A set of sample points $\Lambda_S \subseteq \Lambda_{\kappa,\tau}$ can be selected to capture the structure of $\Lambda_{\kappa,\tau}$, for instance through intersection with a sufficiently fine grid of points. Introducing a variable η to bound the response magnitude at the sample points (i.e., $|p(\lambda)| < \eta$), the problem in (4.7) can be approximately solved through the following optimization.

$$\begin{aligned} \min_{\substack{\eta \in \mathbb{R}_+ \\ p \in P_d}} \quad & \eta \\ \text{s.t.} \quad & p(1) = 1 \\ & p(\lambda) < \eta \\ & -p(\lambda) < \eta \\ & \text{for all } \lambda \in \Lambda_S \end{aligned} \tag{4.8}$$

Collecting the coefficients $\{a_k\}_{k=0}^{k=d}$ of p into the column vector $\mathbf{a} = [a_0, \ldots, a_d]^\top$, this can be rewritten as the linearly constrained linear program (LP)

$$\begin{aligned} \min_{\substack{\eta \in \mathbb{R}_+ \\ \mathbf{a} \in \mathbb{R}^{d+1}}} \quad & \eta \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \\ & \mathbf{v}(\lambda)^\top \mathbf{a} < \eta \\ & -\mathbf{v}(\lambda)^\top \mathbf{a} < \eta \\ & \text{for all } \lambda \in \Lambda_S \end{aligned} \tag{4.9}$$

where $\mathbf{v}(\lambda)$ is the $(d+1) \times 1$ Vandermode column vector

$$\mathbf{v}(\lambda) = \left[\lambda^0, \ldots, \lambda^d\right]^\top \tag{4.10}$$

such that $p(\lambda) = \mathbf{v}(\lambda)^\top \mathbf{a}$.

To evaluate the presented filter design method, this section provides simulation results for two undirected random network models, one with independent links and one with independent block structure. For the first simulation results that appear in Figures 4.1-4.3, consider a stochastic block model that follows the structure described in Example 3.1 of Section 3.2 in Chapter 3. The specific parameters used for this simulation are $M_1 = 4$, $M_2 = 5$, $S = 100$, $\theta_0 = 0.05$, $\theta_1 = 0.03$, $\theta_2 = 0.03$, $\theta_3 = 0.01$. This random network model can be analyzed through Girko's K1 method in Theorem 3.1 as described by Example 3.1.

For the second group of simulation results that appear in Figures 4.4-4.6, consider a modified stochastic block model with dependencies within node blocks that follows the structure described in Example 3.2 of Section 3.3 in Chapter 3. The specific parameters used for this simulation are $M = 10$, $S_1 = 20$, $S_2 = 5$, $\{\phi_{0,k}\} = \{0.50, 0.50\}$, $\{\theta_{0,k}\} = \{0.50, 1.00\}$, $\{\phi_{1,k}\} = \{0.50, 0.50\}$, $\{\theta_{1,k}\} = \{0.05, 0.10\}$, $\{\phi_{2,k}\} = \{0.50, 0.50\}$, $\{\theta_{2,k}\} = \{0.01, 0.02\}$. This random network model can be analyzed through Girko's K27 method in Theorem 3.2 as described by Example 3.2.

For each of these random network models, a deterministic approximation $\widehat{f}_{\Xi_N}$ to the empirical spectral distribution of Ξ_N was computed (first simulation: $\epsilon = 10^{-6}$, second simulation: $\epsilon = 10^{-5}$) to approximate the empirical spectral distribution of $\mathcal{A}_R$. This was transformed to form an approximation $\widehat{f}_{W_N}$ to the empirical spectral distribution f_{W_N} of the consensus iteration matrix through (4.6) (where $\alpha = 1$). Figure 4.1 and Figure 4.4 show both the approximate spectral density $\widehat{f}_{W_N}$ (black curve) and expected spectral density (blue shaded) as simulated through 1000 Monte-Carlo trials with 500 histogram bins. As expected, these plots closely correspond, suggesting the approximation should be useful.

Consensus acceleration filters of degrees $d = 1, \dots, 10$ were designed according to the method proposed in this section (first simulation: $\kappa = 10^{-2}$, $\tau = 10^{-3}$, second simulation: $\kappa = 10^{-1}$, $\tau = 10^{-3}$) as the optimization problem (4.9). For purpose of comparison, filters of degrees $d = 1, \dots K-1$ were designed using only the mean iteration matrix eigenvalues (equivalent to the mean matrix semi-definite program (SDP) method proposed in [16]), where K is the number of distinct mean iteration matrix eigenvalues ($K = 10$ for the first model, $K = 4$ for the second model). Note that optimal mean matrix SDP method filters cannot be uniquely defined for $d > K - 1$. Hence, results are only shown for $d \leq K - 1$ for the mean matrix SDP filters.

For the two simulations, Figure 4.2 and Figure 4.5 display the expected convergence rates per iteration on a logarithmic scale to compare the performance of the proposed filter (blue square curve) and the mean matrix SDP filter (purple triangle curve) along with the trivial filter (black circle curve, no filter applied) and the optimal filter designed with the exact matrix known for each matrix drawn from the distributions (green diamond curve). Note that a smaller value of this plot indicates faster convergence and that the proposed filters perform nearly as well as the optimal filters, which improve significantly over no filtering. Furthermore, the convergence rates achieved using only the mean matrix eigenvalues do not compare well for this model (and can fail to produce convergence in other models), indicating a need to properly model the spread of the eigenvalues under these conditions. For each filter type and degree, the empirical distribution of convergence rate results over the Monte-Carlo

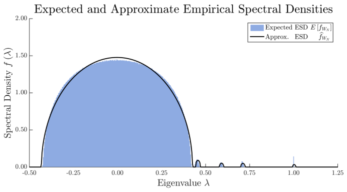


Figure 4.1: Expected density $\mathrm{E}[f_{W_N}]$ and approximate density $\widehat{f}_{W_N}$ for first simulation (SBM described in Sec. 3.2)

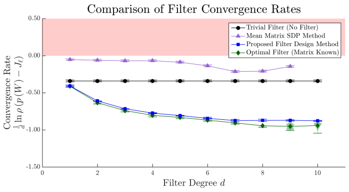


Figure 4.2: Expected convergence rates for each filter type with result distribution shown vertically (first simulation)

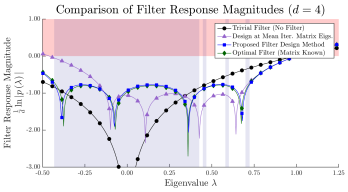


Figure 4.3: Filter response magnitudes for degree $d = 4$ filters of each type with $\Lambda_{\kappa,\tau}$ shaded blue (first simulation)

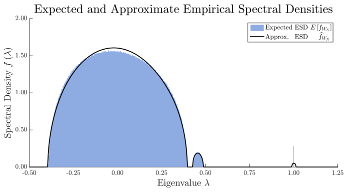


Figure 4.4: Expected density $\mathrm{E}[f_{W_N}]$ and approximate density $\widehat{f}_{W_N}$ for second simulation (SBM described in Sec. 3.3)

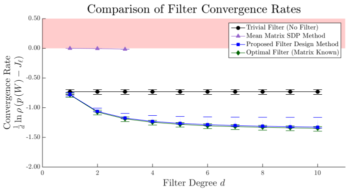


Figure 4.5: Expected convergence rates for each filter type with result distribution shown vertically (second simulation)

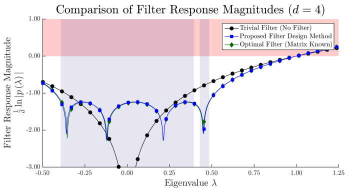


Figure 4.6: Filter response magnitudes for degree $d = 4$ filters of each type with $\Lambda_{\kappa,\tau}$ shaded blue (second simulation)

sample is plotted as a vertical histogram with the extreme results marked by bars. The results have small spread, so these are somewhat difficult to see. Finally, filter responses for each filter type are visualized in Figure 4.3 at degree $d = 4$ for the first simulation and in Figure 4.6 at degree $d = 4$ for the second simulation, with the blue shaded region showing the filtering region $\Lambda_{\kappa,\tau}$. The figures show that the response for the proposed filter is nearly identical to the response for the optimal filter for an example matrix drawn from the distribution.

The filter design method proposed in this section for undirected random networks is subject to some practical limitations that should be noted.

- First, the method may only be applied to random network models for which an asymptotic spectral approximation can be derived for the random network description, for instance by using Girko's methods. Pragmatic solution to Girko's equations relies on the computational reduction justified by node-transitivity. This favors random network distributions with many permutation symmetries such that no two nodes are statistically distinguishable, requiring costly brute force for more general models.

- Second, numerical precision and tolerances of the optimization software limit the degree of filters that can be accurately designed. Further increase in the filter degree may not improve the convergence rate of the filters output from the optimization once the convergence rate has been reduced close to these tolerances and may produce results that cannot be trusted.

- Third, the method relies on the support of the approximate spectral distribution capturing the full set of eigenvalues of the actual consensus iteration matrix. The presence of outlier eigenvalues could lead to reduced worst-case performance or, if the outliers are far from the filtering regions, even possible loss of convergence. However, this problem is much more pronounced when only the eigenvalues of the mean consensus iteration matrix are known. As evident from the small spread of convergence rates in the simulations, this is not a severe problem and only occurs with low probability. Furthermore, it would be possible to add constraints to force convergence robustness by limiting the response magnitude over all possible eigenvalues (i.e, for $W = I - \alpha\mathcal{L}_R$ the interval $[1-2\alpha, 1)$ contains all eigenvalues apart from $\lambda = 1$) at the cost of reducing optimality.

4.3 Constant, Directed Random Networks

For large-scale constant (not time-varying) directed random networks, this section introduces an optimization problem for consensus acceleration filter design based on spectral asymptotics. This section adopts the convention that each node in the directed network receives data from its in-neighbors and sends data to its out-neighbors. Let $\mathcal{G}_N$ be the random graph with N nodes describing the random network, where $\mathcal{A}(\mathcal{G}_N)$ is the unnormalized adjacency matrix such that $\mathcal{A}_{ij}(\mathcal{G}_N) = 1$ if node j is an in-neighbor of node i and such that $\mathcal{A}_{ij} = 0$ otherwise. Also let $\mathcal{D}_{\text{in}}(\mathcal{G}_N)$ be the diagonal matrix of node in-degrees and $\mathcal{L}_{\text{in}}(\mathcal{G}_N) = \mathcal{D}_{\text{in}} - \mathcal{A}(\mathcal{G}_N)$ be the in-degree directed Laplacian matrix. Finally, let $\mathcal{A}_R(\mathcal{G}_N) = \mathcal{D}_{\text{in}}(\mathcal{G}_N)^{-1}\mathcal{A}(\mathcal{G}_N)$ be the row-normalized adjacency matrix and $\mathcal{L}_R(\mathcal{G}_N) = I - \mathcal{D}_{\text{in}}(\mathcal{G}_N)^{-1}\mathcal{A}(\mathcal{G}_N)$ be the row-normalized directed Laplacian matrix. For this filter design problem, the consensus iteration matrix will be derived from the row-normalized directed Laplacian matrix via

$$W_N = W(\mathcal{G}_N) = I - \alpha\mathcal{L}_R(\mathcal{G}_N) \tag{4.11}$$

for suitable values of α. As with the undirected case, while this selection typically results in a weighted average, advantages include the ability to approximate the spectral density using Girko's methods (as discussed in Chapter 3) and good spectral density localization properties for many large-scale percolation matrices. Unlike the undirected case, it is not simple to compute and adjust for the left eigenvector inducing the weighted average.

In order to analyze the spectral asymptotics of W_N, consider the random matrix

$$\Xi_N = \frac{1}{\gamma_1}\left(\mathcal{A}(\mathcal{G}_N) - \mathrm{E}\left[\mathcal{A}(\mathcal{G}_N)\right]\right) + \frac{1}{\gamma_2}\mathrm{E}\left[\mathcal{A}(\mathcal{G}_N)\right] \tag{4.12}$$

where $\gamma_1 = \rho\left(\mathrm{E}\left[\mathcal{A}(\mathcal{G}_N)\right]\right)$ and $\gamma_2 = \mathrm{E}\left[\rho\left(\mathcal{A}(\mathcal{G}_N)\right)\right]$. A deterministic approximation for the empirical spectral distribution of Ξ_N can be used to approximate the empirical spectral distribution of $\mathcal{A}_R(\mathcal{G}_N)$. Note that

$$\mathcal{A}_R(\mathcal{G}_N) = \frac{1}{\alpha}(W(\mathcal{G}_N) - I) + I \tag{4.13}$$

relates $W(\mathcal{G}_N)$ to $\mathcal{A}_R(\mathcal{G}_N)$. Therefore, an approximate empirical spectral density $\widehat{f}_{W_N}$ for W_N relates to an approximate empirical spectral density $\widehat{f}_{\Xi_N}$ for Ξ_N via the following expression.

$$\widehat{f}_{W_N,\beta}(x, y) = \frac{1}{\alpha^2}\widehat{f}_{\Xi_N,\beta}\left(\frac{x-1}{\alpha} + 1, \frac{y}{\alpha}\right) \tag{4.14}$$

The approximate density $\widehat{f}_{\Xi_N}$ may be obtained through methods such as from Section 3.4. For suitable random network distributions. To optimize the convergence rate of the filtered consensus process, the following optimization minimizes the filter response magnitude over regions where the approximate density is nonzero.

$$\begin{aligned} &\min_{p\in P_d}\max_{\lambda\in\Lambda_{\kappa,\tau}} |p(\lambda)| \\ &\text{s.t.} \qquad p(\lambda)=1 \\ \Lambda_{\kappa,\tau}=&\left\{\lambda\in\mathbb{C}\middle| |1-\lambda|>\kappa, \widehat{f}_{W_N,\beta}(\operatorname{Re}\{\lambda\},\operatorname{Im}\{\lambda\})>\tau\right\} \end{aligned} \tag{4.15}$$

The set $\Lambda_{\kappa,\tau} \subseteq \mathbb{C}$ defines the filtering regions, where κ and τ are small constants (e.g., $\tau = 10^{-3}$, $\kappa = 10^{-2}$ in the simulations for this section). The constants κ and τ serve the same purposes as in the formulation for undirected networks in Section 4.2, respectively representing a transition distance and a non-zero density numerical threshold.

As before, a set of sample points $\Lambda_S \subseteq \Lambda_{\kappa,\tau}$ can be selected to capture the structure of $\Lambda_{\kappa,\tau}$, for instance through intersection with a sufficiently fine grid of points. Introducing a variable η to bound the squared response magnitude at the sample points (i.e., $|p(\lambda)|^2 < \eta$), the problem in (4.15) can be approximately solved through the following optimization.

$$\begin{aligned} \min_{\substack{\eta\in\mathbb{R}_+\\ p\in P_d}} \qquad & \eta \\ \text{s.t.} \qquad & p(1) = 1 \\ & |p(\lambda)|^2 < \eta \\ & \text{for all } \lambda \in \Lambda_S \end{aligned} \tag{4.16}$$

Collecting the coefficients $\{a_k\}_{k=0}^{k=d}$ of p into the column vector $\mathbf{a} = [a_0, \ldots, a_d]^\top$, this can be rewritten as the quadratically constrained linear program (QCLP)

$$\begin{aligned} \min_{\substack{\eta\in\mathbb{R}_+\\ \mathbf{a}\in\mathbb{R}^{d+1}}} \qquad & \eta \\ \text{s.t.} \qquad & \mathbf{1}^\top\mathbf{a} = 1 \\ & \mathbf{a}^\top Q(\lambda)\,\mathbf{a} < \eta \\ & \text{for all } \lambda \in \Lambda_S \end{aligned} \tag{4.17}$$

where $Q(\lambda)$ is the $(d+1) \times (d+1)$ real, positive semi-definite matrix

$$Q(\lambda) = \frac{1}{2}\left(\mathbf{v}(\lambda)\,\mathbf{v}(\lambda)^* + \mathbf{v}\left(\overline{\lambda}\right)\mathbf{v}\left(\overline{\lambda}\right)^*\right) \tag{4.18}$$

and $\mathbf{v}(\lambda)$ is the $(d+1) \times 1$ Vandermode column vector

$$\mathbf{v}(\lambda) = \left[\lambda^0, \dots, \lambda^d\right]^\top \tag{4.19}$$

such that $|p(\lambda)|^2 = \frac{1}{2}\left(|p(\lambda)|^2 + \left|p\left(\overline{\lambda}\right)\right|^2\right) = \mathbf{a}^\top Q(\lambda)\,\mathbf{a}$. The matrices $Q(\lambda)$ in the constraints are positive semidefinite, making this a convex optimization problem [50].

To evaluate the presented filter design method, this section provides simulation results for two directed random network models, one with transpose-symmetric structure and one with transpose-asymmetric structure but normal mean matrix. For the first simulation results that appear in Figures 4.7-4.10, consider a stochastic block model that follows the structure described in Example 3.3 of Section 3.4 in Chapter 3. The specific parameters used for this simulation are $M = 6$, $S = 200$, $\theta_0 = 0.05$, $\theta_1 = 0.01$. This random network model can be analyzed through Girko's K25 method in Theorem 3.3 as described by Example 3.3.

For the second group of simulation results that appear in Figures 4.11-4.14, consider a stochastic block model with that follows the structure described in Example 3.4 of Section 3.4 in Chapter 3. The specific parameters used for this simulation are $M = 8$, $S = 200$, $\theta_0 = 0.05$, $\theta_1 = 0.03$, $\theta_2 = 0.00$. This random network model can be analyzed through Girko's K25 method in Theorem 3.3 as described by Example 3.4. Because the mean matrix is normal, the transpose-asymmetry of the distribution does not incur additional computational cost.

For each of these random network models, a deterministic approximation $\widehat{f}_{\Xi_N}$ to the empirical spectral distribution of Ξ_N was computed ($\beta = 10^{-6}$, $u_{\max} = 10^2$, 200 logarithmically spaced integration points) to approximate the empirical spectral distribution of $\mathcal{A}_R$. This was transformed to form an approximation $\widehat{f}_{W_N}$ to the empirical spectral distribution f_{W_N} of the consensus iteration matrix through (4.14) (where $\alpha = 1$). Figure 4.7 and Figure 4.11 show the expected spectral densities $\mathrm{E}\left[f_{W_N}\right]$ for each model as simulated through 1000 Monte-Carlo trials

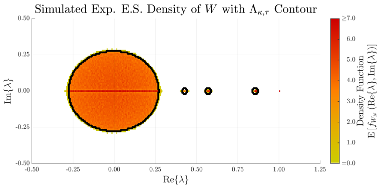


Figure 4.7: Expected density $\mathrm{E}\left[f_{W_N}\right]$ for first simulation (SBM described in Sec. 3.4) with $\Lambda_{\kappa,\tau}$ boundary marked

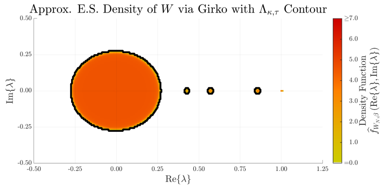


Figure 4.8: Approximate density $\widehat{f}_{W_N,\beta}$ for first simulation (SBM described in Sec. 3.4) with $\Lambda_{\kappa,\tau}$ boundary marked

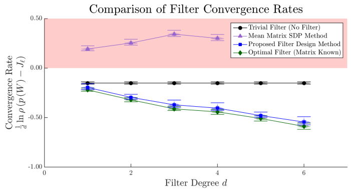
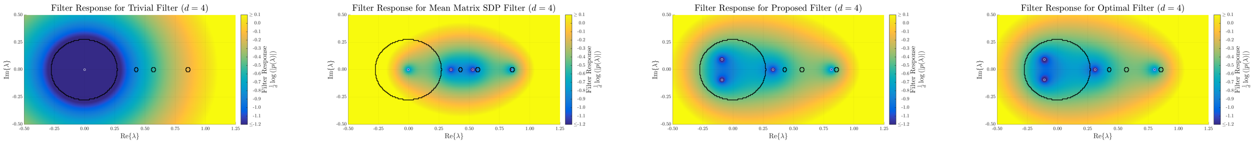


Figure 4.9: Expected convergence rates for each filter type with result distribution shown vertically (first simulation)



(a) Trivial Filter (for $d = 4$)

(b) Mean SDP Filter (for $d = 4$)

(c) Proposed filter (for $d = 4$)

(d) Optimal filter (for $d = 4$)

Figure 4.10: Filter response magnitudes for filters of degree $d = 4$ of each type with boundary of the filtering region $\Lambda_{\kappa,\tau}$ marked by black closed curves and filter zeros marked by white circles located in the dark blue regions (first simulation)

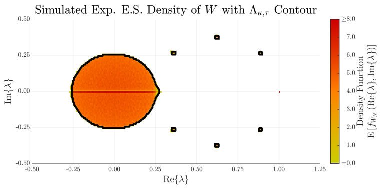


Figure 4.11: Expected density $\mathrm{E}\left[f_{W_N}\right]$ for second simulation (SBM described in Sec. 3.4) with $\Lambda_{\kappa,\tau}$ boundary marked

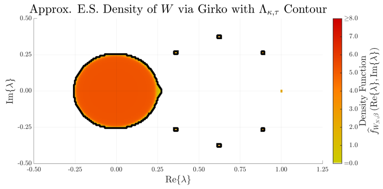


Figure 4.12: Approximate density $\widehat{f}_{W_N,\beta}$ for second simulation (SBM described in Sec. 3.4) with $\Lambda_{\kappa,\tau}$ boundary marked

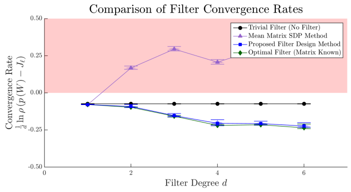
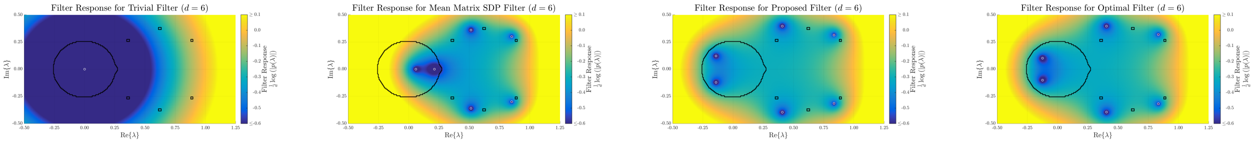


Figure 4.13: Expected convergence rates for each filter type with result distribution shown vertically (second simulation)



(a) Trivial Filter (for $d = 6$)

(b) Mean SDP Filter (for $d = 6$)

(c) Proposed filter (for $d = 6$)

(d) Optimal filter (for $d = 6$)

Figure 4.14: Filter response magnitudes for filters of degree $d = 6$ of each type with boundary of the filtering region $\Lambda_{\kappa,\tau}$ marked by black closed curves and filter zeros marked by white circles located in the dark blue regions (second simulation)

with 300 real histogram bins and 100 imaginary histogram bins. Figure 4.8 and Figure 4.12 show the approximate spectral densities $\widehat{f}_{W_N}$ as computed. For each of these plots, the boundary of the filtering region $\Lambda_{\kappa,\tau}$ is encircled by closed black curves ($\kappa = 10^{-2}$, $\tau = 10^{-3}$). As expected, these plots closely correspond, suggesting the approximation should be useful.

Consensus acceleration filters of degrees $d = 1, \ldots, 10$ were designed according to the method proposed in this section as the optimization problem (4.17). For purpose of comparison, filters of degrees $d = 1, \ldots K - 1$ were designed using only the mean iteration matrix eigenvalues (equivalent to the mean matrix semi-definite program (SDP) method proposed in [16]), where K is the number of distinct mean iteration matrix eigenvalues ($K = 5$ for the first model, $K = 9$ for the second model). Note that optimal mean matrix SDP method filters cannot be uniquely defined for $d>K-1$. Hence, results are only shown for $d\leq K-1$ for the mean matrix SDP filters.

For the two simulations, Figure 4.9 and Figure 4.13 display the expected convergence rates per iteration on a logarithmic scale to compare the performance of the proposed filter (blue square curve) and the mean matrix SDP filter (purple triangle curve) along with the trivial filter (black circle curve, no filter applied) and the optimal filter designed with the exact matrix known for each matrix drawn from the distributions (green diamond curve). Note that a smaller value of this plot indicates faster convergence and that the proposed filters perform nearly as well as the optimal filters, which improve significantly over no filtering. Furthermore, the convergence rates achieved using only the mean matrix eigenvalues do not compare well for this model and can fail to produce convergence, indicating a need to properly model the spread of the eigenvalues under these conditions. For each filter type and degree, the empirical distribution of convergence rate results over the Monte-Carlo sample is plotted as a vertical histogram with the extreme results marked by bars. The results have small spread, so these are somewhat difficult to see. Finally, filter responses for each filter type are visualized in Figure 4.10a-4.10d at degree $d = 4$ for the first simulation and in Figure 4.14a-4.14d at degree $d = 6$ for the second simulation, with the boundary of the filtering region $\Lambda_{\kappa,\tau}$ indicated with closed black curves. The figures show that the response for the proposed filter is nearly identical to the response for the optimal filter for an example matrix drawn from the distribution.

The filter design method proposed in this section for directed random networks is subject to some practical limitations that should be noted. Many of these are shared by the method for undirected networks in Section 4.2, but are more pronounced for the directed networks case as will be described.

- First, the method may only be applied to random network models for which an asymptotic spectral approximation can be derived for the random network description, for instance by using Girko's methods. Pragmatic solution to Girko's equations relies on the computational reduction justified by node-transitivity. This favors random network distributions with many permutation symmetries such that no two nodes are statistically distinguishable, requiring costly brute force for more general models. Note that compared to Girko's K1 system of equations for undirected graphs, the reduced form of Girko's K25 system of equations has twice as many equations.

- Second, numerical precision and tolerances of the optimization software limit the degree of filters that can be accurately designed. Further increase in the filter degree may not improve the convergence rate of the filters output from the optimization once the convergence rate has been reduced close to these tolerances and may produce results that cannot be trusted. Because the quadratic constraints for the directed case are on the magnitude squared, which falls more quickly than the magnitude in the linear constraints for the undirected case, this may become problematic more quickly for the optimization problem proposed in this section.

- Third, the method relies on the support of the approximate spectral distribution capturing the full set of eigenvalues of the actual consensus iteration matrix. The presence of outlier eigenvalues could lead to reduced worst-case performance or, if the outliers are far from the filtering regions, even possible loss of convergence. However, this problem is much more pronounced when only the eigenvalues of the mean consensus iteration matrix are known. As evident from the small spread of convergence rates in the simulations, this is not a severe problem and only occurs with low probability. The results for the directed case experience more variability due to outlying eigenvalues, and these outlying are somewhat more likely to be real due to a phenomenon of the circular law (and related laws) sometimes called the "Saturn effect", in which eigenvalues concentrate on the real line for finite matrix size [51]. This effect disappears asymptotically as the matrix size increases. At the cost of some optimality, the optimization problem could compensate for this effect by considering the set $\Lambda_{\kappa,\tau}$ as a lower bound for the filtering region, which can be slightly expanded to capture close outlying eigenvalues, especially on the real axis. Furthermore, it would be possible to add constraints to force convergence robustness by limiting the response magnitude over all possible eigenvalues at the cost of reducing optimality.

4.4 Weighted Filter Response Optimization

More generally, the filter design problem can be modified by applying penalties dependent on λ that weight the frequency response. This leads to the more general problems

$$\begin{aligned} &\min_{p\in P_d}\max_{\lambda\in\Lambda_{\kappa,\tau}} \left|g\left(\lambda\right)p\left(\lambda\right)\right| \\ &\text{s.t.} \qquad p\left(1\right)=1 \\ \Lambda_{\kappa,\tau} = &\left\{\lambda\in\mathbb{R}\middle|1-\lambda>\kappa, \widehat{f}_{W_N}\left(\lambda\right)>\tau\right\} \end{aligned} \tag{4.20}$$

for undirected networks and

$$\begin{aligned} &\min_{p\in P_d}\max_{\lambda\in\Lambda_{\kappa,\tau}} \left|g\left(\lambda\right)p\left(\lambda\right)\right| \\ &\text{s.t.} \qquad p\left(\lambda\right)=1 \\ \Lambda_{\kappa,\tau} = &\left\{\lambda\in\mathbb{C}\middle|\left|1-\lambda\right|>\kappa, \widehat{f}_{W_N,\beta}\left(\mathrm{Re}\left\{\lambda\right\},\mathrm{Im}\left\{\lambda\right\}\right)>\tau\right\} \end{aligned} \tag{4.21}$$

for directed networks where g is a nonnegative weight function. For undirected networks, the following linear program results from inclusion of the weights

$$\begin{aligned} \min_{\substack{\eta\in\mathbb{R}_+ \\ \mathbf{a}\in\mathbb{R}^{d+1}}} \quad & \eta \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \\ & g(\lambda)\,\mathbf{v}(\lambda)^\top \mathbf{a} < \eta \\ & -g(\lambda)\,\mathbf{v}(\lambda)^\top \mathbf{a} < \eta \\ & \text{for all } \lambda \in \Lambda_S \end{aligned} \tag{4.22}$$

where $\mathbf{v}(\lambda)$ is the $(d+1)\times 1$ Vandermode column vector

$$\mathbf{v}(\lambda) = \left[\lambda^0, \ldots, \lambda^d\right]^\top \tag{4.23}$$

such that $p(\lambda) = \mathbf{v}(\lambda)^\top \mathbf{a}$. For directed networks, the following linear objective with quadratic constraints results from inclusion of the weights

$$\begin{aligned} \min_{\substack{\eta\in\mathbb{R}_+ \\ \mathbf{a}\in\mathbb{R}^{d+1}}} \quad & \eta \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \\ & (g(\lambda))^2\,\mathbf{a}^\top Q(\lambda)\,\mathbf{a} < \eta \\ & \text{for all } \lambda \in \Lambda_S \end{aligned} \tag{4.24}$$

where $Q(\lambda)$ is the $(d+1)\times(d+1)$ real, positive semi-definite matrix

$$Q(\lambda) = \frac{1}{2}\left(\mathbf{v}(\lambda)\,\mathbf{v}(\lambda)^* + \mathbf{v}\left(\overline{\lambda}\right)\mathbf{v}\left(\overline{\lambda}\right)^*\right) \tag{4.25}$$

and $\mathbf{v}(\lambda)$ is the $(d+1)\times 1$ Vandermode column vector

$$\mathbf{v}(\lambda) = \left[\lambda^0, \ldots, \lambda^d\right]^\top \tag{4.26}$$

such that $|p(\lambda)|^2 = \mathbf{a}^\top Q(\lambda)\,\mathbf{a}$.

When optimizing worst case distributed consensus convergence rate as in (4.9) and (4.17), the proper weight function is simply

$$g_1(\lambda) = 1. \tag{4.27}$$

This choice for g minimizes the error for the worst eigenvector to produce fastest asymptotic convergence. Alternatively, the $\boldsymbol{\ell}_2$ graph total variation of the states after a single application of the filter may be considered. Intuitively, signals of low total variation have node values close to the average of neighboring node values. In terms of the row-normalized Laplacian, the total variation $\mathrm{TV}_{\mathcal{G}}(\mathbf{x}) = \left\|\mathcal{L}_R\left(\mathcal{G}\right)\mathbf{x}\right\|_2$ of the filter output corresponding to normalized eigenvector $\mathbf{v}_i$ of $W\left(\mathcal{G}\right) = I - \alpha\mathcal{L}_R\left(\mathcal{G}\right)$ with eigenvalue $\lambda_i\left(W\left(\mathcal{G}\right)\right)$ is

$$\mathrm{TV}_{\mathcal{G}}\left(p\left(W\left(\mathcal{G}\right)\right)\mathbf{v}_i\right) = \left|\tfrac{1}{\alpha}\left(1 - \lambda_i\left(W\left(\mathcal{G}\right)\right)\right)p\left(\lambda_i\left(W\left(\mathcal{G}\right)\right)\right)\right|, \tag{4.28}$$

penalizing high $\left|\lambda_i\left(\mathcal{L}_R\left(\mathcal{G}\right)\right)\right| = \left|\tfrac{1}{\alpha}\left(1 - \lambda_i\left(W\left(\mathcal{G}\right)\right)\right)\right|$. The worst case total variation of the filter output normalized by the magnitude of the initial error relates to the spectral radius by

$$\begin{aligned} &\rho\left(\mathcal{L}_R\left(\mathcal{G}\right)p\left(W\left(\mathcal{G}\right)\right) - J_{\boldsymbol{\ell}}\right) \leq \ldots \\ &\qquad \ldots \left\|\mathcal{L}_R\left(\mathcal{G}\right)p\left(W\left(\mathcal{G}\right)\right) - J_{\boldsymbol{\ell}}\right\|_2 \leq \ldots \\ &\qquad\qquad \ldots \left\|V\right\|_2\left\|V^{-1}\right\|_2\rho\left(\mathcal{L}_R\left(\mathcal{G}\right)p\left(W\left(\mathcal{G}\right)\right) - J_{\boldsymbol{\ell}}\right) \end{aligned} \tag{4.29}$$

where V diagonalizes $W\left(\mathcal{G}\right)$ and $\mathcal{L}_R\left(\mathcal{G}\right)$. Hence, the proper weight function for this problem is

$$g_2\left(\lambda\right) = \left|\tfrac{1}{\alpha}\left(1 - \lambda\right)\right|. \tag{4.30}$$

If the filter was, instead, designed to minimize expected total variation, the frequency response magnitude should be minimized for values of λ for which the empirical spectral distribution is more dense as well as large values of $\left|\tfrac{1}{\alpha}(1 - \lambda)\right|$. Hence, a reasonable weight function to optimize expected total variation would be

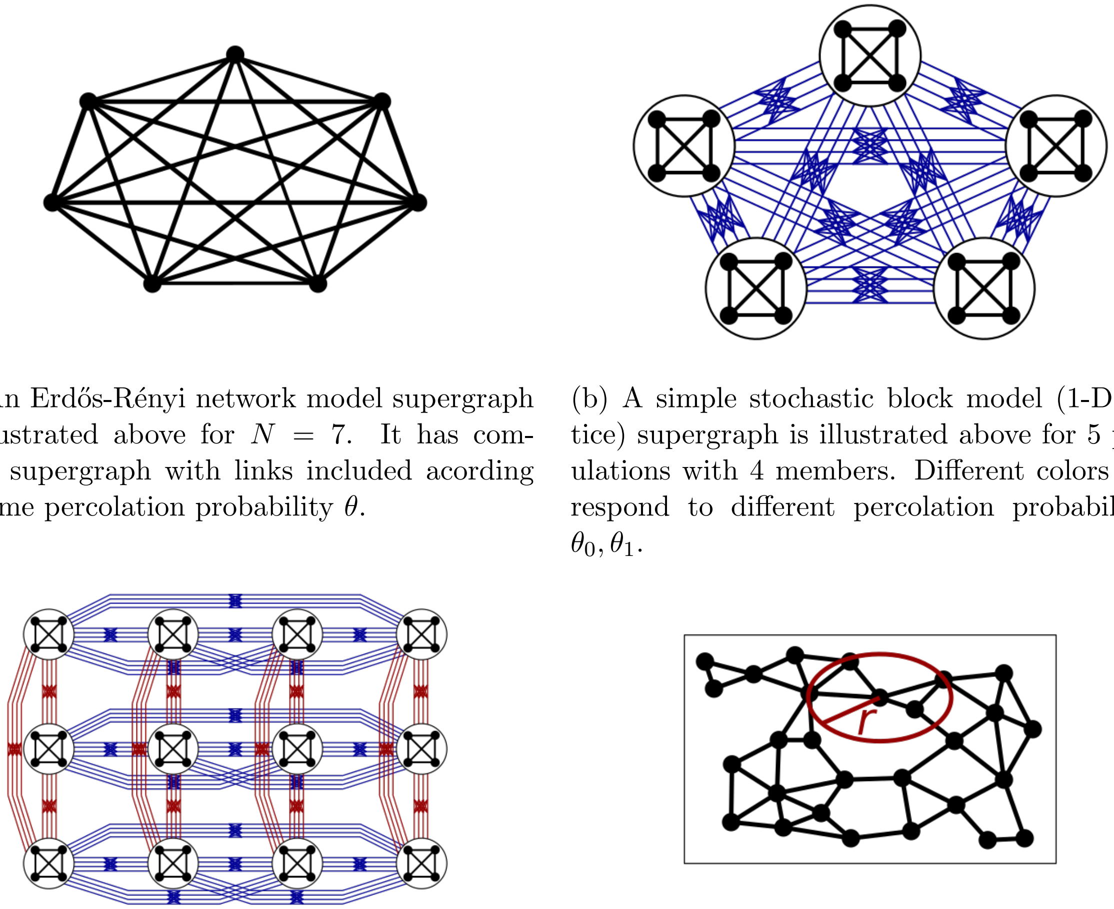
$$g_3\left(\lambda\right) = \left|\tfrac{1}{\alpha}\left(1 - \lambda\right)\right| f_N\left(\lambda\right). \tag{4.31}$$

Note that the factor of $\frac{1}{\alpha}$ could be removed from g_2 and g_3 without changing the optimization problem but is included to more clearly show the relationship to the total variation. Simulated results for the undirected case of each of these design problems appear in Section 4.5 for several random network models.

4.5 Simulations for Weighted Problems

To demonstrate the efficacy of the filter design methods proposed in the previous section, this section provides extensive simulations establishing performance improvement over existing methods for several random network models and filter design objectives. First, this section introduces each network model considered and briefly discusses derivation of deterministic equivalents for the iteration matrix empirical spectral distribution, which is primarily in the scope of [22, 23]. Subsequently, it describes the simulation procedure for each design objective and interprets results that appear in Figures 4.16-4.23.

Several practical random network models lead to random graph matrices with eigenvalues that are amenable to analysis through the methods of random matrix theory. The Erdős-Rényi model and stochastic block models have already been introduced. The simulations in this section will make use of stochastic block models with the following structure. (This serves to introduce a new example, but it requires no new analysis beyond those already presented in Chapter 3.) Consider a D-dimensional lattice structured stochastic block model in which node populations correspond to D-tuples and in which nodes can connect only if their populations differ by at most one tuple symbol. Thus, the populations form a lattice in the sense of [52]. Nodes connect through a link with a fixed probability that depends on the tuple symbol, if any, that differs between populations. Hence, nodes in the same population connect with probability θ_0 and nodes in different populations differing in the kth tuple connect with probability θ_k. For instance, this could arise if the populations were organized by responses to D categorical random variables and nodes only communicate with sufficiently similar nodes. For the $D = 2$ case, this percolation model leads to supergraph and different percolation probabilities like those shown in Figure 4.15c. The adjacency matrix spectral statistics of this model were analyzed in [22, 23], using Girko's K1 method (see Section 3.2), simplifying the problem using model symmetry and simultaneous diagonalizability. Transforming the computed deterministic equivalents for the adjacency matrix empirical spectral distribution to approximate the row-normalized Laplacian empirical spectral distribution, simulations using each of these random network models test the proposed filter design methods.



(a) An Erdős-Rényi network model supergraph is illustrated above for $N = 7$. It has complete supergraph with links included acording to some percolation probability θ.

(b) A simple stochastic block model (1-D lattice) supergraph is illustrated above for 5 populations with 4 members. Different colors correspond to different percolation probabilities θ_0, θ_1.

(c) A 2-D lattice stochastic block model supergraph is illustrated above for 3×4 populations with 4 members. Different colors correspond to different percolation probabilities $\theta_0, \theta_1, \theta_2$.

(d) A random geographic location model is illustrated above. Nodes are independently placed at random in a unit area and connected through a link if within the communication radius r.

Figure 4.15: Illustrations for example random network models

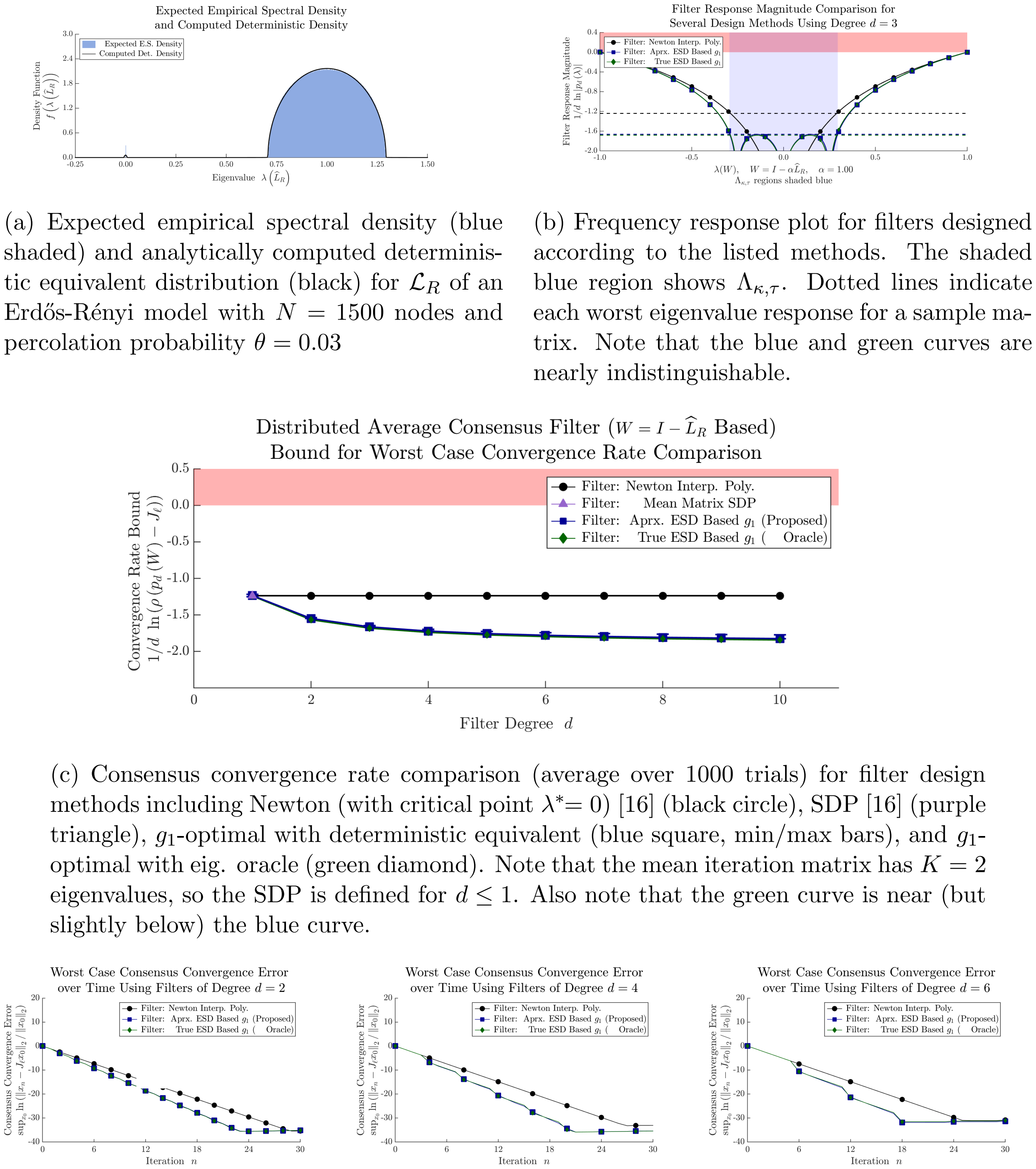
Additionally, the paper [16] examines a random geographic location model [53] when testing the filter design methods it introduces. Therefore, the simulations in this section will include a slight generalization of this model to enable more direct comparison. Under this model, nodes are randomly placed in a unit area and connected by a link if the distance is less than communication radius $r = c\sqrt{\ln(N)/N}$ [16]. An illustration depicting this appears in Figure 4.15d. Because this network distribution is not amenable to analysis using Girko's methods, the eigenvalues will be characterized by simulating the expected histogram over 10^4 Monte-Carlo samples for test purposes. Similarly the mean row-normalized Laplacian is computed by simulation for this model. For comparison, the Monte-Carlo approximation of the expected empirical spectral density requires several minutes ($\approx$55.1 minutes on a 2014 Macbook Pro laptop computer in the simulations), depending linearly on the number of samples and scaling rapidly with the number of network nodes, while the analytic computation for the other graph models require a few seconds ($\approx$5.0 seconds on a 2014 Macbook Pro laptop computer in the simulations), depending on the number of computation points and required accuracy but not scaling with the number of network nodes.

In the first group of simulations, filters of varying degree were designed to improve the convergence rate of the distributed average consensus process using iteration matrix $W(\mathcal{G}_N) = I - \alpha\mathcal{L}_R(\mathcal{G}_N)$ for $\alpha = 1$. For each network model, a deterministic equivalent for the adjacency matrix empirical spectral distribution was computed (or simulated) as described above [22, 23, 33]. After transformation, this information was used to approximate the empirical spectral distributions of $\mathcal{A}_R(\mathcal{G}_N)$ and, thus, $W(\mathcal{G}_N)$ [22, 23]. Using this information, filters of degree $d = 1, \ldots, 10$ were designed to minimize the convergence rate bound using (4.22) with weight function g_1. The resulting convergence rates are compared against those of the Newton interpolating polynomial (with critical point at $\lambda^* = 0$) and mean matrix SDP method from [16]. Note that the maximum degree for the mean matrix SDP methods is limited by the number of distinct mean matrix eigenvalues. Thus, for the Erdős-Rényi and random geographic location models, which have highly symmetric distributions, only first order filters are available for that method. Additionally, the results are compared against filters designed with full knowledge of the iteration matrix eigenvalues after the network is realized.

Figures 4.16-4.22 show results for an Erdős-Rényi model, 2-D lattice stochastic block model, 3-D lattice stochastic block model, and random location network, respectively, with model parameters listed in the corresponding caption. Figures 4.16a-4.22a show the computed deterministic approximation and the simulated expected empirical spectral distribution of W. Figures 4.16b-4.22b show filter responses for degree $d = 4$ filters. Figures 4.16c-4.22c compare the expected convergence rate bounds for the filter design methods. That is, the plotted curves show the filter response for the worst eigenvalue averaged over 1000 simulation trials. The outlying maximum and minimum over the 1000 trials for the proposed method are plotted as whiskers to show the extent of the extremes (whiskers for other methods omitted for clarity). Note that the minimum is often too close to the mean to be easily visible, but the maximum outliers are sometimes substantial. However, most results are extremely close to the mean, resulting in a sample standard deviation size too small to effectively include in the plot. Note that the proposed method (blue square) performs almost exactly as well as when full knowledge of the eigenvalues is available (green diamond) and outperforms the Newton polynomial method (black circle) and mean matrix SDP method (purple triangle) from [16] in all cases. Figures 4.16d-4.22d, 4.16e-4.22e, and 4.16f-4.22f show the consensus convergence error over time for filter lengths $d = 2$, $d = 4$, and $d = 6$.

In the second group of simulations, filters of varying degree were designed to improve the worst case ℓ_2 graph total variation of the data after a single filter application for iteration matrix $W(\mathcal{G}_N) = I - \alpha\mathcal{L}_R(\mathcal{G}_N)$ with $\alpha = 1$. Using the previously computed deterministic empirical spectral distribution approximations, filters of degree $d = 1, \ldots, 10$ were designed to minimize the worst case total variation bound using (4.22) with weight function g_2. The results are compared against those of the Newton interpolating polynomial (with critical point at $\lambda^* = 0$) and mean matrix SDP method from [16] as well as filters designed with full knowledge of the iteration matrix eigenvalues after the network is realized.

Figures 4.17a-4.23a compare the expected worst case total variation after a single application of each filter type for varying filter degrees, each averaged over 1000 simulation trails. Again, note that the proposed method (blue square) performs almost exactly as well as having full knowledge of the eigenvalues (green diamond) and outperforms the Newton polynomial method [16] (black circle) and mean matrix SDP method [16] (purple triangle) in all cases. This simulation was also repeated for filters designed to improve the expected ℓ_2 graph total variation of the data after a single filter application for iteration matrix $W(\mathcal{G}_N) = I - \alpha\mathcal{L}_R(\mathcal{G}_N)$ with $\alpha = 1$. For this problem, filters of degree $d = 1, \ldots, 10$ were designed using (4.22) with weight function g_3. The results, averaged over 1000 simulation trials, appear in Figures 4.17b-4.23b and display the same performance patterns among the filter types.



(a) Expected empirical spectral density (blue shaded) and analytically computed deterministic equivalent distribution (black) for $\mathcal{L}_R$ of an Erdős-Rényi model with $N = 1500$ nodes and percolation probability $\theta = 0.03$

(b) Frequency response plot for filters designed according to the listed methods. The shaded blue region shows $\Lambda_{\kappa,\tau}$. Dotted lines indicate each worst eigenvalue response for a sample matrix. Note that the blue and green curves are nearly indistinguishable.

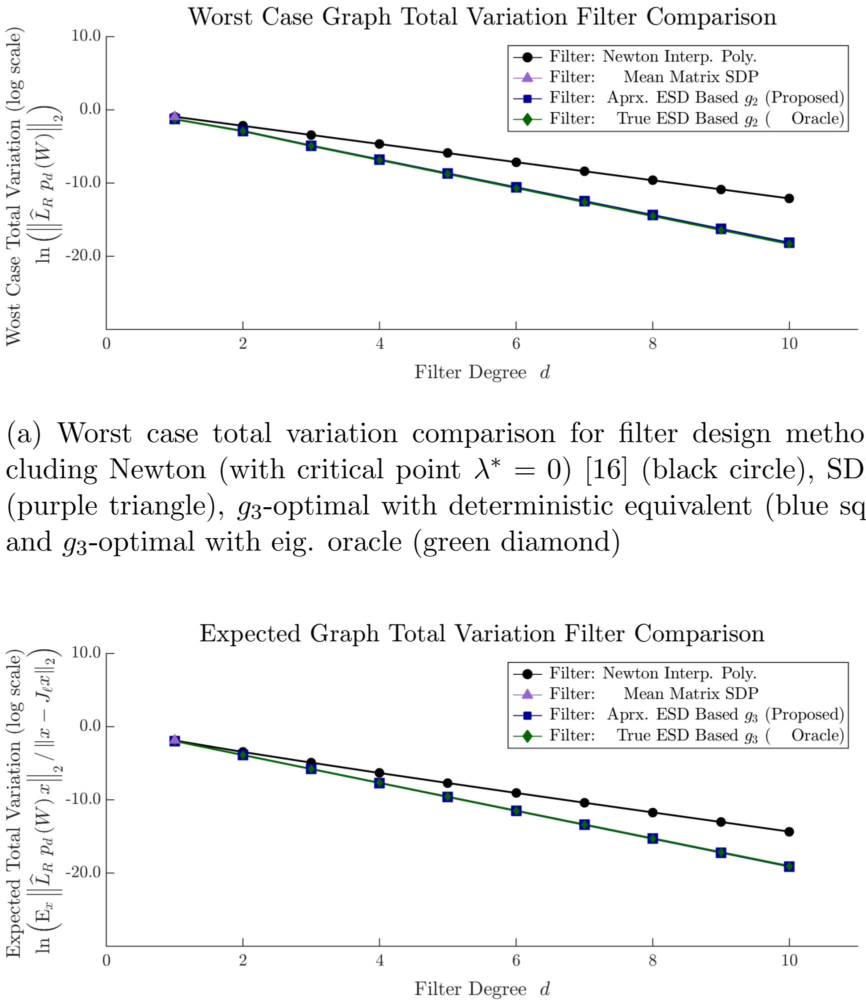
(c) Consensus convergence rate comparison (average over 1000 trials) for filter design methods including Newton (with critical point $\lambda^*= 0$) [16] (black circle), SDP [16] (purple triangle), g_1-optimal with deterministic equivalent (blue square, min/max bars), and g_1-optimal with eig. oracle (green diamond). Note that the mean iteration matrix has $K = 2$ eigenvalues, so the SDP is defined for $d \leq 1$. Also note that the green curve is near (but slightly below) the blue curve.

(d) Example worst case consensus convergence error over time using degree $d = 2$ filter for design methods listed in Figure 4.16c

(e) Example worst case consensus convergence error over time using degree $d = 4$ filter for design methods listed in Figure 4.16c

(f) Example worst case consensus convergence error over time using degree $d = 6$ filter for design methods listed in Figure 4.16c

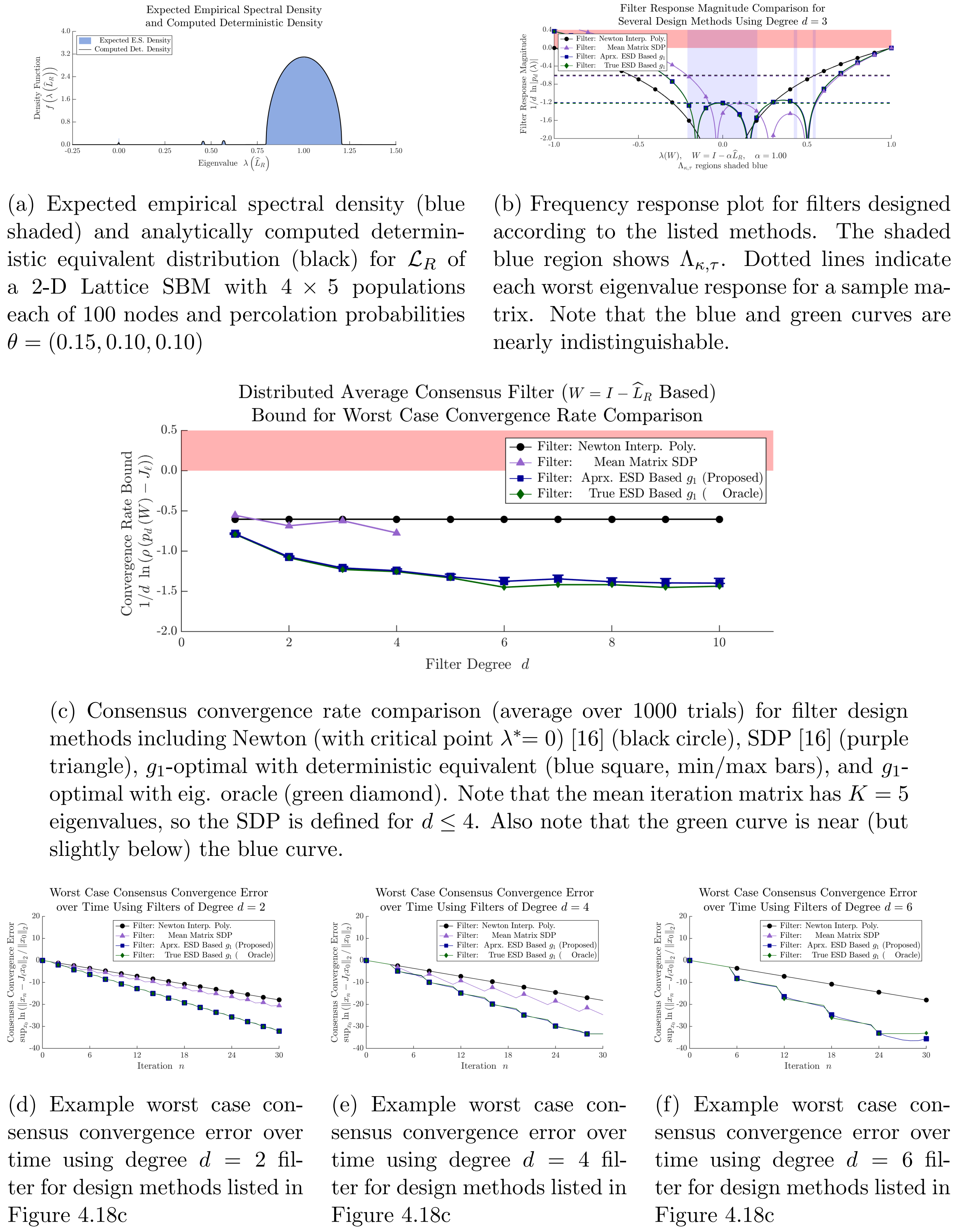
Figure 4.16: Consensus filter design for Erdős-Rényi network



(a) Worst case total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

(b) Expected total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

Figure 4.17: Worst case and expected total variation filter design for Erdős-Rényi network (parameters from Figure 4.16). Note that $K = 2$ so the mean iteration matrix SDP method is only defined for $d \leq 1$. (Results are averaged over 1000 trials.)



(a) Expected empirical spectral density (blue shaded) and analytically computed deterministic equivalent distribution (black) for $\mathcal{L}_R$ of a 2-D Lattice SBM with 4×5 populations each of 100 nodes and percolation probabilities $\theta = (0.15, 0.10, 0.10)$

(b) Frequency response plot for filters designed according to the listed methods. The shaded blue region shows $\Lambda_{\kappa,\tau}$. Dotted lines indicate each worst eigenvalue response for a sample matrix. Note that the blue and green curves are nearly indistinguishable.

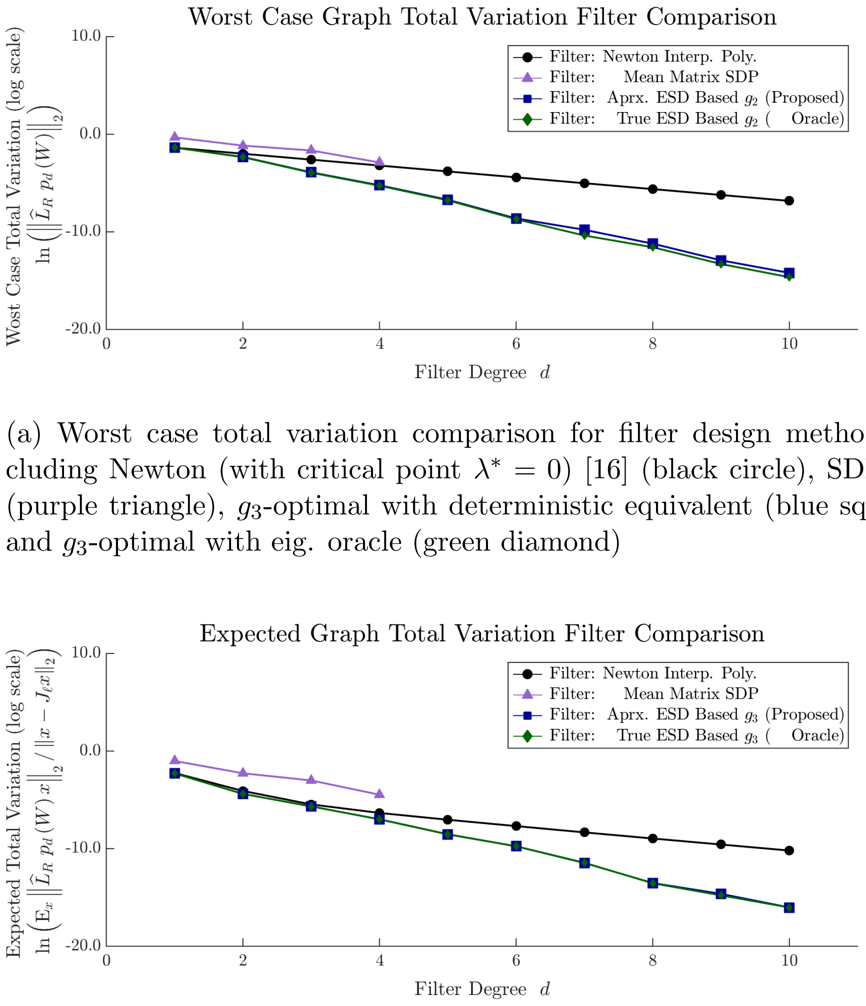
(c) Consensus convergence rate comparison (average over 1000 trials) for filter design methods including Newton (with critical point $\lambda^*= 0$) [16] (black circle), SDP [16] (purple triangle), g_1-optimal with deterministic equivalent (blue square, min/max bars), and g_1-optimal with eig. oracle (green diamond). Note that the mean iteration matrix has $K = 5$ eigenvalues, so the SDP is defined for $d \leq 4$. Also note that the green curve is near (but slightly below) the blue curve.

(d) Example worst case consensus convergence error over time using degree $d = 2$ filter for design methods listed in Figure 4.18c

(e) Example worst case consensus convergence error over time using degree $d = 4$ filter for design methods listed in Figure 4.18c

(f) Example worst case consensus convergence error over time using degree $d = 6$ filter for design methods listed in Figure 4.18c

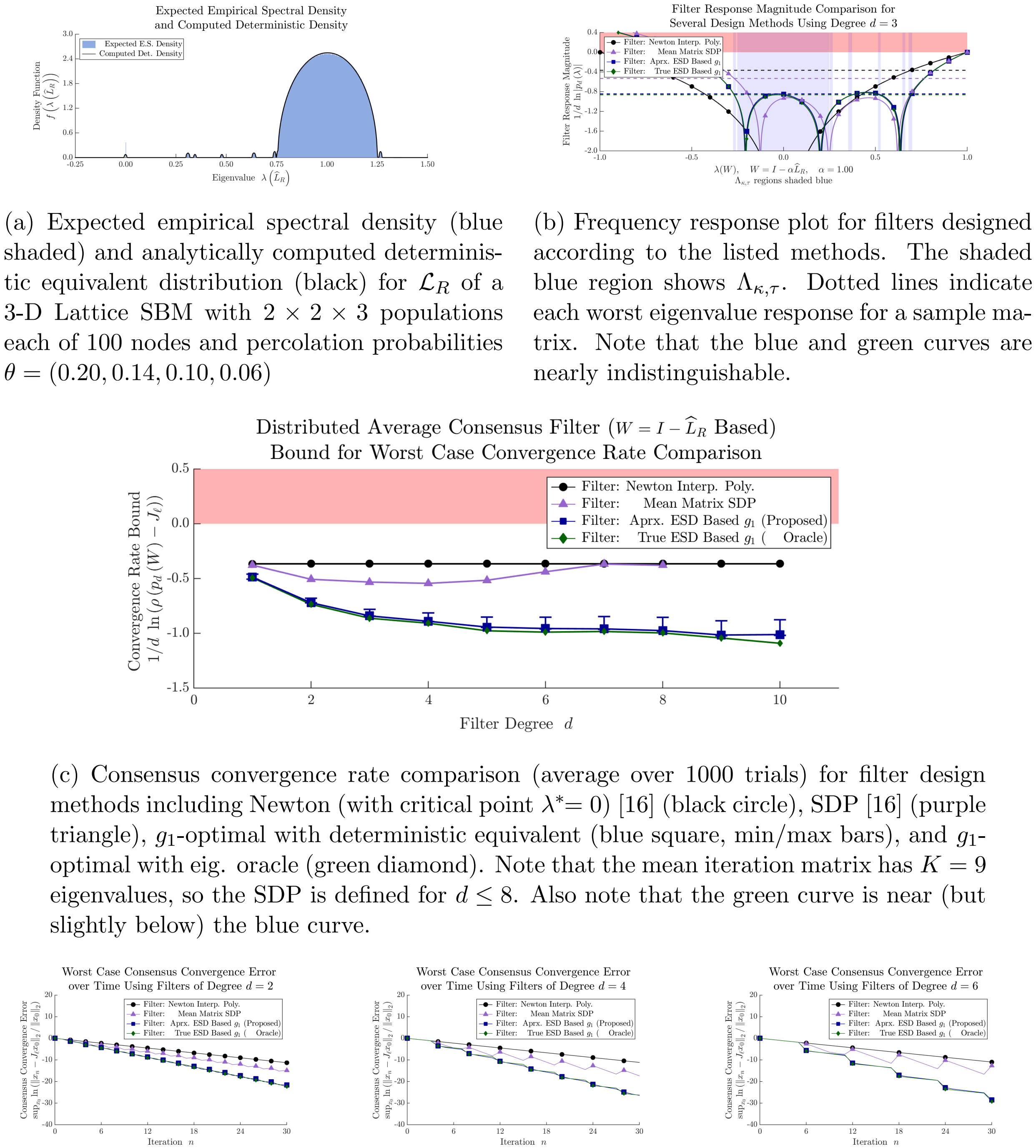
Figure 4.18: Consensus filter design for 2-D Lattice SBM network



(a) Worst case total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

(b) Expected total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

Figure 4.19: Worst case and expected total variation filter design for 2-D Lattice SBM network (parameters from Figure 4.18). Note that $K = 5$ so the mean iteration matrix SDP method is only defined for $d \leq 4$. (Results are averaged over 1000 trials.)



(a) Expected empirical spectral density (blue shaded) and analytically computed deterministic equivalent distribution (black) for $\mathcal{L}_R$ of a 3-D Lattice SBM with $2 \times 2 \times 3$ populations each of 100 nodes and percolation probabilities $\theta = (0.20, 0.14, 0.10, 0.06)$

(b) Frequency response plot for filters designed according to the listed methods. The shaded blue region shows $\Lambda_{\kappa,\tau}$. Dotted lines indicate each worst eigenvalue response for a sample matrix. Note that the blue and green curves are nearly indistinguishable.

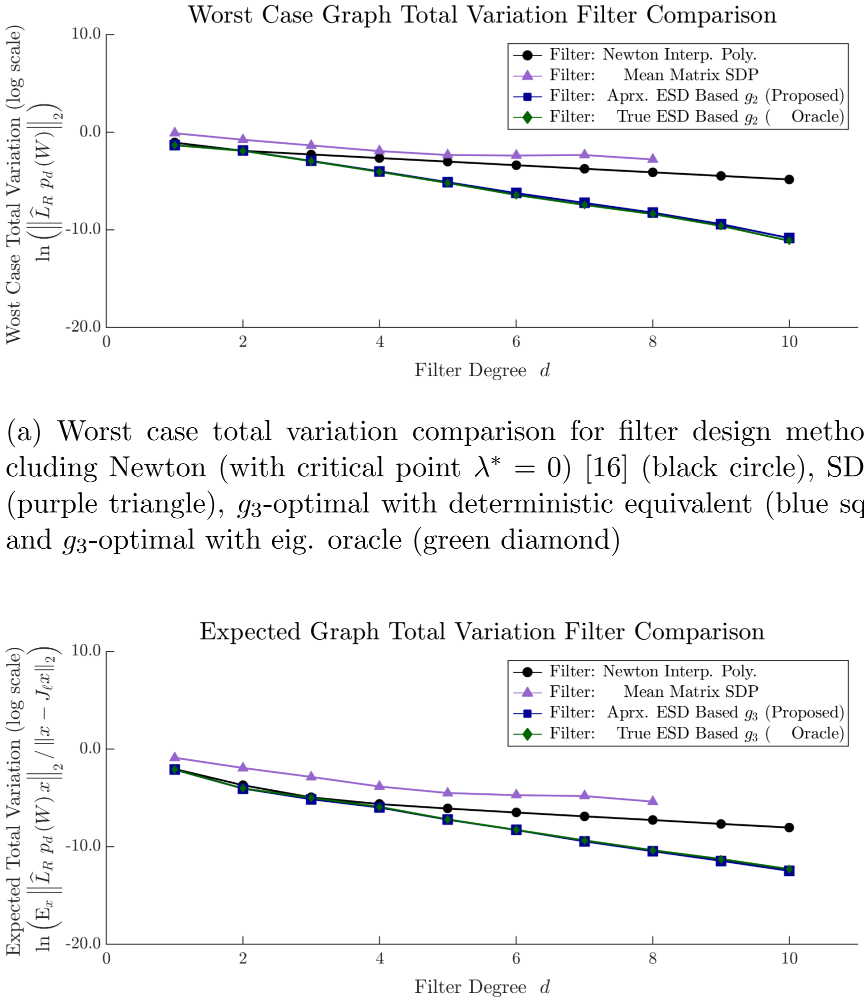
(c) Consensus convergence rate comparison (average over 1000 trials) for filter design methods including Newton (with critical point $\lambda^*= 0$) [16] (black circle), SDP [16] (purple triangle), g_1-optimal with deterministic equivalent (blue square, min/max bars), and g_1-optimal with eig. oracle (green diamond). Note that the mean iteration matrix has $K = 9$ eigenvalues, so the SDP is defined for $d \leq 8$. Also note that the green curve is near (but slightly below) the blue curve.

(d) Example worst case consensus convergence error over time using degree $d = 2$ filter for design methods listed in Figure 4.20c

(e) Example worst case consensus convergence error over time using degree $d = 4$ filter for design methods listed in Figure 4.20c

(f) Example worst case consensus convergence error over time using degree $d = 6$ filter for design methods listed in Figure 4.20c

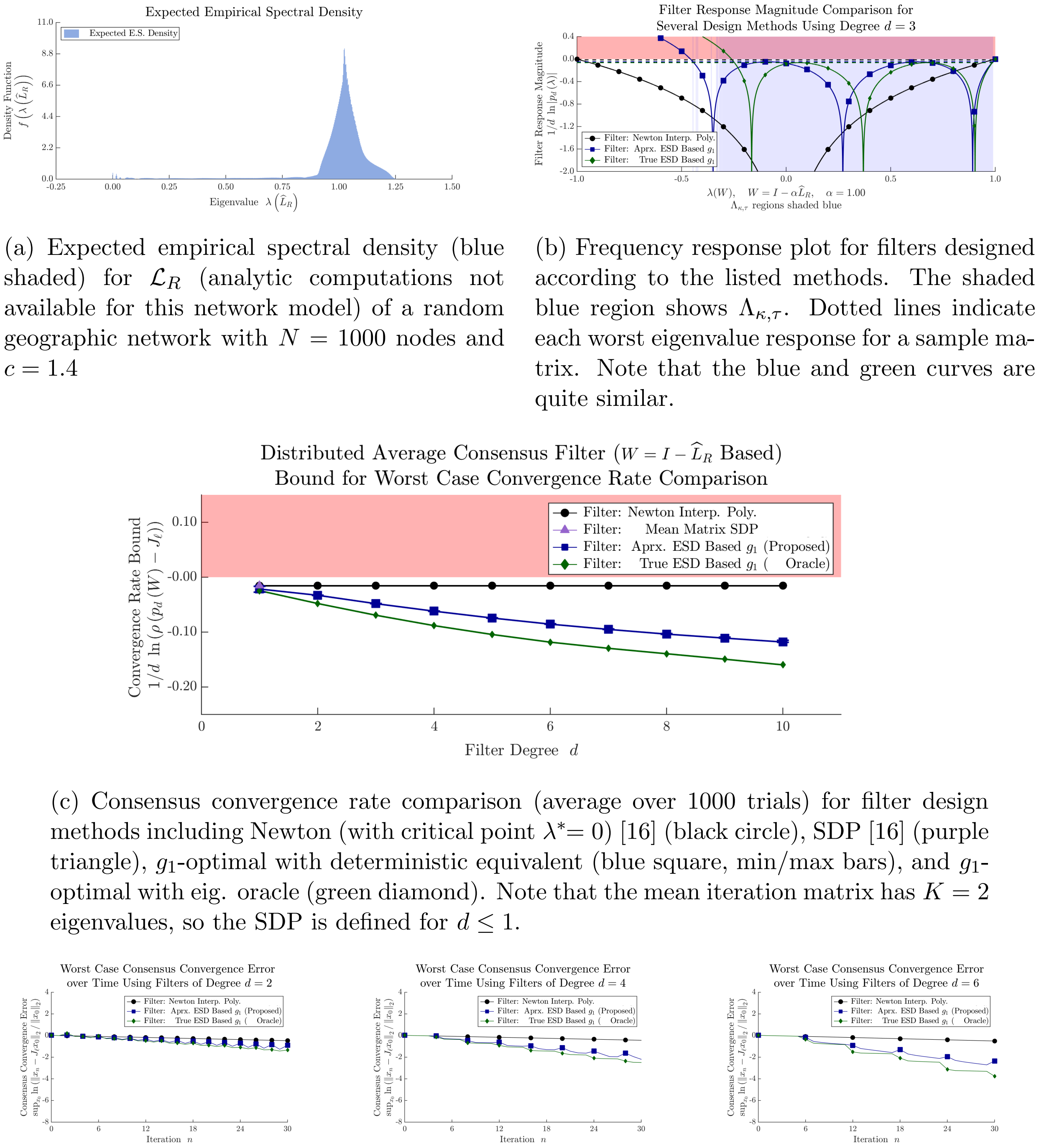
Figure 4.20: Consensus filter design for 3-D Lattice SBM network



(a) Worst case total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

(b) Expected total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

Figure 4.21: Worst case and expected total variation filter design for 3-D Lattice SBM network (parameters from Figure 4.20). Note that $K = 9$ so the mean iteration matrix SDP method is only defined for $d \leq 8$. (Results are averaged over 1000 trials.)



(a) Expected empirical spectral density (blue shaded) for $\mathcal{L}_R$ (analytic computations not available for this network model) of a random geographic network with $N = 1000$ nodes and $c = 1.4$

(b) Frequency response plot for filters designed according to the listed methods. The shaded blue region shows $\Lambda_{\kappa,\tau}$. Dotted lines indicate each worst eigenvalue response for a sample matrix. Note that the blue and green curves are quite similar.

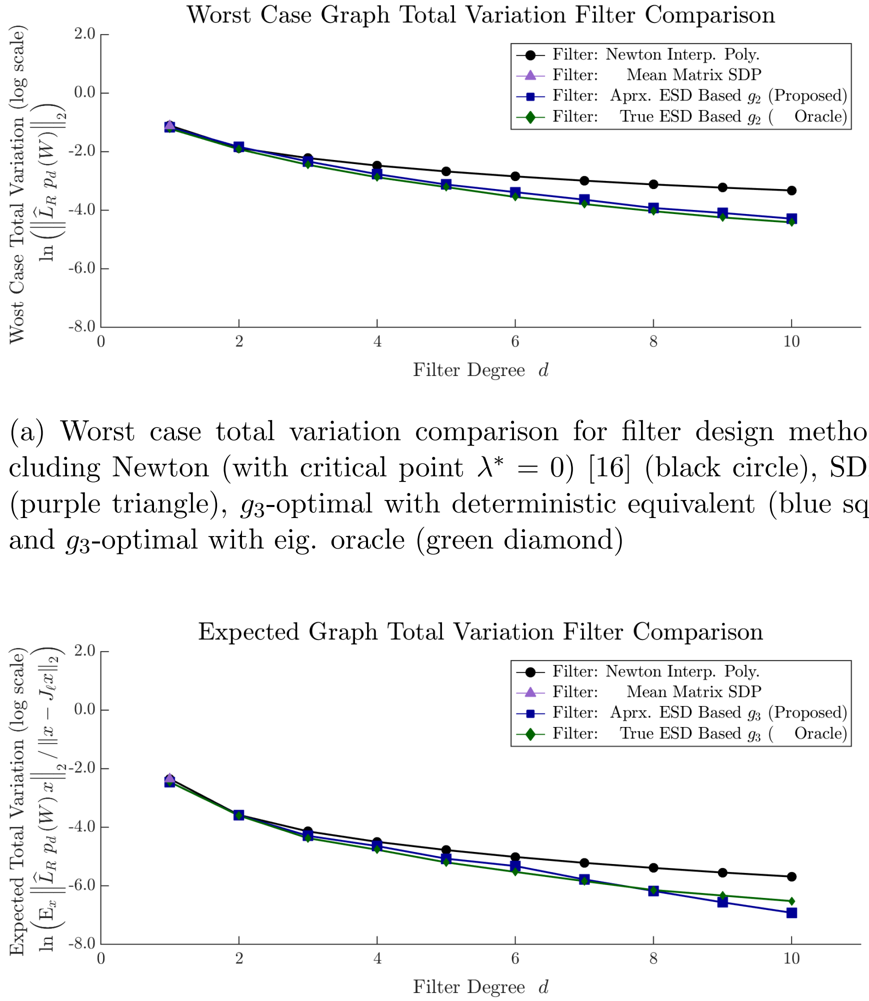
(c) Consensus convergence rate comparison (average over 1000 trials) for filter design methods including Newton (with critical point $\lambda^*= 0$) [16] (black circle), SDP [16] (purple triangle), g_1-optimal with deterministic equivalent (blue square, min/max bars), and g_1-optimal with eig. oracle (green diamond). Note that the mean iteration matrix has $K = 2$ eigenvalues, so the SDP is defined for $d \leq 1$.

(d) Example worst case consensus convergence error over time using degree $d = 2$ filter for design methods listed in Figure 4.22c

(e) Example worst case consensus convergence error over time using degree $d = 4$ filter for design methods listed in Figure 4.22c

(f) Example worst case consensus convergence error over time using degree $d = 6$ filter for design methods listed in Figure 4.22c

Figure 4.22: Consensus filter design for random geographic network



(a) Worst case total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

(b) Expected total variation comparison for filter design methods including Newton (with critical point $\lambda^* = 0$) [16] (black circle), SDP [16] (purple triangle), g_3-optimal with deterministic equivalent (blue square), and g_3-optimal with eig. oracle (green diamond)

Figure 4.23: Worst case and expected total variation filter design for random geographic network (parameters from Figure 4.22). Note that $K = 2$ so the mean iteration matrix SDP method is only defined for $d \leq 1$. (Results are averaged over 1000 trials.)

4.6 Summary

In summary, this chapter presented filter design methods for periodic consensus acceleration filters and other closely related graph filtering problems on constant large-scale random networks. These methods rely on the limiting behavior that can emerge in the eigenvalues of large-scale matrices, using the support of the approximate spectral densities for the row-normalized adjacency matrix obtained using the methods in Chapter 3 to define filtering regions with respect to the consensus iteration matrix. For undirected networks, minimizing the convergence rate leads to a linear program that minimizes a bound on the filter response magnitude over the support of the approximate density function. Similarly, for directed networks, minimizing the convergence rate leads to a quadratically constrained linear program that minimizes a bound on the squared filter response magnitude over the support of the approximate density function. Significant improvements for the filters designed with the proposed method over the filterless convergence rate and over the convergence rate of the mean SDP filter were observed in simulations for each of these design problems. Furthermore, the performance of the proposed filter methods is near that of the optimal filters designed with exact knowledge of the iteration matrix after the network is drawn from its distribution. Practical limitations of this filtering approach, including computational requirements, possible loss of robustness, and numerical issues. Subsequently, these optimization problems were generalized to minimize worst case weighted filter response magnitude, which handles optimization objectives such as the worst case total variance and the expected total variation. Simulation results for the undirected case of these related problems were also provided. For each filter design problem, the filters derived using the proposed method and the approximate spectral density obtained via Girko's methods perform nearly as well as the optimal solution designed with exact knowledge of the iteration matrix after the network is drawn from its distribution.

Chapter Five

Consensus Filter Design: Switching Networks

5.1 Introduction

For networks that change over time, the concept of a shift-invariant graph filter breaks down because the graph shift operator changes along with the network. This significantly complicates the graph filter design process for consensus acceleration, as the same notion of filter response does not apply. However, for some random network models analysis is still possible with respect to the sequence of graph shift matrix eigendecompositions, just as graph signal processing for constant network analyzes data with respect to the eigendecomposition of the constant graph shift operator. For instance, this chapter considers switching random networks, a time-varying random network model. In the switching random network models considered in this thesis, the marginal distribution of the network model at each iteration has some known distribution. At each iteration, the network either remains unchanged or potentially changes to an independently drawn sample from the distribution according to the result of a Bernoulli random variable with switching probability θ_{sw}.

This chapter presents optimization methods for consensus filtering on large-scale random undirected switching networks. These optimization methods are designed with the objective to approximately minimize the consensus error norm after a single filter application. Section 5.2 focuses on filter design for unnormalized Laplacian-based consensus iteration matrices of undirected graphs based on simulated expected empirical spectral distribution data. The filter design optimization problem to minimize the expected consensus error vector norm squared results in a quadratic program, where the matrix in the quadratic objective function can be approximated from the empirical spectral distribution moments given certain assumptions on the eigenvectors. Section 5.3 focuses on filter design for row-normalized Laplacian-based consensus iteration matrices of undirected switching networks based on the approximate spectral density obtained through Girko's methods as described in Chapter 3. The filter design optimization problem to minimize the expected consensus error vector norm squared results in a quadratic program nearly identical to that for the unnormalized Laplacian-based consensus iteration matrix case in Section 5.2, with the exception of employing the approximate spectral density. However, the derivation must be modified to account for the asymmetry of the iteration matrices and resulting non-orthogonal bases of eigenvectors. Finally, Section 5.4 summarizes the chapter in conclusion.

5.2 Switching, Undirected Random Networks: Symmetric Iteration Matrix Case ($\boldsymbol{W}=\boldsymbol{I}-\boldsymbol{\alpha}\boldsymbol{\mathcal{L}}$)

This section proposes design criteria for consensus acceleration filters on random switching networks, a relatively simple class of time-varying network models, for certain large-scale random network distributions using the unnormalized Laplacian based consensus iteration matrix $W(\mathcal{G}) = I - \alpha\mathcal{L}(\mathcal{G})$. Each network is drawn from some specified random network distribution. At each time iteration, the network either remains constant or switches to a new, independent sample from the random network distribution according to a Bernoulli trial with fixed probability. The proposed quadratic optimization objective involves an approximation of the expected Gram matrix of error in the state vectors over a filtering window. The derived approximation depends only on the moments of the expected empirical spectral distribution (obtainable via Monte-Carlo simulation) of the iteration matrices and on the switching probability.

Consider distributed average consensus with respect to a time-varying sequence of iteration matrices $\{W(\mathcal{G}_{N,n})\}$ arising from a random switching network with N nodes and switching probability θ_{sw}. To compute an unweighted mean of the initial data $\mathbf{x}_0$, the network implements a dynamic system with state $\mathbf{x}_n$ at time iteration n described by

$$\mathbf{x}_n = W\left(\mathcal{G}_{N,n}\right)\mathbf{x}_{n-1} \tag{5.1}$$

Provided the iteration matrices satisfy the consensus conditions

$$\begin{gathered} W\left(\mathcal{G}_{N,n}\right)\mathbf{1} = \mathbf{1},\ \mathbf{1}^\top W\left(\mathcal{G}_{N,n}\right) = \mathbf{1}^\top, \\ \rho\left(W\left(\mathcal{G}_{N,n}\right) - J_{\mathbf{1}}\right) < 1 \end{gathered} \tag{5.2}$$

where $J_{\mathbf{1}} = \mathbf{1}\mathbf{1}^\top/\mathbf{1}^\top\mathbf{1}$, the agreement error in the state vector is asymptotically eliminated, converging to the unweighted mean. This section examines undirected graphs and employs the iteration matrix scheme

$$W_{N,n} = W\left(\mathcal{G}_{N,n}\right) = I - \alpha\mathcal{L}\left(\mathcal{G}_{N,n}\right), \tag{5.3}$$

which satisfies the properties in (5.2) when the network graphs $\{\mathcal{G}_{N,n}\}$ are connected and α is chosen suitably. To improve the convergence rate, a degree d filter with coefficients $\{a_k\}_{k=0}^{k=d}$ will be periodically applied to previous state values to update the current state vector according to the following equation.

$$\mathbf{x}_n := \sum_{k=0}^{k=d} a_k\mathbf{x}_{n-d+k}, \quad n \equiv 0 \pmod{d} \tag{5.4}$$

for $k = 0, \ldots, d$. Thus, for initial vector $\mathbf{x}_0$ the state vector terms used for filtering (written for the first filtering operation only to simplify notation) are given by $\mathbf{x}_k = \phi_k\left(\{W_{N,n}\}_{n=1}^{n=d}\right)\mathbf{x}_0$ where

$$\begin{gathered} \phi_k\left(\{W_{N,n}\}_{n=1}^{n=d}\right) = W_{N,k}\ldots W_{N,1} \\ \phi_0\left(\{W_{N,n}\}_{n=1}^{n=d}\right) = I_N \end{gathered} \tag{5.5}$$

Because each iteration matrix has eigenvalue $\lambda = 1$ corresponding to the consensus eigenvector $\mathbf{1}$, the filter coefficients must have unit sum to preserve this eigenvalue in the filtered transformation. Collecting the filter coefficients into a vector

$$\mathbf{a} = [a_0, \ldots, a_d] \tag{5.6}$$

this constraint can be expressed as $\mathbf{1}^\top\mathbf{a} = 1$.

Attempting to directly optimize the expected norm of the filter output error for the worst case input proves challenging. Instead of directly optimizing the convergence rate, the filters designed in this section approximately minimize the expected norm of the filter output consensus error vector with respect to the random iteration matrix sequence and with respect to the initial error vector. Let $\mathbf{x}_0 = \overline{\mathbf{x}}_0\mathbf{1} + \mathbf{x}_{\mathrm{e}}$ where $\mathbf{x}_{\mathrm{e}}$ is orthogonal to $\mathbf{1}$, and assume for simplicity that $\mathbf{x}_{\mathrm{e}}$ is uniformly distributed on unit norm vectors orthogonal to $\mathbf{1}$. By Jensen's inequality, the square root of the expected norm squared provides a lower bound for the expected norm. Thus, rather than minimizing the expected norm of the filter error directly, the expected norm squared will be minimized as follows, where $\mathbf{x}_{\mathrm{e}}$ is uniformly distributed on $\left\{\mathbf{x}_{\mathrm{e}} \in \mathbb{R}^N | \mathbf{x}_{\mathrm{e}} \perp \mathbf{1}, \|\mathbf{x}_{\mathrm{e}}\| = 1\right\}$.

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathrm{E}_{\mathbf{x}_{\mathrm{e}},\left\{W_{N,n}\right\}_{n=1}^{n=d}} \left[\left\| (I - J_{\mathbf{1}}) \sum_{k=0}^{k=d} \mathbf{a}_k \phi_k \left(\{W_{N,n}\}_{n=1}^{n=d} \right) \mathbf{x}_{\mathrm{e}} \right\|_2^2 \right] \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \end{aligned} \tag{5.7}$$

The objective function in the above equation can be written in terms of the Gram matrix of consensus error vectors corresponding to each filtering term. The entries of this matrix are given by

$$\begin{aligned} (Q_{\mathrm{err}})_{ij} = \Big\langle & (I - J_{\mathbf{1}})\, \phi_{i-1} \left(\{W_{N,n}\}_{n=1}^{n=d} \right) \mathbf{x}_{\mathrm{e}}, \\ & (I - J_{\mathbf{1}})\, \phi_{j-1} \left(\{W_{N,n}\}_{n=1}^{n=d} \right) \mathbf{x}_{\mathrm{e}} \Big\rangle . \end{aligned} \tag{5.8}$$

Thus, the optimization problem can be rewritten in the following quadratic form.

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathbf{a}^\top \mathrm{E}_{\mathbf{x}_{\mathrm{e}},\left\{W_{N,n}\right\}_{n=1}^{n=d}} \left[Q_{\mathrm{err}} \right] \mathbf{a} \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \end{aligned} \tag{5.9}$$

Denote by $\mathbf{s}$ the network switching sequence where $\mathbf{s}_1 = 1$ and, for all $n > 1$, $\mathbf{s}_n = 1$ if $W_{N,n} = W_{N,n-1}$, and $\mathbf{s}_n = 0$ otherwise. That is, $\mathbf{s}_n$ determines whether $W_{N,n}$ is a new iteration matrix (within the filtering window), with $W_{N,1}$ always counted. The space of all possible switching sequences are the d-tuples

$$S_d = \{1\} \times \{0,1\} \times \cdots \times \{0,1\}\,, \tag{5.10}$$

and the total number of independent networks is

$$\#\mathbf{s} = \sum_{k=1}^{k=d} \mathbf{s}_k, \qquad \mathbf{s} \in S_d \tag{5.11}$$

with $\#\mathbf{s} - 1$ total switching events. For a switching process governed by independent Bernoulli trials with switching probability θ_{sw}, switching sequence $\mathbf{s} \in S_d$ has probability mass given by

$$f_S\left(\mathbf{s}\right) = \left(\theta_{\mathrm{sw}}\right)^{(\#\mathbf{s}-1)}\left(1-\theta_{\mathrm{sw}}\right)^{(d-1)-(\#\mathbf{s}-1)} \tag{5.12}$$

The distribution of the iteration matrix sequence $\{W_{N,n}\}_{n=1}^{n=d}$ can be factored into the distribution of the switching sequence $\mathbf{s}$ and distribution of $\{W_{N,n}\left(\mathbf{s}\right)\}_{n=1}^{n=d}$ given the switching sequence. The Gram matrix of filter term consensus error vectors can also be conditioned on the switching sequence as follows.

$$\begin{aligned}\left(Q_{\mathrm{err}}\left(\mathbf{s}\right)\right)_{ij} = \Big\langle &\left(I - J_{\mathbf{1}}\right)\phi_{i-1}\left(\{W_{N,n}\left(\mathbf{s}\right)\}_{n=1}^{n=d}\right)\mathbf{x}_{\mathrm{e}}, \\ &\left(I - J_{\mathbf{1}}\right)\phi_{j-1}\left(\{W_{N,n}\left(\mathbf{s}\right)\}_{n=1}^{n=d}\right)\mathbf{x}_{\mathrm{e}}\Big\rangle\end{aligned} \tag{5.13}$$

In terms of the factored distributions, the optimization problem in (5.9) can be reformulated as follows.

$$\begin{aligned}\min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathbf{a}^\top \mathrm{E}_{\mathbf{s}}\left[\mathrm{E}_{\mathbf{x}_{\mathrm{e}},\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}}\left[Q_{\mathrm{err}}\left(\mathbf{s}\right)|\mathbf{s}\right]\right]\mathbf{a} \\ \text{s.t.} \quad & \mathbf{1}^\top\mathbf{a} = 1\end{aligned} \tag{5.14}$$

The above optimization problem could be used for filter design by computing the values of $\mathrm{E}_{\{W_{N,n}(\mathbf{s})\},\mathbf{v}}\left[Q\left(\mathbf{s}\right)|\mathbf{s}\right]$ through simulation. However, the intent of this section is to connect information regarding the empirical spectral distribution of the iteration matrices to filter design for large-scale random switching networks, as done for constant random networks in

Chapter 4. Therefore, an analytical approximation (under suitable conditions) based on the expected empirical spectral distribution will be described below.

Before proceeding, some notation must first be introduced. Note that each switching sequence $\mathbf{s}$ corresponds to an integer composition of d with $\#\mathbf{s}$ partitions, namely

$$c(\mathbf{s}) = (c_1(\mathbf{s}), \ldots, c_{\#\mathbf{s}}(\mathbf{s})) \tag{5.15}$$

where each $c_m(\mathbf{s})$ is the number of iterations the mth network is used before switching. Let $c'_m(\mathbf{s}, n)$ be the number of iterations the mth network is used up to (and including) iteration $n \geq 1$ with $c'_m(\mathbf{s}, 0) = 0$. More explicitly,

$$c'_m(\mathbf{s}) = \begin{cases} 0 & n < \sum_{k=1}^{m-1} c_k(\mathbf{s}) \\ c_m(\mathbf{s}) & n > \sum_{k=1}^{m} c_k(\mathbf{s}) \\ n - \sum_{k=1}^{m-1} c_k(\mathbf{s}) & \text{otherwise} \end{cases} \tag{5.16}$$

Let $\{\mathbf{u}_{n,k}(\mathbf{s})\}_{k=1}^{k=N-1} \cup \{\mathbf{u}_{n,N}(\mathbf{s}) = \mathbf{1}/\sqrt{N}\}$ be a orthonormal basis of eigenvectors for $W_{N,n}(\mathbf{s})$ with $\{\mathbf{v}_k\}_{k=1}^{k=N-1} \cup \{\mathbf{v}_N = \mathbf{1}/\sqrt{N}\}$ an orthonormal basis of eigenvectors for $I - J_{\mathbf{1}}$. Note that if the network does not switch, the basis of eigenvectors remains the same. This motivates redefining some terms with respect to the number of independent networks to avoid duplication. Let $W'_{N,m}(\mathbf{s})$ be the mth independently drawn network for $m = 1, \ldots, \#\mathbf{s}$. Let the corresponding orthonormal bases of eigenvectors be $\left\{\mathbf{u}'_{m,k}(\mathbf{s})\right\}_{k=1}^{k=N-1} \cup \left\{\mathbf{u}'_{m,N}(\mathbf{s}) = \mathbf{1}/\sqrt{N}\right\}$. Let the basis change coordinates for this sequence of eigenvector bases be $\psi'_{(m,m+1),r_m r_{m+1}}$, which gives the coordinate of $\mathbf{u}'_{m,r_m}$ corresponding to $\mathbf{u}'_{m+1,r_{m+1}}$ for $1 \leq m < \#\mathbf{s}$. Furthermore, let the coordinate of $\mathbf{x}_e$ in $\mathbf{u}'_{1,r_1}$ be $\psi'_{(0,1),r_1}$. Finally, let the coordinate of $\mathbf{v}_\ell$ in $\mathbf{u}'_{\#\mathbf{s},r_{\#\mathbf{s}}}$ be $\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}\ell}$. Note that because all bases are orthogonal, these are all inner products.

These basis change coordinates are random (except for the ones corresponding to the consensus eigenvector, which remains constant), and assumptions must be made regarding their nature to enable computation. The results will vary in quality depending on how well these simplifying assumptions hold. Assume that the basis change coordinates have the following properties (divided into four sets). While it may look like there are many cases, they are generated from simple intuitions.

Assumption 1: First, note that the consensus eigenvector is always present and does not change. This will give the following conditions

$$\psi'_{(0,1),N} = 0 \tag{5.17}$$

$$\psi'_{(m,m+1),NN} = 1, \quad m = 1, \ldots, \#\mathbf{s} - 1 \tag{5.18}$$

$$\psi'_{(m,m+1),Nr_{m+1}} = 0, \quad m = 1, \ldots, \#\mathbf{s} - 1,\ r_{m+1} \neq N \tag{5.19}$$

$$\psi'_{(m,m+1),r_m N} = 0, \quad m = 1, \ldots, \#\mathbf{s} - 1,\ r_m \neq N \tag{5.20}$$

$$\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}N} = 1, \quad r_{\#\mathbf{s}} = N \tag{5.21}$$

$$\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}N} = 0, \quad r_{\#\mathbf{s}} \neq N \tag{5.22}$$

$$\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)N\ell} = 0, \quad \ell \neq N \tag{5.23}$$

The remaining assumptions deal with the coordinates corresponding to the other eigenvectors.

Assumption 2: Second, assume that basis change coordinates (other than the ones that remain constant) for different switching events are statistically independent.

Assumption 3: Third, assume that pairs of basis change coordinates for the same switching event are uncorrelated.

$$\mathrm{E}\left[\psi'_{(0,1)r_1}\psi'_{(0,1)t_1}\right] = 0, \quad r_1 \neq t_1,\ r_1, t_1 \neq N \tag{5.24}$$

$$\mathrm{E}\left[\psi'_{(m,m+1)r_m r_{m+1}}\psi'_{(m,m+1)t_m t_{m+1}}\right] = 0, \quad \begin{array}{l} r_m \neq t_m \text{ and/or } r_{m+1} \neq t_{m+1}, \\ m = 1, \ldots, \#\mathbf{s} - 1\ ,\ r_m, r_{m+1}, t_m, t_{m+1} \neq N \end{array} \tag{5.25}$$

$$\mathrm{E}\left[\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}\ell_1}\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)t_{\#\mathbf{s}}\ell_2}\right] = 0, \quad \begin{array}{l} r_{\#\mathbf{s}} \neq t_{\#\mathbf{s}} \text{ and/or } \ell_1 \neq \ell_2, \\ r_{\#\mathbf{s}}, t_{\#\mathbf{s}}, \ell_1, \ell_2 \neq N \end{array} \tag{5.26}$$

Assumption 4: Fourth, assume the following second moments for each basis change coordinate. This is motivated by assumption of even distribution of the signal energy among the eigenvectors.

$$\mathrm{E}\left[\left(\psi'_{(0,1)r_1}\right)^2\right] = \frac{1}{N-1}, \quad r_1 \neq N \tag{5.27}$$

$$\mathrm{E}\left[\left(\psi'_{(m,m+1)r_m r_{m+1}}\right)^2\right] = \frac{1}{N-1}, \quad r_m, r_{m+1} \neq N,\ m = 1, \ldots \#\mathbf{s} - 1 \tag{5.28}$$

$$\mathrm{E}\left[\left(\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}\ell}\right)^2\right] = \frac{1}{N-1}, \quad r_{\#\mathbf{s}}, \ell \neq N \tag{5.29}$$

Then $\mathrm{E}_{\{W_n(\mathbf{s})\},\mathbf{v}}\left[Q\left(\mathbf{s}\right)|\mathbf{s}\right]$ can be approximated as follows.

Proposition 5.1 (Approximate Gram Matrix)

Given a random switching network that satisfies the listed assumptions and the expected empirical spectral density $\mathrm{E}\left[f_{W_N}\right]$, *the following expression approximates the entries of the expected Gram matrix of consensus error in the filter terms conditioned on the switching sequence.*

$$\left(\widehat{Q}_{\mathrm{err}}\left(\mathbf{s}\right)\right)_{ij} = \prod_{m=1}^{m=\#\mathbf{s}} \mathrm{E}_{\widehat{f}^*_{W_N}}\left[\lambda^{c'_m(\mathbf{s},i-1)+c'_m(\mathbf{s},j-1)}\right] \tag{5.30}$$

$$\widehat{f}^*_{W_N}(x) = \begin{cases} \frac{N}{N-1}\,\mathrm{E}\left[f_{W_N}(x)\right] & x \neq 1 \\ 0 & x = 1 \end{cases} \tag{5.31}$$

The approximation depends only on the moments of the transformed approximate empirical spectral distribution and on the number of independent networks in the sequence. The unconditional expected Gram matrix of consensus error in the filter terms is then as follows.

$$\widehat{Q}_{\mathrm{err}} = \mathrm{E}_{\mathbf{s}}\left[\widehat{Q}_{\mathrm{err}}\left(\mathbf{s}\right)\right] \tag{5.32}$$

The matrix $\widehat{Q}_{\mathrm{err}}\left(\mathbf{s}\right)$ *is positive semidefinite for any switching sequence* $\mathbf{s}$ *and spectral distribution* $\widehat{f}^*_{W_N}$. *Thus, the matrix* $\widehat{Q}_{\mathrm{err}}$ *is also positive semidefinite.*

Proof

For the proof of this theorem, refer to the proof of Proposition 5.2 in Section 5.3. Proposition 5.2 provides the statement of the analogous result using the row-normalized Laplacian, which has non-orthogonal right-eigenvector bases. The proof of this theorem is nearly identical to that for Proposition 5.2. Note that the only difference is the use of the expected empirical spectral distribution in place of an approximation derived form Girko's methods. Additionally, for this case the basis change coordinates are inner products due to orthogonality. ■

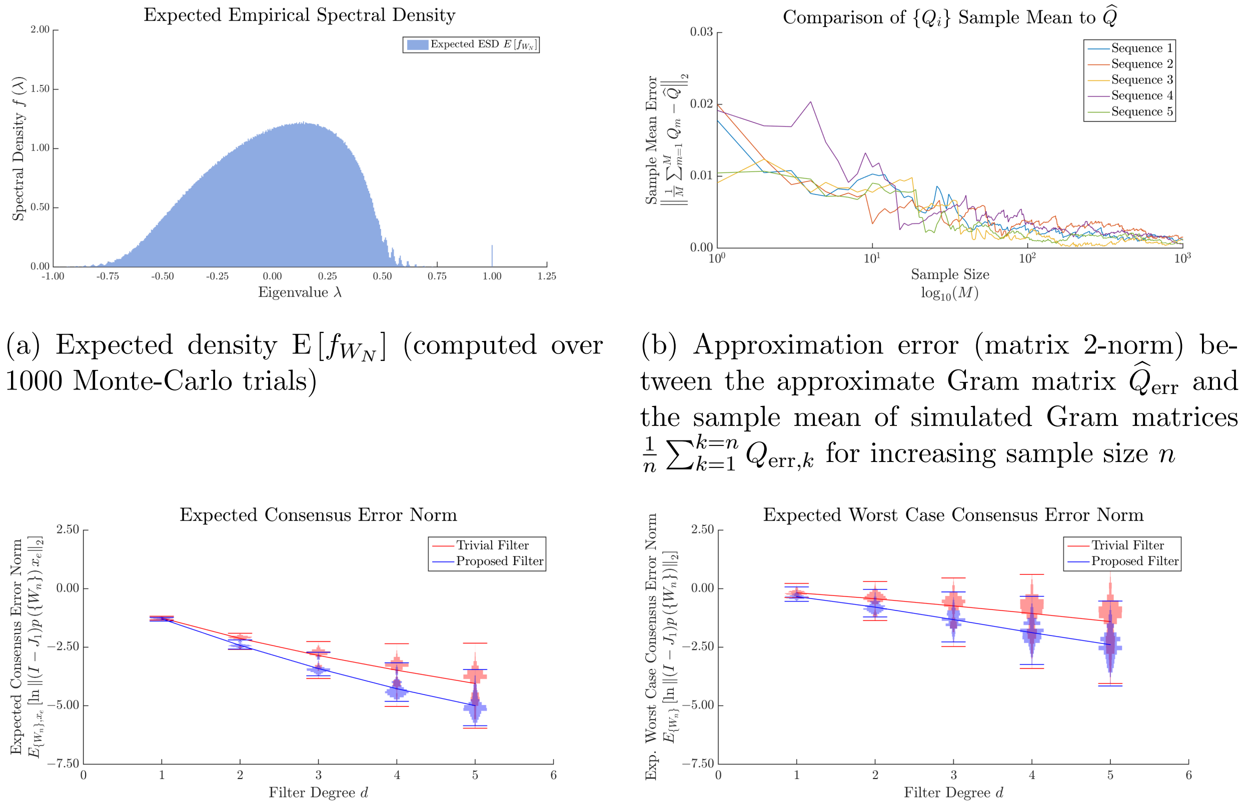
Substitution of the computed value of $\widehat{Q}\left(\mathbf{s}\right)$ for $\mathrm{E}_{\{W_n(\mathbf{s})\},\mathbf{v}}\left[Q\left(\mathbf{s}\right)|\mathbf{s}\right]$ in equation (5.14) results in the final form of the optimization problem.

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathbf{a}^\top \widehat{Q}_{\mathrm{err}}\, \mathbf{a} \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \end{aligned} \tag{5.33}$$

This formulation is a positive semidefinite linearly constrained quadratic program (LCQP). Because the expected norm squared is the sum of the squared expected norm and the variance of the norm, this formulation tends to reduce both mean and variance.

In order to evaluate the proposed Gram matrix approximation and filter design method, this section also provides simulation results. Figure 5.1 demonstrates the results on an Erdős-Rényi model. Figure 5.2 demonstrates the results on a stochastic block model. The precise details of each network model are listed in the figure captions. For each respective group of simulations, Figure 5.1a and Figure 5.2a shows the expected empirical spectral distributions. Figure 5.1b and Figure 5.2b show the approximation error (matrix 2-norm) between the sample mean $\frac{1}{n}\sum_{k=1}^{k=n} Q_{\mathrm{err},k}$ of simulated consensus error Gram matrices and the estimate for the expected consensus error Gram matrix $\widehat{Q}_{\mathrm{err}}$ over increasing sample size n. Figure 5.1c and Figure 5.2c show the expected consensus error norm achieved with the proposed filter against the results with no filter applied (simulated over 1000 Monte-Carlo trials). Figure 5.1d and Figure 5.2d show the expected consensus error for worst case input (expected consensus error matrix 2-norm) achieved with the proposed filtered system against the results with no filter applied (simulated over 1000 Monte-Carlo trials). To show the behavior of the filtering results as the switching probability is varied, Figure 5.3 repeats the results for an Erdős-Rényi network model at three different switching probabilities. Note that it is not possible to claim that this improves the asymptotic convergence rate, but only the worst case response (in expectation with respect to the network) over the first filtering window.

While the presented method represents an interesting result, it is subject to several limitations that must be acknowledged here. The choice of iteration matrix weights based on the row-normalized Laplacian, chosen to enable analysis using Girko's methods, produces a weighted average consensus. Furthermore, the method can only be applied to models for which the empirical spectral distribution can be approximated (or simulated as done in the preceding section). Furthermore, the assumptions of Proposition 5.2 must approximately hold, making it mostly applicable to random network models with a high degree of symmetry. As shown in Figure 5.3 the reduction in the expected consensus error norm provided by filters designed with the proposed method rapidly diminishes with increasing switching probability.



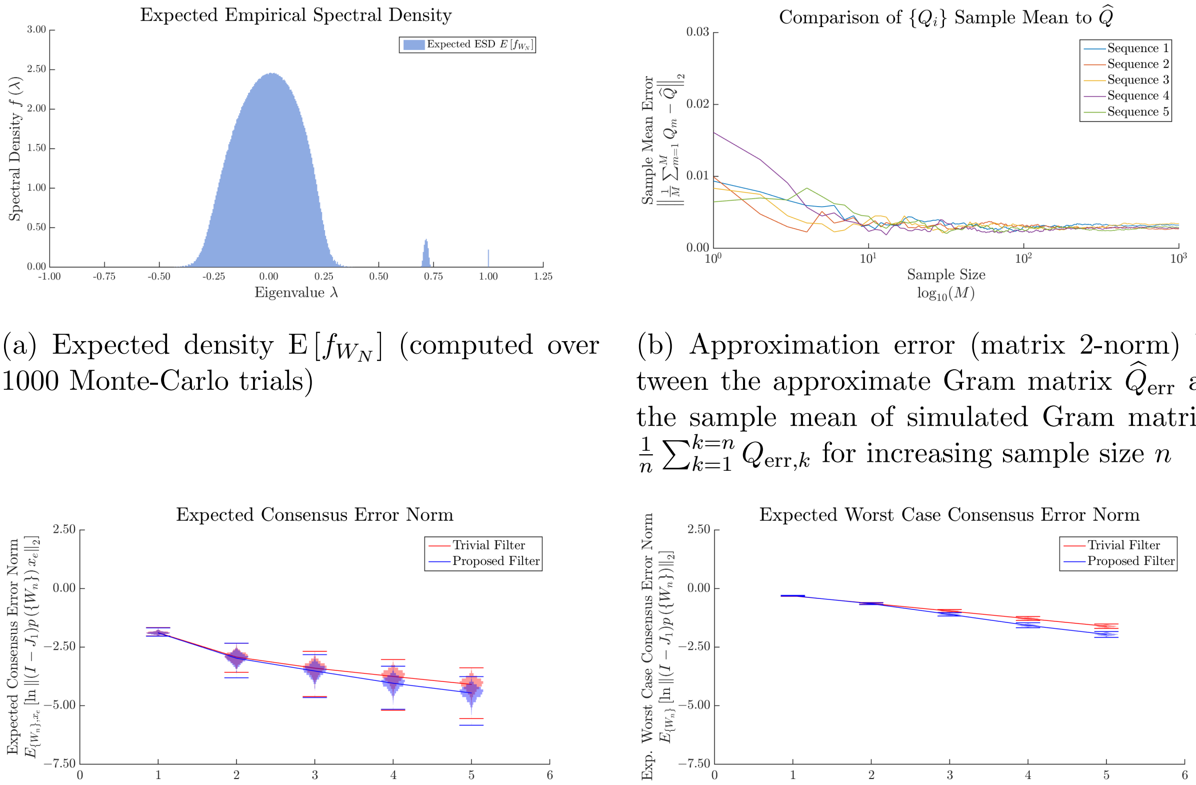
(a) Expected density $\mathrm{E}[f_{W_N}]$ (computed over 1000 Monte-Carlo trials)

(b) Approximation error (matrix 2-norm) between the approximate Gram matrix $\widehat{Q}_{\text{err}}$ and the sample mean of simulated Gram matrices $\frac{1}{n}\sum_{k=1}^{k=n} Q_{\text{err},k}$ for increasing sample size n

(c) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

(d) Expected worst case consensus error of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

Figure 5.1: The plots show results for an Erdős-Rényi model with $N = 1200$ nodes and percolation probability $\theta = 0.02$. The network switching probability is $\theta_{\text{sw}} = 0.10$. The iteration matrix $W = I - \alpha\mathcal{L}$ is used with parameter $\alpha = 1/\gamma_1$ where $\gamma_1 = \theta(N-1)$.



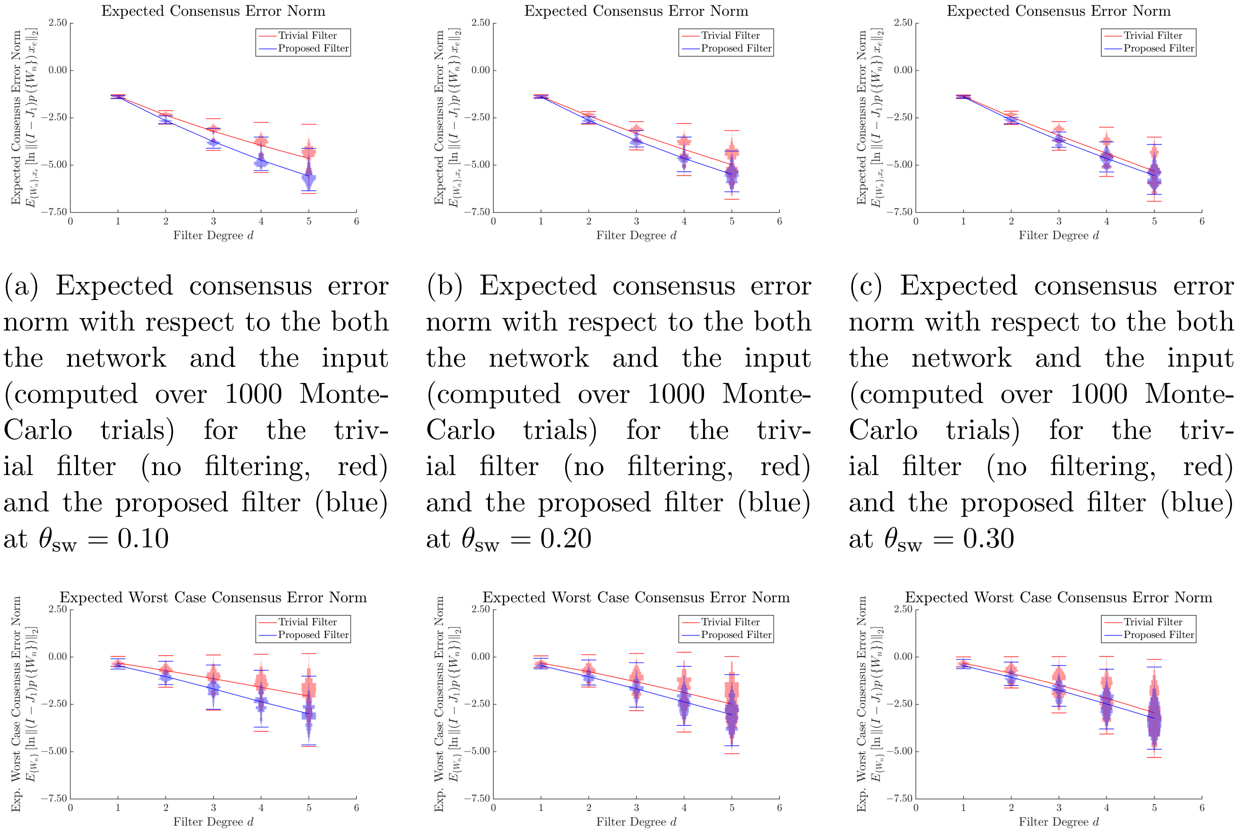
(a) Expected density $\mathrm{E}[f_{W_N}]$ (computed over 1000 Monte-Carlo trials)

(b) Approximation error (matrix 2-norm) between the approximate Gram matrix $\widehat{Q}_{\text{err}}$ and the sample mean of simulated Gram matrices $\frac{1}{n}\sum_{k=1}^{k=n} Q_{\text{err},k}$ for increasing sample size n

(c) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

(d) Expected worst case consensus error of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

Figure 5.2: The plots show results for a stochastic block model with $M = 10$ populations each of size $S = 100$ nodes and percolation probabilities $\theta_0 = 0.50$ within populations and $\theta_1 = 0.02$ between different populations. The network switching probability is $\theta_{\text{sw}} = 0.20$. The iteration matrix $W = I - \alpha\mathcal{L}$ is used with parameter $\alpha = 1/\gamma_1$ where $\gamma_1 = \theta_0(S-1) + \theta_1(M-1)S$.



(a) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\mathrm{sw}} = 0.10$

(b) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\mathrm{sw}} = 0.20$

(c) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\mathrm{sw}} = 0.30$

(d) Expected worst case consensus error of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\mathrm{sw}} = 0.10$

(e) Expected worst case consensus error of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\mathrm{sw}} = 0.20$

(f) Expected worst case consensus error of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\mathrm{sw}} = 0.30$

Figure 5.3: The plots show results for an Erdős-Rényi model with $N = 1000$ nodes and percolation probability $\theta = 0.03$. Three network switching probability values $\theta_{\mathrm{sw}} = \{0.10, 0.20, 0.30\}$ are tested to show the changing degree of benefit as the switching probability increases. The iteration matrix $W = I - \alpha\mathcal{L}$ is used with parameter $\alpha = 1/\gamma_1$ where $\gamma_1 = \theta(N - 1)$.

5.3 Switching, Undirected Random Networks: Asymmetric Iteration Matrix Case ($\boldsymbol{W}=\boldsymbol{I}-\boldsymbol{\alpha}\boldsymbol{\mathcal{L}_R}$)

This section proposes design criteria for consensus acceleration filters on random switching networks, a relatively simple class of time-varying network models, for certain large-scale random network distributions using the row-normalized Laplacian based consensus iteration matrix $W(\mathcal{G}) = I - \alpha\mathcal{L}_R(\mathcal{G})$. Each network is drawn from some specified random network distribution. At each time iteration, the network either remains constant or switches to a new, independent sample from the random network distribution according to a Bernoulli trial with fixed probability. The proposed quadratic optimization objective involves an approximation of the expected Gram matrix of error in the state vectors over a filtering window. The derived approximation depends only on the moments of the approximate empirical spectral distribution of the iteration matrices (obtainable via Girko's methods) and on the switching probability.

Consider distributed average consensus with respect to a time-varying sequence of iteration matrices $\{W(\mathcal{G}_{N,n})\}$ arising from a random switching network with N nodes and switching probability θ_{sw}. To compute a weighted mean of the initial data $\mathbf{x}_0$, the network implements a dynamic system with state $\mathbf{x}_n$ at time iteration n described by

$$\mathbf{x}_n = W(\mathcal{G}_{N,n})\,\mathbf{x}_{n-1} \tag{5.34}$$

Provided the iteration matrices satisfy the consensus conditions

$$\begin{gathered} W(\mathcal{G}_{N,n})\,\mathbf{1} = \mathbf{1},\ \boldsymbol{\ell}_n^\top W(\mathcal{G}_{N,n}) = \boldsymbol{\ell}_n^\top, \\ \rho\left(W(\mathcal{G}_{N,n}) - J_{\boldsymbol{\ell}_n}\right) < 1 \end{gathered} \tag{5.35}$$

where $J_{\boldsymbol{\ell}_n} = \mathbf{1}\boldsymbol{\ell}_n^\top / \boldsymbol{\ell}_n^\top\mathbf{1}$, the agreement error in the state vector is asymptotically eliminated. This section examines undirected graphs and employs the iteration matrix scheme

$$W_{N,n} = W(\mathcal{G}_{N,n}) = I - \alpha\mathcal{L}_R(\mathcal{G}_{N,n}), \tag{5.36}$$

which satisfies the properties in (5.35) when the network graphs $\{\mathcal{G}_{N,n}\}$ are connected and α is chosen suitably. To improve the convergence rate, a degree d filter with coefficients $\{a_k\}_{k=0}^{k=d}$ will be periodically applied to previous state values to update the current state vector according to the following equation.

$$\mathbf{x}_n := \sum_{k=0}^{k=d} a_k \mathbf{x}_{n-d+k}, \quad n \equiv 0 \pmod{d} \tag{5.37}$$

for $k = 0, \ldots, d$. Thus, for initial vector $\mathbf{x}_0$ the state vector terms used for filtering (written for the first filtering operation only to simplify notation) are given by $\mathbf{x}_k = \phi_k\left(\{W_{N,n}\}_{n=1}^{n=d}\right)\mathbf{x}_0$ where

$$\begin{gathered} \phi_k\left(\{W_{N,n}\}_{n=1}^{n=d}\right) = W_{N,k} \ldots W_{N,1} \\ \phi_0\left(\{W_{N,n}\}_{n=1}^{n=d}\right) = I_N \end{gathered} \tag{5.38}$$

Because each iteration matrix has eigenvalue $\lambda = 1$ corresponding to the consensus eigenvector $\mathbf{1}$, the filter coefficients must have unit sum to preserve this eigenvalue in the filtered transformation. Collecting the filter coefficients into a vector

$$\mathbf{a} = [a_0, \ldots, a_d] \tag{5.39}$$

this constraint can be expressed as $\mathbf{1}^\top \mathbf{a} = 1$.

Attempting to directly optimize the expected norm of the filter output error for the worst case input proves challenging. Instead of directly optimizing the convergence rate, the filters designed in this section approximately minimize the expected norm of the filter output consensus error vector with respect to the random iteration matrix sequence and with respect to the initial error vector. Let $\mathbf{x}_0 = \overline{\mathbf{x}}_0\mathbf{1} + \mathbf{x}_{\mathrm{e}}$ where $\mathbf{x}_{\mathrm{e}}$ is orthogonal to $\mathbf{1}$, and assume for simplicity that $\mathbf{x}_{\mathrm{e}}$ is uniformly distributed on unit norm vectors orthogonal to $\mathbf{1}$. By Jensen's inequality, the square root of the expected norm squared provides a lower bound for the expected norm. Thus, rather than minimizing the expected norm of the filter error directly, the expected norm squared will be minimized as follows, where $\mathbf{x}_{\mathrm{e}}$ is uniformly distributed on $\left\{\mathbf{x}_{\mathrm{e}} \in \mathbb{R}^N | \mathbf{x}_{\mathrm{e}} \perp \mathbf{1}, \|\mathbf{x}_{\mathrm{e}}\| = 1\right\}$.

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathrm{E}_{\mathbf{x}_{\mathrm{e}}, \left\{W_{N,n}\right\}_{n=1}^{n=d}} \left[\left\| (I - J_{\mathbf{1}}) \sum_{k=0}^{k=d} \mathbf{a}_k \phi_k \left(\{W_{N,n}\}_{n=1}^{n=d} \right) \mathbf{x}_{\mathrm{e}} \right\|_2^2 \right] \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \end{aligned} \tag{5.40}$$

The objective function in the above equation can be written in terms of the Gram matrix of consensus error vectors corresponding to each filtering term. The entries of this matrix are given by

$$\begin{aligned} (Q_{\mathrm{err}})_{ij} = \Big\langle & (I - J_{\mathbf{1}})\, \phi_{i-1} \left(\{W_{N,n}\}_{n=1}^{n=d} \right) \mathbf{x}_{\mathrm{e}}, \\ & (I - J_{\mathbf{1}})\, \phi_{j-1} \left(\{W_{N,n}\}_{n=1}^{n=d} \right) \mathbf{x}_{\mathrm{e}} \Big\rangle . \end{aligned} \tag{5.41}$$

Thus, the optimization problem can be rewritten in the following quadratic form.

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathbf{a}^\top \mathrm{E}_{\mathbf{x}_{\mathrm{e}}, \left\{W_{N,n}\right\}_{n=1}^{n=d}} \left[Q_{\mathrm{err}}\right] \mathbf{a} \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{a} = 1 \end{aligned} \tag{5.42}$$

Denote by $\mathbf{s}$ the network switching sequence where $\mathbf{s}_1 = 1$ and, for all $n > 1$, $\mathbf{s}_n = 1$ if $W_{N,n} = W_{N,n-1}$, and $\mathbf{s}_n = 0$ otherwise. That is, $\mathbf{s}_n$ determines whether $W_{N,n}$ is a new iteration matrix (within the filtering window), with $W_{N,1}$ always counted. The space of all possible switching sequences are the d-tuples

$$S_d = \{1\} \times \{0, 1\} \times \cdots \times \{0, 1\}, \tag{5.43}$$

and the total number of independent networks is

$$\#\mathbf{s} = \sum_{k=1}^{k=d} \mathbf{s}_k, \qquad \mathbf{s} \in S_d \tag{5.44}$$

with $\#\mathbf{s} - 1$ total switching events. For a switching process governed by independent Bernoulli trials with switching probability θ_{sw}, switching sequence $\mathbf{s} \in S_d$ has probability mass given by

$$f_S(\mathbf{s}) = (\theta_{\mathrm{sw}})^{(\#\mathbf{s}-1)} (1-\theta_{\mathrm{sw}})^{(d-1)-(\#\mathbf{s}-1)} \tag{5.45}$$

The distribution of the iteration matrix sequence $\{W_{N,n}\}_{n=1}^{n=d}$ can be factored into the distribution of the switching sequence $\mathbf{s}$ and distribution of $\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}$ given the switching sequence. The Gram matrix of filter term consensus error vectors can also be conditioned on the switching sequence as follows.

$$\begin{aligned}(Q_{\mathrm{err}}(\mathbf{s}))_{ij} = \Big\langle & (I - J_{\mathbf{1}})\, \phi_{i-1}\left(\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}\right) \mathbf{x}_{\mathrm{e}}, \\ & (I - J_{\mathbf{1}})\, \phi_{j-1}\left(\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}\right) \mathbf{x}_{\mathrm{e}} \Big\rangle\end{aligned} \tag{5.46}$$

In terms of the factored distributions, the optimization problem in (5.42) can be reformulated as follows.

$$\begin{aligned}\min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathbf{a}^\top \mathrm{E}_{\mathbf{s}}\left[\mathrm{E}_{\mathbf{x}_{\mathrm{e}},\left\{W_{N,n}(\mathbf{s})\right\}_{n=1}^{n=d}}\left[Q_{\mathrm{err}}(\mathbf{s})\,|\mathbf{s}\right]\right]\mathbf{a} \\ \text{s.t.} \quad & \mathbf{1}^\top\mathbf{a} = 1\end{aligned} \tag{5.47}$$

The above optimization problem could be used for filter design by computing the values of $\mathrm{E}_{\left\{W_{N,n}(\mathbf{s})\right\},\mathbf{v}}\left[Q(\mathbf{s})\,|\mathbf{s}\right]$ through simulation. However, the intent of this section is to connect information regarding the empirical spectral distribution of the iteration matrices to filter design for large-scale random switching networks, as done for constant random networks in Chapter 4. Therefore, an analytical approximation (under suitable conditions) based on a deterministic approximation of the empirical spectral distribution will be described below.

Before proceeding, some notation must first be introduced. Note that each switching sequence $\mathbf{s}$ corresponds to an integer composition of d with $\#\mathbf{s}$ partitions, namely

$$c(\mathbf{s}) = (c_1(\mathbf{s}), \ldots, c_{\#\mathbf{s}}(\mathbf{s})) \tag{5.48}$$

where each $c_m(\mathbf{s})$ is the number of iterations the mth network is used before switching. Let $c'_m(\mathbf{s}, n)$ be the number of iterations the mth network is used up to (and including) iteration

$n \geq 1$ with $c'_m(\mathbf{s}, 0) = 0$. More explicitly,

$$c'_m(\mathbf{s}) = \begin{cases} 0 & n < \sum_{k=1}^{m-1} c_k(\mathbf{s}) \\ c_m(\mathbf{s}) & n > \sum_{k=1}^{m} c_k(\mathbf{s}) \\ n - \sum_{k=1}^{m-1} c_k(\mathbf{s}) & \text{otherwise} \end{cases} \tag{5.49}$$

Let $\{\mathbf{u}_{n,k}(\mathbf{s})\}_{k=1}^{k=N-1} \cup \{\mathbf{u}_{n,N}(\mathbf{s}) = \mathbf{1}/\sqrt{N}\}$ be a basis of unit norm (not necessarily orthogonal) right-eigenvectors for $W_{N,n}(\mathbf{s})$ with $\{\mathbf{v}_k\}_{k=1}^{k=N-1} \cup \{\mathbf{v}_N = \mathbf{1}/\sqrt{N}\}$ an orthonormal basis of eigenvectors for $I - J_{\mathbf{1}}$. Note that if the network does not switch, the basis of eigenvectors remains the same. This motivates redefining some terms with respect to the number of independent networks to avoid duplication. Let $W'_{N,m}(\mathbf{s})$ be the mth independently drawn network for $m = 1, \ldots, \#\mathbf{s}$. Let the corresponding bases of unit norm (not necessarily orthogonal) right-eigenvectors be $\left\{\mathbf{u}'_{m,k}(\mathbf{s})\right\}_{k=1}^{k=N-1} \cup \left\{\mathbf{u}'_{m,N}(\mathbf{s}) = \mathbf{1}/\sqrt{N}\right\}$. Let the basis change coordinates for this sequence of eigenvector bases be $\psi'_{(m,m+1),r_m r_{m+1}}$, which gives the coordinate of $\mathbf{u}'_{m,r_m}$ corresponding to $\mathbf{u}'_{m+1,r_{m+1}}$ for $1 \leq m < \#\mathbf{s}$. Furthermore, let the coordinate of $\mathbf{x}_e$ in $\mathbf{u}'_{1,r_1}$ be $\psi'_{(0,1),r_1}$. Finally, let the coordinate of $\mathbf{v}_\ell$ in $\mathbf{u}'_{\#\mathbf{s},r_{\#\mathbf{s}}}$ be $\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}\ell}$.

These basis change coordinates are random (except for the ones corresponding to the consensus eigenvector, which remains constant), and assumptions must be made regarding their nature to enable computation. The results will vary in quality depending on how well these simplifying assumptions hold. Assume that the basis change coordinates have the following properties (divided into four sets). While it may look like there are many cases, they are generated from simple intuitions.

Assumption 1: First, note that the consensus eigenvector is always present and does not change. This will give the following conditions

$$\psi'_{(0,1),N} = 0 \tag{5.50}$$

$$\psi'_{(m,m+1),NN} = 1, \quad m = 1, \ldots, \#\mathbf{s} - 1 \tag{5.51}$$

$$\psi'_{(m,m+1),Nr_{m+1}} = 0, \quad m = 1, \ldots, \#\mathbf{s} - 1,\ r_{m+1} \neq N \tag{5.52}$$

$$\psi'_{(m,m+1),r_m N} = 0, \quad m = 1, \ldots, \#\mathbf{s} - 1,\ r_m \neq N \tag{5.53}$$

$$\psi'_{(\#\mathbf{s},\#\mathbf{s}+1) r_{\#\mathbf{s}} N} = 1, \quad r_{\#\mathbf{s}} = N \tag{5.54}$$

$$\psi'_{(\#\mathbf{s},\#\mathbf{s}+1) r_{\#\mathbf{s}} N} = 0, \quad r_{\#\mathbf{s}} \neq N \tag{5.55}$$

$$\psi'_{(\#\mathbf{s},\#\mathbf{s}+1) N\ell} = 0, \quad \ell \neq N \tag{5.56}$$

The remaining assumptions deal with the coordinates that do not correspond to the consensus eigenvector.

Assumption 2: Second, assume that basis change coordinates (other than the ones that remain constant) for different switching events are statistically independent.

Assumption 3: Third, assume that pairs of basis change coordinates for the same switching event are uncorrelated.

$$\mathrm{E}\left[\psi'_{(0,1)r_1}\psi'_{(0,1)t_1}\right] = 0, \quad r_1 \neq t_1,\ r_1, t_1 \neq N \tag{5.57}$$

$$\mathrm{E}\left[\psi'_{(m,m+1)r_m r_{m+1}}\psi'_{(m,m+1)t_m t_{m+1}}\right] = 0, \quad \begin{array}{l} r_m \neq t_m \text{ and/or } r_{m+1} \neq t_{m+1}, \\ m = 1, \ldots, \#\mathbf{s} - 1\ ,\ r_m, r_{m+1}, t_m, t_{m+1} \neq N \end{array} \tag{5.58}$$

$$\mathrm{E}\left[\psi'_{(\#\mathbf{s},\#\mathbf{s}+1) r_{\#\mathbf{s}}\ell_1}\psi'_{(\#\mathbf{s},\#\mathbf{s}+1) t_{\#\mathbf{s}}\ell_2}\right] = 0, \quad \begin{array}{l} r_{\#\mathbf{s}} \neq t_{\#\mathbf{s}} \text{ and/or } \ell_1 \neq \ell_2, \\ r_{\#\mathbf{s}}, t_{\#\mathbf{s}}, \ell_1, \ell_2 \neq N \end{array} \tag{5.59}$$

Assumption 4: Fourth, assume the following second moments for each basis change coordinate. This is motivated by assumption of even distribution of the signal energy if the eigenvectors were orthogonal.

$$\mathrm{E}\left[\left(\psi'_{(0,1)r_1}\right)^2\right] = \frac{1}{N-1}, \quad r_1 \neq N \tag{5.60}$$

$$\mathrm{E}\left[\left(\psi'_{(m,m+1)r_m r_{m+1}}\right)^2\right] = \frac{1}{N-1}, \quad r_m, r_{m+1} \neq N,\ m = 1, \ldots \#\mathbf{s} - 1 \tag{5.61}$$

$$\mathrm{E}\left[\left(\psi'_{(\#\mathbf{s},\#\mathbf{s}+1) r_{\#\mathbf{s}}\ell}\right)^2\right] = \frac{1}{N-1}, \quad r_{\#\mathbf{s}}, \ell \neq N \tag{5.62}$$

Then $\mathrm{E}_{\{W_n(\mathbf{s})\},\mathbf{v}}\left[Q\left(\mathbf{s}\right)|\mathbf{s}\right]$ can be approximated as follows.

Proposition 5.2 (Approximate Gram Matrix)

Given a random switching network that satisfies the listed assumptions, an approximate empirical spectral density $\widehat{f}_{W_N}$, and small constant κ used to isolate the eigenvalue $\lambda = 1$ so it can be removed, the following expression approximates the entries of the expected Gram matrix of consensus error in the filter terms conditioned on the switching sequence.

$$\left(\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)\right)_{ij} = \prod_{m=1}^{m=\#\mathbf{s}} \mathrm{E}_{\widehat{f}^{*}_{W_N,\kappa}}\left[\lambda^{c'{}_m(\mathbf{s},i-1)+c'{}_m(\mathbf{s},j-1)}\right] \tag{5.63}$$

$$\widehat{f}^{*}_{W_N,\kappa}(x) = \begin{cases} \frac{N}{N-1}\widehat{f}_{W_N}(x) & x < 1-\kappa \\ 0 & x > 1-\kappa \end{cases} \tag{5.64}$$

The approximation depends only on the moments of the transformed approximate empirical spectral distribution and on the number of independent networks in the sequence. The unconditional expected Gram matrix of consensus error in the filter terms is then as follows.

$$\widehat{Q}_{\text{err}} = \mathrm{E}_{\mathbf{s}}\left[\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)\right] \tag{5.65}$$

The matrix $\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)$ is positive semidefinite for any switching sequence $\mathbf{s}$ and spectral distribution $\widehat{f}^{*}_{W_N,\kappa}$. Thus, the matrix $\widehat{Q}_{\text{err}}$ is also positive semidefinite.

Proof

In order to prove the result, first express $\left(I - J_{\mathbf{1}}\right)\phi_k\left(\left\{W_{N,n}\left(\mathbf{s}\right)\right\}_{n=1}^{n=d}\right)\mathbf{x}_e$ in terms of the sequence of bases for each independent matrix given the switching sequence. That is, express $\left(I - J_{\mathbf{1}}\right)\phi_k\left(\left\{W_{N,n}\left(\mathbf{s}\right)\right\}_{n=1}^{n=d}\right)\mathbf{x}_e$ in terms of the normalized (but not necessarily orthogonal) right-eigenvector bases $\left\{\left\{\mathbf{u}'_{m,r_m}\left(\mathbf{s}\right)\right\}_{r_m=1}^{r_m=N}\right\}_{m=1}^{m=\#\mathbf{s}}$ for $\left\{W'_{N,m}\right\}_{m=1}^{m=\#\mathbf{s}}$ and the orthonormal eigenvector basis $\left\{\mathbf{v}_\ell\left(\mathbf{s}\right)\right\}_{\ell=1}^{\ell=N}$ for $I - J_{\mathbf{1}}$ using the basis change coordinates $\psi'_{(m,m+1),r_m r_{m+1}}$ for $m = 1, \ldots \#\mathbf{s} - 1$ and $r_m, r_{m+1} = 1, \ldots, N$. (Recall that $\left\{\psi'_{(0,1),r_1}\right\}$ are the coordinates of $\mathbf{x}_e$ in $\left\{\mathbf{u}'_{1,r_1}\left(\mathbf{s}\right)\right\}_{r_1=1}^{r_1=N}$. Additionally, recall that $\left\{\psi'_{(\#\mathbf{s},\#\mathbf{s}+1),r_{\#\mathbf{s}}\ell}\right\}$ are the coordinates of the vectors $\left\{\mathbf{u}'_{\#\mathbf{s},r_{\#\mathbf{s}}}\right\}$ in $\left\{\mathbf{v}_\ell\right\}_{\ell=1}^{\ell=N}$.) Note that the mth independent matrix is used for the network iteration

$c'_m(\mathbf{s},k)$ times before switching up to (and including) the kth iteration, implying its eigenvalues are raised to exponent $c'_m(\mathbf{s},k)$ in that portion of the product. Thus, the following expression emerges. Note that the first and last sums range from 1 to $N-1$ because $\mathbf{x}_e$ is orthogonal to $\mathbf{u}_{1,N} = \mathbf{1}/\sqrt{N}$ by hypothesis and because $(I - J_{\mathbf{1}})$ annihilates $\mathbf{v}_N$.

$$
\begin{aligned}
&(I - J_{\mathbf{1}})\,\phi_k\left(\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}\right)\mathbf{x}_e = \\
&= \sum_{r_1=1}^{r_1=N-1} \psi'_{(0,1),r_1}\left[\lambda_{r_1}\left(W'_{N,1}\right)\right]^{c'_1(\mathbf{s},k)} \\
&\qquad \cdots \left(\sum_{r_2=1}^{r_2=N} \psi'_{(1,2),r_1 r_2}\left[\lambda_{r_1}\left(W'_{N,2}\right)\right]^{c'_2(\mathbf{s},k)} \right. \\
&\qquad\qquad\qquad\qquad \cdots \\
&\qquad\qquad \cdots \left(\sum_{r_{\#\mathbf{s}}=1}^{r_{\#\mathbf{s}}=N} \psi'_{(\#\mathbf{s}-1,\#\mathbf{s}),r_{\#\mathbf{s}-1}r_{\#\mathbf{s}}}\left[\lambda_{r_{\#\mathbf{s}}}\left(W'_{N,\#\mathbf{s}}\right)\right]^{c'_{\#\mathbf{s}}(\mathbf{s},k)} \right. \\
&\qquad\qquad\qquad \left. \left. \cdots \left(\sum_{\ell=1}^{\ell=N-1} \psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}\ell}\mathbf{v}_\ell \right) \right) \cdots \right) \\
&= \sum_{r_1=1}^{r_1=N-1} \sum_{r_2=1}^{r_2=N} \cdots \sum_{r_{\#\mathbf{s}}=1}^{r_{\#\mathbf{s}}=N} \left(\prod_{m=1}^{m=\#\mathbf{s}} \psi'_{(m-1,m),r_{m-1}r_m}\left[\lambda_{r_m}\left(W'_{N,m}\right)\right]^{c'_m(\mathbf{s},k)} \right) \cdots \\
&\qquad\qquad\qquad\qquad\qquad\qquad \cdots \sum_{\ell=1}^{\ell=N-1} \psi'_{(\#\mathbf{s},\#\mathbf{s}+1),r_{\#\mathbf{s}}\ell}\mathbf{v}_\ell
\end{aligned}
\tag{5.66}
$$

Next, approximate the value of $\mathrm{E}_{\mathbf{x}_e,\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}}\left[(Q_{\text{err}}(\mathbf{s}))_{ij}\middle|\mathbf{s}\right]$ through computing the expected inner product

$$
\mathrm{E}\left[\left\langle (I - J_{\mathbf{1}})\,\phi_{i-1}\left(\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}\right)\mathbf{x}_e, (I - J_{\mathbf{1}})\,\phi_{j-1}\left(\{W_{N,n}(\mathbf{s})\}_{n=1}^{n=d}\right)\mathbf{x}_e \right\rangle\right] \tag{5.67}
$$

using the stated assumptions regarding the basis change coordinates. (Note: Several simplifying steps will be made at once due to the size of the equations.) First, observe that because the basis $\{\mathbf{v}_\ell\}$ is orthonormal, only like terms with respect to ℓ need be considered in the inner product. By independence of the basis change coordinates at different time iterations, expectations of like terms with respect to time can be factored from other time iterations. Finally, by the assumption that $\mathrm{E}\left[\psi'_{(m,m+1)k\ell}\psi'_{(m,m+1)rt}\right] = 0$ for different basis change coordinates ($k \neq r$ or $r \neq t$) at the same time iteration, many terms can be removed. The remaining terms in the approximation are as follows.

$$\begin{aligned}\left(\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)\right)_{ij} = \sum_{r_1=1}^{r_1=N-1}\sum_{r_2=1}^{r_2=N}\cdots\sum_{r_{\#\mathbf{s}}=1}^{r_{\#\mathbf{s}}=N}\Bigg(\prod_{m=1}^{m=\#\mathbf{s}}\mathrm{E}\left[\left(\psi'_{(m-1,m),r_{m-1}r_m}\right)^2\right]\cdots\\ \cdots\left[\lambda_{r_m}\left(W'_{N,m}\right)\right]^{c'_m(\mathbf{s},i-1)}\left[\lambda_{r_m}\left(W'_{N,m}\right)\right]^{c'_m(\mathbf{s},j-1)}\Bigg)\sum_{\ell=1}^{\ell=N-1}\mathrm{E}\left[\left(\psi'_{(\#\mathbf{s},\#\mathbf{s}+1),r_{\#\mathbf{s}}\ell}\right)^2\right]\end{aligned} \tag{5.68}$$

Applying the assumptions that $\psi'_{(m,m+1)r_mN} = 0$ for $r_m < N$, that $\mathrm{E}\left[\left(\psi'_{(m,m+1)r_mr_{m+1}}\right)^2\right] = \frac{1}{N-1}$ for $m < \#\mathbf{s}$ and $r_m, r_{m+1} < N$, and that $\mathrm{E}\left[\left(\psi'_{(\#\mathbf{s},\#\mathbf{s}+1)r_{\#\mathbf{s}}\ell}\right)^2\right] = \frac{1}{N-1}$ for all $r_{\#\mathbf{s}}, \ell < N$, the following is derived. (Note: Because of the first assumption for this step, some of the upper ranges of the sum could be reduced.)

$$\begin{aligned}\left(\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)\right)_{ij} = \sum_{r_1=1}^{r_1=N-1}\sum_{r_2=1}^{r_2=N-1}\cdots\sum_{r_{\#\mathbf{s}}=1}^{r_{\#\mathbf{s}}=N-1}\Bigg(\prod_{m=1}^{m=\#\mathbf{s}}\frac{1}{N-1}\cdots\\ \cdots\left[\lambda_{r_m}\left(W'_{N,m}\right)\right]^{c'_m(\mathbf{s},i-1)+c'_m(\mathbf{s},j-1)}\Bigg)\sum_{\ell=1}^{\ell=N-1}\frac{1}{N-1}\end{aligned} \tag{5.69}$$

The last term simply sums to one. Finally, observe that within this equation, moments of the empirical spectral distribution (modified to exclude $\lambda = 1$) are hidden. These moments with respect to the true empirical spectral distribution will be approximated by the moments of the approximate empirical spectral distribution $\widehat{f}_{W_N}$ (after suitably modified). Thus, the following expression

$$\left(\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)\right)_{ij} = \prod_{m=1}^{m=\#\mathbf{s}}\mathrm{E}_{\widehat{f}^*_{W_N,\kappa,m}}\left[\left(\lambda\left(W'_{N,m}\right)\right)^{c'_m(\mathbf{s},i-1)+c'_m(\mathbf{s},j-1)}\right] \tag{5.70}$$

is derived where

$$\widehat{f}^*_{W_N,\kappa,m}(x) = \begin{cases}\frac{N}{N-1}\widehat{f}_{W_N,m}(x) & x < 1-\kappa\\ 0 & x > 1-\kappa\end{cases} \tag{5.71}$$

and κ is a small constant to exclude the $1/N$ of the density near $\lambda = 1$. (Assume it excludes exactly $1/N$ of the density such that the above is a valid probability density function.) Finally, noting that each of the matrices are identically distributed, the final form

$$\left(\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)\right)_{ij} = \prod_{m=1}^{m=\#\mathbf{s}} \mathrm{E}_{\widehat{f}^{*}_{W_N,\kappa}} \left[\lambda^{c'_m(\mathbf{s},i-1)+c'_m(\mathbf{s},j-1)}\right] \tag{5.72}$$

is derived where

$$\widehat{f}^{*}_{W_N,\kappa}(x) = \begin{cases} \frac{N}{N-1}\widehat{f}_{W_N}(x) & x < 1-\kappa \\ 0 & x > 1-\kappa \end{cases} \tag{5.73}$$

and κ is as small constant as before. Thus the result is derived given the approximating assumptions.

It remains to be proven that each matrix $\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)$ is positive semidefinite. This will be accomplished by showing it is an expected Gram matrix (not just an approximation of one). Let $\lambda_1, \ldots, \lambda_{\#\mathbf{s}}$ be independent, identically distributed random variables with distribution $\widehat{f}^{*}_{W_N,\kappa}$. And consider the column vector $\boldsymbol{\lambda}$ with

$$\boldsymbol{\lambda}_i = \prod_{m=1}^{m=\#\mathbf{s}} \lambda_m^{c'_m(\mathbf{s},i-1)}. \tag{5.74}$$

Note that

$$\widehat{Q}_{\text{err}}\left(\mathbf{s}\right) = \mathrm{E}\left[\boldsymbol{\lambda}^{\top}\boldsymbol{\lambda}\right] \tag{5.75}$$

making it an expectation of a Gram matrix and, therefore, positive semidefinite. Consequently, $\widehat{Q}_{\text{err}} = \mathrm{E}_{\mathbf{s}}\left[\widehat{Q}_{\text{err}}\left(\mathbf{s}\right)\right]$ is also positive semidefinite. ■

Substitution of the computed value of $\widehat{Q}\left(\mathbf{s}\right)$ for $\mathrm{E}_{\{W_n(\mathbf{s})\},\mathbf{v}}\left[Q\left(\mathbf{s}\right)|\mathbf{s}\right]$ in equation (5.47) results in the final form of the optimization problem.

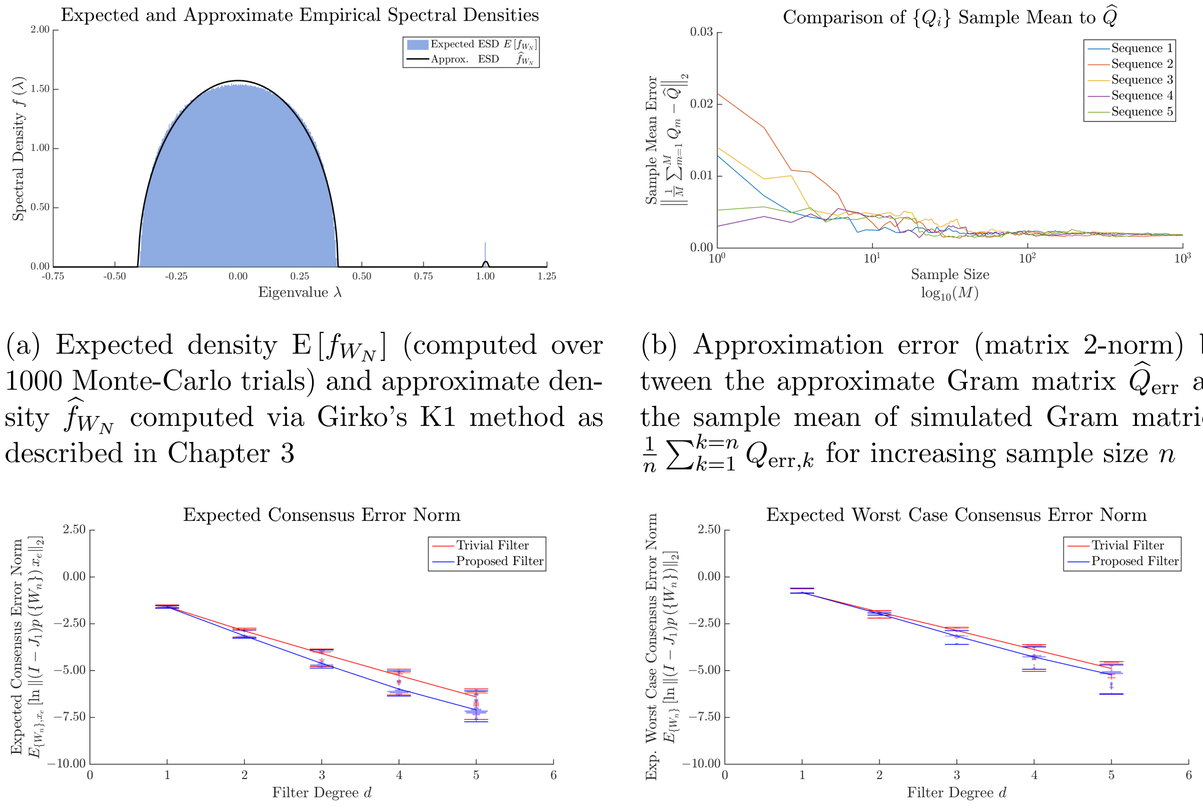
$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad & \mathbf{a}^{\top}\widehat{Q}_{\text{err}}\,\mathbf{a} \\ \text{s.t.} \quad & \mathbf{1}^{\top}\mathbf{a} = 1 \end{aligned} \tag{5.76}$$

This formulation is a positive semidefinite linearly constrained quadratic program (LCQP).

Because the expected norm squared is the sum of the squared expected norm and the variance of the norm, this formulation tends to reduce both mean and variance.

In order to evaluate the proposed Gram matrix approximation and filter design method, this section also provides simulation results. Figure 5.4 demonstrates the results on an Erdős-Rényi model. Figure 5.5 demonstrates the results on a stochastic block model. The precise details of each network model are listed in the figure captions. For each respective group of simulations, Figure 5.4a and Figure 5.5a shows the expected and approximate empirical spectral distributions. Figure 5.4b and Figure 5.5b show the approximation error (matrix 2-norm) between the sample mean $\frac{1}{n}\sum_{k=1}^{k=n} Q_{\mathrm{err},k}$ of simulated consensus error Gram matrices and the estimate for the expected consensus error Gram matrix $\widehat{Q}_{\mathrm{err}}$ over increasing sample size n. Figure 5.4c and Figure 5.5c show the expected consensus error norm achieved with the proposed filter against the results with no filter applied (simulated over 1000 Monte-Carlo trials). Figure 5.4d and Figure 5.5d show the expected consensus error for worst case input (expected consensus error matrix 2-norm) achieved with the proposed filtered system against the results with no filter applied (simulated over 1000 Monte-Carlo trials). To show the behavior of the filtering results as the switching probability is varied, Figure 5.6 repeats the results for an Erdős-Rényi network model at three different switching probabilities. Note that it is not possible to claim that this improves the asymptotic convergence rate, but only the worst case response (in expectation with respect to the network) over the first filtering window.

While the presented method represents an interesting result, it is subject to several limitations that must be acknowledged here. The choice of iteration matrix weights based on the row-normalized Laplacian, chosen to enable analysis using Girko's methods, produces a weighted average consensus. Furthermore, the method can only be applied to models for which the empirical spectral distribution can be approximated (or simulated as done in the preceding section). Furthermore, the assumptions of Proposition 5.2 must approximately hold, making it mostly applicable to random network models with a high degree of symmetry. As shown in Figure 5.6 the reduction in the expected consensus error norm provided by filters designed with the proposed method rapidly diminishes with increasing switching probability.



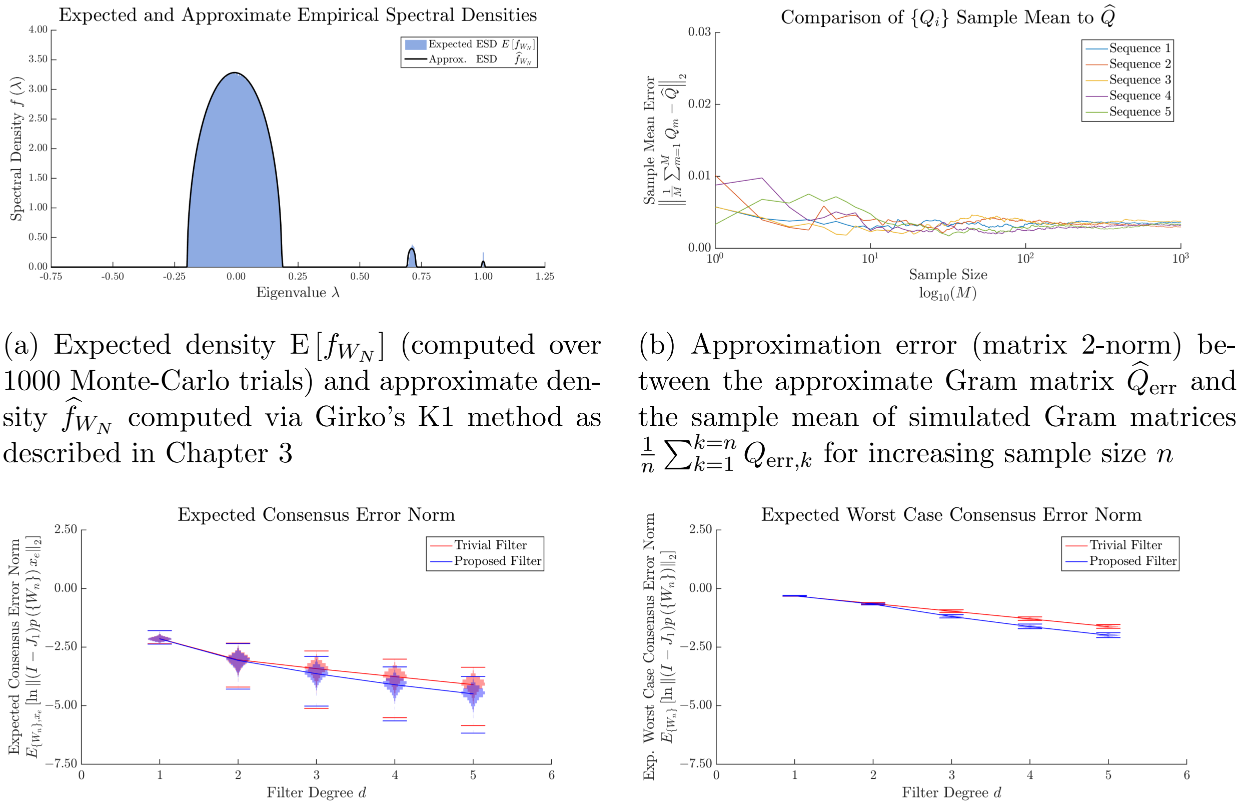
(a) Expected density $\mathrm{E}\left[f_{W_N}\right]$ (computed over 1000 Monte-Carlo trials) and approximate density $\widehat{f}_{W_N}$ computed via Girko's K1 method as described in Chapter 3

(b) Approximation error (matrix 2-norm) between the approximate Gram matrix $\widehat{Q}_{\text{err}}$ and the sample mean of simulated Gram matrices $\frac{1}{n}\sum_{k=1}^{k=n} Q_{\text{err},k}$ for increasing sample size n

(c) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

(d) Expected worst case consensus error norm of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

Figure 5.4: The plots show results for an Erdős-Rényi model with $N = 1200$ nodes and percolation probability $\theta = 0.02$. The network switching probability is $\theta_{\text{sw}} = 0.10$. The iteration matrix $W = I - \alpha\mathcal{L}_R$ is used with parameter $\alpha = 1$.



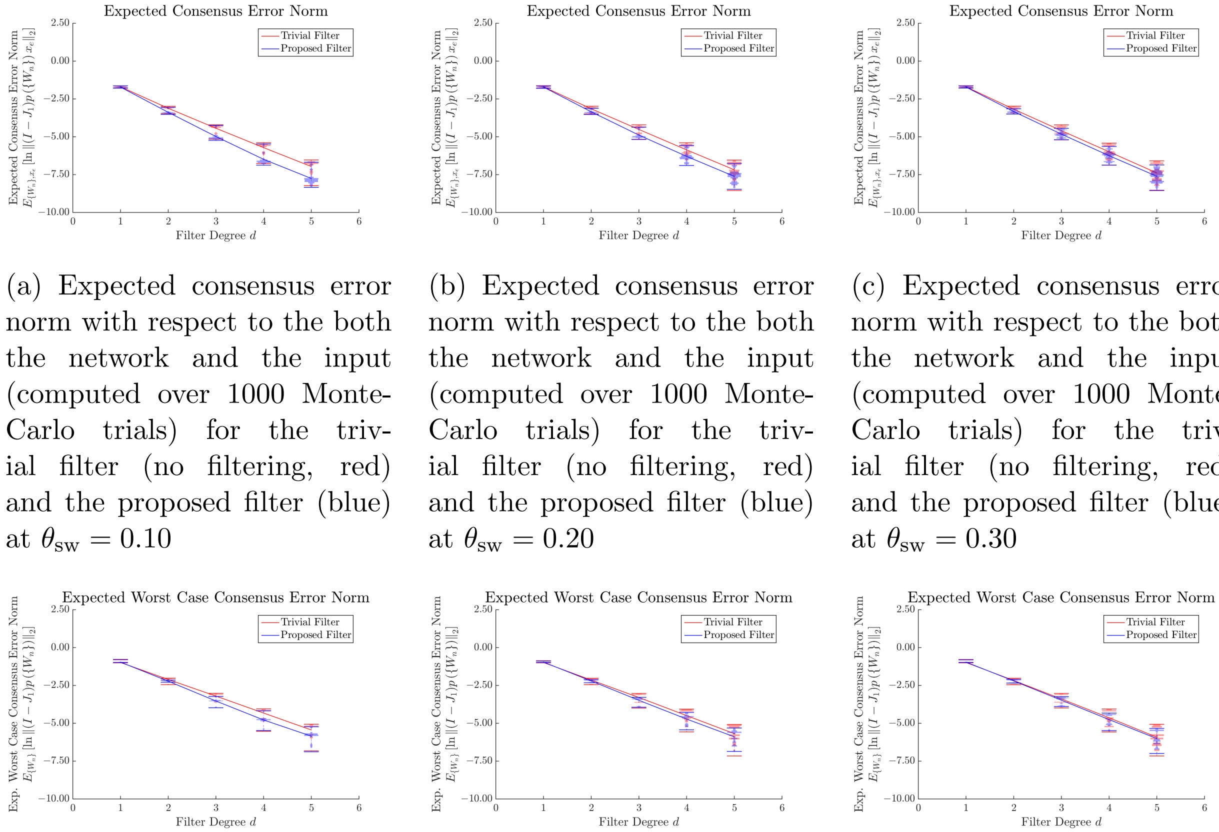
(a) Expected density $\mathrm{E}\left[f_{W_N}\right]$ (computed over 1000 Monte-Carlo trials) and approximate density $\widehat{f}_{W_N}$ computed via Girko's K1 method as described in Chapter 3

(b) Approximation error (matrix 2-norm) between the approximate Gram matrix $\widehat{Q}_{\text{err}}$ and the sample mean of simulated Gram matrices $\frac{1}{n}\sum_{k=1}^{k=n} Q_{\text{err},k}$ for increasing sample size n

(c) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

(d) Expected worst case consensus error norm of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue)

Figure 5.5: The plots show results for a stochastic block model with $M = 10$ populations each of size $S = 100$ nodes and percolation probabilities $\theta_0 = 0.50$ within populations and $\theta_1 = 0.02$ between different populations. The network switching probability is $\theta_{\text{sw}} = 0.20$. The iteration matrix $W = I - \alpha\mathcal{L}_R$ is used with parameter $\alpha = 1$.



(a) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\text{sw}} = 0.10$

(b) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\text{sw}} = 0.20$

(c) Expected consensus error norm with respect to the both the network and the input (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\text{sw}} = 0.30$

(d) Expected worst case consensus error norm of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\text{sw}} = 0.10$

(e) Expected worst case consensus error norm of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\text{sw}} = 0.20$

(f) Expected worst case consensus error norm of the filtered transform with respect to the network (computed over 1000 Monte-Carlo trials) for the trivial filter (no filtering, red) and the proposed filter (blue) at $\theta_{\text{sw}} = 0.30$

Figure 5.6: The plots show results for an Erdős-Rényi model with $N = 1000$ nodes and percolation probability $\theta = 0.03$. Three network switching probability values $\theta_{\text{sw}} = \{0.10, 0.20, 0.30\}$ are tested to show the changing degree of benefit as the switching probability increases. The iteration matrix $W = I - \alpha\mathcal{L}_R$ is used with parameter $\alpha = 1$.

5.4 Summary

In summary, this chapter presented filter design methods for minimizing the expected consensus error norm squared for large-scale random switching networks, a particular time-varying random network model. This problem formulation leads to an optimization problem in the form of a quadratic program where the positive semidefinite matrix in the objective function is the expected Gram matrix of consensus error in each filtering term. The expected Gram matrix can be approximated by conditioning the expectation on the switching sequence and making some simplified. The expected Gram matrix given each switching sequence can be approximated in terms of the consensus iteration matrix empirical spectral distribution moments, as shown by Proposition 5.1 and Proposition 5.2. This was first accomplished for Laplacian-based consensus iteration matrices of the form $W(\mathcal{G}) = I - \alpha\mathcal{L}(\mathcal{G})$ using the expected empirical spectral distribution as computed through simulation. Subsequently, this was extended to row-normalized Laplacian-based consensus iteration matrices of the form $W(\mathcal{G}) = I - \alpha\mathcal{L}_R(\mathcal{G})$ using the approximate expected empirical spectral distributions computed through Girko's methods (in Section 3.2 or Section 3.3 of Chapter 3). Simulation results support the proposed filter optimization method for both cases. The filters designed using the proposed method tend to reduce both the mean and the variance of the consensus error norm. While consistently observed in the simulation, the reductions in the mean error norm compared to the results with no filtering quickly diminish with increasing switching probability, limiting the applicability of this method. It is possible that models with greater correlation between adjacent random matrices in the sequence would have greater alignment between sequential eigenvector bases and thus admit more substantial filtering results. However, these models are much more difficult to analyze and are, therefore, reserved for future work along with extension to directed models.

Chapter Six

Conclusion

6.1 Thesis Summary

The uncertain conditions of random networks represent a significant challenge for graph signal processing applications. Graph signal processing techniques rely on the eigendecomposition of graph shift matrices that conform to the graph structure. For instance, the graph shift matrix eigenvalues define the domain for the filter response of shift-invariant graph filters. Consequently, graph filter design methods that operate with respect to the random network distribution must account for the induced spectral uncertainty. However, statistical knowledge regarding the eigendecomposition is difficult to obtain in a stochastic setting for most random network distributions.

This thesis leverages the predictability that emerges from large-scale random networks for suitable network models to enable such analysis. Specifically, the empirical distribution built from the eigenvalues of a large-scale random matrix may admit a deterministic limit obtainable through methods from random matrix theory literature. For graph shift matrices that conform to the random network structure, random matrix theory methods that accommodate non-identically distributed matrix entries and deterministic explicitly zero matrix entries provide the most relevance. In particular, methods published by Girko [9] provide results suitable for analysis of the adjacency matrix spectral asymptotics for random link-percolation networks subject to mild regularity conditions. Several such results are available covering cases relevant to both directed networks and undirected networks.

After Chapter 2 detailing basic background concepts and definitions used throughout the thesis, Chapter 3 introduced and discussed these spectral asymptotics theorems, focusing on three specific results: Girko's K1 method, Girko's K27 method, and Girko's K25 method.

Girko's K1 method handles symmetric random matrices with independent entries (except as related by symmetry), and thus undirected Bernoulli link-percolation networks. Girko's K27 method handles symmetric random matrices with independent block submatrices (except as related by symmetry), and thus link-percolation networks with some localized correlations. Girko's K25 method handles potentially non-Hermitian random matrices with independent entries, and thus directed Bernoulli link-percolation networks. Each of these methods involves solving a system of nonlinear equations, the solution to which yields the approximate spectral distribution. Introduction of each result was supplemented by providing a detailed tutorial explaining the application of each method to approximate spectral densities for row-normalized network adjacency matrices. For each case, this included describing computational simplification conditions that reduce the system of equations. Node-transitivity of the random network model represents the most important of these. Each method was applied to an example random network model, with simulation visually demonstrating the quality of the approximation.

For shift-invariant filter design with respect to constant random networks, the graph shift matrix spectral asymptotics informed the filter design approach proposed in this thesis. Given a target filter response specification, the methods proposed design a filter to optimally approximate the target filter response on the support of the approximate spectral distribution. Use of the approximate spectral density support represents the critical intuition of this work (for constant networks), from which the results are derived. This enables filter design with respect to the random network distribution, such that the network nodes can be preprogrammed with the filter before deployment.

The particular graph filter application considered in this thesis is filter design for accelerated convergence of distributed average consensus state dynamics. Such filters produce more rapid or more accurate results in a variety of practical network agreement scenarios. For a constant network topology, the rate of convergence is governed by the largest eigenvalue modulus, apart from the consensus eigenvalue $\lambda = 1$. Therefore, the target graph filter response for consensus acceleration maps all eigenvalues to zero and filter optimality is defined in the minimax sense, resulting in a Chebyshev filter design problem. Related problems such at worst case graph total variation minimization and expected graph total variation minimization were approached by generalizing to formulations of the optimization problems with weighted filter response error.

Chapter 4 posed periodic shift-invariant graph filter design optimization problems for consensus acceleration (and closely related problems) informed by the spectral asymptotics of large-scale networks (both undirected and directed). Included simulation results for each case support the proposed filter design methods. The main contributions of this thesis that appear in Chapter 4 are as follows:

- For undirected networks, Section 4.2 posed a linear program (LP) to minimize the worst case filter response magnitude over the filtering region defined by the approximate spectral density (obtained by the methods in Section 3.2 or Section 3.3). Simulation results for several random network models demonstrated convergence rates nearly equal to the optimal filter designed with exactly known consensus iteration matrix. Furthermore, the improvement over results obtained for filters designed to minimize response at only the mean eigenvalues is substantial.

- For directed networks, Section 4.3 posed a quadratically constrained linear program (QCLP) to minimize the worst case filter response magnitude squared over the filtering region defined by the approximate spectral density (obtained by the methods in Section 3.4). Simulation results for several random network models demonstrated convergence rates nearly equal to the optimal filter designed with exactly known consensus iteration matrix. Furthermore, the improvement over results obtained for filters designed to minimize response at only the mean eigenvalues is substantial.

- Variants of the above optimization problems with weighted error minimization were posed in Section 4.4, including both a linear program (LP) for undirected networks and a quadratically constrained linear program (QCLP) for directed networks. This allows filter design for minimized graph total variation and expected total variation on large-scale random graphs. Simulation results for the undirected network case are provided and support the proposed method by demonstrating increased filter performance.

For filter design with respect to time-varying networks, many intuitions from shift-invariant graph signal processing break down. This scenario results in a sequence of graph shift operators with a sequence of eigendecompositions that should be used to perform graph signal processing analysis. For random networks, assumptions must be made regarding how the eigenvector bases relate to each other to enable analysis. To that end, this thesis examined switching networks, in which the network changes randomly to an independently drawn random network. The critical intuition of this work (for switching networks) consists of eigenvector assumptions proposed for switching networks and use of the approximate spectral density, specifically its moments. The particular graph filter application for switching again relates to accelerated consensus but differs slightly in objective. More precisely, the problem seeks to minimize the consensus error vector norm in expectation with respect to both the input consensus error vector and the network sequence.

Chapter 5 posed the graph filter design optimization problems for expected consensus error norm minimization (after a single filter application) informed by the spectral asymptotics of large-scale undirected switching networks . Included simulation results for each case support the proposed filter design methods. The main contributions of this thesis that appear in Chapter 5 are as follows:

- For large-scale switching undirected random networks, the norm squared of the consensus error was expressed as a quadratic objective function. The matrix in this objective function is the expected (with respect to the network sequence and input error vector) Gram matrix of consensus error in the filter terms. This led to posing a linearly constrained quadratic program (LCQP) for approximate expected consensus error norm minimization on switching networks (in both Section 5.2 and Section 5.3).

- An approximation of the expected Gram matrix of consensus error in each term was derived (Proposition 5.1 and Proposition 5.2) based on assumptions regarding the sequence of eigenvector bases and the moments of the approximate spectral distribution. This was done for unnormalized Laplacian-based weights $W(\mathcal{G}) = I - \alpha\mathcal{L}(\mathcal{G})$ in Section 5.2 using the simulated expected empirical spectral distribution and done for row-normalized Laplacian-based weights $W(\mathcal{G}) = I - \alpha \mathcal{L}_R(\mathcal{G})$ in Section 5.3.

In conclusion, this thesis examined graph signal processing on large-scale random networks, performing filter design optimization with respect to the spectral distribution asymptotics as obtained through random matrix theory results. For constant random network models, shift invariant graph filters for accelerated distributed average consensus convergence were designed to minimize the filter response over the support of the approximate spectral density. For switching random network models, graph filters were designed to minimize the expected consensus error using assumptions on the sequence of consensus iteration matrix eigenvectors and the approximate spectral distribution moments. Simulation results for each case demonstrate the utility of asymptotic spectral methods for graph signal processing on large-scale random networks. This suggests similar approaches would benefit other graph filtering applications on large-scale random networks.

6.2 Future Work

This thesis closes by presenting several possible directions for continuing research efforts on graph signal processing using the methods described in this thesis as a point of departure. The suggested topics can be divided into the following three groups: expansion of asymptotic spectral analysis of networks, further analysis of filter design for constant random networks, and further analysis of filter design for time-varying random networks. This section briefly discusses potential continuations within each category.

To apply the methods presented in this thesis to broader classes of large-scale random networks, further search random matrix theory literature for applicable methods. The three methods for empirical spectral distribution approximation from Girko that were employed in this thesis pertain to networks in which a directed link and the reverse of the link are either completely dependent (bidirectional) or completely independent. However, partial correlation between these links represents a well motivated but unhandled case. The conditions imposed by Girko's methods that pertain to matrices with such partial correlations are too severely limiting to allow application to network adjacency matrices. However, the existence of these nearly applicable theorems suggests extensive search through more recently published literature could reveal suitable methods. Additionally, Girko's methods apply to large-scale matrices, but the results do not specify convergence error bounds for the spectral approximation. Such information, if found, would important in more precisely quantifying the meaning of "large-scale" networks. Furthermore, it could potentially extend the practical relevance of the filter design methods in this thesis to networks of smaller size.

For large-scale constant random networks continuing efforts could focus on connecting spectral asymptotics with non-minimax filter design objectives. It is possible to combine ℓ_1, and ℓ_2 (least-squares), and ℓ_∞ (minimax) constraints and objectives in various ways to produce several different linear program (LP), linearly constrained quadratic program (LCQP), and quadratically constrained quadratic program (QCQP) formulations, some of which may be more suitable for given problems. For instance, filter response approximation with respect to least squares optimization has particular relevance to minimizing norm squared objectives in expectation with respect to random inputs. Finally, because the presented optimization methods for large-scale constant random networks can be used to approximate any specified target filter response, the methods in this thesis could find use in additional graph signal processing applications beyond consensus acceleration.

For large-scale time-varying random networks, the results provided in this thesis demonstrate that the improvement in consensus error minimization derived from the designed filters rapidly diminishes with increasing switching probability. However, network models characterized by more gradual change (compared to abrupt switches to an independently drawn network) may maintain greater alignment between eigenvector bases at consecutive network changes and produce more significant improvements using graph filters. For instance, percolation networks with links governed by independent Markov processes, which are well motivated in terms of link failure and recovery processes, may experience more gradual change and represents a potential opportunity for continuing research.

Bibliography

[1] D. Shuman, S. Narag, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs,” *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.

[2] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs,” *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, April 2013.

[3] A. Sandryhaila and J. M. F. Moura, “Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure,” *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, Sept. 2014.

[4] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs: Frequency analysis,” *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[5] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, “Signal recovery on graphs: Variation minimization,” *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4609–4624, Sept. 2015.

[6] E. Wigner, “On the distribution of the roots of certain symmetric matrices,” *The Annals of Mathematics*, vol. 67, no. 2, pp. 325–327, Mar. 1958.

[7] V. Marčhenko and L. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Mat. Sb.*, vol. 1, no. 4, pp. 507–536, 1967.

[8] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications.* Cambridge University Press, 2011.

[9] V. Girko, *Theory of Stochastic Canonical Equations.* Springer Science+Business Media, 2001, vol. 1-2.

[10] R. Olfati-Saber, A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 98, no. 7, pp. 1354–1355, June 2010.

[11] S. Kar and J. M. F. Moura, "Consensus+innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 99–109, May 2013.

[12] A. Loukas, A. Simonetto, and G. Leus, "Distributed autoregressive moving average graph filters," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1931–1935, Nov. 2015.

[13] A. Sandryhaila, S. Kar, and J. M. F. Moura, "Finite-time distributed consensus through graph filters," *Proceedings of ICAASP 2014*, pp. 1080–1084, May 2014.

[14] D. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev polynomial approximation for distributed signal processing," *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, June 2011.

[15] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Filtering random graph processes over random time-varying graphs," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4406 – 4421, May 2017.

[16] E. Kokiopoulou and P. Frossard, "Polynomial filtering for fast convergence in distributed consensus," *IEEE Transactions on Signal Processing*, vol. 57, pp. 342–354, Jan. 2009.

[17] E. Montijano, J. I. Montijano, and C. Sagues, "Chebyshev polynomials in distributed consensus applications," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 693–706, Feb. 2013.

[18] F. Gama and A. Ribeiro, "Weak law of large numbers for stationary graph processes," *2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017)*, pp. 4124–4128, March 2017.

[19] S. Sundaram and C. Hadjicostis, “Finite-time distributed consensus through graph filters,” *Proceedings of the 26th American Control Conference (ACC 2007)*, pp. 711–716, July 2007.

[20] S. Apers and A. Sarlette, “Accelerating consensus by spectral clustering and polynomial filters,” *IEEE Transactions on Control of Network Systems*, vol. 4, no. 3, pp. 544–554, Sept. 2016.

[21] S. Segarra, A. Marques, A. Ribeiro, “Optimal graph-filter design and applications to distributed linear network operators,” *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4117–4131, May 2017.

[22] S. Kruzick and J. M. F. Moura, “Spectral statistics of lattice graph percolations,” *Arxiv: https://arxiv.org/abs/1611.02655*, Sept. 2016.

[23] S. Kruzick and J. M. F. Moura, “Spectral statistics of lattice graph structured, non-uniform percolations,” *42nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017)*, pp. 5930–5934, Sept. 2016.

[24] S. Kruzick and J. M. F. Moura, “Optimal filter design for signal processing on random graphs: Accelerated consensus,” *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1258 – 1272, Mar. 2018.

[25] S. Kruzick and J. M. F. Moura, “Graph signal processing: Filter design and spectral statistics,” *2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2017)*, Dec. 2017.

[26] S. Kruzick and J. M. F. Moura, “Consensus state Gram matrix estimation for stochastic switching networks from spectral distribution moments,” *2017 Asilomar Conference on Signals, Systems, and Computers (ACSSC 2017)*, Oct. 2017.

[27] S. Kruzick and J. M. F. Moura, “Optimal filter design for consensus on random directed graphs,” *Accepted: 2018 IEEE Statistical Signal Proceessing Workshop, Arxiv: https://arxiv.org/pdf/1802.10152v1.pdf*, Feb. 2018.

[28] S. Kruzick and J. M. F. Moura, “Spectral statistics of directed networks with random link model transpose-asymmetry,” *Submitted: 2018 IEEE Date Science Workshop, Arxiv: https://arxiv.org/pdf/1802.10159v1.pdf*, Mar. 2018.

[29] B. Bollobás, *Random Graphs.* Cambridge University Press, 2001.

[30] G. Grimmett, *Percolation*, 2nd ed. Springer, 1999.

[31] X. Ding and T. Jiang, “Spectral distributions of adjacency and Laplacian matrices of random graphs,” *The Annals of Applied Probability*, vol. 20, no. 6, pp. 2086–2117, 2010.

[32] E. Gilbert, “Random graphs,” *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.

[33] K. Avrachenkov, L. Cottatellucci, and A. Kadavankandy, “Spectral properties of random matrices for stochastic block model,” *4th International Workshop on Physics-Inspired Paradigms in Wireless Communications and Networks*, pp. 537–544, May 2015.

[34] C. A. Tracey and H. Widom, “The distribution of the largest eigenvalue in the Gaussian ensembles,” *Calogero-Moser-Sutherland Models,* CRM Series in Mathematical Physics, vol. 4, pp. 461–472, Springer Science+Business Media, 2000.

[35] G. Cybenko, “Dynamic load balancing for distributed memory multiprocessors,” *Journal of Parallel and Distributed Computing*, vol. 7, no. 2, pp. 279–301, Oct. 1989.

[36] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, pp. 63–70, April 2005.

[37] R. Olfati-Saber, “Flocking for multi-agent dynamic systems: Algorithms and theory,” *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, Mar. 2006.

[38] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” *Systems and Control Letters*, vol. 53, pp. 65–78, Feb. 2004.

[39] S. Kar and J. M. F. Moura, "Topology for global average consensus," *40th Asilomar Conference on Signals, Systems and Computers (ACSSC 2006)*, May 2007.

[40] P. Lax, *Functional Analysis.* Wiley-Interscience, 2002.

[41] P. Lax, *Linear Algebra and Its Applications, second edition.* Wiley-Interscience, 2007.

[42] R. Horn and C. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[43] Z. Bai, "Methodologies in spectral analysis of large random matrices, a review," *Statistica Sinica*, vol. 9, no. 3, pp. 611–677, July 1999.

[44] A. M. Tulino and S. Verdú, *Foundations and Trends in Communications and Information Theory: Random Matrix Theory and Wireless Communications*, 2004, vol. 1, no. 1.

[45] R. Olfati-Saber, "Ultrafast consensus in small-world networks," *Proceedings of the 2005 American Control Conference (ACC 2005)*, pp. 2371–2378, June 2005.

[46] R. Pachón and L. N. Trefethen, "Barycentric-Remez algorithms for best polynomial approximation in the chebfun system," *Springer Science + Business Media*, pp. 1–21, Oct. 2009.

[47] L. Hogben, *Handbook of Linear Algebra.* Chapman & Hall/CRC, 2007.

[48] D. Lyubshin and S. Savchenko, "Cayley digraphs with normal adjacency matrices," *Discrete Mathematics*, vol. 309, no. 13, pp. 4343–4348, July 2009.

[49] Gregoire Allaire and Sidi Mahmoud Kaber, *Numerical Linear Algebra.* Springer, 2008.

[50] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[51] A. Edelman and Y. Wang, "Random matrix theory and its innovative applications," in *Advances in Applied Mathematics, Modeling, and Computational Science (Fields Institute Communications vol. 66).* Springer Science+Business Media, 2013, pp. 91–116.

[52] R. Laskar, “Eigenvalues of the adjacency matrix of cubic lattice graphs,” *Pacific Journal of Mathematics*, vol. 29, no. 3, pp. 623–629, July 1969.

[53] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.